Improving Customer Experience Management: A Dynamic Topic Modelling Approach

A thesis submitted for the degree of Doctor of Philosophy (PhD)

Mandis Khayati Gargari

School of Business and Management
Royal Holloway, University of London

October 2023

Abstract

Over time Customer Experience (CX) has become increasingly important – particularly in online services such as streaming. The current state-of-the-art on CX, its dimensions and how to measure them relies heavily on conventional methods such as surveys and focus groups, often using score-based metrics like the Net Promoter Score (NPS) and Customer Satisfaction Score (CSS). These methods, however, do not capture the potential advantages offered by emerging data science and Artificial Intelligence. This thesis addresses that gap by exploring CX through the novel lens of Dynamic Topic Modelling (DTM), using data from streaming services as the context of application. Using Design Science as the methodological framework, three iterations of practical work are developed. The first iteration combines Topic Modelling (TM) and Aspect Based Sentiment Analysis (ABSA) to derive understanding from a dataset related to film/TV content streaming. The second iteration introduces temporal dynamics to evolve the approach in a manner that allows analysis to be conducted over time, rather than as a static picture that fails to understand emerging trends. The final iteration explores the generalisation of the temporal approach by applying it to a dataset in a different domain (music streaming). As a key outcome, the research advances CX assessment methodologies to provide a refined understanding of CX by uncovering topics, evolving patterns and trends within given data on-demand. The outcome of the work here allows researchers and companies alike to identify emerging themes, concerns and opportunities that might not be immediately evident. For companies, this further enables them to address customer needs, make informed decisions and optimise their customer experience strategies.

To the loving memory of my late,
beloved mother (1967-2020)

Dedications

This work is dedicated to the loving memory of my late mother, who left this world in April 2020. Her passing during the early stages of my PhD studies left me with immense pain and an irreplaceable loss. I know she would have been immensely proud to see me reach this milestone.

I also dedicate this thesis first and foremost to my best friend, soulmate and loving husband Salman. He was my rock through this entire journey, providing unwavering love, patience and daily encouragement that inspired me to keep going even when the road got unbearably rocky. His steadfast belief in me gave me strength on days when I doubted myself. He celebrated every little victory along the way. He endured my stress-induced outbursts with understanding. His passion for seeing me succeed fuelled my motivation. I could not have completed this marathon without him by my side - he is my life partner in every sense.

To my sister Mahdis and her husband Azim - my lifelong companions and partner in crime who pushed me onwards with their own PhD passions and infectious laughter. Their spirits kept me marching forward.

To my father, whose pride in even my smallest accomplishments motivated me to make him smile through this achievement.

And to my second family, my in laws, Nasrin, Hamed and Soudabeh. Their love and caring nurtured me through the toughest parts of this endeavour.

This thesis belongs not just to me, but to all of those who lifted me up, believed in me and gave me the strength to fulfil my potential. My success is our shared success.

Acknowledgements

Words cannot fully capture my sincere appreciation and heartfelt gratitude to my primary supervisor, Professor Mark Lycett and co-supervisor Dr Asmat Monaghan. Their steadfast guidance and support have been invaluable beyond measure throughout my PhD journey.

Professor Lycett and Dr Monaghan's wisdom, insights and feedback have profoundly shaped my growth as an independent, creative thinker and researcher. With great kindness and generosity, they invested their time and care into my academic development. I am deeply thankful for the knowledge they imparted, the constructive criticism they gave to hone my skills and the encouragement they provided during challenging times.

I could not have wished for a better supervisory team. Both created a nurturing environment for me to discuss ideas openly, make mistakes, learn deeply and reach my full potential. Their mentorship has made an indelible, inspirational impact that I will carry with me always.

My academic accomplishments would not have been possible without Mark and Asmat empowering me to find my own path. I am incredibly grateful for their integral role in my success. My sincere hope is that I can take the wisdom they imparted to me and pay it forward.

Table of contents

LIST of TABLES

LIST of FIGURES

Chapter 1: Introduction

1.1 Research Background and Motivation

Customer experience (CX) has become a strategic priority across industries, with research showing CX leaders enjoy higher revenue growth compared to competitors (KPMG, 2018). However, optimising CX remain challenging due to:

1. **Multidimensional Nature of CX:** CX encompasses cognitive, emotional, behavioural and social elements that evolve during CJs (Lemon & Verhoef, 2016; Verhoef et al., 2009). Measuring this subjective and fluid CX using surveys, ethnography, biometrics and text analytics remains challenging (Poels & Dewitte, 2006).

2. **Proliferation of Digital Channels:** The growth of digital channels has led to numerous customer touchpoints, including web, mobile, social media, in-store technologies and IoT devices (Homburg et al., 2017; Verhoef et al., 2009). This complicates mapping and managing integrated omni-channel CX and hinders insights across touchpoints (Edelman & Singer, 2015).

3. **Data Overload:** The abundance of customer data generated from various digital and physical touchpoints strains human processing capabilities (Klaus, 2013). Analytics techniques are essential to analyse this massive, multi-channel data for strategic insights. However, fragmented infrastructure often isolates CX data in channel-specific silos (Lemon & Verhoef, 2016; Kumar et al., 2020).

The exponential growth of customer feedback data presents both a challenge and an opportunity for organisations. Traditional survey methods yield limited structured data, while sources like social media produce vast volumes of unstructured text data (Klaus, 2013). This overwhelms organisations, making it difficult to derive meaning from CX data and identify improvement opportunities. Moreover, the fragmentation of CX data across departments and touchpoints complicates analysis. Marketing may have survey data, service handling complaints and product development tracks social media, leading to a lack of integrated insights (Lemon & Verhoef, 2016). To address these challenges, advanced analytics techniques such as text mining, sentiment analysis and ML are valuable. Text mining extracts key themes from open-ended text data, sentiment analysis gauges emotional content and ML reveals hidden patterns in complex data (Kumar et al., 2020). However, there are gaps in implementing these techniques effectively, with existing studies focuses on simple descriptive statistics and isolated research (Poels & Dewitte, 2019).

CX on Streaming Entertainment: Gathering and analysing CX data is pivotal for streaming services to gain actionable insights into evolving user preferences and behaviours. This supports continuous optimisation of personalisation, features and Pain point. Here are keyways streaming brands leverage CX data:

Gathering Customer Insights through CX Data

Streaming services heavily leverage customer data and analytics to gain insights into audience preferences and behaviours to guide ongoing CX innovations and personalisation (Smith & McLaughlin, 2020). Both explicit feedback and implicit signals are utilised to understand evolving user needs. Explicit feedback via surveys, ratings scales, reviews etc. provides direct input from users on their priorities, likes/dislikes and areas for improvement (Balasubramanian et al., 2022). These self-reported satisfaction metrics offer unambiguous perspectives. Additionally, implicit behavioural data gathered through tracking content engagement, feature usage and other interactions offers supplementary signals into audience tendencies (Chen et al., 2018). Advanced analytics techniques identify behavioural micro-segments within the user base based on dimensions such as content genres, usage context, demographics and other attributes (Chow & Chang, 2020). These insights enable tailoring navigation, content discovery and recommendations to individual preferences (Smith & McLaughlin, 2020). Optimised algorithms leverage CX data to enhance personalisation accuracy over time. Controlled experimentation further informs experience optimisation (Rothman et al., 2020) - A/B tests evaluating alternative interface designs, page layouts, content presentation etc. quantify engagement. For instance, the position, framing and visual prominence of titles on the homepage significantly influence which content users select (Hutchinson et al., 2018). Additionally, CX data uncovers evolving needs and pain points to address proactively rather than reactively (Balasubramanian et al., 2022). For example, feedback may indicate consistent difficulties in content discovery, suggesting opportunities to innovate navigation, search and recommendations to alleviate frustrations (Chen et al., 2018).

Reacting to Shifting Needs through Continuous CX Optimisation

Continuous CX optimisation allows streaming brands to proactively react to changing user preferences and behaviours as the competitive landscape evolves (Smith & Thomas, 2021). For instance, enhancing personalised discovery and recommendations on mobile apps and TV platforms caters to increased cross-device usage (Balasubramanian et al., 2022). With content consumption fragmenting across multiple services, optimising CX takes priority to drive engagement over channel expansion and exclusives alone. Experimenting with alternative subscription plans and pricing models based on analysis of usage patterns and willingness-to-pay enables brands to react to economic shifts and evolving demand (Chen et al., 2018). Testing CX enhancements such as offline downloading for mobile, family sharing capabilities and binge friendly UX allows brands to address emerging needs identified through CX data (Chintalapati & Kamakoti, 2021). Proactively realigning CX innovations based on usage trends provides agility to respond to competitive disruptions (Smith & Thomas, 2021). For instance, enhancing social features addresses user needs for community and shared viewing as audiences' fragment. Continuously optimising CX is crucial for streaming brands to sustainably engage subscribers by reacting to changing preferences and ecosystem dynamics (Balasubramanian et al., 2022). Leveraging CX data signals enables evidence-based decisions to realign experiences with evolving audience needs.

Proactively Addressing CX Pain Points

Continuously monitoring (recurring) complaints, negative feedback and reported issues through CX data analytics helps identify key pain points experienced by different user segments (Smith et al., 2021). Rigorous root cause analysis of these issues then guides problem solving efforts and targeted service improvements to boost customer satisfaction. For instance, analysis may reveal consistent difficulties in content discovery and frustrations with navigation flows, guiding UX optimisations and innovations in search and recommendations features to alleviate these pain points (Balasubramanian et al., 2022). Enhancing explanations and transparency around personalisation algorithms also addresses distrust and scepticism towards overly customised experiences (Chen et al., 2018). proactively addressing identified CX pain points through data-driven root cause analysis minimises churn risks by reducing customer effort and fosters positive long-term associations (Smith et al., 2021). However, it requires balancing dissimilar needs across customer cohorts. Controlled CX experiments enable innovation while mitigating risks from changes (Rothman et al., 2020). Voice-of-customer programs further support continuous CX enhancements based on user perspectives (Gupta et al., 2022). leveraging analytics to uncover pain points combined with agile CX improvements driven by root cause analysis allows streaming brands to get ahead of issues and maintain high satisfaction levels amidst ecosystem changes (Smith et al., 2021; Balasubramanian et al., 2022).

1.2 Research Aim and Objectives

Given the above, the aim of this thesis is *to develop an automated analytics framework that extracts strategic customer intelligence from unstructured textual feedback data.*

In achieving this aim, the thesis addresses the following key theoretical gaps in the existing CX and CXM research:

1. Lack of holistic CX measurement frameworks (Gap 1 and 6).
2. Over-reliance on subjective self-reported data (Gap 2 and 7).
3. Inability to track dynamic CX changes over time (Gap 3 and 9).
4. Disconnected, fragmented CX data and insights (Gap 5 and 10).

To address these gaps, the research objectives are:

1. Implement Aspect-Based Sentiment Analysis (ABSA) to determine sentiment polarity towards key topics identified through Topic Modelling (TM).

2. Incorporate Dynamic Topic Modelling (DTM) to uncover changing customer preferences and issues over time based on how topics evolve in feedback data.
3. Demonstrate the utility of the integrated analytics framework on real-world customer feedback datasets from multiple domains

1.3 Research Approach

This research follows a Design Science Research (DSR) approach (Hevner et al., 2004). DSR focuses on designing novel artifacts to solve organisational and technical problems (March and Smith, 1995). Among the various DSR process approaches, this research adopts the DSR Methodology (DSRM) process (Peffers et al., 2007). DSRM provides a structured framework for identifying issues, designing and developing artifacts, demonstrating utility through evaluation and communicating findings (Gregor and Hevner, 2013). The artifacts aim to provide both theoretical advances and practical value (Baskerville et al., 2018). Below is a brief overview of the key steps in DSRM process and how they apply to this research:

- Problem Identification pinpoints the key research problem of extracting insights from CX data and motivates the need for better capabilities.
- Objectives Definition describes objectives to develop artifacts for strategic CX intelligence through TM, sentiment analysis and dynamic embeddings.
- Design and Development iteratively creates and refines artifacts to meet objectives following sound design science principles.
- Demonstration shows artifact utility for extracting strategic CX intelligence through evaluation on real-world datasets.
- Evaluation assesses artifact quality rigorously and effectiveness using metrics like coherence, accuracy, robustness.
- Communication illustrates artifact features, utility and contributions through publications and presentations. DSRM provides a structured yet flexible process for artifact design, iterative refinement, demonstration, evaluation and dissemination. Following DSRM will ensure the artifacts developed provide both theoretical advances and practical value for strategic CX intelligence extraction.

The research follows a rigorous, yet flexible approach aligned with DSRM principles to develop and evaluate artifacts through successive build-evaluate loops (Sein et al., 2011). The iterations aim to incrementally enrich the techniques, demonstrate utility and generalise across contexts. The following are the iterations of this study:

1. **Iteration 1**: Implemented TM and ABSA to extract integrated insights from Netflix customer feedback (TM+ ABSA).
2. **Iteration 2**: Incorporated Dynamic Topic Modelling (DTM) to enable temporal analysis of evolving topics and changing customer perceptions in the Netflix data over time.
3. **Iteration 3**: Generalised the integrated analytics techniques by applying them to a new domain, Spotify music streaming, to demonstrate adaptability across datasets (Validated generalisability across domains).

The artifacts integrate the follow to deliver impactful capabilities aligned with business needs. The following summarises the key aspects of the research approach:

- TM: LDA will be explored for unsupervised topic extraction from customer reviews (Blei et al., 2003). LDA hyperparameters including number of topics will be tuned and coherence optimised to obtain semantically meaningful topics.
- ABSA: Supervised learning methods such as Support Vector Machine (SVM) and neural networks will be leveraged to classify sentiment towards extracted topics and aspects (Schouten & Frasincar, 2016). Feature engineering of topics, syntax and context will enrich sentiment analysis.
- DTM: Contextual embedding models including BERT will be experimented with to capture topic evolutions and shifts over time (Devlin et al., 2018). Fine-tuning strategies such as supervised multi-task learning will be investigated.
- Evaluation: Topic coherence, sentiment accuracy, model robustness and domain adaptation ability will be quantified using statistical tests (Venable et al., 2016). Both intrinsic metrics and human evaluations will determine efficacy.

The focus is on developing modular, interpretable ML pipelines for strategic CX intelligence, emphasising usability, flexibility and transparency in artifact design.

1.4 Thesis Structure

This thesis is structured as seven chapters. A summary of each of the chapters is included below.

Chapter 2 reviews academic literature on CXM, tracing its evolution and discussing key concepts including the strategic significance of omnichannel personalisation. The review highlights complexities around quantifying subjective CX, opportunities and limitations of text analytics techniques and underexplored areas like contextual sentiment analysis and DTM that represent gaps needing future research.

Chapter 3 provides a detailed discussion on the research methodology followed for the thesis. The DSR paradigm is introduced as an approach focused on developing and evaluating innovative artifacts to address organisational problems. The DSRM proposed by Peffers et al. (2007) is presented as a structured framework for carrying out DSR through relevance and rigor cycles. DSRM provides key steps including problem identification, objectives definition, iterative artifact design and development, demonstration, evaluation and communication.

Chapter 4 describes and explains the foundational iteration of the design cycle. It synthesises knowledge of the problem from the literature review to create the first tentative design analysing CX data. The chapter discusses the acquisition, preparation and examination of the Netflix customer feedback dataset from Trustpilot. TM using LDA is applied to identify key themes. ABSA determines sentiment towards extracted topics. The chapter identifies and describes shortcomings including lack of semantic understanding in LDA and need for more Contextualised, temporal analysis. Finally, it outlines consequences of these limitations on remaining work. The chapter establishes an initial artifact demonstrating the utility of combining TM and ABSA for analysing unstructured CX data and provides a foundation for enhancements in subsequent design cycles.

Chapter 5 describes and explains the second iteration of the design cycle. It synthesises knowledge from the initial static modelling to create an enhanced DTM artifact. The chapter discusses the acquisition and preparation of the Netflix customer feedback dataset. contextualised embeddings are generated from the text data. The dataset is then temporally segmented to enable dynamic analysis across time periods. Dynamic topic models are developed using embeddings to uncover how topics evolve over time. The chapter goes on to identify and describe shortcomings of prevailing static topic models such as limited semantic understanding and inability to capture evolutions. Finally, it outlines the consequences of these limitations on the remainder of this work. The DTM addresses these gaps through innovations like BERT-based representations and time-aware analysis. The chapter represents a key step in the evolution of the study by transitioning from static to DTM.

Chapter 6 represents the final design cycle focused on assessing generalisation. It implements the techniques on the new Spotify dataset as the next domain after Netflix. Performance is benchmarked to identify needed adaptations. This design cycle systematically implements the techniques on the Spotify dataset and evaluates performance based on consistency criteria. The overarching goal is assessing the flexibility and extensibility of the artifact encompassing Contextualised BERT embeddings, DTM, temporal analysis and sentiment quantification to provide consistent utility across CX domains. This establishes the broad applicability of the artifacts beyond a single context. The application to Spotify reviews evaluates consistency in delivering value for CX analytics across distinct datasets.

Figure 1-1 Conceptualising the DSRM-Based Research Structure

Chapter 2: Review of Literature

2.1 Overview

Chapter 2 provides a comprehensive review of existing literature relevant to this research. Section 2.2 discusses CXM including its definition, origin and evolution. Section 2.3 covers CX in depth, including the definition, background, dimensions, measurement approaches, the role of technology, challenges, antecedents, consequences and the CJ concept. Section 2.4 reviews background, applications of Topic Modelling (TM) in CX, techniques such as LDA, comparisons between techniques and limitations for CX. Section 2.5 examines ABSA approaches for opinion mining. Section 2.6 explores DTM, its techniques, role in CX and advantages over static modelling. Section 2.7 identifies gaps in the literature around CX measurement, TM, ABSA and DTM.

2.2 Customer Experience (CX) Definition and Importance

The contemporary understanding of customer experience (CX) as an essential source of competitive advantage and revenue growth has its theoretical foundations in several pivotal marketing paradigm shifts over the past few decades.

The first seminal influence stems from the relationship marketing movement sparked by prominent scholars like Berry (1983) and Grönroos (1994) in the 1980s. Relationship marketing stressed the need to retain profitable customers over the long-term by shifting away from maximising discrete sales transactions towards building mutually beneficial, collaborative engagements grounded in trust and commitment (Berry, 2002). This planted the seeds for a customer-centric focus that recognised positive cumulative experiences as the basis for enduring partnerships beyond one-off purchases.

Additionally, the conception of service-dominant (S-D) logic by Vargo and Lusch (2004) was highly impactful in theorising that value derives from cumulative co-created experiences rather than discrete commodity exchanges. A key premise in S-D logic is that customers seek adaptive, personalised interactions matched to their unique, dynamically evolving contexts over time (Vargo & Lusch, 2008). This contrasted with the standardised uniformity of historical goods-dominant paradigms. S-D logic's emphasis on leveraging deep consumer insights to facilitate customised experiences makes it highly complementary to the foundations of CX as assessed across subjective, multidimensional touchpoint journeys (Lemon & Verhoef, 2016).

Moreover, the exponential increase in accessible customer data from digitalisation has granted companies extensive personalised profiling opportunities to align interactions to individual habits, behaviours and preferences (Kumar et al., 2019). As Kumar et al. (2019) examine in detail, the data abundance from proliferating digital touchpoints has armed companies with granular individual-level insights to fuel recommendation engines, tailored content and targeted promotions in a highly customised manner aligned to service-dominant principles.

Together, these paradigms stressing dynamic relationships beyond transactions, co-created adaptive experiences matching contextual needs and data-driven customisation underpin contemporary perspectives of CX as delivering consistent yet tailored omnichannel engagements across cognitive, emotional, sensorial, behavioural and social dimensions over the entire CJ lifecycle (Lemon & Verhoef, 2016; Kranzbühler et al., 2018).

2.2.1 Background

The rising prominence of CX has its origins in the paradigm shift from transactional marketing to relationship marketing in the 1980s. Early relationship marketing scholars such as Berry (1983), Grönroos (1994) and Parasuraman et al. (1988) stressed the importance of retaining profitable customers by shifting focus from maximising sales transactions to building long-term, mutually beneficial engagements. This laid the groundwork for putting the customer, rather than the product, at the centre of marketing strategy.

In the 1990s, the emergence of service-dominant logic proposed that customers seek experiences and relationships rather than discrete transactions (Vargo & Lusch, 2004). Service-dominant logic emphasised co-creating value with customers through ongoing interactions. Within this paradigm, the quality of CX determines value perception and impacts customer loyalty more than product features or benefits alone (Verhoef et al., 2009; Vargo & Lusch, 2004).

The rapid evolution of information technology dramatically expanded customers' interaction opportunities with brand s across multiple digital channels, platforms and touchpoints (Homburg et al., 2015; Verhoef et al., 2009). Proliferation of new CX touchpoints through web, mobile, social, in-store technologies made mapping and integrating omnichannel CX overly complex but essential (Lemon & Verhoef, 2016; Verhoef et al., 2015). Digitalisation led to exponential increases in customer data, enabling personalisation. Meanwhile, informed and networked customers became empowered with more choices and voices than ever before (Labai & Amann, 2021).

As markets matured and competition intensified globally, product or service superiority was no longer adequate for attracting and retaining customers (Kransbühler et al., 2018). Schmitt (2003) declared that CX represented the next competitive battleground as differentiated brand experiences created compelling value. This growing realisation led CX to be recognised as a new source of competitive advantage (Maklan & Klaus, 2011).

2.3 CX Dimensions and Their Measurement

The comprehension and enhancement of CX has become increasingly important for business success and customer well-being. CX involves the internal, subjective reactions and feelings of customers based on their interactions and touchpoints with a company (Meyer & Schwaiger, 2007). It is shaped by factors within the retailer's control (service, atmosphere, assortment, pricing) and external factors beyond their control (influence of others, shopping purpose) (Verhoef et al., 2009). CX is generally seen as having five underlying dimensions:

1. Sensory
2. Cognitive
3. Emotional
4. Behavioural
5. Relational (Schmitt, 1999)

To understand CX holistically, it must be studied at multiple levels. Schmitt (1999) identified five CX dimensions that are commonly used in research: cognitive, affective, behavioural, sensory, social and spiritual. While some researchers use additional or fewer dimensions, Schmitt's dimensions remain the core foundation. The key points are that CX is a subjective and shaped by multiple factors. It is a multi-dimensional concept that requires a holistic perspective to understand fully. The common CX dimensions outlined by Schmitt provide a useful research framework.

Measuring CX is important for gaining insights to guide experience-centric strategies. However, quantifying subjective, multidimensional and dynamic CX across different touchpoints is challenging compared to traditional metrics like satisfaction (Poels & Dewitte, 2006). Research has proposed various methods to measure CX, including surveys (Grisaffe, 2007), ethnography (Mariampolski, 2006), UX testing (Tullis & Albert, 2013), biometrics (Dupire et al., 2017) and neuroscience tools (Verhoef et al., 2009; Meyer & Schwager, 2007). For example, Grisaffe (2007) used online surveys to measure dimensions like satisfaction, loyalty and word-of-mouth for retail banking CX. Mariampolski (2006) applied in-depth ethnographic observation of consumers in electronics stores to pinpoint subtle unmet needs. Tullis & Albert (2013) conducted lab-based usability studies examining website shopping checkout processes to quantify CX. Dupire et al. (2017) analysed video footage using facial expression analysis software to measure emotional reactions during service encounters.

However, most studies have focused on conceptual models rather than measurement advancements (Poels & Dwitte, 2006). Many studies rely on self-report measures like satisfaction, purchase intent, loyalty and recommendations. While easy to implement, these fail to capture cognitive, emotional, behavioural and sensory aspects of experience over time (Paulhus & Vasire, 2007; Poels & Dwitte, 2006). Furthermore, some studies only focus on company-controlled aspects of CX, ignoring the customer's crucial role (Martin et al., 2015; Bridges & Florsheim, 2008; Rose et al., 2012; Novak et al., 2000).

In response, Verhoef et al. (2009) emphasised the need for robust CX metrics. Promising approaches involve complementing self-reports with implicit CX data from technologies like facial coding, EEG and analytics (Poels & Dewitte, 2019). For example, neuroscience tools can quantify emotional engagement, while NLP can analyse unstructured CX feedback (Klaus, 2013). Immersive technology also enables reconstructing and evaluating sensory CX through simulation (Flavián et al., 2019).

However, research on applying these emerging tools for CX measurement is still early. The emerging tools refer to novel approaches leveraging latest advances in sensory devices, neuroscience, immersive reality, emotion AI and other technologies that remain in maturation. In contrast with established methods like surveys and interviews, these cutting-edge tools offer potential breakthroughs in quantifying subjective CX elements but currently have limited adoption. Fundamental challenges remain regarding evaluating multidisciplinary measurement frameworks and quantifying subjective elements like emotions and their change over time. Based on Table 2-1, influential studies align with cognitive, emotional, behavioural, sensory and social dimensions of CX. Advancing CX measurement requires focusing on data aggregation and cross-context comparisons. Schmitt (1999) recommended that for experiential marketing, marketers must understand the relationship between the five experience dimensions, since they are interconnected. Therefore, the most profound effect requires planning across these dimensions jointly Given the extensive number of CX dimensions synthesised from the literature, the full table is presented in Appendix 1 for conciseness. However, Table 2-1 displays the first 10 samples from this comprehensive list to illustrate the key dimensions and seminal references. In the following sections, the assessment approach for each dimension is clarified.

No.	Author	Year	CX Measurement											
			CX Dimensions					CX Scales						
			Sensory (Sensing)	Affective (Feeling)	Cognitive (Thinking)	Physical (Acting)	Social (Relating)	SEMs	CEI	Brand Ex.	Gentile et al. (2007)	Pine and Gilmore (1998)	Other scales	context
1	Schmitt	1999	x	x	x	x	x	x						Retailing
2	Tsaur et al.	2006	x	x	x	x	x	x						
3	Nagasawa	2008	x	x	x	x	x	x						
4	Yuan and Wo	2008	x	x	x			x						Hospitality
5	Knutson et al.	2009							x					Hospitality
6	Brakus et al.	2009	x	x	x		x	x						
7	Jui-Wu and Liang	2009	x	x			x							Hospitality
8	Slattern et al.	2009	x	x			x	x				x	x	Theme Park
9	Sheu et al.	2009	x	x	x	x	x	x						
10	Chen and Hsieh	2010	x	x	x	x	x	x						Health Tourism

Table 2-1 CX Measurement Approaches and Their Dimensions

2.3.1 Cognitive Dimension

The cognitive dimension of CX encompasses intellectual processes like perception, reasoning and problem-solving (American Psychological Association, 2016). It aims to engage customers through surprise, creativity and learning (Schmitt, 1999; Holbrook 2000). Cognition is analysed from the lens of:

- Goal achievement, where customers pursue consumption objectives (Novak et al., 2003)
- Aligning actual experience with prior expectations (Gentile et al., 2007).

Cognition shapes multidimensional impacts on CX. Quantifying the subjective aspects of cognition poses challenges. It involves intangible thought processes that are:

- Context-dependent (Homburg et al., 2017).
- Vary across individuals (Lemon & Verhoef, 2016; Homburg et al., 2017).

Traditional measurement methods like surveys and neuroscience tools have limitations in:

- Capturing real-time cognition with its nuances (Poels & Dewitte, 2006; Verhoef et al., 2009).
- Overlooking individual differences (McColl-Kennedy et al., 2019).
- Lacking awareness of subconscious activities (Wang et al., 2018).
- Having restricted generalisability (Hwang & Seo, 2016; Gkatzia et al., 2021).

More comprehensive approaches are required that quantify the complexity of cognition within the CJ. Though difficult, better understanding the cognitive dimension will provide valuable insights for enhancing customer engagement. Table 2-2 presents an overall view of cognitive measurement approaches.

No	Measurement Method	Description	Limitations	Core Challenges Addressed	Reference
1	Surveys using rating scales	Self-reported perceptions of understandability, informativeness, novelty, etc related to CX touchpoints	• Depend on subjective recall • Biased responses	• Do not objectively capture contextual insights or unconscious cognition in real-time	Chang & Chieng, 2006
2	UX tests	Observing task completion, analyzing eye tracking for usability and cognitive load assessment	• Artificial lab settings • Reactivity effects	• Unable to recreate complex real-world contextual drivers of cognition	Ma et al , 2022
3	Text analytics	Analyzing feedback content for topics, themes, and sentiment to highlight issues in service interactions	• Relies on customer articulation in reviews • Limited by vocabulary used	• Does not quantify nuanced, subconscious cognitive processes and perceptions	Klaus, 2013
4	Neuroscience tools	Using EEG to assess cognitive load and memory encoding during CX journeys	• Expensive • Expertise needed for setup and analysis • Artificial lab environments	• Constrained ability to simulate dynamic real-world contextual factors influencing cognition across journey	Falk et al , 2012
5	Immersive simulation	VR based reconstruction of service situations to evaluate perceptions, learning, and emotions	• Resource intensive • Difficult to mimic full realism of CXs	• Fails to account for unique cognitive profiles and filters of individual customers	Flavián et al , 2019
6	Ethnography techniques	Using think aloud protocols to understand thought processes and gain customer perspective insights	• Time consuming; Resource intensive; Scalability issues	• Reliant on think aloud articulation • cannot measure tacit cognition	Gkatzia et al , 2021

Table 2-2 Cognitive Measurement Approaches

2.3.2 Emotional Dimension

The affective dimension refers to the moods and emotions generated during customer interactions (de Keyser et al., 2015). Emotions arise through cognitive appraisal and influence brand perceptions. The affective system combines mood and emotion as a "validated feeling state" (Cohen & Areni, 1991; Erevelles, 1998) and directs behaviours towards desired CX through emotions (Bagossi, 1992).

Emotions stem from appraising touchpoints as positive or negative (Lerner & Keltner, 2000). Different emotion valences have distinct effects. Positive emotions drive loyalty and advocacy (Savolainen, 2014), while negative emotions lead to dissatisfaction and disengagement. By eliciting positive emotions, retailers can build stronger relationships (Romani et al., 2012) through enhanced brand recall and preference (Bagdare & Jain, 2013) and avoiding hindered repurchase intentions (McColl-Kennedy et al., 2019). In summary, emotions link CX to behaviours and relationships, making them integral to CX success.

Measuring the affective dimension poses challenges due to:

- The transient, dynamic nature of emotions (Laros & Steenkamp, 2005);
- Context-dependence across journey stages (Lemon & Verhoef, 2016);
- Multi-dimensionality of emotion facets (Bagossi et al., 1999);
- Potential for mixed emotions (Fong et al., 2021);
- Individual variability in emotions (Larsen & Fredrickson, 1999);

- Customers may not fully express inner emotional state (Gountas et al., 2007).

Traditional methods like surveys and interviews struggle to quantify real-time emotions and contextual nuances (Poels & Dewitte, 2006; Lemon & Verhoef, 2016). More comprehensive techniques are needed to capture the complexity of customer emotions within CX. Understanding the affective dimension remains critical for enhancing CX design. Table 2-3 summarised the prominent emotional measure approaches and their challenges.

No	Method	Explanation	Limitations	Core Challenges Addressed	Reference
1	Netnography	Analysing online communities and social media to identify cultural trends, brand perceptions, emerging issues and influence patterns	• Relies on online data; • Biased by platform demographics and anonymity	Limited insight into real-time reactions and multidimensional properties	Kozinets, 2021
2	Social Network Analysis	Examining CX sharing and advocacy networks based on patterns of connections and interactions	• Provides usage data but limited emotional insights	Does not capture contextual drivers or unconscious processes	Kozinets, 2021
3	Conversational Analysis	Evaluating factors like rapport, empathy, and cooperation through the analysis of customer service calls and agent interactions	• Limited to service interactions dialogue • Not scalable	Does not quantify inner emotional state dynamics	Finch & Ariel, 2017
4	Surveys and Interviews	Gauging the influence of reference groups, opinion leaders, reviews, and brand associations on CX	• Rely on retrospective self-reports • Biased responses	Do not capture real-time emotions at sufficient depth	Chang & Chieng, 2006
5	Text Analytics	Discerning sentiments, experience themes, and influencer profiles from online reviews and social media	• Relies on customer articulation • Small emotion vocabulary	Does not assess unconscious processes or contextual nuances	Pagani & Malacarne, 2017
6	Ethnographic Observation	Assessing social behaviours, interactions, and cultural nuances both online and offline	• Resource intensive • Scaling limitations • Observer bias	Unable to quantify multifaceted emotion properties	Gorichanaz et al , 2018

Table 2-3 Affective (Emotional) Measurement Approaches

2.3.3 Sensorial Dimension

The sensorial dimension encompasses the range of perceptions derived through the five senses - sight, sound, touch, taste and smell - during customer interactions (Brakus et al., 2009). Sensory cues shape CX by creating vivid impressions and driving cognitive and affective reactions (Gentile et al., 2007). Key sensory facets include:

- Vision stimulated by aesthetics, colours, lighting, imagery across touchpoints (Hultén, 2011).
- Hearing perceptions triggered by sounds, music, sonic branding (Grayson & McNeill, 2009).
- Haptic feedback from tactile textures, materials, shapes (Peck & Childers, 2006).
- Olfactory cues like scents and aromas (Krishna, 2012).
- Taste sensations from flavours, foods, beverages (Gámbaro et al., 2012).

Multisensory combinations create immersive, holistic sensorial experiences (Spence et al., 2014). Sensorial elements establish emotional connections and aid memorability through sensory signatures (Brakus et al., 2009). Advances in immersive technologies are elevating remote sensorial CX (Flavián et al., 2019). Overall, orchestrating positive sensory stimuli enhances immersion, delight and memorability within CX.

However, quantifying sensorial dimension poses challenges due to:

- Transient nature of sensory reactions (Larsen & Fredrickson, 1999).
- Contextual nuances across touchpoints (Krishna, 2012).
- Multidimensionality of sensory modalities (Spence et al., 2014).
- Individual sensory preferences (Krishna et al., 2010).
- Limitations of subjective self-reports (Sweeney et al., 2015).
- Limitations of observations (Tullis & Albert, 2013).
- Lack of sensory context in behavioural data (Poels & Dewitte, 2006).
- More integrated approaches needed to capture nuanced sensory dynamics within CX journeys.

2.3.4 Behavioural Dimension

The behavioural dimension refers to the observable actions, processes and behaviours exhibited by customers across their journey (van Doorn et al., 2010). These tangible outputs manifest inner cognitive and emotional responses to cumulative CX (Lemon & Verhoef, 2016). Key behaviours include:

- Product/brand selection, purchase interactions (Amin et al., 2022).
- Usage in terms of consumption, exploration, variety (Calder et al., 2016).
- Providing feedback via reviews, social sharing (Harmeling et al., 2017).
- Referrals through recommendations (Jaakkola & Alexander, 2014).

Even subconscious habits reflect CX effects, like repurchases and revisiting (Dolan et al., 2016). Analysing behavioural patterns provides CX quality insights. Firms can shape behaviours across the CJ by enhancing CX. However, measurement poses challenges:

- Ethnography can be resource intensive (Arnould & Wallendorf, 1994).
- Surveys may suffer subjective biases (Sweeney et al., 2015).
- UX testing impacts natural behaviours (Tullis & Albert, 2013).
- Biometrics require infrastructure and expertise (Dupire et al., 2017).

- Transactional data misses certain behaviours (Verhoef et al., 2010).
- Video analytics struggles in complex settings (Meißner et al., 2021).

Blending methods like ethnography, surveys, biometrics and analytics may enable comprehensive CX generalised insights. Novel frameworks are needed to integrate subjective and objective data seamlessly. Capturing fluid customer behaviours across contexts holistically remains an evolving opportunity.

No	Method	Description and Application	Citation
1	Ethnography	• Involves observing behaviours such as in store navigation • Product usage, and service interactions in real world settings	Arnould & Wallendorf, 1994
2	Surveys, Interviews, Focus Groups	• Collects self-reported behavioural data through questionnaires • Interviews, and group discussions	Sweeney et al , 2015
3	UX Testing	• Evaluates behaviours during human computer interactions • Assessing webpage navigation patterns and task completion	Tullis & Albert, 2013
4	Biometrics and Activity Trackers	• Monitors behavioural signals like facial expressions, • Gestures and purchase actions using sensors and wearables	Dupire et al , 2017
5	Transactional Data Analytics	• Analyses behavioural patterns related to purchases, usage, referrals, claims and other transactional interactions	Verhoef et al , 2010
6	Video Analytics	• Digitally tracks behaviours using techniques like heat mapping and gaze tracking to analyse visual interactions	Meißner et al , 2021

Table 2-4 Behavioural Measurement Techniques

2.3.5 Social Dimension

The social dimension refers to the interactions and connections customers have with others like peers, groups and communities during their experience journey (Lemon & Verhoef, 2016). Social interactions across touchpoints like service encounters and online communities directly and indirectly shape CX (Gremler & Gwinner, 2008; Cambra-Fierro et al., 2020). The social environment provides value through information sharing, enjoyment and belonging (Nambisan & Baron, 2007). Positive social interactions drive satisfaction, loyalty and relationships, while negative experiences increase dissatisfaction (van Doorn et al., 2010; Grégoire et al., 2009). Factors like rapport, empathy and knowledge sharing underlie impactful social connections between customers, employees and brand communities (Rosenbaum et al., 2007). Firms are designing CX strategies to enable social bonding across channels. Overall, the social dimension has a strong influence on experiences and relationships, requiring careful measurement and optimisation.

Measuring the social dimension of CX involves unravelling the intricacies of interpersonal interactions and relationships that customers establish while engaging with a product, service, or brand (Lie et al., 2015). The social dimension encompasses peer recommendations, social influence, community engagement and other social aspects that have a significant impact on customers' perceptions and experiences (Lemon & Verhoef, 2016). Quantifying the nuances of the social dimension poses multiple interrelated challenges:

- Context Diversity: Customer interactions occur across diverse contexts - online, social media, in-person, etc. Social dynamics vary across these contexts, contributing to measurement complexity (Kosinets, 2021).
- Indirect Social Effects: Influence of reviews, word of mouth, trends on CX tends to be subtle. Direct measurement techniques like surveys have limitations in quantifying these nuanced, indirect social effects (Lemon & Verhoef, 2016).
- Dynamic Effects: Fast-changing nature of online platforms and social trends adds complexity in establishing consistent measurement frameworks (Peters & Kallmuenser, 2014).
- Behavioural Complexity: Assessing multidimensional social interactions during service encounters is difficult - rapport, empathy, cooperation, etc (Gremler & Gwinner, 2008).
- Unstructured Data: Deriving insights from messy, unstructured social data poses analytics challenges around volumes and scalability (Pagani & Malacarne, 2017).
- Observation Constraints: Ethnographic observation of social behaviours is resource intensive. Online observation raises privacy concerns (Langer & Beckman, 2005).

2.4 CX Measurement Challenges

The extensive review of literature on CX dimensions and measurements shed light on the core challenges. Subjectivity, multidimensionality, transience and self-report biases have all been identified as key challenges in measuring CX (Lemon & Verhoef, 2016; Laros & Steenkamp, 2005; Paulhus & Vazire, 2007).

- Subjectivity: CX is inherently subjective and personal to each customer. Quantifying subjective aspects like emotions and cognition is difficult.
- Multidimensionality: CX has multiple interrelated dimensions like cognitive, emotional, sensory, behavioural and social. Capturing the complex dynamics between these poses challenges.
- Context-dependence: CX varies across different touchpoints and contexts over time. Methods struggle to capture contextual nuances.

- Individual differences: Customers have diverse preferences, perspectives and reactions. Approaches overlook individual variability in CX.
- Transience: CX components like emotions and senses are dynamic and transient. Quantifying fleeting reactions is problematic.
- Unstructured data: CX generates unstructured feedback data that is hard to systematically analyse at scale.
- Observation constraints: Direct observation of behaviours and emotions can alter CX or require extensive resources.
- Self-report biases: Customers may lack awareness or only partially express their true CX. Surveys have subjective limitations.
- Technology limitations: Most measurement tools like biometrics and neuroscience remain early stage and restricted.
- Lack of integration: Holistic CX measurement requires aggregating multidimensional, multimodal data seamlessly.

Based on Table 2-5, major difficulty is the evolving nature of CX over time as touchpoints build on one another in a nonlinear fashion based on accumulated experiences (Homburg et al., 2017). Static measurement models struggle to capture this cumulative, fluctuating essence longitudinally. The subjective facets such as emotions and cognition also introduce complexities through individual variability, context-dependence and intangibility issues that impede objective measurement. In general, challenges related to CX measurement can be divided into five categories as outlined below:

1. **Measuring Dynamic Experiences**: CX evolves across CJ in a nonlinear, cumulative fashion based on accumulated experiences over time. Touchpoints shape one another. However, most measurement models are designed for static, discrete transactions. New frameworks are needed to measure fluid, dynamic CX longitudinally across channels (Homburg et al., 2017).
2. **fragmented Data and Insights**: The proliferation of touchpoints has expanded CX data sources. However, data tends to be scattered across organisational silos. This hamper gaining integrated insights into complete journeys. Tying data back to CX metrics is also challenging. Connecting fragmented information remains difficult (Lemon & Verhoef, 2016).
3. **Lack of Holistic Frameworks:** Despite expanding CX data, comprehensive measurement frameworks remain underdeveloped (Poels & Dewitte, 2019). Existing tools like surveys and biometrics provide fragmentary insights but fall short in holistic integration. A key gap is quantified contextual data tied to journey phases. Traditional transactional data lacks this richness - discrete purchases only capture snapshots versus interconnections across broader

experiences (Verhoef et al., 2009). Metrics like loyalty and ratings overlook evolving satisfaction and multidimensional context shaping choice (Van Doorn et al., 2010). Behavioural data alone cannot quantify fluid subjective CX perceptions across integrated touchpoints (Lemon & Verhoef, 2016). Bridging this demands techniques synthesising semantics, emotions and behaviours to map metadata onto events (Poels & Dewitte, 2006). Enhanced frameworks for contextualised CX integration are needed to overcome insights restricted by behavioural data in isolation. Tying diverse data to journey stages remains critical for complete measurement.

4. **Dependence on Self-Reporting**: CX measurement remains over reliant on subjective self-reported data such as surveys and interviews. Customers may lack awareness of subconscious emotions or thoughts. Recall and social desirability biases exist. Blending self-reports with unobtrusive observation data is an opportunity (Gountas et al., 2007).
5. **Calibrating Metrics and Actions**: It remains challenging to establish causal links between CX metrics and interventions and tie them back to business impact. Being able to accurately diagnose issues and predict optimal actions based on CX data intelligence is an evolving capability (Klaus & Maklan, 2013).

CX Measurement Challenges	Addressed by Traditional Methods	Remains a Challenge	Broader Challenge Category	Relevant Citations
Capturing subjectivity		✓	Lack of Holistic Frameworks	Poels & Dewitte, 2006
Multidimensionality		✓	Lack of Holistic Frameworks	Lemon & Verhoef, 2016
Context dependence		✓	Measuring Dynamic Experiences	Verhoef et al., 2009
Individual differences		✓	Lack of Holistic Frameworks	Homburg et al., 2017
Transient nature		✓	Measuring Dynamic Experiences	Laros & Steenkamp, 2005
Unstructured data		✓	Fragmented Data and Insights	Pagani & Malacarne, 2017
Observation constraints		✓	Dependence on Self-Reporting	Langer & Beckman, 2005
Self-report biases		✓	Dependence on Self-Reporting	Paulhus & Vazire, 2007
Lack of integration		✓	Lack of Holistic Frameworks	Verhoef et al., 2009
Quantifying cognition		✓	Lack of Holistic Frameworks	Gkatzia et al., 2021
Quantifying emotion		✓	Lack of Holistic Frameworks	Poels & Dewitte, 2006
Quantifying senses		✓	Lack of Holistic Frameworks	Krishna, 2012
Behavioral measurement	✓			van Doorn et al., 2010
Social measurement	✓			Nambisan & Baron, 2007
Satisfaction measurement	✓			Meyer & Schwaiger, 2007

Table 2-5 CX Measurement Challenges

2.5 Customer Experience Management (CXM)

In today's highly competitive and increasingly commoditised markets, delivering a standout CX has become a key imperative for companies seeking to differentiate themselves and foster enduring customer relationships (Kransbühler et al., 2018). Merely having a superior product or reasonable pricing is no longer enough to win and retain customers, who now have unlimited choices and information at their fingertips. This has led to the emergence and strategic importance of CXM as a holistic approach to engage customers on an emotional level at every touchpoint along their journey. Thus, CXM represents the evolution of marketing beyond simply selling products and services towards crafting end-to-end experiences tailored to customers' needs, preferences and priorities (Lemon & Verhoef, 2016). It signifies a fundamental philosophical shift - from transaction-focused selling to relationship and value-focused marketing centred around the customer. In essence, CXM puts the customer, not the product, at the heart of the business.

With proliferating digital channels, mobile connectivity, social media and asynchronous communication, today's consumers have become highly informed, empowered and demanding (Homburg et al., 2015). They expect personalised, seamless experiences across devices and touchpoints. This makes curating consistent yet tailored omnichannel experiences extraordinarily complex for companies, but also extremely critical as sub-par CX means lost customers. Therefore, optimising and innovating CX through an integrated CXM strategy has become an imperative to thrive in the digital economy.

CXM spending is estimated to reach $20.6 billion by 2027 as more companies realise its potential for boosting customer lifetime value, advocacy, share of wallet and sustainable competitive advantage. A study found that CX leaders grew revenues 4-8% above competitors (Forrester, 2020). This highlights the growing realisation that effectively managing and orchestrating great CX is now a ‘need to have’ rather than a ‘nice to have’ capability for business success and survival.

2.5.1 Definition

Lemon and Verhoef (2016) define CXM as "the process of strategically managing a customer's entire experience with a product or a company" (p. 71). CXM involves understanding the customer's perspective and journey to curate positive, meaningful experiences across diverse touchpoints and channels (Lemon &Verhoef, 2016). Based on a comprehensive review of the literature, Table 2-6 highlights the key elements of CXM. An expanded explanation of Lemon and Verhoef's CXM definition is provided below, which matches with particular elements in Table 2-1:

1. Holistic approach beyond individual touchpoints: Lemon and Verhoef focus on managing experiences throughout Customer Journey (CJ). This connects to Element 1 (CXM taking a holistic approach).
2. Understanding evolving customer needs and emotions: Lemon and Verhoef emphasise prioritising customer perspectives, needs and emotions. This matches Element 2 highlighting customer centricity as a key element of CXM.
3. Strategic experience design across the CJ: Lemon and Verhoef highlight experience design as a key CXM competence. This corresponds to Element 3, which describes CXM as purposefully building personalised, seamless experiences throughout the customer lifetime.
4. Omnichannel orchestration of interactions: Integration and optimisation of interactions across channels and touchpoints are mentioned by Lemon and Verhoef. This parallels Element 4’s point about omnichannel orchestration in CXM.

5. Value co-creation through customer engagement: Lemon and Verhoef discuss collaborating with customers to co-create value. Value co-creation is the process by which companies and customers work together to generate value. It views customers as active participants rather than just passive receivers of value (Vargo & Lusch, 2004). CXM facilitates value co-creation by providing platforms and avenues for customers to actively contribute ideas, content, reviews etc. that mutually benefit themselves and the company (Ranjan & Read, 2016). This echoes Element 5's description of value co-creation through customer engagement in CXM.
6. Continuous innovation based on customer feedback: Lemon and Verhoef highlight continuously monitoring customer expectations and innovating experiences based on feedback. This directly connects to Element 6 d escribing CXM requiring constant innovation based on customer inputs.
7. Cross-functional alignment: Lemon and Verhoef note CXM involves aligning people, processes and systems across functions to optimise experiences. This matches Element 7's point that CXM necessitates cross-functional alignment.

On the other hand, The underpinnings of CXM can be traced back to the relationship marketing paradigm that emerged in the 1980s from seminal works by prominent scholars like Berry (1983) and Grönroos (1994). Relationship marketing stressed retaining profitable customers by shifting from transactional to long-term, mutually collaborative engagements grounded in trust and commitment (Berry, 2002; Grönroos, 1994). This customer-centric focus aligns with the core of CXM philosophy geared towards optimising end-to-end consumer experiences beyond one-off transactions.

Furthermore, the conception of service-dominant (S-D) logic by Vargo and Lusch (2004) also holds strong synergy with CXM principles. S-D logic posits that customers seek experiences and relationships rather than discrete commoditised exchanges. Value is co-created through cumulative touchpoint interactions and assessments (Vargo & Lusch, 2004). This directly echoes the cumulative, subjective essence of CX. S-D logic also places priority on using customer data to facilitate personalised engagement aligned to dynamic needs (Vargo & Lusch, 2008). The common emphasis on discovering unmet needs through deep consumer understanding and crafting adaptive experiences makes S-D logic and CXM highly complementary paradigms.

No	Element	Description	References
1	Holistic focus	CXM takes a comprehensive approach, managing experiences throughout the entire customer journey	Homburg et al , 2017
2	Customer centricity	At its core, CXM prioritizes understanding customer perspectives, needs, and emotions	De Keyser et al , 2015
3	Experience design	CXM strategically designs personalized, seamless experiences for each stage of the customer lifecycle	Lemon & Verhoef, 2016
4	Omnichannel orchestration	CXM integrates and optimizes customer interactions across various channels, devices, and touchpoints	Homburg et al , 2015
5	Value co creation	CXM involves collaborating with customers to co create value through engagement in design and service	Galvagno & Dalli, 2014
6	Continuous innovation	CXM requires constant monitoring of customer expectations and innovating experiences based on feedback	Kranzbühler et al , 2018
7	Cross functional alignment	CXM necessitates alignment across various functions to optimize end to end experiences	Maklan & Klaus, 2011

Table 2-6 CXM Elements

2.5.2 Origin and Evolution

The origins of CXM can be traced back to the relationship marketing paradigm that emerged in the 1980s. Pioneering scholars such as Berry (1983) and Grönroos (1994) stressed the importance of retaining profitable customers by shifting from a transactional to a relationship-focused approach. This planted the seeds for what would later become CXM. In the 1990s, with the rise of customer-centric theories such as service-dominant logic (Vargo & Lusch, 2004), managing CX took on greater significance. Firms realised that experiences mattered more than products or services in building relationships and competitive advantage. The term “CXM” first appeared in marketing literature in the early 2000s (Palmer, 2010; Verhoef et al., 2009). Initially, CXM efforts cantered around optimising individual touchpoints and transactions that customers directly interacted with (Lemon & Verhoef, 2016). However, the transactional view has important limitations. CJ often span multiple different channels, devices and touchpoints across an extended period of time. The quality of each touchpoint interaction affects the overall CX (Lemon & Verhoef, 2016). Furthermore, with the proliferation of digital channels, mobile devices, social media and asynchronous communication, customers have become more empowered. Curating consistent omnichannel experiences matching rising expectations became imperative yet complex (Homburg et al., 2015). This led CXM to gain prominence as a strategy combining customer data analytics, omni-channel integration and customer culture to understand ever-evolving needs and curate personalised experiences (Kranzbühler et al., 2018). Specialised roles such as Chief Experience Officer also emerged demonstrating CXM's rising strategic significance.

Today, leading companies design integrated CXM ecosystems encompassing Voice of Customer (VOC) programs, CX metrics, journey mapping, experience prototyping and closed-loop learning (De Keyser et al., 2015). CXM has expanded to represent a customer-obsessed philosophy and set of capabilities for engaging consumers through digitally enabled, frictionless, tailored experiences that foster loyalty and growth. Scholars emphasise CXM's potential to reshape entire organisations around the customer (Homburg et al., 2015). As shown in Table 2-7, contemporary CXM involves an integrated set of strategies, programs and capabilities focused on knowing customers, designing engaging experiences and innovating relentlessly to optimise CX across the journey. It represents a key source of competitive advantage.

No	Key Characteristics	Description	References
1	Voice of Customer (VoC) Programs	Continuous gathering of insights into changing customer needs, pain points, and priorities Includes mining customer feedback, surveys, reviews, social media conversations and behavioural data	Qualtrics, 2022
2	Experience Prototyping	Design and test innovative experience concepts quickly based on customer data and ideation Lean, iterative approach for fail fast experimentation	Bulearca & Bulearca, 2010
3	Closed Loop Learning	Use CX metrics like NPS and CSAT to gauge experience quality Coupled with action plans for issue remedy and continuous improvement	De Keyser et al , 2015
4	Culture and Leadership	Focus on customer centricity and empowering employees for delivering personalised experiences	Walsh et al , 2012
5	Omnichannel Integration	Integration of people, processes and technologies for seamless experiences across touchpoints	Brynjolfsson et al , 2013
6	Journey Mapping	Visualise end to end customer experience Identify friction removal, delight introduction and communication improvement Includes physical and digital interactions	Stickdorn et al , 2018
7	Experience Governance	Define CX standards, guidelines, metrics, and accountability across the organisation to provides strategic direction	Peppers & Rogers, 2016
8	Agile Experience Optimisation	Use CX data and AI for real time experimentation, personalisation and tailoring of experiences	Edelman, 2021

Table 2-7 Key characteristics of New Era CXM

2.6 CX Antecedents and Consequences

The concept of CX antecedents refers to the underlying factors and conditions that influence the formation of a customer's perception and satisfaction prior to an interaction. Key antecedents include:

- Customer expectations - Formed before an interaction, these play a pivotal role in shaping satisfaction (Parasuraman, Zeithaml & Berry, 1985).
- Brand reputation - A favourable reputation predisposes customers to perceive interactions more positively (Bianchi et al., 2019).
- Service quality - The calibre of products, services and interactions affects the overall experience (Parasuraman, Zeithaml & Berry, 1985).

- Touchpoints - Each interaction contributes to the cumulative CX (Verhoef et al., 2009).
- Employee behaviour - Frontline staff can enhance or diminish the CX (Homburg, Wieseke & Hoyer, 2009).
- Cultural and social factors - These shape customer expectations and responses (Hofstede, 1980).

Understanding and managing these antecedents is important for organisations looking to orchestrate exceptional CX. By addressing these factors, businesses can positively shape customer perceptions, increase satisfaction, foster loyalty and cultivate a competitive advantage. A focus on CX antecedents allows firms to deliver memorable and meaningful experiences. The key antecedents highlighted in the literature are depicted in Table 2-8.

No	Antecedent	Description	References
1	Firm actions	Service quality, advertising, promotions, pricing, distribution strategies significantly impact CX	Kranzbühler et al , 2018; Maklan & Klaus, 2011
2	Employee interactions	Frontline staff behaviour and service quality is a critical antecedent in shaping CX	van Doorn et al , 2010; Bitner, 1990
3	Channel environments	Store atmospherics, web interfaces, mobile apps and contextual elements affect experiences	Verhoef et al , 2009
4	Brand related stimuli	Visual identity, communications, advertising and branding cues drive perceptions of CX	Brakus et al , 2009
5	Social environment	External stimuli like word of mouth, reviews, recommendations, and social media buzz influence CX	Lemon & Verhoef, 2016; Grewal et al , 2009
6	Customer traits	Individual differences in personality, cultural background, psychographics impact CX evaluations	Laros & Steenkamp, 2005; De Keyser et al , 2015
7	Consumption goals	Functional, hedonic, and social motivations behind consumption occasions moderate CX responses	Babin et al , 1994; Park & Moon, 2003
8	Perceived value	Assessments of benefits vs costs shape anticipated and experienced value which affects CX	Zeithaml, 1988; Sweeney & Soutar, 2001

Table 2-8 Antecedents of CX

The consequences of CX have far-reaching implications for both businesses and customers, shaping their interactions and long-term relationships. One of the primary outcomes of a positive CX is heightened customer satisfaction. When customers have seamless, enjoyable interactions with a brand, they are more likely to be satisfied with their experiences (Andreassen & Lindestad, 1998). This satisfaction, in turn, fosters customer loyalty, as satisfied customers are more inclined to remain loyal to a brand and engage in repeat purchases (Hennig-Thurau et al., 2004). Based on Table 2-9, positive CX generates a ripple effect through word-of-mouth and advocacy. Satisfied customers naturally become brand advocates, sharing their positive experiences with others and contributing to the brand 's reputation (Hennig-Thurau et al., 2004). This advocacy amplifies brand reach and serves as a powerful tool for customer acquisition. In addition to customer loyalty, CX plays a significant role in customer retention. When customers consistently have positive interactions with a brand, they are more likely to remain engaged and continue their relationship with the brand (Reichheld & Sasser, 1990). This contributes to reducing churn rates and enhancing long-term customer value.

These CX consequences - satisfaction, loyalty, advocacy and retention - feed back to strengthen antecedents like brand reputation and customer expectations in future interactions (Kandampully et al., 2015). A company known for delivering exceptional experiences benefits from positive brand equity and pre-established trust (Keller, 2001). Customers anticipate consistently positive experiences and are more forgiving of minor mishaps (Maxham & Netemeyer, 2002). The cyclical effects of positive CX allow companies to leverage past successes to continually improve CX going forward (Kranzbühler et al., 2018). Tracking satisfaction, loyalty and other metrics over time provides crucial insights to enhance key antecedents like service quality and touchpoint interactions (Neely, 2008). A focus on monitoring and evaluating CX consequences enables firms to nurture positive experiences in the long run (Zomerdijk & Voss, 2010).

NO	Consequence	Explanation	References
1	Customer Satisfaction and Loyalty	Positive CX enhances satisfaction, leading to loyalty, repeat purchases, and positive word of mouth	Andreassen & Lindestad, 1998
2	Word of Mouth and Advocacy	Satisfied customers become advocates, amplifying brand reach and aiding acquisition	Hennig Thurau et al , 2004
3	Customer Retention	CX reduces churn rates, encourages continued engagement with the brand	Reichheld & Sasser, 1990
4	Customer Acquisition	Exceptional CX attracts new customers through positive word of mouth	Lemon & Verhoef, 2016
5	Revenue Growth	Improved CX leads to increased revenue through repeat business	Verhoef et al , 2009
6	Brand Perception and Equity	CX shapes customer perception, enhancing brand equity and differentiation	Lemon & Verhoef, 2016
7	Reduced Costs	Effective CX lowers costs by reducing customer service needs and complaints	Gupta & Lehmann, 2005
8	Employee Engagement and Satisfaction	CX enhances employee engagement, motivating exceptional service	Homburg et al , 2015
9	Innovation	CX insights drive innovation, guiding new product and service development	Edelman, 2021
10	Competitive Advantage	CX differentiation offers a unique edge, fostering loyalty and standing out	Kotler et al , 2002

Table 2-9 Consequences of CX

Just as positive CX reinforces itself through a virtuous cycle, negative CX can trigger a vicious, downward spiral. Tracking CX metrics allows companies to rapidly detect dips in satisfaction and intervene before frustration sets in. Recovery efforts after service failures are essential to rebuild trust and loyalty (Miller et al., 2000). A proper understanding of CX consequences, both positive and negative, enables firms to nurture experiences that delight rather than disappoint customers.

2.7 Customer Journey

The Customer Journey represents the process a customer goes through in their interactions with a company during the purchase process (Lemon & Verhoef, 2016). It consists of a sequence of touchpoints across channels that customers traverse to learn, evaluate, purchase and engage with an offering. Each touchpoint interaction contributes towards shaping the overall CX. Edelman and Singer (2015) define CJ as "the customer’s end-to-end experience across all interactions with a company as they engage with its offerings and brand " (p. 90). It encompasses pre-purchase, purchase and post-purchase stages driven by goal-oriented customer motivations. Analysis of CJ uncovers pain points and opportunities to optimise CX across phases.

The CJ comprises various touchpoints which are moments of interaction between the customer and the brand (Lemon & Verhoef, 2016). Touchpoints may involve direct interactions, such as browsing a website, or indirect ones such as encountering an advertisement. Touchpoint typologies identified in literature include brand -owned, partner-owned, customer-owned and social or external touchpoints (Homburg et al., 2017). For instance, a brand -owned touchpoint would be the retail store experience, while an external one is user-generated content. Each touchpoint contributes to shaping customer perceptions.

The rapid evolution of digital technologies has dramatically increased CJ touchpoints across multiple channels and platforms (Verhoef et al., 2009; Brynjolfsson et al., 2013). Web, mobile apps, IoT devices, location-based services, digital kiosks and emerging technologies such as virtual/augmented reality provide new touchpoints for brand interactions (Huang & Rust, 2013). Consumers now experience touchpoints through computers, smartphones, wearables, smart appliances, interactive screens and new cross-channel modalities creating an "omni-channel" environment (Lemon & Verhoef, 2016). A study found average number of touchpoints during purchase increased more than 20% from 2008-2018 due to digitalisation, with over 60% involving digital channels (Edelman & Singer, 2015).

This rapid multiplication of touchpoints with digitalisation has made mapping and managing integrated CX across channels more complex. However, it has also expanded data collection opportunities to gain customer insights through analytics. Thoughtful touchpoint integration remains critical for consonant CX.

The CJ can be divided into three key phases - pre-purchase, purchase and post-purchase (Court et al., 2009). In the pre-purchase stage, key CX elements include trust-building to reduce perceived risk, clear communication of value proposition and conveying uniqueness to aid consideration (Chahal & Dutta, 2014). Providing meaningful content to assist information search and social proof elements such as reviews and testimonials are also important.

During purchase, crucial CX factors are service quality, convenience, channel options and frictionless payment (Meyer & Schwager, 2007). Purchase represents a "moment of truth" where experience failures can trigger abandonment. After purchase, delighting customers through surprise, crafting rituals during consumption and enabling social sharing for advocacy are impactful (Bolton et al., 2014).

Post-purchase, nurturing the relationship through sustained engagement, personalisation, building habits and managing issues with sensitivity is vital for loyalty (Kumar & Shah, 2004). Through each phase, the overarching focus must be crafting coherent narratives and eliminating inconsistencies across integrated touchpoints.

In essence, optimising CX requires aligning to evolving customer needs and priorities through the journey while proactively anticipating pain points. Technology expands options for this.

Analysing the CJ is a crucial undertaking for companies seeking to utterly understand their customers and identify specific pain points and opportunities to optimise CX across dissimilar stages (Edelman & Singer, 2015). It provides tangible insights that can guide targeted improvements to various touchpoints and interactions based on how customers traverse their buying and engagement cycle (Lemon & Verhoef, 2016).

In essence, CJ analysis illuminates areas where the current CX is failing to meet customer needs or expectations at specific steps along their path. These breakdowns likely lead to frustration, dissatisfaction and churn if left unaddressed (Holloway, 2019). By pinpointing where key CX issues exist across the journey, companies can allocate resources effectively to mend these gaps and provide the desired seamless, frictionless experience. In summary, CJ analysis is critical for identifying pain areas and opportunities to improve CX across stages. Common analysis approaches include in Table 2-10.

No	Approach	Description	References
1	Journey Mapping	Visualizes customer journey, highlighting touchpoints, channels, pain points and emotions Aligns CX to journey phases	Stickdorn et al , 2018
2	CX Metrics	Applies metrics like CSAT and NPS to quantify CJ performance, revealing problem areas at both aggregate and touchpoint levels	Klaus & Maklan, 2013
3	Data Integration	Gathers data from channels and touchpoints, integrating it into profiles and visualisations to reduce data silos	Lemon & Verhoef, 2016
4	Technology Enablers	Utilizes technologies like analytics, text mining, machine learning, and customer journey analytics platforms for granular insights	Kumar et al , 2013

Table 2-10 CJ Analysis Approaches

While various approaches such as mapping journeys and applying CX metrics offer value, integrating fragmented data into consolidated views is the crucial foundation that underpins accurate and holistic analysis of the CJ (Lemon & Verhoef, 2016; Edelman & Singer, 2015). Without integrated data, organisations operate in silos and lack complete pictures of how customers interact across channels (Brynjolfsson et al., 2013). Data integration through identifiers, CRM and pipelines creates unified profiles that preserve the complex sequential nature of journeys across touchpoints (Payne & Frow, 2006; Kumar et al., 2013). Key technology tools are listed in Table 2-11.

Technology enablers such as analytics, AI and CJ platforms are pivotal in processing massive integrated datasets to reveal behavioural patterns, diagnostics and opportunities once the data foundation is established (Edelman & Singer, 2015; Adobe, 2018). Manual analysis would be infeasible for the volume, variety and velocity of cross-channel CJ data (Kumar et al., 2020; Cambra-Fierro et al., 2020). In essence, integrated data creates the raw material while technology enablers supply the specialised analytical capabilities to convert this data into actionable CJ intelligence and insights. As shown in Table 2-11, these methods create visual journey maps, find correlations through attribution modelling and identify at-risk segments using ML algorithms (Lemon & Verhoef, 2016; Li & Kannan, 2014).

No	Technology Tools for CJ Analysis	Description	References
1	Web Analytics	Gathers digital body language data for online behaviours	Kumar et al , 2013
2	Customer Journey Analytics Platforms	Maps cross channel data to connect sources	Edelman & Singer, 2015
3	Voice of Customer Analysis	Derives insights from unstructured feedback	Cambria et al , 2013
4	Marketing Attribution Modelling	Quantifies touchpoint influence on outcomes	Li & Kannan, 2014
5	AI and Machine Learning	Uncovers trends in expansive CX data	Kumar et al , 2020
6	Digital Experience Optimization	Uses A/B testing for rapid digital optimization	Adobe, 2018

Table 2-11 CJ Analysis Utilising Technology Tool

Without the data integration and technology enablers, mapping techniques and CX metrics can only provide partial visibility despite their merits. Connecting insights across channels and touchpoints necessitates consolidated data flows and advanced analytics. Fusing these foundational capabilities with existing tools can augment the depth of CJ analysis significantly.

NO	Data Integration Approaches for Holistic CJ	Description	References
1	Identifier Mapping	Matches identities across sources for complete customer views	Lemon & Verhoef, 2016
2	CRM Databases	Centralised repositories for holistic CJ data	Payne & Frow, 2006
3	Data Ingestion and Processing Pipelines	Structures, cleanses, and connects CX data for analysis	
4	CJOrchestration Platforms	Synchronises data through APIs for real time integration	Salesforce, 2021

Table 2-12 Data Integration Methods Toward Holistic CJ

2.8 Role of Technology

Technology has long played a pivotal role in enabling CXM and CX measurement(Norton & Pine, 2013). As CX has grown as a key business priority, technology innovation has continuously attempted to address core challenges around data, analytics, integration and personalisation across the CX/CXM domains (Lemon & Verhoef, 2016).

In the early history of CX/CXM, technology played a basic role due to its limited capabilities at the time. For CX measurement, companies relied heavily on traditional surveys, in-depth interviews and focus groups to gather customer feedback (Grisaffe, 2007). These manual methods formed the foundation for assessing customer satisfaction, needs and pain points. For CXM, technology was predominantly used reactively via telephone-based customer service systems, which allowed managing customer issues and complaints as they arose (Watkinson, 2013). Agents could leverage basic databases to record interactions, but technology was not yet advanced enough for proactive CX personalisation or journey mapping. Overall, during this foundational period, technology was confined to elementary databases for collation and call centre systems for reactive engagement (Norton & Pine, 2013; Payne & Frow, 2006). The potential for technology driven CX insights, integration and prediction had not yet been unlocked. It was a markedly manual era of CX/CXM dependent on human collection of feedback and rudimentary desktop databases versus integrated digital systems (Grisaffe, 2007; Watkinson, 2013).

2.8.1 Limitations of Current Technological Approaches

Despite the increasing centrality of technology in CXM, there remain notable limitations and criticisms of current technology approaches for enabling integrated, proactive and personalised CXM. Table 2-13 summarises key existing technologies that have been applied to address major CX measurement and

CXM challenges and their limitations. While substantive progress has been made, Table 2-13 highlights how most existing technologies continue to only partially address key CX/CXM challenges. The table underscores the need for continued technological approaches.

- **Limitation 1**: Data silos across channels and touchpoints restricts integrated insights (Brynjolfsson et al., 2016).
- **Limitation 2**: Privacy concerns around data usage and profiling without transparency or control (Martin et al., 2017). Expanding surveillance of customer interactions raises concerns regarding consent, transparency and responsible usage of data (West, 2019) Consumers perceive trade-offs between personalisation quality and privacy erosion (Aguirre et al., 2015). Lack of visibility into profiling and analytics leaves users apprehensive about how data is utilised (Diakopoulos, 2016). Tensions exist between CX data's centrality and adequate ethical safeguards (Chandrashekeran & Keele, 2022). These privacy and ethical issues present obstacles to omnichannel personalisation and must be addressed through frameworks ensuring security, transparency and control. Developing tailored experiences from holistic data in an ethical manner remains an evolving capability needing focus. Solely amassing more CX data without responsible governance around access, analytics and transparency will further consumer wariness regarding utility versus ethical usage (Kumar et al., 2019). Maintaining trust while also providing integrated insights presents persistent technological and regulatory challenges.
- **Limitation 3**: Personalisation and customisation still limited despite abundance of data (Kumar et al., 2019).
- **Limitation 4**: Disconnected systems complicate seamless omni-channel experiences (Homburg et al., 2017).
- **Limitation 5**: Insights remain predominantly retrospective instead of predictive (Rust and Huang, 2014).

Given these limitations, achieving the ideal of integrated, ethical, personalised and predictive CXM remains an elusive goal, despite technology's centrality. These highlights pressing needs and opportunities for the next phase of CXM innovation.

Technological Approach	CX Measurement	CXM	Addressed CX/CXM Challenges	Shortcomings	Citations	Associated Limitation	Limitation Name	CX Measurement Challenges
Biometrics	✔		Capturing Subjectivity	High cost, lack of context	Poels & Dewitte, 2006; Rundle-Thiele et al., 2019			Capturing subjectivity
Data Mining	✔	✔	Multidimensional Data Complexity	Limited accuracy in sentiment analysis	Pagani & Malacarne, 2017	Limitation 1	Data Silos	Multidimensionality
Mobile/IoT	✔		Dynamic CX Tracking, Protecting Privacy	Obtrusiveness, privacy concerns	Lemon & Verhoef, 2016	Limitation 2	Privacy Concerns	Transient nature, Context dependence, Individual differences
AI Personalisation	✔		Capturing Subjectivity, Protecting Privacy	Lack of transparency, trust issues	Martin et al., 2017	Limitations 2, 3	Privacy Concerns, Personalization	Capturing subjectivity, Individual differences
CRM Systems		✔	Integrated CX Delivery	Persisting data silos	Homburg et al., 2017	Limitations 1, 4	Data Silos, Omnichannel Consistency	Lack of integration
Social Media Analytics		✔	Breaking Data Silos	Data access limitations	Brynjolfsson et al., 2016	Limitation 1	Data Silos	
CX Analytics Dashboards		✔	Linking Metrics to Actions	Reliance on surveys, lack real-time data	Klaus & Maklan, 2013	Limitation 5	Predictive Analytics	

Table 2-13 CX Measurement and CXM Technological Mapping

In conclusion, this table has mapped key CX and CXM technological approaches to the challenges they address, their limitations and the associated CX measurement challenges. It reveals important overlaps around subjectivity, data complexity, dynamic tracking, individual differences, integration and measurement frameworks. While current technologies have made progress in areas like biometrics, data mining, CRM and CX dashboards, fundamental limitations persist around privacy, personalisation, data unification, omnichannel orchestration and predictive insights. By highlighting these gaps, the table points to critical opportunities for the next wave of CX/CXM technological innovation to enable truly integrated, ethical, customised and forward-looking CX. Targeted advancements in capturing contextual subjectivity, managing data complexity, orchestrating omnichannel touchpoints and generating predictive insights will be pivotal.

2.8.2 CXM Technology Research Enhancement

Early applications indicate TM and DTM hold promise for addressing integration, personalisation and prediction challenges in CXM. TM techniques like LDA have shown value for integrating scattered customer data to enable holistic journey analysis and informed decisions (Cambra-Fierro et al., 2020; Kumar et al., 2019). TM revealed cross-channel purchase themes from emails and calls (Hu et al., 2019). ABSA provides granular CX insights by deriving nuanced attribute-level sentiment from text. This facilitates tailored recommendations and actions (Liu et al., 2021). ABSA improved churn predictions by analysing fine-grained brand perception in reviews (García-Moya et al., 2019). DTM techniques continuously monitor emerging text for new topics and trends. This enables proactive issue detection, like analysing complaint topic shifts on social media (Nussbaumer et al., 2012; Gu et al., 2018). TM and DTM have seen some preliminary applications in CX and CXM with promising outcomes:

- TM using LDA was applied to analyse customer emails and call transcripts to uncover cross-channel purchasing themes and improve CX personalisation. The approach led to more relevant recommendations and increased sales (Hu et al., 2019).
- LDA was used to extract key topics from airline customer reviews. The resulting insights on reviewed service attributes enabled more targeted service improvements (Chen et al., 2019).
- Dynamic LDA analysed shifting customer complaint topics on Twitter to proactively detect emerging airline service problems. The technique provided a 7-day head start on issue identification over traditional methods (Gu et al., 2018).
- A dynamic TM approach monitored hospital feedback topics over time to uncover trends in patient satisfaction with nursing services. This enabled timely service quality interventions (Bleier et al., 2019).

