DEVELOPING AN IMPROVED COMPLAINT MANAGEMENT SYSTEM THROUGH THE USE OF TEXT MODELLING TECHNIQUES

Lok Shan Lam

Supervisors: Prof. Alan Wilson & Prof. John Quigley

Submitted in fulfilment of the requirements for the degree of Master of Philosophy (MPhil)

Department of Marketing / Management Science, Strathclyde Business School
University of Strathclyde, Glasgow, UK

December 2023

Declaration

This thesis is the result of the author's original research. It has been composed by the author and has not been previously submitted for examination which has led to the award of a degree.

Signed: Lok Shan Lam

Date: 14th March 2025

Abstract

This thesis investigates the integration of textual complaint data into customer churn prediction (CCP) models, addressing a critical gap in the intersection of service recovery and predictive analytics. While traditional CCP models primarily rely on structured variables, textual complaints encapsulate rich customer sentiments and behavioural signals that remain underutilised in churn modelling. This research, therefore, explores how advanced text representation techniques can enhance CCP models, bridging the gap between service failure recovery theories and data-driven predictive modelling.

Grounded in service recovery and customer relationship management (CRM) theories, this study evaluates a decision-support framework that incorporates textual complaints and structured variables to improve churn prediction. Using real-life customer complaint data from a UK-based data-driven product company, the research benchmarks traditional count-based text representations (e.g., TF-IDF) against modern embedding-based methods (e.g., word embeddings, Transformer models). Additionally, the study investigates data fusion techniques, examining their role in leveraging multimodal information for improved churn prediction.

Key findings reveal that:

1. Incorporating textual complaint data significantly enhances CCP models, confirming the value of textual analysis in understanding customer behaviours.
2. Word embedding models outperform TF-IDF-based models in overall CCP performance, indicating a shift towards more sophisticated text representation techniques.
3. TF-IDF-based models perform better at predicting retained customers, while word embedding models excel in identifying churn instances, underscoring the importance of task-dependent model selection.
4. An ensemble approach combining count-based and latent feature representations improves retained case prediction but slightly decreases churn prediction accuracy, suggesting that model selection should align with specific business objectives.

5. Data fusion techniques play a crucial role in optimizing predictive accuracy, demonstrating the need for well-designed multimodal integration strategies.
6. Structured variables remain essential in CCP models, providing additional insights into customer retention dynamics.

This thesis advances both theoretical and practical understanding of customer churn prediction by demonstrating how textual complaint data can be strategically leveraged to enhance CCP models. It contributes to service recovery and predictive analytics literature by offering a data-driven approach to customer retention strategies, equipping businesses with more effective and intelligent service recovery mechanisms.

List of Figures

List of Tables

List of Acronyms

Adam	Adaptive Moment Estimation
AI	Artificial Intelligence
BERT	Bidirectional Encoder Representations from Transformers
BoW	Bag of Words
CBoW	Continuous Bag-of-Words
CBR	Case-Based Reasoning
CCP	Customer Churn Prediction
CNN	Convolutional Neural Network
CRM	Customer Relationship Management
CSSR	Customer Satisfaction and Service Recovery
DL	Deep Learning
F1-score	F1 Score (Harmonic Mean of Precision and Recall)
LSTM	Long Short-Term Memory
MAG	Multimodal Adaptation Gate
ML	Machine Learning
MLP	Multi-Layer Perceptron
NLP	Natural Language Processing
NSP	Next Sentence Prediction
ReLU	Rectified Linear Unit
RF	Random Forest
RNN	Recurrent Neural Network
TF-IDF	Term Frequency-Inverse Document Frequency
WOM	Word of Mouth

Table of Contents

1. Introduction

1.1 The Significance of Textual Data in Business Analytics and Service Recovery

Unstructured types of data have become more readily available over the past two decades, especially those in the form of texts that have led to a tremendous amount of valuable information requiring to be analysed for extracting business insights (Xiang et al., 2015; Humphreys, 2021). Textual data can come from various sources, for example, user posts can be collected from social media platforms, and customer comments/discussion sharing service experiences can be found on review sites, feeds or message boards (George et al., 2016).

Just as the channels for generating, transmitting and recording texts have grown more diversified and complicated, so too have the approaches for deriving knowledge from textual data, this is because more companies are recognising the great value of textual data and applying the techniques that enable them to extract insights for improving business processes (Xiang et al., 2015; Raguseo, 2018). In the context of marketing, textual data contains knowledge regarding service experience, emotions and feedback that can reflect customers' minds, this knowledge could be used as additional information to complement traditional market research. Despite being a real challenge, companies that succeed in effectively taking advantage of customer-generated texts can better understand the market structure, satisfy customers' needs and recover dissatisfaction, therefore they are more likely to maintain a healthy relationship with customers (McAfee et al., 2012; Humphreys and Wang, 2018).

A particularly textual-data-intensive subfield in marketing is service recovery. Miller, Craighead and Karwan (2000, p.388) described service recovery as a series of actions designed by the service provider *"to resolve problems, alter negative attitudes of dissatisfied customers and to ultimately retain these customers"*. Service failure, complaining behaviour, and service recovery are interrelated elements that significantly influence customer retention and business performance. Service failure refers to a breakdown in delivering the expected level of service, which can stem from product defects, poor customer support, service delays, or billing errors. When customers experience service failure, they may react in several ways: some will silently switch to competitors, while others will voice their dissatisfaction by filing complaints. This complaining behaviour serves as a critical juncture in customer relationship management, as it provides companies with an opportunity to resolve the issue before the customer churns. However, not all customers choose to complain; studies indicate that only a small percentage of dissatisfied customers formally report their grievances, making those who do complain particularly valuable to businesses seeking to improve service quality (Tax, Brown, and Chandrashekaran, 1998).

Once a complaint is lodged, the company's service recovery efforts determine whether the customer remains loyal or ultimately leaves. Effective service recovery involves acknowledging the failure, providing fair compensation, and delivering timely resolutions (Smith, Bolton, and Wagner, 1999). A more detailed discussion of these concepts, supported by empirical studies, is presented in Chapter 2, where theoretical frameworks on service failure, complaining behaviour, and service recovery are examined in depth.

Indeed, companies have clear operational objectives and financial incentives to establish robust complaint handling programs as part of their customer relationship management strategy (Fox, 2008; Pinto and Mansfield, 2012). First, successful companies encourage customer complaining as opportunities to improve. Prior studies have revealed that most consumers can tolerate a service failure, and the failure event alone may not result in much discontent, but ignorance of complaints and a refusal to take recovery measures has been a primary cause of customer dissatisfaction (Bitner et al., 1990; Wirtz and Mattila, 2004). Second, literature regarding customer evaluation of service recovery (Rawls, 1971; Tax et al., 1998; Smith et al., 1999) suggests that the performance of service recovery often tends to be more influential than that of the initial product/service to overall customer satisfaction. Complainants are said to be especially both cognitively and emotionally 'mindful' of the outcomes and procedures of service recovery (Blodgett et al., 1997; McCollough, 2009). Third, an exceptionally outstanding service recovery effort is believed to result in a higher level of overall customer satisfaction as compared to that in which a situation with no service failure occurs, this phenomenon is defined by Etzel and Silverman (1981) as the service recovery paradox. Finally, failure in resolving complaints is found to be the main cause of losing customers (Maxham and Netemeyer, 2002). Successful service recovery can avoid churn, which in turn can increase the customer retention rate that plays a prominent role in earning a good reputation (Kim, Ok, and Canter 2012), establishing a loyal customer base (Miller, Craighead and Karwan 2000), and facilitating repeat business (Bendall-Lyon and Powers, 2001).

However, despite this growing attention, companies are still struggling to put effective service recovery into practice (Cui et al., 2017). In the current era that is characterised by higher volumes, velocities and varieties of data, it is challenging and time-consuming for those who rely heavily on traditional manual approach (i.e., human reading and resolving complaints) to manage customer complaints due to limited human resources available (Maia, Carvalho, Ladeira, Rocha, and Mendes, 2014; HaCohen-Kerner, Dilmon, Hone, and Ben-Basan, 2019). The increasing number of daily complaints requires more efficient handling, and thus, being able to resolve complaints in a swift and effective manner becomes a core competence for companies to retain existing customers (Sultan, Abidin and Abdullah, 2008; Dey, Hariharan, and Ho, 2009; Schwab, 2015).