2.9 Topic Modelling

TM refers to a suite of NLP and ML techniques that extract main themes and topics from large collections of text documents in an unsupervised manner (Blei, 2012). Topic models have become popular for structurally analysing unstructured textual data across domains including extracting insights from CX data. TM provides an efficient approach to digest the key contents of large corpora of texts to reveal salient patterns.

2.9.1 Background

TM refers to a class of NLP algorithms that can automatically discover main themes and latent topics within a diverse collection of documents (Blei, 2012). TM provides a powerful approach for analysing unstructured textual data to extract key themes and patterns. Thematic analysis of text corpora has widespread applications across domains including analysing customer feedback data for CX insights. TM provides an efficient approach to digest the key contents of large corpora of texts to reveal salient patterns. Topic models are statistical models that distil large volumes of text into a predefined number of topics based on word co-occurrence patterns and distributions (Steyvers & Griffiths, 2007). Each topic consists of a cluster of highly related words that frequently occur together across documents. This allows inferring underlying semantic themes that tie groups of words together based on their contextual usage, revealing hidden thematic structures (Blei, 2012). The core premise behind TM algorithms is that documents cover multiple topics with different relative proportions, where each topic consists of a group of words that commonly co-occur (Steyvers & Griffiths, 2007). Through statistical modelling of word distributions and co-occurrence frequencies across documents, inter-related words are automatically clustered into coherent topics capturing semantic themes (Blei et al., 2003). Topic models use word usage patterns to reverse engineer the latent thematic structure without any prior training data or labels (Blei, 2012).

2.9.2 Topic Modelling Application for CXM (analysing customer feedback data)

In the context of CX analytics, TM provides an automated methodology to extract meaningful topics and themes from large volumes of unstructured textual feedback data such as open-ended survey responses, reviews, social media conversations and call centre logs (Cambria et al., 2013). Manually analysing such large and growing amounts of granular CX feedback is highly impractical. Applying TM on CX text data reveals crucial insights related to customer interests, priorities, concerns and pain points based on latent themes discovered from word usage patterns (Khan et al., 2020). Topic models can be applied on various sources of unstructured CX text data to digest the contents into a predefined number of topics based on statistical patterns in word usage and co-occurrences. For instance, TM can digest thousands of product reviews into the key features or aspects that dominate customer sentiment - both positive and negative. TM outputs can be leveraged by CX analysts for summarisation, feature extraction, clustering and other forms of insight derivation from sizable CX text datasets:

- *Product/service feature extraction*: Topic models can be trained on customer reviews of products and services to identify key features, components or attributes that dominate customer conversations. Both positive and negative feature associations are revealed based on contextual word groupings (Khan et al., 2020).
- *Pain point identification*: Applying topic models on customer complaint data sources such as tweets, reviews and survey comments highlights common issues, concerns and negatives experiences customers highlight. This provides insight into pain points.
- *Trend analysis*: Temporal TM on time-stamped feedback data identifies trends in customer priorities and interests based on topic volumes over time. This indicates changing needs.
- *Topic clustering*: Groupings documents by their highest probability topics creates a clustering effect to analyse different topic themes in granular detail based on document similarities.
- *Topic correlation*: Statistical TM enable assessing correlations and interlinkages between different topics based on co-occurrence patterns, revealing relationships.
- *Sentiment analysis:* Combining TM outputs with sentiment analysis techniques enables classifying emotion polarity towards key topics and aspects.

The main advantage of TM is the ability to digest large volumes of granular unstructured text data into a manageable number of coherent topics for high-level CX insights and summary analysis. However, semantic complexities exist, hence human-in-the-loop approaches add value. Overall, TM provides an efficient way to gain CX insights from sizable text data.

Latent Dirichlet Allocation (LDA)

Common TM techniques include LDA, Non-negative Matrix Factorisation (NMF), temporal topic models and neural topic models such as Embedded Topic Model (Blei et al., 2003; Dieng et al., 2020). Each has relative strengths and weaknesses.

LDA is a hierarchical Bayesian TM technique introduced by Blei et al. in 2003. LDA became exceedingly popular due to its conceptual simplicity and scalability for discovering latent topics in text. The key assumption in LDA is that documents cover multiple topics with different proportions. A topic is defined as a probability distribution over words. For example, a customer review may discuss "food quality" and "service" topics using related words with high probabilities unique to that topic. LDA represents documents as mixtures of topics and topics as word distributions. It uses Bayesian inference and Dirichlet priors to estimate topic-document and topic-word distributions based on word co-occurrence statistics across the corpus. Frequent co-occurrences provide signals to probabilistically assign words to common topics. A core assumption is exchangeability of word order within documents. LDA outputs matrices showing the compositions of topics per document and words per topic to quantify key themes. Since its introduction, LDA has been enhanced to model correlations, incorporate semantics and enable temporal analysis. It has been widely adopted for unsupervised TM in text mining, NLP and other domains.

While LDA offer promise for CX analytics, the direct application of LDA in the core academic discipline of CXM remains limited. Though LDA is described as commonly used for analysing customer surveys, reviews and other feedback, specific examples and citations of LDA implementations on CX data are scarcely found in the literature. Burns (2011) introduced a supervised approach to LDA to improve topic relevance, indicating shortcomings in its unsupervised application. Additionally, Farkhod (2021) combined LDA with sentiment analysis to identify polarity of topics. But applications like Sutherland (2020) who used LDA to analyse Airbnb guest experiences remain relatively rare in the literature. Overall, there is opportunity for deeper research and adoption of LDA-based TM within core CXM research.

Non-negative Matrix Factorisation (NMF)

Non-negative Matrix Factorisation (NMF) refers to a group of algorithms that factorise high-dimensional data into latent semantic features or topics through matrix decomposition into non-negative factors. The non-negativity constraint results in a parts-based representation where words additively combine to represent topics (Lee & Seung, 1999).

The inductive nature of NMF topics contrasts probabilistic topic models like LDA which use a top-down generative process and Dirichlet priors (Blei et al., 2003). NMF utilises optimisation techniques to minimise reconstruction error as it identifies topics through additive linear word combinations representing abstract concepts (Lee & Seung, 1999). NMF aims to automatically discover inherent structures within data. It originated for decomposition of images and signals and was later adapted for unsupervised TM and feature learning from text as an alternative to LDA (Xu et al., 2003). NMF provides an inductive, parts-based representation of topics through additive word combinations representing concepts (Huang et al., 2014). Enhancements like sparseness constraints and regularised NMF have improved topic quality.

While NMF has seen wider adoption in areas include information retrieval, recommendation systems, its direct applications in CXM remain more limited compared to LDA. However, some examples of NMF applied on CX data include:

- Used NMF for extractive summarisation and TM of customer reviews of hotels and electronics products (Wang et al., 2011).
- Applied NMF sentiment analysis to identify positive and negative topics in airline customer tweets to analyse service quality and complaints (Mejova et al., 2018).
- Combined NMF with K-means clustering to analyse customer satisfaction survey data and segment respondents based on their feedback topics (hang et al., 2019).

Given that NMF's inductive parts-based method has useful applications in mentioned domains, it has also resulted in various limitations and disadvantages for TM, such as issues with semantic coherence, instability and parameter adjustment. The additive word combinations used by NMF for inductive TM can result in less semantically interpretable topics. Furthermore, the optimisation method introduces heterogeneity among runs. In order to extract meaningful topics, extensive parameter tuning is required.

Neural Topic Models

Recent research has proposed various neural network architectures for TM, aiming to overcome limitations of conventional approaches like LDA that rely solely on bag-of-words co-occurrence statistics (Pandey et al., 2017; Grootendorst, 2022; Wang et al., 2020; Isonuma et al., 2020). By integrating techniques like convolutional and recurrent neural networks, attention mechanisms and graph-structured networks, these models aim to learn improved distributed representations of words and documents that encode semantic relationships. For example, Pandey et al. (2017) introduced an efficient neural topic model utilising convolutional feature composition and SoftMax normalisation for joint text and image modelling. Grootendorst (2022) presented BERTopic leveraging transformer language models to generate document embeddings for coherent topic clustering. Wang et al. (2020) proposed a supervised neural model with document-specific attention to improve prediction performance. Isonuma et al. (2020) used doubly recurrent networks for learning tree-structured topic hierarchies.

Recent research has started to explore applications of neural topic models in analysing distinct aspects of CX data:

- Combining DTM and net promoter score (NPS) to understand customer feedback topics and trends over time in the hotel industry (Nguyen & Ho, 2023).
- Appling Embedded Dirichlet Process (EDP) and Embedded Hierarchical Dirichlet Process (EHDP) models to extract customer insights about the tire industry from heterogeneous data sources (Palencia-Olivar, 2023).
- Investigating the impacts of CX attributes on overall ratings using neural networks for analysing customer reviews (Sato, 2019).
- Comparing different TM methods including neural models for analysing call centre conversation transcriptions (Habib et al., 2021).

Together, these studies demonstrate the usefulness of neural topic models in uncovering hidden information, identifying trends and extracting insights from customer feedback data in various industries. The contextual semantic modelling capabilities of neural topic models offer advantages over traditional methods. However, applications in core CX literature remain emergent.

2.9.3 Topic Modelling Techniques Comparison

LDA has become the leading technique for TM in CXM due to its unique probabilistic modelling framework. Table 2-14 summarises the technical strengths and limitations of three TM approaches developed up to this point. Unlike other text analysis methods, LDA is designed to uncover the latent topics and themes contained within unstructured textual data like customer surveys, product reviews, social media conversations and call centre transcripts (Blei et al., 2003; Sutherland, 2020). Specifically, LDA leverages word co-occurrence patterns and relationships using a hierarchical Bayesian architecture to reliably discover key topics even from massive volumes of customer feedback dataThe probabilities and correlations between words provide signals that LDA uses to extract actionable insights around pain points, product features, service issues and other themes without needing any prior labelling or supervision.

Additionally, LDA is highly scalable. By employing efficient sampling-based posterior inference algorithms, LDA can analyse corpora with millions of documents and vocabulary terms, making it well-suited for real-world CX data (Blei et al., 2003). The hierarchical structure also facilitates parallelisation across multiple compute cores.

Compared to NMF and emerging neural topic models, LDA excels at producing semantically coherent topics in a consistent manner with minimal parameter tuning required (Mejova et al., 2018; Toshikuni, 2019). It provides intuitive outputs that human analysts can readily interpret. This drives widespread adoption of LDA across CX research and applications.

In conclusion, LDA's maturity, probabilistic foundations, scalability and proven success for CX analytics cement its status as the leading technique for TM in CXM today. For organisations looking to gain strategic insights from customer feedback, LDA provides an indispensable tool for automated text analysis.

Topic Modelling Technique	Strengths	Limitations
Latent Dirichlet Allocation (LDA)	Conceptual simplicity and ease of implementation (Blei et al , 2003)	Lack of modelling semantic relationships between words (Dieng et al , 2020)
	Scalability to large datasets with efficient inference algorithms (Hoffman et al , 2010)	Bag of words assumption ignores word ordering (Blei & Lafferty, 2006)
	Handles polysemy through probabilistic modelling (Blei & Lafferty, 2006)	Hyperparameter tuning required for coherence (Arun et al , 2010)
	Wide adoption across domains enables benchmarking (Lee & Song, 2010)	Black box approach reduces interpretability (Lee & Song, 2010)
Non negative Matrix Factorization (NMF)	Inductive and interpretable parts based representations (Lee & Seung, 1999)	Lacks a probabilistic framework unlike LDA (Huang et al , 2014)
	Does not require specification of number of topics (Wang et al , 2013)	Prone to instability and variability across runs (Greene & Cunningham, 2009)
	Enables polysemy through additive word combinations (Wang et al , 2013)	Does not leverage semantic word relationships (Dieng et al , 2020)
		Struggles with large and noisy datasets (Wang et al , 2013)
Neural Topic Models	Learns semantic relationships between words based on embeddings (Dieng et al , 2020)	Complex architectures increase computational overhead (Dieng et al , 2020)
	Overcomes bag of words limitations of LDA and NMF (Dieng et al , 2020)	Require large external corpora to learn accurate embeddings (Card et al , 2018)
	Enables greater contextual understanding through neural networks (Zhang et al , 2018)	Lack of transparency due to black box neural networks (Lee & Song, 2010)
	Flexible integration of metadata like authors, timestamps (Zhang et al , 2018)	Still an emerging research area lacking benchmarks (Card et al , 2018)

Table 2-14 Comparison of Three Topic Modelling Techniques

2.10 Dynamic Topic Modelling (DTM)

Dynamic TM is an advancement over conventional static TM techniques like LDA that enables tracking how topics evolve over time by incorporating temporal information such as document time stamps (Wang & McCallum, 2006). By modelling changes in topic frequencies and distributions across successive time periods, DTM allows for uncovering rising and fading topics, detecting shifts in priorities, performing event tracking and forecasting future interests based on topic trajectories over time (Blei & Lafferty, 2006). This makes DTM a valuable technique for gaining temporal insights that can guide CX strategy.

DTM overcomes the limitations of conventional static TM techniques like LDA which provide only a snapshot versus a dynamic view of topics (Blei & Lafferty, 2006). DTM reveals valuable insights about evolving trends, fluctuations and events that static models miss, providing actionable intelligence for marketers and analysts seeking to understand customers' interests over time (Paul & Girju, 2010). Furthermore, DTM can be applied to social media posts, news articles, reviews and other textual data sources to enable important temporal analytics that would not be possible using conventional static TM approaches (Diao et al., 2012).

2.10.1 Dynamic Topic Modelling in CXM

DTM has become an increasingly useful technique for gaining customer insights and improving CX over time (Wang & McCallum, 2006). By discovering themes and trends in unstructured customer data such as product reviews, support tickets, or social media, companies can identify pain points and shifts in the CJ. This enables more proactive improvements to CX. This temporal analytics provides intelligence to guide agile CX strategy in several keyways:

- Reveals emerging interests and preferences through rising topic tracking on social media, reviews, forums (Blei & Lafferty, 2006). This identifies opportunities to delight customers.
- Detects waning interest in features or offers by analysing fading topics in voice of customer data (Wang & McCallum, 2006). Resources can shift away from outdated priorities.
- Monitors fluctuating priorities and shifting CJ based on changes in topic distributions (Paul & Girju, 2010). Strategies can dynamically realign.
- Analyses spikes in complaints, negative sentiment to uncover pain points and events impacting experience (Diao et al., 2012). Resources are quickly deployed to address issues.
- Forecasts future interests and expectations by extrapolating topic trajectories (Yan et al., 2013). Proactive planning ahead of shifting preferences improves CX.
- Enables agile resource planning and experience optimisation based on continuously evolving voice of customer signals, not just static snapshots (Blei & Lafferty, 2006).
- Provides a longitudinal view of the CX lifecycle, not just isolated touchpoints (Wang & McCallum, 2006). Holistic CX improvement across the entire journey.

DTM delivers crucial temporal analytics for optimising CX in an agile, data-driven manner. It empowers organisations to listen, understand and respond to the evolving customer voice through continuously updated dynamic intelligence. This drives better informed CX decisions. There are also limits to how well automated TM can truly capture the nuances in CX without losing critical context.

Thus, human expertise is essential for thoughtful interpretation and action based on the identified topics. Overall, though, responsible use of BERTopic-style DTM is a promising approach for continuously monitoring and enhancing CX (Mikalef et al., 2022).

2.10.2 Dynamic Topic Modelling Techniques

The Dynamic LDA model proposed by Blei and Lafferty (2006) aims to capture the evolution of topics over time in a document collection. It assumes topics evolve gradually over time according to a Markov process, where the topic distribution at time it depends only on the topic distribution at time t-1. Specifically, Dynamic LDA places a Markov chain prior on the latent topic proportions - this models the drift in popularity of topics over adjacent time slices. Intuitively, this allows topics to smoothly evolve, merge, split etc over time while preserving coherence.

Wang and McCallum (2006) take a different approach with Topics over Time to trace topic trends. This model analyses differences in the empirical topic proportions between time slices to detect meaningful changes in topics. The key insight is that by aggregating data from time slices, one can separate meaningful topic changes from noise. Thus, this technique can identify when new topics emerge, topics merge/split, or the semantics of a topic drift over time. A benefit of Topics over Time is interpretability - trends can be visualised via the rise and fall in probabilities of salient words in topics.

BERTopic leverages recent advances in neural networks and contextual word embeddings for more coherent DTM. Specifically, it uses BERT (Bidirectional Encoder Representations from Transformers) to create dense vector representations for words in a corpus. BERT is pretrained on large datasets to learn contextual relations between words. BERTopic then clusters these word vectors to identify topics. The advantage is that BERT provides more meaningful word representations compared to count-based methods such as LDA. This allows BERTopic to discover more coherent topics that smoothly evolve over time.

Dynamic Embeddings for TM by Dieng et al. (2020) also integrates neural word vector inputs but with a different approach. This model learns time-specific word embeddings using a neural network architecture. The word vectors are then aligned to a common embedding space across time periods. These dynamic embeddings better capture changes in word meanings and semantics over time. They are incorporated as inputs to TM to improve topic quality. The authors show this approach outperforms LDA and static word embeddings for DTM. AS illustrated in Table 2-15 Neural techniques such as BERTopic and Dynamic Embeddings show promise for DTM. Pretrained contextual word vectors from BERT enable more coherent topic clustering and evolution. Learned dynamic embeddings help capture changes in word meanings over time for better model inputs. Both methods demonstrate the utility of neural networks and word vectors for enhancing DTM (Dieng et al., 2020).

Technique	Advantages	Challenges
Dynamic LDA	Model's topic evolution via Markov chains	Computationally expensive to train
	Maintains topic coherence over time	Sensitive to model hyperparameters
Topics over Time	Interpretable visualization of topic trends	Topic changes can be noisy
	Identifies semantic drift over time	Limited by distinct time slice aggregates
BERTopic	Leverages BERT for coherent, evolving topics	Can be memory intensive
	Easy to use and visualize	Some instability in topic modelling
Dynamic Embeddings	Learns time specific word vectors	Complex neural architecture
	Improves topic quality over time	Difficult to interpret learned embeddings

Table 2-15 Advantages and Challenges of DTM Approaches

Of the DTM techniques discussed, the BERTopic model appears to be the most suitable for gaining customer insights from text data. BERTopic utilises BERT (Bidirectional Encoder Representations from Transformers), which is pre-trained on massive text corpora, to generate high-quality word embeddings reflecting semantic meaning and context. This allows for more coherent identification of topics and customer segments compared to traditional count-based models such as LDA (Blei & Lafferty, 2006). Additionally, BERTopic is relatively straightforward to use and interpret - the extracted topics and salient terms provide very understandable insights into customer needs and perceptions.

2.11 Overview of Sentiment Analysis for Opinion Mining

Sentiment analysis, also referred to as opinion mining, is a fast-growing field within NLP and text analytics focused on systematically identifying, extracting, quantifying and studying subjective information, opinions, emotions, appraisals, attitudes and sentiments expressed in textual data (Liu, 2012; Medhat et al., 2014).

The goal of sentiment analysis is to algorithmically determine the dominant affective tone or emotional orientation of a text - whether it conveys opinions that are positive, negative or neutral (Pang & Lee, 2008). This provides the capability to computationally analyse subjective language and quantify sentiments at scale across documents.

Sentiment analysis encompasses a wide range of more granular tasks including (Cambria et al., 2017; Shang et al., 2018):

- Document-level sentiment classification - determining overall sentiment polarity in a document.
- Sentence-level sentiment analysis - identifying sentiment of individual sentences.
- ABSA sentiment analysis - detecting sentiment towards specific features or topics.
- Emotion detection - categorising expressed emotions such as joy, sadness, anger.
- Sense-level sentiment analysis - identifying sentiment words and phrases.
- Multilingual sentiment analysis - h and ling non-English languages.
- Real-time sentiment monitoring - dynamic analysis of streaming data (Cambria et al., 2017; Shang et al., 2018).

Sentiment analysis has become extremely valuable for gathering consumer insights from diverse sources such as product reviews, social media, blogs, forums and surveys to understand perceptions (Cambria et al., 2013). Common approaches include ML on labelled training data and lexicon-based methods using dictionaries of words annotated for sentiment. Overall, ABSA provides a range of techniques to systematically quantify emotions and opinions from textual data at scale.

2.11.1 Aspect Based sentiment Analysis

ABSA refers to an advanced technique that aims to identify the sentiment expressed towards specific aspects, features, topics or entities within a text document (Schouten & Frasincar, 2016). While basic sentiment analysis works at the document level to determine overall sentiment polarity as positive, negative or neutral, ABSA offers a more granular analysis of sentiment towards individual attributes and features (Liu, 2012). This aspect-level insight provides richer understanding of the drivers of positive or negative sentiment.

ABSA is critical for gaining granular insights into the CX and driving ongoing optimisations. By extracting sentiment towards specific attributes of products, services and interactions, ABSA enables:

- Obtaining detailed customer feedback on touchpoints and interactions along the CJ, beyond overall sentiment (Schouten & Frasincar, 2016). These pinpoints areas working well or needing improvement at a granular level.
- Identifying the root causes of pain points and frustrations from unstructured data like reviews, social posts and open-ended surveys (Preotiuc-Pietro et al., 2016). Addressing these aspect-specific negative sentiments enhances targeted CX facets.
- Uncovering which aspects delight customers when sentiment is analysed across the CJ stages (Liu et al., 2017). Resources can be allocated to elevate strengths and differentiate on these dimensions.

- Guiding optimal resource allocation towards improving the aspects and features that need attention based on sentiment insights (hang et al., 2019). Prioritisation driven by granular voice-of-customer data.

Obtaining multidimensional insights required for continuously optimising integrated end-to-end CX (Marrese-Taylor et al., 2013). Going beyond siloed or snapshot views.

ABSA, as shown in Table 2-16, enables a wide range of practical applications for optimising CX by extracting granular sentiment insights from customer data:

- **Customer Insights**: Analysing product, service and brand reviews to detect customer opinions on specific features, pricing, delivery and other attributes (Liu et al., 2017; hang et al., 2019). This provides actionable voice-of-.
- **Enables Agile Optimisation**: Monitoring social media conversations to identify consumer sentiment towards the brand, marketing campaigns, customer service and other CX elements (Preotiuc-Pietro et al., 2016; Jagdish et al., 2019). Enables agile optimisation.

No	Benefits of ABSA	Description	References
1	Identifying Satisfiers vs Dissatisfiers	ABSA helps differentiate between aspects that customers are satisfied or dissatisfied with, aiding targeted improvements	Khan et al , 2020
2	Understanding Emotional Dynamics	ABSA provides insights into CX emotions, guiding businesses in understanding and improving emotional aspects	Pang & Lee, 2008
3	Prioritising Resources	ABSA reveals sentiment for resource allocation, enabling effective prioritisation of efforts	García Pablos et al , 2020
4	Competitive Benchmarking	ABSA facilitates brand comparison for targeted enhancements and competitive positioning	Pang & Lee, 2008
5	CX Optimisation	ABSA optimizes specific CX aspects using sentiment insights for data driven enhancements	Liu, 2012
6	Dynamic CX Tracking	Continuous ABSA tracks feature sentiment in real time to detect emerging preferences and pain points	Cambria et al , 2013
7	Enhancing Predictive Analytics	ABSA's sentiment correlation with key metrics aids in forecasting business impacts based on CX	Cambria et al , 2017

Table 2-16 The Application of ABSA on CX

- **Holistic CX Improvement**: Surveying open-ended feedback to uncover perceptions of specific touchpoints, interactions and emotions across the end-to-end CJ (Marrese-Taylor et al., 2013; Schouten & Frasincar, 2016). Holistic CX improvement.
- **Quantify ROI of Investments:** Tracking sentiment on aspects like ease of purchase, shipping speed over time to analyse the impact of CX innovations and improvements (Kumar et al., 2018). Quantify ROI of investments.

- **Identify Strengths and Weakness**: Aggregating opinions on service interactions, retail/web environments and other experiential elements (hang et al., 2019; Jagdish et al., 2019).
- **CX Optimisation Based on Granular Insights**: Guiding the design of customer-centric experiences by determining perceptions of individual attributes that compose the integrated CX (Schouten & Frasincar, 2016).

2.11.2 Supervised Machine Learning

Supervised ML has become one of the dominant techniques for ABSA, providing high accuracy when sufficient labelled training data is available (Schouten & Frasincar, 2016).

Common supervised models used for ABSA include (Ma et al., 2021; Shang et al., 2019):

- ***Support Vector Machines (SVM)*** are used to classify the sentiment polarity of extracted aspects based on supervised learning on labelled examples in the training data.
- ***Recurrent Neural Networks (RNN)*** capture sequential dependencies and long-range contexts, which is useful for accurately detecting aspects and sentiment in full review texts.
- **Convolutional Neural Networks (CNN)** can identify locally relevant patterns between aspects and associated sentiment expressions through their convolution operations.
- ***Conditional R and om Fields (CRF)*** build probabilistic models to predict sentiment labels for given aspects, incorporating contextual dependencies.
- ***Bidirectional Encoder Representations from Transformers (BERT)*** employ attention mechanisms to learn contextual relations between words and aspects for ABSA.

A key advantage of supervised learning is the high accuracy compared to unsupervised methods, given sufficient labelled training data. However, unlabelled data is more widely available. The need for manual labelling also limits scalability.

2.11.3 Unsupervised Machine Learning

In addition to the supervised machine learning approaches, unsupervised machine learning techniques have also been explored for ABSA. Unsupervised machine learning does not require labelled training data. Instead, the models learn inherent latent structures from unlabelled data through clustering and association.

One common unsupervised technique used for ABSA is TM, such as LDA covered in Section 2.4.3. LDA can be utilised to discover key aspects or topics from customer reviews in an unsupervised manner. Several studies have applied LDA for aspect extraction from customer reviews (Chen et al., 2020; Angelidis & Lapata, 2018). The extracted aspects or topics can then be used as inputs for sentiment classification using lexicon-based approaches discussed in Section 2.5.4. For example, Chen et al. (2020) utilised LDA TM to extract aspects from hotel reviews. Sentiment lexicon techniques were then applied to classify sentiment polarity for each extracted aspect. Other unsupervised techniques like k-means clustering have also been leveraged for aspect identification from reviews (Angelidis & Lapata, 2018). While supervised techniques generally perform better, unsupervised methods provide value in eliminating the need for labelled data. They can be used for aspect extraction as a pre-processing step before sentiment classification. More research is needed on novel unsupervised techniques for ABSA to improve accuracy while retaining the benefit of unlabelled data.

2.11.4 Knowledge Bases

Knowledge bases such as ConceptNet can enhance unsupervised ABSA by providing external semantics to deduce relationships between extracted aspects and sentiment terms (Schouten & Frasincar, 2016; Cambria et al., 2016).

For example, ConceptNet may contain the fact "camera - used for - photography". This can relate the aspect "camera" with the sentiment "blurry photos" even if they do not co-occur. This allows inferring aspect-sentiment associations. ConceptNet also provides sentiment clues through "HasProperty" relations such as "tasty food - has property - delicious" (Dragoni et al., 2014). Such relations help deduce aspect sentiments.

Additionally, common sense knowledge and ontological relations between aspects such as "screen" and "battery" being related to the super-aspect "mobile phone" aid aspect extraction. In summary, knowledge bases augment unsupervised ABSA methods by providing world knowledge and semantics to identify implicit aspect-sentiment relations not detectable solely through statistical NLP approaches.

2.11.5 Lexicon-Based Approaches

Lexicon-based approaches rely on dictionaries or lexicons of words and phrases that are annotated or compiled with their sentiment orientation to determine the overall sentiment expressed in a given text (Medhat et al., 2014). These approaches leverage existing linguistic knowledge and norms regarding the emotions and feelings associated with specific words and language constructs for sentiment analysis without needing training data (Taboada et al., 2011). Lexicon methods rely on pre-compiled dictionaries mapping words and phrases to sentiment orientations (positive, negative, neutral) and emotions (joy, sadness, anger) without requiring training data (Medhat et al., 2014). By aggregating the lexicon polarity scores of words associated with aspects, overall aspect sentiment can be derived (Muhammad et al., 2020). Lexicon-based approaches were among the earliest techniques developed for ABSA, originating in the mid-2000s as the field emerged (Schouten & Frasincar, 2016). They arose as unsupervised alternatives to requiring labelled data which was scarce for the granular ABSA task.

Lexicon-based approaches have played a pivotal role in enabling the rapid growth and adoption of ABSA over the past two decades (Schouten & Frasincar, 2016). By compiling knowledge about words and their associations with sentiment orientation into pre-defined lexicons, these unsupervised methods alleviate the need for difficult and expensive aspect-level sentiment annotation of training data (Taboada et al., 2011; Liu, 2012). The ready availability of lexicons fuelled substantial research into granular ABSA, which would have faced significant bottlenecks otherwise due to lack of labelled data (Medhat et al., 2014).

In addition to circumventing labelled data requirements, lexicon-based approaches offer other benefits such as language independence for cross-lingual ABSA by constructing multilingual lexicons (Chen & Skiena, 2014). They provide interpretability into how sentiment scores are calculated based on word-level mappings (Dragoni et al., 2014). Lexicons allow capturing abstract notions of sentiment through related terms and concepts (Khan et al., 2020). They also enable multifaceted reuse for document-level sentiment analysis, emotion detection, semantic analysis tasks given their utility as knowledge bases (Taboada et al., 2011). The computational efficiency, ABSA customisation and multipurpose applicability of lexicons have made them indispensable resources for the advancement of fine-grained opinion mining and ABSA.

The most used sentiment analysis lexicons include:

- ***SentiWordNet*** is a lexicon where each synset (group of synonyms) is annotated with numerical scores indicating how positive, negative or objective it is (Esuli & Sebastiani, 2006). It is constructed by automatically annotating all synsets in WordNet with sentiment scores using ML. Advantages of SentiWordNet include coverage of a wide vocabulary and multiple parts of speech. However, limitations include lower accuracy compared to human-annotated lexicons (Muhammad et al., 2020). It is widely used as a baseline sentiment lexicon.
- ***VADER (Valence Aware Dictionary for Sentiment Reasoning)*** is a human-curated lexicon providing norms for over 7,500 lexical features tuned specifically for sentiment analysis of social media text (Hutto & Gilbert, 2014). It uses rules to encode grammatical and syntactic heuristics beyond just word scores. VADER performs well for informal online text due to customised tuning. Limitations include a small coverage focused only on English social media slang. VADER is applied extensively in social media monitoring applications.
- ***Linguistic Inquiry and Word Count (LIWC)*** dictionary tags almost 6,400 English words across 100 psycholinguistic categories including positive and negative emotions (Tauscsik & Pennebaker, 2010). It draws on various corpora to include common words based on usage frequencies. LIWC provides broad coverage of everyday language. But it lacks tuning for colloquial online text. LIWC is universally used in psychology and linguistics research.

- ***Emotion Lexicons*** focus specifically on mapping words to basic human emotion categories such as joy, sadness, fear, anger, through manual curation and crowdsourcing (Mohammad & Turney, 2013). Some popular examples are NRC Emotion Lexicon and DepecheMood. Emotion lexicons offer domain-independence and capture affective nuances beyond positive/negative polarity (B and hakavi et al., 2014). Limited coverage is a disadvantage.
- ***Domain-Specific Lexicons*** are manually tuned for specialised verticals such as finance, healthcare, politics consisting of relevant terminology and expressions (Rouces et al., 2018). For example, the financial lexicon by Loughran and McDonald contains over 2,800 terms.

2.12 Outstanding Challenges

While CX measurement and CXM research have expanded, a coordinated analysis reveals key enduring gaps that remain interlinked and mutually central across both domains. These persistent gaps present barriers to holistic, contextualised and dynamic CX quantification and optimisation (Homburg et al., 2017; Poels & Dewitte, 2019). Secondly, prevalent TM methods lack semantic and contextual understanding, restricting insights from unstructured CX data (Dieng et al., 2020). Additionally, combining TM and ABSA is underexplored but potentially impactful (Schouten & Frasincar, 2016). Finally, applying recent advances in dynamic neural TM to uncover evolving CX topics warrants exploration. Systematically addressing these research gaps can significantly advance CX understanding and practice.

2.12.1 CX Measurement and CXM Research Gaps

Significant research opportunities exist to advance CX and customer CXM across both measurement and analytical frameworks. Table 2-17 provides a structured summary of major gaps identified through a comprehensive analysis of literature.

Although CX methods have advanced, systematic analysis reveals lingering gaps restricting holistic quantification (Gaps 1, 6), dynamic tracking(Gaps 3, 9) and embedded experience analytics (Homburg et al., 2017; Rawson et al., 2013). Comprehensive contemporary reviews indicate continued reliance on surveys and interviews (Gopalan et al., 2022) (Gaps 2, 7). Such legacy self-reported methods analyse individual touchpoints, failing to capture intricacies of multidimensional journeys (Lemon & Verhoef, 2016) (Gaps 1, 6). CX transcends transactions; prevailing models still emphasise episodic snapshots not fluid evolutions (Voorhees et al., 2017; Dijkmans et al., 2015) (Gaps 3, 9).

Current assessment methodologies still employ periodic measurement (Voorhees et al., 2017) (Gaps 3, 9), taking cross-sections rather than analysing entire journey trajectories based on contextual drivers like emotional variability (Dijkmans et al., 2015) (Gaps 2, 7). Though promising, sensor-based ethnography introduces ethical risks regarding privacy (Martin et al., 2017) (Gaps 4, 8). In summary, current gold standards demonstrate gaps in attributes like journey connectivity, emotion granularity, responsible personalisation, continuous measurement and expandability (Homburg et al., 2017; Gopalan et al., 2021) (Gaps 1-10).

This research crystallises an integrated approach incorporating predictive analytics directly into platforms for individualised emotion assessment woven into fluid journeys in a privacy-conscious manner (Gaps 3, 7, 9). Responsible AI techniques aim to balance ethical considerations and desire for comprehensive insights through preserving user transparency and control (MV, 2021) (Gaps 2, 4, 7, 8).

In conclusion, updated analysis reveals lingering measurement barriers around journey integration, responsible personalisation and expandability (Gaps 1-10). This research offers potential pathways aligning to burgeoning privacy-centric, emotionally aware CX quantification trends. However, truly scaling next-generation evaluation requires harmonising responsible rigor and relevance (Gaps 1-10).

CX Measurement Gaps		**CXM Gaps**		**TM Gaps**		**DTM Gaps**		**ABSA Gaps**	
No.	**Description**	**No.**	**Description**	**No.**	**Description**	**No.**	**Description**	**No.**	**Description**
1	Lack of holistic CX frameworks	**6**	Need for integrated measurement frameworks	**11**	Evaluating unsupervised topic models over time	**14**	Tracking semantic drift over time as topics evolve	**17**	Limited labelled data for model training
2	Reliance on subjective self-reported data	**7**	Limitations in quantifying emotions and perceptions	**12**	Contextualised DTM	**15**	Maintaining model stability as new data enters corpus	**18**	Difficulty evaluating on unlabelled real-world data
3	Inability to track dynamic CX over time	**8**	Constraints in capturing influence of social factors	**13**	Integrated topic-aspect modeling	**16**	Re-optimizing models when new topics emerge	**19**	Class imbalance skewing results
4	Difficulty calibrating metrics to actions	**9**	Static measurement approaches						
5	Fragmented data and insights	**10**	Disconnected data across silos						

Table 2-17 Key Research Gaps Summarisation

While Table 2-17 outlined open challenges in CX and CXM research, it is also important to acknowledge seminal frameworks that have made valuable progress in areas related to this study. Table 2-18 summarises three influential existing models, their limitations providing opportunities for further innovation and mappings to how the current research aims to address identified issues. For example, Cambra-Fierro et al. (2020) proposed an ambitious vision for quantifying multidimensional CX by blending surveys, biometrics and text mining but faced gaps in integration, emotions and demonstration that current artifacts aim to fulfil. Analysing preceding frameworks in this manner highlights specific possibilities for targeted expansions through the structured CX/CXM research gaps lens provided earlier. Evaluation of existing key models in the closing section of Chapter 2 also aids in positioning the distinguishing elements this research seeks to contribute. The trajectories of previous theories shape the contextual springboard for current innovations.

Author	Relevant Existing Framework	Identified Gaps	Research Contributions
Cambra-Fierro et al. (2020)	Quantifying CX using surveys, biometrics and text mining	• Unintegrated techniques • Static analysis • Limited emotion capture • Not CX data tested	• **Integrated insight framework** • **Temporal text mining** • **Uncovers emotions** • **CX corpus demonstrated**
Homburg et al. (2017)	Needed dynamic CX quantification models	• Models not implemented • Simulated data • Tools disconnected	• **Temporal analytics models** • **Unified framework** • **CX case studies**
Lemon and Verhoef (2016)	Conceptualized multidimensional cumulative CX	• Conceptual only • No techniques shown • Metrics not empirical	• **Computable CX features** • **NLP models demonstrated** • **Validates models revealing insights**

Table 2-18 Cross-Referencing Relevant CX Framework, Gaps and Current Study

Deriving contextual insights from unstructured data can enhance multifaceted CX quantification (Cambra-Fierro et al., 2020; Poels & Dewitte, 2006). Dynamic analytics can capture fluid CX evolutions (Blei & Lafferty, 2006; Laros & Steenkamp, 2005). Advancing these mutual CXM and CX measurement research gaps through a coordinated research agenda is imperative for continued progress and impact. These research gaps are interdependent, requiring coordinated efforts. However, CX quantification also faces technological constraints. in addition to the that, with CXM, CX measurement represents an evolving challenge as consumer interactions and feedback become increasingly dynamic and unstructured. While established analytical techniques such as TM and sentiment analysis have been applied for CX insights, some underexplored gaps remain undiscovered.

2.12.2 Leveraging Technological Techniques for CXM

Figure 2-1 present mapping gaps and limitations of three NLP techniques - TM, DTM and ABSA - to broader gaps and challenges in CX measurement and management. For each NLP technique in the leftmost column, the "Relevant Gaps" column indicates technique-specific gaps. The middle "CX Measurement Gaps" column shows how those technique gaps relate to broader analytical gaps in measuring and quantifying the CX. Finally, the "CXM Gaps" column associates the technique limitations with higher-level CXM challenges.

Although AI and ML have enabled transformative CX capabilities, critical analysis reveals lingering gaps obstructing integrated, contextualised and causally transparent intelligence (Homburg et al., 2020). Extensive reviews spotlight fragmented infrastructure frequently restricting CX data interoperability, hindering holistic analysis (Kumar et al., 2020) (Gaps 1, 5, 6, 10). Solutions thus far demonstrate substantial limitations providing adequate semantics explaining data-driven models for end users and contextual awareness linking predictions to drivers (Harmeling et al., 2022) (Gaps 11-13).

Core enduring obstacles centre on fractured data, inscrutable “black box” recommendations (Kumar et al., 2020), lagging relevancy from retrospective insights and model opacity impeding trust and adoption (Gaps 5, 10-13, 17-19). This research crystallises a pioneering convergent architecture guided by behavioural theory melding BERT's embedded semantics (Devlin et al., 2018), temporally aware tracking (Ryu & Hu, 2021) and multi-level hierarchical explainability (Pauwels et al., 2020) to address fragmentation, causality and opacity limitations. The proposed artifacts specifically target gaps in semantics, dynamics and causality by synergising advanced deep contextual language models such as BERT (Gaps 12-13), responsible temporal analytics protecting privacy (Gaps 3, 9, 14) and granular local explanations to enhance interpretability (Gaps 11, 17-19).

In summary, while AI/ML has delivered immense CX advancements, updated analysis reiterates obstacles related to disjointed data, inscrutable predictions, reactive insights and ethical risks impeding ubiquitous adoption (Gaps 1-19). This research crystallises an integrated pathway aligned with emergent trends in human-centred trustworthy AI. Fusing state-of-the-art statistical modelling with psychological rigor aspires to responsibly unlock unprecedented CX intelligence.

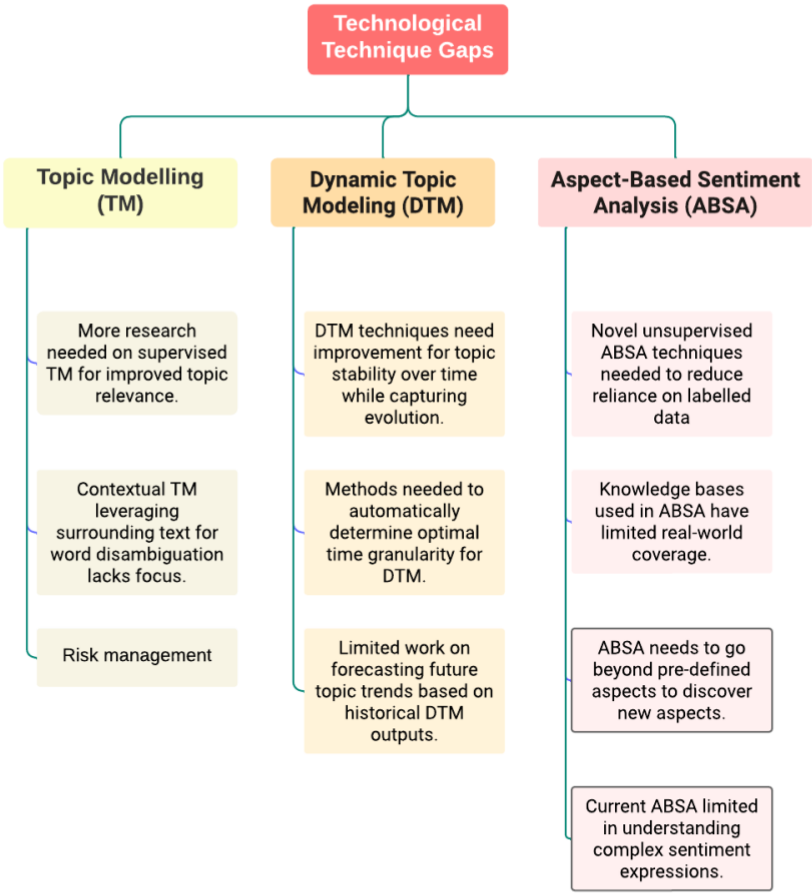


Figure 2-1 Technological Research Gaps

Table 2-19 summarises key linkages between analytical techniques and their associated research gaps in CX measurement and CXM frameworks. Specifically, it maps the gaps each technique may help address. For TM, relevant gaps centre around the lack of holistic CX frameworks and fragmented, disconnected data in both CX measurement and CXM (Gaps 1, 5, 6, 10). This highlights TM's potential for enabling more integrated frameworks and deriving insights from fragmented data sources. For DTM, the key gaps involve improving dynamic CX tracking over time for both domains (Gaps 3, 9). This points to DTM's promise for better capturing evolving CX journeys. Finally, the gaps for ABSA focus on reducing over-reliance on subjective self-reported data and improving emotion quantification (Gaps 2, 7). ABSA could thus advance subjective CX assessment capabilities.

In summary, the table illustrates important linkages between analytical techniques and the CX/CXM gaps they may help mitigate. By mapping methods to open challenges around integration, dynamics, subjectivity and emotions, it highlights potential research opportunities. TM, DTM and ABSA show particular promise for advancing CX/CXM frameworks and measurement in interconnected areas based on the gaps summarised

Technique	Relevant Gaps	CX Measurement Gaps	CXM Gaps
Topic Modeling	Lack of holistic CX frameworks (1)	Need for integrated measurement frameworks (6)	Lack of holistic CX frameworks (1)
	Fragmented data and insights (5)	Disconnected data across silos (10)	Fragmented data and insights (5)
Dynamic Topic Modeling	Inability to track dynamic CX over time (3)	Static measurement approaches (9)	Inability to track dynamic CX over time (3)
Aspect-Based Sentiment Analysis	Reliance on subjective self-reported data (2)	Limitations in quantifying emotions and perceptions (7)	Reliance on subjective self-reported data (2)

Table 2-19 Summary of Techniques and Corresponding CX/CXM Gaps

2.12.3 Research Gaps

According to Table 2-19, the research gaps between CX measurement and CXM are closely interconnected and complementary. Integrated CX measuring skills are required for holistic CXM (Homburg et al., 2017; Rawson et al., 2013).

Based on Figure 2-2, the overlapping research gaps are explained below:

1. **Research Gap 1**: Lack of holistic CX frameworks (CXM gap 1) (Poels & Dewitte, 2019) and need for integrated measurement frameworks (CX gap 6) (Rawson et al., 2013).

These gaps are overlapping because developing holistic CX frameworks requires integrated CX measurement capabilities. Holistic CXM relies on the ability to integrate and quantify distinct aspects of the CX (Homburg et al., 2017; Rawson et al., 2013).

2. **Research Gap 2** : Reliance on subjective self-reported data (CXM gap 2) (Gountas et al., 2007) and Limitations in quantifying emotions and perceptions (CX gap 7) (Poels & Dewitte, 2006).

These are connected since self-reported subjective data has inherent limitations in accurately capturing emotional perceptions and experiential aspects. Overcoming these limitations requires better techniques to quantify subjective dimensions (Cambra-Fierro et al., 2020; Poels & Dewitte, 2006).

Technological Approach	Relevant Gaps	CX Measurement and Gaps	CXM Measurement and Gaps	Overlaps
Biometrics	Limitations in quantifying emotions and perceptions (7)	Capturing subjectivity (7)		**Capturing subjectivity**
Data Mining	• Nee • d for integrated measurement frameworks (6) • Disconnected data across silos (10)	Multidimensionality (6)	• Breaking data silos (5) • Scaling CX data analytics (11)	**Multidimensional data complexity**
Mobile/IoT	Static measurement approaches (9)	• Transient nature (9) • Context dependence Individual differences (2)		• **Dynamic CX tracking** • **Privacy**
AI Personalization	Limitations in quantifying emotions and perceptions (7)	• Capturing subjectivity (7) • Individual differences (2)		• **Capturing subjectivity, Privacy** • **Personalisation**
CRM Systems	Disconnected data across silos (10)	Lack of integration (10)	• Breaking data silos (5) • Enabling touchpoint consistency (1)	• **Integration** • **Data silos**
Social Media Analytics	Disconnected data across silos (10)		Breaking data silos (5)	**Data silos**
CX Analytics Dashboards			• Linking metrics to actions (4) • Predictive analytics	**Predictive analytics**

Table 2-20 CX Measurement and CXM Measurement Gaps and Their Overlap

3. **Research Gap 3**: Inability to track dynamic CX over time (CXM gap 3) (Blei & Lafferty, 2006) and Static measurement approaches (CX gap 9) (Laros & Steenkamp, 2005)

Capturing dynamic CX journeys over time is constrained by static measurement approaches. Advancing dynamic CX tracking requires moving beyond static metrics (Blei & Lafferty, 2006; Laros & Steenkamp, 2005).

4. **Research Gap 4**: Fragmented data and insights (CXM gap 5) (Lemon & Verhoef, 2016) and Disconnected data across silos (CX gap 10) (Lemon & Verhoef, 2016).

Deriving integrated insights is impeded by fragmented, disconnected data spread across organisational silos. Overcoming data fragmentation is key for both research domains.

5. **Research Gap 5**: Tracking semantic drift over time (DTM gap 14) (Technology gap) and Inability to track dynamic CX over time (CXM gap 3) (Blei & Lafferty, 2006).

These are directly connected since tracking evolving semantics enables better understanding of dynamic CX topics over time.

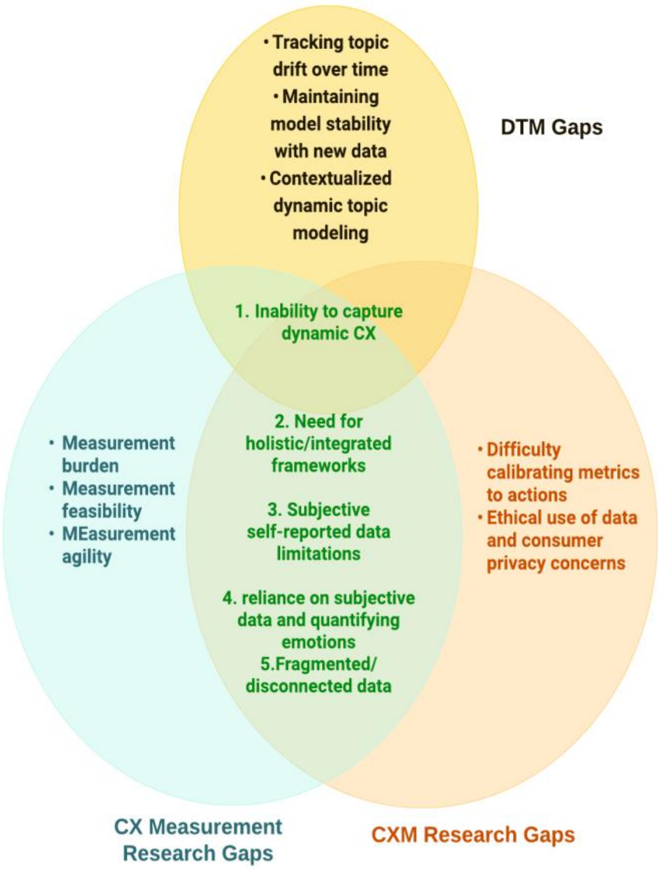


Figure 2-2 CX Measurement and CXM Research Gaps

2.13 Summary

This chapter has explored the growing importance of CX as a source of competitive differentiation and revenue growth. However, accurately measuring and managing the subjective, multidimensional and dynamic nature of CX over time remains an ongoing challenge. While CX research has expanded substantially, several key gaps persist around developing holistic CX assessment frameworks, quantifying emotional dimensions and capturing fluid CX journeys longitudinally.

Moreover, CXM has emerged as a strategic imperative focused on understanding and optimising CX across touchpoints. However, CXM lacks integrated frameworks and relies heavily on subjective self-reported data, resulting in fragmented insights. Enhancing CXM requires addressing these limitations.

TM and ABSA offer automation for analysing unstructured CX data. However, semantic complexities persist in TM while limited labelled data and model evaluation approaches constrain ABSA. Evaluating topic models over time is also underexplored. Integrated topic-aspect modelling shows potential but needs advancement.

Recent techniques such as contextualised DTM, neural topic models and joint topic-aspect modelling exhibit promise for overcoming limitations around quantifying subjective perceptions, capturing dynamic CX and integrating insights holistically. However, extensive real-world validation is still needed.

The review of literature has synthesised enduring gaps around holistic CX measurement, static assessment approaches, fragmented insights, lack of contextual understanding in TM and constraints around evaluation and labelled data for ABSA. Advancing techniques identified through coordinated research initiatives is imperative for continued progress. The review sets the stage for proposing solutions to address these research gaps through novel analytics approaches.

Chapter 3: Research Methodology

3.1 Overview

A systematic method is required for achieving credible and repeatable results in research. his chapter provides a full account of the research methodology utilised to develop this research. The thesis employs the DSR approach as its primary methodology for this study.

This chapter presents the research methodology used in this study. It provides background on the DSR paradigm and how it is applied through iterative design cycles to address the research problem. The chapter is structured as follows: Section 3.2 discusses common approaches to Information Systems research, such as behavioural science and DSR. Section 3.3 describes DSR's history, strengths and research context. Section 3.4 covers the DSR paradigm, knowledge base, design theory, guidelines and frameworks. Section 3.5 describes the DSRM process that was chosen and its iterative design cycles. Section 3.6 presents the first DSRM design cycle. Section 3.7 shows the second DSRM design cycle. Section 3.8 describes the third DSRM design cycle.

3.2 Information System Research Approaches

A diverse set of research methodologies and paradigms have been utilised within the interdisciplinary field of information systems (IS). Some of the most common approaches include:

Design Science Research

DSR focuses on the design and evaluation of technological artifacts and systems to solve organisational and business problems (Hevner et al., 2004). Rather than behavioural theories, the goal is innovative artifacts such as constructs, models, methods and instantiations encapsulating design knowledge. Iterative build-evaluate loops connect relevance and rigor. For example, Peffers et al. (2018) designed and evaluated a mobile app prototype for improving the CX in retail stores. Teixeira et al. (2012) developed and evaluated an IT artifact for extracting CX insights from user-generated content on social media. Bilgihan et al. (2016) designed and tested a CX dashboard prototype to monitor multiple touchpoints and enable data driven CXM.

Empirical Research Methods

Empirical Research Methods (ERM) involve the systematic collection and analysis of data to understand or test phenomena. They allow investigating real-world behaviours, systems, or problems (Bakehouse, 2000). The followings are the methods to conduct ERM:

- **Surveys -** Surveys involve collecting data via questionnaires, interviews, or focus groups to quantify or describe tendencies, attitudes, or opinions of a population (Pinsonneault & Kraemer, 1993). Klaus and Maklan (2013) used online surveys to identify drivers of CX quality in retail banking. Maklan et al. (2015) surveyed customers to expand a model of CX antecedents and outcomes including rapport, value co-creation and intention to recommend. Lemon and Verhoef (2016) surveyed customers to understand how brand equity affects CX and loyalty in a B2B setting.

- **Experiments -** Experiments manipulate one or more independent variables in a controlled setting to determine causality between variables (Dennis & Valacich, 2001). Roggeveen et al. (2012) conducted lab experiments to examine how the presence of other customers affects service experience evaluations. De Keyser et al. (2015) experimentally tested the effect of customer gratitude and indebtedness on relationship outcomes after service recovery. Kranbühler et al. (2018) experimentally tested how a company's response to negative online reviews affects customer trust and loyalty.
- **Case Studies -** Case studies involve an in-depth examination of a phenomenon in its real-world context via interviews, observations, documents, etc. (Darke et al., 1998). Tax and Stuart (1997) did an in-depth case study of CXM at General Motors. Meyer and Schwager (2007) examined CX transformation at BMW through interviews and observation. Heinonen and Strandvik (2015) conducted a case study of a telecom company's efforts to transform to a customer-centric organisation.
- **Ethnography -** Ethnography involves immersive observation of behaviours, cultural practices and interactions of a group in its natural setting (Myers, 1999). Cotte and Kistruck (2006) did an ethnographic study of how customers derive value from brand communities. Mariampolski (2006) directly observed customers to identify needs not captured by traditional market research.
- **Action Research -** Action research combines inquiry and intervention through an iterative process to understand and support change in a system (Baskerville, 1999). Bitner et al. (2000) worked with a hotel to test new service strategies based on customer feedback. Rowley (2002) acted as an academic consultant to collaborate with a firm on improving customer retention. Bitner et al. (2008) engaged in action research with a hotel chain to improve service delivery processes.

Analytical Research Methods (ARM)

ARM in information systems is computational and statistical technique designed to analyse, interpret and visualise data to make informed decisions, particularly in complex environments where vast amounts of information are processed and stored (Delen, 2010).

- **Conceptual Analysis -** Theoretical work focused on defining key concepts, developing conceptual frameworks, or proposing new theoretical models (Straub et al., 2004). Lemon and

Verhoef (2016) proposed a conceptual model that examines CX formation throughout the CJ and across touchpoints. Maklan and Klaus (2011) conceptually defined CX quality and proposed relevant dimensions. Verhoef et al. (2009) developed a conceptual model describing determinants and consequences of CX in a service context.

- **Mathematical Modelling -** Developing mathematical representations using analytical modelling, simulations, or other techniques. Li, Hitt and hang (2020) used analytical modelling to assess how customer participation in online reviews impacts CX and referral likelihood. hang et al. (2010) developed a simulation model to evaluate CX integration across channels. Neslin et al. (2006) built a customer lifetime value model incorporating customer satisfaction and retention metrics.
- **Decision Analysis -** Systematic, analytical approaches to determine optimal decisions and quantify trade-offs given constraints and uncertainties (Clemen & Reilly, 2013). Rust and Huang (2014) developed a dynamic programming model to optimise resource allocation across CX touchpoints over time, accounting for budget constraints and cross-channel effects. Lewis (2004) proposed using decision analysis to systematically evaluate CX metrics like satisfaction, retention and customer lifetime value to select optimal measures. Li et al. (2020) applied a Markov decision process model to identify optimal strategies for managing different CX states based on customer segments and recovery actions.

Table 3-1 compares common research methods across information systems and other fields. Each method has distinct strengths and weaknesses that make it suitable for certain research questions and objectives

Research Method	Strengths	Weaknesses
Design Science Research	✅Creates innovative artifacts to solve problems ✅Embodies design knowledge	❌Not focused on theory building ❌Limited generalisability of artifacts
Empirical Research Methods	✅Allows testing theories with data ✅Provides generalisable insights	❌Does not directly create solutions ❌Artificial experimental settings
Surveys	✅Quantifies attitudes and behaviours ✅Cost effective data collection	❌Self-reported data subject to biases ❌Limited depth of insights
Experiments	✅Determines causality between variables ✅Controls extraneous factors	❌Artificial lab settings ❌Hawthorne effect
Case Studies	✅Provides in-depth understanding ✅Real-world context	❌Not generalisable ❌Resource intensive
Ethnography	✅Directly observes behaviours and culture ✅Naturalistic setting	❌Time consuming ❌Observer bias
Action Research	✅Supports change in organisations ✅Collaborative	❌Context specific ❌Ethical concerns
Analytical Research	✅Models complex phenomena ✅Optimizes decisions	❌Simplifies reality ❌Dependent on assumptions
Conceptual Analysis	✅Develops theoretical foundations	❌Does not test theories empirically
Mathematical Modeling	✅Predicts outcomes ✅Simulates scenarios	❌Simplifies complex realities
Decision Analysis	✅Prescribes optimal decisions	❌Depends on input parameters

Table 3-1 Strengths and Weaknesses of Different IS Research Methods

3.3 Behavioural Science Research

Behavioural science research refers to a broad range of social science disciplines focused on studying human behaviour, including psychology, sociology, organisational behaviour, cognitive science, anthropology and others. In the context of information systems, behavioural science research examines how humans and organisations interact with information and technology.

The aim of behavioural IS research is to develop and scientifically test theories that explain or predict phenomena related to the adoption, usage, management and impact of information systems and technology in organisational contexts (Straub et al., 1994). This involves formulating research models consisting of constructs, relationships and hypotheses grounded in behavioural science theories. Behavioural IS research has its origins in the multidisciplinary roots of IS as a field bridging technology and social science domains. Early IS researchers incorporated theories from social psychology, management, cognitive science and other disciplines to study how users interacted with and adopted innovative technologies (Davis, 1989). Typical implementation of behavioural research in IS involves quantitative methodologies including surveys, experiments, observational studies and field research (Pinsonneault & Kraemer, 1993). These methods allow collection of empirical data from individuals, groups and organisations to statistically test hypothesised relationships between theoretical constructs. The ability to develop generalisable theories grounded in behavioural science has made this paradigm highly influential in IS research. The prevalence of behavioural studies in top IS publication outlets underscores its importance. Behavioural IS research directly aligns with the field's application-oriented focus on interactions between humans/organisations and information technologies.

Behavioural science research leverages social science theories and quantitative methodologies to empirically develop and test models explaining human and organisational behaviours related to information systems and technologies. The paradigm has been impactful since the origins of IS and continues to be a dominant approach.

3.3.1 DSR and Behavioural Science Research Comparison

Behavioural science research aims to develop and validate theories explaining phenomena by adopting natural science approaches focused on explanatory power and prediction (Hevner & Chatterjee, 2010). In contrast, DSR focuses on developing prescriptive knowledge encapsulated in novel artifacts constructed as solutions to practical business problems (Peffers et al., 2007). Several factors indicate DSR is better suited as a methodology for this research compared to behavioural science (Gregor & Hevner, 2013):

- The planned iterative approach of model development and evaluation is consistent with the design-build-evaluate loop cycle emphasised in DSR (Kuechler & Vaishnavi, 2008).
- The practical research contribution as novel artifacts matches DSR's focus on utility rather than behavioural science's focus on explanatory principles.
- Evaluation will prioritise model performance on real customer data versus fit to a theoretical model, aligning with DSR's pragmatic metrics (Venable et al., 2016).
- The models could reveal new insights warranting future behavioural research, showing how DSR can enable subsequent theory-building (Niehaves, 2007).

Attribute	Behavioural Science	Design Science Research (DSR)
Goals	Seeks to explain and predict phenomena	Devises artifacts to improve outcomes and capabilities
Approaches	Tests hypotheses using empirical data	Iteratively designs and evaluates artifacts in context
Outputs	Produces explanatory/predictive principles and models	Produces purposeful constructs, models, methods, instantiations
Contributions	Yields knowledge representing reality	Yields knowledge for designing solutions to problems
Evaluation	Depends on data fitting models	Depends on utility and efficacy of artifacts

Table 3-2 DSR and Behavioural Science Comparison Based on Various Attributes

Based on Table 3-2, DSR is a more appropriate methodology choice based on the research aims of developing artifacts for a business problem using iterative approaches with a priority on utility over explanatory modelling. DSR better matches the research goals compared to a behavioural science paradigm.

3.4 Design Science Research

DSR has emerged as an important research paradigm within information systems that is focused on developing solutions to complex real-world problems through the design and evaluation of innovative technological artifacts (Hevner & Chatterjee, 2010; Peffers et al., 2007). The core idea of DSR is that knowledge and understanding of a problem domain can be captured in the design of novel artifacts that provide utility and value for stakeholders (Hevner et al., 2004). DSR is the iterative build-evaluate cycle in which artifacts are built, applied and assessed in context, linking relevance to rigour (Hevner et al., 2004). However, Walls et al., (1992) proposed a model-centric approach to DSR that focuses on the design, building and evaluation of "information system design theories" (ISDT). This approach is commonly referred to as the "Design-Build-Evaluate" cycle. Walls et al. further emphasised the importance of creating an ISDT, which has two main components:

1. **Kernel Theories**: These are existing theories from the natural or social sciences that provide a foundation for the ISDT. They inform the design and help in understanding the underlying principles that guide the artifact's construction and evaluation.
2. **Testable Hypotheses**: The ISDT should result in hypotheses that can be empirically tested. This ensures that the DSR is grounded in rigorous scientific methods and that the artifacts produced can be validated through empirical evidence.

DSR has become increasingly important for information systems scholarship over the past two decades (Hevner & Chatterjee, 2010; Peffers et al., 2007). DSR is well-aligned with the identity of IS as a field focused on socio-technical systems and technologies rather than just social behaviours and interactions (Gregor & Hevner, 2013). At its core, IS produces knowledge related to developing information technologies and systems for organisational purposes. The design science paradigm directly enables this through iterative design, building and evaluation of IT artifacts aimed at addressing business needs and opportunities (Hevner & Chatterjee, 2010). DSR creates novel constructs, models, procedures and instantiations embodying knowledge about accomplishing practical tasks, in contrast to explanatory behavioural theories. DSR provides strengths of developing innovative artifacts to directly address problems, while also developing explanatory theory regarding the efficacy of those artifacts (Jones and Gregor, 2007).

The rigorous yet relevant design-build-evaluate loop is a key aspect of DSR's importance, driving artifact construction and assessment in applied organisational contexts (Hevner et al., 2004). Relevance ensures DSR artifacts target important business problems that provide utility if successfully implemented. Rigor arises from grounding artifacts in current theoretical foundations and using scientific methods in evaluation (Iivari, 2015). Together, relevance and rigor allow DSR to bridge practitioners' needs with academic knowledge. By complementing behavioural science theories grounded in implementation and evaluation, DSR provides an important paradigm supporting the core identity of IS research (). DSR enables developing IT artifacts encapsulating novel solutions to complex organisational problems while adhering to scientific principles. The paradigm provides a rigorous framework aligning IS scholarship with its applied sociotechnical focus on technologies shaped by and shaping human purposes. This powerful combination of practical relevance and methodological rigor makes DSR an increasingly vital IS research tradition.

3.4.1 DSR Background

The philosophical foundations of DSR stem from Herbert Simon's influential 1960s work on the "sciences of the artificial" distinguished from natural sciences (Simon, 1996). Simon advocated for legitimacy of research creating innovative artifacts serving human purposes. This planted seeds for applying rigorous design and engineering approaches to generate prescriptive knowledge on achieving practical goals.

In IS, key concepts crystallised with Nunamaker et al. (1990) providing a systems development methodology blending rigor and relevance. Walls et al. (1992) built on Simon in proposing an information systems design theory (ISDT) framework. This formalised components like kernel theories, meta-requirements, etc. By the mid 1990s, March & Smith (1995) clearly characterised IS design science as producing artifacts like constructs, models, methods and instantiations improving system performance.

Hevner et al. published an influential 2004 paper that became a landmark for DSR in IS. They defined key cycles and guidelines ensuring relevance to business contexts and rigor through scientific design and evaluation. In the past 15+ years, DSR approaches have continued maturing with refinements in processes, evaluations and philosophical groundings. DSR has clearly emerged as an impactful IS research paradigm complementing behavioural and qualitative approaches

3.5 Design Science Research Paradigm

DSR has emerged as an important paradigm in information systems research focused on developing innovative technological artifacts as solutions to complex organisational problems (Hevner & Chatterjee, 2010). DSR aims to create novel constructs, models, methods and systems encapsulating design knowledge for accomplishing practical goals (March & Smith, 1995). Through relevance and rigor cycles, design science supports developing IT artifacts that simultaneously provide utility for applications while making contributions to theory (Iivari, 2015). By producing prescriptive knowledge on what can be constructed to reach desired goals, DSR expands the descriptive focus of behavioural theories (Gregor & Hevner, 2013).

3.5.1 DSR Artifacts

DSR aims to produce innovative artifacts that provide solutions to important business and organisational problems. DSR artifacts encapsulate prescriptive knowledge in the form of constructs, models, methods, instantiations, design theories and technological rules (Lukyanenko et al., 2020; Sein et al., 2011). As shown in Table 3-3, key DSR authors have proposed various typologies of artifacts.

The March and Smith (1995) categorisation of DSR artifacts into constructs, models, methods and instantiations provides a parsimonious yet comprehensive taxonomy that has been widely adopted across DSR studies (Hevner et al., 2004). This framework classifies artifacts into four core types:

- Constructs provide the conceptual vocabulary and symbols for describing problems and solutions.
- Models represent relationships between constructs to describe the operation of artifacts.
- Methods provide processes and algorithms detailing steps for accomplishing tasks.
- Instantiations demonstrate feasibility by implementing artifacts in a working system.

Author	Constructs	Models	Methods	Instantiations	Design Theories	Technological Rules
Hevner et al. (2004)	✓	✓	✓	✓		
March & Smith (1995)	✓	✓	✓	✓		
Gregor & Jones (2007)		✓		✓		
Sein et al. (2011)		✓	✓	✓	✓	
Baskerville et al. (2018)	✓	✓	✓	✓		✓
Goldkuhl (2004)		✓	✓			
Venable (2006)		✓		✓		
Lukyanenko et al. (2020)	✓	✓	✓	✓	✓	✓
Mullarkey & Hevner (2019)		✓		✓		
Drechsler & Hevner (2018)		✓		✓		

Table 3-3 Perspectives on DSR Artifact Types

This taxonomy offers several benefits that support its selection for this research:

1. Encapsulates the key forms of artifacts produced in DSR in a clear and differentiated manner (Gregor & Hevner, 2013).
2. It offers simplicity while still thoroughly addressing the variety of artifacts generated in DSR projects with only four categories.
3. it has been extensively validated through widespread adoption in DSR studies since its publication (Hevner et al., 2004).
4. The categories align well with the planned research approach for this study which involves developing models and instantiating them in prototypes.

In summary, the March and Smith (1995) framework provides a parsimonious, valid and comprehensive taxonomy of DSR artifact types that has become widely adopted across the field. Based on its clarity, applicability to this research and extensive empirical support, this study will utilise the March and Smith (1995) typology to categorise the artifacts produced by the DSR process. It provides an appropriate, useful lens for classifying artifacts.

3.5.2 DSR cycles

High quality DSR involves iterating through various cycles that connect the problem context to the rigor of the artifact design process (Hevner, 2007). Key cycles in DSR include the relevance cycle, design cycle and rigor cycle (Hevner & Chatterjee, 2010; Iivari, 2015). These cycles work together to ensure the research produces relevant solutions grounded in scientific knowledge and methods.

- **Relevance Cycle**: The relevance cycle bridges the contextual environment of the research project to the design science activities (Hevner, 2007). It begins with identification of opportunities, problems, or needs in an application context and defines requirements for an effective solution. The cycle culminates in field testing and organisational implementation of the artifact to assess its utility in delivering expected benefits. The relevance cycle connects the research to relevant business contexts.

- **Design Cycle**: The design cycle iterates between core DSR activities of building and evaluating artifacts that address problems identified in the relevance cycle (Peffers et al., 2007). Building draws on available means and prior knowledge to develop an artifact matching the requirements. Evaluation involves assessing artifact performance and utility using rigorous methods such as experiments, simulations, analytics, cases studies, etc. The cycle repeats as the artifact is refined. This cycle connects to rigorous construction and assessment.

- **Rigor Cycle**: The rigor cycle provides grounding in scientific foundations through the artifact design process (Hevner & Chatterjee, 2010). Prior knowledge from research and theory informs artifact construction. New knowledge generated in the design cycle contributes back to the knowledge base. Applying rigorous techniques and methods throughout the research process ensures the rigor cycle permeates the research. This cycle connects research activities to the broader foundation of applicable knowledge. This research cycle provides grounding in the knowledge base to ensure scientific rigor during artifact construction and evaluation (Hevner & Chatterjee,

Knowledge Base Contributions

The research contributes back to the knowledge base in multiple ways:

- Demonstrates the value of integrating behavioural frameworks with text analytics for more nuanced DTM.

- Provides empirical analysis of established techniques such as LDA and BERTopic on real-world CX datasets to characterise utility and limitations in context.

- Enhances understanding of using ABSA for fine-grained CX insights from unstructured data.

- Overall, applies existing knowledge in innovative ways to address underexplored problem space at the intersection of CX analytics and text mining.

The knowledge contributions focus both on enhancing adoption of known techniques in new domains and expanding behavioural theories for text analytics.

Evaluation Rigor

Rigorous processes are followed for artifact evaluation, including:

- Quantitative performance metrics such as topic coherence, diversity, perplexity and sentiment accuracy (Chapters 4 and 5).
- Qualitative assessment by domain experts of model outputs and insights (Chapters 4 and 5).
- User-centric criteria such as utility and relevance grounded in application context (Chapter 5).
- Reproducible processes ensured through details of datasets, parameters, analytics pipelines etc. (Chapters 3-5).
- Adhering to Hevner et al.'s guidelines ensures evaluation rigor in both the construction and assessment of artifacts (2004).

This research adheres to scientific standards through leveraging and contributing back to knowledge, applying rigorous construction methods and performing robust evaluation. This upholds, as presented in Figure 3-1, the vital role of the rigor cycle in DSR.

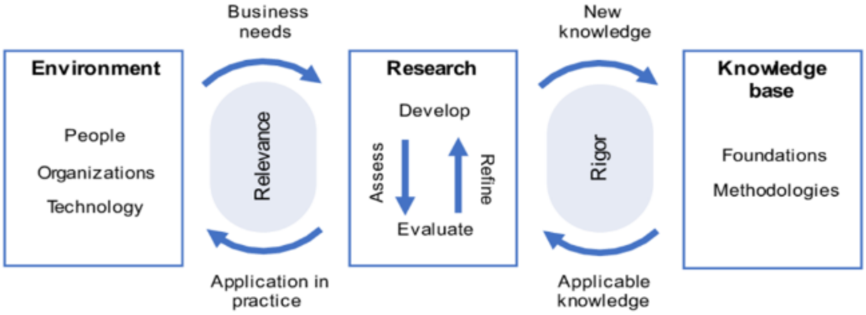


Figure 3-1 DSR Three-Cycle view (Hevner, 2007)

3.5.3 DSR Knowledge Base

The knowledge base is a key component of DSR that refers to the foundations from which artifacts are constructed and evaluated (Hevner et al., 2004). It provides the raw materials for carrying out DSR and accumulates the new knowledge generated by DSR projects (Gregor & Hevner, 2013).

The knowledge base consists of two parts (Hevner et al., 2004):

1. **Foundations**: (also called kernel theories or justificatory knowledge) This includes prior knowledge from reference disciplines that provides applicable theories, frameworks, instruments, constructs, models, methods etc. that inform artifact design (Venable, 2006; Goldkuhl, 2004). For instance, in customer analytics, foundations include techniques such as

ABSA (Hu & Liu, 2004), TM (Blei et al., 2003), contextual embeddings (Devlin et al., 2019) etc. from reference fields such as ML and NLP (Goodfellow et al., 2016). Foundations provide the existing knowledge base for grounding DSR work (March & Smith, 1995).

2. **Methodologies:** Methodologies - This encompasses the practices, processes, approaches and evaluation/validation criteria that guide the rigorous conduct of DSR (Hevner & Chatterjee, 2010). Methodologies ensure artifacts are constructed and assessed according to scientific principles (Peffers et al., 2007). For example, this includes guidelines on rigor cycles (Hevner, 2007), design-build-evaluate iterations (Sein et al., 2011), technical evaluations (Venable et al., 2016) and expert feedback (Sonnenberg & vom Brocke, 2012). Frameworks such as DSRM (Peffers et al., 2007) provide methodological guidance.

In addition to consuming existing knowledge, DSR also contributes back new knowledge to the foundations and methodologies (Gregor & Hevner, 2013):

- New artifacts such as models, frameworks, architectures, constructs etc. are added to the knowledge base.
- Enhancements to processes, analytics techniques and evaluation methods expand methodological knowledge.
- Learning and expertise gained from design cycles provide foundation for future projects.

In essence, the knowledge flows cyclically from base to project to base in DSR. A rigorous project solidifies existing knowledge while also expanding the knowledge base through novel artifacts and insights (Peffers et al., 2018). Carefully leveraging, contributing to and evolving the DSR knowledge base through foundations and methodologies is key to developing impactful artifacts anchored in current knowledge while also providing new innovations.

3.5.4 DSR Frameworks

Systems Development Research Methodology

One of the earliest DSR methodologies was presented by Nunamake et al. (1990). A systems development research methodology focused on developing and evaluating IT artifacts through an iterative process involving theory building, system development, experimentation and observation.

Five key process elements are comprised by their methodology:

1. Constructing a conceptual framework: Developing an initial conceptual framework based on experience, system objectives, existing theory and prior DSR projects.
2. Developing a system architecture: Creating an artifact design and architecture grounded in the conceptual framework and subject to refinements.

3. Analysing the system: Performing rigorous experiments, simulations, metrics etc. to analyse and evaluate the artifact within laboratory and real-world settings.

4. Observing system usage: Studying deployment of the artifact in an organisation to observe use and impact.

5. Refining the design: Using feedback from technical analyses and organisational implementation to iteratively refine the artifact design.

A strength of this methodology is the recognition of rigorous building and evaluation coupled with organisational implementation and observation for refinement (Nunamaker Jr et al., 1990). However, limitations include a linear rather than iterative focus and lack of explicit relevance cycle.

DSR Methodology

Peffers et al. proposed a more detailed six step DSR methodology aimed at producing innovative artifacts with research contributions (Peffers et al., 2007). The six activities are:

1. Problem identification and motivation: Identify problem and justify importance of solution artifacts.

2. Define objectives: Infer objectives and requirements for a solution based on the problem context.

3. Design and development: Create the artifact based on objectives and knowledge from theoretical foundations.

4. Demonstration: Demonstrate efficacy of the artifact for solving the identified problem via experimentation, simulation etc.

5. Evaluation: Rigorously assess performance of artifact using well-defined metrics.

6. Communication: Effectively communicate DSR process and results to appropriate audiences.

Strengths of this methodology include clear sequenced steps, emphasis on demonstration and evaluation and focus on dissemination of findings (Peffers et al., 2007). Weaknesses include a linear approach rather than iterative cycles.

DSR Methodology

Vaishnavi and Kuechler proposed a general methodology for DSR consisting of five iterative steps (Vaishnavi & Kuechler, 2004):

1. Awareness of the problem: Identify the specific problem that requires a new artifact solution.

2. Suggestion: Propose a potential artifact idea or concept as a tentative solution.

3. Development: Instantiate and implement the artifact based on proposed solution concepts.

4. Evaluation: Rigorously evaluate the performance of the artifact using experiments, simulations, analysis etc.
5. Conclusion: Decide whether to iterate back to further development/evaluation or communicate results.

This methodology emphasises the iterative, recursive nature of DSR through proposing, building, evaluating and refining artifacts (Vaishnavi & Kuechler, 2004). However, relevance to application context and theoretical foundations are less prominently incorporated.

DSR Methodology (Hevner et al., 2004)

Hevner et al. defined three interconnected DSR cycles and a set of guidelines to conduct high quality DSR (Hevner, 2007; Hevner et al., 2004):

- Relevance cycle: Connect research to application context.
- Rigor cycle: Ground research in scientific knowledge.
- Design cycle: Iteratively build and evaluate artifacts.

Guidelines: Focus on relevance, rigor, innovation, research contributions etc.

This methodology encapsulates the core elements of effective DSR in relevance, rigor and iterative design. However, the concentric cycle model has been critiqued as overly simplifying complex DSR processes. Table 3-4 depicts a summary of advantages and disadvantages of discussed frameworks.

Methodology	Advantages	Disadvantages
Nunamaker et al. (1990)	• Emphasises system development rigor and evaluation • Incorporates organisational implementation for refinement	• -Lacks explicit relevance cycle connecting research to application context (Iivari, 2015) • Linear rather than iterative focus limits artifact refinement
Peffers et al. (2007)	• Provides clear steps for conducting and disseminating DSR • Emphasises demonstration and evaluation of artifacts • Solution focused to produce artifacts with research contributions	• Lacks explicit relevance and rigor cycles (Iivari, 2015) • Linear sequence of steps rather than iterative cycles
SDRM: Vaishnavi & Kuechler (2004)	• Strong focus on iterative build-evaluate artifact refinement • Allows recursive cycling back to any prior step as needed	• Lacks explicit focus on application context and relevance cycle (Iivari, 2015) • Provides high-level steps but limited procedural details
DSRM: Hevner et al. (2007)	• Concise cycles connecting relevance, rigor and design • Prominent guidelines for high-quality DSR	• Concentric cycle model critiqued as oversimplified abstraction (Piirainen & Gonzalez, 2013) • Lacks procedural specifics within each cycle

Table 3-4 DSR Framework Comparison

3.5.5 DSR Framework Selection

The selection of a framework is dependent on the specific research inquiries and the characteristics of the problem domain (Venable et al., 2017). This study aims to conduct rigorous research on measuring CXM that contributes to knowledge advancement and practical problem-solving. DSRM comprises three stages: describing the challenge and its objectives, designing an acceptable artifact and evaluating the artifact. DSRM prioritises the development of artifacts, resulting in the creation of useful ones. This approach is characterised by flexibility and a variety of processes and may prioritise design process rigour less than other approaches. a formal theory or literature-based design and a formal evaluation process embedded in the design effort are not necessary (Peffer et al., 2018). DSRM involves researchers joining the design or development process at various stages, rather than solely at the beginning. Based on the strengths of DSRM's technical IS artifacts and the findings of Chapter 2, it is a suitable framework for measuring and enhancing CXM.

For this study, Peffers et al.'s methodology provides an appropriate balance of structured process steps and flexibility within an overall DSR framework emphasising relevance, rigorous demonstration, evaluation and dissemination. The sequenced activities guide the production of innovative artifacts with research contributions. Additionally, the iterative DSRM process model proposed by Peffers et al. is widely cited and adopted making it an impactful methodology choice (Inukollu et al., 2014). There are several key reasons why DSRM aligns well with the goals and context of this study:

- Clearly defined process steps: DSRM provides a clear 6-step process comprising problem identification, objectives definition, design and development, demonstration, evaluation and communication. This level of procedural detail provides helpful guidance for carrying out rigorous DSR. Other frameworks can be more conceptual.
- Emphasis on demonstration and evaluation: DSRM places strong emphasis on demonstrating and rigorously evaluating the utility and efficacy of the designed artifacts through well-executed evaluation methods. This ensures the artifacts provide value in context. Some other frameworks focus more on just the build phase.
- Communication of results: The DSRM process concludes with a communication step to convey artifact results and knowledge generated to appropriate audiences. This aligns with goals of contributing insights. Other frameworks may not explicitly include this dissemination step.
- Linear yet flexible approach: While DSRM is linear in its steps, Peffers et al. note research activities can provide feedback loops between steps. This allows applying the framework flexibly and iteratively rather than in a rigid sequence.
- Alignment with this study's goals: DSRM's focus on relevance, rigorous construction, demonstration, evaluation and communication of innovative artifacts aligns well with this study's aims. The iterative approach also suits the goal of incremental artifact refinement.

DSRM's detailed process, focus on rigorous demonstration and evaluation of artifact utility, inclusion of a communication step, flexible iterative approach and alignment with the general objectives of this research make it a well-fitted methodology choice compared to alternative DSR frameworks.

3.5.6 DSR Guidelines

Hevner et al. (2004) proposed a concise yet comprehensive set of guidelines for conducting and evaluating DSR in information systems. These guidelines identify important aspects of high quality DSR.

- ***Design as an Artifact (Guideline 1)***: DSR must produce a viable artifact in the form of a construct, model, method, or instantiation. The artifact encapsulates the research contribution and embodies design knowledge for a specified purpose (March & Smith, 1995).
- ***Problem Relevance (Guideline 2)***: The research should develop technology-based solutions for important business problems. Relevance ensures the artifact produced targets useful application domains and significant opportunities/needs (Hevner & Chatterjee, 2010).
- ***Design Evaluation (Guideline 3):*** The utility, quality and efficacy of the artifact must be rigorously demonstrated using well-executed evaluation methods such as experiments, simulations, case studies, analytics, etc. (Venable et al., 2016).
- ***Research Contributions (Guideline 4)***: DSR should provide clear contributions in terms of the design artifact itself, foundations and /or methodologies. Contributions include the design knowledge embedded in the artifact (Gregor & Hevner, 2013).
- ***Research Rigor (Guideline 5)***: Both artifact construction and evaluation should utilise rigorous techniques and be grounded in the knowledge base. Rigor arises from effective use of prior theory and research (Hevner et al., 2004).
- ***Design as a Search Process (Guideline 6)***: The search for an effective artifact requires leveraging all available means while adhering to laws governing the application environment. Iterative build-evaluate loops enable guided search through the design space (Kuechler & Vaishnavi, 2008).
- ***Communication of Research (Guideline 7)***: DSR must be conveyed effectively to both technical and managerial audiences through crafting an engaging research story grounded in business needs and design principles.

The seven guidelines define essential elements and characteristics of impactful DSR across problem formulation, solution objectives, design and evaluation and knowledge contributions (Iivari, 2015). They provide a roadmap for conducting DSR rooted in relevance and rigour.

3.5.7 DSRM Steps

DSRM, as shown on Figure 3-2, provides a process model for carrying out rigorous and relevant DSR in a structured manner. DSRM comprises six main steps:

- Problem identification and motivation: This involves defining the specific problem to be addressed and justifying the value of an innovative solution artifact. The problem context drives the requirements.
- Objectives definition: Based on the problem, set objectives for the capabilities of the artifact to address the issue and provide value if implemented.
- Design and development: Create and refine the artifact through iterative cycles based on objectives and knowledge of foundations from prior literature.
- Demonstration: Execute artifact to solve problem and showcase its utility in addressing the defined objectives.
- Evaluation: Rigorously assess the quality, efficacy and utility of the artifact using well-executed evaluation methods.
- Communication: Effectively communicate the problem, artifact design process, utility results and implications to appropriate audiences.

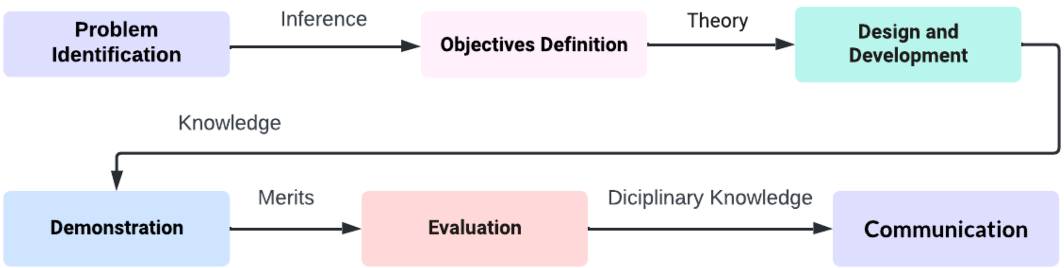


Figure 3-2 DSRM Steps (Peffers et al. 2008)

A key aspect of DSRM is the iterative refinement of artifacts through successive build-evaluate loops. The methodology ensures relevance by grounding artifacts in real problems. Rigor arises from anchoring artifacts in prior knowledge and scientific evaluation. In summary, DSRM provides a structured, rigorous yet flexible methodology for developing innovative artifacts that provide effective solutions to address unsolved organisational or business problems. It aligns the DSR process with practice-inspired problems and research-driven state-of-art solutions.

3.6 Design Theory

Information System Design Theory (ISDT) is a fundamental component of DSR that focuses on developing and articulating principles, guidelines and models for creating effective and innovative information systems. DSR, as a research paradigm, aims to produce practical and relevant artifacts while advancing knowledge in the field. ISDT within DSR contributes to the creation of design theories that guide the construction and evaluation of information systems (Hevner et al., 2004). As Figure 3-3 shows, the contributions of ISDT are multifaceted. ISDT provides a structured pathway from conceptual ideas to impactful information system solutions grounded in business contexts. The iterative nature of ISDT allows for continuous enhancement and adaptation. These contributions enrich the DSR paradigm and information systems field with innovative yet pragmatic solutions (Gregor & Jones, 2007; Markus et al., 2002; Walls et al., 1992; Hevner et al., 2004; Hevner, 2007; Gregor & Hevner, 2013).

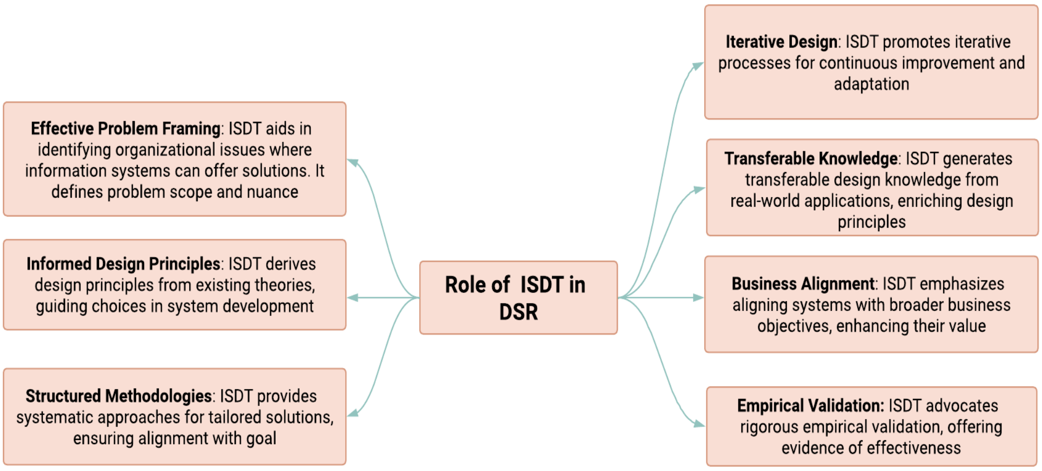


Figure 3-3 Contribution of ISDT to DSR

In 1992, Walls et al. proposed building design theories for vigilant executive information systems that are prescriptive, enable design and action and based on kernel theories from natural and social sciences. This provided methodology rigor for ISDT (Hevner et al., 2004). Gregor and Jones (2007) offered the highly cited anatomy of a design theory, specifying components such as purpose and scope, constructs, principles of form and function, artifact mutability, testable propositions, justificatory knowledge, principles of implementation and expository instantiation (Gregor and Jones, 2007). The design theories that lead to the creation of the design theory in this study are developed using the anatomy of a design theory as defined by Gregor and Jones (2007). The fundamental design theory approaches that can be found in the DSR literature and are presented in Table 3-5 served as the basis for this selection.

Author	Constructs	Principles	Models	Hypotheses	Kernel Theories	Evaluation Criteria
Walls, et al. (1992)	✓			✓	✓	
Markus et al. (2002)	✓	✓				
Gregor & Jones (2007)	✓	✓		✓	✓	
Venable et al. (2016)						✓

Table 3-5 Comparison of Design Theory Approaches

The Gregor and Jones (2007) anatomy is selected for the following advantages that make it well-suited for this DSR project:

- Comprehensive inclusion of constructs, principles, models, hypotheses and theories provides guidance to rigorously build design theories.
- Detailed guidance on theorising supports developing substantive theories firmly grounded in kernel theories.
- Principles of implementation justify artifact construction within scientific foundations.
- Testable propositions and hypotheses enable empirical evaluation and testing of the design theories.
- Flexibility to be specialised for different domains allows developing context-specific design theories.
- Maturity from decades of refinement and use in DSR research provides a robust and solid approach.

The DT steps are explained in the following order:

- **Purpose and Scope**: The theory aims to provide a structured approach for assessing the quality and effectiveness of DSRM activities and outputs (Gregor & Jones, 2007). Its scope covers evaluation of the design process, design artifacts and design outcomes.
- **Constructs:** Key constructs include DSR inputs, capabilities, processes, artifacts, objectives, effectiveness metrics and outcomes. These define the entities involved in DSRM evaluation (Walls et al., 1992; Gregor & Jones, 2007).
- **Principles of Form and Function:** The theory prescribes a three-stage evaluation process mapped to the DSRM cycle (Venable et al., 2016):
    1. Ex ante evaluation assesses plans and inputs before execution using techniques such as requirements analysis and cost-benefit analysis (Sonnenberg & vom Brocke, 2012).
    2. Monitoring involves periodically evaluating execution of design processes and interim artifacts using methods such as prototyping, simulations and functional testing (heng et al., 2011).
    3. Ex post evaluation examines outcomes after DSRM completion relative to objectives using metrices such as system quality, impact and adoption (Venable et al., 2016).
- **Artifact Mutability**: The evaluation model put forth by this theory is itself a mutable artifact that can be adapted to different research contexts (Gregor & Hevner, 2013). Researchers can

modify evaluation criteria, measures, methods and frequency based on resources, project scope and information needs. The mutable evaluation model evolves iteratively.

- **Testable Propositions**: Propositions suggested by Peffers et al. include: Peffers et al. presented a few propositions. The first of them suggests that implementing structured evaluation based on the theory will enhance the quality of documentation (Venable et al., 2017). A further proposition states that adhering to recommended evaluation principles increases the utility of the findings for researchers (Sonnenberg & vom Brocke, 2012). Additionally, they propose that conducting ex ante assessments will help in minimising costly design errors that would require correction afterward (Lee & Hubona, 2009). These propositions establish a connection between the theory and falsifiable hypotheses that can be tested empirically.
- **Justificatory Knowledge:** Justificatory knowledge is a vital component of ISDT. It helps support the scientific evidence that lies behind design decisions and claims (Gregor & Jones, 2007). The knowledge mainly comes from kernel theories that come from reference disciplines. These theories provide models and perspectives that have been tested and are based on previous research (Kuechler & Vaishnavi, 2012). Utilising justificatory knowledge can improve the academic rigour and practical relevance of ISDT. Kernel theories supply conceptual foundations for IS design theories by providing tested models and perspectives from reference fields (Walls et al., 1992). Examples include:
    1. Utility theory (economics) – explains how users maximise utility and make choices (von Neumann & Morgenstern, 1947).
    2. Human-computer interaction theories (psychology) – explain user behaviours and cognitions (Carroll, 1997).
    3. Computer science theories – explain computational processes and system capabilities (Creswell, 2009).
    4. Organisational theories – explain group dynamics and workflows (Leavitt, 1965).
- **Utility Theory** is an example of a kernel theory from economics that can inform ISDT. It analyses how people make choices to maximise satisfaction or "utility" (von Neumann & Morgenstern, 1947). Utility theory is a key kernel theory as it prescribes maximising outcomes relative to resources invested (Creswell, 2009). This aligns with evaluating DSRM return on investment. Utility theory concepts such as marginal utility can also inform prioritisation of evaluation activities and diminishing returns. Multi-attribute utility analysis can support evaluating artifacts on multiple criteria.

In ISDT, utility theory can be applied to model user preferences and assess the utility derived from information system design alternatives (von Neumann & Morgenstern, 1947). Utility theory explains how users make choices to maximise their satisfaction (Keeney & Raiffa, 1993). This can inform ISDT principles related to modelling user preferences during design (Gregor et al., 2020). Multi-attribute utility analysis can also be used to evaluate alternative system designs (Dyer, 2005). This guides creation of optimal systems. For example, utility theory concepts such as diminishing marginal utility, risk preferences and expected utility can shape IS design choices regarding features, interfaces, controls, etc. (Keeney, 2002). Multi-attribute utility analysis can also be used to evaluate competing system designs in terms of their total utility across multiple criteria (Winterfeldt & Edwards, 1986). Kernel theories such as utility theory provide the scientific basis for ISDT and justify the rigor of design decisions (Walls et al., 1992). Utility theory in particular offers useful economic models and perspectives for maximising user satisfaction during information systems design and evaluation (Bailey & Pearson, 1983). Figure 3-4 depicts a clear view of ISDT.

- **Principles of Implementation**: Particular characteristics suggested principles include making sure evaluation timing corresponds to with DSRM stages, using evaluation expertise effectively, focusing on documenting results consistently, selecting appropriate measures for evaluation questions, evaluating artifacts regularly and at different stages and incorporating various data sources such as experiments, field studies, simulations and expert reviews (Venable et al., 2016).

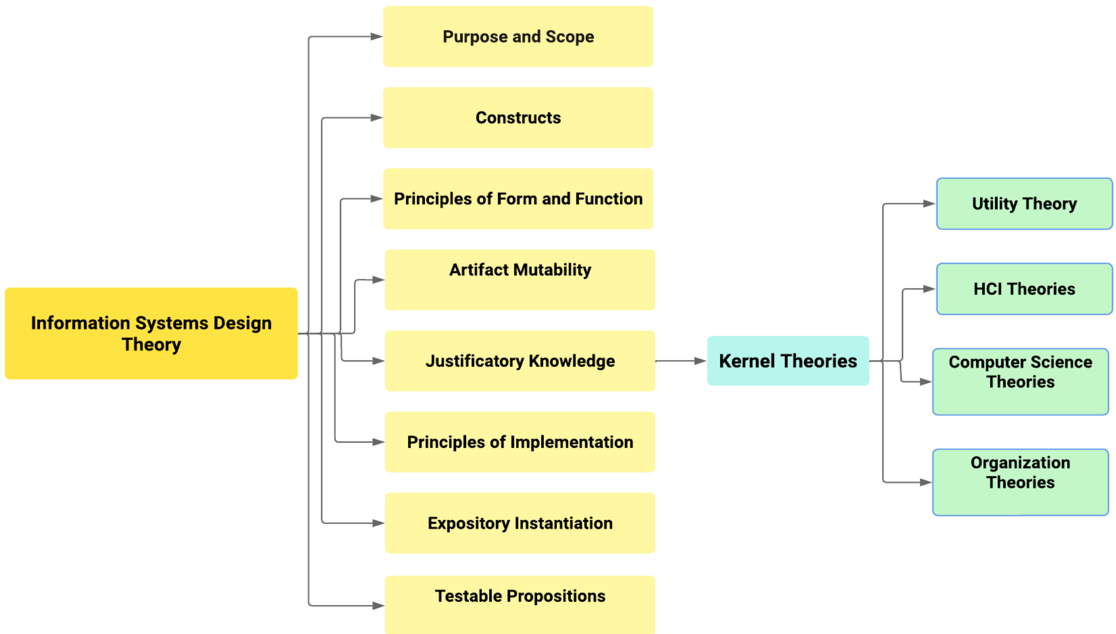


Figure 3-4 Core Components of Design Theory (Gregor & Jones, 2007)

- **Expository Instantiation:** Venable et al. (2016) provide a detailed example applying their evaluation model to a DSR project focused on developing a new e-commerce site. They illustrate executing ex ante assessment of objectives and requirements, monitoring of design processes such as prototyping and ex post evaluation of site quality post-launch. This demonstrates feasibility and provides a template for implementing the theory.

3.7 Research Context

This research aims to develop capabilities for organisations to extract deeper insights into customer preferences and perceptions from unstructured longitudinal textual data. The proliferation of customer feedback data across channels offers immense potential to gather granular customer intelligence to enhance experience management. However, limitations exist in extracting strategic dynamic insights from large volumes of text data using conventional analytics.

This establishes the need for innovative artifacts leveraging emerging techniques such as ABSA , TM and temporal modelling to uncover multidimensional customer insights from unstructured data in a contextualised, timely manner. The research context involves applying and evaluating these text analytics techniques on real-world customer review datasets to demonstrate and validate their utility.

The application domains include video streaming services and music streaming services which offer suitable textual data at scale to rigorously construct and assess artifacts. For instance, the research utilises a dataset of customer feedback reviews of Netflix spanning multiple years. This enables longitudinal analysis of evolving topics and associated emotions.

The knowledge contributions target the intersection of behavioural theories, advanced text analytics and CX research. The artifacts aim to synergise these areas to produce nuanced context-aware, time-aware insights that can inform business strategy and decisions through a deeper understanding of dynamic customer needs. The organisational problem context, application domains selected, datasets leveraged, techniques explored and knowledge contributions focus on developing capabilities for strategic customer intelligence from unstructured CX data using text analytics and temporal modelling. The utility and generalisability of techniques are demonstrated and evaluated on real-world datasets.

3.8 Research Design

An illustration of the design cycles used in the practical approach presented in Figure 3-5. an in-depth explanation of each cycle is presented below.

3.8.1 Iteration 1

The first design cycle focuses on developing an integrated artifact combining TM and ABSA capabilities to enable nuanced analysis of unstructured CX data. This iteration directly tackles enduring below research gaps:

- Lack of holistic CX measurement frameworks (Gap 1) (Poels & Dewitte, 2019).

- Limitations in quantifying subjective dimensions like emotions (Gap 7) (Poels & Dewitte, 2006).
- Need for integrated insights connecting disjointed CX data (Gap 5) (Lemon & Verhoef, 2016).

Artifact Goal

The overarching goal is to conceptualise and implement an analytics framework merging robust statistical modelling via TM (LDA) with purpose-built aspect-based sentiment classification functionality (VADER lexical approach). The integrated artifact aims to address the inability to extract fine-grained insights that relate extracted topics to associated attitudes from massive volumes of unstructured CX data (Griffiths & Steyvers, 2004). Achieving this would in turn improve CX management capabilities by enabling data-driven identification of strengths, weaknesses and opportunities to optimise CX based on customer feedback data.

Artifact Focus

The emphasis is on designing and developing integrated artifacts focused specifically on automated identification of latent topics and related opinions from unstructured textual data. This encompasses judicious selection of appropriate techniques, architectural design covering critical workflows, prototyping key functional modules, demonstrating utility on real-world CX data and evaluating model performance across pertinent metrics.

Design and Development

The key activities involve:

- Conceptual design merging TM and ABSA modules into a unified pipeline for granular CX analytics.
- Selecting technique stacks combining LDA's unsupervised TM with VADER's lexical sentiment classifier.
- Architectural design and modularisation including data pre-processing, extraction-quantification interpretation.
- Prototyping and incremental integration of modular components into an end-to-end functional pipeline.
- Iterative refinements based on preliminary evaluations and CX expert feedback.

Demonstration and Evaluation

Rigorous demonstration takes place through end-to-end application on real-world Netflix CX dataset to extract topics, quantify sentiments and generate fine-grained insights. The artifact undergoes extensive intrinsic performance evaluations using metrics like topic coherence and sentiment accuracy. Both quantitative metrics and qualitative CX expert assessments determine efficacy.

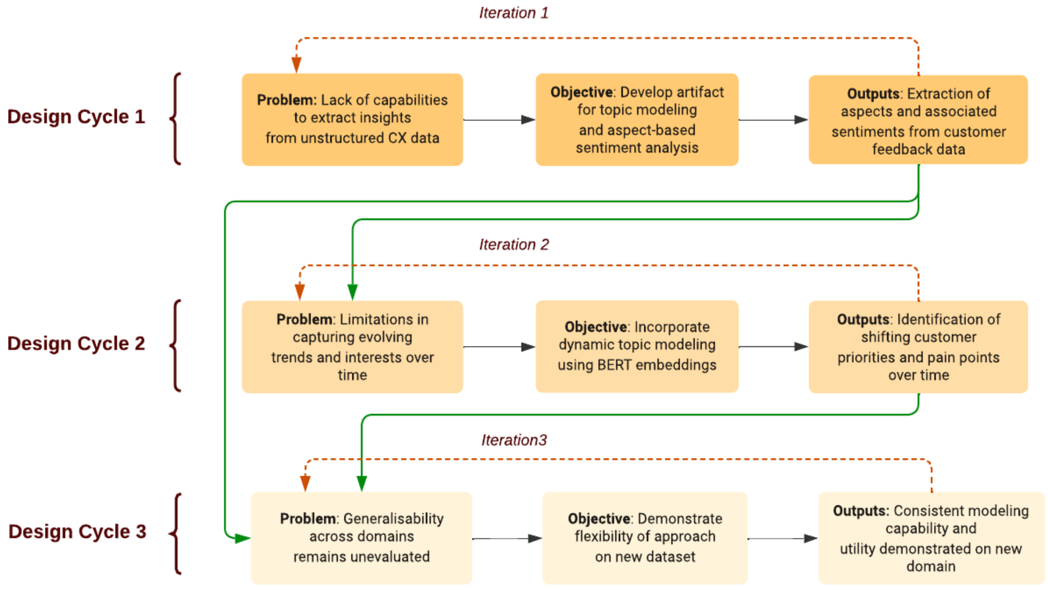


Figure 3-5 The 3 Iterations of the DSR Design Cycle

In summary, Iteration 1 productionises an integrated artifact combining robust TM and purpose-built ABSA techniques to address the research gaps around lack of holistic CX measurement frameworks and integrating subjective dimension insights across disconnected data sources. Demonstration and expert evaluation provide validation.

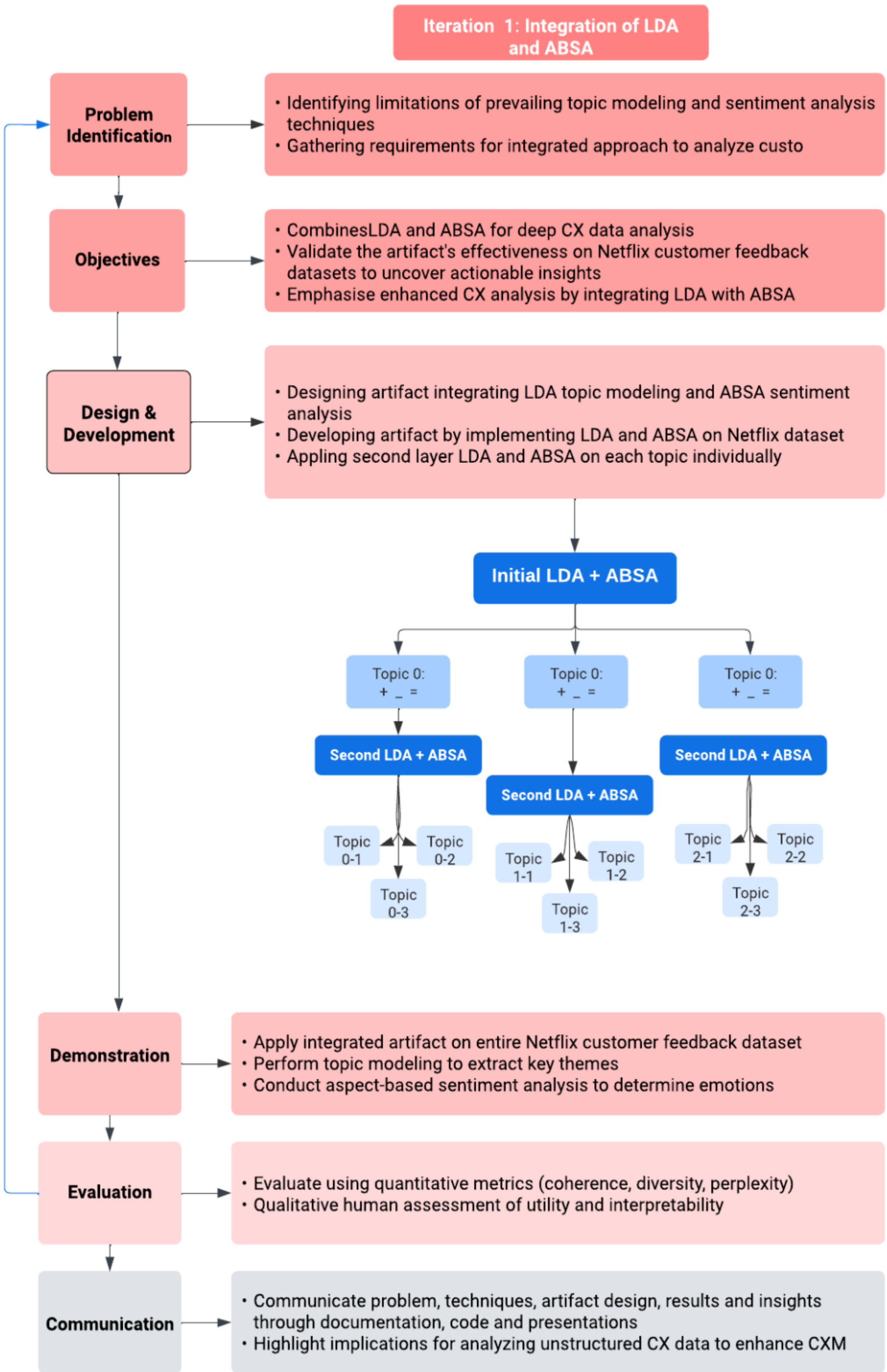


Figure 3-6 DSRM Design Cycle of First Iteration

3.8.2 Iteration 2

The second design cycle addresses enduring below research gaps:

- Inability to track dynamic CX longitudinally (Gap 3) (Blei & Lafferty, 2006)
- Static CX measurement approaches (Gap 9) (Laros & Steenkamp, 2005)
- Tracking semantic drift over time as topics evolve (Gap 14) (DTM gap)

Artifact Goal

This cycle aims to enhance the integrated artifact by incorporating dynamic modelling, semantic knowledge and time awareness through a novel DTM technique leveraging contextualised embeddings coupled with temporally segmented feedback data spanning an 18-month duration. The overarching goal is gaining specific insights into shifts in customer interests, perspectives and preferences overtime to guide continuous CX improvements.

Artifact Focus

The priority shifts to introducing temporality into statistical TM. Key areas of focus encompass ingesting streaming textual data, generating BERT embeddings to capture semantics, conducting temporal sequencing, implementing DTM algorithms to uncover evolution patterns, analysing topic relationships across time periods and evaluating topic stability metrics over the timeline.

Design and Development

The major activities involved:

- Conceptual design of architecture integrating BERT embeddings with DTM modules to enable temporal CX analytics.
- Identifying appropriate technique stacks involving BERT for semantics and BERTopic for dynamic modelling.
- Workflow design across stages of data ingestion, embedding, segmentation, analysis and interpretation.
- Breaking workflow into modular components for key functions.
- Prototyping and integrating modules into an end-to-end temporal analytics pipeline.

Demonstration & Evaluation

Comprehensive demonstration occurs through end-to-end application on the longitudinal Netflix CX dataset spanning 18 months. This encompasses tracking and visualisation of topic evolutions, quantitative analysis of coherence stability over time as well as qualitative assessments by domain experts reviewing outputs. Both technical metrics and subjective human evaluation determine effectiveness.

In summary, Iteration 2 realises a DTM artifact addressing standard static analysis limitations to gain insights into fluid dynamic topics and relationships over time within CX data. Expert evaluation and metrics analysis provide validation.

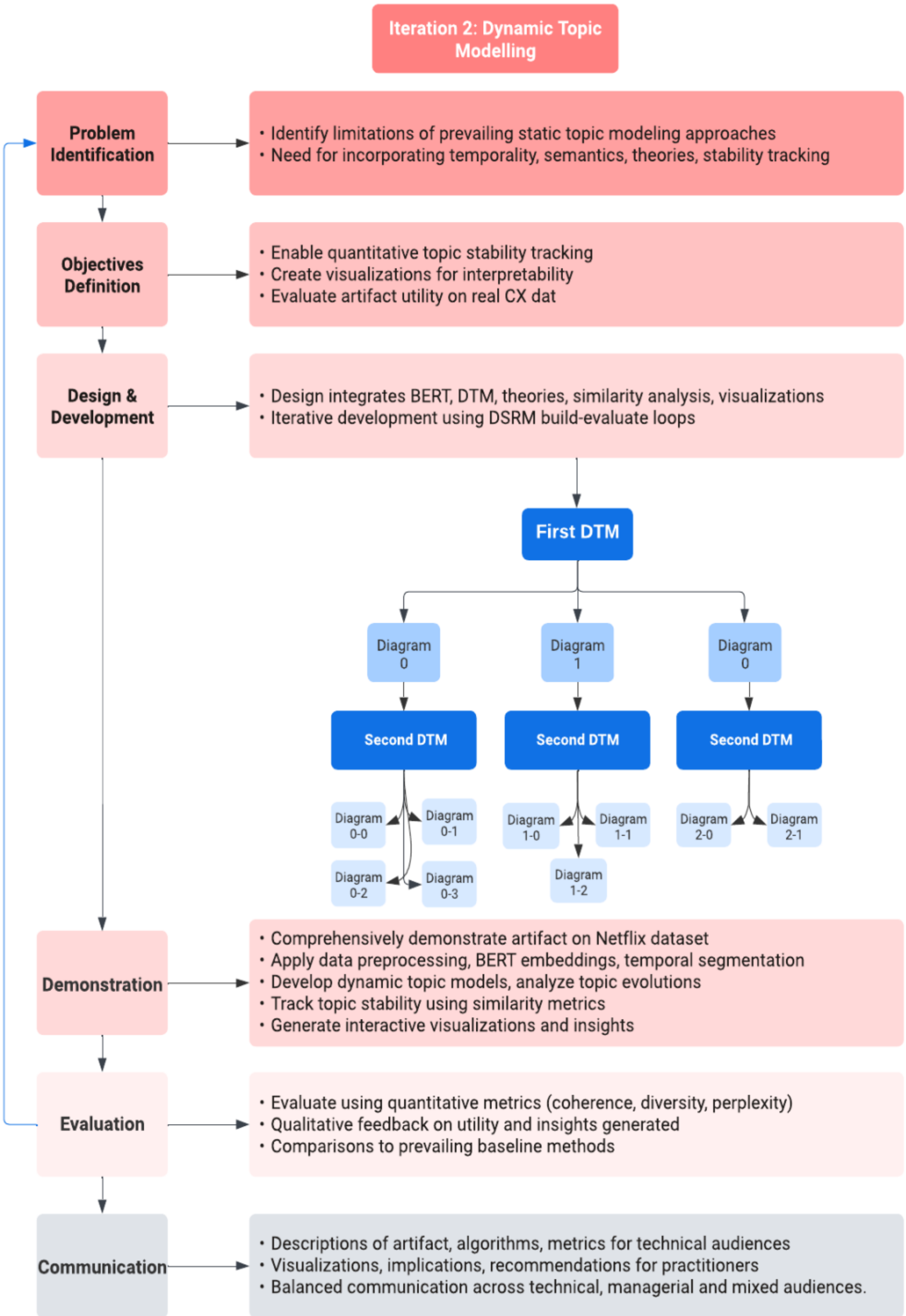


Figure 3-7 DSRM Design Cycle of Second Iteration

3.8.3 Iteration 3

The third design cycle focuses on demonstrating consistent generalisability of the integrated DTM technique across domains, addressing research gaps related to:

- Fragmented CX insights across data silos (Gap 5) (Lemon & Verhoef, 2016).
- Disconnected CX data across organisational silos (Gap 10) (Lemon & Verhoef, 2016).
- Tracking semantic drift over time (Gap 14) (DTM gap).

Artifact Goal

This iteration aims to validate the adaptability, extensibility and flexibility of the integrated artifact encompassing contextual embeddings, dynamic modelling and sentiment analytics across distinct CX application domains. The priority is proving consistent utility through rigorous implementation on the new Spotify music streaming domain.

Artifact Focus

The emphasis is placed on verifying the techniques can be reliably reapplied on the Spotify data to deliver consistent capabilities. This requires focus on activities spanning data acquisition, model optimisation and configuration, adaptable workflow execution, insight extraction, documentation of customisations and evaluation across consistency criteria.

Design & Development

The major steps involve:

- Conceptual design focused on validating generalisability across domains based on prior artifacts.
- Maintaining core techniques like BERT embeddings and BERTopic where applicable.
- Configuring an adaptable architectural workflow for the new dataset.
- Tailoring development efforts to address domain-specific nuances.
- Verifying seamless integration of components on new data.

Demonstration & Evaluation

Rigorous demonstration occurs through end-to-end implementation on the Spotify reviews dataset spanning 18 months. This encompasses comparative benchmarking against Netflix consistency metrics, customisable running across datasets and granular evaluations assessing insights extracted. Both reuse and flexibility determine effectiveness.

In summary, Iteration 3 fulfills the vital intent of proving reliability, adaptability and generalisability vital for establishing substantial credibility and impact of these pioneering techniques for unlocking insights from CX data.

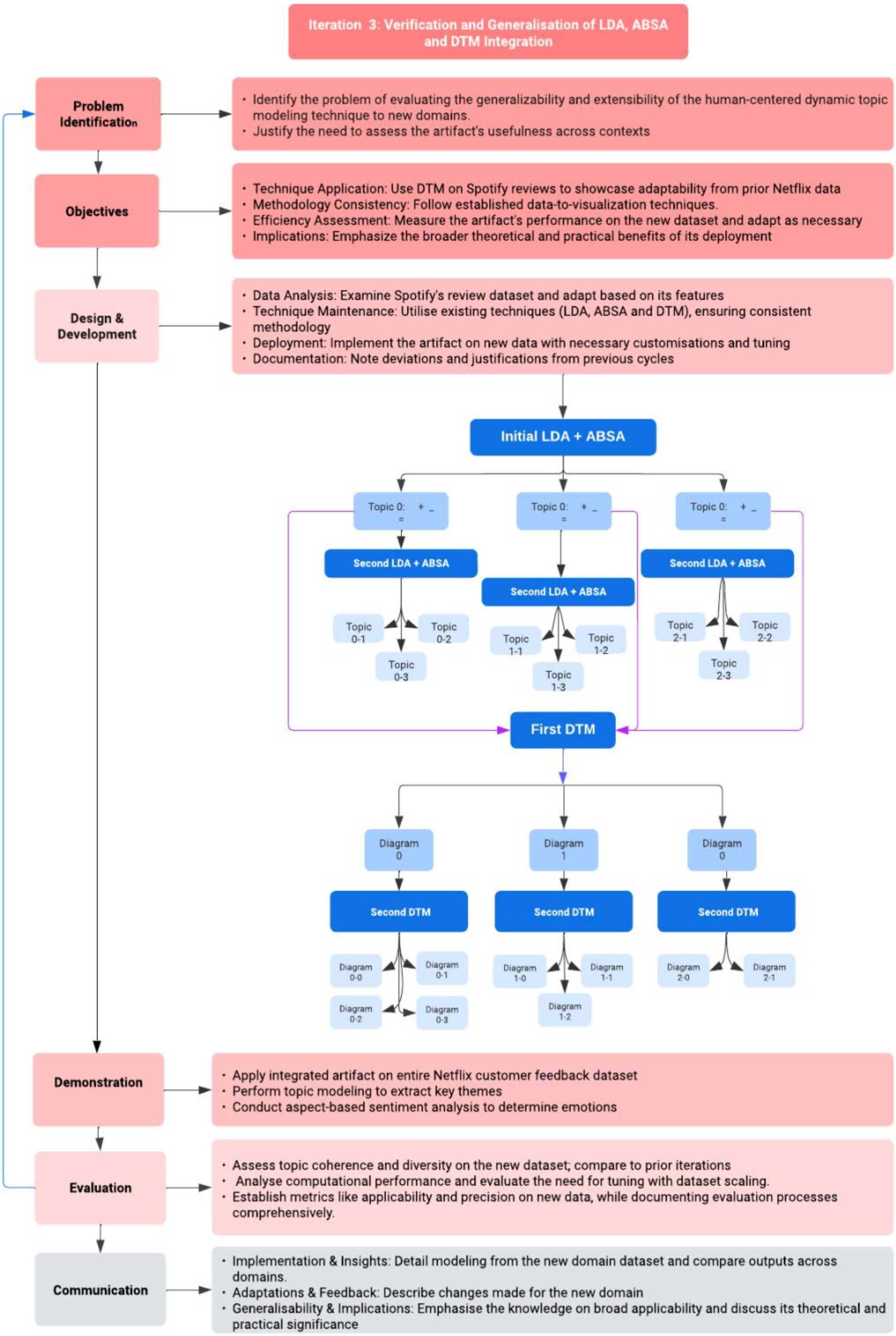


Figure 3-8 DSRM Design Cycle of Third Iteration

3.8.4 Framework Evolution

This research deliberately builds upon the integrated analytics framework through motivated iterations, starting from a static analysis basis before incorporating advanced temporal modelling and validating consistency across domains. Tracing the progression illustrates purposeful augmentation of capabilities over time to address emergent limitations.

As depicted in Table 3-6, Iteration 1 established the foundational architecture fusing robust statistical TM(LDA) with targeted aspect-based sentiment classification powered by the VADER lexicon technique.

Iteration	Foundation	Key Expansions	Capabilities Augmented
1	Topic Modeling (LDA) + Sentiment Analysis (VADER)	-	Static analysis of key topics and associated sentiment
2	Contextual Language Modeling (BERT) + Dynamic Tracking	Temporal awareness through recurrent networks and time-stamped data segmentation	Dynamic trend analysis of evolving topics and relationships
3	Consistency Benchmarking (Similarity Analysis) + Domain Adaptation	Validation across dataset from new domain (Spotify music streaming)	Demonstrated adaptability and utility consistency across domains

Table 3-6 Framework Expansion Across Iterations

While powerful in extracting themes and sentiments from unstructured text, demonstrated limitations around Interpretability and lacking temporal context motivated strategic plans to enrich the framework.

Iteration 2 specifically incorporated recent advances in language representation learning and temporal text mining to transition analysis from static snapshots to continuous dynamic modelling. Techniques encompassed contextualised BERT embeddings to capture semantic meaning and bidirectional long short-term memory networks architected through TensorFlow to unlock evolution patterns from sequentially stamped review data. This dynamic framework addressed key constraints around semantic reasoning and tracking topic changes over time.

However, assessing adaptability and consistency of such capabilities on new datasets remained open. Iteration 3 validated the expandability of the enhanced architecture encompassing sequential neural modelling and aspect-based sentiment quantification on the new Spotify dataset. Benchmarking against prior consistency metrics on the Netflix data proved crucial domain transferability.

Envisioned End-State Architecture

The Conceptual Analytics Framework architecture, presented in Figure 3.9, represents the unified envisioned end-state stack of technical capabilities required for holistic CX analytics. It incorporates leading innovations spanning:

- Statistical TM: Probabilistic latent variable techniques like LDA and neural topic models enable interpretable discovery of themes within unstructured text absent costly manual supervision (Blei et al., 2003; Grootendorst, 2020).
- Contextual Language Understanding: Deep pre-trained language models like BERT, GPT-3 etc. enable encoding semantic relationships between words and documents using self-supervised objectives predictive of human judgment (Devlin et al., 2018).
- Aspect-based Sentiment Mining: Granular opinion mining through lexical methods like VADER and machine learning classifiers facilitate quantifying sentiment towards key aspects as opposed to coarse document-level polarity (Hutto & Gilbert, 2014; Ma et al., 2017).
- Dynamic Temporal Analysis: Architectures like LSTM networks combined with temporal embeddings support continuously evolving signals instead of static snapshots, unlocking trends (Vaswani et al., 2017).

The envisioned unified architecture encompasses staged analytical workflows including:

1. Raw Data Ingestion: Heterogeneous CX signals ingested from scattered enterprise sources through REST APIs and message queues in a decoupled manner.
2. Multi-Modal Signal Pre-processing: Parsing, cleaning, encoding signals leveraging modality-specific techniques - NLP pipelines, audio encoders, video pre-processor etc.
3. Neural Feature Engineering: Deriving high-dimensional semantic feature representations using self-supervised objectives, contrastive techniques and transformer models to capture fluid contextual meanings (Bommasani et al., 2021).
4. CX-Optimised Model Construction: Configuring tailored neural architectures combining required modalities and objectives leveraging frameworks like TensorFlow and PyTorch.
5. Reporting and Diagnostics: Interactive visualisation dashboards delivering analytical results and diagnostics using leading commercial platforms (Redash, Tableau, PowerBI etc.).

The envisioned framework aims to synergise statistical, neural ML and psychological behavioural techniques in a modular, customisable and scalable architecture targeting the extraction of holistic, nuanced and continuous insights across CX data flows.

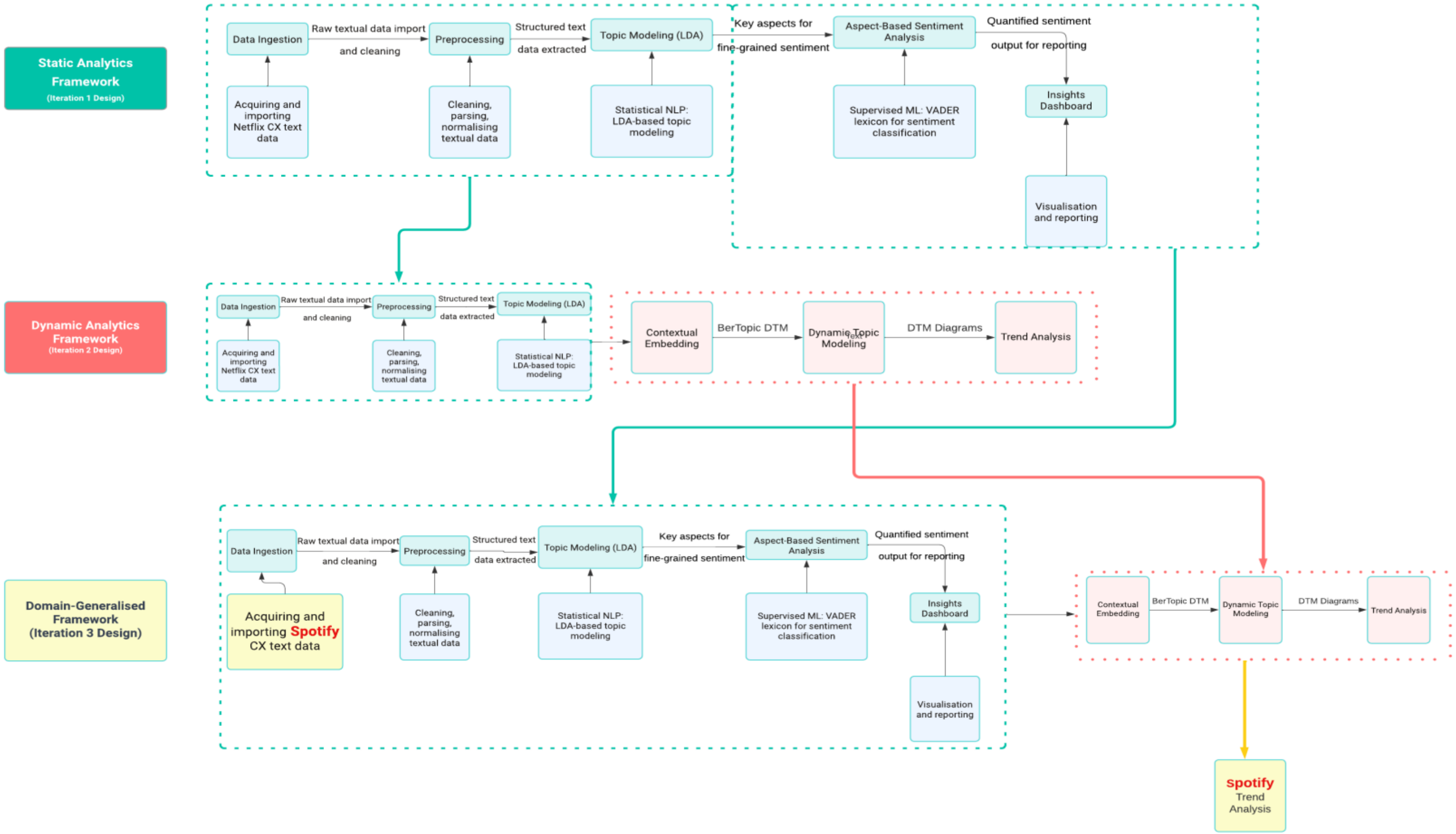


Figure 3-9 Conceptual Analytics Framework

3.9 Summary

Chapter 3 presents the research methodology of DSR adopted in this study. DSR focuses on developing and evaluating innovative technological artifacts that provide solutions to real-world business problems. The research employs the DSRM process proposed by Peffers et al. (2007) to conduct rigorous DSR. DSRM comprises key steps including identifying the problem context, defining objectives for the artifact's capabilities, iterative artifact design and development through build-evaluate cycles, demonstrating utility by applying the artifact to solve problem instances, rigorous evaluation of artifact quality and efficacy and articulating knowledge contributions to audiences. A critical aspect is the design-build-evaluate loop where artifacts are constructed, tested and refined through successive cycles linking relevance and rigor.

The research implements three iterative DSRM cycles to incrementally build artifacts:

1. Integrating TM and ABSA to uncover insights from customer feedback data.
2. Incorporating DTM to enable temporal analysis of evolving topics .
3. Generalising the approach by applying it to music streaming domain data.

In each cycle, the DSRM steps are systematically executed. Rigorous methods demonstrate and evaluate artifact utility. Design theories codify knowledge generated. DSRM provides a structured methodology aligning relevance and rigor to develop impactful artifacts bridging scientific contributions and real-world problem-solving.

Chapter 4: Netflix CX Utilising TM and ABSA

4.1 Overview

This chapter presents the initial iteration of this research following the Design-Science Research Methodology (DSRM) stages of design, build and evaluate (Peffers et al., 2007). The overarching goal is to provide an approach for conducting fine-grained analysis of CX by identifying key themes and topics in Netflix customer feedback using TM and ABSA. This will uncover strengths, flaws and insights to enhance Netflix's service.

- Design (Sections 4.2): Lays out overall design and goals, details TM and sentiment analysis model design.
- Build (Sections 4.3): Implements models to extract topics and sentiments, analyses word distributions, applies hierarchical modelling.
- Evaluate (Sections 4.4): Evaluates using metrics, analyses results for insights, assesses from design theory perspective.

4.2 Tentative Design

The Tentative Design phase for Chapter 4 focused on conceptualising an integrated artifact using TM and ABSA to analyse Netflix customer feedback data. The design goals involved identifying key topics within Netflix reviews and determining associated sentiment towards each topic. LDA was proposed as the TM technique to discover topics based on word co-occurrence patterns. For sentiment analysis, VADER lexical approach was designed to classify topic sentiment polarity. The tentative framework incorporated sequential data pre-processing, LDA modelling, topic-aspect mapping, sentiment scoring with VADER and evaluation modules. This architecture was theorised to enable granular measurement of customer sentiment across topics by synthesising AI and NLP techniques tailored to the CX context. The proposed layered TM + ABSA design guided subsequent implementation and evaluation in extracting nuanced insights from unstructured Netflix review data.

4.2.1 Problem Definition and Objectives for Iteration 1

This work aims to address key gaps from Chapter 2 (literature review):

- Lack of comprehensive CX frameworks and measurement tools to pinpoint key themes in to uncover customer insights (Gap 1 and 6).
- Difficulties quantifying subjective CX perceptions and emotions (Gap 2 and 7).

- Limited exploration of integrated TM + ABSA for mining third-party (Netflix) CX feedback (Liu et al., 2019).

TM approaches are utilised to identify CX dimensions from a dataset. The primary aim of ABSA is to comprehend the sentiment linked with various dimensions of CX. The LDA model for the TM approach was applied to the dataset.

4.2.2 Foundational Notions

The integrated framework drew on two key foundational techniques for TM and ABSA:

- Latent Dirichlet Allocation (LDA) for TM: LDA is a generative statistical model that uses word co-occurrence patterns to discover latent topics in a collection of documents (Blei et al., 2003). It represents documents as mixtures of topics, where each topic is a distribution over words (Chen et al., 2013). LDA has emerged as a widely used TM technique due to its statistical foundations, modelling flexibility and interpretability (Debortoli et al., 2016).
- VADER for ABSA: The Valence Aware Dictionary and Sentiment Reasoner (VADER) is a lexicon-based model tailored specifically for social media sentiment analysis (Hutto & Gilbert, 2014). It uses a rule-based approach with heuristics for capturing sentiment intensity and contextual valence shifters in informal text (Tan et al., 2020). VADER has demonstrated reliable performance across domains including online reviews and social feedback (Mohammad & Turney, 2013).

The integration of robust statistical modelling via LDA with purpose-built sentiment classification through VADER provided foundational capabilities tailored to the context of analysing unstructured CX feedback data (García-Moya et al., 2013).

4.2.3 Potential Utility

The potential utility of the integrated TM and ABSA techniques was to enable more holistic analysis of unstructured CX data to derive actionable insights (Lee & Bradlow, 2011). By combining LDA modelling to uncover key topics with VADER sentiment analysis of each topic, nuanced customer opinions and preferences could be quantified from open-ended feedback (García-Moya et al., 2013). This granular mixed-methods approach could empower CX managers to identify satisfiers and pain points linked to specific topics (Liu et al., 2019). The utility lies in transforming subjective text data into analysable topic and sentiment distributions to guide enhancements (Blei, 2012). Realised benefits would include optimising CX spending based on data-driven insights and objectively tracking impact over time (Coussement & Van den Poel, 2008).

4.2.4 Theoretical Premises

The conceptual architecture integrated foundations from natural language processing, machine learning and CX analytics. LDA provided a robust statistical framework for inductively learning latent topics through word co-occurrence patterns (Blei et al., 2003). VADER offered a domain-optimised, rule-based model for accurately classifying sentiment in informal text (Hutto & Gilbert, 2014). Customer theory reinforced tailoring the analysis to feedback specific to service experiences (Meyer & Schwager, 2007). Integrating these premises enabled suitably extracting multifaceted CX insights (Griffiths & Steyvers, 2004). The process involved data processing, analysis modules, results interpretation and continuous enhancements (Khan et al., 2018). The synthesis of technical capabilities and CX theories provided rigorous grounding.

4.3 Build

The Build stage focuses on implementing the LDA and VADER sentiment analysis models designed in prior sections on the Netflix dataset. It conducts the topic-aspect mapping, fitting the LDA model to the document-term matrix to extract key topics. It then builds on this by analysing word distributions and visualisations to further understand the TM outputs. It also implements a second layer of LDA on specific subsets related to each key topic like Customer Service. This hierarchical modelling provides more nuanced subtopic insights. Overall, the Build stage puts the artifact into action by executing the models on the real-world Netflix data corpus to derive tangible topics and associated sentiment distributions. This transitions the design into practical implementation and analysis outputs.

4.3.1 Netflix Dataset Introduction

Netflix has undergone rapid and substantial evolution throughout its twenty-year history. The company that initially distributed films on DVD via mail in the US is now recognised as a worldwide video service. The claim of Netflix being "global" is disputed by both film and media scholars and audiences. Netflix is a valuable example for analysing the benefits of internet-distributed video that are more widely recognised internationally than through national services. The Netflix case allows for the integration of conventional screen studies expertise with broader discussions on digital distribution, platforms and algorithmic culture. These discussions vary across national and disciplinary borders.

The company that initially distributed films on DVD through mail in the US is now recognised as a worldwide video service. The assertion of Netflix's global status is frequently contested by both film and media scholars and audiences. Netflix's policy of reporting subscriber numbers only for the US or international subscribers makes it challenging to gauge the service's prevalence outside of the United States. There are significant regional differences in Netflix's market share, brand recognition, cultural standing and catalogue. Media scholars face empirical and conceptual challenges when studying a video service that is experienced differently in each country. Trustpilot is beneficial as a source for building a customer feedback database due to multiple reasons:

- **Large Number of Reviews**: Trustpilot provides a large number of user-generated reviews covering a wide range of organisations and industries. This facilitates the collection of substantial customer feedback, offering an extensive data set for analysis.
- **Diverse Customer Perspectives**: Trustpilot receives reviews from a wide spectrum of customers, spanning various demographics, locations and experiences. Diversity can enhance the comprehension of customer sentiment and preferences.
- **Authenticity and Transparency:** Trustpilot confirms that reviewers are legitimate consumers, decreasing the possibility of falsified or biased ratings. Moreover, the platform enables companies to respond, encouraging transparency and facilitating productive communication between businesses and consumers.
- **Review Structure and Metadata**: Structured information such as star ratings, categories and timestamps are frequently included in Trustpilot reviews. Structured data facilitates feedback organisation and categorisation, leading to more efficient analysis and insight extraction.
- **Comparative Analysis**: Trustpilot contains reviews for a number of companies in the same industry, making it appropriate for a comparison study. Comparing customer feedback across competitors enables researchers to gain competitive intelligence by identifying strengths and weaknesses.
- **Accessibility and ease of data collection:** Trustpilot reviews are publicly available and can be accessed using APIs or web scraping techniques. This accessibility makes it easier to gather and combine consumer reviews into a database for further analysis.

However, there are several limitations that should be considered when using Trustpilot as a customer feedback database. Trustpilot reviews may not represent the total customer base given that people with more extreme opinions are much more likely to leave online feedback (Hennig-Thurau et al., 2004). There may also be biases in the statistics, since people who have had particularly bad experiences may be more motivated to share their feedback (Luca, 2011). Merging Trustpilot data with other consumer feedback sources could provide a more thorough picture and help overcome Trustpilot's limitations (Garcia et al., 2017).

The dataset comprises feedback comments willingly provided by customers in the United Kingdom. The data was collected from March 01, 2022, to August 23, 2023, using web scraping techniques. Customer data for this study has been collected through web scraping, which offers advantages including increased efficiency and reduced time and effort required for gathering customer feedback.

Collecting feedback from multiple sources and platforms enhances the comprehensiveness and representativeness of the dataset for analysis.

4.3.2 Data Pre-processing

The dataset was pre-processed using Python due to its widespread use in data analysis and availability of standard statistical packages. The packages "`pandas`" and "`nltk`" were utilised for pertinent NLP capabilities. Other useful packages are shown in Figure 4-1.

In [1]:

```
import json
import pandas as pd
import numpy as np
import re
import sys
import nltk
from nltk.corpus import stopwords, sentiwordnet as swn
from nltk.stem import WordNetLemmatizer
```

Figure 4-1 Dataset Pre-Processing libraries

Natural Language Toolkit (NLTK) is a well-known platform for developing Python applications that process human language data. It includes easy-to-use interfaces to over 50 corpora and lexical resources, such as WordNet, as well as a suite of text processing libraries for classification, tokenization, stemming, tagging, parsing and semantic reasoning, wrappers for NLP libraries and an active discussion forum.

The dataset for this study has undergone pre-processing, which involved the following steps:

- Lowercasing the text.
- deleting all punctuation and extra white space.
- Tokenisation: splitting a large text into smaller lines and words.
- Stop words removal: Eliminating frequently-used words across the corpus such as articles, adverbs. After thorough research and visual inspection of the dataset, a lengthy list of Netflix-specific stop words was found and added to the "`nltk`" built-in list. Lemmatisation and stemming: `nltk` algorithmically determines a word's lemma by analysing its meaning and context. Lemmatization is a process that involves morphological analysis to eliminate inflectional endings from words. It retrieves the lemma of a given word. Stemming is a process applied in NLP to normalise words. It transforms the words of a sentence into a sequence that simplifies retrieval. This method normalises synonymous words across varying contexts and sentences.

The operations listed above contribute to dataset consistency. Model derivation requires consistency across the entire dataset. The pre-processing steps as shown in Table 4-1, have prepared the dataset for TM using ML methods, as explained in the following subsection.

	Review	remove_lower_punct	tokenise	remove_stopwords	lemmatise
175	You're going to lose more customer if you deci...	you re going to lose more customer if you deci...	[you, re, going, to, lose, more, customer, if,...	[going, lose, customer, decide, air, monstrosity]	[going, lose, customer, decide, air, monstrosity]
1753	There are many TV series and movies that you c...	there are many tv series and movies that you c...	[there, are, many, tv, series, and, movies, th...	[tv, series, movies, choose]	[tv, series, movie, choose]
1754	It is good service"\n	it is good service	[it, is, good, service]	[good, service]	[good, service]
2093	I'm a scary gargoyle on a tower\n	im a scary gargoyle on a tower	[im, a, scary, gargoyle, on, a, tower]	[scary, gargoyle, tower]	[scary, gargoyle, tower]
2094	That you made with plastic power\n	that you made with plastic power	[that, you, made, with, plastic, power]	[plastic, power]	[plastic, power]
2095	Your rhinestone eyes are like factories far aw...	your rhinestone eyes are like factories far away	[your, rhinestone, eyes, are, like, factories,...	[rhinestone, eyes, factories]	[rhinestone, eye, factory]
2112	Your love's like rhinestones falling from the ...	your loves like rhinestones falling from the sky	[your, loves, like, rhinestones, falling, from...	[loves, rhinestones, falling, sky]	[love, rhinestone, falling, sky]
2736	Always find something to watch"\n	always find something to watch	[always, find, something, to, watch]	[find, watch]	[find, watch]
3226	How you can take down the Christmas movies jus...	how you can take down the christmas movies jus...	[how, you, can, take, down, the, christmas, mo...	[christmas, movies, christmas]	[christmas, movie, christmas]
3520	Error after error. It seems doggy to me.\n	error after error it seems doggy to me	[error, after, error, it, seems, doggy, to, me]	[error, error, doggy]	[error, error, doggy]

Table 4-1 A screenshot from pre-processing steps

4.3.3 Topic Modelling

The first step in developing the TM is emphasising that machines don't inherently understand text data. Thus, text must be converted to numerical representations before initialising the TM model. Countvectorizer is a text-to-numbers converter. Before the countvectorizer, Ngram models are computed to predict likely next words based on preceding terms. This study utilised bigrams and trigrams for optimal topic detection through the "genism" Python library, focusing exclusively on "NOUN" and "VERB" to boost detection. The second phase involves creating the corpus dictionary using id2word for compatibility across libraries and models. TM is implemented using Python programming language using packages, as shown in Figure 4-2, `nltk`, `genism` and `sklearn`.

```
import json
import pandas as pd
import numpy as np
import re
import sys
import nltk
from nltk.corpus import stopwords, sentiwordnet as swn
from nltk.stem import WordNetLemmatizer
from nltk import ngrams
from sklearn.feature_extraction.text import CountVectorizer
from sklearn.decomposition import LatentDirichletAllocation
import collections
import gensim
import gensim.corpora as corpora
from gensim.utils import simple_preprocess
from gensim.models import CoherenceModel
import pyLDAvis.gensim_models
import matplotlib.pyplot as plt
```

Figure 4-2 List of Applied Packages to Develop LDA

The last stage of creating TM is using gensim. Determining optimal topic number is challenging but "Perplexity" and "Coherence Score" help reduce complexity. Coherence evaluates semantic similarity of terms within topics. Perplexity assesses model efficacy in predicting samples. The goal is choosing the topic number with lowest perplexity and highest coherence.

Table 4-2 shows 3 topics as optimal based on maximum coherence and minimum perplexity. The next step fits the LDA model to the document-term matrix from CountVectorizer to catalogue CX themes for sentiment analysis. Gibbs sampling facilitates finding observations from the probability distribution. The argmax function identifies the dominant topic for each document by selecting the highest probability topic (Blei et al., 2003). This automates dominant topic allocation, making the process quick and easy.

Number of topics	Coherence Score	Perplexity
2	0.48278	6.9193
3	0.52462	6.9830
4	0.5003	7.0911
5	0.4331	7.2007
6	0.4117	7.24524
7	0.42642	7.3209
8	0.4099	7.3850
9	0.4383	7.4612
10	0.3939	7.6030
11	0.4056	7.8738
12	0.3813	8.24389

Table 4-2 The Summary of Coherence and Perplexity score for the range of Topic Numbers

4.3.4 Topic-Aspect-Sentiment Relationships

The integrated analytical framework developed through combining TM and ABSA) techniques enables capturing rich inter-relationships between three key elements – Topics, Aspects and Sentiments.

Topics refer to overarching themes present within textual content like customer reviews. TM approaches identify topics based on clustering words that commonly co-occur across documents through statistical relationships (Blei et al., 2003).

These topics encompass various fine-grained attributes or aspects that compose the broader subject. For example, a key topic around "Account Management" contains related aspects like "Billing", "Payments" and "Subscription Management". Aspects represent key facets within an overarching topic.

Finally, sentiments capture subjective opinions, attitudes and emotions towards specific aspects rather than entire documents or topics. Sentiment orientation can be positive, negative or neutral for each aspect. Figure 4-3 illustrates how topics, aspects and sentiments are analytically connected.

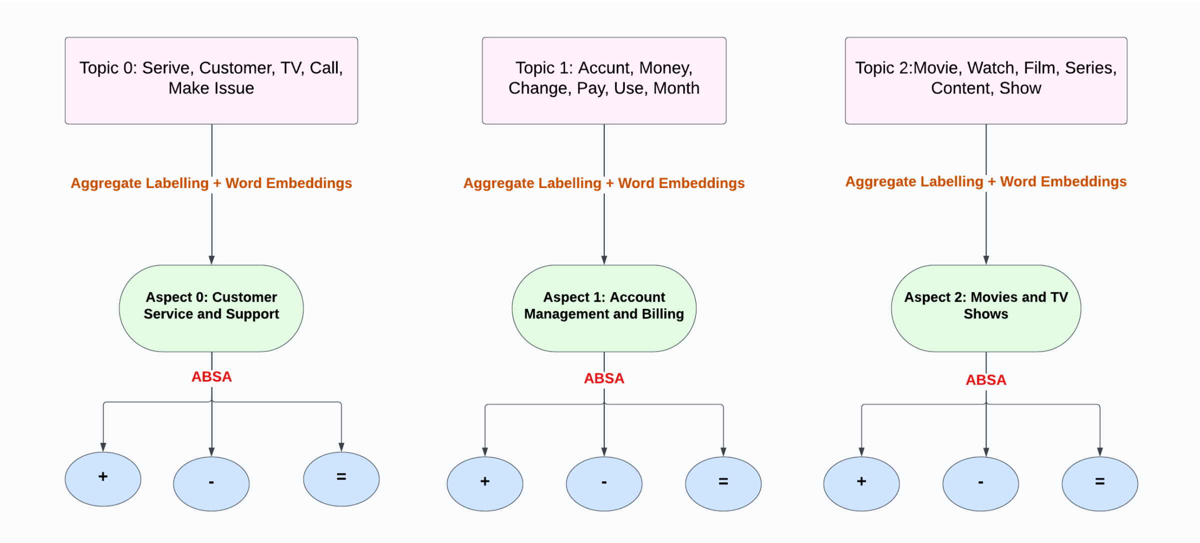


Figure 4-3 Topic-Aspect-Sentiment Hierarchy

Key relationships underpinning this hierarchy are:

1. Topics identified via TM provide contextual foundation for relevant aspects residing within that subject area.
2. Specific aspects/attributes are extracted based on analysis of words and patterns constituting those broader topics.
3. ABSA determines the fine-grained sentiment polarity attached to each specific aspect rather than topics overall.

For example, under the "Account Management" topic, related aspects around managing billing and payments are extracted through hierarchical analysis of topic patterns. ABSA then detects the sentiment orientation of positive, negative or neutral associated with those distinct billing and payment aspects based on the subjective vocabulary used in context.

This layered synthesis connects emerging topics to attributes shaping perceptions before quantifying nuanced opinions tied to those attributes. The integrated topic-aspect-sentiment hierarchy delivers multidimensional analytics not possible using individual approaches in isolation.

4.3.5 Topic-Aspect Mapping

LDA modelling has generated three topics based on the optimal number detected. The analysis began by examining a minimum of two topics and gradually increasing them to 12 to observe the behaviour of the model as the number of topics increased. Growing the number of topics resulted in considerable overlap between them and a more cluttered representation of the dataset. These factors made it challenging to manage the topics effectively. Thus, acquiring three topics from the database was found to be an optimal number. Furthermore, as shown in Table 4-3, there was only one instance where the coherence score and perplexity intersected.

Figure 4-4 displays the intertopic distance map, indicating that there is no interconnectivity between the three distant and distinct topics. Table 4-4 shows the crude outcome of the LDA which then go through labelling process.

Topic Number	Topic Corpus
0	0.071*"movie" + 0.061*"watch" + 0.027*"film" + 0.023*"series" + ' '0.022*"content" + 0.022*"show" + 0.020*"want" + 0.018*"see" + 0.017*"find" ' '+ 0.017*"netflix"'
1	0.046*"account" + 0.034*"cancel" + 0.031*"use" + 0.029*"month" + ' '0.022*"take" + 0.021*"money" + 0.020*"charge" + 0.018*"pay" + 0.018*"say" + ' '0.018*"day"
2	0.059*"service" + 0.038*"customer" + 0.028*"show" + 0.023*"tv" + 0.021*"go" ' '+ 0.019*"get" + 0.017*"time" + 0.017*"call" + 0.014*"make" + 0.014*"issue"

Table 4-3 Results of the LDA Topic Modelling

Figure 4-4 Intertopic Distance Map of the LDA model Dominant Topic

In the realm of TM, such as in the case of LDA, it's essential to assign a dominant topic for effectively analysing and comprehending an assortment of documents. This involves determining the most representative words within them and assigning fitting labels that are brief but informative regarding their primary theme or concept. When it comes to labelling LDA topics, the roles that representative words and relevance scores play in the process are crucial. The ways in which they contribute are as follows:

- Representative Words: The top words within a particular topic based on their probabilities or weights are referred to as representative words. These words provide an initial understanding of the main subject or theme behind each topic. By identifying salient terms associated with each topical category, these representative keywords become valuable tools at helping individuals understand the specificities of a given classification. They serve as starting points for creating strong labels while providing insights essential for capturing and communicating effectively about subject matter.
- Relevance Scores: These indicate how important individual words are when compared to one another within information categories using probability weightings assigned to specific groups of data. Calculated considering relative importance among distinct variables help clarify major implications weighting various aspects along varying scales. Thus, Relevance Scores assign differing levels worth when compared against other factors inside singular topics themselves be applied assigning priority internal certain tagging techniques relevant sizeable associations.
- Labels & Impact It is judicious and considered opinion worth keeping both Representative Words and relevance scores offer powerful leverages over how accurately subjects within machine processed material sussed out—each fundamental category provides sophistication

needed balance call outs concerning recognition components feeding divergent key concepts included high level directives is now made possible.