1.2 Evolution and Challenges in Automated Complaint Management

Some past review papers into practical service recovery/complaint management (e.g., Zaby and Wilde, 2018), have suggested using advanced data analytics for making the most of the textual complaints together with other related information to enable more intelligent business processes, for instance text representation models can prepare textual data as the input to machine learning algorithms for the underlying downstream tasks. Some recent studies have adapted and focused on improving complaint handling practices through developing automated systems that can assist relevant decision-making. For those that have used textual complaints as the data source, their focal points can be grouped into two streams in terms of the different segments in the complaint handling procedures being investigated.

Research in the first stream seem to be inspired by the viewpoint that the speed of the handling process is the prerequisite for a satisfactory outcome, and thus such studies mainly involve automating the activities that often happen prior to the actual complaint resolution. These automation applications include filtering customer complaints from other types of texts (Preotiuc-Pietro et al., 2019; Jin et al., 2013; Coussement and Van den Poel, 2008) and identifying the themes/topics/possible hazards and risks of the complaint (Merson and Mary, 2017; Laksana and Purwarianti, 2014; Tjandra et al., 2015; Bhat and Culotta, 2017). The implication behind this is straightforward, recognising complaints then directing them to the responsible department/personnel in a swift manner can avoid potential customer dissatisfaction caused by delays. While complaint identification is a binary classification problem, topic modelling is the task of classifying a complaint into one of a few categories which are aligned with predefined topics/themes. These studies aim to achieve a better classification performance while maintaining some level of comprehensibility and interpretability of the developed models, and therefore their experimental designs have focused on benchmarking some simpler text representation models and classification algorithms, such as vector space models and the logistic regression (Forster and Entrup, 2017; Gunawan et al., 2018).

Due to the different characteristics of individual customers and the various complicated nature of service/product issues, deciding on appropriate solutions for complaints has long been a human-resource-intensive process in that its performance is heavily dependent on the skills and experiences of employees learned from the past (Lee, Wang and Trappey, 2015). To facilitate intelligent complaint resolving, the second stream of research has focused on developing reasoning systems that can automatically identify similar past complaint cases as the reference solution to the current complaint, such applications are achieved by computing and comparing the semantic similarity between new and historical complaints (Lee, Wang and Trappey, 2015; Cui et al., 2017). These studies proved that with the appropriate use and derivation of linguistic information from textual complaints, not only can they improve the speed of the complaint handling procedure, but they also have an implicit positive impact on helping to restore customer satisfaction.

1.3 Towards a Novel Framework for Customer Churn Prediction

Since the use of text analytics techniques has been tested in practical service recovery/complaint management, the importance of revealing relevant insights from textual complaints has become increasingly appreciated by researchers and companies (Wu, 2022). Most previous studies in this field have merely considered the dimension of efficiency in the delivery of complaint resolution, by facilitating a reduction in time costs and human participation through the automation processes. Yet the concepts regarding the relation between service recovery and a customer's post-recovery behavioural outcomes has rarely been brought into consideration when these automated models were built (Yang et al., 2019). In fact, successful individual complaint handling can be evidenced by customer behavioural outcomes such as the satisfaction with recovery efforts, the attitudes/emotions about the service, or the customer retention/churn after the service, and so forth. Translating them into corresponding operational indicators in line with certain strategic goals can guide and benefit the development of the textual-data-based complaint management system resulting in a higher effectiveness in dealing with customers. Nonetheless research regarding the use of such customer behavioural outcomes as the measures for verifying complaint handling performance is limited in the literature (Khodadadi et al., 2019).

Until recently, a few studies that investigated other subfields of CRM but partially in association with service recovery or complaint management, have tried and succeeded in incorporating some operational indicators into textual-data-driven predictive frameworks (e.g., Yang et al., 2019; De Caigny, Coussement, De Bock, and Lessmann, 2020). Among the different applications, a highly relevant area has been customer churn prediction (CCP) which refers to as the classification task of identifying customers who show a high inclination to end their service relationship with a company (Ganesh, Arnold, and Reynolds, 2000).

Customer churn poses significant challenges across various industries, impacting both revenue and growth. For instance, the cable television industry in the United States experienced a 25% churn rate in 2020, indicating that a quarter of customers switched providers due to poor customer service (Statista, 2024). Similarly, the financial services sector faces substantial churn, with an average customer retention rate of 81%, implying a 19% annual churn rate (CustomerGauge, 2024a). In the telecommunications industry, the retention rate stands at 69%, reflecting a 31% churn rate, which underscores the competitive nature of the sector (CustomerGauge, 2024b). These statistics highlight the critical need for effective customer retention strategies across these sectors.

A CCP model enables a company to allocate scores to its customers based on an estimation on their future churn intention, such estimation traditionally draws on structured data such as individual customers' demographic data and historical behavioural information (Risselada, Verhoef, and Bijmolt, 2010; De Caigny et al., 2020). Customer churn prediction has been gaining importance and has been widely applied in different industries to inform the decision-making of managing customer relationship and the launch of retention campaigns. However, just as the demographic and the day-to-day transaction-related data have been valuable in enhancing the strength of churn prediction, so too have those unstructured data especially in textual form, considering that written language is filled with complex meaning, ambiguity, and nuance in which massive customer insights can be captured (Humphreys, 2021). Not surprisingly, the latest developed churn prediction frameworks have adapted to incorporate textual data (e.g. contact emails between the company and the customer), modelled with modern text representation models and deep learning architectures (De Caigny et al., 2020).

Regarding the integration of complaint management and customer churn prediction, it is realised that there has been analysis into customer complaint data in past CCP studies, yet many of which just collected structured information such as the number of complaints recorded within a certain length of period, and analysis into textual complaints has remained scarce (Wu, 2022). This might be because collecting and storing the complaint texts for every individual customer had been technically challenging and not financially cost-effective. In fact, textual complaints can be helpful to churn prediction due to their unique reflection of the perception of customers regarding service failure events, and moreover on some occasions such unsatisfactory experiences are often the trigger points/root cause of customer churning (Spanoudes and Nguyen, 2017).

Based on a review of the literature, research regarding the automation of complaint handling activities in association with achieving certain CRM strategic goals such as CCP, has been scarce (Yang et al., 2019). A combination of the concepts and practical techniques from the two disciplines can be explored as the link for filling such a research gap. Thus, this paper proposes a novel framework for customer churn prediction, which in turn can support intelligent decision-making in service recovery campaigns. In this study, the collected multimodal complaint data is analysed through the proposed framework. By drawing on advanced text analysis techniques this paper aims to offer insights into factors that contribute to automate churn prediction. The following section outlines the research aim, objectives, and questions.

1.4 Research motivation

In the era of big data and artificial intelligence, businesses are increasingly leveraging predictive analytics to enhance customer relationship management (CRM) strategies. However, existing customer churn prediction (CCP) frameworks predominantly rely on structured data, such as demographic attributes, transaction history, and service usage patterns. While such models provide valuable insights, they overlook unstructured textual data, particularly customer complaints, which contain rich sentiment, behavioural signals, and dissatisfaction patterns. Given that service failures and subsequent complaint handling play a pivotal role in shaping post-recovery customer behaviour, the integration of textual complaint data into churn prediction models remains an underexplored area with significant research potential.

The aim of this research is:

"*To investigate how Natural Language Processing (NLP) techniques can be applied to customer complaint texts to predict churn rates and develop a predictive framework that integrates advanced NLP techniques with structured data to enhance customer churn prediction accuracy and decision-making in service recovery strategies*."

To achieve this aim, the study sets forth three key research objectives:

1) To identify and evaluate the most effective text representation and feature extraction methods for analysing textual complaint data in customer churn prediction.
2) To compare the performance of different machine learning models in predicting various classes of churn behaviour.
3) To examine the impact of incorporating multimodal data (textual complaints and structured variables) on the performance of churn prediction models.

From these objectives, the study formulates the following five research questions (RQs):

1. What are the most effective text representation and feature extraction methods for analysing textual complaint data in customer churn prediction?
2. How do different machine learning models compare in terms of predictive performance across various classes of churn behaviour?
3. What is the impact of incorporating multimodal data (textual complaints and structured variables) on the performance of churn prediction models?
4. How effective are different data fusion techniques (early, intermediate, and late fusion) in improving churn prediction accuracy?
5. Which structured variables contribute most significantly to churn prediction, and how do they complement textual features in predictive modelling?