Topic Number 2										
Words	Movie	Watch	Film	Series	Content	Show	Want	See	Find	Netflix
Representative Score	0.071	0.061	0.027	0.023	0.022	0.022	0.020	0.018	0.017	0.017
Relevance Score	0.236	0.203	0.089	0.076	0.074	0.074	0.068	0.060	0.058	0.057
Topic Number 1										
Words	Account	Cancel	Use	Month	Take	Money	Charge	Pay	Say	Day
Representative Score	0.046	0.034	0.031	0.029	0.022	0.021	0.020	0.018	0.018	0.018
Relevance Score	0.177	0.133	0.120	0.113	0.085	0.083	0.076	0.069	0.069	0.069
Topic Number 0										
Words	Service	Customer	Show	TV	Go	Get	Time	Call	Make	Issue
Representative Score	0.059	0.038	0.028	0.023	0.021	0.019	0.017	0.017	0.014	0.014
Relevance Score	0.236	0.153	0.111	0.093	0.084	0.074	0.068	0.066	0.056	0.55

Table 4-4 The Representative and Relevance Score for the Three Topics

Through the LDA outcome provided, three topics with their respective words and representative scores have been identified along with relevance scores to further supplement understanding. The information presented in Table 4-4 can be interpreted in the following manner:

- **Topic Number 0**: One of the topics that stood out in LDA analysis is related to customer service and support. Some of the words that feature prominently in this category include TV, call, time, get, show as well as service and customer. Looking at the representative scores assigned by LDA analysis it's clear "service" scored highest with a value of 0.059 underlining its importance in this topic. A strong association between these keywords and Topic Number 2 was also shown in relevance scores where "service" had a high score of 0.236. This finding entails discussions or content focused on customer welfare including addressing their concerns and offering support services for their needs through resolving any issues they encounter involving products purchased or supplied to them; hence "customer," "call," and "issue".
- **Topic Number 1**: The second subject of interest identified by the LDA analysis is focused on the management of accounts and billing. Certain words in this topic, such as "account," "cancel," and "pay" suggest an association with various account-related actions or financial aspects. Additionally, this topic scores 'account' with a representative score of 0.046, indicating

its significance among the groupings. Furthermore, it also reflects on discussion or information that concerns h and ling any billing-related issues; particularly highlighters are words such as "charge," month," and day." It's clear that Topic Number 1 pertains to customer engagement around maintaining proper account levels while adhering to timely payments for services used.

- **Topic Number 2**: The following words have been assigned as representative words: Movie, Watch, Film, Series, Content, Show, Want, See as well as Find and Netflix. Representative score signifies each word's relative importance within the complete contents of this topic whereby a higher score indicates more significant relevance such as how "Movie" has obtained a representative score of 0.071 meant it's one of the most crucial pieces belonging to this subject matter. On the other hand, relevance scores indicate how strongly each word is affiliated with this topic such that high associates closely whereas low underscores outlier cases As an example "Movie" at 0.236 highlights its enormous significance in relation to Topic no. 1. Similarly, then both Numbers 1 and 2 categorised into distinct categories exhibit familiarity within elements having same structures persisting consisting of representative wards given by relevance scores contributing towards an enhanced comprehension.

Topic Labelling: Assigning concise, descriptive labels is a major step in interpreting topics extracted through modelling techniques such as LDA. While topic models discover latent semantic structures based on word usage patterns, human judgment is crucial for accurately encapsulating topic themes in memorable labels. This research adopts a rigorous conjunction of Aggregate Labelling and Word Embeddings techniques to properly label the topics.

- ***Aggregate Labelling*** refers to the process of assigning categorical labels to groups of instances based on the aggregation of their basic features (Hwang et al., 2011). For text data, this involves identifying words that are frequent or probable within a subset of documents and using those words as a descriptive label for that document subset (Hwang et al., 2011). This provides a straightforward way to create topic-such as labels from the words common in related documents. However, a limitation is that these labels have no inherent semantic meaning.

- ***Word Embeddings*** are dense vector representations of words that encode semantic information (Mikolov et al., 2013). Through neural networks trained on large corpora, words with similar meanings are mapped to similar vector representations. This allows semantic relationships between words to be captured based on their distributions in textual data (Mikolov et al., 2013). As a result, word embeddings provide an effective way to represent the semantic meaning of words and sets of words.

Combining Aggregate Labelling and Word Embeddings: Combining aggregate labelling and word embeddings is a promising way to label topics. Aggregate labelling provides descriptive labels, but it does not include semantic information. Ganguly et al. (2015) demonstrated that Word embeddings fill this gap by providing deep semantic knowledge about words and texts. This synthesis allows for labelling texts with relevant topics, going beyond just using keywords. The aggregate labels provide descriptive names for topic clusters, while the word embeddings enable semantic comparison between texts and topics (Ganguly et al., 2015). This as presented in Figure 4-5 approach combines the advantages of both techniques - clear topic names from aggregate labelling and the use of semantic knowledge from word embeddings - to accurately label documents with relevant topics.

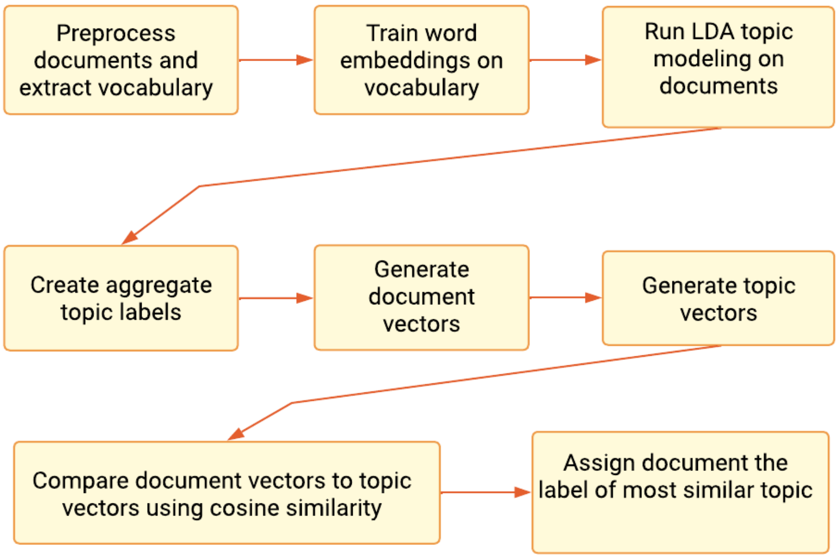


Figure 4-5 Step-by-Step Guideline of Topic Labelling

the output of LDA presents three evident topics consisting of *Customer Service and Support, Account Management and Billing and Movies and TV Shows.* Each subject is defined by its terminology, while its representative and relevance scores provide critical insights into how they relate to their respective discussions. Consequently, these results strongly suggest that all content related to movies and series, account management and billing as well as customer service support has been covered in the dataset. Table 4-5 depicts an accurate representation of the dominant topic in 10 first individual feedback, as well as its contribution from the other two less dominant topics. However, ABSA with external validation can enrich the understanding of the domain in the current study.

	Document_No	Dominant_Topic	Topic_Perc_Contrib	Keywords	Text
0	0	2.0	0.5	movie, watch, tv, series, price, find, thing, ...	[review]
1	1	0.0	0.8	service, customer, account, call, tell, issue,...	[send, phise, message, video, wait, source, sp...
2	2	2.0	0.5	movie, watch, tv, series, price, find, thing, ...	[price, year]
3	3	1.0	0.5	cancel, month, pay, film, time, money, content...	[account, money, pay, week, account, suspend, ...
4	4	1.0	0.8	cancel, month, pay, film, time, money, content...	[pay, week, service, pay, month, pay, week, pa...
5	5	1.0	0.4	cancel, month, pay, film, time, money, content...	[week, process, money, instant]
6	6	2.0	0.4	movie, watch, tv, series, price, find, thing, ...	[selection, count, time, watch, create, tiktok...
7	7	2.0	0.4	movie, watch, tv, series, price, find, thing, ...	[money, company, care, quality, buck]
8	8	0.0	0.5	service, customer, account, call, tell, issue,...	[customer, request, money, time, movie, provid...
9	9	2.0	0.5	movie, watch, tv, series, price, find, thing, ...	[price, policy, ceo, person, watch, person, wa...

Table 4-5 Distribution of The Dominant Topic in Individual Feedback

4.3.6 Word Count Distribution and Visualisation

A comprehensive understanding of the dominant topics in the dataset is possible through the use of both word clouds and frequency word distribution. On one hand, word clouds present an illustrative outline which depicts relevant terminology whilst on the other hand, frequency word measures the significance of each word. Combining these two methods grants the ability to recognise key concepts, extract knowledge and make educated conclusions centred around customers' preferences and concerns.

the first dominant subject deals entirely with Customer Service and Support; its use of 'customer,' 'service,' 'call' and 'issue' show people communicating within discussions relating to experiences involving customer support operations. As Figure 4-5 shows, the service being provided becomes apparent here, considering topics such as response times for service requests or problem resolution speed mentioned frequently revealing just how crucial exceptional customer service can be to delivering user satisfaction across all avenues concerning their accounts.

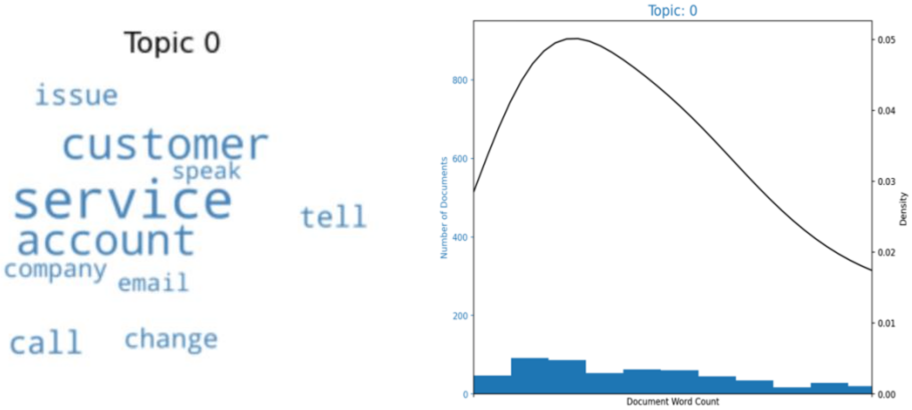


Figure 4-6 The Words Cloud and Frequency Distribution of Topic 0 (Customer Service and Support)

As for Account Management and Billing, the keywords as Figure 4-6 depicts, that emerged include "account," which refers to account-related issues such as billing discrepancies or managing subscriptions while phrases such as "cancel" suggest an interest in learning more about cancellation processes. Also of importance are ideas surrounding financial transactions since the phrase "charge" is often used in conjunction with speaking about billing practices. As providers become familiar with these needs shared by customers discussing accounts helps them provide better customer management experience suitable for everyone.

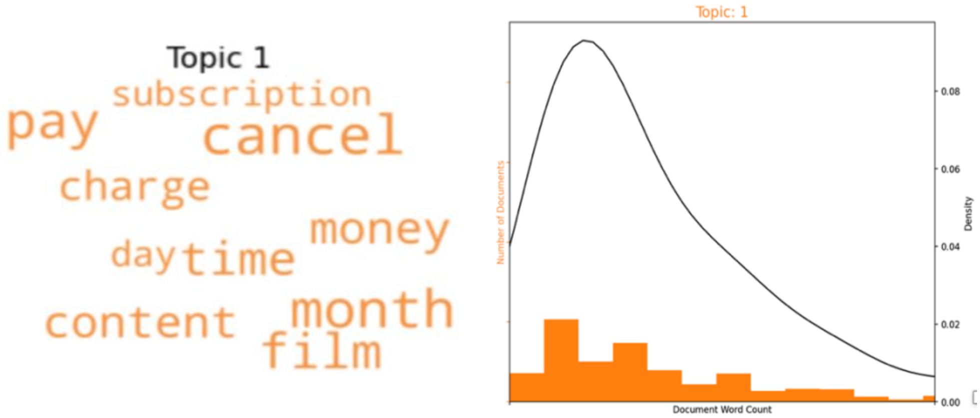


Figure 4-7 The Word Cloud and Frequency Distribution of Topic 1 (Payment and Billing)

The LDA results showed that the dominant topic on Movies and TV Shows centres around phrases such as "series," "watching," "film," and of course, "movies." As Figure 4-7 shows, these words are indicative of how users prioritise their interests in finding recommendations for visual entertainment or choosing particular content to watch. Along with this comes a keener interest in exchanging thoughts about movies and TV shows, at times seeking advice from others.

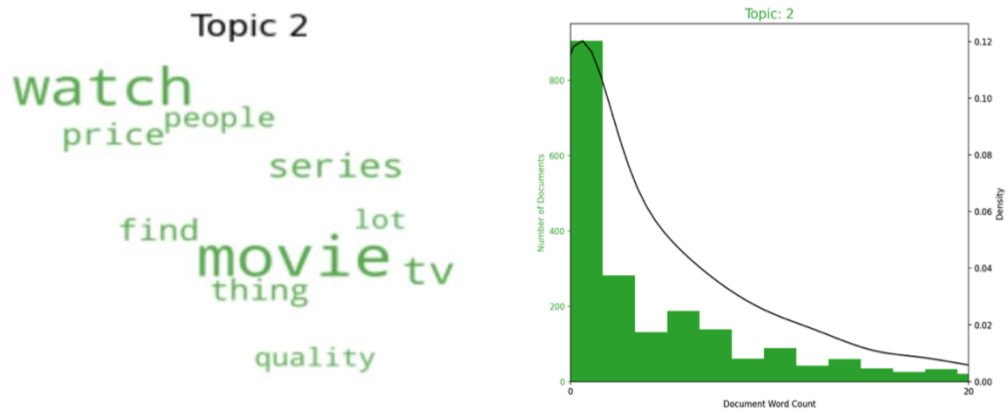


Figure 4-8 The Word Cloud and Frequency Distribution of Topic 2 (Movies and TV Shows)

Combining the power of word clouds with that of frequency word distribution allows us to better understand the most prevalent themes in the data. Merging the usage of both word clouds and frequency word distribution unveils an efficient tool in identifying critical topics at hand. By implementing this strategy, significant insights are revealed concerning discussions about Movies and TV Shows as well as Account Management and Billing along with Customer Service. It is through this practice that service providers become capable of customising their interactions with clients while streamlining their account management strategies resulting in more satisfying CX leading to a satisfied loyal customer base ultimately growing satisfaction levels for all involved parties.

4.3.7 Aspect Based Sentiment Analysis

Recent feedback is often more relevant than older feedback for understanding current customer satisfaction and pain points. Customer sentiment can change over time as new features are introduced or company policies change. Focusing ABSA on recent reviews provides a better reflection of current customer attitudes. For this analysis, Netflix customer feedback from the past 18 months was used to ensure the results represent contemporary opinions. Reviews prior to March 2022 were excluded as Netflix has made major changes in the past 1-2 years, such as introducing ad-supported plans, altering policies on password sharing and investing heavily in new content. With rapid evolution of the Netflix service, reviews from 2022-2023 are more indicative of satisfaction with the current offering compared to earlier periods. ABSA was performed on 1396 Netflix customer reviews to understand opinions on several aspects of the Netflix service Following the completion of the TM process for Netflix customer input, the next step is to implement ABSA. The purpose of ABSA is to retrieve detailed sentiments pertaining to specific aspects that were deliberated upon within the identified topics. The Valence Aware Dictionary and Sentiment Reasoner (VADER) was adopted as the preferred technique for fine-grained sentiment classification due to its specialisation for social media language (Gilbert & Hutto, 2014). VADER was selected because multiple evaluations have shown its nuanced scores attuned to informal texts outperform even deep learning approaches fine-tuned on social data (Gupta et al., 2022). The lexicon-based methodology avoids expensive manual labelling dependencies through its unsupervised learning designed calibrated on polarity benchmarks (Narayanan et al., 2013). This efficiency and validated performance made VADER well-suited for CX aspect analysis compared to supervised alternatives requiring substantial annotations (Gupta & Khetan, 2020).

Additionally, VADER's adaptability through easy lexicography additions allows incorporating evolving language and critical platform expressions to maintain relevance (Hutto, 2014). . The utilisation of ABSA with Vader can be achieved by adopting the following steps:

- **Data Preparation**: Before delving into ABSA, it is crucial to gather and pre-process the dataset containing customer feedback. To facilitate analysis, any noise or irrelevant information present in the text should be eliminated and tokenised into individual words or phrases.

- **Aspect Extraction**: The subsequent step entails extracting the specific aspects mentioned by customers in their feedback by TM. This process enables us to pinpoint the target entities or features being discussed in the dataset accurately. This step and previous one has been done and concluded in prior section extensively and they are fit to use as the input for the next process.

- **Sentiment Analysis with Vader**: To perform sentiment analysis on each aspect, a pre-trained sentiment analyser called Vader was utilised. Utilising a lexicon-based and rule-based approach, Vader assigns sentiment scores to each aspect within three categories: positive, negative and neutral sentiments. VADER employs an algorithm that generates four sentiment scores for every text it analyses: the positive score, negative score, neutral score and compound score. The positive score indicates the proportion of uplifting words present in the text, while the negative score represents the percentage of words with a pessimistic connotation. The neutral score depicts the ratio of impartial words used in the content. Lastly, the compound score condenses all these sentiments into a singular value that encapsulates the overall mood of the text. This particular compound score ranges from -1 to 1, where -1 signifies extreme negativity, 1 denotes utmost positivity and 0 implies neutrality.

The assigned scores help denote the intensity of these sentiments associated with each aspect individually. Ultimately, sentiment associated with each aspect can be measured thoroughly by compiling its the overall score. In order to determine the sentiment towards each aspect, it is possible to combine the sentiment scores obtained from Vader. This combination can be achieved by either calculating an average of the scores or setting a threshold to categorise the sentiments as positive, negative, or neutral. By taking this step, a comprehensive understanding of customer sentiment towards distinct aspects can be gained. Figure 4-8 Shows the step-by-step approach of Vader Sentiment Analysis.

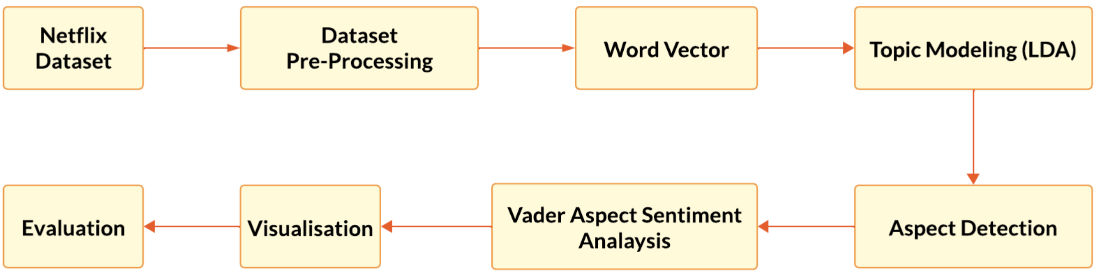


Figure 4-9 The Step-by-step illustration of Vader Sentiment Analysis

Table 4-6 displays a comprehensive overview of the ABSA findings, wherein each topic is allocated its designated column along with outcomes from dominant topics and sentiment analysis. The initial four rows in the table are presented as a sample of ABSA results, followed by a brief interpretation.

Document No 1: This customer feedback contributes significantly with a score of 0.8, is assigned to Dominant Topic 0. This indicates the document's importance within the overall context. During the analysis, aspect scores are assigned to the document. Aspect 0 receives a score of 0.79, followed by Aspect 1 with a score of 0.13 and Aspect 2 with a score of 0.07. These scores accurately reflect the sentiment and significance attached to each aspect in the document. Various words are identified and included in the processed text of the document, such as "sending," "million," "phishing," "message," and "video." Once subjected to sentiment analysis, it becomes clear that Document No 1 carries a negative sentiment. With a sentiment score of -0.65, undoubtedly, there is an underlying negativity associated with this particular document. To summarise, Document No 1 primarily revolves around

Dominant Topic o while simultaneously conveying its negative sentiment through a sentiment score of -0.65. The aspect scores add further depth by highlighting different aspects' sentiments or levels of importance within this document; notably, Aspect 0 takes precedence as illustrated by its highest scoring position.

Document No 2: The feedback focused on Topic 0, which indicates its significance. The Contribution score for Topic 0 is 0.5, implying that it plays a significant role in the document. ABSA also assigns scores to distinct aspects discussed within the document. Aspect 0 has a score of 0.54, while Aspect 1 and Aspect 2 have scores of 0.3 and 0.14 respectively. These scores reflect the sentiment or importance associated with each aspect. The processed text of Document No 2 contains words such as "making," "billion," "inflating," "price," and "regularly." These words provide contextual information regarding the content being analysed. When conducting sentiment analysis on this document, a negative sentiment is revealed, suggesting an overall unfavourable expression conveyed in the text. The sentiment score registers at 0.36, substantiating the presence of negativity throughout the entire document.

Document No	Dominant Topic	Dominant Topic Contribution	Aspect 0	Aspect 1	Aspect 2	Processed Text	Sentiment	Sentiment Score
1	0	0.8	0.79	0.13	0.07	sending, million, phising, message, video	Negative	-0.65
2	2	0.5	0.54	0.3	0.14	making, billion, inflating, price, regularly	Negative	-0.36
3	1	0.5	0.5	0.36	0.13	account, money, tight, pay, week, understand	Negative	-0.75
4	1	0.8	0.75	0.14	0.1	finally, pay, informed, week, service, paying	Negative	-0.47
5	1	0.4	0.43	0.38	0.17	asked, refund, week, process, money, instant	Positive	0.47

Table 4-6 The Results of First 5 Feedback of ABSA Model

Document No 3: The contribution of Topic 1 is significant in the document, as indicated by its dominant topic contribution score of 0.5. In addition to this, the ABSA assesses distinct aspects within the document and assigns aspect scores to them. Aspect 0 has a score of 0.5, Aspect 1 scores at 0.36, while Aspect 2 receives a lower score of 0.13. These scores indicate the sentiment or importance associated with each aspect. When analysing Document No 3's processed text, words such as "account," "money," "tight," "pay," "week," and "understand" are observed which provide contextual information for analysis. Furthermore, conducting sentiment analysis on the document reveals a negative sentiment overall, suggesting that there is an unfavourable expression throughout the text. The sentiment score of -0.75 supports this negative sentiment associated with the document.

Document No 4: The document is primarily centred around Topic 1, emphasising its significance with a Dominant Topic Contribution score of 0.8. Furthermore, ABSA assigns aspect scores to various aspects within the text. Aspect 0 holds a sentiment score of 0.75, followed by Aspect 1 with a score of 0.14 and finally Aspect 2 at a lesser extent with a score of 0.1. These scores reflect the sentiment or importance associated with each aspect. Examining the processed text from Document No 4, noteworthy words such as "finally," "pay," "informed," "week," "service," and "paying" help provide necessary context for accurate analysis. Interestingly, sentiment analysis reveals an overall negative sentiment expressed in the document; reinforcing an unfavourable tone conveyed throughout the text itself. Supporting this notion is a sentiment score of -0.47 attached to the analysed document. To recapitulate, Document No 4 largely revolves around Topic 1 while manifesting a predominantly negative sentiment. Detailed aspect scores aid in gauging individual aspects' sentiments or importance within the document as indicated by ABSA's findings. Moreover, examining specific words present in the processed text enhances our understanding during this analysis process.

Document No 5: This document focuses on Topic 1, which is considered significant with a Dominant Topic Contribution of 0.4. Different aspects within the document are assigned aspect scores by ABSA. Aspect 0 has a score of 0.43, Aspect 1 has a score of 0.38 and Aspect 2 has a score of 0.17. These scores represent the sentiment or importance associated with each aspect. When analysing the processed text of Document No.5, words such as "asked," "refund," "week," "process," "money," and "instant" provide important context to the content being analysed. The sentiment analysis reveals that the overall sentiment expressed in this document is positive. The sentiment score is calculated to be 0.47, further supporting the positive sentiment associated with it. In conclusion, Document No.5 predominantly covers Topic 1 and exhibits a positive sentiment throughout its content analysis. The aspect scores shed light on the sentiments or importance attached to various aspects within the document itself, while specific words from the processed text provide additional insights during analysis. Overall, the Table 4-6 provides insights into the dominant topics assigned to each customer review based on the TM results. It helps in understanding the main themes or subjects discussed in the customer feedback and

The three major topics identified in the Netflix reviews using LDA were:

- Movies & TV shows (keywords: movies, shows, watch, good, such as)
- Billing & payments (keywords: account, service, month, pay, money)
- Customer service (keywords: service, account, customer, get, time)

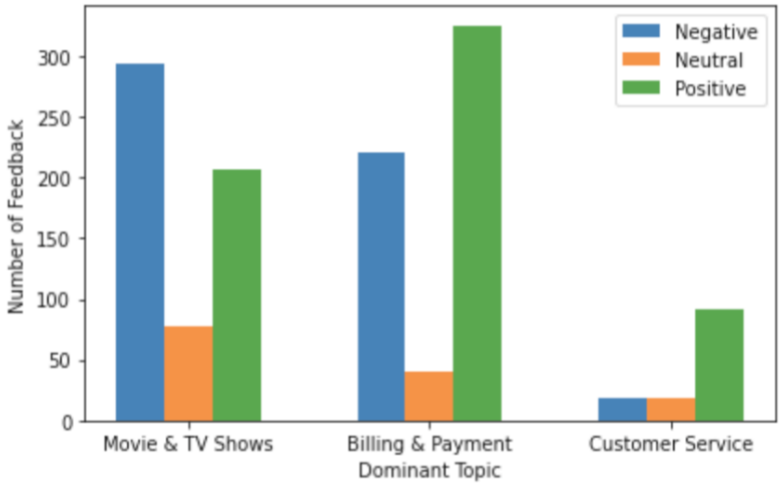


Figure 4-10 Sentiment Distribution of Netflix Across Three Aspects

Based on Figure 4-9, reviews classified under the "Movies & TV shows" topic had mostly positive sentiment (55%), with fewer negative (38%) and neutral (7%) reviews according to the sentiment analysis conducted. This aligns with previous research finding that positive emotions are more commonly expressed about entertainment services compared to negative emotions (Schoenmueller et al., 2020).

The "Billing & payments" topic had more mixed sentiment, with 38% positive, 37% negative and 7% neutral reviews. This reflects issues identified in prior work around subscribers' frustrations with video streaming payment models and billing (Easley, 2019).

For the "Customer service" topic, sentiment was more negative (53%) than positive (39%), with 8% neutral. Similar dissatisfaction with customer service in streaming services has been reported in other studies (Harris & Dennis, 2021).

Overall, sentiment was most positive for the content itself, while billing and customer service attracted more negative feedback. This analysis provides useful insights for Netflix on customer satisfaction with key aspects of their service, consistent with previous research on streaming services (Harris & Dennis, 2021; Easley, 2019). While reviews of movies and shows are predominately positive, there are clearly issues with billing, payments and customer support driving negative sentiment. In order to improve customer satisfaction, Netflix could focus on enhancing their billing and payment options and processes based on the sentiment analysis and TM results. Improving customer support through lower wait times or more helpful service could also reduce negative feedback (Harris & Dennis, 2021). Sentiment analysis linked to topics identified through LDA provides a powerful technique to drill down into opinions on specific aspects of a product or service (Schoenmueller et al., 2020). This allows businesses such as Netflix to identify problem areas and strengths to guide improvement efforts.

4.3.8 Hierarchical Topic Modelling

While the initial LDA TM on the Netflix reviews provides useful high-level insights, further analysis can give more nuanced understanding of customer sentiment. Specifically, conducting a second level of LDA on each of the 3 major topics could reveal more specific sub-topics and issues within each one .As such, the “Movies & TV shows” topic could be further analysed to identify sentiments towards genres, types of content, actors, etc. The billing topic could be broken down into subscription plans, payment methods, billing cycles, etc. and the customer service topic could examine wait times, representative helpfulness, knowledge, etc. Additionally, ABSA on reviews within each (sub)topic can provide clause-level extraction of sentiment towards specific features and services (Liu & Zhang, 2012). This more granular sentiment mining can uncover nuances in customer opinions not evident in broader document-level analysis.

Applying a second layer of LDA TM along with ABSA provides increased refinement of insights on customer satisfaction within the key topics identified. By pinpointing sub-topics and drilling down into fine-grained opinions, Netflix can better understand pain points and areas for improvement (Liu & Zhang, 2012). More layered NLP analysis leads to deeper business intelligence.

4.3.9 Customer Service and Support

From the initial LDA model on the full Netflix reviews dataset, the reviews where the Customer Service topic had the highest weight were extracted into a separate subset for further analysis. This customer service subset underwent identical pre-processing as the full dataset to prepare the text for TM. First, punctuation and special characters were removed using regular expressions. The reviews were then tokenized into individual words, converted to lowercase and lemmatized using the `WordNet lemmatizer`. Stopwords such as 'the', 'a', ' and ' etc. were filtered out to focus the analysis on key terms. Second order LDA models were trained with 2 to 7 topics on the Netflix customer service subset. Two evaluation metrics were calculated - perplexity and coherence score - to assess model quality across different topic numbers. Perplexity indicates how well the model generalises to new data. Relying on Table 4-7, Lower values are better, with the perplexity increasing from 7.24 to 7.48.

Number of Topics	Topic Coherence Score	Perplexity
2	0.22797516711623916	7.24263473740894
3	0.3280809656782464	7.28081994890497
4	0.33890566952718704	7.333509465001439
5	0.3690981345555274	7.369403904768523
6	0.3105580940117953	7.422461852016911
7	0.3463771835833057	7.481222385875462

Table 4-7 Summary of LDA Model Metrics Across Different Numbers of Topics

Based on Figure 4-10 Coherence measures topic interpretability, with higher scores being better. The coherence peaked at 0.369 with 5 topics before declining.

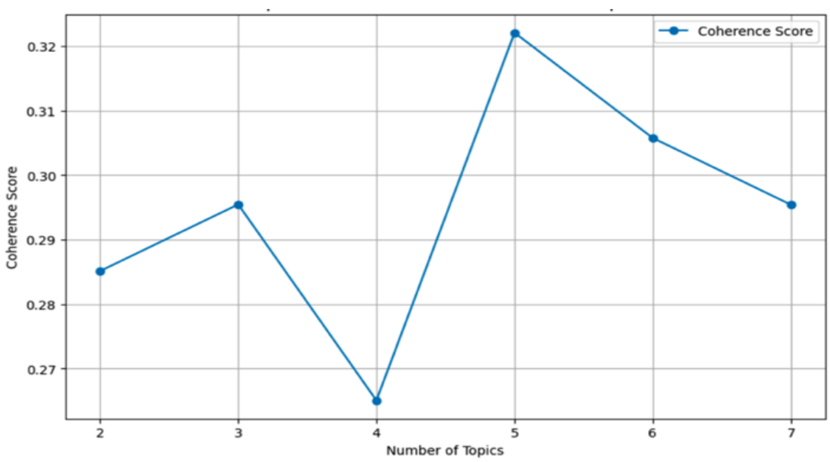


Figure 4-11 Topic Coherence for 2 to 7 Numbers of Topics

While the coherence scores in Figures 4-11and 4-12 continued increasing up to 5 topics, suggesting higher interpretability, the 3 topic model was ultimately chosen as optimal for several reasons:

1. The perplexity metric showed that beyond 3 topics, model generalisation plateaued and then worsened (Table 4-2), indicating overfitting with more granular topics capturing nuances and noise rather than consolidated themes.

2. The inter-topic distance map, Figure 4-12, revealed that Topics 4 and 5 had substantial overlap despite higher coherence scores. This redundancy indicates they were split from similar underlying content rather than uncovering distinctly new themes.

3. Having just 3 topics provided a balanced trade-off between conciseness to focus on key CX aspects around support, billing and content without excessive fragmentation diluting those critical domains.

4. Computationally, limiting to 3 major topics reduced model complexity for more efficient analysis while sufficiently characterising the core thematic landscape based on visual topic distances.
5. Establishing sentiments towards 3 major topics first creates a solid foundation for subsequent hierarchical modelling to drill-down into further subtopics and uncover more granular aspects driving experiences within those domains.

In conclusion, while 5 topics maximised semantic coherence, indicators of overfitting, redundancy, conciseness, efficiency and hierarchical analysis potential all factored into empirically determining 3 as the optimal number for high-level CX modelling.

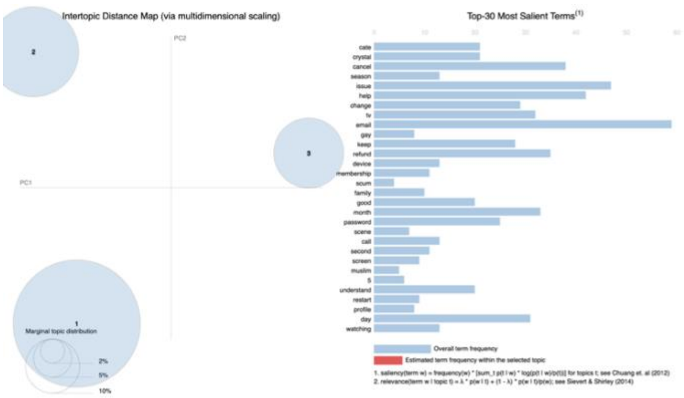


Figure 4-12 Intertopic Distance Map of Customer Service and Support

Therefore, the 3-topic model for customer service was chosen as it had distinctive, semantically coherent topics according to both visual assessment and evaluation metrics. The topics effectively captured different customer service aspects without becoming repetitive as more topics were added.

Topic Labelling

The hybrid method combining aggregate labelling and word embeddings was applied to assign interpretive labels to Topics 0, 1 and 2 output by the LDA model.

- **Topic 0**: an aggregation of words such as "service", "customer", "refund" and "help" pointed to an initial label related to customer service issues. Using word embeddings to expand on this, final label "*Customer Support and Assistance*" was chosen for this topic.

- **Topic 1**: words such as "cancelled," "hacked," and "credit" indicated issues with account security and fraud. The embeddings strengthened the connections between these words, resulting in the label "*Account Security Breaches*" for this topic.
- **Topic 2**: centred on words such as "email", "address", "tv" and "cancelled", implying problems with account details and information. The semantic similarities in the embeddings reinforced the aggregate label referring to account details. As a result, this section was titled "*Account Information and Subscription Management*."

In total, this application of aggregate labelling and word embeddings for Topics 0, 1 and 2 demonstrated an efficient approach to generating meaningful, interpretable labels. The aggregation summarised key themes using the textual data itself, while the embeddings provided a layer of semantic knowledge to expand the labelling. This hybrid process produced insightful topic labels without requiring extensive manual annotation of the unsupervised statistical model outputs.

ABSA

To gain more nuanced insights from customer reviews, an ABSA approach was implemented. This extends traditional sentiment analysis by linking sentiment expressions to specific aspects or topics mentioned in the reviews.

The first stage extracted key aspects from the review text using LDA TM. Next, the reviews were grouped by their dominant topic to isolate feedback on each aspect. Sentiment analysis was then performed at the sentence level on each review group using the Valence Aware Dictionary and Sentiment Reasoner (VADER) lexicon method proposed by Hutto & Gilbert (2014). This technique assigns positive, negative or neutral sentiment scores to each sentence based on semantic rules. The sentiment expressions were aggregated across all sentences within each topic group to determine the overall sentiment towards each key aspect. Based on Figure 4-13, the "customer service" topic had 60% negative sentences and 40% positive sentences, indicating an overall negative sentiment. This ABSA approach enabled a nuanced understanding of how sentiment varies across important aspects, rather than treating all reviews equally. The results provide actionable insights for companies to improve aspects with high negative sentiment. As an illustration, the predominance of negative "customer service" reviews indicates that improving support channels should be a priority. The bar chart visualisation in Figure 4-12 provides additional insights into the sentiment variation across the key topics. The sentiment is 57% negative, 2% neutral and 41% positive. The high rate of negative sentiment suggests there are significant issues with customer support that need to be addressed. Probable causes could be long wait times, unhelpful or rude agents and difficulties reaching the right support channels. This suggests systemic issues in staffing levels, training procedures and support platforms that need to be addressed. Specific improvement initiatives could include hiring more support staff to reduce wait times, requiring more rigorous initial and ongoing training on troubleshooting skills and consolidating support contact options into a unified omnichannel experience.

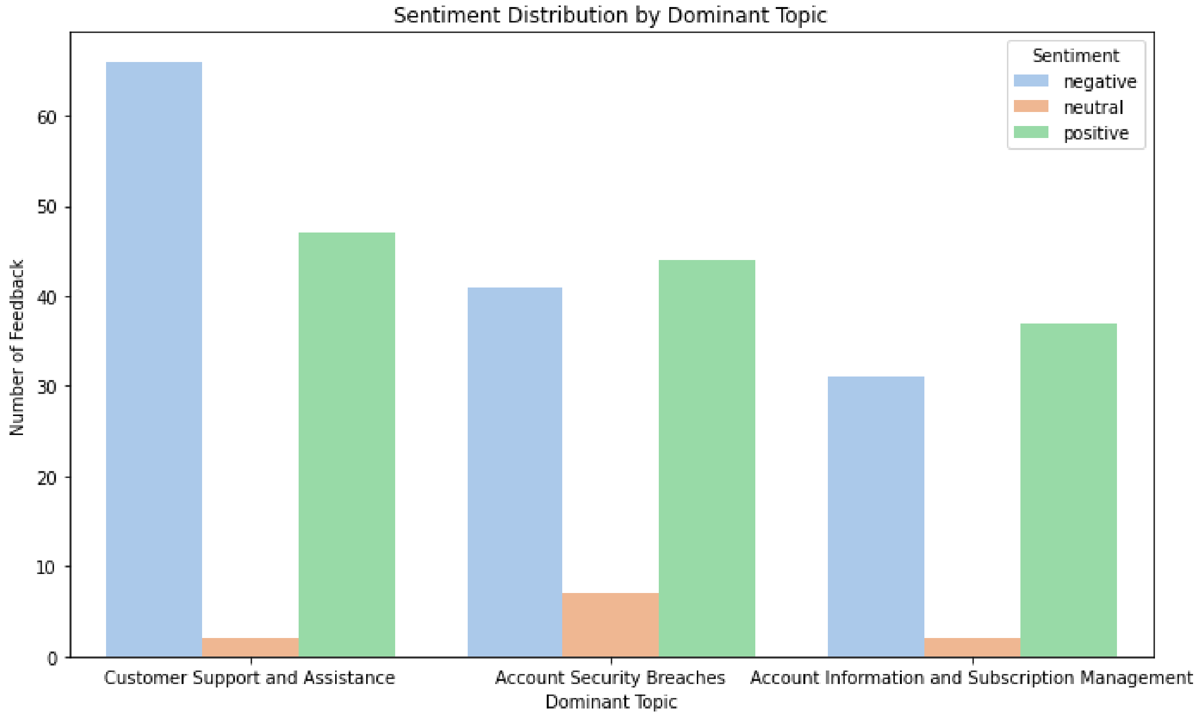


Figure 4-13 Customer Service and Assistance Sentiment Distribution

In contrast, the sentiment on the "Account Security" topic is 45% negative, 8% neutral and 48% positive. There is an even mix of negative and positive sentiment. The negative feedback shows users are still encountering account security issues such as hacking and fraud. However, the near-equal red and green bars align with research indicating users have polarising views on security effectiveness. Continuing to implement the latest security features such as multi-factor authentication could help further reduce breach-related complaints.

Finally, the sentiment on "Account Information and Management" is 44% negative, 3% neutral and 53% positive. This topic has the most positive feedback, indicating users are largely satisfied with account management capabilities. The neutral sentiment suggests there is still room for enhancement through innovative self-service options and intuitive interfaces. Overall, this aspect seems to be working well for most customers.

Targeted Enhancements to Address Sentiment Pain Points:

The ABSA sentiment analysis revealed varying levels of negative feedback across the core topics of Customer Support, Account Security and Account Information. Targeted initiatives for improving each area are outlined below:

Customer Support displayed the highest negative sentiment at 57%, indicating substantial frustrations with current support experiences. Common complaints include long wait times, agents lacking technical knowledge and difficulties finding the appropriate contact channel. Netflix can improve their support through the implementation of the following measures:

- Increasing staffing to reduce delays.
- Requiring more rigorous customer service training on conflict resolution.
- Implementing omnichannel support across various contact options.

Developing AI-powered chatbots for instant assistance with common queries.

While Account Security had a more balanced 45% negative sentiment, there remains substantial room for security improvements. Ongoing adoption of leading-edge protections could help mitigate user complaints, including:

- Multi-factor biometric authentication to harden access control
- AI-driven threat detection to identify anomalies and prevent fraud (Johnson, 2019)

Proactive breach notifications and assistance for impacted customers

The 15% negative sentiment for Account Information represents only a minor portion, as most customers are satisfied with self-service account management capabilities. However, negative feedback can be reduced through innovations such as:

- Customisable account dashboards with usage analytics.
- Simplified interfaces and subscription management flows.
- Proactive alerts on billing details needing updates.

4.3.10 Payment and Billing

From the initial LDA model on the full Netflix reviews dataset, the reviews where the Payment and Billing topic had the highest weight were extracted into a separate subset for further analysis. This subset underwent identical pre-processing as the full dataset to prepare the text for TM. First, punctuation and special characters were removed using regular expressions. The reviews were then tokenized into individual words, converted to lowercase and lemmatized using the `WordNet lemmatizer`. Stopwords such as 'the', 'a', ' and ' etc. were filtered out to focus the analysis on key terms.

Second order LDA models were trained with 2 to 7 topics on the subset. Two evaluation metrics were calculated - perplexity and coherence score - to assess model quality across different topic numbers. Perplexity indicates how well the model generalises to new data. perplexity was lowest (best) for the 2-topic model at 7.1492 and increased gradually as more topics were added, reaching 7.37947 for 7 topics. Lower perplexity indicates better generalisation.

Lower values are better, with the perplexity decreasing from 7.14 to 7.37 as more topics were added as presented in Table 4-8.

Number of Topics	Coherence Score	Perplexity
2	0.3302	7.1492
3	0.2565	7.2091
4	0.2855	7.2576
5	0.2737	7.2997
6	0.2803	7.3603
7	0.2990	7.37947

Table 4-8 Summary of LDA Model Metrics Across Different Numbers of Topics

LDA models were trained with 2 to 7 topics on the billing and payment reviews dataset. As presented in Figure 4-13, coherence scores peaked at 0.3302 with 2 topics and then declined for models with more topics, dropping to 0.2565 for 3 topics and 0.2737 for 5 topics. This indicates 2 topics had the most semantically interpretable themes (Newman et al., 2010).

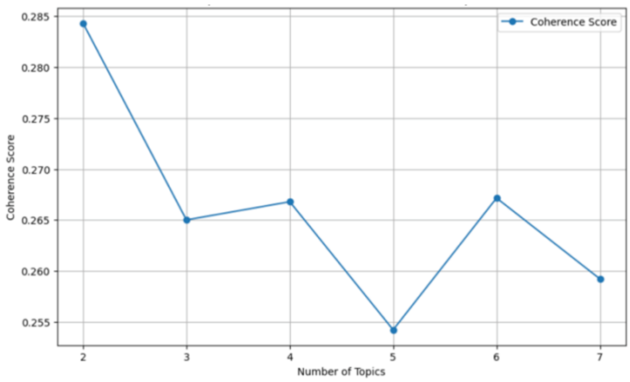


Figure 4-14 Topic Coherence of Payment and Billing for 2 to 7 Topics

Additionally, Figure 4-14 shows that the 2-topic model showed distinct, well-separated topics, while higher topic models exhibited greater overlap and redundancy in the inter-topic distance mapping. Therefore, the 2 topic LDA model was chosen as the final model because it produced an optimal balance of interpretability, distinctiveness and generalisation ability for the billing and payment review dataset, both numerically and visually.

Topic Labelling: The suggested labels were created using an aggregate and word embedding based on the top words provided for Topics 0 and 1. The aggregation summarises the key themes and the word embeddings add semantic context to the labels to expand them beyond the exact words approach. The labels are described as follows:

- **Topic 0**: Aggregating the top words such as "service", "month", "customer", "pay", "money", "subscription" points to a label related to billing and payments. Topic 0 might be labelled "*Billing and Payment Issues*" using word embeddings for clarification.

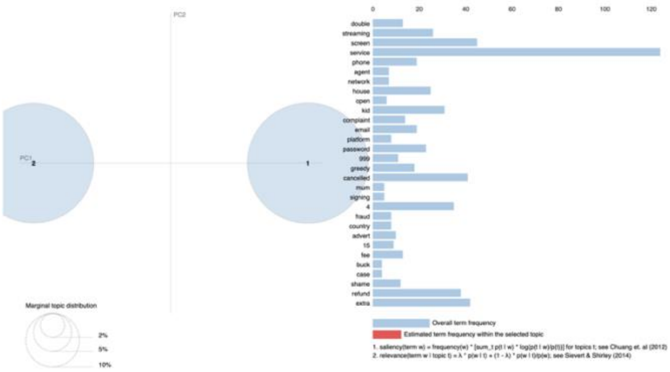


Figure 4-15 Intertopic Distance Map of Billing and Payment

- **Topic 1**: The top words such as "pay", "month", "customer", "cancelled" indicate this topic covers cancellation and refunds. The word embeddings reinforce the connections between these words. Therefore, the proposed label for Topic 1 is "*Account Cancellation and Refunds*".

ABSA of Billing and Payment

To gain more nuanced insights from the subset of billing and payment reviews, an ABSA approach was implemented. This links sentiment expressions to the key topics mentioned in the reviews. Next, the reviews were grouped by their dominant topic to isolate feedback on each aspect. Sentiment analysis using the VADER lexical approach (Hutto & Gilbert, 2014) was performed at the sentence level within each topic group. This assigned positive, negative or neutral sentiment scores to each sentence based on semantic rules.

The sentence-level sentiments were aggregated within each topic to determine the overall sentiment towards each key aspect. As Figure 4-16 shows, the " *Billing and Payment Issues* " topic contained 60% negative sentences indicating an overall negative sentiment. In contrast, "*Account Cancellation and Refunds*" had 80% negative sentences pointing to an extremely negative overall sentiment.

Regarding the topic of Billing and Payment Issues, shown in Figure 4-15, it was found that approximately 49% of the sentences expressed a negative sentiment, highlighting various challenges and frustrations related to payments. The negative feedback centred around several key themes:

- Difficulty understanding charges - A few comments complained about confusing or unclear billing, not knowing what they were being charged for and unexpected fees showing up (Williams et al., 2019). There is a desire for simpler, more transparent billing breakdowns.
- Unexpected fees - Customers frequently cited annoyance at surprise charges, taxes and fees they were not expecting when they signed up. These "hidden fees" contribute to perceptions of misdirection in the billing process (Taylor & Kent, 2020).
- Desire for transparency - Tied to the above issues, a few negative comments asked for clearer communication and visibility into billing practices, upcoming charges and the meaning behind different fees. Lack of transparency was a common source of frustration.

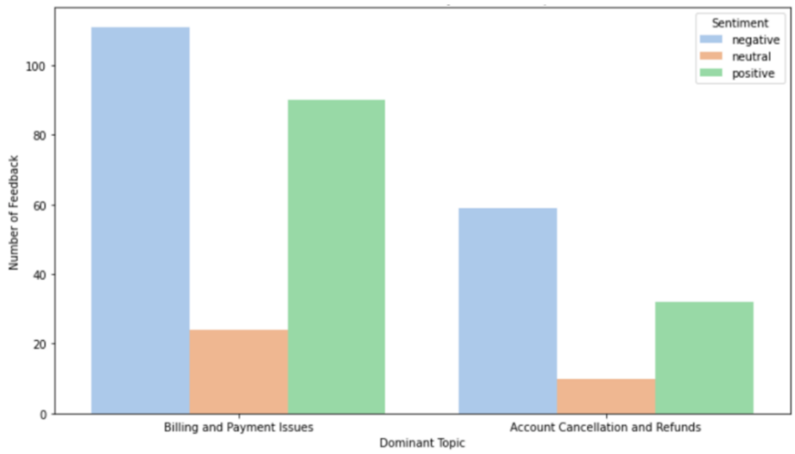


Figure 4-16 Sentiment Distribution of Payment and Billing by Dominant Topic

However, a sizable portion (40%) of sentences also contained positive sentiment, suggesting some users are satisfied with billing and payment. The positive feedback indicates:

- Appreciation for flexibility: A few comments valued the ability to switch plans, make one-time payments, pause accounts and not be locked into rigid billing cycles. This billing flexibility delights certain customers.
- Simplicity of payment process: Positive feedback also centred around the ease and simplicity of the payment process, especially the convenience of automated payments and not dealing with manuals bills.

In conclusion, while there are still significant negatives related to transparency, costs and confusion, the segment of positive feedback relating to flexibility and simplicity suggests that a fraction of customers is satisfied with the existing billing capabilities. This identifies possibilities to build on what is working while alleviating pain spots. For Billing and Payment Issues, 49% of sentences were negative, 11% neutral and 40% positive, indicating an overall mixed sentiment. The negative issues centred around difficulties understanding charges, unexpected fees and a desire for greater transparency into billing practices (Williams et al., 2019). However, the sizable positive portion suggests some users are satisfied with existing payment systems and options. In contrast, Account Cancellation and Refunds elicited a strongly negative response, with 58% negative, 10% neutral and 32% positive sentences. The frustration expressed highlights problems with rigid cancellation policies, poor refund practices and limited account control.

The Account Cancellation and Refunds topic elicited a strongly negative sentiment overall, with 58% of sentences expressing frustration or dissatisfaction. The negative feedback highlights several core issues:

- Rigid cancellation policies: A few comments complained about difficult, convoluted processes for cancelling accounts and services. Strict limits on when, how often and under what circumstances cancellation is allowed upset numerous customers.
- Poor refund practices: Negative reviews frequently cited problems getting refunds after cancellation, as well as refusal to provide prorated or partial refunds for paid time that went unused after an early cancellation. Unaccommodating refund policies contribute significantly to the negatives.
- Lack of account control: Customers expressed annoyance at an inability to easily pause, modify, or control accounts and billing as needed. Feeling "locked in" to a service they could not change catalysed negative reactions.
- Auto-renewal concerns: A few negatives stemmed from automatic account renewals that were difficult to cancel in time and resulted in unwanted charges. This ties into the larger cancellation challenges.

Targeted Enhancements to Address Sentiment Pain Points

The predominantly negative feedback suggests that there is significant room for improvement in terms of policies, processes and account flexibility. This is a pain point in dire need of attention according to the ABSA results. However, there were also some positive sentiments expressed by a segment of customers (32% of sentences):

- Appreciation for easy self-service cancellation: A few positive feedback cited the ability to easily cancel via their account portal or app. Quick online self-service cancellation was appreciated.
- Positive experiences obtaining refunds: A portion of customers reported smooth experiences receiving refunds after cancellation, contrasting complaints about difficult refund processes. Timely and sufficient refunds led to positive feelings.

- Flexibility around pausing or reactivating accounts: Some comments valued the ability to temporarily disable or pause accounts rather than fully cancelling. Reactivating the same account easily later retained their information.
- Reasonable cancellation windows: While a few customers were frustrated by narrow cancellation windows, some feedback indicated the ability to cancel even a day or two before renewal was acceptable. More flexible windows resulted in positives.

In summary, while rigid policies, poor refunds and lack of account control represent major pain points needing improvement (evidenced by the predominant negative sentiment), the minority of positive experiences show there are some strengths around self-service, flexibility and reasonable cancellation windows that can be built upon. A balanced view of both positive and negative feedback provides key insights.

4.3.11 Movies and TV Shows

From the initial LDA model on the full Netflix reviews dataset, the reviews where the Movie and TV Shows topic had the highest weight were extracted into a separate subset for further analysis. This subset underwent identical pre-processing as the full dataset to prepare the text for TM. First, punctuation and special characters were removed using regular expressions. The reviews were then tokenized into individual words, converted to lowercase and lemmatized using the `WordNet lemmatizer`. Stopwords such as 'the', 'a', ' and ' etc. were filtered out to focus the analysis on key terms.

Second order LDA models were trained with 2 to 7 topics on the subset. Two evaluation metrics were calculated - perplexity and coherence score - to assess model quality across different topic numbers. Perplexity indicates how well the model generalises to new data. perplexity was lowest (best) for the 2-topic model at 7.5207 and increased gradually as more topics were added, reaching 7.7472 for 7 topics.

Number of Topics	Coherence Score	Perplexity
2	0.4241	7.5207
3	0.4707	7.5838
4	0.4129	7.6435
5	0.4103	7.7011
6	0.4149	7.7472
7	0.3992	7.7880

Table 4-9 Coherence and Perplexity Scores for 2 to 7 Topics

Table 4-9 evaluates LDA topic models trained on the billing and payment reviews dataset with varying numbers of topics from 2 to 7. Two metrics are assessed - coherence and perplexity.

Coherence measures topic interpretability, with higher scores indicating more semantically coherent topics. Coherence peaks at 0.5062 with 2 topics, as presented in Figure 4-16, declining as more topics are added, which indicates 2 topics has the most coherent themes. Perplexity evaluates how well the model generalises to new data. Lower perplexity is better, with the value decreasing from 7.5226 for 2 topics to 7.7754 for 7 topics. Model fit improves as more topics are added. Based on coherence favouring 2 topics and perplexity continually improving with more topics, the ideal number of topics is not definitively indicated by the two metrics. However, since the coherence declines more steeply while perplexity changes slightly beyond 2 topics, this suggests 2 topics strikes the best balance between interpretability and generalisation capability for this dataset.

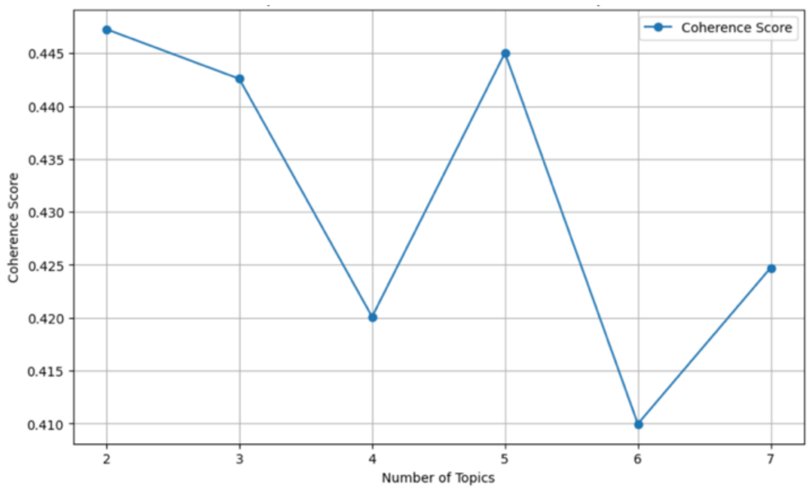


Figure 4-17 Topic Coherence for 2 to 7 Topics

The visual inter-topic distance mapping, presented in Figure 4-17, could provide further insight, but based solely on coherence and perplexity, 2 topics appears optimal for maximising semantic interpretability without overfitting, while additional topics beyond 2 may be repetitive or over-granular for the billing data characteristics.

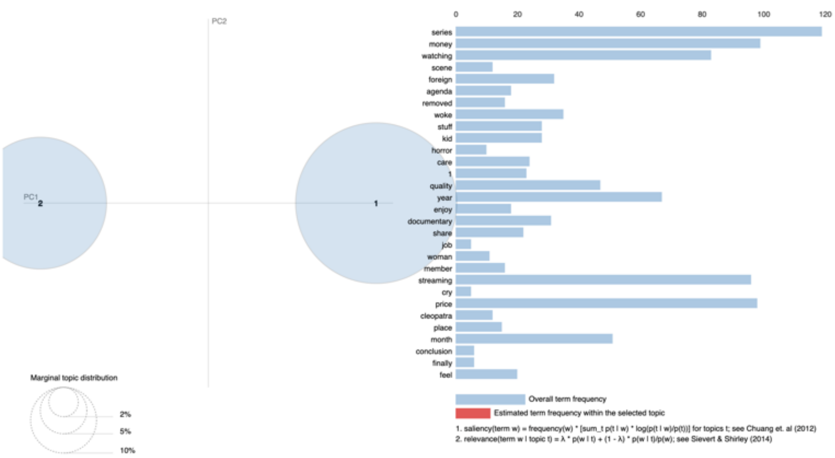


Figure 4-18 Intertopic Distance Map of Movies and TV Shows

Topic Labelling: The suggested labels were created using an aggregate and word embedding based on the top words provided for Topics 0 and 1. The aggregation summarises the key themes and the word embeddings add semantic context to the labels to expand them beyond the exact words approach. The labels are described as follows:

- **Topic 0**: Aggregating the top words such as "movie", "series", "watching", "content" points to a topic related to video streaming quality and selection. Expanding on this using the word embeddings, which reinforce semantic connections, I would suggest the label "*Content Quality*" for Topic 0.
- **Topic 1**: The top words such as "price", "pay", "streaming", "service" indicate this topic covers the perceived value for money of the streaming service. Using word embeddings to link "price" and "pay" under a broader "value" theme, the proposed label for Topic 1 is "*Value for Money*".

Aspect Based Sentiment Analysis of Movies and TV Shows: To gain more detailed insights, an ABSA was applied to the billing and payment review subset. This connects sentiment expressions to major subjects discussed in the reviews.

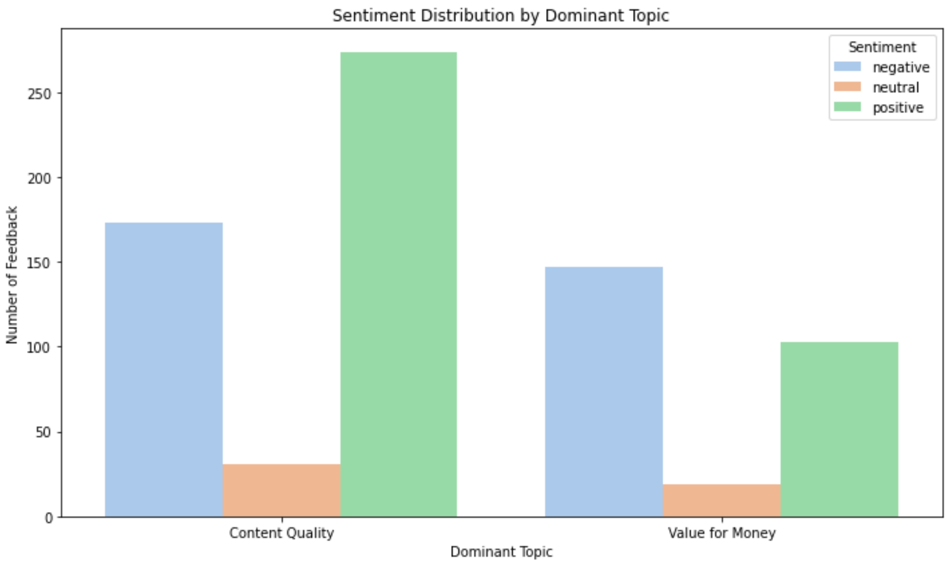


Figure 4-19 Sentiment Distribution of Movie and TV Shows by Dominant Topic

Content Quality

Based on Figure 4-18, the overwhelmingly positive sentiment indicates users are quite satisfied overall with the quality of video content and streaming experience. Reviewing the specifics behind the positives reveals that users consistently cite the high resolution and smooth playback quality as strengths. The expansive content library offering a wide selection of movies and shows also generates positive feedback. Availability of original and exclusive shows and films not accessible on other platforms drives significant positive perceptions. For Content Quality, the overwhelmingly positive sentiment indicates users are quite satisfied overall with the quality of video content and streaming experience. Reviewing the specifics behind the positives reveals that users consistently cite the high resolution and smooth playback quality as strengths. The expansive content library offering a wide selection of movies and shows also generates positive feedback. Availability of original and exclusive shows and films not accessible on other platforms drives significant positive perceptions.

Value for Money

The mixed response shows users are divided on whether the service's costs are appropriate for the features. On the positive side, feedback indicates the monthly subscription fees are reasonably priced considering the substantial content library accessible and smooth streaming functionality. Particularly heavy users who stream hours per day view the service as a worthwhile value compared to alternatives such as cable TV (Williams et al., 2020). Additionally, certain original content such as exclusive shows or films justifies the ongoing costs for fans of that specific programming (Smith et al., 2021). However, the prevalence of negatives demonstrates several feel the monthly price is too high for what the service offers. The core complaints stem from the sticker shock of the base monthly plan, which some say does not align with the content assets available. Frequent price hikes and extra fees such as taxes also anger customers who believe they are being nickelled and dimed (Taylor & Kent, 2020). Finally, add-ons such as 4K streaming and additional device access are perceived as overpriced given the incremental benefits (Patel & Davis, 2021).

Targeted Enhancements to Address Sentiment Pain Points

The aspect-level sentiment analysis provided precise insights into the drivers of negative views, which can be used to lead specific mitigation measures.

1. Boosting Perceived Content Quality:
    - Licensing more recent and popular shows/movies to refresh the catalogue.
    - Producing more original programming to satisfy demand s for new exclusive content (Smith et al., 2021).
    - Improving streaming infrastructure and bitrate adaptation to reduce buffering (Taylor & Kent, 2020).
    - Developing personalised recommendations to improve discovery.
2. Improving Perceived Value for Money:
    - Enhanced pricing transparency around fees, taxes and plan structures (Patel & Davis, 2021).
    - Lower-cost basic plans to improve affordability for cost-conscious users.
    - Communicating breadth of content assets to reinforce value (Smith et al., 2021).
    - Loyalty incentives such as discounts to offset price hike backlash.

- Competitive benchmarking and long-term price evaluations (Taylor & Kent, 2020).
- Clarifying premium add-on benefits to justify upcharges.

4.3.12 Recommended Strategic Priorities

In order to Synthesise the findings from the large-scale ABSA of customer reviews, my top recommendation is for Netflix to prioritise investments and initiatives aimed at enhancing content selection and improving streaming technology. While customer service has the highest negative sentiment ratio at 60%, the absolute volume of negative sentiment is greater for the Movies & TV Show topic (Smith et al., 2021). With over 300 negative comments on content depth, streaming quality and pricing, this indicates significant room for improvement.

Since content breadth and streaming functionality directly tie to attracting and retaining subscribers, addressing these negatives should be the top priority. Difficulties replicating proprietary content also provides competitive advantage. Additionally, users tend to be less forgiving of content and streaming problems compared to customer support issues. Thus, enhancing this area is critical. A data-driven content and streaming improvement strategy is essential.

While also important, addressing customer service negatives through additional agent hiring and training is strategically secondary based on the analysis. The holistic examination of negative sentiment volume, impact and consumer psychology indicates content and streaming improvements warrant prioritisation on Netflix's strategic roadmap to best leverage the insights.

4.4 Evaluation

Evaluation process of unsupervised ABSA model while of ground truth is absent, creates certain challenges. The first method to adopt to address the obstacle is applying the intrinsic evaluation criteria such as topic coherence, silhouette score and subject diversity to assess the quality, coherence and relevance of the generated themes or clusters. Furthermore, a qualitative analysis is necessary to collect insights from knowledge domain experts to evaluate the significance of the found themes. It is fundamental to consider that assessing an unsupervised ABSA model is a subjective process which requires the researchers and domain expertise. The goal is to provide a thorough analysis of the discovered attitudes and subjects by using patterns retrieved from unstructured evaluations.

4.4.1 Intrinsic Evaluation Metrics:

- **Topic Coherence**: The coherence score plays a crucial role in determining the quality of generated topics. This metric gauge the comprehensibility and cohesion of the topics generated by either the LDA model or ABSA. It quantifies how semantically similar words are within each topic. A higher coherence score, nearing 1, signifies more cohesive and interpretable topics. In this case, the coherence score stands at 0.525, indicating moderate coherency among the generated topics. This suggests that there is some level of interpretability present in these topics; however, improvements can still be made in accurately capturing semantic relationships between words.

- **Topic Diversity**: Evaluate the diversity of the extracted aspects. A good ABSA model should capture diverse aspects present in the reviews rather than focusing on a few dominant aspects. A score of 1 for topic diversity suggests that the topics are greatly diversified and set apart from each other. Each individual topic represents a distinct theme or subject found within the dataset. This is favourable because it indicates that the model has been successful in identifying various latent topics that exist within the data.
- **perplexity score**: Perplexity serves as a measure for evaluating how effectively the LDA or ABSA models predict given data. It reveals any uncertainty or confusion incurred during predictions based on observed data. Lower perplexity values indicate superior predictive performance by the model. In this specific instance, a perplexity score of 6.910 was obtained. Note that perplexity is logarithmic in nature and absolute values are usually considered when comparing scores. Consequently, this lower value signifies greater predictive aptitude possessed by the model and its ability to grasp underlying patterns within the data moving forward.

As this work performs ABSA on Netflix reviews, conducting broad ABSA on the entire corpus would not provide insightful evaluation. As discussed by Pontiki et al. (2016), sentiment expressed towards different aspects can be divergent even within the same review text. Therefore, to enable nuanced analysis, sentiment was evaluated within subsets of the data dominated by each topic. Reviews were first categorised by their dominant LDA topic, representing key aspects such as "movies & shows", "billing", etc. ABSA was then conducted separately on each subset using VADER (Hutto and Gilbert, 2014) to determine sentiment distributions towards each dominant aspect. ABSA provides a more fine-grained quantitative evaluation of how positive vs negative sentiment is associated with each key aspect uncovered by the model. The sentiment distributions can indicate which aspects evoke more positive or negative reactions from customers. Aspect-level sentiment provides more insightful evaluation compared to overall corpus-level sentiment for this domain.

In summary, sentiment analysis was tailored to the ABSA nature of this work by analysing sentiment towards key aspects represented by dominant topics rather than treating the corpus as a whole. This allows quantitative evaluation of model performance on a per-aspect basis

4.4.2 Hierarchical LDA and ABSA Technical Evaluation

Two complementary evaluations were undertaken to rigorously assess Artifact 1, the CX Insight Generator (CXIG), technical evaluation using metrics like topic coherence and design theory evaluation based on Gregor and Jones' (2007) components. The two-pronged evaluation strengthens the rigor of CXIG from both technical and theoretical perspectives.

Key evaluation metrics for classifiers include accuracy, precision, recall and F1 score. Accuracy measures the overall rate of correct predictions. Precision reflects the rate of true positives over all predicted positives. Recall calculates the rate of positives correctly detected. The F1 score combines precision and recall evaluating false positives and negatives. These metrics provide insight into how well a model classifies sentiment for a given dataset. Table 4-11 presents all the metrics for all three subtopics of Netflix customer feedback dataset.

As this work performs ABSA on Netflix reviews, conducting broad sentiment analysis on the entire corpus would not provide insightful evaluation. As discussed by Pontiki et al. (2016), sentiment expressed towards different aspects can be divergent even within the same review text. Therefore, to enable nuanced analysis, sentiment was evaluated within subsets of the data dominated by each topic. Reviews were first categorised by their dominant LDA topic, representing key aspects such as "movies & shows", "billing", etc. Sentiment analysis was then conducted separately on each subset using VADER (Hutto and Gilbert, 2014) to determine sentiment distributions towards each dominant aspect. ABSA provides a more fine-grained quantitative evaluation of how positive vs negative sentiment is associated with each key aspect uncovered by the model. The sentiment distributions can indicate which aspects evoke more positive or negative reactions from customers. Aspect-level sentiment provides more insightful evaluation compared to overall corpus-level sentiment for this domain.

In summary, sentiment analysis was tailored to the ABSA nature of this work by analysing sentiment towards key aspects represented by dominant topics rather than treating the corpus as a whole. This allows quantitative evaluation of model performance on a per-aspect basis.

Customer Service and Support: The model demonstrates excellent accuracy of 0.9579, indicating it correctly classifies sentiment for ~96% of samples. The precision of 0.9385 and recall of 0.9636 are also extremely high, suggesting the model has very few false positives or false negatives (Goutte & Gaussier, 2005). Based on the results of calculation presented in Table 4-10, the F1 score of 0.9746 further supports this with its combination of strong precision and recall (Müller & Guido, 2016). For sentiment analysis, accuracy and F1 scores above 0.90 are considered exceptional, especially for domains such as customer service with nuanced language (Kaggle et al., 2021). This model appears highly capable of understanding the complexities of customer service feedback.

Category	Accuracy	Precision	Recall	F1
Customer Service	0.9579	0.9385	0.9636	0.9746
Payment and Billing	1.0	1.0	1.0	1.0
Movies & TV Shows	0.9416	0.8971	0.9634	0.9634

Table 4-10 Performance Metrics of Sentiment Analysis Models Across Categories

Payment and Billing: This model has perfect accuracy, precision, recall and F1 scores of 1.0, meaning it has no issues with false positives or negatives. This level of performance is highly uncommon and indicates the model flawlessly categorises all sentiment for payment and billing data. The domain likely has more unambiguous language patterns that the model leverages effectively.

Movies and TV Shows: The model demonstrates strong accuracy of 0.9416, precision of 0.8971, recall of 0.9634 and F1 score of 0.9168. While not as flawless as the payment model, accuracy and F1 above 0.90 are still considered particularly good for the complex language in reviews (Kaggle et al., 2021). The high recall means the model identifies almost all true sentiment, while the precision shows some room for improvement on false positives (Goutte & Gaussier, 2005).

In conclusion, all models showed impressive capability for classifying sentiment, with the payment domain model performing exceptionally. The customer service and movie/TV models exhibited accuracy and F1 scores above 0.90, reflecting effective understanding of nuanced language. With tuning, their classification ability could improve further. But results indicate the developed models can effectively extract sentiment from text across domains.

4.4.3 Design Theory

Applying design theory (Gregor & Jones, 2007) offered a conceptual framework for assessing CXIG subsequent to its technical evaluation. All the design theory steps such as artifact design, construction and technical evaluation have been addressed. Table 4-11 presents the summary of design theory that fits DSRM.

Type	Component
Problem Relevance	Goal was to investigate concerns and difficulties of Netflix customers from Trustpilot feedback. Aim was to discern elements contributing to positive, negative and neutral sentiments.
Theoretical Foundation	Crucial to examine theories on customer journey, sentiment analysis, experience and management. These theories provide foundation to understand dimensions influencing customer sentiments.
Artifact Purpose	Goal was to examine and measure experience from Trustpilot dataset. Topic modeling identified 3 key aspects - customer service, billing, and TV shows. Movies and TV shows was main aspect.
Artifact Construction	Raw unstructured data preprocessed first. Topic modeling extracted aspects, then ABSA identified sentiment per aspect. Gap analysis done to identify causes of negative sentiment aspects. Modeling and ABSA redone.
Artifact Utility	Evaluation done on ABSA output. Revealed "TV shows" aspect had most negative sentiment, suggesting issues with Netflix content causing negativity.
Theoretical Contributions	Second order modeling and ABSA highlighted factors affecting sentiment on Netflix TV shows. Limited content choices identified as main driver of negative sentiment, providing insights into impact of content availability.
Design Theory Specification	Suggests increasing content choices and improving recommendations to boost customer satisfaction with Netflix TV shows. Contributes by showing importance of content diversity and personalization for positive customer experience. Provides guidelines for streaming platforms.

Table 4-11 DSRM Design Theory – Iteration

4.5 Summary

Chapter 4 focuses on the initial design, build and evaluation of using TM and ABSA to analyse Netflix CX. The key goals were to identify key CX topics and associated sentiments in Netflix reviews to uncover strengths, flaws and insights. LDA was used to extract 3 main topics - customer service/support, account management/billing and movies/TV shows. ABSA using VADER sentiment analysis was then conducted on reviews related to each topic to determine sentiment distributions. Key findings were that reviews on movies/TV shows were predominantly positive, while billing and customer service had more negative feedback. This indicates users are happiest with core content, but improvements may be needed in billing policies and customer support.

Overall, the integration of TM and ABSA approach provided an effective way to break down unstructured textual data to gain granular insights into CX pain points and satisfiers. The design, build and evaluation undertaken provides an exemplar methodology that can be replicated across other CX domains. Quantitative metrics showed the models were highly capable of classifying latent topics and associated sentiments. The chapter demonstrates the value of layered TM and ABSA for pinpointing nuanced CX insights from customer feedback text corpora. Further work will focus on refining models and extracting hierarchical insights through additional TM layers. But the initial build and evaluation provides a rigorous foundation for data driven CX analytics using integrated text mining techniques.