This research contributes to both theoretical advancements and practical applications by developing a multimodal analytical framework that integrates textual complaints with structured data, allowing businesses to proactively identify high-risk customers and optimize service recovery strategies. The findings of this study offer insights into the automation of churn prediction and provide a data-driven approach to improving customer retention management. For a more detailed discussion of the research aim, objectives, and questions, refer to Chapter 6 (Methodological Approaches), where these elements are systematically elaborated and linked to the study's experimental framework.

1.5 Thesis approach

This research adopts a positivist philosophical stance, aligning with its objective to develop an empirically validated predictive framework for customer churn based on machine learning techniques. The study follows a quantitative research approach, leveraging a structured and systematic methodology to analyse multimodal customer complaint data. The primary aim is to assess the predictive power of different text representation techniques, machine learning architectures, and data fusion strategies in customer churn prediction. To achieve this, an experimental research design is employed, involving the construction and evaluation of various predictive models trained on real-world customer complaint datasets. This structured approach ensures that findings are replicable and generalizable, contributing to both academic knowledge and practical business applications.

The study is based on secondary data sourced from a third-party complaint mediation platform in the United Kingdom. The dataset comprises structured customer data and unstructured textual complaint narratives. These multimodal data sources enable a comprehensive analysis of how customer experiences influence churn decisions. Methodologically, this research integrates deep learning and traditional machine learning models, comparing feature extraction techniques such as TF-IDF, CNN-based embeddings, and BERT contextual embeddings. Furthermore, the study evaluates different data fusion methods, including early fusion, intermediate fusion, and late fusion, to examine how structured and unstructured data interact in predictive modeling. The detailed research methodology, including data preprocessing, model development, and evaluation criteria, is further elaborated in Chapter 6.

1.6 Thesis overview

The remainder of this paper is organised as follows. Chapter 2 lays the groundwork with an in-depth literature review, critically examining customer reactions and coping strategies in response to service failures and setting the stage for the need for advanced churn prediction methods. Chapter 3 transitions into an analysis of traditional versus modern complaint management strategies, highlighting the evolution and impact of big data analytics in this domain. In Chapter 4, the focus shifts to multimodal data fusion in complaint analysis, exploring the integration of varied data types for enhanced predictive accuracy. Chapter 5 explores deeper into advanced techniques in textual data analysis, assessing various text representation models and their applicability to churn prediction. The methodological core of the thesis is presented in Chapter 6, where a detailed explanation of the chosen research designs, including data collection and processing techniques, is provided. Chapter 7 presents the culmination of the research with an analysis of the results, offering insights into the efficacy of different models and data fusion methods in predicting customer churn. Chapter 8 discusses the implications of these findings for both theory and practice in customer relationship management.

2. Customer Reactions to Service Failures

This chapter delves into the intricate dynamics of customer behaviours in response to service failures, highlighting the criticality of understanding these behaviours for effective service recovery. By examining the multifaceted nature of customer evaluations during their consumption journey, particularly in the wake of service failures, this chapter aims to unravel the complex interplay between service experiences and customer churn. Drawing on extensive literature, it seeks to identify key factors that influence customer perceptions and behaviours, thereby offering insights into effective strategies for managing and mitigating customer churn.

2.1 Service Encounter

The service encounter is often seen as a process involving a series of incidents, and research has consistently shown that customers' perceptions of service encounters are critical in determining their satisfaction and subsequent behavioural responses (Grönroos, 1984; Grönroos, 1994; Crosby and Stephens, 1987; Brown and Swartz, 1989). Much attention has been given to service encounter in Service Marketing, Consumer Behaviours and Relationship Marketing literature, primarily through the exploration of customer satisfaction models and theories. Numerous leading researchers, including Homans (1961), Surprenant and Solomon (1987), Danaher and Mattson (1994), Mattsson and den Haring (1998), Zeithaml, Bitner and Gremler (2018) and Shostack (1985), have attempted to define and conceptualize service encounters.

Surprenant and Solomon (1987, p.88) define a service encounter as a "*dyadic interaction between a customer and service provider*", reflecting the view that service encounters are interpersonal interactions where customers and front-line employees assume distinct roles. However, this definition has been criticized for its limited scope, as it assumes that all service encounters involve direct human interaction. Many service experiences now occur through technology-mediated channels or even fully automated systems, where traditional interpersonal dynamics are absent (Grönroos, 1994). This raises the question of whether traditional service encounter theories remain applicable in an era where digitalization is reshaping customer-firm interactions.

Some scholars argue that a broader conceptualization of service encounters, which views them as a sequence of interrelated incidents rather than isolated interactions, provides a more accurate representation of the modern service experience (Danaher and Mattson, 1994). For example, in multi-channel service environments, a single transaction may span multiple touchpoints, requiring companies to integrate these interactions into a coherent service experience rather than treating them as independent encounters.

Langeard and Eiglier's (1987) 'Servuction (service + production)' model provides a more comprehensive view by integrating all elements that shape customer experiences during service encounters. These include internal and external customers, new technologies such as email communication, websites, and call centres (Shostack, 1985); the physical service environment, and interactions among employees that contribute to the overall service process. However, this model has also faced criticism for not fully capturing the complexities of modern service ecosystems, where AI-driven and omnichannel interactions challenge traditional assumptions about service delivery (Zeithaml, Bitner, and Gremler, 2018).

Other researchers, such as Blume (1988), also suggest that a service encounter does not only involve the direct interaction between customers and front-line employees. Instead, it may also involve indirect interaction where back-office employees can be part of the organisational-wide effort in influencing the customers' experience.

Homans (1961, cited in Solomon et al., 1985), from an economic point of view, depicts service encounter as A service encounter can be understood as a reciprocal exchange process in which individuals aim to maximize their benefits while simultaneously minimizing the associated costs of the interaction. With the emphasis that the benefit is the core element within the exchange process, Bateson (1999, p.127) describe a service encounter as "*a compromise between partially conflicting parties: the customer, the server, and the service firm as embodied in the environment and rules and procedures it creates for the service encounter*".

The concepts of time and space have also been integrated into defining service encounter. It has been defined as “a period of time” (Shostack, 1985) or “*the time frame during which the customer directly interacts with a servic*e” (Bitner, 1990). This definition is more encompassing as it considers ‘time’ as an attribute associated with service delivery. In addition, Mattsson and den Haring (1998, p.418) relate service encounter to the concept of place, pointing out that “*service encounter constitutes an environment for specific dyadic relationships*”.

For any service provider, a complete service encounter involves interactions with customers over a period at a variety of places. Zeithaml, Bitner and Gremler (2018) propose that multiple service encounters can be likened to a cascade, where there is a series of interaction that customers experience. These interactions, in turn, shape service encounter perceptions, which influence customers’ evaluation of the service. Gabbott and Hogg (1998, p.2), focusing on customers’ evaluation, define service encounter as “*consumer's expectations, past experiences, and previous information about the service product are validated by first-hand experience*”.

Service encounters are not isolated events but rather interconnected experiences that shape customer perceptions over time. The definitions proposed by different scholars highlight the complexity of these interactions, ranging from traditional face-to-face engagements to digital, multi-touchpoint environments. As service ecosystems continue to evolve with advancements in technology and changing consumer behaviours, understanding service encounters requires a more flexible and integrated approach. Future research and business practices must adapt to these developments, ensuring that service encounters remain effective in delivering customer value and satisfaction.

Recent literature has increasingly emphasized the importance of customer experience and customer journeys as fundamental concepts that extend beyond individual service encounters. While traditional research on service encounters has focused primarily on isolated interactions between customers and service providers, modern perspectives advocate for a more holistic, longitudinal approach that considers the entire customer journey (Lemon and Verhoef, 2016). Customer experience is now understood as the cumulative effect of multiple service encounters across various touchpoints, rather than a single dyadic exchange. This broader perspective is particularly relevant in today's service ecosystems, where interactions occur across digital, physical, and omnichannel platforms (De Keyser et al., 2020).

Customer journeys consist of pre-purchase, purchase, and post-purchase stages, each of which encompasses multiple service encounters that collectively shape customer perceptions and behaviours (Becker and Jaakkola, 2020). Within this framework, service encounters serve as key moments within a broader journey, influencing customer trust, satisfaction, and brand loyalty. For example, service failures occurring at different touchpoints can have varying implications depending on when and where they arise. Research suggests that negative service experiences are more likely to result in churn if they occur at critical moments in the customer journey, such as during onboarding or high-stakes transactions (Jaakkola and Terho, 2021). Understanding these patterns allows businesses to develop strategic recovery mechanisms tailored to specific journey stages, rather than addressing complaints in isolation.

Furthermore, the increasing availability of big data and AI-driven analytics enables firms to track and optimize customer journeys in real time (De Keyser et al., 2020). By leveraging predictive analytics, companies can anticipate pain points and proactively implement service recovery strategies before customers decide to churn. This shift from reactive to proactive service recovery highlights the evolution of service management paradigms, where organizations are not only responding to customer complaints but also designing preventative service interventions that enhance overall experience.

Given the evolving nature of service encounters and their role within customer journeys, this study adopts a multimodal analytical approach to investigate how service failures, complaints, and recoveries influence churn decisions across different industries. The following sections will further elaborate on the theoretical foundations and empirical evidence supporting this perspective, linking service encounter research with the broader customer experience and service recovery literature.

2.2 Customer Satisfaction

Customer satisfaction has long been of great interest to consumer behaviour and marketing research. Customer satisfaction is also one of the most crucial indicators to an organisation's success in terms of its effect on customer loyalty and profitability (Fornell, 1992; Anderson, Fornell, and Lehmann, 1994; Anderson, Fornell and Rust, 1997; Rust, Moorman, and Dickson, 2002). Despite its centrality, the concept remains highly debated, with no universally accepted definition. This lack of consensus underscores the necessity of thoroughly reviewing existing research to establish a solid foundation for the current study. A comprehensive understanding of how customers evaluate satisfaction is essential for predicting their subsequent behavioural responses, particularly in service recovery contexts.

However, despite extensive academic inquiry, researchers continue to grapple with a fundamental question: What exactly determines customer satisfaction (or dissatisfaction)? Walker (1995) notes that the antecedents of satisfaction remain inconclusive, as the construct has been examined from various perspectives, including psychological, economic, and behavioural dimensions. Szymanski and Henard (2001), in their meta-analysis of empirical studies, identify five primary antecedents of customer satisfaction: 1) expectations; 2) perceived performance; 3) disconfirmation of expectations; 4) affect (emotional responses); 5) equity (perceived fairness of exchange). This categorization, while comprehensive, is not exhaustive, as customer satisfaction is shaped by an intricate interplay of multiple factors. A detailed examination of these five key antecedents follows.

1) **Expectations: The role of anticipation in customer satisfaction.** Expectations have been widely regarded as a key determinant of customer satisfaction. Scholars conceptualize expectations as either a benchmark for evaluating service performance or as an independent driver of satisfaction. When expectations are treated as an independent factor, they shape customers'

perceptions before their actual experience with a service (Szymanski and Henard, 2001). This implies that customers enter a service encounter with preconceived notions of quality, which significantly influence their post-consumption evaluation.

Several empirical studies support the notion that higher expectations correlate with higher satisfaction, provided the service meets or exceeds those expectations (LaTour and Peat, 1980; Oliver and DeSarbo, 1988). Conversely, unmet expectations often result in disappointment, even if the service quality is objectively reasonable. This highlights the asymmetry in customer satisfaction formation, where negative disconfirmation has a stronger impact than positive experiences (Bearden and Teel, 1983).

2) **Perceived performance: The value-percept diversity perspective.** The performance-based model posits that customer satisfaction is directly related to how well the service performs, independent of prior expectations. This model, rooted in value-percept diversity theory, suggests that customers evaluate satisfaction based on a cost-benefit analysis—assessing whether the service delivers sufficient value relative to its cost.

   Westbrook (1981) and Swan and Oliver (1991) provide strong empirical evidence for a positive correlation between perceived performance and satisfaction. However, an important limitation of this model is its assumption of rationality in consumer decision-making. Research suggests that customers often rely on heuristics and emotional responses rather than purely rational evaluations (Westbrook and Oliver, 1991).

3) **Disconfirmation of expectations: A Dominant Framework in Satisfaction Research.** The disconfirmation paradigm remains one of the most widely accepted models in customer satisfaction research. According to this model, satisfaction is determined by the gap between expected and perceived performance. Positive disconfirmation occurs when performance exceeds

expectations, leading to satisfaction. Negative disconfirmation occurs when performance falls short, leading to dissatisfaction. Zero disconfirmation implies that performance matches expectations, resulting in a neutral response (Oliver, 1980; Bearden and Teel, 1983; LaBarbera and Mazursky, 1983).

Despite its popularity, the disconfirmation model has been criticized for its heavy reliance on cognitive evaluations while overlooking emotional and situational influences. Some scholars argue that satisfaction judgments are not merely a result of expectation-performance comparison but also shaped by past experiences, brand perceptions, and psychological states (Westbrook, 1980; Oliver and DeSarbo, 1988).

4) **Affect: The Emotional Dimension of Satisfaction.** The role of emotions in satisfaction judgements has gained increasing attention in recent research. Affect is often divided into two categories: a) emotions (short-term affective reponses) and b) attribution-dependent reactions (longer-term attitudinal shifts). Bagozzi et al. (1999, p.184) define emotions as "*a mental state of readiness that arises from cognitive appraisals of events or one's own thoughts*". In service encounters, emotions play a crucial role in shaping customer memory and subsequent evaluations. Positive emotions such as joy and excitement enhance satisfaction, while negative emotions such as frustration and anger contribute to dissatisfaction (Westbrook and Oliver, 1991; Mattila and Enz, 2002).

   Several studies confirm that customer emotions, both during and after a service interaction, influence satisfaction levels. For instance, Mattila and Wirtz (2000) demonstrate that emotional responses can sometimes outweigh rational evaluations, causing customers to feel either disproportionately satisfied or dissatisfied. This suggests that businesses must manage not only service quality but also customer emotions throughout the service journey.

In the service consumption context, emotions are understood as a series of affective responses that customers generate and experience during the service encounter (Westbrook and Oliver, 1991). Emotions elicited during service provision are revealed to leave emotional traces and to become part of customers' memory that are available to be recalled and integrated into satisfaction judgements (e.g., Westbrook, 1980; Westbrook and Oliver, 1991; Lemmink and Mattson, 2002; Mattila and Enz, 2002).

It is generally accepted in consumer behaviour research (e.g., Westbrook, 1987; Mano and Oliver, 1993; Dube´ and Morgan, 1998; Liljander and Strandvik, 1997; Mattila and Wirtz, 2000; Phillips and Baumgartner, 2002) that consumption emotion is a multidimensional construct that includes positive and negative types of emotions as two separate and independent dimensions. In the extant literature, numerous studies have examined the relationship between customer consumption emotions and customer satisfaction in normal service context. Based on prior research findings, Wen and Chi (2013) summarise those positive emotions (e.g., happiness, joy) have remarkable positive influence on customer satisfaction while negative emotions (e.g., sadness, disgust, anger) have significant negative impact on customer satisfaction.

5) Equity: The Fairness Perception in Service Exchange. The equity theory, derived from (Adams, 1963), posits that customer satisfaction is influenced by perceptions of fairness in the service exchange. Equity is assessed by comparing one's input-output ratio (effort vs. reward) to that of other consumers or industry norms. Research shows that customers are satisfied when they perceive fair treatment and dissatisfied when they believe they have received an unfair deal (Swan and Oliver, 1989; McCollough, Berry and Yadav, 2000). This is particularly relevant in service recovery scenarios, where customers evaluate whether compensation is proportionate to their perceived loss.

However, the applicability of equity theory is debated, as **fairness perceptions are highly subjective and influenced by cultural norms, personality traits, and past experiences** (Woodruff, Cadotte, and Jenkins, 1983; Oliver and DeSarbo, 1988). For instance, customers from collectivist cultures may place greater emphasis on social justice and group harmony, while those from individualist cultures may focus more on personal benefits.

The cognitive foundations of customer satisfaction research have been significantly shaped by theories such as discrepancy theory, contrast theory, equity theory, and the concept of desire congruency (Parker and Mathews, 2001). Among these, the disconfirmation model has emerged as a dominant approach to defining satisfaction (Caruana, 2002). However, despite its widespread adoption, this model has faced criticisms. Scholars argue that it overly relies on situational factors, failing to incorporate deeper intra-personal influences, such as affective states and general attitudes, which can shape customer satisfaction independently of immediate service interactions (Westbrook, 1980). Additionally, the model is critiqued for neglecting prior experiences that influence customer expectations, including past consumption experiences, word-of-mouth recommendations, and marketing influences (Woodruff et al., 1983; Pieters et al., 1995). Furthermore, it does not fully account for customers' aspirational desires—expectations that extend beyond direct service encounters to broader perceptions of ideal service performance (Spreng et al., 1996).