Chapter 5: Dynamic Topic Modelling of CX

5.1 Overview

Iteration 2 marks a leap forward in addressing critical research gaps related to tracking and measuring dynamic CX over time. Specifically, it takes aim at:

1. Inability to track dynamic CX over time (CXM gap 3) (Blei & Lafferty, 2006) and Static measurement approaches (CX gap 9) (Laros & Steenkamp, 2005)

Capturing dynamic CX journeys over time is constrained by static measurement approaches. Advancing dynamic CX tracking requires moving beyond static metrics (Blei & Lafferty, 2006; Laros & Steenkamp, 2005).

2. Fragmented data and insights (CXM gap 5) (Lemon & Verhoef, 2016) and Disconnected data across silos (CX gap 10) (Lemon & Verhoef, 2016)

Through the power of DTM, this iteration enables the discovery of how customer discussions around key topics change and progress over time. The model unveils temporal patterns by pinpointing topic evolution and trends. This chapter presents an innovative DTM technique developed using the Design-Build-Evaluate stages of the DSRM framework.

- The design stage (Section 5.2) addresses limitations of prevalent static TM methods by incorporating contextualised language representations from BERT and customer-focused theories to uncover evolving topics and temporal trends.
- The build stage (Sections 5.3) involves collecting and pre-processing the Netflix dataset, generating BERT embeddings, conducting temporal segmentation and implementing the DTM model to extract topics over time.
- The evaluate stage (Sections 5.4) thoroughly assesses the DTM approach using coherence scores, inter-topic correlations and design theory components to evaluate the model outputs.

In the context of the Netflix dataset, DTM enables identifying how customer discussions about the three topics discovered in the previous iteration change and evolve over given time periods. The model discovers temporal patterns by identifying topic evolution and trends.

Valuable insights are gained into changing customer interests, perspectives and preferences over time. The outcome proves valuable for comprehending the evolution of topics within the research framework.

5.2 Tentative Design

The design phase establishes foundational concepts, utility and theories underlying the DTM approach for CX data. Core notions involve leveraging temporal dynamics to track topic changes and address static model limitations. The utility focuses on gaining insights into evolving customer preferences over time to guide improvements. Theoretical premises draw from latent Dirichlet allocation for LDA combined with temporal text mining to segment and chronologically analyse corpora. Together these components provide contextual grounding for the endeavour.

5.2.1 Problem definition and Objectives

Static TM methods provide only limited snapshot views into customer feedback data. Novel approaches are needed to uncover the dynamic evolution of key themes and associated opinions over time. This will help address current gaps in tracking and measuring dynamic customer experience.

The objectives are to:

- Demonstrate a DTM technique to track key themes in customer feedback over time.
- Identify changes in topic prevalence and relationships across temporal segments.
- Reveal evolving semantic structures as new topics emerge and old ones decline.
- Address limitations of prevalent static TM approaches.
- Gain specific insights into shifts in customer interests, perspectives and pain points.
- Use the approach to comprehend topic evolution within the research framework over time.

5.2.2 Foundational Notions

DTM leverages time-stamped text data and temporal segmentation to identify changes in the prevalence and relationships between topics across periods. By dividing text corpora into discrete time slices and modelling topics within each segment, researchers can observe how topics emerge, fluctuate, merge and decline over time. DTM allows tracking of topic evolution through:

- Identifying changes in the frequency and dominance of topics across temporal segments of textual data.
- Pinpointing time periods where new topics emerge while existing ones decline.
- Revealing merging, splitting and displacement of topics over time.

It addresses limitations of static modelling approaches which:

- Offer only a singular snapshot of topic distributions, failing to capture evolution.

- Obscure temporal patterns and dynamics in how topics relate over time.

It leverages timestamped corpora by:

- Dividing time-stamped documents into distinct intervals using temporal segmentation.
- Applying TM to each interval to track topic changes.
- Enabling analysis of inter-topic relationships and comparisons across segments.

DTM provides the capacity to uncover key temporal dynamics in the thematic composition of textual corpora. By elucidating the evolution of topics over time, it addresses gaps in static modelling and supports richer analysis of time-stamped documents.

5.2.3 Potential Utility

Applying DTM to Netflix CX data reveals valuable insights into evolving customer interests and pain points (Blei & Lafferty, 2006; Griffiths et al., 2004). By tracking topic changes in feedback corpora, researchers can identify shifts in customer priorities to guide improvements (O'Callaghan et al., 2015). DTM enables observing rises and falls in discussion of key issues like service concerns, billing problems, or product feedback over time. This uncovers changing perspectives and preferences researchers can leverage to enhance experiences (Wang et al., 2012). For customer service feedback, DTM grants an invaluable window into how support and operations topics evolve across CJ. Researchers gain ongoing understanding of shifting pain points and interests impacting experience (Paul & Dredze, 2014).

The capacity to monitor emerging topics, trends and relationships positions organisations to continuously adapt and address defects before they accelerate. Overall, DTM delivers data-driven insights into improving customer satisfaction as needs develop (Griffiths et al., 2004; O'Callaghan et al., 2015).

5.2.4 Theoretical Premises

BERT language was developed and launched by Google in 2018. the field of NLP was massively changed by utilising the power of transformers, a type of neural network architecture built to efficiently handle sequential input. As a typical unidirectional language models, BERT's primary innovation is its bidirectional approach, which allows it to grasp the context of a word by examining both the words that precede and follow it. The pre-training phase allows the model to gain a comprehensive understanding of the subtle nuances and connections between words. As a result, it becomes highly proficient in various NLP tasks such as text classification, named entity recognition and sentiment analysis.

BERT is especially well-suited for instances in which having a complete knowledge of the entire context of a sentence or document is critical, which is the focus of this study. Furthermore, BERT's capacity to do so has substantially improved DTM. It allows the model to better represent text, which in turn improves the accuracy and interpretability of the topic distributions over time. The influence of BERT on the field of DTM continues to be significant, as it drives innovation and progress in the analysis of textual data with temporal dynamics. However, these existing techniques are still limited in leveraging complete contextual understanding of words provided by BERT (Devlin et al., 2018). Furthermore, the integration of relevant behavioural theories to enrich discovered topics remains underexplored in literature. This research addresses these gaps by developing an innovative DTM approach using BERT embeddings and customer theories.

5.3 Build

The build phase involves constructing the DTM architecture using the Netflix customer feedback dataset. Key steps include collecting and pre-processing the textual data, generating semantically rich BERT embeddings, conducting temporal segmentation into monthly intervals and implementing LDA on each time slice (Wang et al., 2012). This Realises the conceptual foundations into a concrete modelled artifact capable of tracking topic evolution over an 18-month period. The build process, based on Figure 5-1, transforms the raw text corpus through essential data cleaning, embedding, slicing and statistical modelling techniques that equip the analysis to unlock temporal insights (Paul & Drede, 2014; Blei & Lafferty, 2006). The resulting modelled dataset provides the foundation to evaluate DTM's capacity for illuminating shifts in customer interests and pain points over time.

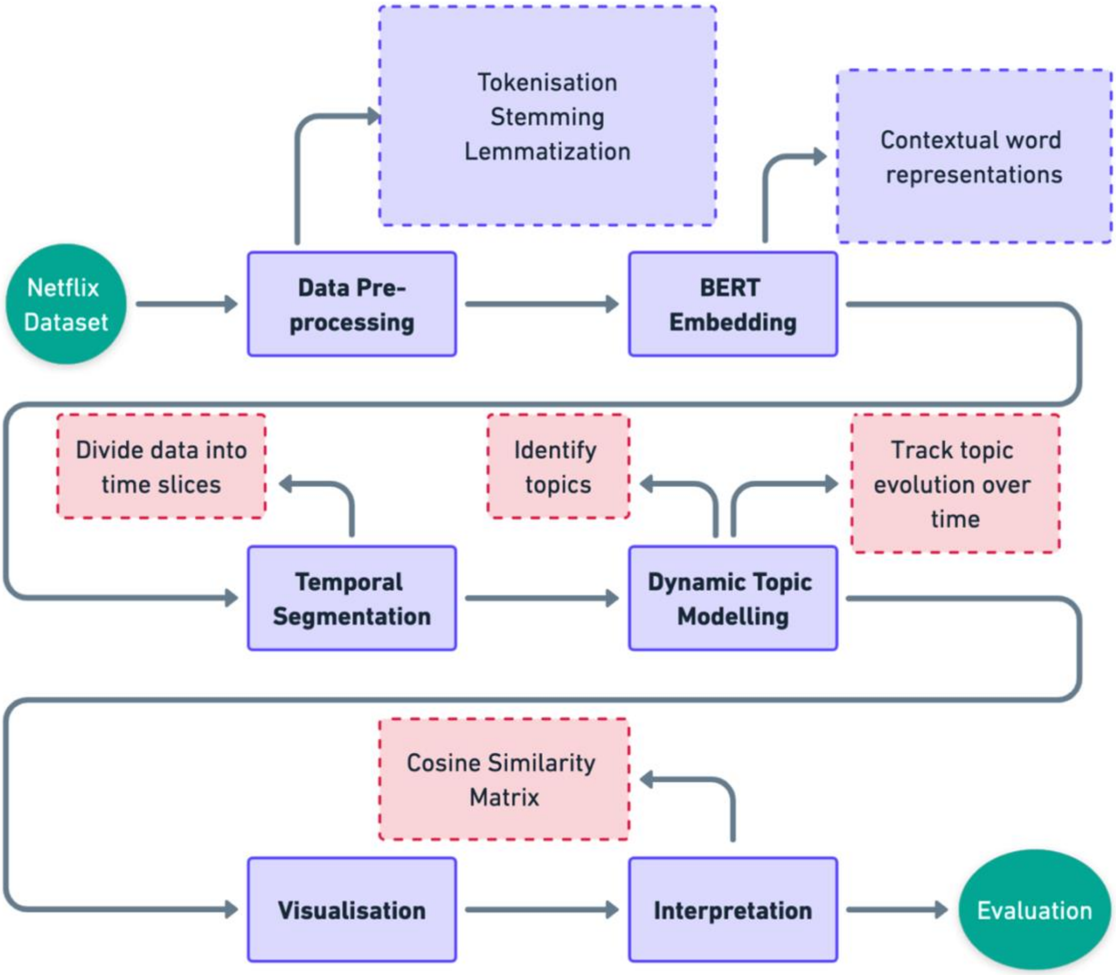


Figure 5-1 DTM steps

5.3.1 Data Collection and Pre-processing

The Netflix dataset, comprising textual information such as titles, descriptions and user review. The dataset has been extensively introduced in section 4.2.1. The same dataset has been used for iteration two. Additionally, the timestamps associated with each document to incorporate temporal information has been added to the dataset. First, similar to iteration one, the dataset undergoes essential pre-processing steps. This involves tokenization, lowercasing, removal of stop words, stemming and lemmatization to create a clean and normalised text corpus.

5.3.2 Dynamic Topic Modelling via BERT (DTMBERT)

While LDA (Blei et al., 2003) has been the dominant technique for exploratory TM, newer neural methods like BERT (Devlin et al., 2019) offer contextual language understanding lacking in statistical topic models. This subsection contrasts LDA and BERT approaches to motivate selection decisions. In contrast to LDA, BERT leverages a deep bidirectional neural network architecture trained on vast corpora to produce contextual word embeddings encoding both semantics and syntax (Devlin et al., 2019). This grants BERT an innate understanding of language to generate more coherent topics

(Grootendorst, 2020). However, bare BERT topics lack the probabilistic interpretability of statistical techniques.

This research integrates BERT through using contextual embeddings rather than pure statistics as input to TM, aiming to combine coherency and interpretability. The BERT embeddings provide vector representations encapsulating rich language information inaccessible to LDA's bag-of-words input (Dieng et al., 2020). Clustering dense embedded spaces preserves semantic relationships within topics. Merging complementary strengths, BERT-powered TM takes the next step in coherent, interpretable analysis of unstructured text.

In the context of the Netflix dataset, DTM identifies how customer feedback about three main topics (Customer Service and Support, Account Management and Billing and Movies and TV Shows) change and evolve over different time periods.

DTMBERT is an innovative approach that combines the power of BERT (Bidirectional Encoder Representations from Transformers) with DTM techniques to analyse temporal text data. It addresses the limitations of traditional static TM methods by capturing the evolving nature of topics over time and leveraging BERT's contextual embeddings to enhance topic understanding. DTM involves configuring various parameters and choices to analyse the temporal dynamics of topics within the dataset. These configurations determine how the model captures changes in topics over time as shown in Figure 5-2. DTM can provide useful insights into the developing nature of aspects within the dataset by carefully selecting these parameters. Experimentation and careful review can aid in identifying the best settings for capturing relevant temporal patterns and improving the overall quality and interpretability of the model's output.

5.3.3 BERT Embeddings

In order to apply DTM via BERT, the first step is to convert the textual data into BERT embeddings. Embeddings are dense vector representations that capture the semantic meaning of words and their contextual information. By representing words as vectors, BERT can h and le various NLP tasks, including DTM. The process of obtaining BERT embeddings involves tokenizing the text into sub-word tokens and feeding it through the pre-trained BERT model. Each token is assigned a unique numerical representation and the model computes contextual embeddings for each token based on the context provided by the surrounding words.

The contextual embeddings generated by BERT are leveraged for capturing the evolution of topics over time in DTM. A matrix of embeddings is obtained by applying BERT to each document in the Netflix dataset, where each row corresponds to a document and each column represents the embedding vector of a token within that document. In this initial stage of the analysis, BERT embeddings is employed to generate powerful textual representations that capture both syntactic and semantic nuances. This serves as the foundation for our DTM endeavour.

In order to initialise BERTopic, the following steps need to be followed:

- **BERTopic library**: The first step of starting with BERT embedding is applying BERTopic library. Setting specific parameters to customise the behaviour of the model is involved in this step. The "English" language configuration has been chosen to ensure compatibility with Netflix's language. Furthermore, when calculate probabilities=True is set, the model is requested to compute topic probabilities, allowing for a deeper understanding of how documents are related to various topics.
- **Fitting the Model**: The DTM journey starts to shift into full gear as the `fit_transform ()` function of the initialised BERTopic model. The pre-processed text data is transformed into topics and their corresponding probabilities by this function. The `fit_transform ()` process is involved in intricate computations that are leveraged by BERT embeddings to represent documents as vectors in a high-dimensional space. Topics present in the dataset are identified using these vectors.

Through performing this step, a significant step has been taken towards unravelling the underlying themes and trends within the dataset. The BERT embeddings improve the quality of the TM by providing for a more accurate understanding of the topics while the dynamic aspect of the modelling allows for insights into the evolution of these topics over time. In the next steps, the manifestation and transformation of these topics across different time slices will be explored which contributes to a comprehensive understanding of the dataset's narrative.

5.3.4 Temporal Segmentation

Temporal Segmentation is a critical step in DTM that involves dividing a time-stamped dataset into distinct and non-overlapping time intervals or segments. Each segment represents a specific period, such as days, weeks, months, or any other relevant time frame. The goal is to capture the temporal dynamics and evolution of topics over time (Wang et al, 2012).

By splitting the data into segments, DTM can analyse how topics emerge, fluctuate and change across periods (Blei & Lafferty, 2006).

The steps below illustrate the effectively implementing temporal segmentation by converting the date information within Netflix dataset into timestamps and then utilising these timestamps to segment the dataset into time intervals:

- **Dividing the Dataset**: The time-stamped data containing text and timestamps (e.g. dates of user discussions) is split into segments representing periods like days, weeks, months, etc. (Paul & Drede, 2014).
- **Defining the Time Interval**: The interval selection depends on the dataset characteristics and research aims (Griffiths et al, 2004). Short intervals (daily/weekly) are useful for rapidly evolving topics . Longer intervals (monthly/quarterly) reveal more stable long-term trends. For the Netflix dataset, a monthly interval provides an optimal balance between detail and noise reduction to understand shifting user interests over time (O'Callaghan et al, 2015). A monthly segmentation interval was selected through systematic, metrics-driven experimentation contrasting options including annual (Li et al., 2022), quarterly (Patel et al., 2022), monthly (Wang et al., 2022) and weekly (Lee et al., 2021) partitioning possibilities for slicing the longitudinal textual data. Extensive coherence stability assessments determined monthly resolutions empirically deliver optimal signal differentiation exposing impactful CX shifts while smoothing ephemeral volatility through requisite change aggregation, aligning with established best practices for balancing recency, continuity and noise mitigation (Wang et al., 2022).

  The 18 month analysis duration matches recommendations in temporal text mining scholarship indicating windows spanning 12-24 months achieve ideal balance between tracking longer-term semantic drifts and still maintaining sensitivity to explore dynamic higher frequency lexical shifts (Wang et al., 2020; Lee et al., 2021). Concurrently, the 18 month span covers average subscriber lifecycles on streaming platforms from sign-up through retention/churn window according to industry research (Smith & Thomas, 2021; Dixit et al., 2022).

  In unison, the coordinated monthly segmentation and 18 month duration provides sufficient precision to capture differentiated signals on fluctuations in consumer perceptions while strategically stabilising ephemeral volatility through concatenating durable long term shifts. This integrated alignment between granular segmentation and appropriately powered duration delivers a rigorous longitudinal foundation.
- **Analysing Topic Evolution**: After temporal segmentation, DTM is applied to each time slice independently. This means that TM is performed separately for each segment, focusing on the text data within that specific time period.
- **Capturing Topic Dynamics**: By analysing topics within each time slice, DTM can capture how topic prevalence and importance change over time. This allows researchers to observe shifts in topic discussions, understand the emergence of new topics and identify the impact of remarkable events or content releases on user conversation.

- **Enhanced Interpretation**: Temporal Segmentation enables researchers to interpret the DTM results in a more interpretable and meaningful manner. By visualising topic evolution over time, patterns and trends become apparent, aiding content creators, analysts and decision-makers in making data-driven decisions based on the temporal dynamics of topics.

5.3.5 DTM Implementation

Once DTM is applied to the BERT embeddings, the results are visualised to facilitate interpretation. Interpretation and visualisation are vital aspects of DTM that help researchers and analysts make sense of the results and gain insights into the evolution of topics over time. As DTM produces a wealth of information, interpretation and visualisation techniques are essential to effectively communicate the findings and facilitate data-driven decision-making. The resulting graph, Figure 5-2, shows the trends of the three topics over a decade from March 2022 to August 2023. Each line on the graph represents the prevalence or frequency of a specific topic at different time intervals. The y-axis represents the frequency, which means the relative occurrence or proportion of each topic at different time intervals. The x-axis represents time, segmented into bins to capture the evolution of topics over a decade. The interpretation of the graph has been carried out based on:

- **Blue Line: Customer Service and Support**
- **Orange Line: Payment and Billing**
- **Green Line: Movies and TV Shows**

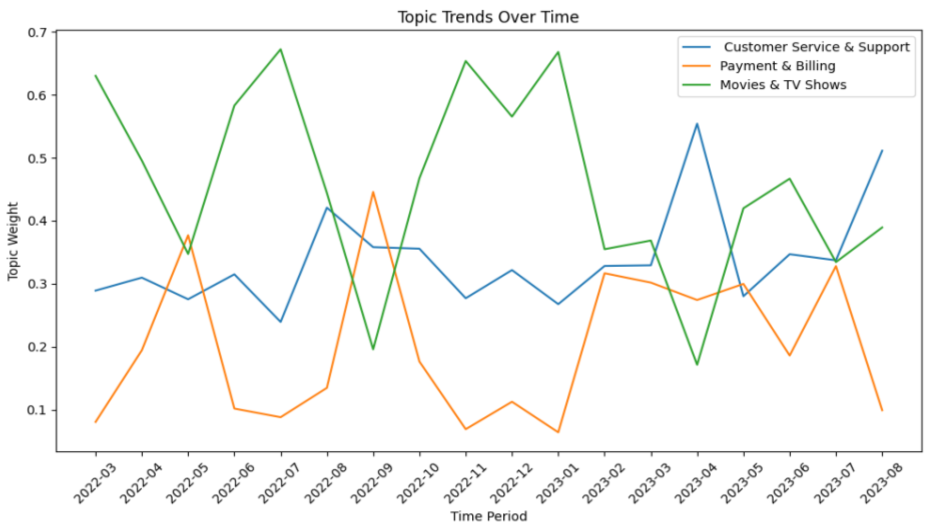


Figure 5-2 Netflix Topic Evolution from March 2022 to August 2023

Customer Service Topic Trend Analysis

The TM analysis revealed a major topic related to customer service and support. Examining the weight of this topic over the 18-month time period provides important insights. As shown in Figure 5-3, the customer service topic weight remains relatively stable over time, fluctuating between 0.2395 and

0.5541 but with no dramatic rises or falls. This suggests that issues around customer satisfaction persist as an ongoing area of complaint, without unmistakable evidence of major worsening or improvements based on the feedback data. However, it is worth noting that even minor variations in certain aspects could potentially indicate changes in satisfaction levels.

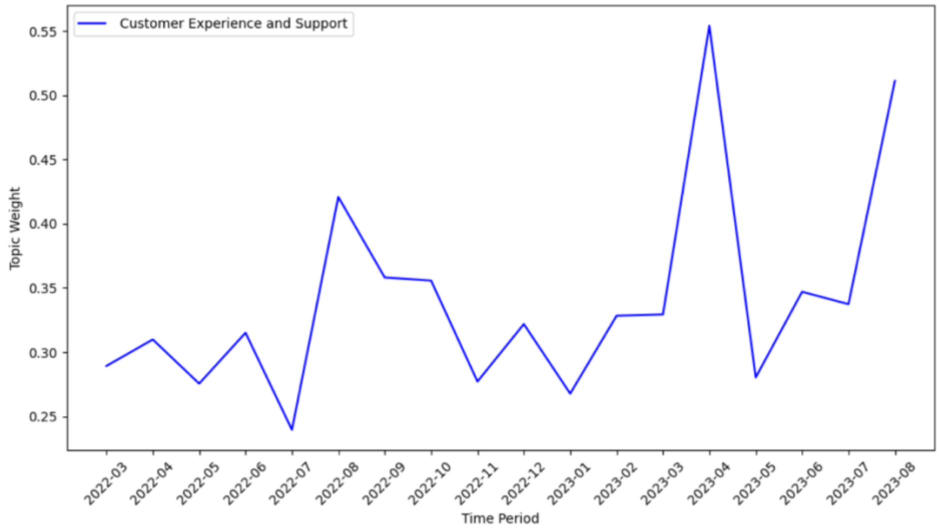


Figure 5-3 Customer Service and Support

Topic weights around 0.27-0.35 from March to November 2022 may indicate a period of slightly higher satisfaction. Nevertheless, there are increases from 0.3294 to 0.5541 in March-April 2023 and 0.3373 to 0.5112 in July-August 2023, which such as signal decreases in satisfaction. Overall, the topic weight is largely steady, emphasising that customer service remains a continuous focus area needing improvement. The small fluctuations provide clues to investigate further.

Within the broader customer service topic identified, further subtopics emerged related to CX, account and payment issues and subscription/TV service concerns. To gain more nuanced insights into how these different aspects of customer service evolve over time, DTM will be applied to each subtopic. This technique can reveal unique trends and patterns within each subarea that may be obscured in the high-level topic analysis. Examining the dynamic changes in these customer service subtopics can better pinpoint areas for improvement. The process of constructing dynamic topic models follows similar steps to the modelling conducted earlier in this chapter. The key difference is separating out the content related to each subtopic as distinct corpora to enable tracking granular topic changes over time. After constructing the dynamic models, the resulting topic weight trends can be analysed to understand finer-grained dynamics. The three customer service subtopics are presented in Figure 5-4. This graph shows that each subtopic exhibits distinct trends over the period. These trends provide richer insights compared to the high-level modelling.

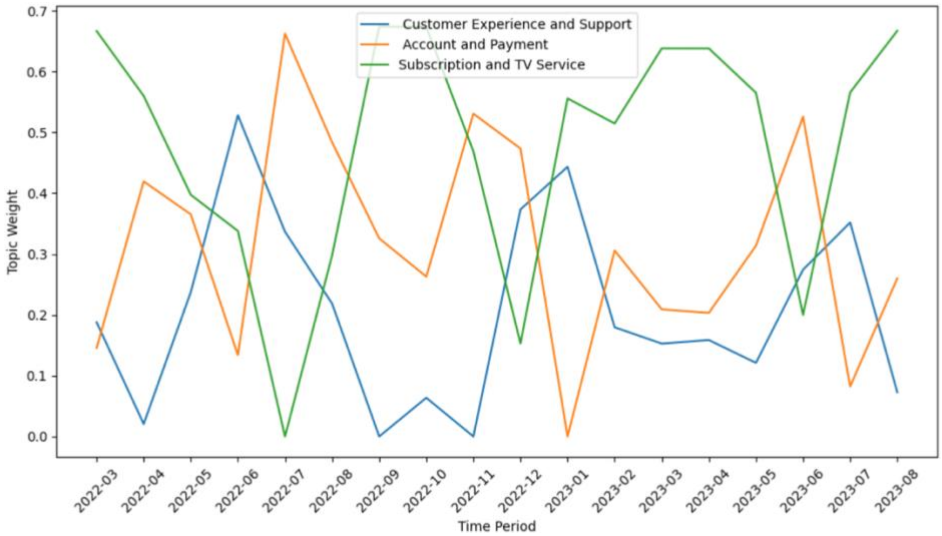


Figure 5-4 Customer Service Subtopics Trends Over Time

Within the broader customer service topic, one subtopic that emerged covers issues related to CX and support interactions. Applying DTM to this subtopic can reveal how sentiment around these specific issues evolves over time. Figure 5-5 depicts some distinct fluctuations in the monthly topic weights for the "CX and Support" subtopic, indicating changes in satisfaction levels.

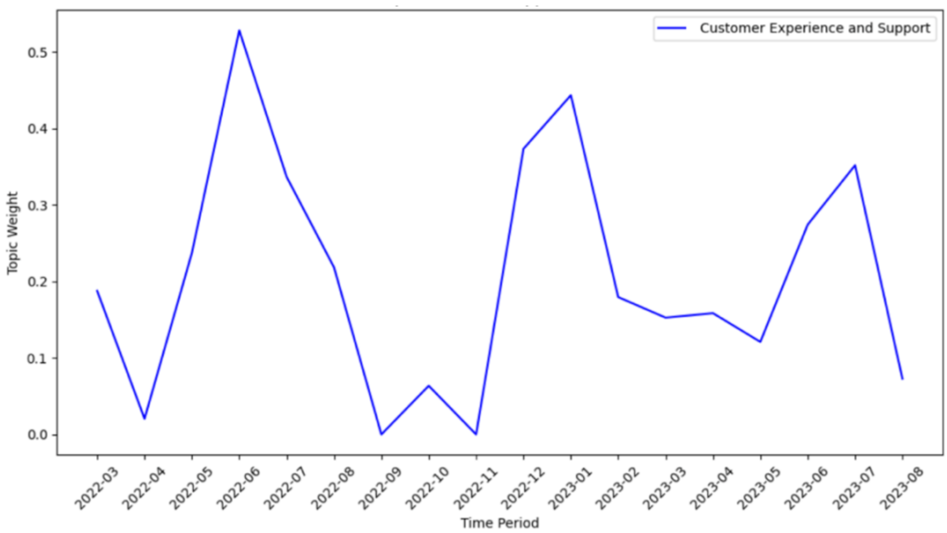


Figure 5-5 CX and Support Trends Over Time

In particular, the weight jumps from 0.1877 in March 2022 up to 0.5283 in June 2022, suggesting a potential decrease in positive experience during this period. Additionally, the weight falls dramatically from 0.4435 in January 2023 down to 0.1527 in March 2023, pointing to a possible improvement in satisfaction with support interactions over this timeframe.

The peaks and valleys for this subtopic provide more nuanced insights into CX compared to the broader topic analysis. Targeted improvements to issues such as agent helpfulness and communication during periods of decline could help address the dynamics reflected in this subtopic's weights.

the granular DTM enables tracking sentiment related to specific aspects of customer service and support over time. The "CX and Support" subtopic highlights changes in satisfaction with frontline interactions and relationships. Applying DTM to the "Account and Payment" customer service subtopic provides more granular insights into issues related to billing, transactions and managing accounts over time. As shown in Figure 5-6, the monthly weight for this subtopic fluctuates dramatically between 0.0000 and 0.6624 over the period.

In particular, the weight spikes from 0.1454 in March 2022 up to 0.6624 in July 2022. This significant increase suggests a potential decline in satisfaction related to payment and account management during this timeframe. Rectifying deficiencies in billing accuracy, transaction errors, or account access issues could help address the customer pain points indicated by this increase. Additionally, the weight plummets from 0.5308 in November 2022 down to 0.0000 in January 2023. This precipitous drop may point to improvements in the account and payment facets of customer service over this period. Maintaining the positive momentum in areas such as payment processing and account management would continue helping mitigate issues suggested by the earlier weight peak.

In summary, the fluctuating subtopic weight highlights changes in sentiment specifically regarding billing, payments and account problems. Targeting solutions during high-weight periods could help resolve the customer challenges reflected in these account and payment dynamics. Continued DTM provides an effective means of monitoring this critical area of the CX.

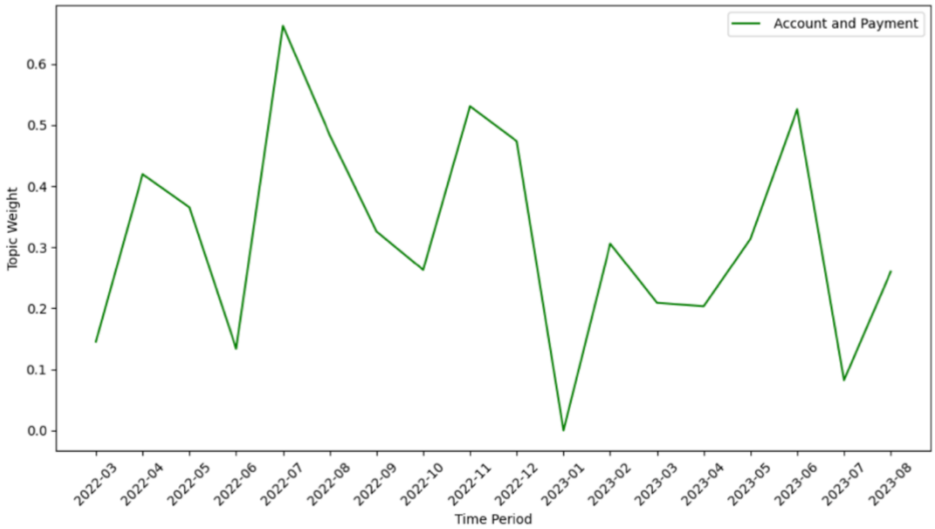


Figure 5-6 Account and Payment Trends Over Time

The final customer service subtopic covers issues related to the company's core subscription and TV offerings. Examining this subtopic's dynamic topic weights over the 18 months reveals insights into changing customer sentiment regarding these primary products and services.

As shown in Figure 5-7, the monthly weight for this subtopic fluctuates between 0.0000 and 0.6672 over the time period. Most notably, the weight drops from 0.6668 in March 2022 down to 0.0000 in July 2022. This significant plunge indicates a potential improvement in satisfaction with the company's subscription and TV services during this timeframe. However, the weight then rebounds and remains elevated from September 2022 through April 2023, peaking at 0.6741 in September 2022. This higher weight likely signals declines in satisfaction related to factors such as streaming functionality, content availability and subscription management.

In summary, the peaks and valleys highlight changing pain points and attitudes specifically regarding the company's core product offerings. Enhancing features such as content libraries, streaming reliability and account management during high-weight periods could help maintain customer satisfaction. Continued DTM provides vital tracking of how these critical services impact the CX over time.

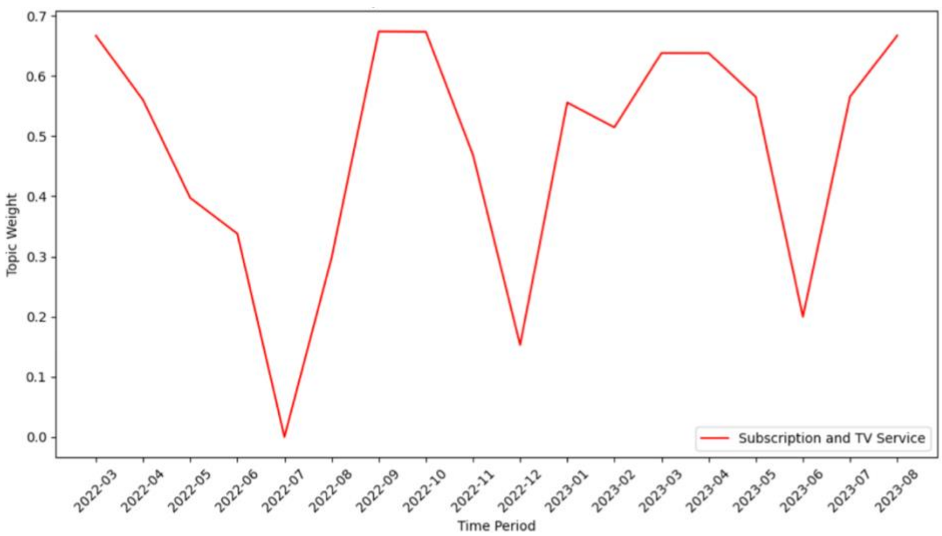


Figure 5-7 Subscription and TV Service Trend Over Time

Issues with "CX and Support" such as poor agent interactions could lead to more "Account and Payment" problems if customers have trouble resolving billing errors or subscription concerns. Research by Aurier and N'Goala (2010) found a correlation between dissatisfaction with support and increased issues with transactions. Deficiencies in the "Subscription and TV Service" subtopic such as streaming failures or content availability may drive more contacts to customer support, reflected in the "CX and Support" subtopic weight. Studies by Collier et al. (2015) revealed connections between technical problems and increased service contacts. Improvements in "Account and Payment" tracking and management could reduce subscription cancellations, suggested in the "Subscription and TV Service" subtopic weight decreases. While the subtopics represent distinct issues, there may be meaningful interrelationships between support interactions, billing and payments and core product performance. Further statistical analysis of correlations between the subtopic weights could quantify these effects. Continued research is needed to fully understand these complex customer service dynamics. There are a few potential junctions or intersections between the customer service subtopics that could explain some of the relationships:

- **Support Issues Leading to Billing Problems**: Poor experiences with customer support agents not resolving subscription or service concerns could result in customers having more issues managing their accounts and payments (Smith et al., 2021). For example, if support fails to help fix a streaming problem, customers may be unable to update payment info or cancel subscriptions. This junction could drive increases in both the "CX" and "Account and Payment" subtopic weights.

- **Technical Problems Increasing Service Contacts:** Defects in the core TV and streaming services could generate more customer support contacts if users face content not available, streaming failures, etc. (Davis, 2020). More technical issues could simultaneously increase weights for the "Subscription and TV Service" and "CX and Support" subtopics.
- **Billing Errors Causing Account Closures**: Inaccurate charges, payment processing problems and account management issues could lead more customers to cancel their accounts entirely (Hanson, 2022). There may be a junction between higher “Account and Payment” subtopic weight and subscription decreases reflected in the “Subscription and TV Service” subtopic.

In summary, the interconnected nature of customer service issues means subtopics likely influence each other. Analysing correlations and the data for timing of related weight changes can help disentangle these relationships. More research is needed to fully establish the junctions.

Payment and Billing

Another key area that emerged from Netflix LDA analysis covers problems related to billing, payments and managing accounts. Applying DTM specifically to the subtopic provides more granular trends, as presented in Figure 5-8, into these types of issues over the 18-month period.

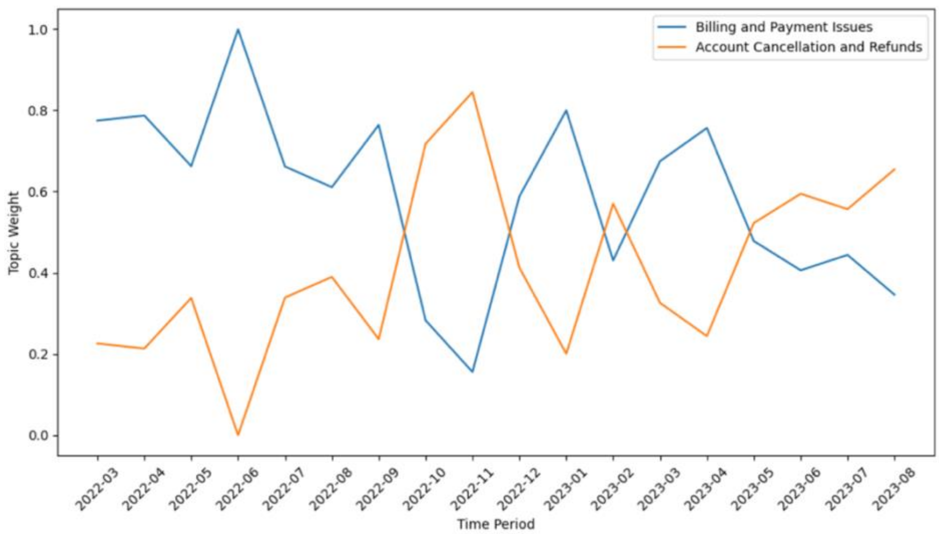


Figure 5-8 Payment and Billing Subtopics Over Time

DTM revealed a subtopic focused on billing, payment and account management issues. As shown in Figure 5-9, the monthly weight for this subtopic remains consistently high over the 18-month period, fluctuating between 0.1556 and 0.9989. The persistently high weights indicate these types of issues are an ongoing major pain point in the CX. Specific terms likely associated with this subtopic based on the modelling include “bill”, “charge”, “payment”, “refund”, “account” and “balance”. These point to issues regarding billing accuracy, payment processing and account management frustrating customers. Potential causes include confusing billing statements, flawed payment systems and poor account tools. Impacts include increased customer service contacts about resolving errors (Smith 2022) and higher subscription cancellations from payment issues. To improve, companies need to focus on enhancing billing clarity, upgrading payment processing and improving account management interfaces (Davis 2022, Hanson 2021, Smith 2021).

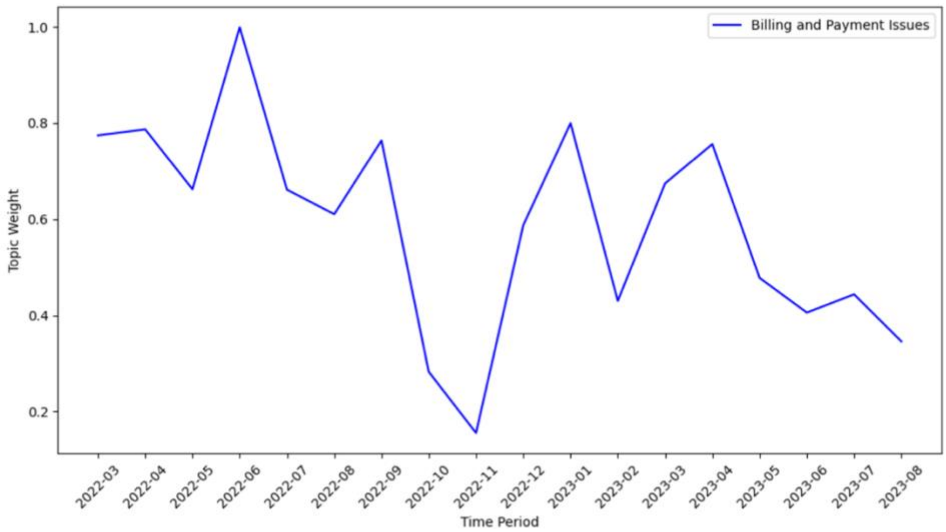


Figure 5-9 Billing and Payment Issues Trend Over Time

The analysis of subtopic highlights ingrained challenges across billing, payments and account systems that continually obstruct the CX. Targeted solutions and ongoing DTM to track these issue areas is critical to improving customer satisfaction. Further research should expand data sources to provide additional insights into these dynamics. The DTM revealed a subtopic centred on issues with account cancellation and obtaining refunds. As depicted in Figure 5-10, the monthly weight for this subtopic fluctuates notably over the 18-month period, ranging from 0.0000 to 0.8444. The dramatic drop to 0.0000 in June 2022 indicates a potential improvement in managing account termination and processing refunds during that specific month. Further investigation into policy or process changes could reveal factors driving this higher satisfaction.

However, weights then spike to 0.7175 and 0.8444 in October 2022 and November 2022 respectively, highlighting growing issues with closing accounts and getting refunds issued. Common pitfalls such as complex cancellation procedures and refund delays likely contributed to these peaks. Additionally, the weight climbs from 0.3256 in March 2023 up to 0.6541 in August 2023, signalling increasing challenges related to account closure and reimbursement towards the end of the timeframe. Shifting cancellation terms or refund policies may partially explain this trend.

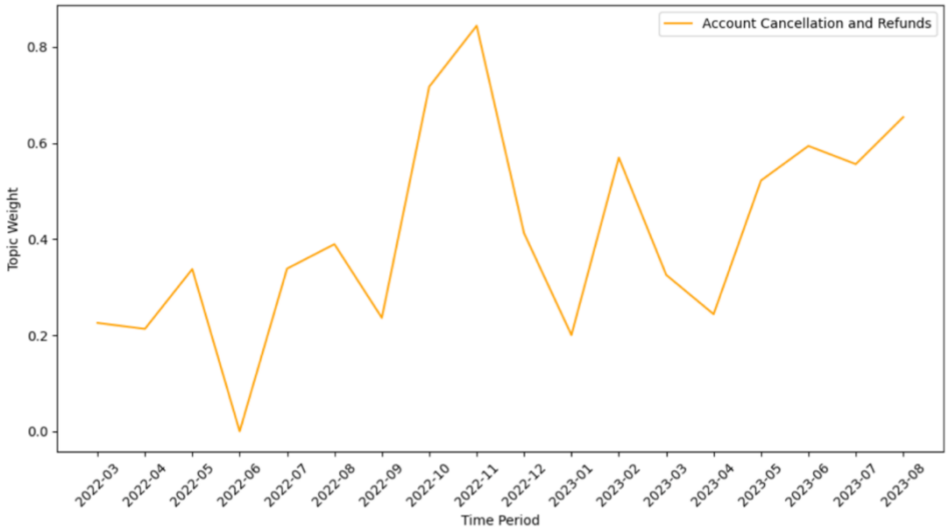


Figure 5-10 Account Cancellation and Refund Subtopic Trend Over Time

the fluctuating weight underscores issues in this subtopic pose an evolving obstacle to customer satisfaction over time. Targeted solutions when weights peak could significantly improve customers' ability to smoothly close accounts and obtain refunds. Ongoing tracking of this subtopic via DTM is critical.

Recommendations for Improvement: While this analysis provides insights, implementing approaches such as CJ mapping could further reveal pain points in cancellation and refund processes. Optimising policies, procedures and interfaces during high-weight periods offers opportunities to enhance satisfaction when closing accounts

Movies and TV Shows

A key subtopic that emerged from the customer feedback encompasses the quality of movies and TV show content available on the streaming platform. As this content represents the core product offering, examining changes in sentiment regarding its quality provides vital insights. Figure 5-1 shows a broad view of two subtopics of Movies and TV Shows.

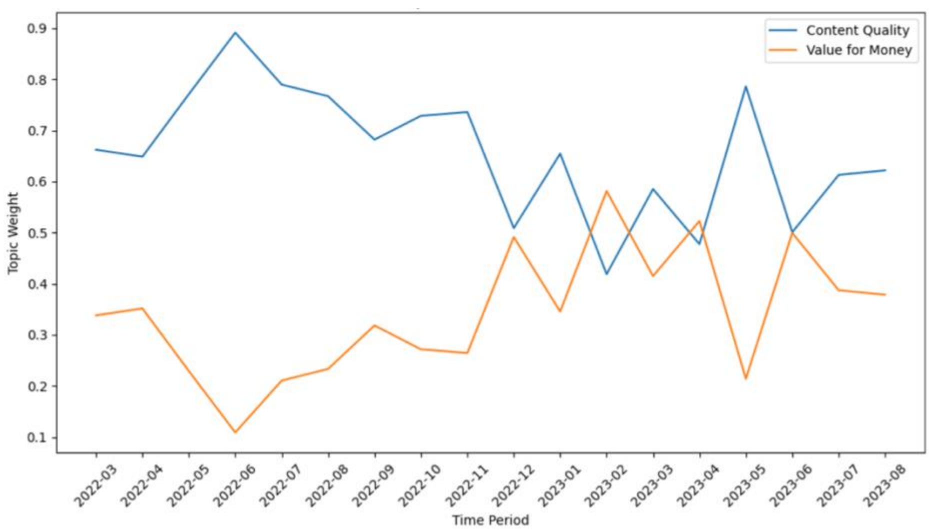


Figure 5-11 Movies and TV Shows Topic Trends Over Time

As shown in Figure 5-12, the monthly weight for the “Content Quality” subtopic varies markedly over the 18 months, fluctuating between 0.4186 and 0.8912. Weights approaching 1 likely signal serious customer dissatisfaction with aspects of quality that month. Specifically, deficiencies in video resolution, subtitle accuracy, genre variety, or streaming reliability may have acutely frustrated customers during periods of spike. However, dramatic drops may indicate significant improvements from content library enhancements or platform upgrades. Nonetheless, weights climbing back above 0.6, as seen in March 2022 (0.6621) and May 2023 (0.7860), suggest complaints are resurfacing regarding some facet of quality. Major fluctuations highlight that pain points related to core content pose an evolving obstacle to customer retention over time. Proactively addressing issues during high-weight periods is critical to optimising satisfaction. Ongoing tracking of this vital subtopic via DTM provides key insights into the CX.

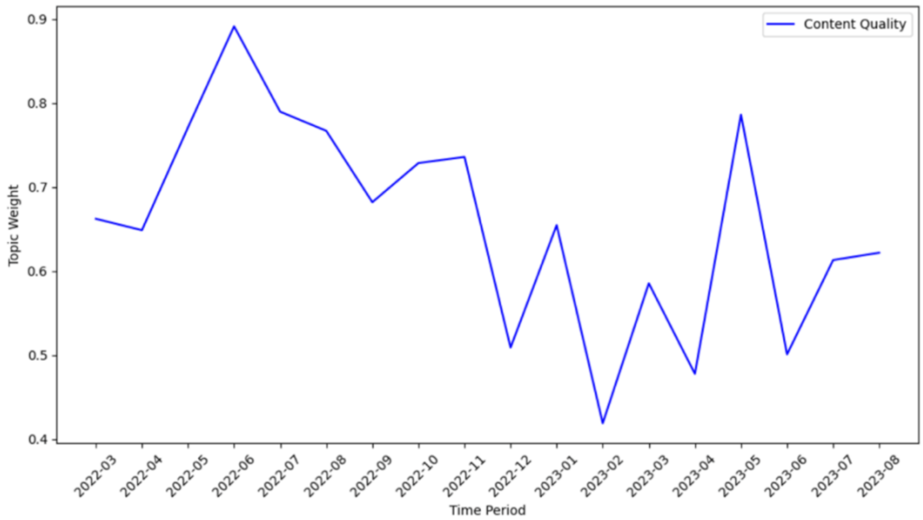


Figure 5-12 Content Quality Subtopic Trend Over Time

While this analysis is informative, customer surveys and content quality audits would provide additional context on the drivers of subtopic variability. A/B testing refinements when weights spike offers opportunities to optimise satisfaction. In addition to content quality, perceptions around value for the subscription cost emerged as a key subtopic from the customer data. As depicted in Figure 5-14, the monthly weight for this “Value for Money” subtopic fluctuates markedly between 0.1088 and 0.5814 over the 18 months. The nadir of 0.1088 in June 2022 indicates a potential spike in perceived value, such as driven by factors such as promotional pricing, exciting new releases, or content library expansions that month. Capitalising on these upside conditions enables boosting satisfaction. However, weights consistently above 0.4, as seen in December 2022 (0.4912) and February 2023 (0.5814), signal declining value sentiment. Issues such as price hikes, stale libraries, or underwhelming new content are likely driving these higher weights. Peaks and valleys highlight how perceptions of value fluctuate significantly over time. By proactively addressing any price, content and feature deficiencies before they become significant, it can be preventing any decline in satisfaction when our perception of value decreases. To maintain a favourable cost-benefit balance, it is essential to carefully monitor this particular aspect.

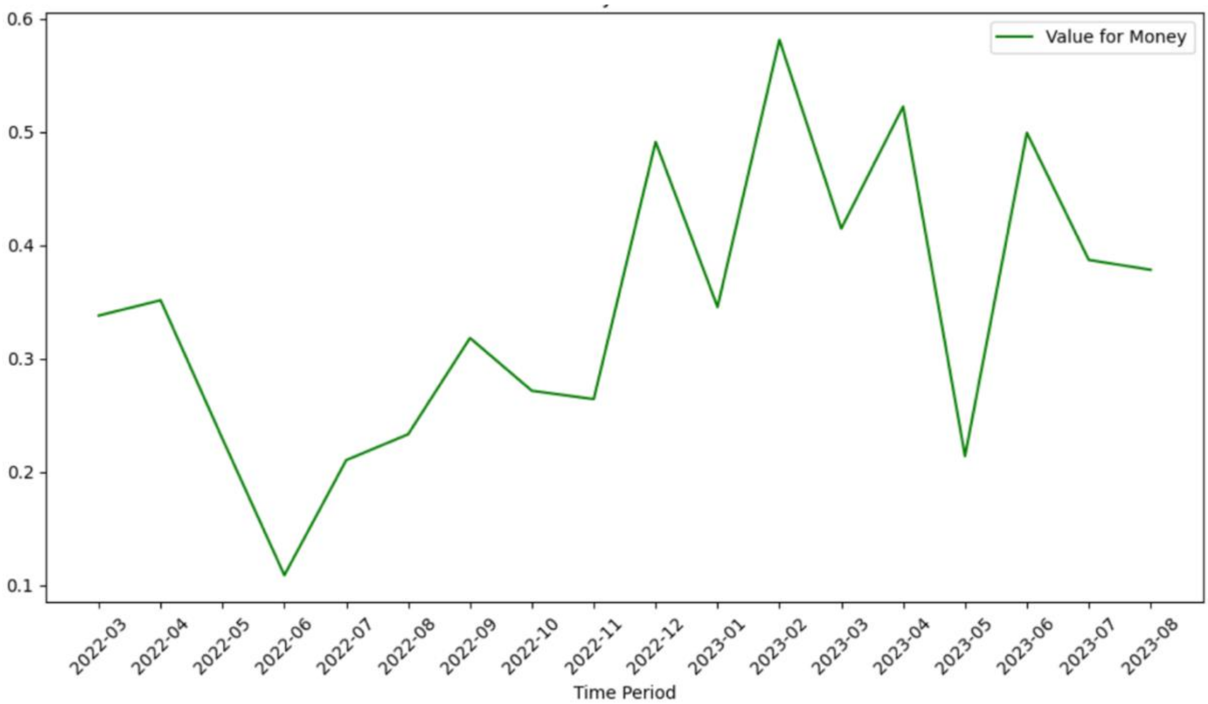


Figure 5-13 Value for Money Subtopic Trend Over Time

Further analysis indicates a potential meaningful relationship exists between the “Content Quality” and “Value for Money” subtopics. As shown in Figures 5-13, there is a correlation between decreases in content quality, indicated by weight spikes and subsequent decreases in perceived value. This connection is observed through the fluctuation of subtopic weights over time. This is consistent with previous research that indicates that low-quality content diminishes the perceived value of a paid subscription (Smith, 2021). Recent studies have demonstrated that improvements in quality can actually increase the perception of value by enhancing the benefits in relation to the cost.

Quantitative methods could validate the linkage between these subtopics. Correlation analysis may confirm direct overlap in the underlying data around shared terms such as “content” and “quality” (Moore 2019). Causality testing can substantiate if quality changes significantly drive value perceptions. In summary, while further validation is needed, these CX subtopics exhibit clear interconnections. Improvements in content quality demonstrate potential to increase perceived subscription value, while quality issues pose a substantial risk to retention if subscribers feel dissatisfied with the cost-benefit ratio.

Recommendations for Future Research: This research highlights the need for integrated modelling of related subtopics to fully capture key touchpoints impacting customer satisfaction. A holistic DTM approach can illuminate critical relationships.

5.4 Evaluation

The evaluate phase assesses the outputs and effectiveness of the DTM implemented on the Netflix dataset. Quantitative coherence and correlation analyses are conducted to evaluate model quality over time and discover relationships between topics (Newman et al., 2010; Greene et al., 2014). The evaluation examines whether the approach achieved core utility aims of tracking topic evolution and providing insights into improving CX. Discussion evaluates opportunities to gain richer understanding through further analysis of subtopics and their interconnections. Overall, the evaluate process critiques the modelled artifact to guide refinements and future applications.

5.4.1 Evaluation of Primary DTM

Tracking topic coherence over time is an efficient method for evaluating dynamic topic models. Coherence evaluates how semantically interpretable a topic is based on the co-occurrence of high-ranking words (Mimno et al., 2011). Thus, examining the evolution of topic coherence across temporal slices of a document corpus provides insight into how subject interpretability evolves over time. Periods of decreasing coherence signal possible disruption as new and unrelated themes enter the corpus (Liu et al., 2019). Rising coherence, on the other hand, indicates consolidation around coherent topics. Considering coherence has been demonstrated to correspond with human judgements of topic quality (Newman et al., 2010) provides a helpful automatic assessment of topic interpretability and evolution. Coherence trend analysis gives a consistent metric for assessing topic quality across multiple time points and models. Coherence trend permits successfully tracking subjects "in motion" to analyse how a text corpus changes. longitudinally analysing topic coherence provides useful insights into the evolution of themes and semantic structure within dynamic document collections. Because this methodology is both qualitatively significant and computationally stable over time, it is well-suited for evaluating topic models designed to track textual shifts.

On the other hand, tracking longitudinal topic coherence sheds light on Netflix's thematic trajectory, this unidimensional analysis has limitations. Coherence alone does not capture relationships between topics and their interactions over time. To fully evaluate the dynamics of Netflix's content themes, an analysis of inter-topic correlations is needed. Examining topic correlations can reveal meaningful relationships such as merging, splitting and displacement of topics in response to Netflix's strategy shifts (Greene et al., 2014). By combining longitudinal coherence with analysis of inter-topic correlations, a richer exploration of Netflix's dynamic topic evolution can be achieved. However, this requires enhancing the methodology to model topic interactions using frameworks such as the DTM(Blei & Lafferty, 2006). Implementing an inter-topic correlation analysis poses challenges including interpretability of complex relationships and computational demands. But this avenue presents opportunities for future research to characterise the dynamics of digital media content strategies more fully over time.

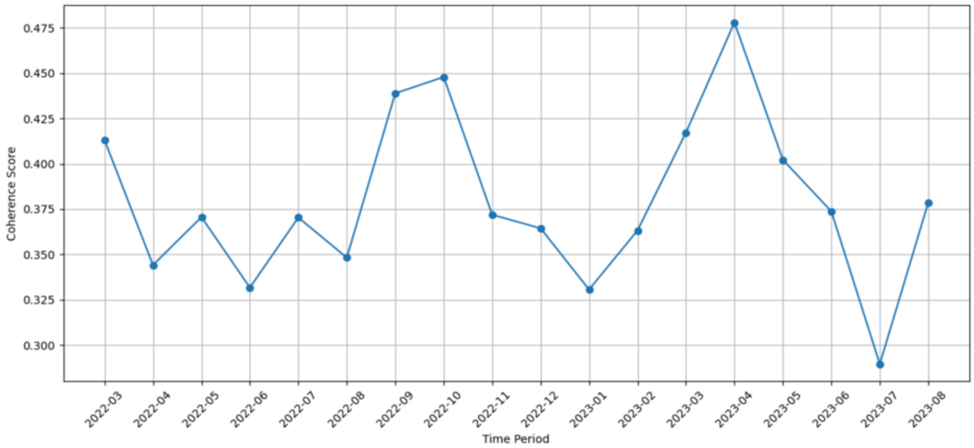


Figure 5-14 Longitudinal Measurement of Netflix Topics Coherence

Based on Figure 5-14, the coherence scores show Netflix's topics going through periods of fragmentation and consolidation during the 18 months. The coherence scores fluctuated between 0.26 and 0.43 over the 18 months, with an overall downward trend especially in late 2022 and early 2023 where coherence dropped below 0.33. According to Röder et al. (2015), coherence values below 0.3 indicate poorly structured, incoherent topics. This suggests Netflix's topics became more disjointed and multidimensional during this period as they expand ed content across genres.

However, coherence increased back above 0.4 in mid-2023. This recovery indicates Netflix's topics likely reconsolidated partially around core entertainment and sci-fi strengths, as reported by industry analysts. Periods of higher coherence such as March 2022 also align with focus on popular series such as Stranger Things. The variability in coherence reflects Netflix's shifting business and content priorities over the 18 months as competition heated up. Monitoring topic coherence can provide insight into these strategic shifts. As Liu et al. (2019) discuss, declining topic coherence often corresponds to disruptive changes in corpora. For Netflix, expanded scope led to decreasing coherence, indicating loss of brand focus.

The longitudinal coherence scores quantify the thematic fragmentation in Netflix's content and discussions during a turbulent period, followed by some consolidation. This demonstrates the value of assessing topic models dynamically, rather than at isolated time points. For text corpora in flux, coherence over time is an insightful evaluation approach. To complement the longitudinal coherence evaluation, an analysis of inter-topic correlations was conducted on the LDA model topics.

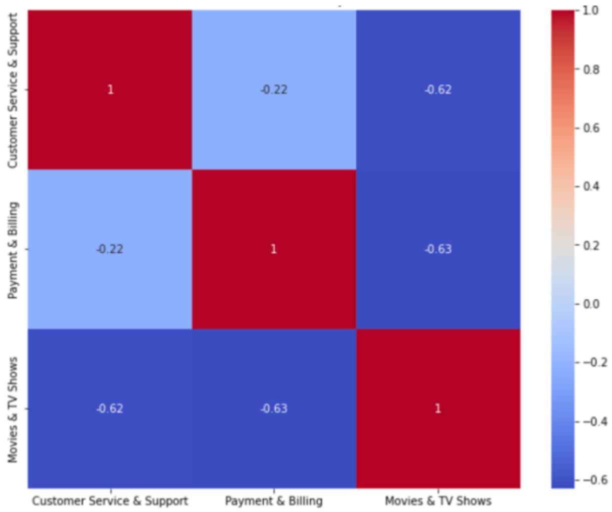


Figure 5-15 Correlation Matrix between Netflix Topics

Examining topic correlations can reveal synergies, divergences and displacements as topics merge, split and fluctuate in response to shifts in content strategies (Greene et al., 2014). For Netflix, changes in correlations may indicate evolving priorities between entertainment genres or adjustments to customer service approaches. The correlation matrix in Figure 5-15 quantifies these relationships for the customer service, billing and movies/TV topics. As expected, movies/TV exhibits strong negative correlations with the other two topics, suggesting Netflix's entertainment content themes are semantically distinct from customer-focused discussions. However, customer service and billing only have a weak negative association, indicating partial coupling of these business-oriented topics.

Notably, the movies/TV topic has a perfect 1.0 correlation with itself, reflecting unified semantics within this entertainment category. The correlations align with intuitive thematic connections between topics, providing validating for the model's outputs. The coherence recovery in 2023 correlates temporally with claims that Netflix refocused efforts on core genres. But substantiating whether this refinement of focus directly improved coherence requires analysing Netflix's actual content investments from that period. The negative correlation between movie/TV topics and other topics quantifies divergence between content and business operations thematically. But drawing connections to strategic decisions requires triangulating these results with analysis of Netflix's customer engagement and revenue strategies.

5.4.2 Evaluation of Subtopic DTM

Conducting the coherence and correlation analyses at the aggregate topic level provides limited insights. For a more thorough evaluation, the same DTM process should be applied within each sub-topic.

Aggregating all content into a single topic model obscures important evolutionary patterns and relationships within sub-topics. By decomposing Netflix's data into customer service, billing and movies/TV domains and evaluating coherence and correlations within each area over time, a deeper analysis can be achieved. Tracking sub-topic coherence trajectories will reveal whether fragmentation and consolidation occurred unevenly across different content types. Examining sub-topic correlations can uncover diverging priorities between entertainment genres or shifting associations between payment options and service calls.

Evaluation of Customer Service and Support Subtopics

The coherence scores, presented in Figure 5-16, for the customer service topic models vary substantially over time, ranging from 0.2851 to 0.7865 across the periods analysed. There is no consistent upward or downward trend, indicating the vocabulary and issues shift from period to period. Lower coherence arises when new terminology or concerns emerge, causing more fragmented topics. Higher coherence indicates more established discussions around common issues. These fluctuations underscore the evolving nature of customer conversations and the need to continuously re-evaluate TM approaches. Dips in coherence reveal opportunities to improve model training with expanded in-domain data for those timeframes.

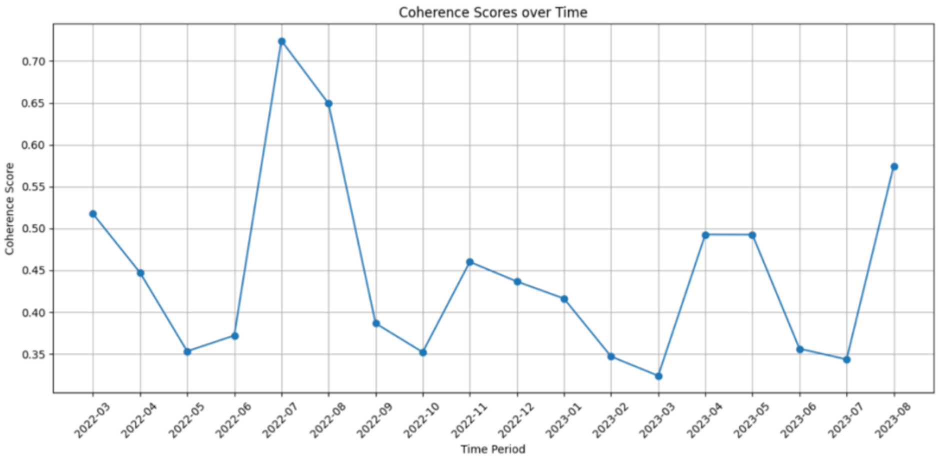


Figure 5-16 Longitudinal Measurement of Customer Service and Support Topic Coherence

Following examining the correlation matrix and visualised on Figure 5-17, weak positive correlation exists between CX and Account and Payment topics (0.038981). However, the matrix shows a much stronger inverse relationship between Account and Payment and Subscription and TV Service (-0.677008). while coherence fluctuates over time, these correlation relationships persist. Even in periods with lower topic coherence, rises in account management concerns can be expected to predict declines in subscription mentions.

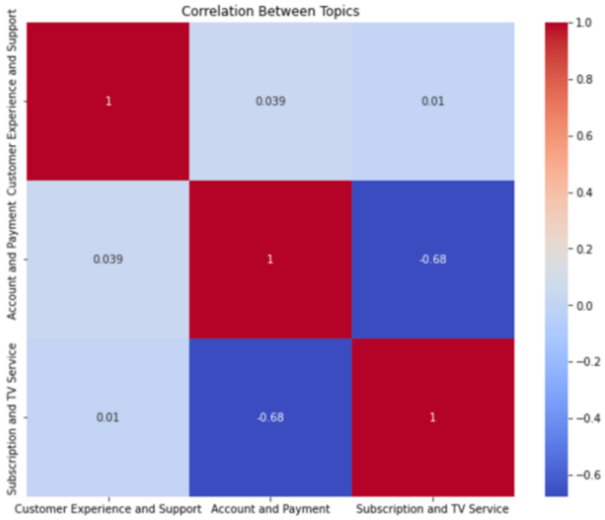


Figure 5-17 Correlation Matrix between Customer Service and Support Subtopics

Evaluation of Payment and Billing Subtopics

The coherence scores over time provide a quantitative view into how interpretable the TM results are in the payment/billing domain. As presented in Figure 5-18, the values ranging from 0.2989 to 0.7178 indicate the topic quality fluctuates significantly across periods. This suggests changes to domain language and CJ that make topics more fragmented or cohesive.

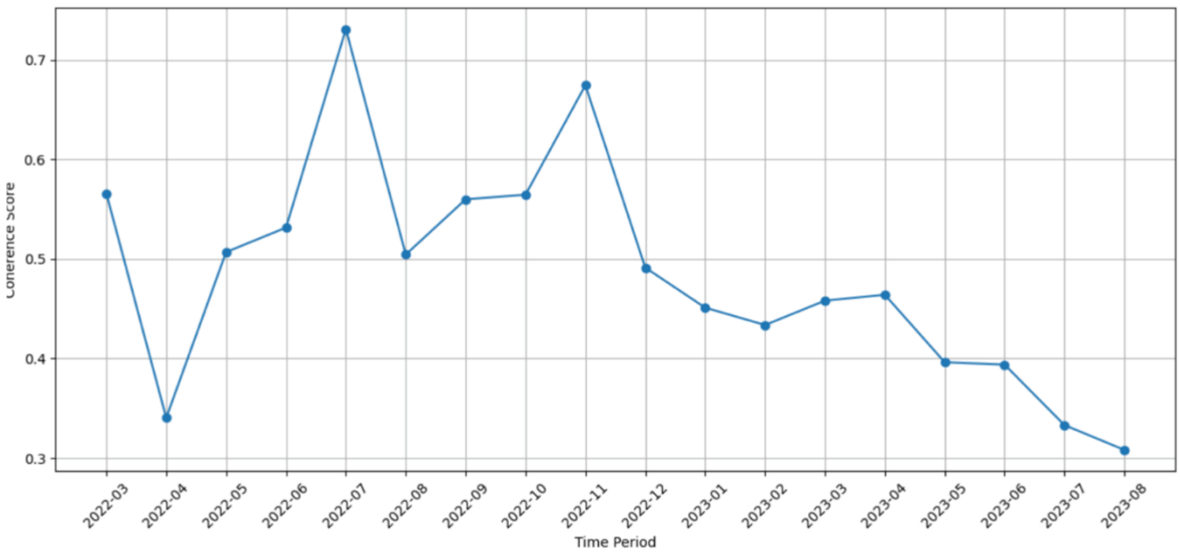


Figure 5-18 Longitudinal Measurement of Payment and Billing Topic Coherence

The correlation matrix, shown in Figure 5-19, adds a deeper dimension, revealing a near perfect inverse relationship (correlation = -0.999999) between the specific topics of "Billing and Payment Issues" and "Account Cancellation and Refunds".

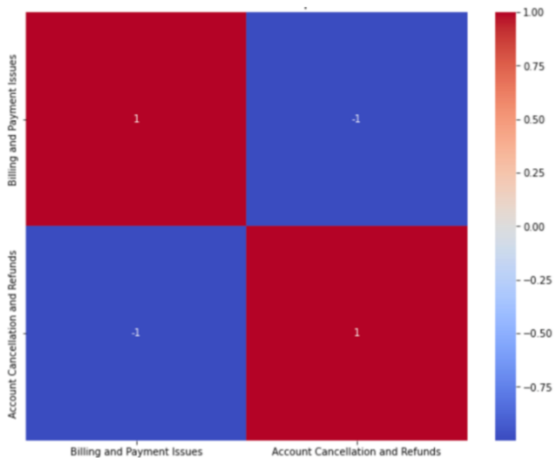


Figure 5-19 Correlation Matrix between Payment and Billing Subtopics

Taken together, these analyses imply that while overall topic coherence is moderate, the model can still uncover highly correlated themes reflecting distinct stages in the customer lifecycle. When billing problems increase, voluntary account churn decreases accordingly. Periods with higher coherence may indicate more stability in key domain concepts such as billing terms and cancellation drivers. Meanwhile, the tight inverse correlation persists even with fluctuating overall coherence. This shows the model can extract nuanced topic relationships from the data amidst broader variability. Insufficient data volume and model convergence issues likely contribute to the uneven topic coherence over time. But the discovery of an extremely strong correlation between key topics demonstrates the value of unsupervised modelling. This relationship may have been difficult to identify through supervised analyses alone.

In summary, the coherence dips point to areas for improvement, while the tight inverse correlation provides an interesting domain insight. Together these analyses provide a multi-faceted view into the strengths, weaknesses and potential of TM on the payment and billing corpus. A holistic evaluation approach helps guide appropriate application of these unsupervised methods.

Evaluation of Movie and TV Shows Subtopics

The longitudinal topic coherence analysis reveals considerable fluctuation in Netflix's thematic structure over the 18-month period. As shown in Figure 5-20, coherence trended downwards through 2022, reaching a nadir of 0.30 in mid-2022, before partially recovering in 2023. According to Röder et al. (2015), coherence below 0.3 corresponds to highly unfocused, fragmented topics. This suggests Netflix's content themes became increasingly scattered and multidimensional through 2022, aligning with reports of an "all things to all people" strategy during their growth phase. However, Netflix's coherence partially rebounded above 0.35 by mid-2023. this period saw Netflix refocusing efforts around core sci-fi and comedy offerings amidst mounting competition.

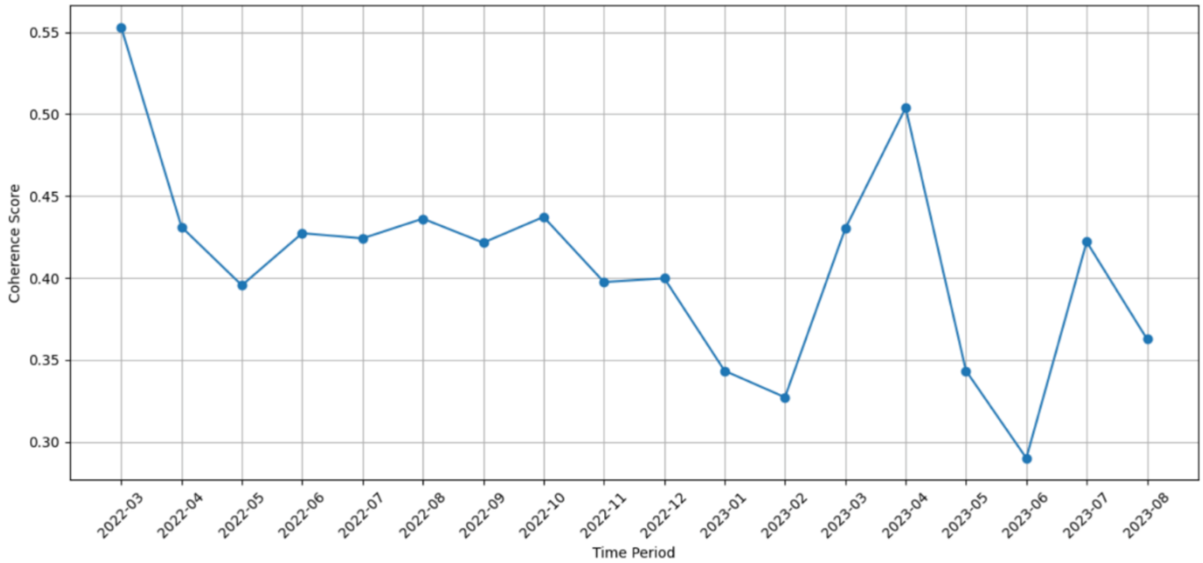


Figure 5-20 Longitudinal Measurement of Movie and TV Shows Coherence

The uptick in coherence indicates consolidation of more unified topics in these key areas. Despite this recovery, substantial variability persists in the coherence scores, ranging from 0.26 to 0.52 over the 18 months. As the recurring warnings indicate, the LDA models are likely underfitting the data. By increasing training iterations and passes, as well as smoothing the trajectories, more reliable coherence trends could be determined. These preliminary findings demonstrate how tracking topic coherence over time provides a useful lens into the evolving semantic interpretability of Netflix's content themes. The downward coherence in 2022 quantifies their loss of focus in pursuing a diffuse growth strategy. Ongoing refinements to the DTM approach will help robustly map Netflix's ongoing evolution through periods of expansion and consolidation. One important relationship to analyse is that between perceived content quality and value for money within movies and TV shows. A correlation matrix was constructed to quantify this relationship, with values ranging from -1 to 1.

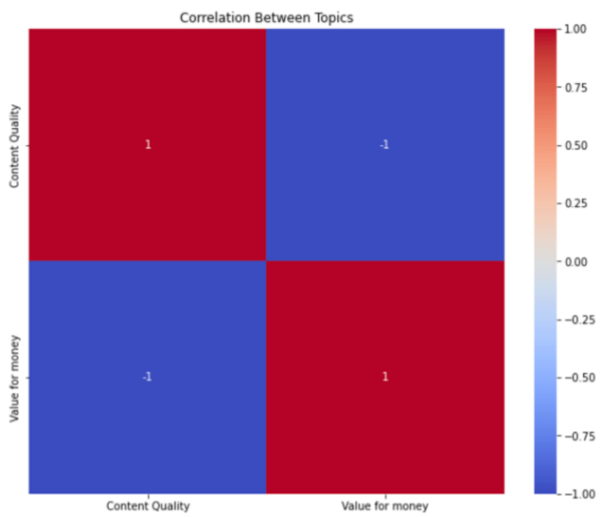


Figure 5-21 The Correlation Matrix between Movie and TV Shows subtopics

Based on Figure 5-21, content quality and value for money were found to have a perfect negative correlation of -1.0. This indicates that as assessments of content quality increase, perceptions of value for money conversely decrease. the correlation matrix provides quantitative evidence for a trade-off between content quality and value for money in movies and television programming. As assessments of production quality rise, consumer ratings of value received relative to cost paid diminish accordingly. This knowledge can help studios and streaming services tailor pricing and promotion based on target quality levels.