Beyond cognitive evaluations, affective factors also play a crucial role in shaping satisfaction judgments. Westbrook and Oliver (1991) emphasize that satisfaction is not merely a rational process but is significantly influenced by emotional responses. This perspective is further supported by Szymanski and Henard's (2001) meta-analysis, which highlights the interplay between cognitive and affective components in satisfaction judgments. Their findings suggest that customer evaluations are not static but evolve dynamically, as past experiences, emotions, and expectations continuously interact to shape satisfaction perceptions over time.

The dynamic nature of satisfaction assessments has been explored extensively in subsequent research. Oliver (1980) first proposed that satisfaction judgments influence future expectations, suggesting that a customer's evaluation of a service encounter informs their baseline expectations for subsequent interactions. Building on this, Bolton (1998) introduced a cumulative satisfaction model, arguing that satisfaction should be viewed as a longitudinal construct rather than a single-instance evaluation. Bolton and Lemon (1999) further reinforce this idea with their customer service usage model, which suggests that satisfaction evolves as customers repeatedly interact with a service provider, adjusting their expectations and perceptions based on each experience.

Tronvoll (2007) extends these perspectives by framing customer evaluation as an ongoing process of adjustment. According to this view, customers continuously refine their satisfaction assessments as they accumulate service experiences, integrating both immediate reactions and long-term perceptions into their overall evaluation. This process underscores the fluid nature of customer satisfaction, highlighting the importance of managing not only isolated service interactions but also the broader customer journey.

In Section 2.2, we explored customer satisfaction as a dynamic and multifaceted construct, shaped by expectations, perceived performance, emotional responses, and perceptions of fairness. Rather than being a static outcome, satisfaction is continuously reassessed based on past interactions and cumulative experiences (Bolton, 1998; Tronvoll, 2007). The expectation-disconfirmation model remains widely used, but its limitations—such as overlooking emotional influences and prior experiences (Westbrook, 1980) - highlight the need for a broader perspective that incorporates both cognitive and affective factors.

Within the journey in a service failure context, customer satisfaction plays a crucial role in shaping responses to service failures. A customer's perceived fairness in service interactions influences whether they choose to remain silent, voice a complaint, or exit the relationship. Moreover, satisfaction assessments impact expectations in future encounters, meaning that recovery efforts must not only address immediate dissatisfaction but also rebuild trust and shape long-term loyalty.

Recognizing that customer satisfaction is a key determinant of complaint behaviour and post-recovery perceptions, the next section delves into customer coping strategies, examining how different factors—such as emotional responses, prior experiences, and perceived justice - affect the ways in which customers react to service failures and engage in the recovery process.

2.3 Customer Coping Strategies

Service failures are an inevitable part of service interactions, occurring even in organizations with highly refined operational processes. Bell and Zemke (1987, p.1) defined service failures as "*situations in which customers are dissatisfied because their perception of the service they have received is worse than their expectation*." When customers experience a failure, their response is influenced by a complex interplay of cognitive and emotional factors, which ultimately shape their coping strategies. These strategies, in turn, determine whether customers escalate the issue into a formal complaint, engage in negative word-of-mouth (WOM), or adopt alternative measures such as switching providers or seeking third-party intervention (Lazarus, 1966; Lazarus and Folkman, 1984).

Researchers attribute the frequency of service errors to the intangible nature of service provision and the often unpredictable "human interaction" necessary to service encounter (e.g., Siu, Zhang, and Yau, 2013). Within the service failure journey, coping strategies represent a critical decision-making juncture - the point at which customers choose whether to formally voice their dissatisfaction or resort to other means of addressing the failure. Understanding these coping mechanisms is essential for businesses to predict customer reactions, minimize negative consequences, and develop effective service recovery strategies.

The study of consumer responses to service failures has been largely shaped by the cognitive-emotive model (Lazarus, 1966; Lazarus and Folkman, 1984), which suggests that individuals first cognitively appraise a negative event before experiencing an emotional response and ultimately deciding on a coping behaviour. This model has been widely adopted in consumer research (e.g., Stephens and Gwinner, 1998; Chebat, Davidow, and Codjovi, 2005) to explain consumer complaining behaviors as forms of coping strategies.

Building upon this framework, Kim, Wang and Mattila (2010) proposes an integrative model that characterises the key stages of customer evaluation following service failures. This model suggests that customers engage in two primary levels of appraisal: Primary Appraisal - an initial assessment of the severity of the failure and its personal impact. Secondary Appraisal - an evaluation of available coping mechanisms, including their feasibility and likelihood of success. Lastly, influence of Past Experience - how prior interactions with the service provider shape expectations and future coping behaviours.

By analyzing these stages, this section aims to provide a structured understanding of customer responses to service failures. The following subsections explore these appraisals in greater detail, examining how cognitive and emotional evaluations influence the coping strategies that ultimately determine whether a service failure progresses to the complaint stage or is handled through alternative means.

2.3.1 The Primary Appraisal

In the initial stage of customer coping strategies, customers engage in a primary appraisal to assess the severity of a service failure and determine its impact on their consumption goals. This cognitive process influences their subsequent emotional responses and coping decisions. Understanding how customers evaluate the significance of a service failure is essential, as it determines whether they escalate their dissatisfaction into formal complaints, engage in negative word-of-mouth (WOM), or choose to remain silent.

One of the primary factors influencing customer reactions is the perceived severity of the service failure. Richins (1983) found that customers assess not only the immediate inconvenience caused by the failure but also its broader implications. Later research has expanded this perspective by identifying additional factors that contribute to perceived severity. For instance, Stephens and Gwinner (1998) argue that a failure is more likely to be perceived as severe if it disrupts an important consumption goal such as a missed flight connection due to airline mismanagement. Similarly, Richins and Verhage (1985) suggest that the price of the service also plays a crucial role, higher-priced products or services tend to heighten dissatisfaction when expectations are not met.

Beyond tangible losses, service failures can also affect customers' self-perception and emotional state. Research has shown that service failures triggering a perceived loss of status, dignity, or self-esteem tend to provoke stronger negative emotions and higher complaint likelihood (Sparks and Fredline, 2007). Customers who feel personally disrespected or undervalued by service employees may be more inclined to seek retribution, either through direct complaints or negative WOM (Grégoire and Fisher, 2008).

The outcome of the primary appraisal determines whether a customer proceeds to secondary appraisal, where they evaluate potential courses of action. If a failure is deemed minor, customers may choose passive coping strategies, such as ignoring the issue or self-resolving it. Conversely, if the failure is considered highly consequential, customers are more likely to escalate their dissatisfaction by seeking redress through direct complaints or external interventions (Voorhees et al., 2006).

At this stage in the service journey, the primary appraisal acts as the first filter - determining which failures are severe enough to progress toward the complaint stage. If a failure does not meet the threshold of perceived severity, it may never reach the recovery phase, remaining an unresolved grievance or leading to silent dissatisfaction.

2.3.2 The Secondary Appraisal

In the secondary appraisal stage, customers evaluate their ability to cope with the service failure, considering factors such as responsibility attribution, expected future service performance, and their own propensity to complain. This appraisal shapes their decision on whether to escalate the issue through complaints or negative word-of-mouth (WOM), or instead adopt passive coping strategies. Understanding how customers assess these elements is critical for service providers seeking to manage post-failure reactions and prevent customer churn.

A key determinant in this stage is how customers assign responsibility for the service failure. Attribution theory (Weiner, 1985) suggests that customers evaluate whether the failure was caused by the service provider (external attribution) or by their own actions or external circumstances (internal attribution). If the failure is perceived as the company's fault (e.g., a delayed flight due to poor scheduling), customers are more likely to complain and demand compensation (Folkes, 1984; Blodgett et al., 1997). If the failure is attributed to external, uncontrollable factors (e.g., a flight delay due to bad weather), customers may show more understanding and be less likely to escalate dissatisfaction (Stephens & Gwinner, 1998). If customers believe they were partially responsible (e.g., arriving late for a flight), they are less likely to blame the company and may attempt to resolve the issue privately (Bitner et al., 1990).

Attribution not only influences complaint behaviour but also affects emotional responses. Service failures perceived as intentional or due to negligence generate stronger negative emotions such as anger and frustration, increasing the likelihood of revenge-seeking behaviours like aggressive complaints or public shaming via online reviews (Grégoire & Fisher, 2008).

Customers also assess whether the service failure is a one-time occurrence or likely to happen again in the future. This stability attribution affects their willingness to continue engaging with the service provider (McCollough, Berry, & Yadav, 2000). If the failure is perceived as a rare incident, customers may be more forgiving and open to service recovery efforts (Mattila, 2004). If failures frequently occur, customers may conclude that the service provider is incompetent, leading to higher customer churn and negative WOM (Voorhees et al., 2006).

This perception is particularly relevant in industries where service consistency is critical, such as airlines, healthcare, and banking. For instance, a single bad meal in a restaurant might not deter future visits, but repeated hygiene issues could lead to permanent disengagement and complaints to regulatory bodies (Davidow, 2003).

The appraisal of one's propensity to complain, on the other hand, depends on several personal characteristics and situational constraints. For instance, it has been widely agreed that costs of money and time, spiritual consumption to complain, and reputation of the service provider are the primary factors that affect one's reaction behaviours (Richins and Verhage, 1985; Huppertz, 2014; Voorhees et al., 2006). Furthermore, the consumer's anticipated likelihood of receiving a satisfactory response from the service provider, is also suggested to impact on their propensity to complain (Hirschman, 1972; Blodgett, Wakefield and Barnes, 1995; Kim et al., 2003). Given the previous example that happened in a gourmet restaurant, the likelihood of voicing dissatisfaction would be dependent on the consumer's personal characteristics and the situational factors.

2.3.3 Customers' Past Experience with Service Providers

Customers' prior experiences with a service provider significantly influence their evaluation of service failures and subsequent responses. While earlier research conceptualized satisfaction as an episodic evaluation, recent studies emphasize its cumulative nature, where past interactions shape present and future service expectations (Bolton et al., 2004). This challenges traditional models that treat failures as isolated incidents rather than as part of a broader history of interactions.

A key debate concerns whether prior positive experiences increase tolerance for failures or heighten dissatisfaction when expectations are violated. Some argue that satisfied customers are more forgiving of minor lapses (Mattila, 2004), while others suggest they are less tolerant, as failures feel like a betrayal of trust (Bitner et al., 1990; Grégoire and Fisher, 2008). This loyalty paradox implies that highly engaged customers may be more likely to complain, seek redress, or even retaliate.

Beyond satisfaction history, previous complaint-handling experiences shape how customers respond to failures. Those who have received fair resolutions are more likely to complain again, expecting a positive outcome (Blodgett et al., 1997), whereas those who experienced unsatisfactory resolutions may either disengage or escalate their dissatisfaction through negative word-of-mouth (Tax et al., 1998). This raises concerns about whether service recovery efforts truly rebuild trust or merely encourage repeat complaints without addressing underlying service deficiencies.

Past experiences also determine whether and how customers react. Some customers, particularly in industries with high switching costs or limited alternatives, may opt for silent dissatisfaction, continuing their patronage while actively seeking alternatives (Davidow, 2003). This hidden churn poses a greater long-term risk than overt complaints, as dissatisfied customers quietly defect without giving firms the opportunity to recover (Hirschman, 1970). Conversely, customers who have experienced effective recoveries may become brand advocates, reinforcing the notion that well-executed complaint handling can turn failures into opportunities.

The role of past experiences within the service failure context is critical. At the failure stage, prior experiences influence whether customers perceive the issue as an anomaly or a pattern of poor service. During the complaint stage, past complaint-handling outcomes shape whether they voice dissatisfaction or disengage. Finally, at the recovery stage, prior service recovery effectiveness determines whether customers remain loyal, defect, or actively damage the brand through negative advocacy.

These dynamics challenge assumptions that loyalty always leads to greater tolerance and that complaint resolution automatically restores trust. Instead, past interactions create a complex framework where service failures are judged in relation to previous experiences. As we transition to Section 2.4 on customer coping strategies, it becomes clear that customers’ reactions - whether through complaining, exiting, or remaining silent - are shaped by their accumulated experiences, emphasizing the need for a longitudinal perspective in service failure and recovery research.

2.4 Types of Customer Coping Strategies

Service failures, though often inevitable, provoke a range of customer responses that significantly impact firms' ability to retain dissatisfied customers and manage brand reputation. The ways in which customers cope with service failures have been extensively explored in consumer complaining behaviour research (e.g., Singh, 1988; Stephens & Gwinner, 1998), yet the diversity and complexity of these responses remain a subject of debate. This section examines how customers react to service failures, comparing competing frameworks and discussing their implications within the journey in a service failure context.

The types of customers' coping strategies to service failures have been widely researched in prior studies. Various taxonomies have been proposed to characterise the different types of coping strategies based on, such as, the distinction between private and public actions (Day et al., 1981), the motivations of complaining (Singh, 1988), behavioural or non-behavioural actions. Lazarus and DeLongis (1983), from a psychological perspective, suggest that coping strategies can be grouped into the avoidance-based, the emotion-focused, and the problem-focused.

A long-standing classification in the literature is Hirschman's (1970) Exit-Voice Theory , which categorizes customer responses into two primary forms. While 'exit' means the customer intends to drop out from the relationship with the service provider by either just stopping buying or switching to another brand, 'voice' implies a complaint is to be sent to the service provider (Hirschman, 1970; Fornell and Wernerfelt, 1987).

However, subsequent research has challenged this binary framework, proposing more nuanced taxonomies. Singh (1988), for example, distinguishes between public and private coping strategies: Public actions include direct complaints to the firm or third-party involvement. Private actions involve spreading negative word-of-mouth (offline or online) or disengaging without formal complaints.

Other scholars (e.g., Lazarus and Folkman, 1984; Maute and Forrester, 1993) take a psychological perspective, categorizing responses into: 1) Problem-focused coping: Customers actively attempt to resolve the failure, often through complaints; 2) Emotion-focused coping: Customers regulate their emotions rather than attempting to change the situation (e.g., venting frustrations to friends or avoiding confrontation); 3) Avoidance coping: Customers disengage entirely, either by switching providers or suppressing dissatisfaction.

Each coping strategy has distinct implications for firms, particularly in relation to the failure-complaint-recovery chain. Customers who choose problem-focused coping contribute to the complaint stage, offering firms an opportunity for recovery. By contrast, those who engage in emotion-focused coping (e.g., venting through negative word-of-mouth) or avoidance coping (e.g., silent defection) bypass the complaint stage, making recovery more difficult. This highlights the need for proactive service recovery measures that encourage complaints rather than allowing dissatisfaction to remain hidden.

The different taxonomies proposed in the literature are presented in Table 2.1. According to these studies, common consumer behaviours can be outlined as: remailing silent, spreading negative word of mouth offline and online, complaining to the service provider, and seeking help from a third party. These are discussed further in later subsections.

Table 2.1 Classification Schemes of Customer Coping Strategies in Service Failure Contexts

Taxonomy	**Description**	**Authors**
Distinction between Private and Public Actions	Differentiates coping strategies as either private (internal to the customer) or public (external and visible).	Day and Landon (1977); Day et al. (1981)
Motivations of Complaining	Categorizes coping strategies based on the motivations behind customers'	Singh (1988)

	complaints, such as seeking redress, venting emotions, or warning other consumers.	
Behavioural or Non-behavioral Actions	Differentiates coping strategies into active behavioural responses or non-behavioural ones.	Day and Landon (1977)
Psychological Perspective	Groups coping strategies into avoidance-based, emotion-focused, and problem-focused categories, considering the psychological mechanisms underlying the coping process.	Lazarus and DeLongis (1983)
Social Support Seeking	Involves customers seeking support from their social circles to manage service failure experiences.	Thoits (1986)
Information Seeking	Categorizes coping strategies based on the customer's tendency to seek additional information post-service failure.	Brashers (2001)
Exit Strategy	Includes coping strategies where the customer chooses to leave or avoid the service provider in future interactions.	Hirschman (1970)
Voice Strategy	Includes strategies where customers actively communicate their dissatisfaction to the service provider or through public platforms.	Hirschman (1970)
Loyalty as a Coping Strategy	Explores the concept of customer loyalty as a coping mechanism in response to service failures.	Oliver (1999)
Ethical Consideration in Coping	Examines how customers' ethical beliefs and values influence their coping strategies, especially in	Vitell and Muncy (2005)

	situations involving perceived injustice or unfair treatment.	