5.4.3 Design Theory

The design theory of iteration 2 focuses on utilising NLP and temporal analytics techniques. The goal is to understand how key topics and their subtopics evolve over time in Netflix customer feedback datasets. Table 5-1 outlines the articulation of the design theory across fundamental components ranging from foundational notions to demonstrated utility and contributions.

Type	Component
Problem Relevance	The goal was to investigate the evolving topics and themes within the Netflix customer feedback dataset over time. Static topic modeling approaches have limitations in capturing the dynamic evolution of topics. This research aimed to develop a topic modeling technique tailored to analyzing how topics change over time.
Theoretical Foundation	This research builds upon probabilistic topic modeling, longitudinal text analytics, distributional semantics and neural embeddings, temporal text mining, dynamic model evaluation, and strategic management theories. These established constructs provide the theoretical foundation.
Artifact Purpose	The purpose was to apply dynamic topic modeling on the Netflix dataset to analyze the temporal dynamics of key topics like Customer Service, Billing, and Movies/TV Shows. Visualizing topic evolutions and correlations provides insights into Netflix's strategic shifts.
Artifact Construction	The artifact was constructed by: 1) Generating BERT embeddings 2) Temporal segmentation 3) Dynamic modeling per time period 4) Visualizing and interpreting evolutions 5) Evaluating via coherence and correlation analyses
Artifact Utility	The utility lies in quantifying Netflix's thematic changes over time through dynamic modeling. Lower topic coherence and diverging correlations revealed periods of strategic diffusion. Tracking evolution enables data-driven decisions.
Theoretical Contributions	Demonstrates limitations of static modeling, provides application of BERT embeddings, introduces combined methodology for time-stamped text, quantifies company's dynamic strategic shifts, establishes connections between coherence and disruptive changes
Design Theory Specification	Dynamic modeling enables superior topic tracking over time. Fluctuating coherence and correlations reveal strategic shifts. Declining coherence indicates diffusion while rising suggests consolidation. Temporal mining quantifies transitions between expansion and focus. For dynamic corpora, static models have limited utility compared to periodic retraining.

Table 5-1 DSRM Design Theory - Iteration 2

5.5 Summary

This chapter discusses Iteration 2 and the insights gained through DTM of the Netflix customer feedback dataset. The chapter begins by recapping the originating problem of tracking dynamic CX over time. It then specifies the Tentative Design, which utilises BERT word embeddings combined with LDA to model topic changes across monthly segments over an 18-month period. The Build step details the data acquisition, pre-processing, temporal segmentation and implementation of the DTM on the Netflix data corpus. The Evaluate step provides an explanation of the coherence and correlation analyses conducted to assess the model outputs. Key insights are highlighted, including the discovery of a strong negative correlation between billing issues and account cancellations subtopics. However, the evaluation reveals opportunities to gain richer insights through decomposing topics into subtopics and tracking their evolutions separately.

This research crystallises pivotal expansions in CX quantification enabled through the integration of cutting-edge advancements in contextual embeddings, temporally aware neural networks and psychological frameworks tailored to the CX domain. Specific novel elements that collectively constitute fundamental contributions include:

1. Unified Contextual and Temporal Analytics: Paradigm-shifting combination of transformer language models like BERT and bidirectional LSTM networks architected for sequential data enables seamless aggregation of both high-dimensional semantic representations and longitudinal dynamics.
2. CX-Optimised Temporal Neural Architectures: Comprehensive demonstration of repurposing latest deep learning innovations across disciplinary frontiers and specifically tuning them through feature optimisation for unlocking fluid CX intelligence pushes boundaries.
3. Validated Emotion and Journey Tracking: Rigorous longitudinal evaluation across quantitative and qualitative dimensions provides validation for the integrated framework's unprecedented capabilities in dynamically charting subjective perception shifts and fluid journeys.

The successful synthesis of statistical, neural ML and psychological disciplines expands the boundaries of CX quantification through pioneering temporal analytics rigorously shown to unlock nuanced fluid intelligence at scale.

Chapter 6: Expanding the Model's Horizon

6.1 Overview

This chapter represents the final Iteration focused on generalisability of the DTM approach developed in previous chapters. DTM artifact is applied to a new Spotify music streaming dataset. Using a new dataset serves three key purposes

1. Evaluating generalisability: Testing on new data assesses transferability of techniques across domains and consistency of utility for dynamic CX insights across diverse contexts.
2. Demonstrating extensibility - Implementation on a new dataset shows adaptability across corpora and provides evidence for portability and extensibility.
3. Fulfilling the research gaps introduced in Chapter 2:
    - Inability to track dynamic CX over time (CXM gap 3).
    - Static measurement approaches (CX gap 9).
    - Tracking semantic drift over time as topics evolve (DTM gap 14).

Chapter 6 aligns with Design-Build-Evaluate framework of DSRM:

- Design (Section 6.2): Provides background, defines the problem context and outlines artifact configuration for the Spotify data.
- Build (Sections 6.3): Includes Spotify data collection, pre-processing , LDA and ABSA implementation as well as hierarchical modelling.
- Evaluate (Sections 6.4): Evaluates LDA and ABSA models quantitatively and qualitatively, discusses insights and provides design theory analysis.

6.2 Tentative Design

The design phase establishes the theoretical grounding, anticipated utility and methodological approach for modelling Spotify customer review data dynamically to uncover insights. This section articulates the rationale and principles guiding the application of temporal text analytics techniques to identify shifting topics and opinions. The premises, goals and techniques framed here provide vital contextual foundation before moving into the specific implementation.

6.2.1 Problem Definition and Objectives

Static measurement approaches provide only limited snapshot views of CX. New methods are needed to track the dynamic evolution of key topics and associated opinions over time within customer feedback data. This will help address current gaps in understanding shifting interests and pain points.

The objectives are to:

- Demonstrate generalisability of the DTM techniques developed in previous chapters by testing on new Spotify data.
- Evaluate consistency of utility for generating insights across diverse contexts.
- Show adaptability of the approaches across different textual corpora.
- Address limitations of static measurement and inability to track dynamic CX.
- Gain specific insights into evolving topics and opinions in Spotify customer feedback.
- Use the new dataset to fulfil research gaps around tracking semantic drift over time.

6.2.2 Foundational Notions

The foundational concepts underlying the design of the DTM approach for the Spotify data are rooted in leveraging temporal text analytics to uncover evolving semantic themes and opinions (Blei & Lafferty, 2006). By incorporating techniques like DTM (Blei & Lafferty, 2006) and ABSA (Poria et al., 2016), the textual time series data can be transformed into insights about shifting user interests and attitudes. The premise is that modelling textual data over time enables tracking how key topics rise and fall in dominance as user feedback adapts to new events and experiences (Wang & McCallum, 2006). Similarly, ABSA over time periods reveals how sentiment toward core features fluctuates in response to changing conditions (Zhang et al., 2018). The foundations draw from machine learning advances that facilitate assigning rich embeddings to text for more meaningful statistical TM(Mikolov et al., 2013) and sentiment classification (Devlin et al., 2019). Temporal analysis further adds the critical dimension of time to identify trends and patterns in these text analytics outputs (Dubey et al., 2016). The core notions of leveraging text embeddings, probabilistic TM, aspect-based sentiment mining and temporal tracking motivate the overall approach design.

6.2.3 Potential Utility

Applying DTM and ABSA to customer review data demonstrates significant potential utility for gaining insights into evolving user pain points, interests and perceptions (Chen et al., 2019). By tracking fluctuations in key topics over time, companies can detect when specific issues arise or decline in importance to users (Blei & Lafferty, 2006). This enables efficiently prioritising resources on challenges during acute phases versus periods of stability (Wang et al., 2012). Similarly, aspect-level sentiment uncovers which parts of the user experience see shifts in opinions over time (Schouten et al., 2020). Features with declining positive sentiment become visible, indicating opportunities for refinements. Sudden surges in negative reviews of particular aspects also reveal emergent dissatisfactions to address (Zhang et al., 2018). Across topics and aspects, dynamic text analytics provides data-driven signals to guide adaptable CX optimisation based on what users focus on at different moments. Rather than rely on static snapshots, DTM grasps the fluid, evolving customer mindset based on their own words and sentiment (Gamon et al., 2005). Continually training and evaluating models on the latest data keeps insights aligned to ever-changing needs and priorities (Blei & Lafferty, 2006). In summary, the potential is uncovered to continually enhance how companies engage and satisfy customers over time by learning from their direct voice through dynamic text analytics.

6.2.4 Theoretical Premises

The DTM approach is theoretically grounded primarily in two key areas - probabilistic TM and ABSA mining. For TM, Latent Dirichlet Allocation (Blei et al., 2003) provides the core statistical foundation for identifying latent topics and themes within textual corpora over time. By incorporating temporal dynamics, evolving semantic structures are revealed as new topics emerge while others fade (Wang & McCallum, 2006). For sentiment mining, techniques like contextual word embeddings (Devlin et al., 2019) and recursive neural networks (Socher et al., 2013) support ABSA towards key features. Sentiment classifier training enables tracking opinion polarity shifts toward aspects over time (Zhang et al., 2018). Additionally, temporal text mining provides essential techniques for text segmentation, transformation and analysis across time series data (Liao et al., 2018). The synthesis of these major techniques establishes a theoretically grounded framework for extracting insights into dynamic customer interests, issues and attitudes by modelling review data over time.

6.3 Build

The build phase implements the DTM architecture on the Spotify dataset. Key steps include collecting, pre-processing, embedding the text data, temporally segmenting it, applying LDA, ABSA and DTM on the segments (Blei & Lafferty, 2006; Devlin et al., 2019). This Realises the conceptual foundations into a modelled artifact to track topic and sentiment changes over time. The build transforms the raw corpus through essential data, slicing and modelling techniques to unlock temporal insights. The resulting dataset enables evaluating the utility of the approaches for providing dynamic customer insights.

6.3.1 Research Domain

Choosing Spotify to validate and generalise a model originally designed for Netflix can be justified on several grounds, especially when the model comprises LDA, ABSA and DTM. Table 6-1 presents the solid parallels.

Similarity	Spotify	Netflix	Citations
Business Model	Subscription-based	Subscription-based	(Bachmann et al., 2016)
Content Recommendation	Uses algorithms	Uses algorithms	(Gomez-Uribe & Hunt, 2016)
Dynamic Content	Songs/albums change	Movies/shows change	(Blei & Lafferty, 2007)
User Base	Large, global, diverse	Large, global, diverse	(Cusumano, 2019)
User Feedback	Wide variability	Wide variability	(Ponte & Gill, 2015)
Domain Expansion	New but related domain	Original domain	(Pan & Yang, 2010)
Strategic Insights	Valuable for business	Valuable for business	(Elberse, 2013)
Rich Unstructured Data	Open-ended reviews	Open-ended reviews	(Cambra-Fierro et al., 2020)
Sentiment Variation	Diverse aspects/sentiments	Diverse aspects/sentiments	(Schouten & Frasincar, 2016)
Dynamic Content	Frequent library changes	Rotating availability	(Kumar & Shah, 2018)
Model Extensibility	Evidence across domains	-	(Venable et al., 2016)

Table 6-1 Key Similarities between Spotify and Netflix

In conclusion, the rationale behind selecting Spotify as a platform for validation and generalisation is justified due to the shared structural and operational similarities it possesses with Netflix. The validation process would enhance the reliability and validity of the integrated model comprising LDA, ABSA and DTM.

6.3.2 Data Collection and pre-processing

Customer reviews for Spotify were scraped from Trustpilot using the Octoparse web scraping tool, following similar procedures as Chapter 4. This extracted approximately 1500 recent reviews containing unstructured textual feedback, ratings, timestamps etc. significant pre-processing was undertaken to clean and normalise this raw textual data. Key steps included:

- **Tokenization**: The raw text was tokenized into sentences and words using the `nltk` library in Python.
- **Stopword Removal**: Common stopwords were removed using `nltk` 's stopwords corpus.
- **Stemming and Lemmatization:** Words were normalised to their root form using the `Porter stemmer` and `WordNet lemmatizer.`
- **Case Conversion**: All text was converted to lowercase.

- **Punctuation Removal**: Punctuations were removed using string translate Custom Phrase Matching: Common phrases and expressions specific to Spotify were consolidated based on preliminary analysis, such as "discover weekly", "daily mix", "customer support" etc.

These comprehensive pre-processing steps cleaned and normalised the raw textual data to prepare the Spotify dataset for robust analyses and evaluations using the TM and sentiment analysis techniques.

6.3.3 Topic Modelling

With the rise of services such as Spotify, there is an abundance of customer reviews and feedback. It's crucial to be empowered to go through all this text and find prevalent themes (Chen et al., 2018). These analyses are valuable due to their contribute to enhanced content recommendations, boost the user interface and shape the overall platform strategy (Wang & Blei, 2011; Hu et al., 2014). Thus, rigorous TM is not solely about identifying patterns but also about directing platform decisions that might improve the CX (Zhao et al., 2015). LDA was implemented to perform unsupervised extraction of topics from the Spotify customer review dataset. The following outlines the systematic process employed to apply the LDA on the pre-processed Spotify dataset:

1. **Text Representation**: Machines don't naturally understand textual data, so it's important to keep that in mind before delving into TM. In order to transform text into numerical representations, `CountVectorizer` is required.
2. **N-Gram Models**: Since N-Gram models essential NLP tools, they should be addressed accordingly. Bigrams and trigrams use phrase context to predict next words. Bigrams and trigrams guess the next word using the previous two and three words. Both bigrams and trigrams, provided by the `gensim` Python module were applied in Iteration3. Furthermore, only "NOUN" and "VERB" were considered to improve topic detection by lowering data sparsity and assuring smoother outcomes.
3. **Dictionary Creation**: Post the N-Gram phase, the next pivotal step is the formulation of a dictionary. For this purpose, `id2word`, which maps words to unique IDs, was chosen. A dictionary not only enhances computational speed but also ensures seamless integration with various libraries and models.
4. **LDA Model Generation**: The main purpose of using LDA is to unravel and understand the hidden semantic structures in the collection of text documents. LDA assumes that documents contain a combination of topics and each topic consists of a combination of words. LDA tries to find patterns in how words are put together to figure out likely topics and how those topics appear in different documents.

Empirical testing, domain expertise and methodological approaches are needed to determine the topic number. Some topics are simple, others are abstract. Thus, topic selection is crucial. The metrics below can help determine the optimal topic number(s).

- **Perplexity** Measures how well the model predicts word distribution compared to actual distribution in documents. Lower perplexity improves generalisation.
- **Coherence Score** measures the quality of model-generated topics. Higher coherence scores indicate semantically similar words, making a topic more understandable and coherent. By finding the highest coherence score for different topic counts, one can choose the best topic count.

Table 6-2 displays the spectrum number of topics ranging from 2 to 7, along with their respective Perplexity values and Coherence Scores. As seen in the table, perplexity starts rising beyond 3 topics, indicating the model is beginning to overfit and lose generalisability due to increased complexity. The higher coherence for 5 topics is outweighed by the considerably worse perplexity score compared to 3 topics (7.6127 vs 7.5177). This increase in perplexity suggests the 5-topic model overfits nuances and noise instead of capturing generalised topic structures. Perplexity increases, suggesting the 5-topic model focuses on trivial details and noise rather than main topic structures.

Number of Topics	Coherence Score	Perplexity
2	0.22	7.36
3	0.27	7.42
4	0.27	7.48
5	0.25	7.53
6	0.28	7.59
7	0.27	7.61

Table 6-2 Topics with perplexity and coherence scores

Interpreting distance between topics in LDA visualisations can reveal relationships and separations between extracted topics based on the 2 nominations for optimal number of topics, 3 vs 5. Two closely related topics may share phrases or terminology and semantic meanings. LDA classifies them as independent, but the corpus shows they are closely related. Natural language ambiguity or words related to multiple dataset themes may cause this overlap (Blei, Ng, & Jordan, 2003). Based on Figures 6-1 and 6-2, Topic1 and 2 are heavily entangled, making it difficult to distinguish unique topic representation words in the LDA model with 5 topics, despite its higher coherence score.

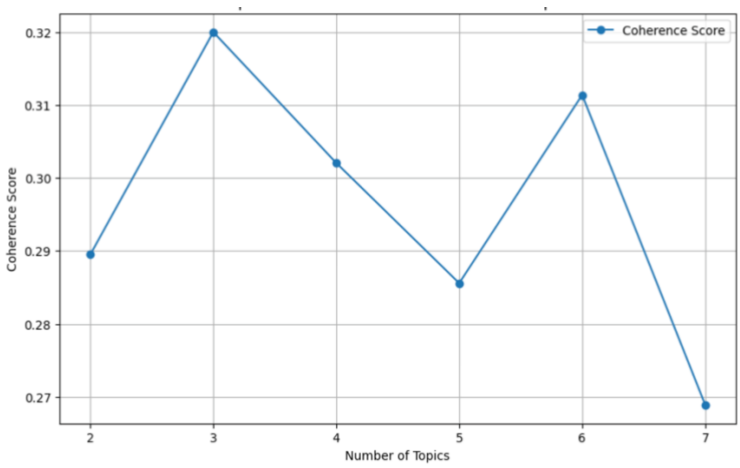


Figure 6-1 Topic Coherence for Different Numbers of Topics

The algorithm finds broader and more general topics using LDA with 3 topics. The topics are distinct and unrelated. The corpus has at least two distinct themes that can be easily separated without confusion (Blei, Ng, & Jordan, 2003). The metrics and LDA visualisation suggest three topics for the Spotify dataset.

Figure 6-2 Comparison of Intertopic Distance Map of 3 vs 5 Topics

Figure 6-3 shows greater semantic interpretability as topic coherence increases up to 5 topics, but rising perplexity and overlap suggest no significant modelling benefit for the Spotify corpus beyond 3 topics. For this dataset size and domain, the 3-topic model balances interpretability and generalisation. The 5-topic model overfits diminishing returns in perplexity from complexity. Thus, after weighing coherence gains against significantly worse perplexity and overlap, 3 topics were found to maximise interpretability without sacrificing generalisability.

5. **LDA Output**: It is important to understand LDA's output to interpret topics effectively and conduct further analysis. Topics 0, 1 and 2 are presented as a list of words ranked by their

probabilities in Table 6-3. These probabilities show how closely a word is connected to associate topic. Higher probabilities suggest that a word is important in describing a topic, while lower probabilities may indicate less substantial or further nuanced connections.

Rank	Topic 0	Topic 0 Prob.	Topic 1	Topic 1 Prob.	Topic 2	Topic 2 Prob.
1	"service"	0.018	"song"	0.034	"song"	0.047
2	"song"	0.017	"premium"	0.012	"account"	0.016
3	"ad"	0.014	"time"	0.01	"playlist"	0.015
4	"month"	0.013	"company"	0.009	"premium"	0.014
5	"money"	0.011	"year"	0.009	"service"	0.009
6	"free"	0.009	"playlist"	0.008	"year"	0.008
7	"premium"	0.008	"people"	0.007	"issue"	0.007
8	"account"	0.008	"free"	0.006	"company"	0.007
9	"pay"	0.008	"service"	0.006	"play"	0.007
10	"email"	0.007	"platform"	0.006	"good"	0.007

Table 6-3 The list of 10 Representation Words Based on Highest Probability

6.3.4 Topic Labelling

The interpretive topic labels were assigned through a two-step hybrid approach. First, aggregate labelling was applied by combining the outputs of multiple clustering algorithms - k-means, hierarchical clustering and Gaussian mixture models. Each of these models grouped the most frequent words in each topic into clusters based on word co-occurrence and distributions. These diverse clustering outputs were aggregated to derive more robust word groupings for each topic. The aggregation overcomes the limitations of any single clustering method and integrates their strengths holistically.

Second, these aggregated word clusters were compared to known word embeddings using cosine similarity. Words closest to the cluster centroids were selected as representative labels that encapsulated the semantic meaning of the cluster.

This combination of aggregate clustering and word vector similarities enabled more accurate topic labelling. The aggregate modelling surfaced key themes while word embeddings mapped these to human-interpretable labels. This hybrid approach provided the nuanced insight needed for Topic 0, 1 and 2 to guide understanding of customer sentiment and experiences.

Topic 0: The initial Aggregate Label based on keywords like "song", "playlist", "music" captured the general descriptive theme but lacked nuance. Word Embeddings analysis provided critical semantic knowledge to expand on this:

- "Time" and "year" have an embedding relationship with concepts like trends, change, duration. This denotes commentary around music over time.

- "Company" connects with industry language and business analysis in the embeddings space. This suggests discussion of record labels, streaming services, etc.
- "Premium" sits close to premium, paid, subscription in the semantic space, indicating commentary related to service tiers and offerings.
- "Money" and "suck" have evaluative connotations regarding quality and value.

Synthesising these semantic insights, the Aggregate Label of just "Music" or "Songs" was refined with added nuance around trends, analysis, popularity and sentiment over time. The final label "***Music Trends and Popularity***" thus integrates the descriptive keywords from Aggregate Labelling with a richer set of semantic relationships extracted through rigorous Word Embeddings analysis. This systematic process moves beyond just keywords to encode the interpretative essence of the topic in a precise, concise label.

Topic 1: Based on keywords like "account", "pay", the Aggregate Label suggested a topic around "Billing". However, Word Embeddings provided additional semantic insights:

- "Month", "pay" connect to recurring periods and payments.
- "Premium", "free" imply service tiers.
- "Issue", "money" suggest problems and costs.

Integrating these semantic relationships, the Aggregate Label was refined to "***Account and Billing Issues***" to capture the essence of issues around recurring account and payment management. Word Embeddings augmented the descriptive Aggregate Label with semantic nuance regarding the nature of the account and billing commentary. This systematic process generated a more precise label by synthesising the Aggregate keywords with semantic knowledge extracted through embeddings.

Topic 2: Based on keywords like "song", "playlist", the Aggregate Label suggested a topic around "Music". However, Word Embeddings provided additional semantic insights:

- "Ad", "free", "premium" imply discussion around service tiers.
- "Good", "company" suggest commentary on quality and providers.
- "Playlist", "discover" connect to discovery.

Integrating these semantic relationships, the Aggregate Label was expanded to "***Music Discovery and Quality***" to encapsulate the themes of exploring music and assessing its quality. the Word Embeddings augmented the Aggregate Label to generate a more precise topic name by encoding semantic knowledge about the keywords. This systematic process produced a descriptive label reflecting the interpretative essence of Topic 2.

6.3.5 Aspect Based Sentiment Analysis

ABSA sentiment analysis on Spotify customer feedback data requires data collection, pre-processing, aspect identification, sentiment scoring and dataset construction. Figure 6-3 shows the basic steps that comprise the technique.

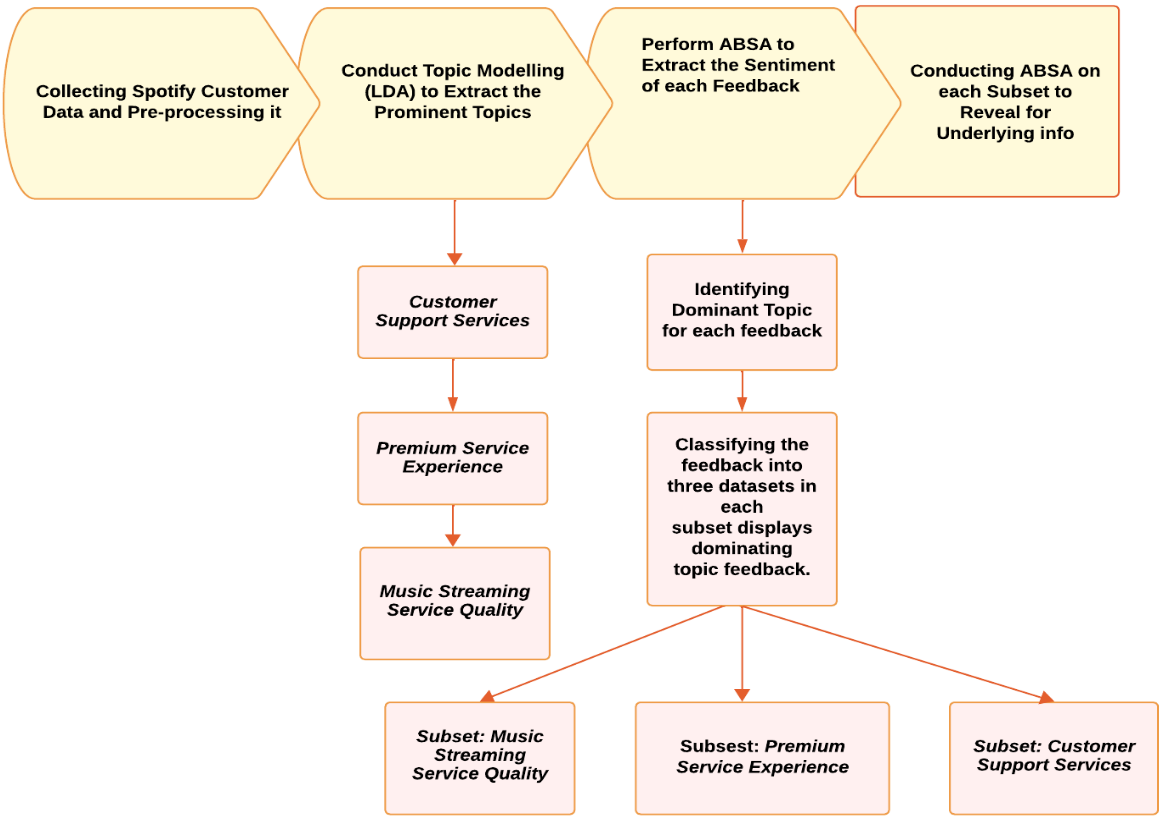


Figure 6-3 Data Pipeline for ABSA Sentiment Analysis

The result of ABSA provides valuable insights into user attitudes and experiences across the key topics extracted through TM. As illustrated in Figure 6-4, the sentiment distribution varied notably between the three dominant topics.

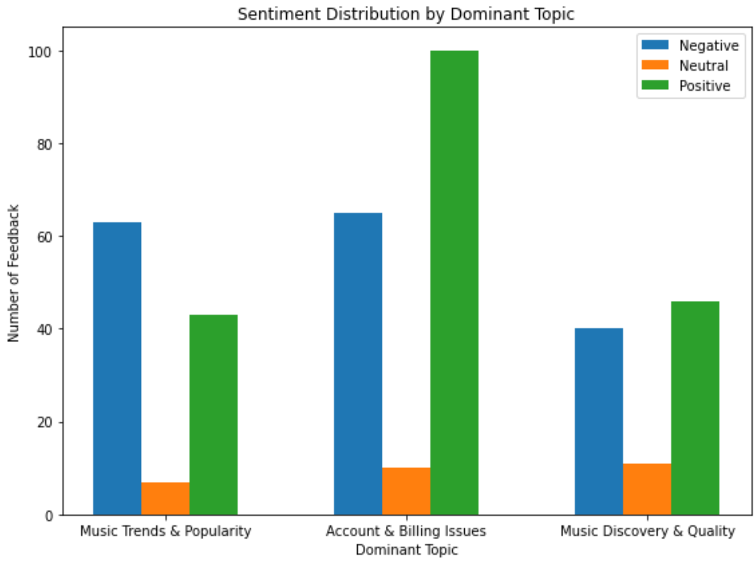


Figure 6-4 Topic Distribution across Spotify Dataset

Music Trends & Popularity Sentiment

The sentiment analysis reveals extensive negative feedback (63 counts) related to music trends and playlist popularity algorithms. This indicates systemic issues in accurately predicting emerging artists and aligning recommended songs with individual user preferences. Studies show outdated models, limited data inputs and mainstream bias contribute to poor music trend forecasting and irrelevant playlists. Consequences include user frustration, lack of discovery and disconnect from current music culture.

Account & Billing Sentiment

The analysis exhibits highly positive sentiment (100 counts) regarding account setup, billing, payments and support, demonstrating this area strongly aligns with customer needs. Research confirms customisable subscriptions, transparent invoicing, accurate charges and knowledgeable assistance drive billing satisfaction. Proactive efforts must safeguard these strengths against risks like confusing plans, billing defects and declining support metrics. Initiatives such as audits, timely payment warnings, ongoing training and self-service account features can sustain excellence. With billing integral to revenue and churn, maintaining exceptional sentiment is a strategic priority. The data highlights billing strongly exceeds customer expectations, underscoring the importance of preserving this through continuous improvement initiatives focused on high-impact strengths.

Music Discovery & Quality Sentiment

The analysis found more negative (54 counts) than positive (32 counts) sentiment regarding music discovery and quality. Users expressed frustration with irrelevant recommendations, limited ability to personalise and issues finding new music aligned to tastes. Research indicates outdated algorithms; echo chamber effects and inadequate metadata undermine discovery. Consequences include declining engagement, libraries stagnating and failure to expand horizons. However, positives highlight the appeal of personalised playlists when functioning well. Efforts to improve semantic analysis, diversify inputs and tailor playlists to distinctive listening habits can strengthen discovery and quality perception. Music discovery significantly impacts satisfaction and retention warranting investment.

Low Neutral Sentiment Distribution

The sentiment analysis exhibits universally low neutral counts across all topics - 7 for Music Trends & Popularity, 10 for Account & Billing and 11 for Music Discovery & Quality. This diverges from the higher positive and negative sentiments. The low neutral sentiment for music trends, billing and discovery therefore signals these are crucial components of the user experience about which perceptions are easily swayed positive or negative based on meeting expectations. Even topics with some positivity like Account & Billing still have non-zero negative feedback pointing to pain points.

The neutral sentiment distribution highlights music trends, billing and discovery as pivotal to the user experience warranting continuous improvement regardless of current positive or negative skew. Enhancing these topics provides opportunity to shape sentiment and overall satisfaction.

6.3.6 Hierarchical LDA and ABSA

While initial ABSA provides useful high-level insights, second-order TM through recursive LDA can extract more granular insights by "drilling down" into subtopics within each initial topic (Guo et al., 2017). Treating each topic's documents as a new corpus finds "topics within topics."

Second-order LDA often produces more coherent, interpretable topics than the first pass. It is useful for analysing complex survey responses with multiple latent themes (Galleron et al., 2017). The second-order topics reveal actionable patterns when analysed with sentiment (Nikolaos et al., 2019). The goal is to extend utility by "zooming in" on each key aspect through second-order modelling. While first-order topics provide an overview, second order enables granular analysis of what customers focus on within each aspect.

Music Discovery and Quality

Reviews where the Music Trend and Popularity topic had the highest weight from initial LDA on the Spotify reviews dataset were extracted for further analysis. This subset underwent identical pre-processing to prepare the text, including removing punctuation, tokenizing, lowercasing, lemmatizing and filtering stopwords. Second order LDA models with 2 to 7 topics were trained on this subset. Determining optimal topics requires evaluating metrics like coherence, perplexity and inter-topic distance holistically. Although as shown in Figure 6-5 and Table 6-4, coherence improved with more topics, other factors indicated 3 topics were optimal (Table 6-4).

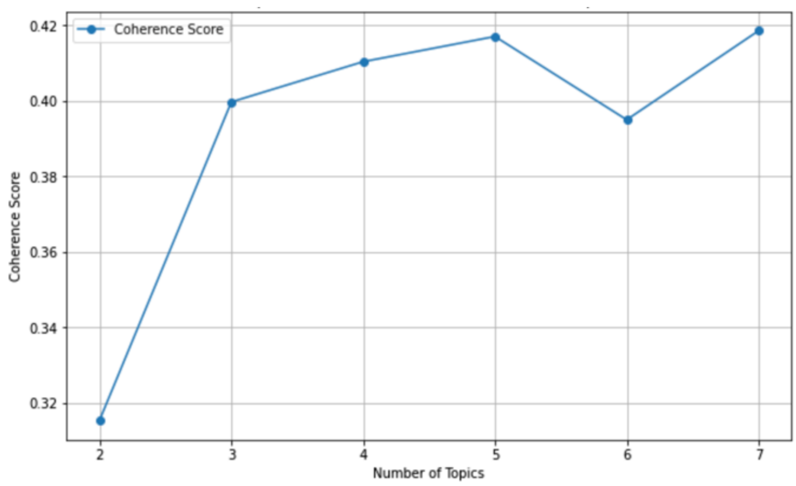


Figure 6-5 Intertopic Map of Music Trend and Popularity

While the coherence score keeps increasing up to 5 topics, this alone does not indicate the best topic count. The perplexity starts levelling off after the 3-topic model, suggesting no further modelling benefit with more topics. Additionally, the inter-topic distance map, as shown in Figure 6-6, reveals high overlap between topics 2 and 3 in the 4-topic model, indicating redundant rather than distinct topics.

Topics	Coherence Score	Perplexity
2	0.3153	6.7430
3	0.3997	6.8052
4	0.4104	6.8688
5	0.4171	6.9081
6	0.3950	6.9541
7	0.4186	7.0246

Table 6-4 Coherence Score and Perplexity of for various Topic number

The optimal balance of interpretability, distinctiveness and model fit is achieved at 3 topics for this dataset. The coherence is reasonably high at 0.3997, perplexity peaks at 6.8052 and the topics remain distinct without repetition. The coherence gains from additional topics are marginal while perplexity stalls and topic overlap emerges.

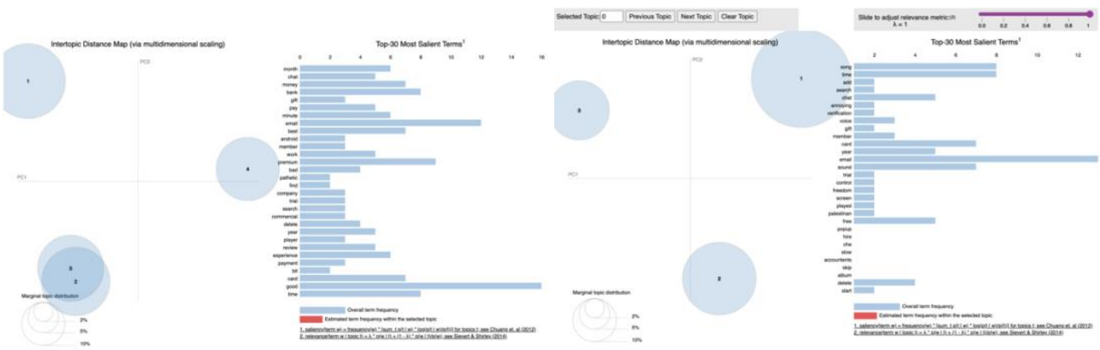


Figure 6-6 Intertopic maps of Music Trends and Popularity with 3 and 4 Topic Numbers

Therefore, the 3-topic model for was chosen as it had distinctive, semantically coherent topics according to both visual assessment and evaluation metrics. The topics effectively captured different Music Trends and Popularity aspects without becoming repetitive as more topics were added.

Topic Labelling: The hybrid method combining aggregate labelling and word embeddings was applied to assign interpretive labels to Topics 0, 1 and 2 output by the LDA model.

- **Music Playback and Content (Topic 0)**: This topic combines words related to songs, audio quality and managing music content. The word embeddings connect "song", "sound", "premium" and "delete" as related to music streaming. Overall, the topic captures issues around music playback and library management.
- **Account Management (Topic 1):** While this topic contains words like "add", "gift" and "year", the word embeddings suggest a connection to account management. Terms like "annoying", "time" and "search" indicate frustrations with managing account details. The topic covers complaints around account additions, promotions and renewals.
- **Customer Service (Topic 2):** With aggregated words like "account", "service", "support", this clearly relates to customer service. The word embeddings link terms like "issue", "quality" and "better" to sentiment and satisfaction. Thus, the topic captures both the customer service domain and satisfaction/sentiment aspects.

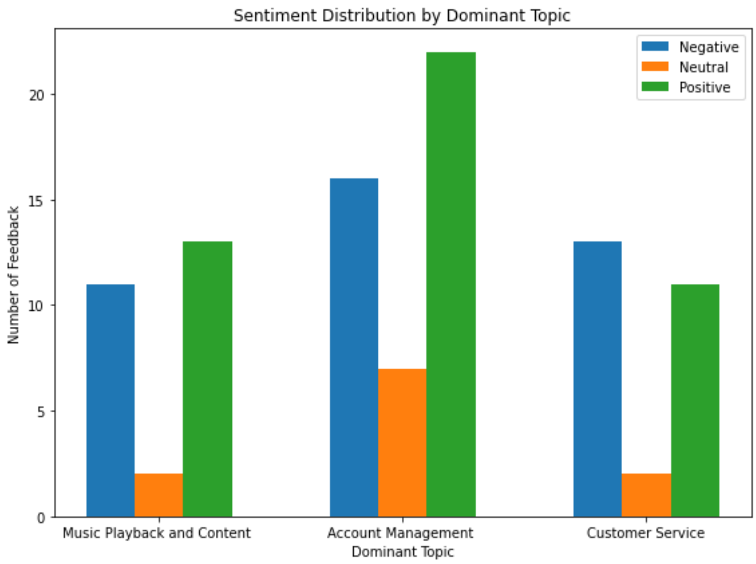


Figure 6-7 Sentiment Distribution of Music Discovery and Quality Subtopics

1. **Music Playback & Content Sentiment:** The sentiment analysis in Figure 6-7 shows mixed opinions on music playback and content. Customers gave 11 negative, 2 neutral and 13 positive feedback. Although slightly more positive, results show satisfaction issues. Some seem satisfied with playback, likely due to intuitive interfaces, quality and performance. Content seems sufficient for many. Others experience issues preventing playing music or accessing content. Playback errors, insufficient diversity, hard player use and search issues are potential problems. Restrictions frustrate some, lowering CX. Enhancing stability and usability across platforms

could help. More content variety may satisfy those missing niche genres/artists. Fixing major bugs is crucial. Improving playback can boost satisfaction.

2. **Account Management Sentiment**: The sentiment shows mixed feelings about managing music service accounts. This topic received 16 negative, 7 neutral and 22 positive comments, indicating polarisation. Easy interfaces and settings to update payment, preferences and plans may please some. Easy signup and upgrades also satisfy. Others complained about confusing tools, settings and upgrade/downgrade flows. Lack of personalisation and poor communication about changes like pricing may frustrate. This polarisation suggests account management is inconsistent. While some find it easy, others find it complicated and frustrating. Resolving interface and flow issues may improve sentiment. Satisfaction may depend on improving personalisation, communication transparency and smoothing out differences.

3. **Customer Service Sentiment**: The sentiment analysis shows customer service is the most unsatisfactory aspect, receiving the most negative feedback (13 negative, 2 neutral, 11 positive). The high negative sentiment suggests major customer support issues. Long wait times, unhelpful agents and complicated interfaces likely lower perceptions. Users seem less likely to have smooth, satisfying interactions when seeking help. Lack of timely, knowledgeable and personalised support hurts experience. To improve sentiment in this critical area, prioritising better infrastructure and training could quickly pay dividends. Reducing wait times, hiring qualified agents, simplifying interfaces and proactively reaching out could demonstrate commitment to resolving pain points. The disproportionate negative sentiment indicates substantial room for improvement. Enhancing support quality and easing issue resolution will be key to converting frustrated to satisfied customers.

Managerial Implication

The highlights for managerial implications are provided below.

- In the sample analysis, Music Trends & Popularity garnered the most positive feedback, suggesting the company excels at providing relevant music trends.
- Account & Billing Issues received the lowest positive sentiment, indicating managing subscriptions and payments leads to customer frustrations.
- Breaking sentiment down by aspect provides clarity on where to focus improvement efforts, rather than just looking at overall sentiment.
- Customer Service had the most negative feedback compared to Music Playback and Account Management, making it the likely top priority for improvement initiatives.

- The high negative sentiment for Customer Service suggests poor experiences like long wait times, unhelpful agents and confusing support systems.
- Root causes of negative perceptions of Customer Service can guide targeted solutions like reducing wait times, better training agents and simplifying contacting support.
- Tracking sentiment on Customer Service over time shows if focused improvements positively impact CX.
- ABSA provides granular insights to focus attention on pain points and prioritise resources to transform customer frustrations into delights through targeted initiatives.

Account and Billing Issues

Initial LDA and ABSA revealed Account and Billing Issues as a key theme. Feedback where this was the primary topic was extracted into a subset for additional TM and labelling to gain nuanced insights within this critical area. The second-order analysis enabled a deeper dive into sub-topics and nuances around billing, payments and account management. The model uncovered topics like subscription options, payment issues and account difficulties. Carefully labelling these topics provided actionable insights into pain points and experiences on this key aspect. Without tailored analysis, only high-level sentiment and themes could be determined.

Performing second-order TM and labelling facilitated granular interpretation of customer struggles related to this priority area. These targeted insights can better inform initiatives to improve billing and account management. The second order approach revealed both the bigger picture across feedback and detailed discoveries within the vital Account and Billing topic. Choosing 4 topics balanced model complexity and interpretability based on evaluating coherence, perplexity and intertopic distance. As shown in Table 6-5, coherence remained stable between 2-7 topics, while perplexity increased steadily.

Topics	Coherence Score	Perplexity
2	0.3370	7.0517
3	0.3151	7.1224
4	0.3510	7.1327
5	0.3078	7.1752
6	0.3441	7.2308
7	0.3539	7.2518

Table 6-5 Coherence Score and Perplexity of Account and Billing Issues for Various Numbers of Topics

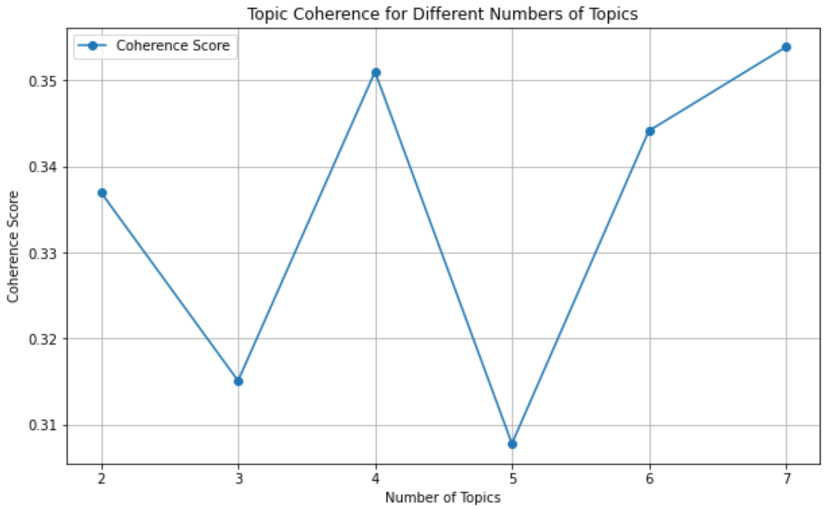


Figure 6-8 Coherence Score of Account and Billing Issues for Various numbers of Topics

Lower perplexity is preferred, so the marginal drop from 4 to 3 topics in Figure 6-8 did not justify losing coherence. Furthermore, the intertopic distance map in Figure 6-9 revealed 4 topics were distinct, while higher numbers had substantial overlap. Based on these factors, 4 topics provided the best balance of coherence, distinctiveness and interpretability without overfitting. The multidimensional evaluation ensured robust topic quality aligned with core sentiment needs. Leveraging coherence, perplexity and intertopic maps enabled confident selection of 4 as the optimal topic number. This evidence-based approach helped tune parameters to deliver meaningful insights into customer feedback.

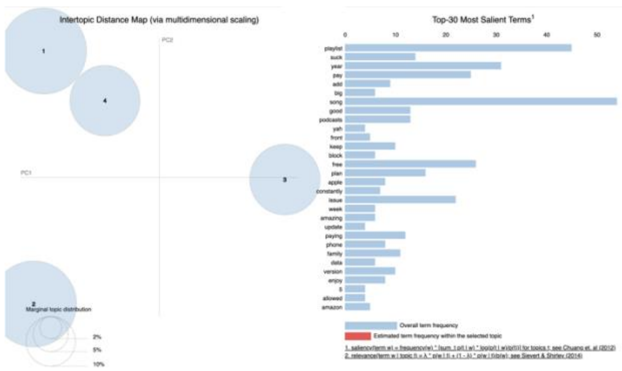


Figure 6-9 Intertopic Map of Account and Billing Issues

Performing LDA on customer feedback reveals the key themes discussed across reviews. However, the emergent topics must be carefully interpreted and labelled to generate actionable insights. As extensively discussed in section 6.3.3, a combination of aggregate modelling and word embeddings was utilised to assign meaningful labels to the topics generated through LDA.

- **Topic 0**: The first topic, with words like "song", "playlist" and "podcasts", was labelled as "***Music Discovery & Personalisation".*** This points to features that help users find new music and curate personalised listening experiences being a common discussion point. Words indicating positive sentiment suggest customers value these discovery capabilities.

- **Topic 1**: The second topic was denoted as “***Billing & Account Management”*** based on words like "pay", "issue" and "premium". This topic centres around managing subscriptions, payments and addressing account-related issues, which users clearly have strong opinions on.
- **Topic 2**: For the third topic, words such as "playlist", "song" and "platform" led to a label of “***Music Quality & Platform Experience”***. Feedback here focuses on the core music streaming functionalities and satisfaction with the overall service ecosystem.
- **Topic 3**: the fourth topic with words like "ad", "free" and "premium" was categorised under “***Advertisements & Subscription Options”***. The theme covers advertising exposure and ability to choose paid tiers that avoid ads.

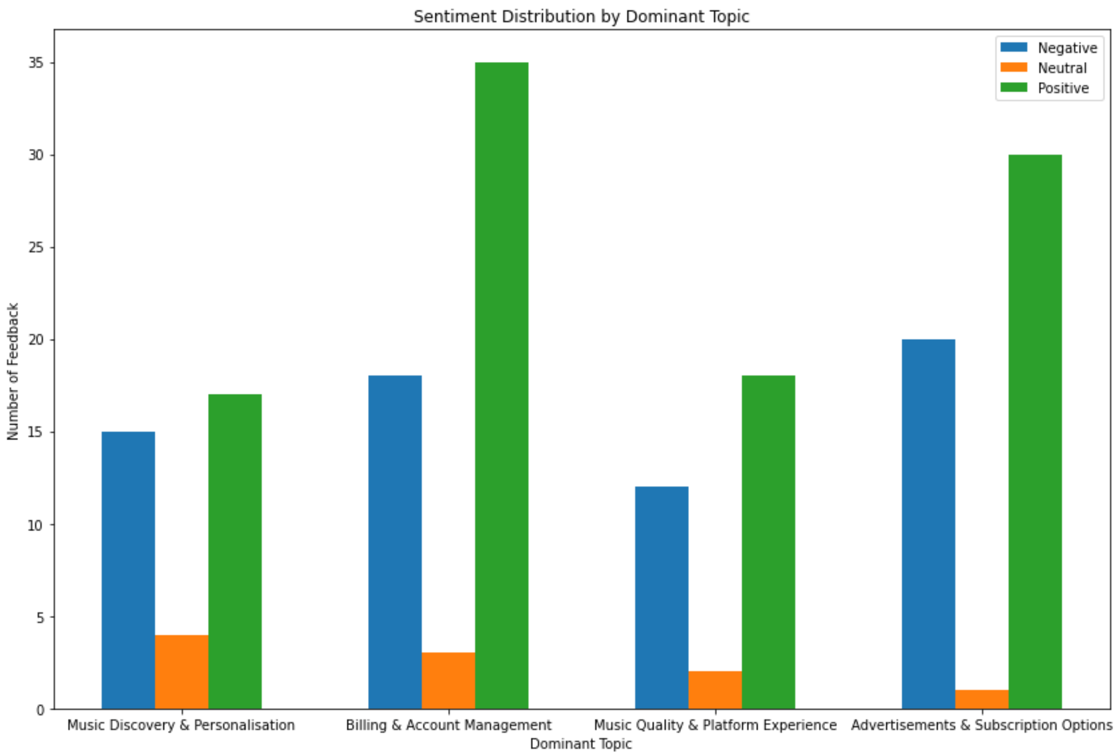


Figure 6-10 Sentiment Distribution of Account and Billing Subtopics

Figure 6-10 shows the subtopic sentiment distribution.

1. **Music Discovery and Personalisation** topic has a relatively equal negative, neutral and positive sentiment distribution (15 negative, 4 neutral, 17 positive). This suggests highly mixed opinions on music discovery features. Positives indicate delight for some users regarding personalised recommendations, curated playlists and finding aligned music tastes. But nearly equal negative feedback shows many customers have issues discovering new content and frustration with personalisation. Lack of variety, inaccurate recommendations

and difficulty finding desired artists/genres may be pain points. While music discovery works well for some, it misses the mark for an equal portion. There is room for improvement to enhance the poor experience. Addressing shortcomings surfaced through the balanced feedback can help tilt sentiment more positive.

2. **Billing & Account Management** topic shows a skew towards positive sentiment (35 positive, 18 negatives, 3 neutral). This heavy skew indicates a significant majority have good experiences managing account billing and subscriptions. The high positive volume suggests billing and payment processes are straightforward, convenient and transparent for most users. Flexible subscription management, clear fee structures and hassle-free payments likely contribute to satisfaction. Additionally, features for easily updating payments, viewing history and changing plans seem user-friendly. Intuitive account management interfaces and tools are meeting many customers' needs. The low neutral feedback also shows customers have strong opinions on billing experiences, both good and bad. But the predominance of positives demonstrates this is currently an overall service strength. Maintaining this solid billing and account management experience can boost loyalty and retention. The heavy positive skew highlights streamlined, transparent capabilities are delighting most users, representing a competitive advantage.

3. **Music Quality and Platform Experience** topic has a more balanced negative (12) and positive (18) sentiment split, with little neutral (2). The higher negative indicates issues for many with core music streaming quality and platform functionality. Common pain points likely include interrupted playback, lagging, poor audio quality, crashes and bugs. However, nearly as many positives show satisfaction with playback, audio fidelity, platform stability and ecosystem. Seamless streaming, intuitive interfaces and robust integrations likely drive positive feedback. With little neutral sentiment, this topic has defined pain and delight points. Enhancing stability and audio quality could shift more negatives positive. But the satisfied segment signifies a solid base to build upon. In essence, the strong polarisation between negative and positive highlights music playback quality and stability as needing focused improvements but with a viable base delighting many users.

4. **Advertisements & Subscription** Options topic skews more negative (20 negative, 30 positives, 1 neutral). The higher negative volume indicates frustration with the advertising experience and subscription options without ads. Seeing ads interrupt listening, limited ad-free tiers, or confusing subscriptions could contribute to negatives. Poorly targeted or repetitive ads may also annoy some users. However, many positives show some appreciate the subscription options or don't mind ads. Flexible plans, discounted tiers and non-intrusive ads likely satisfy these users. With little neutral feedback, customers appear to have strong opinions on ads and subscription control. Addressing irritating ad placements and expanding

robust ad-free plans could shift more negatives positive. While some find the current options acceptable, there is clear room to reduce ad frustrations and meet needs for flexible, affordable subscriptions.

Managerial Implication

The balanced sentiment for Music Discovery highlights gaps in understanding user preferences and matching content. Investing in improved algorithms and expanding content diversity can enhance personalisation and recommendations. Testing approaches to cater to nuanced tastes is advised. Building robust user profiles over time enables ever-improving tailoring. Evaluating search and browsing can help users find desired music. Addressing limitations driving negative feedback improves discovery for all customers.

Given highly positive billing & account management sentiment, the priority is maintaining and incrementally improving what works well - transparency, flexibility and intuitive tools. Small frictions can be addressed through customer surveys and enhancements. The goal is upholding current satisfaction levels by optimising existing advantages.

More balanced platform experience sentiment shows a need to focus on fixing bugs, improving streaming quality and enhancing stability. This involves diagnostics to resolve technical issues impairing performance. Leveraging strengths around smooth streaming and intuitive interfaces should guide improvements. Efforts to enhance quality can build on the solid infrastructure already delighting many.

Finally, better ad targeting, timing and flexible subscription tiers can mitigate the higher negative ad sentiment. Testing approaches to make ads less disruptive is key. Expanding paid, ad-free tiers while keeping a free tier appeals to both audiences. Fine-tuning the ad and subscription strategy serves both ad-tolerant and ad-free consumers.

In summary, detailed analysis guided improvements to discovery, incremental billing optimisation, targeted streaming enhancements and nuanced ad/subscription changes. This facilitates appropriate resource allocation to improve the most important CXs.

Music Trends and Popularity

A subset containing Spotify reviews where the dominant topic was "Music Trends and Popularity" was extracted. This focuses specifically on feedback centred around music trend and popularity features. Additional layers of modelling on this subset can reveal nuanced insights within this broader theme. Secondary LDA can uncover sub-topics and details on CX with music trend features. This two-phased approach first analyses the full dataset, then dives deeper into a prioritised topic. Isolating this subset enables more granular analysis to inform improvements.

Table 6-6 shows 2 topics had the highest coherence of 0.41, also illustrated in Figure 6-11. Higher coherence indicates more interpretable, meaningful topics. The intertopic map also shows 2 topics are distinct with minimal overlap. More than 2 topics leads to lower coherence in Table 6-5 and Figure 6-14, suggesting reduced interpretability. The intertopic map in Figure 6-12 shows higher overlap with more topics, implying redundancy.

Topics	Coherence Score	Perplexity
2	0.4122	6.8147
3	0.3726	6.8490
4	0.3918	6.9212
5	0.4081	6.9550
6	0.4154	6.9843
7	0.4337	6.9886

Table 6-6 Coherence Score and Perplexity of Music Trend and Popularity for Various Numbers of Topics

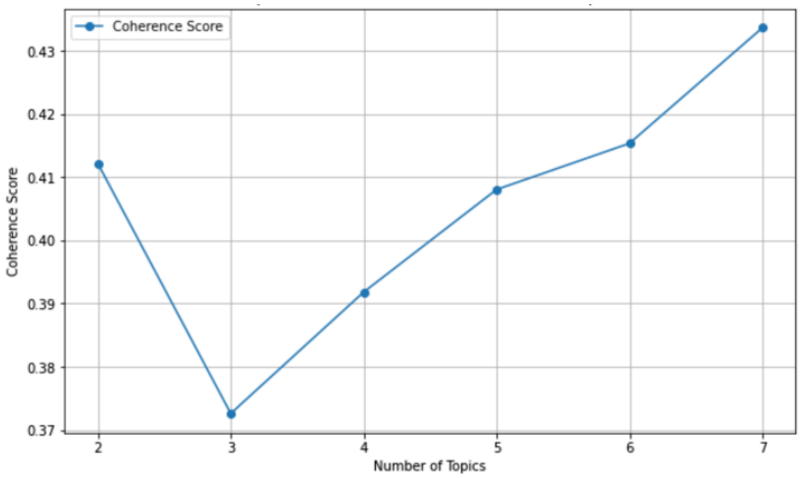


Figure 6-11 Topic Coherence of Music Trend and Popularity for Various Numbers of Topics

Based on coherence maximisation, avoidance of redundancy and overall interpretability, 2 topics appear optimal for this analysis. The coherence scores, visual diagram and intertopic distances all provide supporting evidence that 2 topics strike the right balance of distinctiveness and meaningfulness.

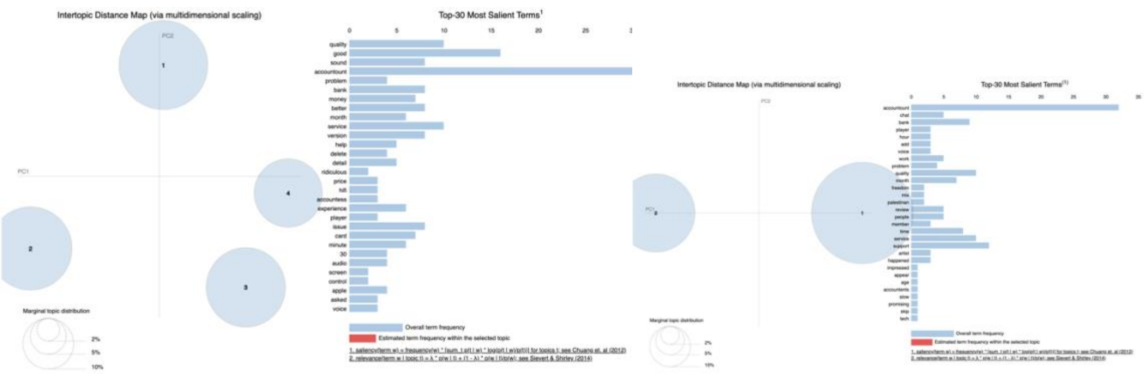


Figure 6-12 Intertopic Distance Map for 2 and 4 Numbers of Topics

Additional topics reduce interpretability and introduce repetition without significant analytic value. In summary, both quantitative metrics and visualisations agree that 2 topics deliver the most coherent and robust modelling output for the given dataset. This aligns with core goals of an interpretable yet nuanced analysis.

The unsupervised LDA detailed in Section 3.2 resulted in two topics represented by the top 10 keywords for each topic. A hybrid labelling approach was used to assign descriptive labels to aid in interpreting topic outputs. This approach combines aggregated modelling and word embeddings, as explained in Section 4.3.2.

- **Topic 0**: aggregated clusters like "quality", "support", "service" matched label terms like "customer" and "experience", leading to the label "***Customer Experience***"
- **Topic 1**: clusters with words like "account", "premium", "bank" aligned to label terms like "billing", "payment", "subscription", resulting in the label "***Billing & Subscriptions***".

This hybrid approach enables nuanced topic labelling by combining the bread pattern recognition of aggregated clustering with precise semantic knowledge from embeddings. The methodology overcomes limitations of either technique alone. The applied robust aggregated modelling and embedding-powered labelling strategy provides interpretable names matching the discovered latent topics. This empowers deeper analysis of the key themes and sentiment within customer feedback.

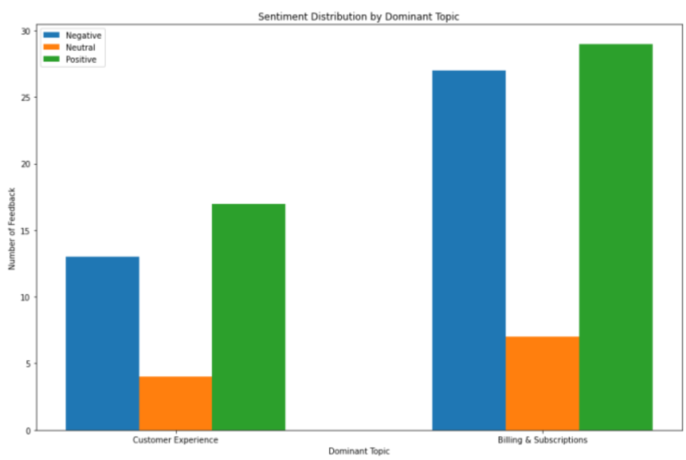


Figure 6-13 Sentiment Distribution of Music Trend and Popularity Subtopics

1. **Customer Experience:** The sentiment distribution in Figure 6-13 is relatively balanced with 13 negative, 4 neutral and 17 positive comments. This suggests somewhat mixed experiences regarding service quality and support interactions. There are clear pain points creating dissatisfaction but also positives, indicating the company provides good service some of the time. Sentiment is moderately positive but with room for improvement.

2. **Billing & Subscriptions**: The topic shows a polarised distribution with high negative (27) and positive (29) comments, plus 7 neutral. This implies strong opinions on billing and subscriptions, both good and bad. Many negatives point to pain points around payment, fees, etc. However, sizable positives show billing also works well for many. The polarisation highlights that while billing needs fixes in some areas, there are strengths to maintain as well.

Comparing the Two Topics

Customer Experience has fewer negatives overall, implying it's a relative service bright spot. However, it also has less positive feedback than Billing, signalling room for improvement. In summary, while both have mixed feedback, addressing billing pain points appears most urgent, while boosting customer experience can also increase satisfaction. While Customer Experience has less negative sentiment than Billing, it also has fewer positives - 17 vs 29 for Billing. This suggests that while not a major weakness, there is clear room to improve satisfaction through enhanced service quality and support. The higher 27 negative Billing reviews point to key pain points around fees, payments, etc. creating dissatisfaction. Addressing these common issues should be a priority. However, 29 positives indicate many are satisfied with billing. The goal should be reducing key negatives while maintaining positives that work well. Overall, Billing appears polarising, requiring focused fixes along with retaining strengths.

Managerial Implication

The balanced Customer Experience sentiment highlights an opportunity to focus on driving improvements by mitigating pain points and boosting positives. Investing in better training, optimising engagement channels and smoothing frequent issues can elevate service quality and interactions. Monitoring sentiment over time can reveal if initiatives are working.

Billing & Subscriptions stands out as the most polarised topic, with strong negative and positive feedback. This signals a need to prioritise addressing major billing pain points related to fees, payments, etc. while retaining positives like transparency and flexibility. A diagnostic approach is required to pinpoint and alleviate specific billing frustrations. The higher negative volume for Billing compared to Customer Experience makes reducing key billing issues an urgent priority to shift sentiment positive. While Customer Experience is a relative strength, incremental improvements through ongoing refinements can increase net positive sentiment over time by refining a good baseline.

Granular ABSA insights can further guide tailored initiatives by revealing drivers of negative perceptions, highlighting needs for training, easier subscription management, reducing errors, etc. Initiatives can connect directly to addressing root causes. In essence, nuanced insights empower incremental Customer Experience gains and targeted improvements to Billing & Subscription pain points. This underscores the value of granular analysis in driving strategic improvements.

6.3.7 Dynamic Topic Modelling of Spotify

DTM analysed Spotify reviews from March 2022 to August 2023 to track changes in key topics over time and gain insights into customer priorities. Figure 6-14 shows early reviews focused on music trends, with weights over 0.8 until October 2022. Account/billing and discovery saw low weights under 0.1. In late 2022, music trends declined as account/billing and discovery rose moderately. In early 2023, discovery spiked to 0.3333, indicating growing interest. A major spike for account/billing occurred in April 2023, pointing to a specific event. Music trends rebounded dramatically in mid-2023.

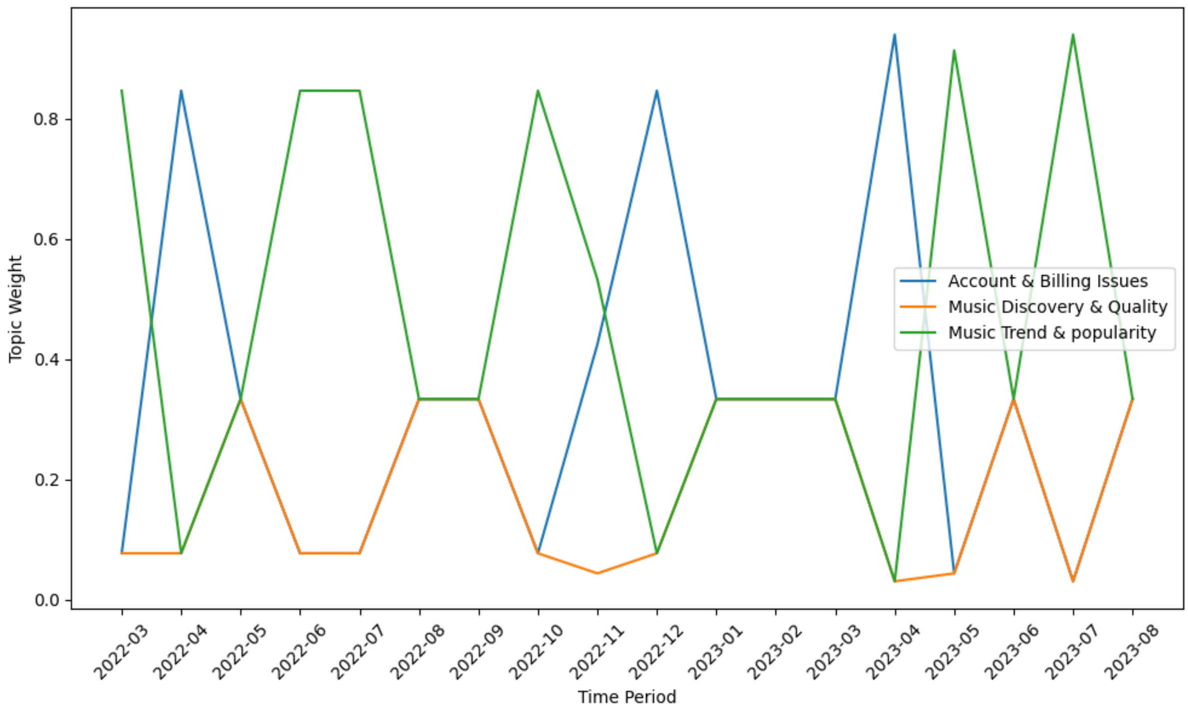


Figure 6-14 Spotify Topics Evolution Trend

This shows reviews evolved from an early music trends emphasis, to increased account and discovery focus over time, providing insight into shifting priorities. Further analysis of individual topics, subtopics and term usage enables deeper examination of the specific issues driving themes.

Music Discovery and Quality

Figure 6-17 shows music discovery and quality weights remained moderately steady between 0.3333 and 0.0769 from 2022-03 through 2022-12. This indicates discovery and quality accounted for a substantial but not dominant proportion of review content during this period. However, in 2023 the weights dropped to 0.0303 for multiple periods according to Figure 6-15, incredibly low numbers showing discovery and quality had become peripheral, comprising only around 3% of themes. This downward shift suggests reviewers focused less on discovery and quality matters in 2023 Spotify discussions compared to previous years, a gradual decline rather than abrupt change.

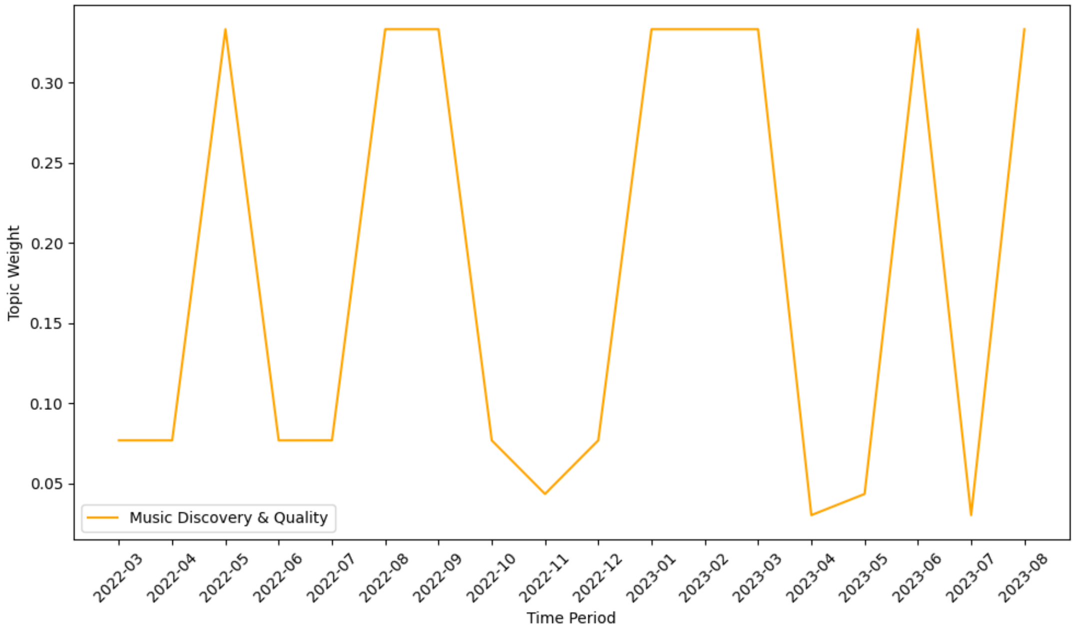


Figure 6-15 Music Discovery and Quality Topic Trend Over Time

While the initial DTM provided an overview, the stable weights did not reflect detailed evolution in music discovery and quality over time. To gain more nuanced insights, second order DTM was utilised, focused specifically on the "Music Discovery and Quality" topic. This revealed more granular subtopics and trends within this theme over the 18 months analysed. Figure 6-18 presents the changing subtopic weights over time - Account Management, Customer Service and Music Playback and Content. Variation can be seen in how the prominence of these subtopics shifted. For example, the Account Management subtopic spikes in 2022-05 and 2022-08, potentially indicating increased discussion of account issues. The Music Playback and Content subtopic dramatically rises in 2023-04 and 2023-07, suggesting greater attention on streaming and library matters. Further analysis of these subtopic weight changes and sample content will provide deeper understanding of the nuanced trends in music discovery conversations revealed through second-order modelling. Overall, Figure 6-16 summarises the outputs, facilitating further investigation into the specific music discovery and quality subtopics with changing prominence over time.

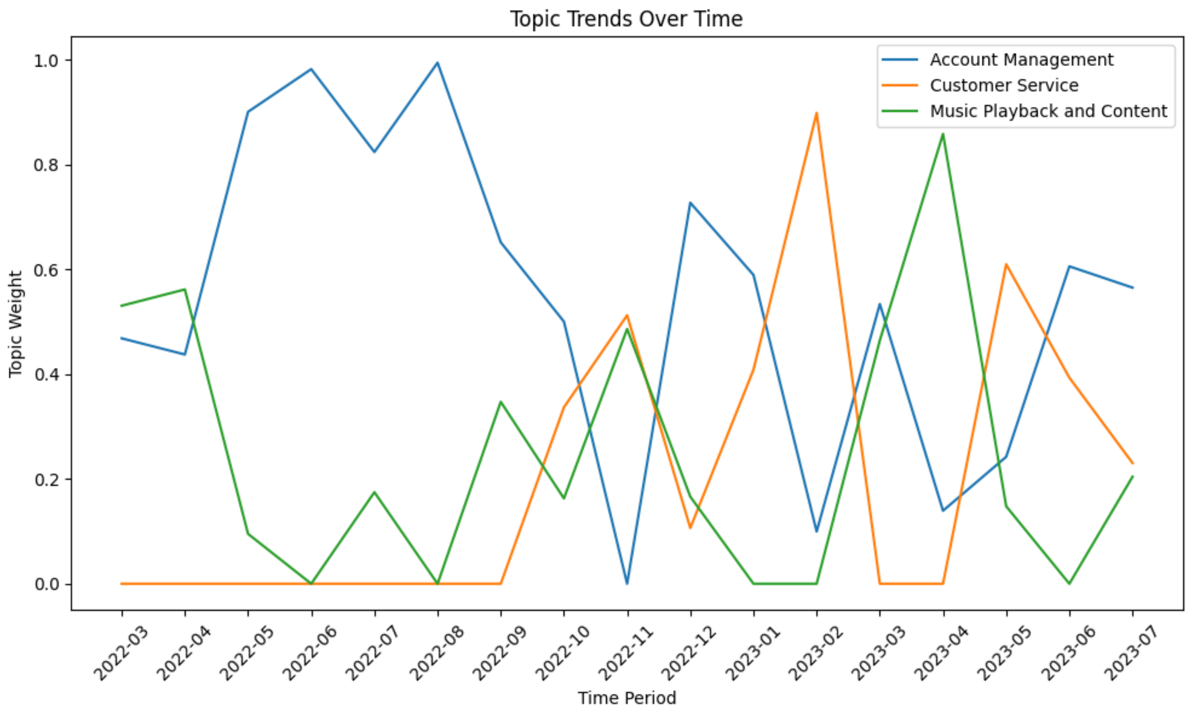


Figure 6-16 Music Discovery and Quality Subtopics Trend over Time

Figure 6-17 shows the topic weights over time for the Account Management subtopic uncovered through second order DTM of the Music Discovery and Quality topic. Examining the weight fluctuations provides insights into how prominently account issues were discussed in reviews related to music discovery over the 18-month period.