2.4.1 Remailing Silent

One of the most prevalent yet least observable responses to service failures is customer silence (Zeelenberg and Pieters, 2004), with studies indicating that over 90% of dissatisfied customers choose not to formally lodge complaints (Lapidus and Schibrowsky, 1994; Vos, Huitema and De Lange-Ros, 2008). This phenomenon presents a significant challenge for service providers, as it limits opportunities for organizational learning and improvement through customer feedback.

The decision to remain silent is often a result of a cost-benefit evaluation, where customers assess whether the potential gains from complaining outweigh the perceived costs, including time investment, emotional effort, and anticipated confrontation (Day, 1984). In many cases, the perceived futility of complaint resolution deters customers from voicing their dissatisfaction, particularly when past experiences suggest that service providers are unresponsive or dismissive (Chebat et al., 2005). Furthermore, personality traits and cultural norms influence complaint behaviors; for instance, individuals with higher assertiveness are more likely to complain (Richins, 1983), whereas those from collectivist cultures may refrain from confrontation to maintain social harmony (Liu and McClure, 2001).

From the perspective of the service recovery framework, customer silence disrupts the feedback loop, preventing firms from identifying and addressing service deficiencies. While some customers may continue their patronage due to switching barriers—such as contractual obligations or limited alternatives (White and Yanamandram, 2004) - their continued engagement should not be conflated with genuine loyalty (Rowley and Dawes, 2000). Unaddressed dissatisfaction may manifest in gradual disengagement, reduced consumption, or eventual defection, particularly when viable substitutes become available.

2.4.2 Spreading negative word of mouth offline and online

In contrast to remaining silent, some dissatisfied customers externalize their grievances by spreading negative word-of-mouth (WOM), which involves sharing unfavorable experiences with others, either through personal conversations or digital platforms (Folkes, 1984; Blodgett et al., 1993). The effects are said to become intensified as the negative message being spread to broader social networks (Lau and Ng, 2001).WOM is considered as a type of private action (Day and Landon, 1977; Singh, 1988), an indirect complaining behaviour (Marquis and Filiatrault, 2002), or an emotion-focused coping strategy through which customers can gain sympathy from others (Lazarus and DeLongis, 1983; Zeelenberg and Pieters, 2004).

Past studies (e.g., Richins, 1983; Blodgett et al., 1993; Sundaram et al., 1998) identify several key motivations behind negative WOM: 1) complaining is unlikely to result in satisfactory consequence; 2) a product failure is perceived to be controllable but not stable (or stable but not controllable); 3) keeping others away from facing the same problem; 4) reducing the anxiety of themselves; 5) retaliating; and 6) seeking for advice.

With the rise of digital communication, the impact of negative WOM has intensified through electronic word-of-mouth (e-WOM), where dissatisfied customers disseminate complaints via online reviews, social media, and complaint forums (Zeelenberg and Pieters, 2004). Unlike traditional WOM, which is limited to personal networks, e-WOM has a broader reach and longer lifespan, making it more damaging to service providers (Lau and Ng, 2001). Studies have shown that consumers are more influenced by negative reviews than positive ones, further exacerbating the impact of service failures on brand perception (Hennig-Thurau et al., 2004).

Given the detrimental impact, service providers are advised to be well-prepared for the substantial usage of Internet through developing specific strategies to deal with both online and offline negative word of mouth.

2.4.3 Complaining to the service provider

Unlike customers who remain silent or spread negative word-of-mouth, some dissatisfied consumers opt for direct complaints to service providers, engaging in a formal communication process to express their dissatisfaction and seek resolution. This form of coping is particularly significant within the failure-complaint-recovery chain, as it provides service providers with an opportunity to rectify failures, regain customer trust, and prevent reputational damage (Hirschman, 1970).

Empirical research identifies several factors influencing customers' likelihood to lodge complaints. McKee et al. (2006) found that expectation of resolution is a key factor - customers who believe the service provider has the capability and willingness to resolve their issue are more inclined to complain directly. Stephens and Gwinner (1998), on the other hand, suggest that consumers with a strong awareness of their rights or a belief in procedural fairness are more likely to file complaints. In addition, desire for compensation or correction can also become a cause for complaining because customers may complain to seek redress (refunds, discounts, replacements) or to ensure future service improvements (Blodgett, Wakefield and Barnes, 1995).

However, not all complaints are motivated by legitimate grievances. Research suggests that opportunistic complaining, driven by financial incentives or manipulative intent, is also prevalent (Prim and Pras, 1999; Reynolds and Harris, 2006). Such cases pose a challenge for businesses in differentiating genuine service failures from exploitative complaints.

In comparison with other coping strategies, complaining is argued to be the least damaging to an organisation because complaints can offer insights into what could have been wrong during the service process (Fornell and Wernerfelt, 1987). However, mishandled complaints can exacerbate dissatisfaction, leading to heightened frustration, loss of goodwill, and intensified negative word-of-mouth (Blodgett et al., 1997). Customers who feel ignored or unfairly treated are more likely to retaliate, spreading negative reviews and disengaging from the brand. Within the service recovery framework, direct complaints play a pivotal role in facilitating service recovery. Unlike negative WOM, which operates outside the organization's control, direct complaints provide a structured opportunity for intervention.

In summary, direct complaints are integral to the 'failure-complaint-recovery' process, offering organizations a crucial opportunity to rebuild trust and strengthen customer relationships. The next section explores an alternative and more escalated coping mechanism - seeking third-party intervention - which signifies an increased level of customer dissatisfaction and loss of confidence in internal resolution mechanisms.

2.4.4 Seeking help from a third party

When customers perceive that direct complaints are ineffective or ignored, they may escalate grievances to external bodies such as consumer protection agencies, regulatory authorities, legal systems, or online platforms (Fisher et al., 1999). This escalation signals a failure in the internal complaint resolution process, reflecting a breakdown in a service journey.

Customers turn to third-party channels when they experience perceived injustice, repeated failures in complaint resolution, or severe financial or contractual disputes (Davidow, 2003; McAlister and Erffmeyer, 2003). For instance, More expensive products or services are considered more likely to lead customers to seek external redress compared to lower-cost purchases (Day and Landon, 1977). Regulatory agencies serve as mediators in such conflicts, imposing penalties where necessary, while legal action, including lawsuits and arbitration, introduces further financial and reputational risks for businesses (Fisher et al., 1999). The rise of digital platforms has further transformed third-party complaints, enabling dissatisfied customers to share their negative experiences with a broad audience in real-time, exacerbating reputational damage (Zeelenberg and Pieters, 2004).

Actions of third-party complaints may involve: regulatory agencies mediating disputes or imposing penalties (Singh, 1989); legal action, including lawsuits and arbitration, with financial and reputational risks (McAlister and Erffmeyer, 2003); media exposure and digital complaints, where dissatisfied customers leverage social platforms, amplifying negative experiences (Hennig-Thurau et al., 2004).

The escalation of complaints to third-party channels presents a serious challenge for service providers. Once external bodies intervene, businesses often lose control over the resolution process, as regulatory authorities or legal systems dictate the outcome. Media exposure or viral social media complaints can amplify reputational harm, influencing prospective customers' perceptions (Hennig-Thurau et al., 2004). Financial implications also arise, as companies may be required to provide compensation, pay fines, or undertake costly service improvements to address regulatory concerns (McAlister and Erffmeyer, 2003).

Despite these challenges, proactive engagement with third-party complaints can serve as an opportunity for companies to restore consumer trust and demonstrate corporate accountability (Singh, 1989). Strengthening internal complaint-handling mechanisms can prevent escalation by ensuring that grievances are resolved at an earlier stage (Tax et al., 1998). Businesses that promptly engage with third-party entities and acknowledge customer concerns can mitigate potential reputational damage. Active monitoring of online complaints and swift public responses help counteract negative sentiment before it gains traction (Blodgett et al., 1997). Publicly demonstrating corrective action, particularly in high-profile cases, not only reassures affected customers but also signals a commitment to continuous service improvement.