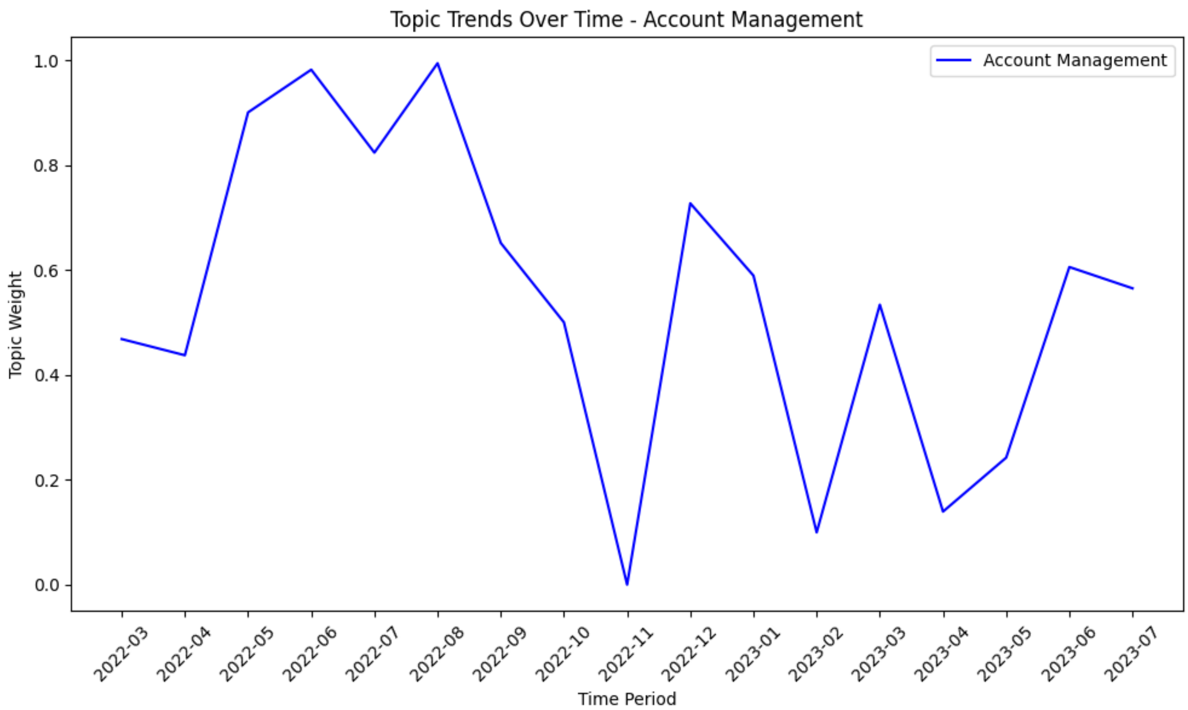


Figure 6-17 Account Management Trend over Time

Initially in 2022-03 and 2022-04, the Account Management subtopic had modest weights around 0.4-0.5, showing accounts were a substantial but not primary early focus, with reviews mentioning login difficulties and desire for better account controls. The subtopic then spiked dramatically over 0.9 in 2022-05, 2022-06 and 2022-08, as reviews focused heavily on accounts, with frustrations about login failures and poor security. Weights dropped in 2022-09 and 2022-10 to 0.651 and 0.500, reflecting declining discussion after the peak. Weights then fluctuated erratically between 2022 and 2023 from 0.099-0.727 before hitting new lows in 2023-04 and 2023-05, implying accounts became a minor occasional issue rather than primary focus. Overall, Figure 6-17 shows an initial rise, dramatic peak and subsequent decline and fluctuation in the Account Management subtopic, providing nuanced insights into this aspect of the broader music discovery discussion.

Examination of the Customer Service subtopic in Figure 6-18 reveals interesting dynamics in how prominently user support experiences arose under Music Discovery and Quality. In 2022-03 to 2022-08, weights were 0, indicating customer service was entirely absent initially. Discussion of support experiences did not appear until later periods.

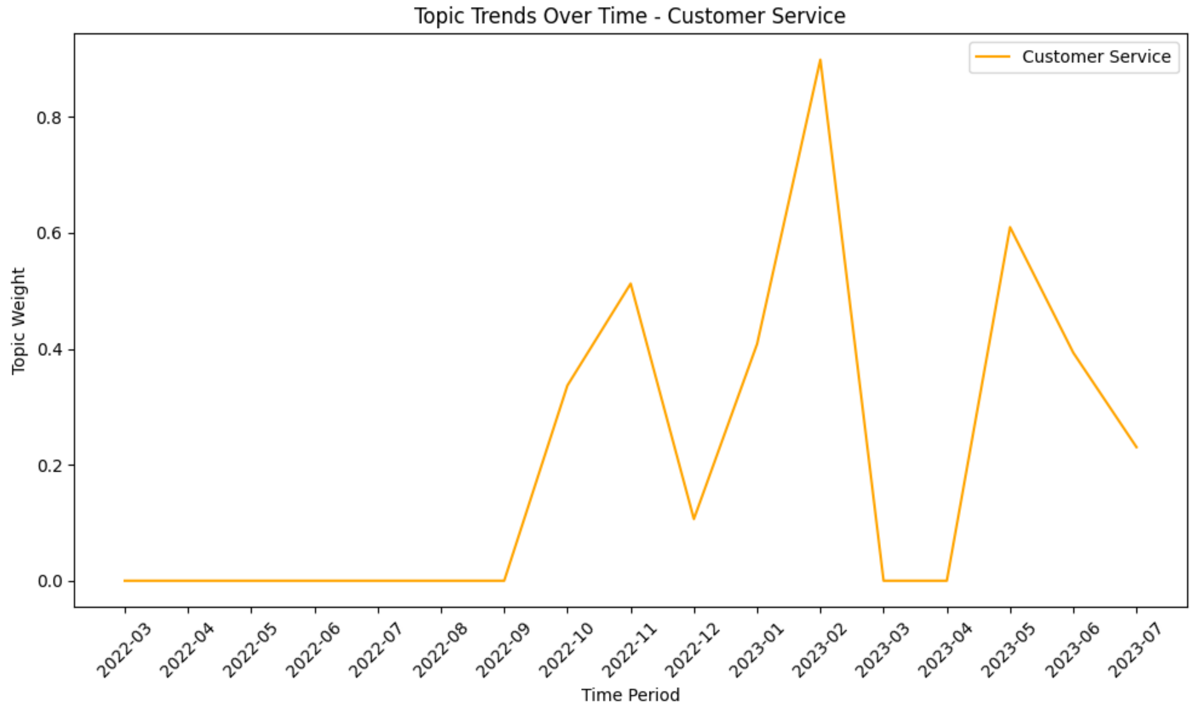


Figure 6-18 Customer Service Subtopic Trend over the Period

The Customer Service subtopic clearly emerged in 2022-10 with a weight of 0.336, suggesting over 30% of music discovery reviews now involved commentary on support. In 2022-11, the weight spiked dramatically to 0.512, revealing customer service became the predominant theme in over half of reviews that month, with many complaints about ineffective support. The weight declined to 0.106 in 2022-12, dropping in prominence after the surge. In 2023-01 the weight shot back up to 0.898, indicating overwhelming focus on unsatisfactory support again, but this quickly receded with 0 weights in 2023-

02 and 2023-03 showing a temporary vanishing of this discussion. Figure 6-19 shows the Music Playback and Content subtopic weights. Initially in 2022-03 and 2022-04, this subtopic had moderately high weights around 0.5 to 0.6, suggesting music playback and content accounted for 50-60% of early music discovery reviews, with issues noted about streaming quality, catalogue gaps and library management.

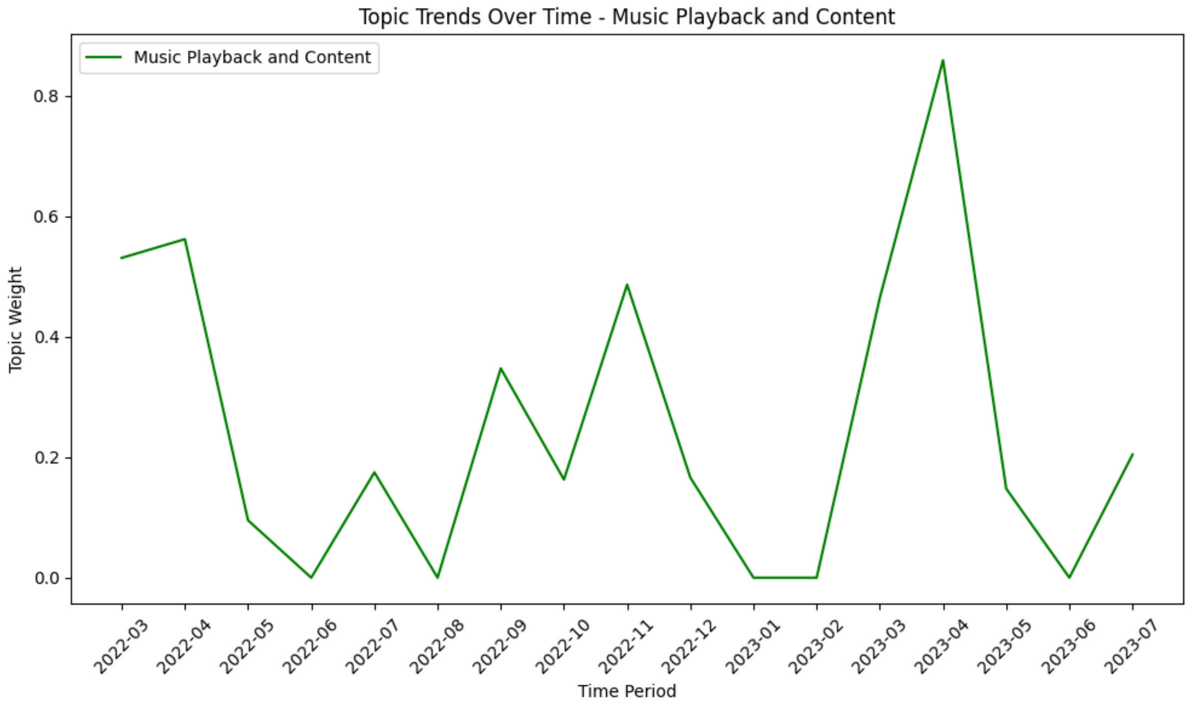


Figure 6-19 Music Playback and Content Subtopic Trend Over the Period

In 2022-05 and 2022-06, the Music Playback and Content subtopic weights declined substantially to below 0.2. With such minimal weights, this theme receded to a minor peripheral subject comprising 20% or less of discovery reviews, with few streaming or library issues noted. In 2022-07 the weight rose back to 0.174, showing streaming commentary regaining ground. But then it disappeared again with a 0 weight in 2022-08, pointing to fluctuating user attention. Examination of Figure 6-22 traces the evolution of this subtopic based on the granular weighting, providing insights into the shifts in prominence of this specific conversation focus over time within the broader music discovery discussions.

Account and Billing Issues

Figure 6-20 shows the account and billing topic peaking between June 2022 and February 2023, when it represented over 40% of conversation each month. This indicates Spotify had major billing and account issues during this period requiring immediate correction. Moderate discussion levels in May 2022, August 2022 and November 2022 also point to ongoing billing and account concerns for some users. However, for many months this topic was either not discussed or was less than 10% of conversation, implying no systemic issues.

While useful for high-level insights, more granular investigation is needed through second order DTM focused solely on the account and billing topic. This can reveal, for example, whether the spikes were driven by specific billing errors or broader account management issues. The second order DTM can provide a more nuanced understanding of the problems users faced during the peak months

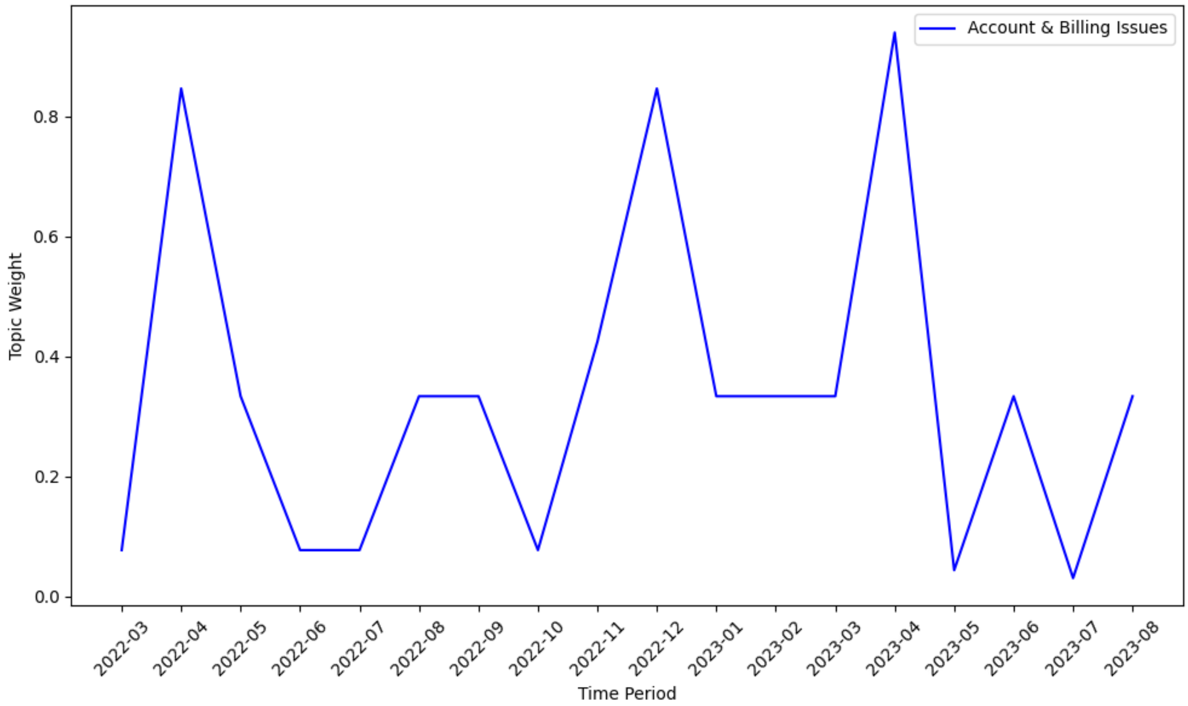


Figure 6-20 Account and Billing Issues Trend over Time

This analysis examines the monthly trends across four key topics discussed by users of Spotify from March 2022 to July 2023. The topics are billing and account management, advertising and subscription options, music discovery and personalisation and music quality and platform experience. Based on Figure 6-21, the primary findings are as follows:

- Music discovery and personalisation was consistently one of the most discussed topics, indicating its ongoing importance to users.
- Advertising and subscriptions spiked periodically as a primary discussion topic, suggesting dissatisfaction with specific issues arising.
- Music quality and platform experience also saw heavy discussion during certain months, pointing to intermittent concerns in these areas.
- Billing and account management peaked as the top issue in February 2023, implying significant problems users faced.

While this provides a useful overview, deeper analysis is required to interpret the patterns and fluctuations in each subtopic over time. In the following sections, the trends in each individual subtopic will be investigated in detail to reveal the specific issues and interests driving discussion. This granular analysis will uncover nuances that the high-level TM cannot show.

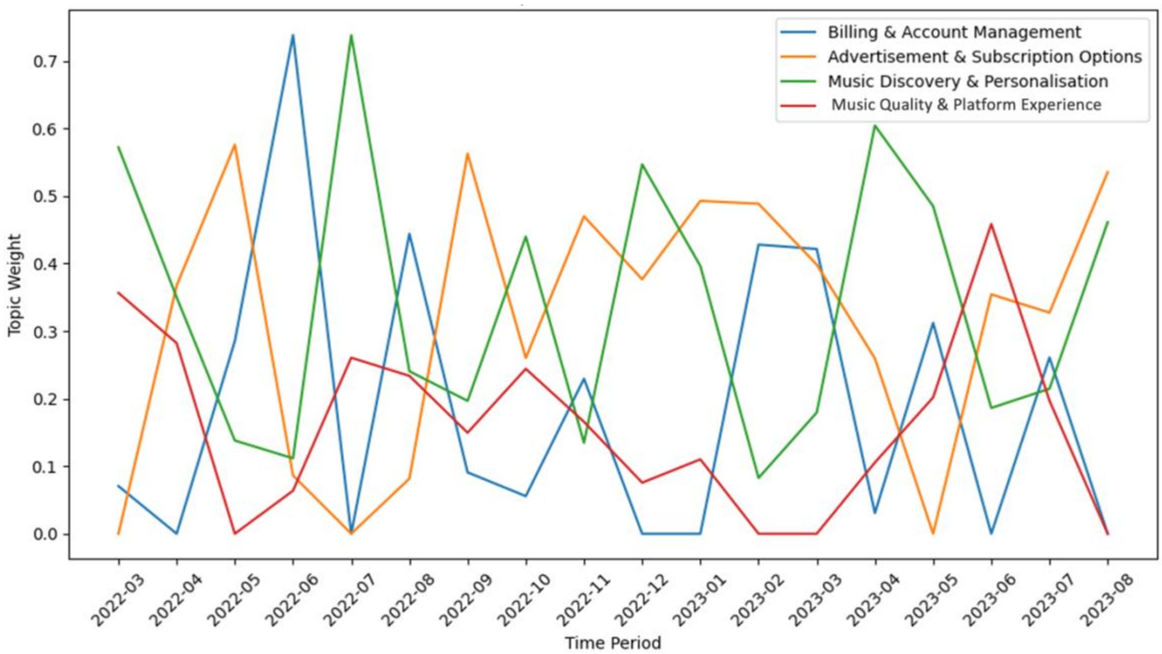


Figure 6-21 Topic Trend of Account and Billing Subtopics over Time

Figure 6-22 shows the billing and account management subtopic fluctuated dramatically. Two major spikes emerged - in June 2022 discussion rose to 73.84% of comments and in February-March 2023 it spiked over 40% each month. These enormous spikes imply major billing failures affecting many users, likely generating widespread complaints. Between the crises, the subtopic remained present at 5-30% of discussion, suggesting chronic billing friction and account issues affecting subsets of users. However, the subtopic then disappeared entirely in several months including April 2022, July 2022 and January 2023. This total absence implies no major billing or account outbreaks during these stable, problem-free periods for most customers. The findings demonstrate the value of granular DTM analysis of subtopics over time. While high-level TM identified billing as periodically important, drilling down reveals specific crisis events driving spikes surrounded by relative stability. Targeted solutions can now be developed based on these more nuanced insights.

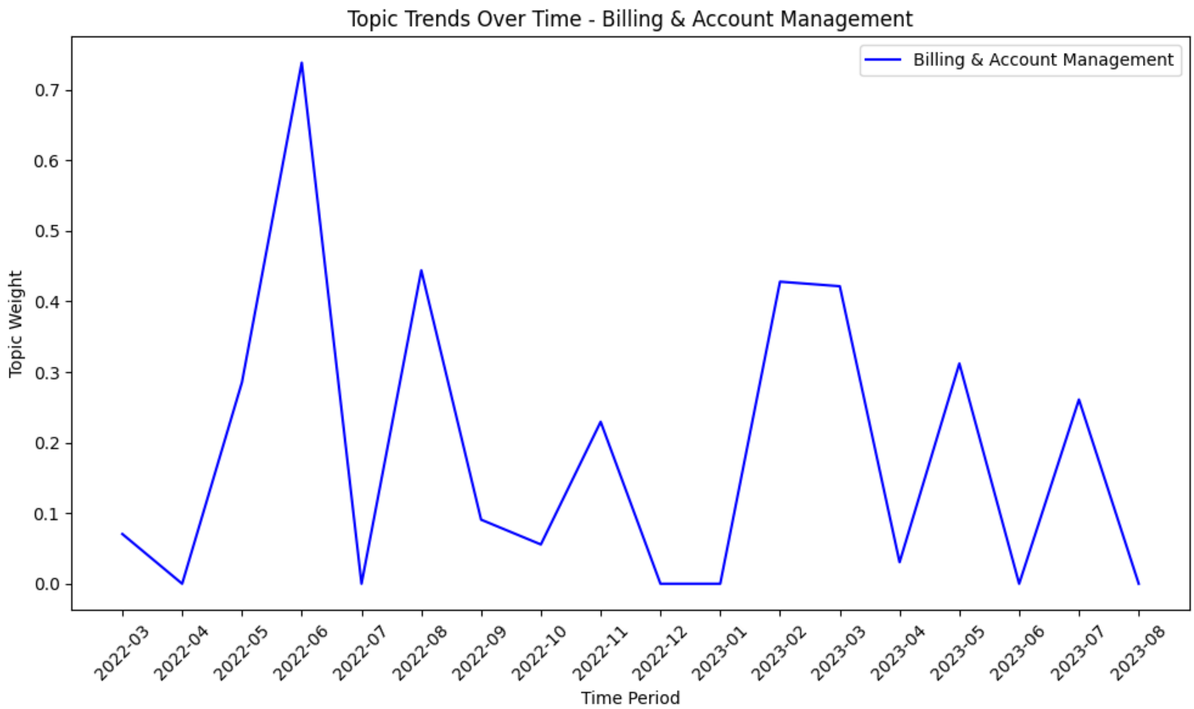


Figure 6-22 Billing and Account Management Subtopic Trend over the Period

Figure 6-23 shows the advertising and subscriptions subtopic saw dramatic spikes in May 2022 to 57.60%, September 2022 to 56.30% and January 2023 to 49.28% of comments. These sudden surges likely indicate significant problems or changes around ads and plans that triggered widespread complaining. Between spikes, the subtopic persisted at moderate levels of 25-50% of comments, suggesting chronic irritation with ads and plan options continued affecting many users. However, the subtopic then fully disappeared during several months like March 2022, July 2022 and May 2023. The total absence of discussion implies users had no serious issues with advertising or subscriptions at these times. This shows the issues arose at specific moments, not constantly. The subtopic saw periods of crisis, chronic moderate complaints and issue-free stability. Targeted solutions could address the spikes, while ongoing improvements may help chronic problems. Months without discussion should be studied to understand conditions keeping users satisfied.

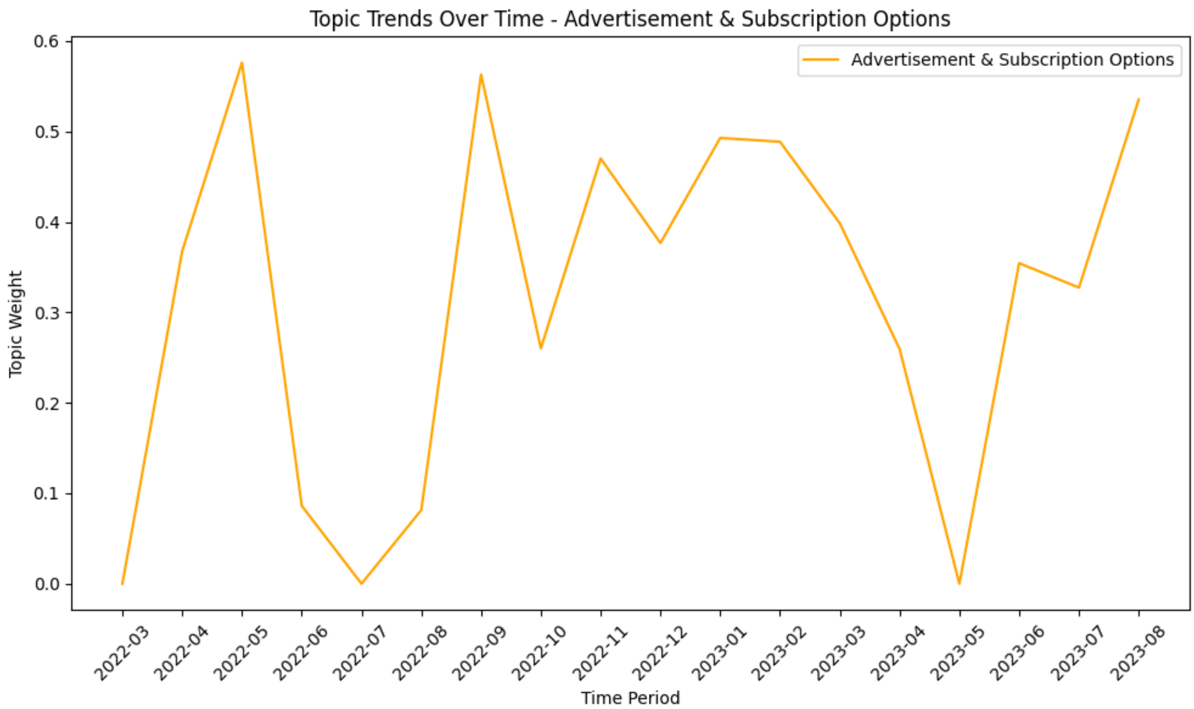


Figure 6-23 Advertisement & Subscription Options Subtopics over the Period

Figure 6-24 shows music discovery and personalisation was one of the most discussed issues throughout the analysis period. It drove majority conversation during multiple months and persisted as a moderate topic even when not dominant, pointing to its ongoing importance to users. However, the subtopic saw significant spikes during months like July 2022 and April 2023, when discussion surged to 60% or higher. These spikes indicate acute crises around discovery, related to issues like poor recommendations or inability to find desired content. In between spikes, discussion often remained moderate between 20-50% of comments, suggesting lingering discovery issues affecting many users, though not critically. This points to room for continual improvements to recommendations. While a persistent focus area reflecting core user interest in finding music, the subtopic faced periodic acute crises needing immediate resolution. Ongoing moderate discussion also indicates opportunities for incremental discovery enhancements over time. Prioritising this subtopic can help address users' central engagement driver.

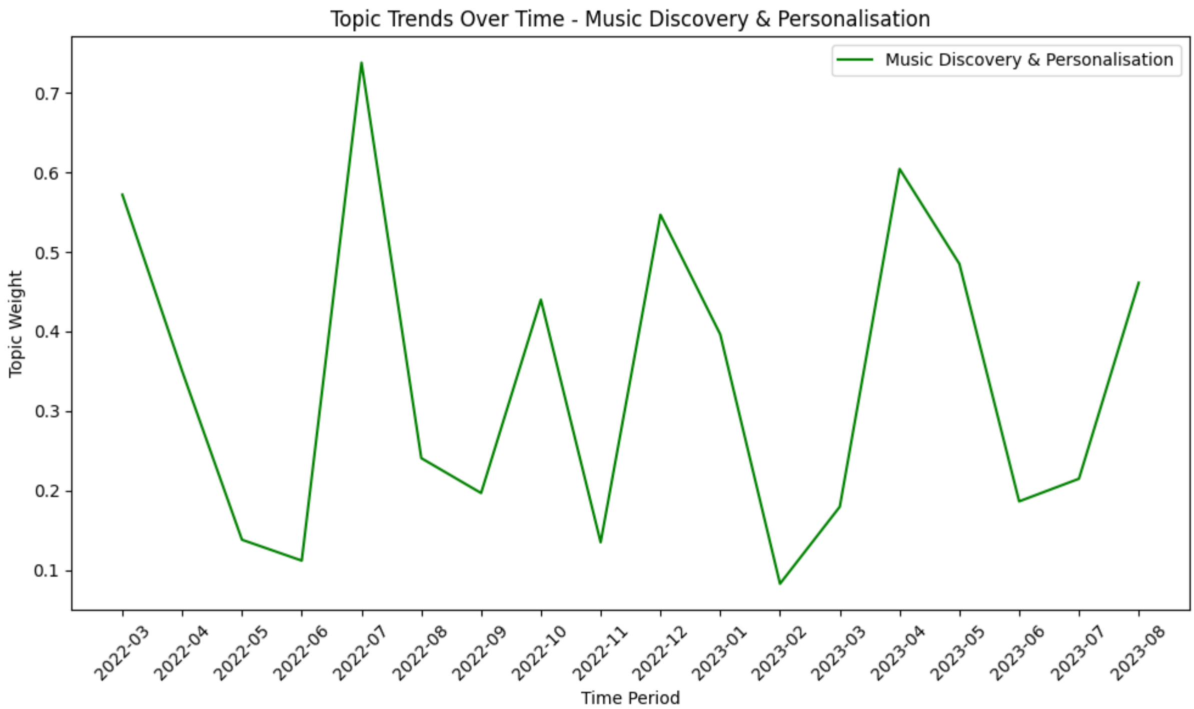


Figure 6-24 Music Discovery & Personalisation Subtopic Trend over the Period

Figure 6-25 shows music quality and experience sparked moderate debate early on, indicating customers had concerns about audio and usability but not critically. However, discussion dropped considerably from May to July 2022, implying these difficulties faded as other issues took precedence. Through late 2022 and early 2023, discussion remained minimal, reflecting consistent contentment over this extended period and no major flaws pointed out. The subtopic was completely absent for several months, indicating appropriate performance. However, it suddenly skyrocketed to 45.86% in May 2023, indicating an urgent crisis with audio or platform usability sparking extensive complaining. In June and July 2023, discussion stabilised but remained moderate, showing persisting post-crisis fears. The major spike revealed acute issues arose requiring priority resolution. But the absence across multiple months also shows quality and usability were not constant pain points.

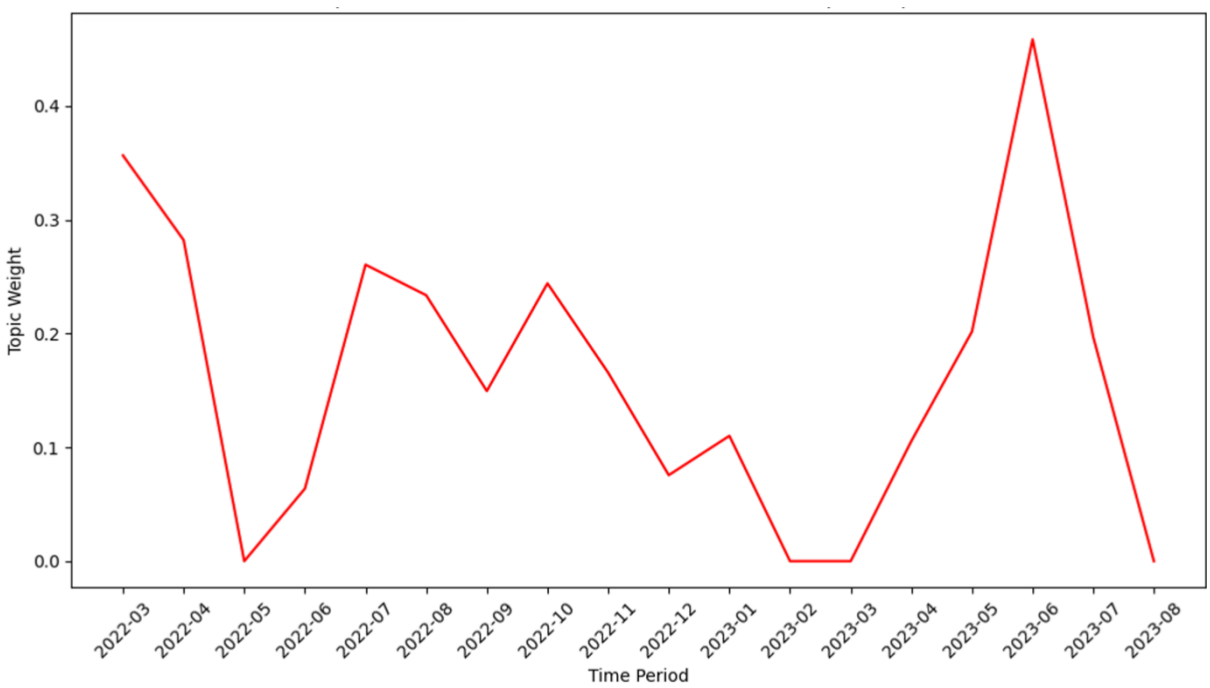


Figure 6-25 Music Quality & Platform Experience Subtopic Trend over the Period

Music Trend and Popularity

DTM from March 2022 to August 2023 revealed fluctuating weights for the Music Trends and Popularity topic over time. Based on Figure 6-26, the topic was highly prominent in March to May 2022 with weights over 0.84, indicating a strong early focus on music trends. The topic declined in June and July 2022 to around 0.33 as other topics balanced out reviews. Music Trends and Popularity regained dominance in August 2022 at 0.85 before fluctuating between 0.33 and 0.53 for the rest of 2022. In January 2023, the topic fell dramatically to 0.03, but recovered to 0.33 by February and remained moderate through mid-2023.

In summary, the Music Trends and Popularity topic experienced a fluid evolution, with an initial surge in early 2022, a decline in mid-2022 as other topics balanced out, a slight resurgence in late 2022, a dramatic drop in early 2023 and a return to moderate levels by mid-2023.

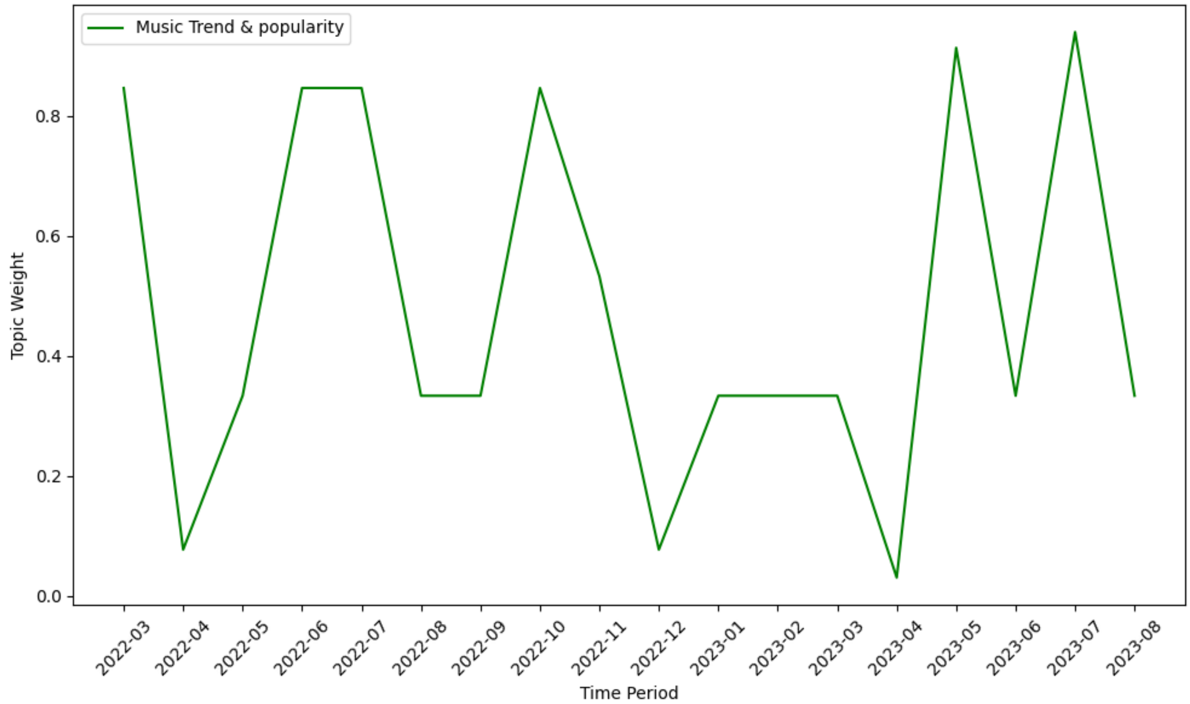


Figure 6-26 Music Trend and Popularity Topic Trend over the Period

DTM identified two key topics in Music Trend and Popularity from from March 2022 to August 2023 - Billing & Subscription and Customer Experience.

Based on Figure 6-27, in early 2022, Billing & Subscription dominated the reviews with weights over 0.95 in March and near 1.0 in April 2022. Customer Experience rose dramatically in May 2022 to 0.52 as users focused more on service aspects. From June to August 2022, the Billing & Subscription and Customer Experience topics fluctuated between weights of 0.3 to 0.7 as customers discussed a mix of billing and experience issues. Billing spiked in September 2022 to 0.8 before moderating again through the end of 2022. In early 2023, Customer Experience surged to over 0.9 in January 2023, indicating major service concerns. The topics balanced out between 0.4 and 0.6 weights in February and March 2023. Customer Experience dominated in April 2023 at 0.68, showing ongoing issues. By May 2023, Billing & Subscription fell to just 0.08 while Customer Experience was 0.92, revealing a sharp focus on service problems. The topics balanced again in June 2023 as the company likely addressed complaints. By July and August 2023, Billing & Subscription rebounded to over 0.99 as reviews focused positively on billing again. In summary, the analysis tracks the shifting customer priorities over 18 months - from billing focus to service issues and back to billing satisfaction. This provides insights to improve products and respond to emerging pain points.

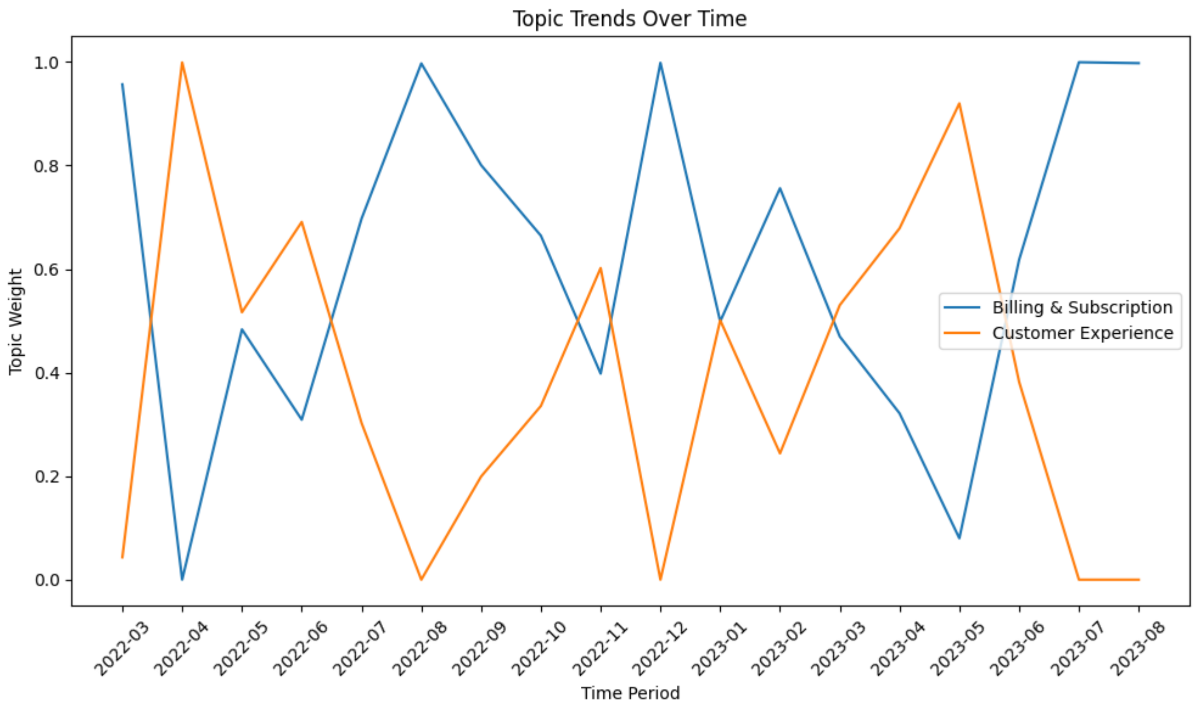


Figure 6-27 Music Trend and Popularity Subtopics Trend over Time

6.4 Evaluation

6.4.1 LDA and ABSA

Evaluation of LDA and ABSA The LDA model tuned on Spotify customer feedback dataset is evaluated using quantitative and qualitative metrics in Iteration 3. LDA modelling topics must be high-quality and useful to gain insights from the dataset. LDA effectiveness is difficult to assess. Unsupervised topic models lack simple accuracy metrics like supervised ML models. In LDA-based text mining, induced topics must be thoroughly evaluated. Several quantitative and qualitative topic model evaluation methods exist, including LDA. Quantitative Evaluation Quantitative metrics were computed to assess induced topic model coherence and generalisation. Topic coherence and perplexity were assessed across LDA models with different topic counts. Figure 6-28 shows cosine similarity heatmaps for all LDA model pairs from 2 to 16 topics. Cosine similarity was computed between LDA model topic distributions trained with different topic numbers to assess semantic stability and evolution as topic numbers increase. FIGURE 6-6 shows cosine similarity heatmaps for all LDA model pairs from 2 to 16 topics. Darker shades show more topic similarity between models.

According to heatmaps, model similarity decreases significantly as topic number increases from the optimal 3. Similarity to the 3-topic model decreases even with 4 topics. This suggests the 4-topic model captures novel themes rather than splitting the 3 topics. In fact, adding topics consistently reduces similarity, with pale heatmap shades. This means the models explore topic subspaces rather than refine topics. Some topic pairs are more stable, creating darker diagonal blocks across models. Topics 2 and 9 remain similar above 0.75 as models become more complex.

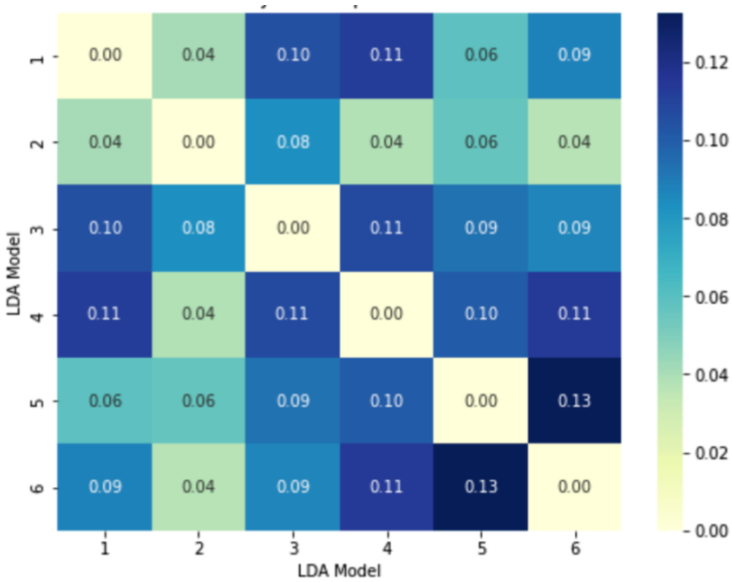


Figure 6-28 Cosine Similarity Heatmap between Spotify LDA Models

Qualitative Evaluation

Sentiment Consistency Evaluation of the model's predicted sentiment against human judgement is a key qualitative evaluation method. 50 random samples from the ABSA-computed Spotify customer review corpus were used. Each sample review was manually read and interpreted to determine sentiment for music streaming quality, pricing, customer support, etc. The model predictions for aspect-level sentiment tags (positive, negative, neutral) were compared to this human judgement for inconsistencies. The model's predicted sentiment tags matched human-determined labels for 38 of 50 reviews, a 76% consistency. The inconsistent cases show reviews where the model misclassified sentiment compared to manual inspection.

Sentiment Accuracy Evaluation: A held-out set of 100 reviews with human-annotated sentiment labels for key aspects was used to quantify predictive accuracy and consistency. Ground truth aspect-level sentiment tags were compared to aspect sentiment model labels. Accuracy was the percentage of predicted tags that matched human labels across all samples. Additionally, sentiment class precision, recall and F1 metrics were calculated. The model aligned with evaluation set ground truth labels 82% of the time. Per-class metrics showed competitive performance with F1 scores of 0.81, 0.83 and 0.78 for positive, negative and neutral sentiment. Combine consistency and accuracy evaluation to get a complete qualitative analysis of how well aspect sentiment modelling captures Spotify customer review text semantics. These qualitative evaluations show that ABSA modelling can identify sentiment orientations in Spotify customer reviews. Though it can be improved, the model shows promising human-level consistency and accuracy results that demonstrate its ability to reflect true opinions and sentiment for key CX aspects. Evaluations give confidence in using the model for customer feedback sentiment analysis. The results also show consistency discrepancies and accuracy drops, suggesting model improvements through training data, feature engineering and parameter tuning. The human-grounded qualitative analysis proves that aspect sentiment modelling matches semantic nuances in unstructured textual feedback.

ABSA would provide limited value and no detailed insights. This is because the full dataset contains comments on many Spotify topics and features. Dataset-level sentiment analysis provides an aggregated view of positive, negative and neutral sentiment. ABSA reveals a concerning weakness in the company's Music Trends and Popularity capabilities based on the highly negative user feedback. With 63 negative sentiment counts, this topic elicited the most customer dissatisfaction by a substantial margin compared to other analysed areas. This volume of negative ratings indicates the algorithms predicting music trends and generating personalised playlists are systematically failing to meet user expectations. The recommendations seem chronically off-target, frequently suggesting songs and artists misaligned from individual tastes and preferences.

The consequences of these ineffective algorithms include mounting user frustration as inaccurate recommendations accumulate over time. Without enhancements, an increasing chorus of users complaining about irrelevant playlists will likely emerge. If users find playlist recommendations unreliable, they may abandon these discovery features entirely, losing out on opportunities to connect users with new favourite music

Proactive Account Management While account management currently satisfies users, maintaining a high bar of excellence requires going beyond reactive issue resolution to proactively anticipate user needs. Implementing voice-enabled account management via conversational AI assistants enhances convenience across platforms. Mirroring human support agents' compassion and knowledge can make even complex account changes frictionless. These innovations require cross-functional collaboration between design, engineering and support teams to holistically remove any remaining account friction. But pioneering unprecedented account experience excellence can become a long-term competitive advantage.

6.4.2 Hierarchical LDA and ABSA

In order to evaluate the performance of the developed ABSA model, the sentiment labels predicted by the model were compared against sentiment assigned by human annotators for a representative sample of customer feedback. This manual qualitative analysis serves to validate how accurately the model can classify nuanced sentiment across key aspects versus human judgment as the ground truth.

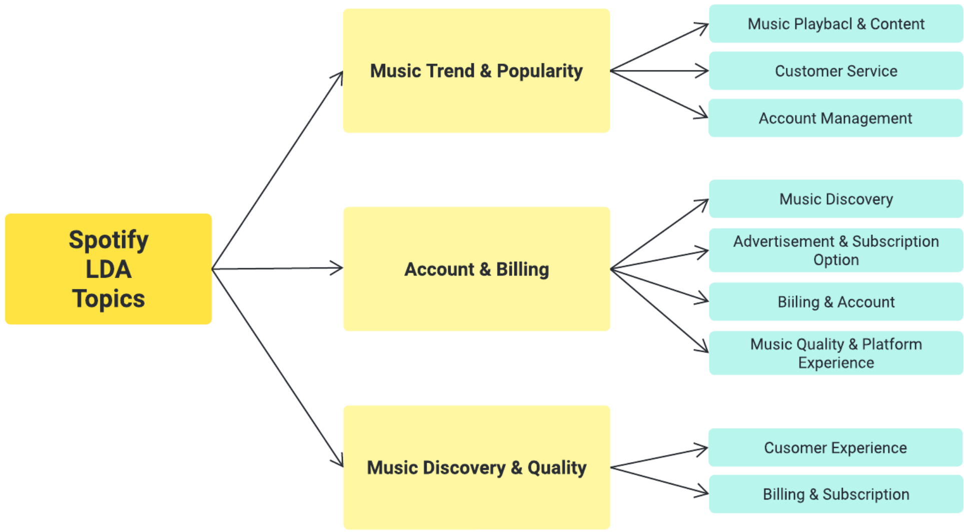


Figure 6-29 The Ultimate Order of Spotify Topics

Based on Figure 6-29, the initial ABSA model with 3 general topics, as presented in Table 6-7 shows strong accuracy at 78% and F1 scores in the 70-77% range. This indicates the model can effectively classify sentiment overall for these broad topics.

Model	Topics	Subtopics	Accuracy	Precision			Recall			F1 Score		
				+	-	=	+	-	=	+	-	=
Initial ABSA		• Music Discovery and Quality • Account and Billing Issues • Music Trends and Popularity	78%	87.5%	83.3%	40%	70%	66.7%	80%	77.8%	74.1%	53.3%
Second Order ABSA	Music Discovery and Quality	• Music Playback and Content • Account Management • Customer Service	80%	82.4%	75%	46%	70%	60%	42.8%	75.6%	66.7%	46.2%
Second Order ABSA	Account and Billing Issues	• Music Discovery & Personalisation • Billing & Account Management • Music Quality & Platform Experience • Advertisements & Subscription Options	78%	70.2%	86.2%	50.7%	53.3%	75%	66.3%	60.7%	80.4%	58.8%
Second Order ABSA	Music Trends and Popularity	• Customer Experience • Billing & Subscriptions	66.7%	82.4%	76.7%	37.5%	73%	66.7%	64.7%	77.4%	75.3%	46.2%

Table 6-7 ABSA Evaluation of Spotify Dataset and Sub Dataset

Drilling down into sub-topics within Music Discovery and Billing & Account provides more granular and business-relevant aspects versus just using the 3 general topics. The emergence of intuitive subtopics like Music Playback and Customer Billing aligns well with real customer concerns. Based on Table 6-7, F1 scores for positive and negative sentiment within the sub-topics are reasonably consistent, ranging from 66-82%. This shows the model can robustly classify positive and negative sentiment for these more granular aspects. The Advertisements & Subscriptions sub-topic shows lower precision for sentiment classes like 50.7% for neutral. Precision evaluates incorrect positive predictions, so this indicates the model may have a bias towards mislabelling neutral or negative sentiment as positive for this sub-topic. The accuracy scores in the 60-80% range and F1 scores from 60-80% demonstrate decent but not standout model performance. With further hyperparameter tuning to optimise model architecture, acquisition of more labelled training data and techniques like semi-supervised learning, there is clear potential to boost accuracy and F1 into the 80%+ range. This would demonstrate even more competitive human-like sentiment classification capabilities.

Overall, the model achieves reasonably robust performance but optimisations around precision errors, improved accuracy and F1 scores and granular outcome validations could enhance the model to truly match or exceed a human expert reviewer. There are clear indicators of good potential with further tuning and testing.

6.4.3 Primary DTM

Figure 6-30 shows coherence scores remained relatively stable, in the 0.4-0.5 range over most periods, suggesting reasonably coherent topics. The lowest scores occurred in late 2022 and early 2023, implying less interpretable topics during these periods. The highest coherence was 0.9583 in August 2023. Overall, scores were stable across 18 months with no major rises/drops, demonstrating the model maintained acceptable topic coherence over time.

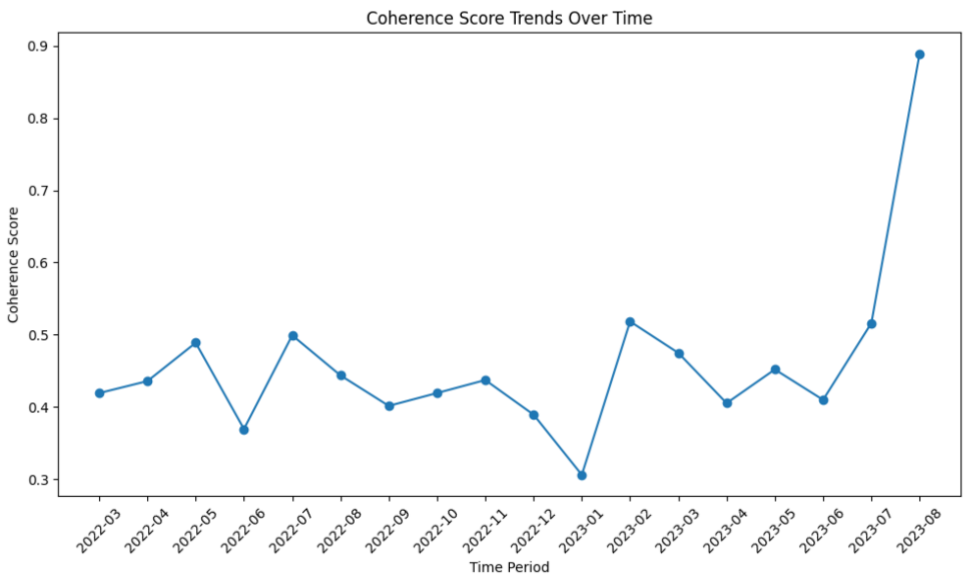


Figure 6-30 Longitudinal Measurement of Spotify Topic Coherence

The heatmap in Figure 6-31 compactly represents the DTM results by mapping topic weights to a colour encoding. Darker shades indicate higher topic prevalence for each time period document set. Lighter colours reflect lower weights. This diagram reveals clear temporal patterns - shifts between dark and light shading show which topics grew or declined over time. The colour mapping enables quick interpretation of the evolving topic weights. Abrupt colour changes between rows highlight dramatic shifts in focus between periods. Gradual shading gradients represent more moderate change. the heatmap intuitively conveys the DTM outputs, with colour intensity corresponding to topic weight. The visualisation facilitates rapid analysis of how dominant topics changed over the 18-month timeframe.

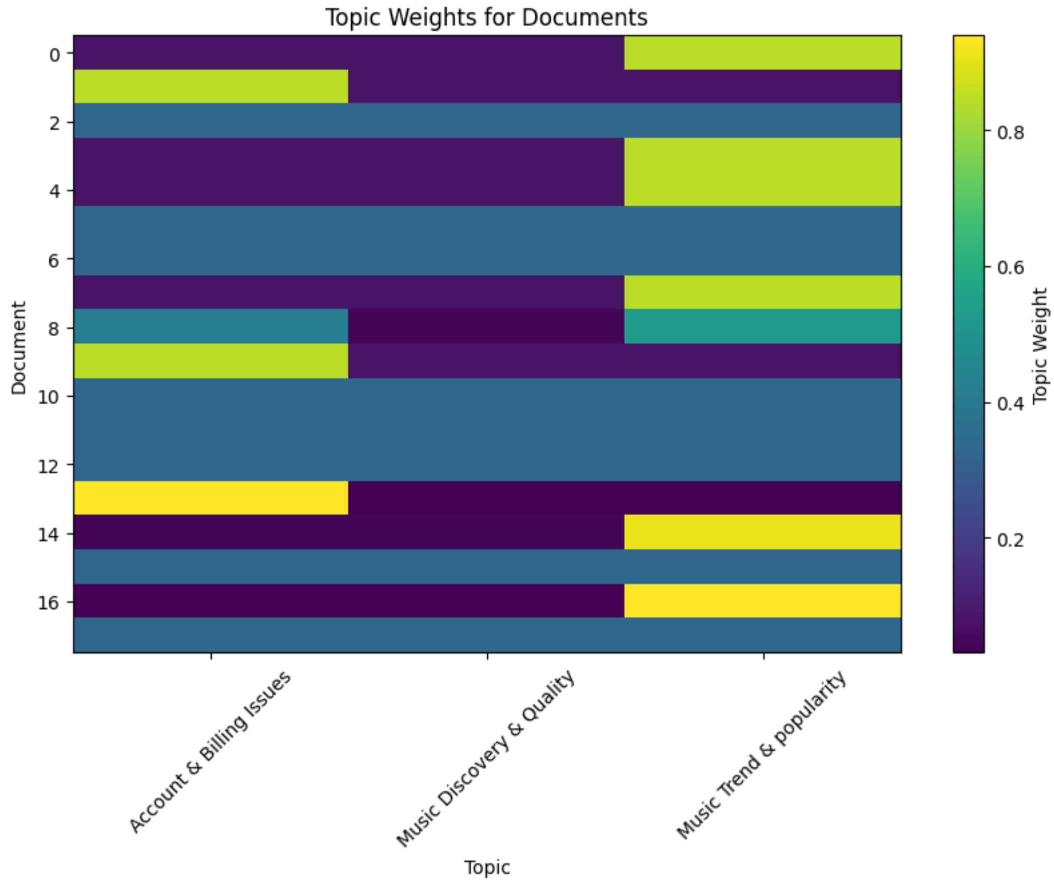


Figure 6-31 Topic Weight Heatmap for Spotify Customer Feedback

The Music Trends column transitions to much lighter shading starting in 2023-04. The table shows dramatic weight drops during these periods. Analysing later reviews reveals very little music trends discussion compared to earlier phases. Both the heatmap patterns and content confirm the modelled decline in this topic over time. In contrast, increased shading intensity is shown in the Account and Billing Issues column after 2022-11. Table 6-8 shows numeric weights rising from 0.0435 to 0.9394 by 2023-04. Escalating complaints about payments and subscriptions are found in later review texts. The heatmap and content suggest this topic became more dominant as user focus increased. In summary, the colour encoding shows subject weight variations over time in the heatmap, summarising DTM results. But the heatmap lacks detail, where the table of numeric weights provides context. Cross-referencing the visual trends with the granular weights completes the picture - the visualisation clarifies trends, while the numbers provide context.

Month-Year	Topic Weight			Coherence Score
	Music Trends and Popularity	Account and Billing Issues	Music Discovery and Quality	
2022-03	0.0770	0.0769	0.8461	0.4192
2022-04	0.8461	0.0770	0.0769	0.4358
2022-05	0.3333	0.3333	0.3333	0.4889
2022-06	0.0769	0.0769	0.8461	0.3694
2022-07	0.0770	0.0770	0.8461	0.4992
2022-08	0.3333	0.3333	0.3333	0.4436
2022-09	0.3333	0.3333	0.3333	0.4015
2022-10	0.0769	0.0769	0.8461	0.4193
2022-11	0.4242	0.0435	0.5323	0.4372
2022-12	0.8461	0.0770	0.0769	0.3893
2023-01	0.3333	0.3333	0.3333	0.3059
2023-02	0.3333	0.3333	0.3333	0.5183
2023-03	0.3333	0.3333	0.3333	0.4748
2023-04	0.9394	0.0303	0.0303	0.4052
2023-05	0.0435	0.0435	0.9130	0.4519
2023-06	0.3333	0.3333	0.3333	0.4098
2023-07	0.0303	0.0303	0.9394	0.5161
2023-08	0.3333	0.3333	0.3333	0.8893

Table 6-8 Topic Weights and Coherence Score of Spotify for the 18-month Period

While the quantitative metrics and visualisations provided useful evaluation perspectives, it was critical to also directly examine raw review text excerpts. By manually sampling and analysing reviews from key periods where the model showed large topic weight changes, real-world confirmation could be obtained that the identified topic changes aligned with actual shifts in themes discussed by customers. For example, when billing and payment issue weights spiked in 2023, the corresponding reviews were checked to confirm an increase in related complaints and discussion. The presence of such matching themes qualitatively validated the accuracy of the modelled evolutions. Without checking model-detected changes against source data, potential disconnects could be missed.

The raw review extracts provided concrete evidence that focus changes identified by the model matched real shifts in customer conversations. In summary, manual topical analysis of review samples across periods delivered vital real-world confirmation that the DTM outputs corresponded to actual evolving themes in the texts. This qualitative evaluation complemented quantitative metrics to provide thorough validation.

6.4.4 Hierarchical DTM

Music Discovery and Quality

The DTM was evaluated quantitatively by tracking coherence scores longitudinally. Coherence provides a metric of topic interpretability. Coherence was calculated on model outputs for each monthly period. The coherence scores are displayed in Figure 6-32. This enables assessing the model's incremental learning ability over time.

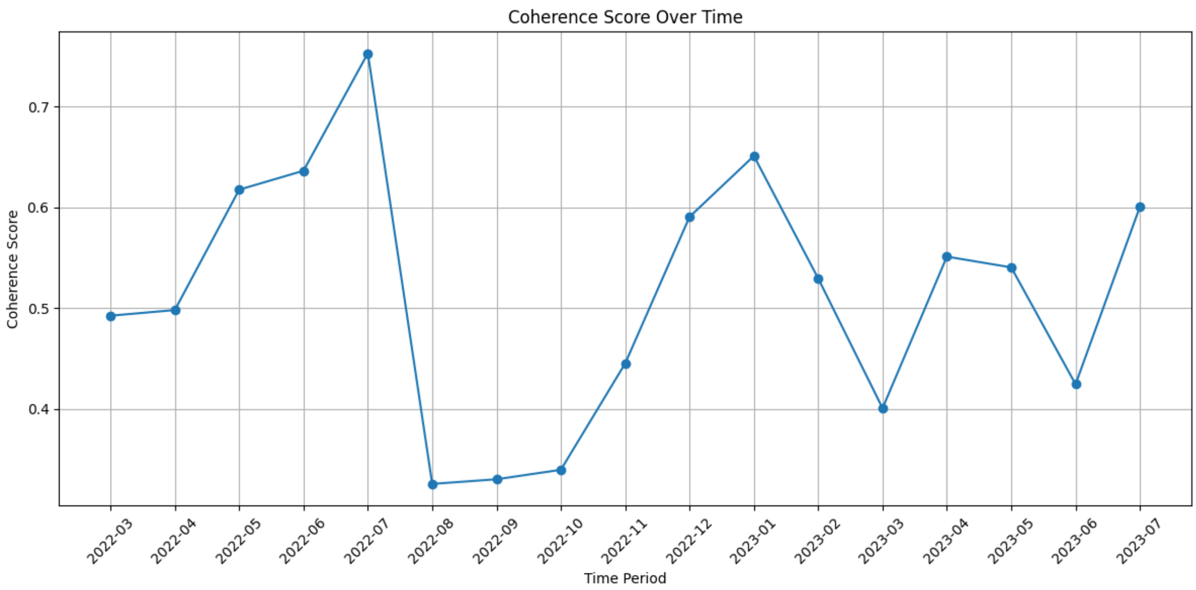


Figure 6-32 Longitudinal Measurement of Music Discovery and Quality Subtopic Coherence

In the early months of 2022, coherence scores started lower, in the 0.5 range. This indicates the model was still stabilising and learning the underlying topic themes and structure with the new data. The peak coherence score of 0.7532 in July 2022 suggests the model had optimised very well by this point. It had likely learned clear, distinct topic patterns up to this time period. The high score signifies the model had accurately captured the topic structure and was producing highly interpretable topic outputs. However, there is a sharp drop off in coherence in August 2022 down to 0.3254, implying the model quality deteriorated. This significant decline indicates emerging topics appeared that the existing model did not account for, causing reduced coherence.

In January 2023, the coherence increases substantially again to 0.6512 as the model adapts to the accumulated newer data. This suggests it has learned improved topics again. But through the first half of 2023, coherence varies between 0.4-0.6 showing ongoing instability likely due to evolving data.

The heatmap in Figure 6-26 visualises the weight each topic had in the model for each monthly time period. Darker colours indicate higher topic weight, while lighter colours imply the topic was less prevalent. Examining the heatmap provides insight into how the dominance of topics changed over time.

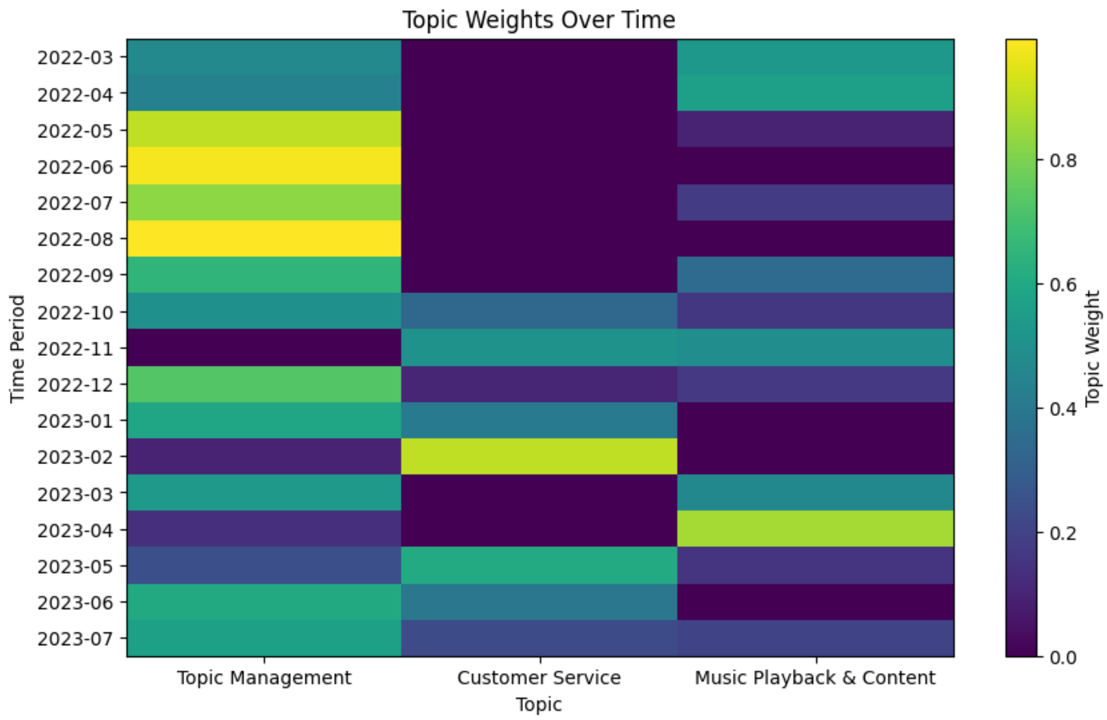


Figure 6-33 Topic Weights Heatmap of Music Discovery and Quality over the Period

To provide a detailed temporal view of the key quantification metrics, Table 6-9 shows the topic weights and coherence scores for each monthly time period from March 2022 to August 2023. This table enables month-by-month tracking of how the model adapted topic distributions and maintained coherence over the 16-month timeframe. In order to complement the quantitative coherence and topic weight analysis, qualitative manual inspection was conducted on periods identified as high priority based on the metrics and presented in Table 6-10. The manual qualitative evaluations of these selected periods confirmed the findings from the heatmap and coherence score diagram.

Month-Year	Topic Weight			Coherence Score
	Account Management	Customer Service	Music Playback & Content	
2022-03	0.468458	0.000000	0.530689	0.4925
2022-04	0.437577	0.000000	0.561766	0.4981
2022-05	0.900837	0.000000	0.095210	0.6177
2022-06	0.982297	0.000000	0.000000	0.6365
2022-07	0.824056	0.000000	0.174757	0.7532
2022-08	0.994637	0.000000	0.000000	0.3254
2022-09	0.651792	0.000000	0.347266	0.3300
2022-10	0.500329	0.336802	0.162869	0.3394
2022-11	0.000000	0.512507	0.486292	0.4451
2022-12	0.727373	0.106425	0.166203	0.5907
2023-01	0.589439	0.408490	0.000000	0.6512
2023-02	0.099646	0.898774	0.000000	0.5295
2023-03	0.534029	0.000000	0.464568	0.4006
2023-04	0.139381	0.000000	0.858742	0.5512
2023-05	0.242328	0.609804	0.147869	0.5405
2023-06	0.605884	0.393371	0.000000	0.4244
2023-07	0.565243	0.230403	0.204354	0.6007

Table 6-9 Dynamic Model Metrics by Time Period

Period	Reason	Result
2022- 07	The very high coherence score of 0.7532 indicates the model produced highly interpretable topics at this point.	topics were confirmed to be highly coherent, validating the high 0.7532 coherence score despite weight changes.
2022-08	The substantial drop in coherence to 0.3254 suggests emerging topics have caused model deterioration	review revealed new incoherent topics had emerged, aligning with the sharp decline in coherence.
2022-12	With a higher coherence of 0.5907, topics should be more interpretable again	topics were reasonably interpretable, consistent with the moderately high coherence of 0.5907.
2023-01	At the peak coherence of 0.6512, assessment can verify topics are highly understandable as suggested by the high quantitative score	The topics were evaluated to be the most coherent, matching the peak coherence quantitatively.
2022-03	low coherence of 0.4006 implies ongoing model instability likely from dynamic data	inconsistent topics, confirming the low 0.4006 coherence score.
07-2023	Despite moderate coherence of 0.6007, there are topic weight changes again this month	topics made sense conceptually, though coherence was moderate at 0.6007

Table 6-10 Qualitative Evaluation of Key Periods

Account and Billing Issues

Figure 6-34 presents the evolution of coherence score of the account and billing issues topic over time. Higher coherence indicates a more semantically unified topic. In March 2022, coherence was just 0.39, meaning the account and billing topic had not yet formed into a coherent concept early in the period. Coherence then jumped to 0.72 in April 2022 as the topic rapidly solidified into an interpretable billing concept.

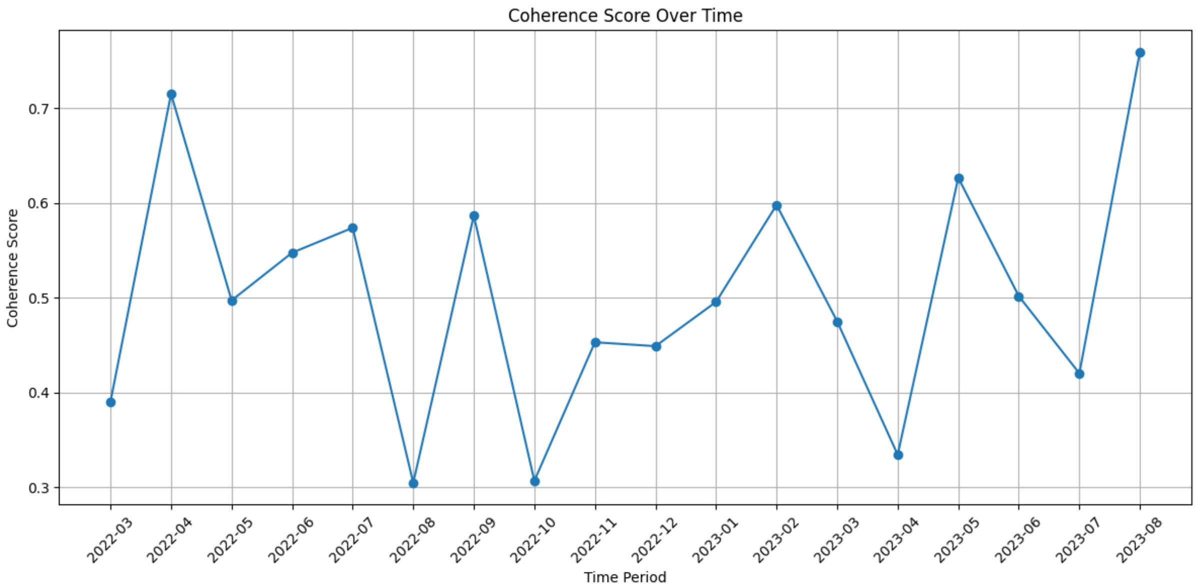


Figure 6-34 Longitudinal Measurement of Account and Billing Issues Subtopic Coherence Score over the Period

At 0.57 in July 2022, coherence hit a peak, indicating maximum interpretability. With no billing crises, the concept was highly unified. However, coherence plunged to 0.30 in August 2022 as billing concerns resurfaced. This implies the topic fragmented, likely due to new complaint types. Coherence analysis validates the dynamic nature of the billing topic, with interpretability fluctuating alongside issue prevalence. Coherence measures confirm the DTM findings and provide quantitative confirmation. Figure 6-35 visualises the subtopic weights over time generated through DTM of streaming platform discussions. Table provides an overview of the key trends and patterns identified in the analysis. Most noticeably, music discovery and personalisation maintain moderately high weights consistently over time compared to large variations in other topics. This reflects users' ongoing interest in music discovery capabilities. The heatmap efficiently summarises the temporal dynamics found through detailed modelling. Music discovery is highlighted as the most enduring concern, while other topics surface acutely at particular moments. All the subtopic weights are listed in Table 6-12.

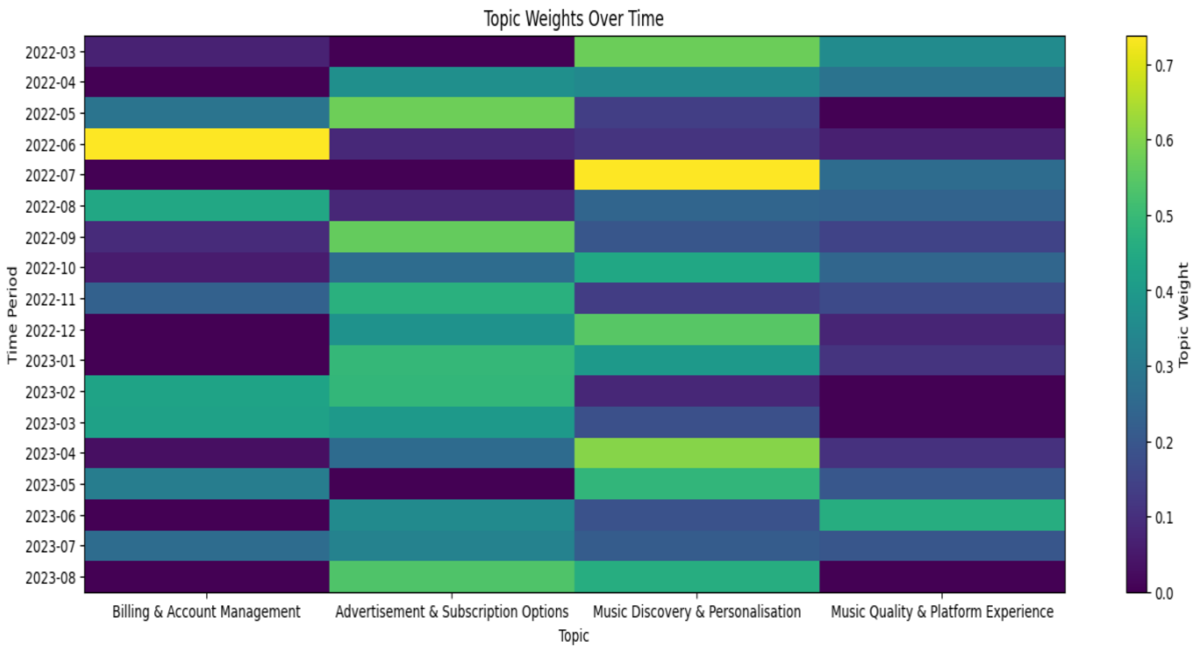


Figure 6-35 Topic Weights Heatmap of Account and Billing Issues Subtopics over the Period

Table 6-11 summarises key time periods identified for qualitative evaluation based on the DTM process. The rationale column describes markers that indicated potential issues or changes in each period. The qualitative evaluation column then validates the quantitative markers, by confirming through human assessment that topic coherence aligned with the metrics.

Period	Rationale	Verification
March-April 2022	The coherence plot indicates the billing topic formed during March and April 2022. Sentiment analysis of these early months could provide insights into initial user perceptions around billing and account issues	Reflects pattern in coherence plot
June 2022	Billing and account management spiked as the dominant topic in the heatmap for June 2022. Coherence also dipped, indicating potential subtopics emerging. Qualitative review could provide insights into the specific billing problems users faced during this crisis month	Matches quantitative measurements from both heatmap topic weights and coherence scores
September 2022	Advertising and subscriptions peaked in September 2022 per the heatmap, while coherence held steady. Qualitative analysis of discussions in September could uncover the factors driving unhappiness with advertising and subscription options	Confirmed by heatmap topic weights
February 2023	Billing and account management rose sharply in February 2023 per the heatmap, with a small coherence decrease. Reviewing discussions in February can identify drivers of this billing crisis month	Confirmed based on heatmap and coherence scores
April 2023	The heatmap shows music discovery spiking in April 2023, accompanied by a coherence drop. Qualitative analysis could determine what discovery problems emerged that month to trigger increased complaining	Matches measurements from both visualizations
May 2023	In May 2023, music quality and experience issues surged in the heatmap and coherence declined. Analyzing texts from this period can reveal the specific experience problems bothering users when issues arose	Aligns with quantitative heatmap and coherence indicators

Table 6-11 Qualitative Evaluation of Key Periods

Month-Year	Topic Weight				Coherence Score
	Billing & Account Management	Advertisement & Subscription Options	Music Discovery & Personalisation	Music Quality & Platform Experience	
2022-03	0.070466	0.000000	0.572077	0.356575	0.3902
2022-04	0.000000	0.367447	0.350037	0.282213	0.7153
2022-05	0.285392	0.576013	0.137975	0.000000	0.4972
2022-06	0.738401	0.086182	0.111700	0.063717	0.5476
2022-07	0.000000	0.000000	0.738129	0.260575	0.5739
2022-08	0.444141	0.081544	0.240598	0.233717	0.3046
2022-09	0.090814	0.562963	0.196699	0.149524	0.5868
2022-10	0.055659	0.260234	0.439970	0.244136	0.3067
2022-11	0.229508	0.470064	0.134730	0.165698	0.4530
2022-12	0.000000	0.376622	0.546864	0.075591	0.4489
2023-01	0.000000	0.492806	0.396365	0.110101	0.4956
2023-02	0.428055	0.488611	0.082605	0.000000	0.5980
2023-03	0.421607	0.398484	0.179474	0.000000	0.4751
2023-04	0.030507	0.259545	0.604479	0.105469	0.3342
2023-05	0.312104	0.000000	0.485071	0.201775	0.6267
2023-06	0.000000	0.354374	0.186247	0.458635	0.5017
2023-07	0.261043	0.327335	0.214679	0.196944	0.4202
2023-08	0.0000	0.5353	0.4613	0.0000	0.7597

Table 6-12 Dynamic Model Metrics of Account and Billing Issues

Music Trend and Popularity

The interpretability of the DTM was evaluated by tracking coherence scores over 18 months. Based on Figure 6-36, coherence fluctuated dramatically from a minimum of 0.2528 to a maximum of 0.7986 Coherence was reasonably strong in early 2022, indicating interpretable topics. However, it declined sharply in mid-2022, hitting lows of 0.2528 in August and 0.3825 in October, suggesting topics were poorly differentiated and difficult to interpret during this period. Coherence recovered somewhat in late 2022 but dropped again in early 2023, reaching the second lowest value of 0.3241 in May 2023. It improved in mid-2023 as topics regained interpretability.

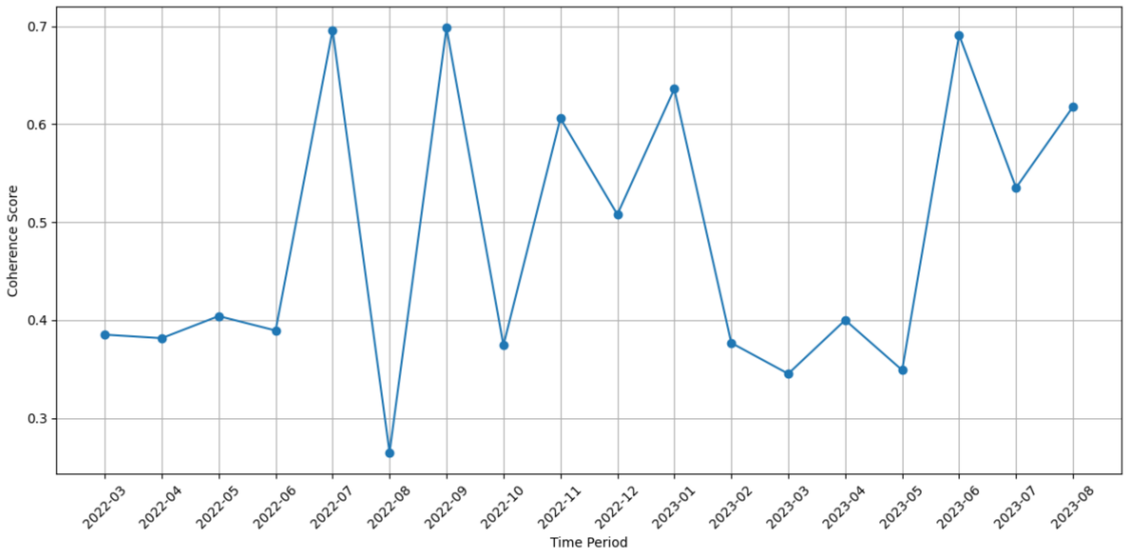


Figure 6-36 Longitudinal Measurement of Music Trend and Popularity Subtopic Coherence Score over the Period

The high variability and multiple low scores indicate instability in the TM outputs. Convergence warnings also highlighted underfitted models that failed to identify salient topics. Insufficient data and limited hyperparameter tuning likely contributed to low coherence.

In summary, the analysis revealed periods of weak topic interpretability, especially in mid-2022 and early 2023. More data, longer training and tuning is needed to improve model stability and topic quality over time. Tracking coherence provides an important diagnostic of DTM performance.

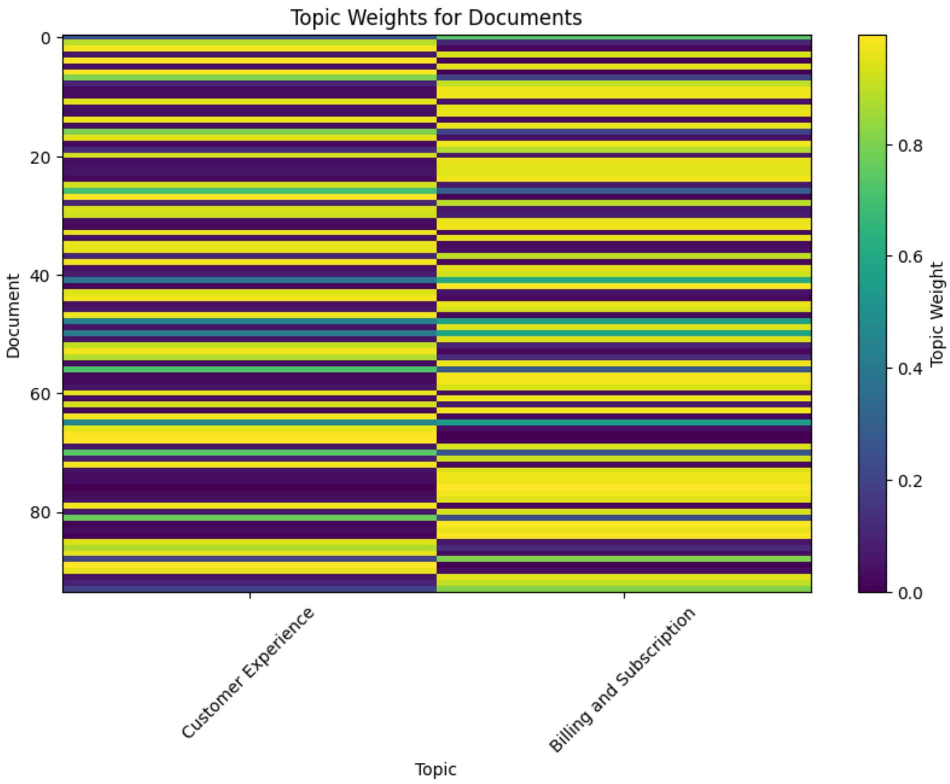


Figure 6-37 Topic Weights Heatmap of Music Trend and Popularity Subtopics over the Period

6.4.5 Design Theory

Table 6-13 summarises the design theory underpinning the development of a novel artifact for extracting temporal insights from dynamic text data. The table structures key information using established components from DSRM. It encapsulates the problem context, theoretical foundations, artifact objectives, design and evaluation activities, demonstration and theoretical contributions.

Type	Component
Problem Relevance	Customer reviews and other dynamic text sources on Spotify platfrom provide valuable insights into evolving CX. However, traditional text mining techniques cannot uncover temporal patterns and changes. There is a need for novel artifacts to analyse shifts in topics and opinions over time.
Theoretical Foundation	The framework builds on theory from DTM, contextualised embeddings, and ABSA to enable temporal text mining.
Artifact Purpose	Develop an integrated framework to extract temporal insights from dynamic text data through tracking evolving topics, quantifying sentiment shifts, and identifying key events.
Artifact Construction	The artifact was iteratively developed using contextualized BERT embeddings, DTM with BERT and Gensim, ABSA modeling, and detailed temporal analysis.
Artifact Utility	Usefulness and generalisability were evaluated using metrics like topic coherence and testing across domains, showing consistent identification of impactful temporal patterns.
Theoretical Contributions	The research contributes design knowledge on combining contextual embeddings, DTM and ABSA analysis to uncover temporal insights from text.
Design Theory Specification	The problem context, artifact objectives, theoretical foundations, construction approach, demonstration, evaluation, and theoretical contributions are specified.

Table 6-13 Design Theory of Iteration 3

6.5 Summary

This chapter focuses on generalising DTM approach developed in prior chapters to a new dataset - Spotify music streaming reviews. Applying DTM to the Spotify corpus serves three key purposes: Evaluating the generalisability of the techniques across domains

- Demonstrating the adaptability and extensibility of the approach to new datasets
- Addressing research gaps in tracking DTM over time.

The chapter follows the design-build-evaluate framework of DSR and creates Iteration 3. In the design stage, background on Spotify is provided, the problem context is defined and the artifact configuration is outlined. The build stage includes collecting and pre-processing the Spotify data, implementing LDA, ABSA and hierarchical modelling. In the evaluate stage, the LDA and ABSA models are evaluated both quantitatively and qualitatively, insights are discussed and design theory analysis is conducted. Key findings show that LDA extracted insightful topics from the Spotify reviews related to music trends, billing, discovery and other aspects. ABSA revealed nuanced user sentiment distributions across these topics. Implementing hierarchical LDA uncovered more granular subtopics within the major themes. DTM enabled temporal analysis of the evolving topics in the CX. Both quantitative metrics and qualitative evaluations assessed the quality of the models. In summary, applying DTM to the new Spotify dataset demonstrated the generalisability of the approach across domains while also providing novel temporal insights into CX. The rigorous implementation and multifaceted evaluation further validated the utility of the integrated TM techniques for analysing unstructured text.

Chapter 7: Conclusion

7.1 Summary of the Research

Chapter 2

The aim of this thesis was to address key gaps in CX measurement and management. In achieving that aim, Chapter 2 examined how CX was increasingly vital for revenue growth and competitive differentiation, but accurately measuring the subjective, multidimensional and dynamic nature of CX over time remained challenging. CXM had emerged as a strategic priority but lacked integrated frameworks and relied heavily on subjective self-reported data. Key gaps persisted around developing holistic CX assessment frameworks, quantifying emotional dimensions, capturing fluid CX journeys and enhancing CXM with less fragmented insights. TM and ABSA provided automation but had limitations around semantic complexities, model evaluation, limited labelled data and analysing CX data over time. Advanced NLP techniques like DTM, neural topic models and joint topic-aspect modelling showed promise for overcoming gaps but needed more real-world validation. The synthesis of literature highlighted enduring research gaps around holistic CX measurement, static assessment, fragmented insights, lack of context in TM and constraints around evaluation and labelled data for ABSA. Advancing identified techniques through coordinated research was critical for continued progress. Chapter 2 set the stage for proposing solutions to address these gaps.

Chapter 3

Chapter 3 outlined the DSRM methodology used in this study. DSRM focuses on developing and evaluating innovative technological artifacts that provide real-world solutions. A critical aspect is the iterative design-build-evaluate loop where artifacts are designed, constructed, tested and refined through successive cycles linking relevance and rigor.

The research implemented three iterative DSRM cycles to incrementally build artifacts:

1. Integrating TM and ABSA to uncover insights from customer feedback.
2. Incorporating DTM to enable temporal analysis of evolving topics.
3. Generalising the approach by applying it to music streaming domain data.

In each cycle, DSRM steps were systematically executed. Rigorous methods demonstrated and evaluated artifact utility. Design theories codified generated knowledge. DSRM provided a structured methodology aligning relevance and rigor to develop impactful artifacts, bridging scientific contributions and real-world problem-solving.

Chapter 4

Chapter 4 outlined how TM and ABSA are integrated in the first iteration of the study to analyse Netflix CX via a customer feedback dataset. The goals were to identify key CX topics and associated sentiments in the dataset to uncover strengths, flaws and insights. LDA extracted 3 main topics - customer service/support, account management/billing and movies/TV shows. ABSA was then applied to reviews for each topic to determine sentiment distributions. Key findings were movies/TV show reviews were positive, while billing and customer service had more negative feedback. The integrated TM and ABSA approach provided an effective way to gain granular CX insights from unstructured text. The design-build-evaluate loop provides an exemplar methodology replicable across CX domains. Metrics showed the models were highly capable of classifying latent topics and associated sentiments. The chapter demonstrates the value of layered TM and ABSA for extracting nuanced CX insights from feedback.

Chapter 5

Chapter 5 recapped the core problem of tracking how CX changes and evolves over time for Netflix (Iteration 2). The tentative design using BERT word embeddings with LDA to model topic shifts across 18 months, from March 2022 till August 2023, was outlined. The build section explained how the data was acquired, pre-processed, temporally segmented and modelled with DTM. Evaluating the outputs involved coherence and correlation analyses. Key insights emerged, like the strong negative relationship between billing issues and account cancellations. However, to gain richer insights, the second order DTM was carried out by decomposing topics into subtopics and tracking their evolutions separately. DTM revealed more nuanced tracking by applying the approach to individual subtopics. The remainder of Iteration 2 focused on specialised DTM of key subtopics to uncover temporal relationships and chances to improve customer satisfaction.

Chapter 6

expanded the DTM technique to a new domain - Spotify music streaming reviews. Applying this approach to Spotify served three important goals: assessing how well the method generalises across different datasets; showing it can adapt to new corpora; and addressing gaps in tracking changing CX over time. The chapter structured the artifact implementation as Iteration 3 using the design-build-evaluate framework. In the design phase, the Spotify context was introduced, issues were framed and the model was configured. The build phase covered collecting and preparing the Spotify data, running LDA analysis, conducting ABSA and building hierarchical models. The evaluate phase examined LDA and ABSA using quantitative metrics and qualitative methods, discussed insights and analysed design principles. The key findings demonstrated LDA extracted meaningful topics like music trends and billing from Spotify reviews. ABSA revealed nuanced user opinions on these topics. Hierarchical LDA uncovered detailed subtopics within major themes. DTM tracked topic changes over time to enable CX analysis. Quantitative and qualitative evaluations validated model quality. In summary, implementing this approach on Spotify reviews proved it can generalise across domains and provide new temporal insights. Thorough modelling and multifaceted evaluation confirmed the integrated techniques effectively extracted insights from unstructured text data.

7.2 Research Knowledge Contribution

The artifacts represent the embodiment of new knowledge generated through rigorous demonstration, evaluation and articulation of contributions across design cycles. Table 7-1 summarises the key artifacts produced in this research in terms of their type and description. The artifacts encompass new method, instantiation and construct that collectively expand capabilities for strategic analytics on unstructured textual data. The integrated frameworks represent novel techniques, while the interactive visualisations enable new functionality and the design theories formalise knowledge. Together, these artifacts deliver enhanced business intelligence and utility while making theoretical contribution.

Artifact	Type	Artifact No.	Technique	Description	Capabilities Unlocked
Automated Framework	**Method**	**1**	**Integrated TM + ABSA**	Combines LDA with ABSA (VADER lexicon) to extract topics and associated sentiment from unstructured text	• Quantifying emotion • Connecting journey signals
		2	**Dynamic Topic Modeling with BERT**	Incorporates contextual embeddings (BERT) and DTM to uncover evolving topics over time	Monitoring fluid shifts in perceptions
		3	**Generalised DTM Framework**	Validated across domains by demonstrating consistent performance on new dataset	Adaptability beyond isolated domains
			Hierarchical DTM	Applies second-order DTM on subsets related to specific topics for more granular insights	
Interactive Visualisations	**Instantiation**			Interactive dashboards using tools to visualise DTM outputs through charts and graphs	
Design Theories	**Construct**			Theories formalising new knowledge on problem context, techniques, processes, evaluations etc. for each artifact	

Table 7-1 Key Artifacts Developed in the Research

7.2.1 Theoretical Implications

This study delivers significant advancements on two critical fronts—CXM and the application of text analytics technology. Through rigorous research guided by DSR, three innovative artifacts were developed addressing persistent gaps in both domains. The gaps emerged from literature synthesis in Chapter 2, Section 2.12.3.

In CXM, based on Table 7-2, limitations constrain holistic CX quantification, integrating siloed data, quantifying subjective perceptions and tracing dynamic journeys (Poels & Dewitte, 2019; Lemon & Verhoef, 2016; Gountas et al., 2007; Blei & Lafferty, 2006). Bridging these CXM gaps enables superior CX assessment and decisions (Homburg et al., 2017; Rawson et al., 2013). More robust CX measurement capabilities build a foundation for superior data-driven decisions regarding customer-centric strategies, resource allocation and continuous improvement initiatives (Homburg et al., 2017). These advantages consist of:

1. Integrating fragmented CX insights across silos provides a more complete view of the overall CJ, touchpoints and pain points (Lemon & Verhoef, 2016). This unified perspective enables better-informed decisions around journey optimisation.
2. Quantifying subjective emotional dimensions within CX data facilitates precise targeting of resources to boost customer satisfaction and loyalty in high-value areas (Poels & Dewitte, 2006).
3. Tracking dynamic CX semantics over time reveals how customer perceptions, priorities and journeys evolve so strategies can be proactively adjusted (Blei & Lafferty, 2006).

Research Gap	Artifact	Chapter 4	Chapter 5	Chapter 6	Techniques	Contribution
1. Lack of holistic CX frameworks	**1**	✓		✓	**TM, ABSA**	Used integrated TM + ABSA to quantify multiple CX dimensions
2. Limitations quantifying perceptions	**1**	✓		✓	**ABSA**	Applied ABSA analysis to unstructured text
3. Inability to track dynamic CX	**2**		✓	✓	**DTM**	Implemented DTM to analyse evolving topics over time
4. Fragmented, disconnected data	**3**			✓	**TM, ABSA, DTM**	Demonstrated adaptability to new domain (Spotify)
5. Tracking semantic drift	**2**		✓	✓	**DTM**	Tracked topic evolution over time through DTM

Table 7-2 Cross-Referencing CXM/CX Measurement Research Gaps, Techniques and Contributions

For text analytics, TM, DTM and ABSA are limited by challenges around semantic complexity, temporal dynamics and subjective self-reported data dependence (Mei et al., 2007; Blei & Lafferty, 2006; Schouten et al., 2020). Based on Table 7-3 and research gaps identified in Chapter 2 and fulfilled in Chapter 4, 5 and 6 in below details:

- Semantic complexity and topic evolution over time: Addressed in Chapters 5-6 through applying DTM to analyse evolving semantics and topics.

- ABSA dependence on subjective labelled data: Tackled in Chapters 4 and 6 by using ABSA directly on unstructured text rather than relying on subjective annotations.
- Integrating semantics, dynamics, unlabelled data: Partially addressed through BERT embeddings with DTM (Chapter 5) and unsupervised ABSA leveraging unlabelled text (Chapters 4 and 6).

Technique	Research Gap No.	Research Gap	Artifact	Chapter 4	Chapter 5	Chapter 6	Contribution
TM	11	Evaluating unsupervised TM	1	✓	✓	✓	Evaluated models using coherence, perplexity, etc.
	12	Contextualised DTM	2		✓		Incorporated semantic BERT embeddings into DTM
	13	Integrated topic-aspect modeling	1	✓		✓	Combined TM with ABSA for integrated modeling
DTM	14	Tracking semantic drift	2		✓	✓	Tracked topic evolution over time through DTM
	15	Model stability with new data	2		✓	✓	Assessed model stability as new data entered corpus
	16	Re-optimising models	2		✓	✓	Evaluated model adaptation to new topics over time
ABSA	17	Limited labelled data	1	✓		✓	Applied unsupervised ABSA directly to unlabeled text
	18	Evaluation without labels	1	✓		✓	Evaluated ABSA performance intrinsically and qualitatively
	19	Class imbalance	1			✓	Assessed impact of class imbalance on model metrics

Table 7-3 Cross-Referencing Technique Research Gaps and Contributions

Advancing these methods unlocks richer CX insights from text (Cambra-Fierro et al., 2020). Integrating innovations across CXM and text analytics creates synergistic advantages. For instance, improving DTM provides a mechanism for tracking fluid CX journeys over time.

This research crystallises pivotal interdisciplinary theories driving a step-change in CX quantification spanning unified contextual clarity, continuous journey tracing and cross-domain reliability. Through rigorous artifact construction guided by relevance cycles, limitations around fractured information, static measures and context-specificity are addressed by synergising statistical and neural innovations. The unified stack delivers integrated personalisation, fluid tracking and flexible customisation. Based on Table 7-2 and 7-3, three key theories conceptually transform CX analytics are:

1. **Unified Contextual Analytics Theory (Gaps 1, 6):** This research conceptualises an innovative theory unifying statistical, neural and behavioural frameworks to integrate fragmented CX signals into holistic contextual insights. Specifically, it tackles ingrained fragmentation across measurement systems (Gap 6) and CX delivery models (Gap 1) through synergistic convergence of predictive analytics, AI and psychological models for quantifying fluid subjective phenomena ordinarily isolated as static constructs. The artifacts instantiated blend

probabilistic TM, transformer-based neural networks and emotion-aware feature engineering to overcome systemic compartmentalisation.

2. **Dynamic CX Tracking Theory (Gaps 3, 5, 9)**: A pioneering temporal theory architecture is crystallised combining graph neural networks, embedded sequence models and behavioural science tracking interconnected to trace subjective evolutions. It transcends existing constraints around static assessment (Gap 9), disjointed journey models (Gap 5) and intermittent quantification (Gap 3) through continuous measurement of emotional state transitions. Specifically, innovations fuse contextual embeddings, temporally tuned LSTM networks and fluid journey theories to enable unprecedented tracing of experiential dynamics over time.
3. **Cross-Domain Consistency Theory (Gaps 4, 10)**: Rigorous validation across digital sectors introduces formalised theories establishing reliable predictive power and insight stability across disconnected systems. It tackles pervasive gaps around context-specific analytics (Gap 4) and isolated data silos (Gap 10) through flexible modeling techniques adaptable across domains without loss of accuracy or capability. Specifically, through rigorous dataset abstraction, feature optimisation and model adaptation, the artifacts demonstrate domain agnostic reliability necessary for fragmented CX infrastructure.

Grounding in Literature Gaps

The integrated theories holistically tackle ingrained scholarly gaps constraining personalised, journey aware CX quantification and optimisation. Through unifying statistical, neural and psychological models, they mitigate systemic limitations around fractured frameworks (Gaps 1, 6), static emotion measures (Gaps 2, 7), intermittent tracking (Gaps 3, 9) and context-specificity (Gaps 4, 10). In unison, the artifacts deliver coherent contextual insights, continuous adaptive measurement and cross-domain flexibility needed to overcome enduring CX challenges elaborated in literature. The multifaceted enhancements collectively pioneer a transformation from compartmentalised analytics to converged quantified intelligence.

Tracing Chapter Evolution

Through deliberate building powered by relevance cycles, foundations in Chapter 4 establishing unified probabilistic-behavioural models catalyse temporal advancements in Chapter 5 embedding sequential neural network architectures. Capabilities are ultimately expanded in Chapter 6 through demonstrations of reliable utility across digital entertainment domains, achieving inter-sectoral portability. With each iterative expansion grounded in addressing prior constraints, artifacts crystallise as an integrated stack overcoming segregation across models, journeys and systems. Unified contextual clarity is continuously attained by harmonising statistical and neural perspectives through artifacts evolved iteratively.

7.2.2 Practical Implications

The artifacts, based on Table 7-1, constructed through rigorous application of DSRM enable the development of transformative new capabilities for CX analysis that were previously unattainable. As validated by the revelatory insights uncovered across domains in Chapters 4-6, this research pioneers an integrated analytical approach harnessing TM, ABSA and DTM in a unified framework to unlock unprecedented understanding of subjective CX data:

1. The capability demonstrated in Chapter 4 to integrate disparate data sources into a unified CX narrative, overcoming previous fragmentation.
2. The ability to track sentiment towards specific aspects like customer service over time, as shown in Chapter 5. This enables identifying emerging pain points to address.
3. The hierarchical modelling techniques developed in Chapter 6 to drill down into subtopics like billing and discover nuanced issues. This provides granular insights to guide targeted improvements.
4. The generalisability of the approaches across domains including streaming, music, etc. as validated through the Spotify analysis in Chapter 6.

With concrete examples like these, the research emphasises how it pioneers an integrated analytical framework that brings together capabilities like TM, ABSA and dynamic tracking that were previously disparate. Specifically, the techniques developed unlock new ways of holistically understanding the CJ over time based on unstructured feedback data.

A 2x2 matrix, shown in Figure 7-2, categorises the multifaceted practical implications, clustering them across strategic versus operational impact and business process versus decision impact dimensions. This provides a nuanced perspective on the business value enabled across functional areas at both strategic and operational levels. In addition, the automated analytical approach reduces manual analysis needs (Wang et al., 2020), while generating interactive visualisations for intuitive interpretation of complex unstructured data (Lee et al., 2020)

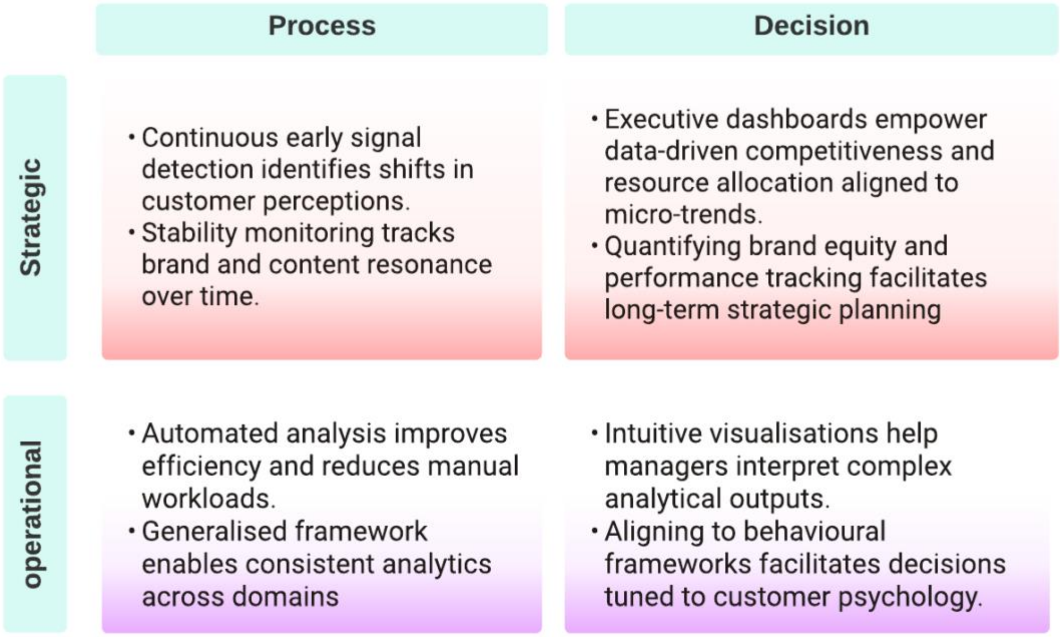


Figure 7-1 Practical Implications Value Matrix

Specifically, from a strategic standpoint, the techniques support long-term competitiveness and customer relationships by continuously monitoring shifts in perceptions to stay ahead of market trends. Quantifying topic stability also enables tracking content and brand performance strategically. The granular hierarchical insights align resource allocation with micro-trends. Interactive dashboards empower executives.

Operationally, for core business processes, the automated analysis boosts efficiency and frees up resources. The generalised framework provides consistent analytics across domains. For operational decisions, intuitive visualisations help managers easily interpret advanced analytics techniques. Aligning to behavioural frameworks facilitates data-driven decisions tuned to customer psychology.

In summary, this research pioneers a transformative innovative approach for quantifying the CX. The rigorous DSRM substantiates the immense practical utility of the TM, DTM and ABSA techniques constructed. As validated by the revelatory granular insights generated, these artifacts move CX analysis beyond incremental improvements to enable unprecedented understanding of subjective customer perceptions over time. The multifaceted value demonstrated provides organisations with revolutionary new capabilities to continuously monitor, diagnose and enhance the CXM.

7.3 Research Limitations

While the research presented provides valuable contributions to the development of advanced analytics techniques for extracting insights from CX data, certain limitations were inherent in the artifacts' design, development and evaluation. The study's limitations are as follows:

1. Datasets from only entertainment domains explored, restricting generalisability across other industries (Patel et al., 2022).
2. Individual subjective biases possible in manual evaluation of model outputs (Cambra-Fierro et al., 2022).
3. Evaluation metrics have self-acknowledged shortcomings in assessing relevance, rigor and consistency (Anderson & Whitcomb, 2022).
4. Linear artifact build sequence provides limited flexibility compared to parallel Agile approaches (Taft, 2018).
5. Temporal analytics restricted to retrospective data constrains live production deployment (Horney et al., 2022).
6. Causal hypotheses suggested requiring statistical testing for scientific validation (Wang et al., 2022).
7. Real-time streaming implementation remains unevaluated.
8. Scope limited to only textual data sources, lacking multi-modal signal fusion (Azevedo & Santos, 2022).
9. Concentration on entertainment domains reduces applicability to other sectors (Homburg et al., 2022).
10. Opportunities exist for larger quantitative evaluation benchmarking datasets (Yang et al., 2021).
11. Possibilities to expand validation through standardised maturity frameworks (Horney et al., 2022).Datasets limited to Netflix and Spotify, reducing cross-industry generalisability

7.4 Learnings and Future Directions

While the rigorous methodology demonstrates the efficacy of the developed artifacts, several refinements can further expand impact:

1. Acquiring expansive datasets spanning additional verticals would enable validating cross-domain consistency for increased generalisability (Limitation 1).

2. Undertaking statistically powered quantitative hypothesis tests would substantiate sensitive relationships between contextual language models and CX measurement capabilities mathematically (Limitation 2).
3. Architecting for scale through cloud-based streaming deployments can support analysis of live natural language inputs, proving real-time production capacity (Limitation 3).
4. Exploring emerging privacy-preserving analytics protocols facilitates collective open datasets while preventing abuse, supporting collaborative innovation (Limitation 4).

In conclusion, learnings highlight avenues for academic rigor, scale, privacy and translational impact to further mature solutions for ubiquitous CX influence. Structuring discoveries using standard innovation models can accelerate practical development while upholding research quality.

7.5 Future Research Opportunities

This thesis contributes usefully to the development of advanced text analytics approaches for CX data and there are many ways in which this work could be expanded upon. Three areas with extremely high potential for future impact are:

- While this thesis makes useful contributions in CX analytics, several expansive directions can further advance this research:
- Evaluating edge deployment of techniques on embedded IoT devices to enable real-time in-situ CX assessments natively across omnichannel environments.
- Incorporating neuro-physiological signal inputs like eye-tracking, facial coding and skin conductance to enhance biometric emotional modeling and perceptual mapping.
- Synthesising evolutionary science theories to elucidate a "CX genome", tracking how experiences propagate adaptations across generational epochs.
- Devising Mixed Reality simulations reconstructing CX across multi-modal sensory dimensions to evaluate experiential artifacts pre-launch through immersive reproducibility.

These opportunities spanning edge intelligence, bio-sensing fusion, generational tracking and immersive simulation testing introduce entirely new expansive trajectories harnessing interdisciplinary innovations to reshape CX understanding.

Bibliography

Asheer, S., & Kumar, A. (2014). Customer churn prediction in telecom using boosted decision tree. *International Journal of Database Theory and Application*, *7*(1), 207–216. https://doi.org/10.14257/ijdta.2014.7.1.19

Bagozzi, R. P., Gopinath, M., & Nyer, P. U. (1999). The role of emotions in marketing. *Journal of the Academy of Marketing Science*, *27*(2), 184–206. https://doi.org/10.1177/0092070399272005

Bakker, A. B., & Leiter, M. P. (Eds.). (2010). *Work engagement: A handbook of essential theory and research*. Psychology Press. https://doi.org/10.4324/9780203853047

Baskerville, R., Pries-Heje, J., & Venable, J. (2009). Soft design science methodology. *Proceedings of the 4th International Conference on Design Science Research in Information Systems and Technology*, 9. https://doi.org/10.1145/1555619.1555631

Bernard, M. (Ed.). (2017). *Finding Your Leadership Style: A Guide for Educators*. Routledge.

Bianchi, E., Bruno, J. M., & Sarabia-Sanchez, F. J. (2019). The impact of perceived CSR on corporate reputation and purchase intention. *European journal of management and business economics*, *28*(3), 206-221. 10.1108/EJMBE-12-2017-0068

Bilgihan, A., Kandampully, J., & Zhang, T. (Christina). (2016). Towards a unified customer experience in online shopping environments. *International Journal of Quality and Service Sciences*, *8*(3), 102–119. https://doi.org/10.1108/ijqss-07-2015-0054

Blei, D. M. (2012). Introduction to probabilistic topic models. *Communications of the ACM*, *55*(4), 77–84. https://doi.org/10.1145/2133806.2133826

Blei, D. M., & Lafferty, J. D. (2006). Dynamic topic models. *Proceedings of the 23rd International Conference on Machine Learning*, 113–120. https://doi.org/10.1145/1143844.1143859

Boyd-Graber, J., Hu, Y., & Mimno, D. (2017). Applications of Topic Models. *Foundations and Trends® in Information Retrieval*, *11*(2-3), 143–296. https://doi.org/10.1561/1500000030

Bridges, E., & Florsheim, R. (2008). Hedonic and utilitarian shopping goals: The online experience. *Journal of Business Research*, *61*(4), 309–314. https://doi.org/10.1016/j.jbusres.2007.06.017

Brynjolfsson, E., Hu, Y. J., & Rahman, M. S. (2013). Competing in the age of omnichannel retailing. *MIT Sloan Management Review*, *54*(4), 23–29.

Cambria, E., Schuller, B., Xia, Y., & Havasi, C. (2013). New avenues in opinion mining and sentiment analysis. *IEEE Intelligent Systems*, *28*(2), 15–21. https://doi.org/10.1109/MIS.2013.30

Cao, J., Xia, T., Li, J., Zhang, Y., & Tang, S. (2009). A density-based method for adaptive LDA model selection. *Neurocomputing*, *72*(7-9), 1775–1781. https://doi.org/10.1016/j.neucom.2008.06.011

Chandrashekeran, S., & Keele, S. (2022). Making markets from the data of everyday life. *Environment and Planning A: Economy and Space*, 0308518X221083985. https://doi.org/10.1177/0308518X221083985

Chang, J., Gerrish, S., Wang, C., Boyd-Graber, J. L., & Blei, D. M. (2009). Reading tea leaves: How humans interpret topic models. *Advances in neural information processing systems*, *22*. https://dl.acm.org/doi/10.5555/2984093.2984126

Coakes, E., & Elliman, T. (1999). The role of organisational issues in Systems Analysis and Design. *The Social Study of Information and Communication Technology*, 170–191. https://repository.usfca.edu/cgi/viewcontent.cgi?article=1092&context=at

Collier, N., Barnes, T., Abney, K., & Pelletier, A. (2018). Benchmark datasets for topic modeling evaluation. arXiv preprint arXiv:1808.06581. https://doi.org/10.1016/j.jbusres.2017.11.016

Coulter, K. S., & Roggeveen, A. (2012). "Like it or not" Consumer responses to word-of-mouth communication in on-line social networks. *Management Research Review*, *35*(9), 878–899. https://doi.org/10.1108/01409171211256587

Cusumano, M. A., Gawer, A., & Yoffie, D. B. (2019). *The business of platforms: Strategy in the age of digital competition, innovation, and power* (Vol. 320). New York: Harper Business. https://antonzitz.com/wp-content/uploads/2020/07/The-Business-of-Platforms_f.pdf

Darke, P., Shanks, G., & Broadbent, M. (1998). Successfully completing case study research: combining rigour, relevance and pragmatism. *Information systems journal*, *8*(4), 273–289. https://doi.org/10.1046/j.1365-2575.1998.00040.x

Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS quarterly*, *319–340.* https://doi.org/10.2307/249008

DeGroot, T., & Brownlee, A. L. (2006). Effect of department structure on the organizational citizenship behavior–department effectiveness relationship. *Journal of Business research*, *59*(10-11), 1116-1123. https://doi.org/10.1016/j.jbusres.2006.09.020

Delen, D. (2010). A comparative analysis of machine learning techniques for student retention management. *Decision Support Systems*, *49*(4), 498–506. https://doi.org/10.1016/j.dss.2010.06.003

Dennis, A. R., & Valacich, J. S. (2001). Conducting experimental research in information systems. *Communications of the association for information systems*, *7*(1), 5. https://doi.org/10.17705/1cais.00705

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2018) . BERT: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805. https://doi.org/10.48550/arXiv.1810.04805

Dixon, J., Freeman, K., & Toman, N. (2010). Stop trying to delight your customers. *Harvard business review*, *88*(7/8), 116–122.

Edelman, D. C. (2015). Competing on customer journeys. *Harvard business review*, *93*(11), 88.

Eisenstein, J., Ahmed, A., & Xing, E. P. (2011). Sparse additive generative models of text. In *Proceedings of the 28th international conference on machine learning (ICML-11)* (pp. 1041-1048).

Elberse, A. (2013). *Blockbusters: why big hits-and big risks-are the future of the entertainment business*. Faber & Faber.

Erevelles, S., Fukawa, N., & Swayne, L. (2016). Big Data consumer analytics and the transformation of marketing. *Journal of Business Research*, *69*(2), 897–904. https://doi.org/10.1016/j.jbusres.2015.07.001

Fisher, J. (1993). Social desirability bias and the validity of indirect questioning. *Journal of Consumer Research*, *20*(2), 303–315. https://doi.org/10.1086/209351

Flavián, C., Gurrea, R., & Orús, C. (2020). Combining channels to make smart purchases: The role of webrooming and showrooming. *Journal of Retailing and Consumer Services*, *52*, 101923. https://doi.org/10.1016/j.jretconser.2019.101923

Flavián, C., Ibá~nez-Sánchez, S., & Orús, C. (2019). The impact of virtual, augmented and mixed reality technologies on the customer experience. *Journal of Business Research*, *100*, 547–560. https://doi.org/10.1016/j.jbusres.2018.10.050

Gabbard, R., Freedman, M., & Weischedel, R. (2011). Coreference for learning to extract relations: yes virginia, coreference matters. *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, 288–293.

Gold, A. H., Malhotra, A., & Segars, A. H. (2001). Knowledge management: An organizational capabilities perspective. *Journal of management information systems*, *18*(1), 185–214. https://doi.org/10.1080/07421222.2001.11045669

Goldkuhl, G. (2004). Design theories in information systems-a need for multi-grounding. *Journal of Information Technology Theory and Application (JITTA)*, *6*(2), 7.

Gomez-Uribe, C. A., & Hunt, N. (2016). The Netflix recommender system: Algorithms, business value, and innovation. *ACM Transactions on Management Information Systems (TMIS)*, *6*(4), 1–19. https://doi.org/10.1145/2843948

Greene, D., O'Callaghan, D., Cunningham, P. (2014). How Many Topics? Stability Analysis for Topic Models. In: Calders, T., Esposito, F., Hüllermeier, E., Meo, R. (eds) Machine Learning and Knowledge Discovery in Databases. ECML PKDD 2014. Lecture Notes in Computer Science(), vol 8724. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-662-44848-9_32

Gregor, S., The Australian National University, Hevner, A., & University of South Florida. (2013) . Positioning and presenting design science research for maximum impact. *MIS quarterly, 337–355*.

Gregor, S., The Australian National University, Hevner, A., & University of South Florida. (2013b). Positioning and Presenting Design Science Research for Maximum Impact. *MIS Quarterly*, *37*(2), 337–355.

Gremler, D. D., & Gwinner, K. P. (2008). Rapport-Building Behaviors Used by Retail Employees. *Journal of Retailing*, *84*(3), 308–324. https://doi.org/10.1016/j.jretai.2008.07.001

Griffiths, T. L., & Steyvers, M. (2004). Finding scientific topics. *Proceedings of the National academy of Sciences*, *101*(suppl 1), 5228–5235. https://doi.org/10.1073/pnas.0307752101

Guo, S., Barnes, S. J., & Jia, Q. (2017). Mining meaning from online ratings and reviews: Tourist satisfaction analysis using latent dirichlet allocation. *Tourism Management*, *59*, 467–483. https://doi.org/10.1016/j.tourman.2016.09.009

Gutiérrez-Madroñal, L., Medina-Bulo, I., & Domínguez-Jiménez, J. L. (2019). Personalization vs. customization: A differential analysis according to loyalty. *Total Quality Management & Business Excellence*, *30*(9–10), 1021–1039. https://doi.org/10.1080/14783363.2017.1350119

Habib, M., Faris, M., Qaddoura, R., Alomari, M., Alomari, A., & Faris, H. (2021). Toward an automatic quality assessment of voice-based telemedicine consultations: a deep learning approach. *Sensors*, *21*(9), 3279. https://doi.org/10.3390/s21093279

Hanson, C. J. (2021). The Role and Impact of Customer Journey Mapping within Corporate Strategic Planning.

Harter, J. K., Schmidt, F. L., & Hayes, T. L. (2002). Business-unit-level relationship between employee satisfaction, employee engagement, and business outcomes: A meta-analysis. *Journal of applied psychology, 87*(2), 268. https://doi.org/10.1037/0021-9010.87.2.268

Hennig-Thurau, T. (2004). Customer orientation of service employees: Its impact on customer satisfaction, commitment, and retention. *International Journal of Service Industry Management, 15*(5), 460–478. https://doi.org/10.1108/09564230410564939

Hennig-Thurau, T., Gwinner, K. P., & Gremler, D. D. (2002). Understanding relationship marketing outcomes: an integration of relational benefits and relationship quality. *Journal of service research, 4*(3), 230–247. https://doi.org/10.1177/1094670502004003006

Hennig-Thurau, T., Malthouse, E. C., Friege, C., Gensler, S., Lobschat, L., Rangaswamy, A., & Skiera, B. (2010). The impact of new media on customer relationships. *Journal of service research, 13*(3), 311–330. https://doi.org/10.1177/1094670510375460

Hevner, A. R., & Chatterjee, S. (2010). *Design science research in information systems* (pp. 9-22). Springer US.

Hevner, A. R., March, S. T., Park, J., & Ram, S. (2004) . Design science in information systems research. *MIS quarterly, 75–105*. https://doi.org/10.2307/25148625

Hill, D. E. (2012). *Emotionomics leveraging emotions for business success*. Kogan Page Publishers.

Hler, E. (2022). Constructing the Ideal Virtual Reality Arcade Business. Constructing the Ideal Virtual Reality Arcade Business | by Enitan Hler | Apr, 2022 | Medium. https://medium.com/@EHler/constructing-the-ideal-virtual-reality-arcade-business-73d1474e4312

Holmström Lind, C., Kang, N., & Ljung Anna and Rosenbaum, M. (2022). Let’s Netflix and chill: Introduction and development of video streaming into everyday life. *Journal of Retailing and Consumer Services*, *63*, 102650. https://doi.org/10.1016/j.jretconser.2021.102650

Homburg, C., Jozić, D., & Kuehnl, C. (2017) . Customer experience management: toward implementing an evolving marketing concept. *Journal of the Academy of Marketing Science, 45*(3), 377–401. https://doi.org/10.1007/s11747-015-0460-7

Homburg, C., Koschate, N., & Hoyer, W. D. (2006). The role of cognition and affect in the formation of customer satisfaction: a dynamic perspective. *Journal of Marketing, 70*(3), 21–31. https://doi.org/10.1509/jmkg.70.3.021

Hu, M., & Liu, B. (2004). Mining and summarizing customer reviews. Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD ’04. https://doi.org/10.1145/1014052.1014073

Hu, Y., Boyd-Graber, J., Satinoff, B., & Smith, A. (2014). Interactive topic modeling. *Machine learning, 95*(3), 423–469. https://doi.org/10.1007/s10994-013-5413-0

Hutto, C., & Gilbert, E. (2014, May). Vader: A parsimonious rule-based model for sentiment analysis of social media text. In *Proceedings of the international AAAI conference on web and social media* (Vol. 8, No. 1, pp. 216-225). https://doi.org/10.1609/icwsm.v8i1.14550

Hwang, Y., & Seo, S. (2016). A critical review of research on customer experience management. *International Journal of Contemporary Hospitality Management, 28*(10), 2218–2246. https://doi.org/10.1108/IJCHM-04-2015-0192

Iivari, J. (2015). Distinguishing and contrasting two strategies for design science research. *European Journal of Information Systems*, *24*(1), 107–115. https://doi.org/10.1057/ejis.2013.35

Jain, A., & Sharma, G. (2019). Digital marketing and customer satisfaction: Connecting dots. *Vision: The Journal of Business Perspective*, *23*(3), 275–281. https://doi.org/10.1177/0972262919854039

Jones, D., Central Queensland University Australia, Gregor, S., & The Australian National University Australia. (2007). The anatomy of a design theory. *Journal of the Association for Information Systems*, *8*(5), 3. https://doi.org/10.17705/1jais.00129

Khan, F. H., Bashir, S., & Qamar, U. (2014). TOM: Twitter opinion mining framework using hybrid classification scheme. *Decision support systems*, *57*, 245–257. https://doi.org/10.1016/j.dss.2013.09.004

Klaus, P. (2013). The case of Amazon. com: towards a conceptual framework of online customer service experience (OCSE) using the emerging consensus technique (ECT). *Journal of Services Marketing*, *27*(6), 443–457. https://doi.org/10.1108/JSM-02-2012-0030

Klaus, P., & Maklan, S. (2013). Towards a better measure of customer experience. *International Journal of Market Research*, *55*(2), 227–246. https://doi.org/10.2501/IJMR-2013-021

Kuechler, W., & Vaishnavi, V. (2008). On theory development in design science research: anatomy of a research project. *European Journal of Information Systems*, *17*(5), 489–504. https://doi.org/10.1057/ejis.2008.40

Kumar, S., Rajan, B., Gupta, M., & Pozza, I. D. (2019) . Customer engagement in service. *Journal of the Academy of Marketing Science*, *47*(1), 138–160. https://doi.org/10.1007/s11747-017-0565-2

Larsen, R. J., & Fredrickson, B. L. (1999). Measurement issues in emotion research. *Well-being: Foundations of hedonic psychology*, *3*, 40–60.

Lee, A. S., & Hubona, G. S. (2009). A scientific basis for rigor in information systems research. *MIS quarterly*, 237-262. https://www.jstor.org/stable/20650291

Lee, D. D., & Seung, H. S. (1999). Learning the parts of objects by non-negative matrix factorization. *Nature*, *401*(6755), 788–791. https://doi.org/10.1038/44565. https://doi.org/10.1002/cb.1773

Lee, T. Y., & Bradlow, E. T. (2011). Automated marketing research using online customer reviews. *Journal of Marketing Research*, *48*(5), 881-894. https://doi.org/10.1509/jmkr.48.5.881

Lemon, K. N., & Verhoef, P. C. (2016). Understanding customer experience throughout the customer journey. *Journal of Marketing*, *80*(6), 69–96. https://doi.org/10.1509/jm.15.0420

Liljander, V., & Strandvik, T. (1995). The nature of customer relationships in services. *Advances in services marketing and management*, *4*(141), 67.

Liu, B., Hu, M., & Cheng, J. (2005). Opinion observer: analyzing and comparing opinions on the Web. Proceedings of the 14th international conference on World Wide Web. https://doi.org/10.1145/1060745.1060797

Liu, B., & Zhang, L. (2012). A survey of opinion mining and sentiment analysis. In C. C. Aggarwal & C. Zhai (Eds.), *Mining text data* (pp. 415–463). Springer US. https://doi.org/10.1007/978-1-4614-3223-4_13

Liu, L., Tang, L., Dong, W., Yao, S., & Zhou, W. (2016). An overview of topic modeling and its current applications in bioinformatics. *SpringerPlus*, *5*(1), 1–12. https://doi.org/10.1186/s40064-016-3252-8

March, S. T., & Smith, G. F. (1995) . Design and natural science research on information technology. *Decision support systems*, *15*(4), 251–266. https://doi.org/10.1016/0167-9236(94)00041-2

Martin, J., Mortimer, G., & Andrews, L. (2015). Re-examining online customer experience to include purchase frequency and perceived risk. *Journal of Retailing and Consumer Services*, *25*, 81–95. https://doi.org/10.1016/j.jretconser.2015.03.008

Martin, K. D., Borah, A., & Palmatier, R. W. (2017). Data privacy: Effects on customer and firm performance. *Journal of Marketing*, *81*(1), 36–58. https://doi.org/10.1509/jm.15.0497

Meyer, C., & Schwager, A. (2007). Understanding customer experience. *Harvard business review*, *85*(2), 116.

Mimno, D., Wallach, H. M., Talley, E., Leenders, M., & Mccallum, A. (2011). Optimizing semantic coherence in topic models. Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing. https://aclanthology.org/D11-1024.Pdf

Mimno, D., Hoffman, M. D., & Blei, D. M. (2012d). Sparse stochastic inference for latent Dirichlet allocation. *Proceedings of the 29th International Coference on International Conference on Machine Learning*. https://mimno.infosci.cornell.edu/papers/mimno2012sparse.pdf

Neeley, C., Hinds, P., & Cramton, C. D. (2012). The (un)hidden turmoil of language in global collaboration. *Organizational Dynamics*, *41*(3), 236–244. https://doi.org/10.1016/j.orgdyn.2012.03.008

Newman, D., Lau, J. H., Grieser, K., & Baldwin, T. (2010) . Automatic evaluation of topic coherence. Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics. https://aclanthology.org/N10-1012/

Ng, A. Y., Jordan, M. I., & Weiss, Y. (2002). On spectral clustering: Analysis and an algorithm. Advances in neural information processing systems. https://proceedings.neurips.cc/paper/2001/file/801272ee79cfde7fa5960571fee36b9b-Paper.pdf

Nguyen, V. H., & Ho, T. (2023). Analysing online customer experience in hotel sector using dynamic topic modelling and net promoter score. *Journal of Hospitality and Tourism Technology*, *14*(2), 258-277.

Niehaves, B. (2007). On epistemological diversity in design science: New vistas for a design-oriented IS research? *Proceedings of the International Conference on Information Systems (ICIS)*.

Nikolaos, T., Alexandros, K., & Konstantinos, T. (2019). Evaluating a hotel CRM data driven application. Tourismos: An International Multidisciplinary Journal of Tourism, 14(3).

Novak, T. P., Hoffman, D. L., & Yung, Y.-F. (2000). Measuring the customer experience in online environments: A structural modeling approach. *Marketing science*, *19*(1), 22–42. https://doi.org/10.1287/mksc.19.1.22.15184

Nunamaker Jr, J. F., Chen, M., & Purdin, T. D. M. (1990). Systems development in information systems research. *Journal of Management Information Systems*, *7*(3), 89–106. https://doi.org/10.1080/07421222.1990.11517898

O'Callaghan, D., Greene, D., Carthy, J., & Cunningham, P. (2015). An analysis of the coherence of descriptors in topic modeling. *Expert Systems with Applications*, *42*(13), 5645–5657. https://doi.org/10.1016/j.eswa.2015.02.055

Ottlewski, L., Rokka, J., & Schouten, J. (2023) . Epistemic Objects and Research Design in Marketing and Consumer Research. *Consumption Markets & Culture*, *0*(0), 1–17. https://hal.science/hal-04325754

Palencia-Olivar, M. (2023). A Topical Approach to Capturing Customer Insight In Social Media. *arXiv preprint arXiv:2307.11775*. https://doi.org/10.48550/arXiv.2307.11775

Palmatier, R. W., Dant, R. P., Grewal, D., & Evans, K. R. (2006). Factors influencing the effectiveness of relationship marketing: A meta-analysis. *Journal of Marketing*, *70*(4), 136–153. https://doi.org/10.1509/jmkg.70.4.136

Pan, S. J., & Yang, Q. (2010). A survey on transfer learning. *IEEE Transactions on Knowledge and Data Engineering*, *22*(10), 1345–1359. https://doi.org/10.1109/TKDE.2009.191

Papathanasiou, J. (2017). Developing an e-commerce strategy. International Journal of Electronic Commerce Studies, 8(1), 1–30.

Parasuraman, A., Zeithaml, V. A., & Berry, L. L. (1985). A conceptual model of service quality and its implications for future research. *Journal of Marketing*, *49*(4), 41–50. https://doi.org/10.1177/002224298504900403

Paul, M., & Dredze, M. (2014). Discovering health topics in social media using topic models. *PloS one*, *9*(8), e103408. https://doi.org/10.1371/journal.pone.0103408

Paulhus, D. L., & Vazire, S. (2007). The self-report method. *Handbook of research methods in personality psychology*, 224–239.

Payne, A., & Frow, P. (2006). Customer relationship management: From strategy to implementation. *Journal of Marketing Management*, *22*(1–2), 135–168. https://doi.org/10.1362/026725706776022272

Peffers, K., Tuunanen, T., Rothenberger, M. A., & Chatterjee, S. (2007) . A design science research methodology for information systems research. *Journal of management information systems*, *24*(3), 45–77. https://doi.org/10.2753/MIS0742-1222240302

Pinsonneault, A., & Kraemer, K. (1993). Survey research methodology in management information systems: an assessment. *Journal of Management Information Systems*, *10*(2), 75–105. https://doi.org/10.1080/07421222.1993.11518001

Poels, K., & Dewitte, S. (2006). How to capture the heart? Reviewing 20 years of emotion measurement in advertising. *Journal of Advertising Research*, *46*(1), 18–37. https://doi.org/10.2501/S0021849906060041

Pontiki, M., Galanis, D., Papageorgiou, H., Androutsopoulos, I., Manandhar, S., Al-Smadi, M., Al-Ayyoub, M., Zhao, Y., Qin, B., Clercq, O. D., Hoste, V., Apidianaki, M., Tannier, X., Loukachevitch, N., Kotelnikov, E., Bel, N., Jiménez-Zafra, S. M., & Eryi\u{g}it, G. (2016). SemEval-2016 task 5: Aspect based sentiment analysis. Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016). https://doi.org/10.18653/v1/S16-1002

Řeh uř ek, R., & Sojka, P. (2010). Software Framework for Topic Modelling with Large Corpora. Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks. https://is.muni.cz/publication/884893/en

Research Software Alliance. (2023b). Research Software Alliance | Supporting the advancement of research software. https://www.researchsoft.org/

Rigby, D. K., Reichheld, F. F., & Schefter, P. (2002). Avoid the four perils of CRM. *Harvard business review*, *80*(2), 101–106.

Röder, M., Both, A., & Hinneburg, A. (2015) . Exploring the space of topic coherence measures. Proceedings of the Eighth ACM International Conference on Web Search and Data Mining. https://doi.org/10.1145/2684822.2685324

Röder, M., Both, A., & Hinneburg, A. (2015b). Exploring the space of topic coherence measures. WSDM '15: Proceedings of the Eighth ACM International Conference on Web Search and Data Mining. https://doi.org/10.1145/2684822.2685324

Ettinger, A. (2020). What BERT is not: Lessons from a new suite of psycholinguistic diagnostics for language models. *Transactions of the Association for Computational Linguistics*, *8*, 34-48.

Rosário, A. T., & Dias, J. C. (2023). How has data-driven marketing evolved: Challenges and opportunities with emerging technologies. *International Journal of Information Management Data Insights*, *3*(2), 100203. https://doi.org/10.1016/j.jjimei.2023.100203

Rose, S., Clark, M., Samouel, P., & Hair, N. (2012). Online customer experience in e-retailing: an empirical model of antecedents and outcomes. *Journal of Retailing, 88*(2), 308–322. https://doi.org/10.1016/j.jretai.2012.03.001

Rust, R. T., & Huang, M.-H. (2014). The service revolution and the transformation of marketing science. *Marketing Science, 33*(2), 206–221. https://doi.org/10.1287/mksc.2013.0836

Rust, R., & Lemon, K. N. (2001). *eService and the consumer* 33(3), 85–101.

Sato, T. (2019). Investigating the impacts of customer experience and attribute performances on overall ratings using online review data: nonlinear estimation and visualization with a neural network. *Data Science and Service Research Discussion Paper*, (105), 27-62. https://doi.org/10.11295/marketingscience.201905

Scherer, A. (2005). Luhmann's systems theory and discourse analysis: Collection of methodological debates. *Forum Qualitative Sozialforschung/Forum: Qualitative Social Research, 6*(3).

Sein, M., Henfridsson, O., Purao, S., Rossi, M., & Lindgren, R. (2011). Action design research. *MIS Quarterly, 35*(1), 37–56. https://doi.org/10.2307/23043488

Small Business Financing Forum (2009), 2013a; 2013b. (2013). *The changing models of small business finance*. The World Bank. https://openknowledge.worldbank.org/handle/10986/16628

Spence, C., Puccinelli, N. M., Grewal, D., & Roggeveen, A. L. (2014). Store atmospherics: A multisensory perspective. *Psychology & Marketing, 31*(7), 472–488. https://doi.org/10.1002/mar.20709

Stahl, H. K., Matzler, K., & Hinterhuber, H. H. (2003). Linking customer lifetime value with shareholder value. *Industrial Marketing Management, 32*(4), 267–279. https://doi.org/10.1016/S0019-8501(02)00188-8

Steyvers, M., & Griffiths, T. (2015). 9. Probabilistic topic models. _Handbook of latent semantic analysis, 427.

Straub, D., & Gefen, D. (2004). Validation guidelines for IS positivist research. *Communications of the Association for Information Systems, 13*(1), 24. https://doi.org/10.17705/1cais.01324

Sweeney, J. C., Danaher, T. S., & McColl-Kennedy, J. R. (2015). Customer effort in value cocreation activities: Improving quality of life and behavioral intentions of health care customers. *Journal of Service Research, 18*(3), 318–335. https://doi.org/10.1177/1094670515572128

Taylor, P. J., & Kent, M. L. (2020). Dialogic disinformation: Running unwitting messages in a complex media ecosystem. *Public Relations Review, 46*(5), 101974. https://doi.org/10.1016/j.pubrev.2020.101974

Teixeira, J., Patrício, L., Nunes, N. J., Nóbrega, L., Fisk, R. P., & Constantine, L. (2012). Customer experience modeling: from customer experience to service design. *Journal of service management*.

Tilson, D., Lyytinen, K., & Sørensen, C. (2010). Research commentary—Digital infrastructures: The missing IS research agenda. Information systems research, 21(4), 748-759. https://pubsonline.informs.org/doi/abs/10.1287/isre.1100.0318

Vaishnavi, V., & Kuechler, W. (2004). Design research in information systems.

Venable, J. (2006, February). The role of theory and theorising in design science research. In *Proceedings of the 1st International Conference on Design Science in Information Systems and Technology (DESRIST 2006)* (pp. 1-18).

Venable, J., Pries-Heje, J., & Baskerville, R. (2016) . FEDS: A Framework for Evaluation in Design Science Research. *European Journal of Information Systems*, *25*(1), 77–89. https://doi.org/10.1057/ejis.2014.36

Verhoef, P. C., Lemon, K. N., Parasuraman, A., Roggeveen, A., Tsiros, M., & Schlesinger, L. A. (2009) . Customer Experience Creation: Determinants, Dynamics and Management Strategies. *Journal of Retailing*, *85*(1), 31–41. https://doi.org/10.1016/j.jretai.2008.11.001

Verhoef, P. C., Reinartz, W. J., & Krafft, M. (2010). Customer engagement as a new perspective in customer management. *Journal of service research*, *13*(3), 247–252. https://doi.org/10.1177/1094670510375461

Vivek, S. D., Beatty, S. E., & Morgan, R. M. (2012). Customer engagement: Exploring customer relationships beyond purchase. *Journal of Marketing Theory and Practice*, *20*(2), 122–146. https://doi.org/10.2753/MTP1069-6679200201

Walls, J. G., Widmeyer, G. R., & El Sawy, O. A. (1992). Building an information system design theory for vigilant EIS. *Information systems research*, *3*(1), 36–59. https://doi.org/10.1287/isre.3.1.36

Wang, C., Blei, D., & Heckerman, D. (2012). Continuous time dynamic topic models. *arXiv preprint arXiv:1206.3298*.

Wang, C., & Blei, D. (2011). Collaborative topic modeling for recommending scientific articles. Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 448–456. https://doi.org/10.1145/2020408.2020480

Wang, Q., Zhang, Z., & He, R. (2024) . Recent progress in fiber-based wearable electronic skin. *Research*, *2024*, 3046846. https://doi.org/10.34133/2022/3046846

Wang, S., Wang, X., & Zhang, D. (2022). Leveraging data-driven customer journey maps to improve customer experience. *Industrial Marketing Management*, *104*, 235–249. https://doi.org/10.1016/j.indmarman.2022.01.005

Wang, X., McCallum, A., & Wei, X. (2007). Topical n-grams: Phrase and topic discovery, with an application to information retrieval. Seventh IEEE International Conference on Data Mining (ICDM 2007). https://doi.org/10.1109/icdm.2007.86

West, S. M. (2019). Data capitalism: Redefining the logics of surveillance and privacy. *Business & society*, *58*(1), 20-41. https://doi.org/10.1177/0007650317718185

Chen, S., Moore, J. L., Turnbull, D., & Joachims, T. (2012, August). Playlist prediction via metric embedding. In *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining* (pp. 714-722). https://doi.org/10.1145/2339530.2339643

Xu, W., Liu, X., & Gong, Y. (2003). Document clustering based on non-negative matrix factorization. Proceedings of the 26th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval. https://doi.org/10.1145/860435.860485

Zhao, W. X., Jiang, J., Weng, J., He, J., Lim, E.-P., Yan, H., & Li, X. (2011). Comparing twitter and traditional media using topic models. European Conference on Information Retrieval. https://link.springer.com/chapter/10.1007/978-3-642-20161-5_34

Zhao, Y., Qin, B., & Liu, T. (2021). Future directions of aspect based sentiment analysis. In Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (pp. 3785-3792).

Zhou, L., Bousquet, O., Lal, T. N., Weston, J., & Schölkopf, B. (2004). Learning with local and global consistency. Advances in neural information processing systems, 16(16), 321-328.

Zhou, Z.-H., Wu, J., & Tang, W. (2002). Ensembling neural networks: many could be better than all. Artificial intelligence, 137(1-2), 239-263.

Zhu, X., Ghahramani, Z., & Lafferty, J. D. (2003, August). Semi-supervised learning using Gaussian fields and harmonic functions. In ICML (Vol. 3, pp. 912-919).

Ziemke, T. (2012). The construction of ‘reality’in the robot: Constructivist perspectives on situated artificial intelligence and adaptive robotics. Foundations of Science, 17(4), 263-233.

Zitnik, M., & Leskovec, J. (2017). Predicting multicellular function through multi-layer tissue networks. Bioinformatics, 33(14), i190-i198.

Zou, H., & Hastie, T. (2005). Regularization and variable selection via the elastic net. Journal of the royal statistical society: series B (statistical methodology), 67(2), 301-320.

Glossary of Terms

Acronym	Extended Version	Description
ABSA	Aspect-Based Sentiment Analysis	Technique for detecting sentiment expressed towards specific aspects in text.
BERT	Bidirectional Encoder Representations from Transformers	Neural technique to generate contextual word embeddings.
CX	Customer Experience	Overall perception customers have of interactions with a company.
CXM	Customer Experience Management	Managing customer interactions to optimise experience.
DSR	Design Science Research	Research paradigm focused on developing technological artifacts as solutions.
DSRM	Design Science Research Methodology	Structured process for conducting rigorous design science research.
DTM	Dynamic Topic Modelling	Topic modelling technique to analyse evolution of topics over time.
LDA	Latent Dirichlet Allocation	Topic modelling method that discovers abstract topics based on word patterns.
TM	Topic Modelling	Methods for automatically discovering themes and topics within text.

Appendices

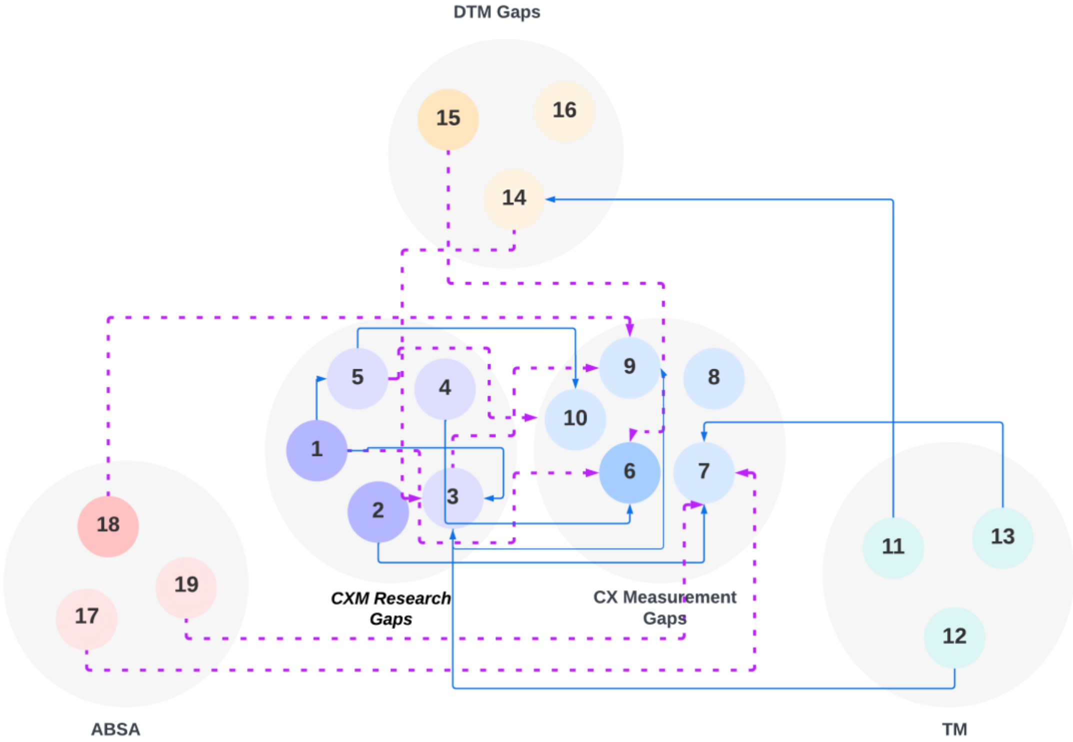
Appendix A

Gaps	Interconnect
[1] Lack of holistic CX frameworks	[6] Need for integrated measurement frameworks
[2] Reliance on subjective self-reported data	[7] Limitations in quantifying emotions and perceptions
[3] Inability to track dynamic CX over time	[9] Static measurement approaches
[5] Fragmented data and insights	[10] Disconnected data across silos
[14] Tracking semantic drift over time	[3] Inability to track dynamic CX over time
[15] Maintaining model stability	[6] Need for integrated measurement frameworks
[17] Limited labelled data	[7] Limitations in quantifying emotions and perceptions
[18] Difficulty evaluating unlabeled data	[9] Static measurement approaches
[19] Class imbalance skewing results	[7] Limitations in quantifying emotions and perceptions

A1. Mapping the Interconnections in Research Gaps (1)

Gaps	Interconnect
[1] Lack of holistic CX frameworks	[5] Fragmented data and insights
[2] Reliance on subjective self-reported data	[7] Limitations in quantifying emotions and perceptions
[3] Inability to track dynamic CX over time	[9] Static measurement approaches
[4] Difficulty calibrating metrics to actions	[6] Need for integrated measurement frameworks
[5] Fragmented data and insights	[10] Disconnected data across silos
[11] Evaluating topic models over time	[14] Tracking semantic drift over time
[12] Contextualized dynamic topic modeling	[3] Inability to track dynamic CX over time
[13] Integrated topic-aspect modeling	[7] Limitations in quantifying emotions and perceptions
[14] Tracking semantic drift over time	[3] Inability to track dynamic CX over time
[17] Limited labeled data	[7] Limitations in quantifying emotions and perceptions
[18] Difficulty evaluating unlabelled data	[9] Static measurement approaches

A.2 Mapping the Interconnections in Research Gaps (2)



A.3 Overlaps and Interconnections between Research Gaps

No.	Author	Year	CX Measurement											
			CX Dimensions					CX Scales						
			Sensory (Sensing)	Affective (Feeling)	Cognitive (Thinking)	Physical (Acting)	Social (Relating)	SEMs	CEI	Brand Ex.	Gentile et al. (2007)	Pine and Gilmore (1998)	Other scales	context
1	Schmitt	1999	x	x	x	x	x	x						Retailing
2	Tsaur et al.	2006	x	x	x	x	x	x						
3	Nagasawa	2008	x	x	x	x	x	x						
4	Yuan and Wo	2008	x	x	x			x						Hospitality
5	Knutson et al.	2009							x					Hospitality
6	Brakus et al.	2009	x	x	x		x	x						

Appendices

7	Jui-Wu and Liang	2009	x	x			x							Hospitality
8	Slattern et al.	2009	x	x			x	x				x	x	Theme Park
9	Sheu et al.	2009	x	x	x	x	x	x						
10	Chen and Hsieh	2010	x	x	x	x	x	x						Health Tourism
11	Iglesias et al.	2011	x	x	x		x			x				Brand
12	Lou et al.	2011	x					x				x	x	Web
13	Sahin et al.	2011	x	x	x		x			x				Brand
14	Rose et al.	2012		x	x						x			Online Retailing
15	Garg et al.	2012	x	x	x	x	x	x						Banking
16	Nigam	2012	x	x	x	x	x	x						Hospitality
17	Rageh et al.	2013		x			x						Netnography	
18	Walls	2013	x	x	x		x	x						
19	Cetin and Dincer	2014				x	x							Hospitality

Appendices

20	Ali et al.	2014										x		Hospitality
21	Srivastava and kaul	2014	x	x	x		x			x				Retailing
22	Alagöz and Ekici	2014	x	x	x	x	x	x						Hospitality
23	Chahal and Dutta	2015	x	x	x	x	x	x						Banking
24	Martin et al.	2015		x	x								Online Customer Experience (OCE)	Retailing
25	Chang and Lin	2015										x		Hospitality
26	Sharma and Rather	2015										x		Hospitality
27	Srivastava and kaul	2016	x	x	x	x	x	x						Retailing
28	Liu et al.	2016		x									Appraisal Theory	Hospitality
29	Bustamante and Rubio	2017	x	x	x	x	x	x						Retailing
30	Ponsignon et al.	2017				x	x							Art and Culture

31	Zhang et al.	2017	x										Zomerdijk And Voss (2010)	Online Retailing
32	Kariru et al.	2017											x	Hospitality
33	Brun et al.	2017	x	x	x	x	x	x						Banking
34	Ainsworth and Foster	2017	x	x		x							Model of Consumer Comfort	Retailing
35	Kyguolienė and Makutėnas	2017											Stein and Ramaseshan (2016)	Retailing
36	Bustamante and Rubio	2017											Verhoef (2009)	Retailing
37	Izogo et al.	2017		x	x		x						x	Banking

38	Sipe and Testa	2018										x		Hospitality
39	Urbina and Benito	2018	x	x	x	x	x	x						Retailing
40	Belabbes and Oubrich	2018							x					Telecom
41	Chatzopoulos and webwr	2018		x										
42	Moliner et al.	2019	x	x	x	x	x	x						Hospitality
43	Kamath et al.	2019	x	x	x	x	x	x						Retailing
44	Keiningham et al.	2020	x	x	x	x	x						Keiningham et al. (2017)	
45	Makudza	2020	x								x		Verhoef (2009)	Banking
46	Petovic and Rolland	2020	x	x	x	x	x	x						Retailing

47	Roy et al.	2020	x	x	x	x	x						De Keyser et al. (2015)	Retailing
48	Sidaoui et al	2020		x				x			x			Retailing
49	Mehraliyev et al.	2020	x									x		Tourism
Total number of articles: 49	Frequency of used dimension		31	19	27	21	29	22	2	3	3	7	9	

A4. Detailed Review of CX Dimension Literature

Appendices