Third-party complaints underscore the limitations of ineffective service recovery efforts, but they also highlight the potential for brand rehabilitation when managed strategically. Addressing such escalations requires an approach that not only resolves individual grievances but also reinforces broader service recovery mechanisms. The subsequent section explores the individual factors shaping customer responses to service failures.

2.5 Influential Individual Factors

The manners in which customers deal with service failures were initially defined by scholars (e.g., Jacoby and Jaccard, 1981; Singh, 1988) as behavioural reactions that are directly driven by customers' judgement on satisfaction. However, later studies have argued that such a direct link is overly simplistic and does not account for the complexity of individual differences. Other research (e.g., Smith et al., 1999; McColl-Kennedy and Sparks, 2003) suggests that customers tend to implicitly integrate their satisfaction evaluation with additional individual factors when deciding their coping strategies. This implies that service failure reactions are not only situational but are also influenced by stable personal characteristics and past experiences.

Susskind (2005) claims that the way customers response to a service failure is dependent not only on the nature of the incident but also on various surrounding personal factors, such as emotions, personality traits, demographics, and cultural influences. These factors may moderate the decision to complain, switch providers, or remain silent. Notably, while these factors provide a valuable framework for understanding customer behaviour, empirical research has not always reached a consensus on their relative importance, leading to ongoing debates within service failure literature. The rest of this section discusses several key individual factors that shape customers' responses to service failures.

2.5.1 Emotions

Prior research has confirmed that customers' affective states impact their satisfaction judgements (Westbrook, 1980; Westbrook and Oliver, 1991; Stephens and Gwinner, 1998; Maute and Dube´s, 1999). More recent studies have shown that specific negative emotions - such as anger, frustration, and disappointment – can significantly shape coping strategies (Yi and Baumgartner, 2004; Zeelenberg and Pieters, 2004, Mattila and Ro, 2008). For example, highly frustrated customers are more likely to engage in aggressive complaints or negative word-of-mouth, whereas disappointed customers may simply disengage without voicing their dissatisfaction.

Furthermore, research suggests that emotions influence not only immediate behaviour but also long-term attitudes towards the brand and recovery efforts (Watkins and Liu, 1996; Stephens and Gwinner, 1998; Mattila, 2004; Schoefer and Diamantopoulos, 2008). However, a limitation of existing studies is that they often categorize emotions into broad positive-negative dichotomies without fully exploring the nuances of different emotional reactions. For instance, anger and frustration may both be negative emotions, but they lead to different coping behaviours - anger is more likely to result in direct complaints, while frustration may lead to avoidance strategies.

2.5.2 Demographics and Personality

Stephens and Gwinner (1998) argue that individual personality significantly impacts customer responses to service failure due to the role of cognition in evaluation processes. Personality traits such as assertiveness, risk aversion, and extroversion influence whether a customer will escalate a complaint or silently withdraw. For instance, Richins (1983) found that assertive individuals are more likely to engage in formal complaints, whereas passive individuals tend to avoid confrontation.

Extroverted customers, on the other hand, are often less likely to complain formally but more likely to share their dissatisfaction through informal channels, such as social networks or online reviews (Richins, 1987; Kowalski, 1996). This suggests that firms need to adopt different service recovery approaches based on customer personality profiles, as traditional complaint-resolution mechanisms may not effectively address the needs of all customer segments.

Demographic factors, such as age, gender, education, and income, have also been linked to variations in complaint behaviour, though findings remain inconsistent. Some studies suggest that older customers are less likely to complain due to generational differences in communication styles and expectations (Day and Landon, 1977; Bearden and Mason, 1984), while others argue that younger consumers—who are more digitally engaged—are more inclined to post negative reviews rather than seek direct resolutions (Hunt, 1991; Heung and Lam, 2003). Gender differences have also been noted, with female customers generally being more vocal in seeking redress, particularly in service failure contexts (McColl-Kennedy et al., 2003).

However, a critical issue with demographic-based studies is their failure to account for individual variability within groups. Not all younger consumers prefer online complaints, and not all older customers avoid confrontation. Future research should shift towards understanding the intersection of personality traits and demographic factors rather than treating them as isolated predictors of behaviour.

2.5.3 Cultural Difference

Cultural variations significantly impact customer complaining behaviour, particularly in terms of individualistic versus collectivistic orientations. Research has shown that individualist cultures (e.g., the United States, Canada) encourage open confrontation and direct complaints, whereas collectivist cultures (e.g., China, Japan) emphasize social harmony, leading to more indirect responses (Richins, 1983; Richins, 1987; Liu and McClure, 2001; Yuksel et al., 2006).

For example, Asian customers may be less likely to voice direct complaints due to face-saving concerns but more likely to engage in negative word-of-mouth as an alternative coping strategy (Le Claire, 1993; Ngai et al., 2007). This aligns with findings that collectivist consumers are more likely to seek redress through peer networks rather than through direct confrontation with service providers. However, some scholars argue that this cultural model oversimplifies consumer behaviour, as younger generations in collectivist cultures are increasingly adopting individualistic communication styles due to globalization and digital engagement.

Additionally, cultural differences extend beyond communication styles to expectations of fairness and justice in service recovery. For example, customers from high-power-distance cultures may be less likely to challenge authority figures in service encounters, whereas those from low-power-distance cultures may demand higher levels of procedural fairness (Blodgett et al., 2006). These nuances suggest that firms operating in global markets must tailor their service recovery approaches based on regional cultural norms rather than applying a one-size-fits-all strategy.

2.5.4 Technologies

With the rapid advancement of self-service technologies, technology-mediated service failures have introduced new complexities to customer reactions. Despite the increasing reliance on digital interfaces, research on technology-based service failures remains relatively scarce (Meuter et al., 2003; Shapiro and Nieman-Gonder, 2006).

Meuter and others (2000) argue that recovery from technology-based failures is more challenging than traditional service failures because customers often lack direct human intervention to resolve issues. For example, a malfunctioning self-checkout machine in a supermarket may frustrate customers, particularly if there is no immediate support available. Holloway and Beatty (2003) found that delays in resolving technology-based failures—such as online booking errors or mobile payment failures—heighten negative emotional responses and reduce trust in the service provider.

However, some scholars suggest that not all technology failures provoke negative reactions—tech-savvy customers may be more forgiving, particularly if they perceive the issue as a temporary glitch rather than a systemic problem (Shapiro and Nieman-Gonder, 2006). This highlights the importance of segmentation in technology-driven service recovery strategies, as customer responses may vary based on digital literacy levels.

2.5.5 Summary

Understanding the role of individual factors in shaping customer responses to service failures is essential for predicting how different customers will react when service expectations are not met. Emotions significantly influence immediate reactions, with heightened negative emotions often leading to escalated complaints or disengagement. Personality traits and demographic factors introduce variability in complaint behaviours, as assertiveness, extroversion, and cultural background shape the likelihood of customers voicing dissatisfaction or remaining silent. Cultural differences further complicate response patterns, as expectations of fairness and confrontation styles vary across regions. Additionally, the increasing reliance on technology-mediated services adds another layer of complexity, as digital failures may provoke frustration but also demand different recovery strategies.

These individual factors not only determine how customers initially respond to service failures but also influence whether they escalate their dissatisfaction, seek redress, or disengage from the service provider altogether. Customers who feel emotionally wronged or perceive systemic unfairness are more likely to voice their concerns publicly or intensify their complaints, while those with low expectations of resolution may opt for passive withdrawal. The effectiveness of service recovery efforts depends on recognizing these differences and tailoring responses accordingly, ensuring that interventions address both the practical and psychological aspects of customer dissatisfaction.

Given that individual differences shape how customers respond to service failures, it is crucial to explore how service providers can implement effective recovery strategies to mitigate dissatisfaction and restore customer trust - a discussion that follows in the next section.

2.6 Service Recovery

Service recovery research has received much attention over recent decades, particularly with the growing concern about its importance to the success of organisations. The ability to recover from service failures is increasingly viewed not only as a damage control mechanism but also as a competitive advantage that can strengthen customer relationships when executed effectively. Service recovery has been well-defined in the service marketing literature, and it is considered to hold a prominent role because of its effects on customer satisfaction and overall image of an organisation (Swanson and Hsu, 2011). Miller, Craighead, and Karwan (2000, p.388) describe service recovery as the actions of a service provider "*designed to resolve problems, alter negative attitudes of dissatisfied customers and to ultimately retain these customers*".

However, despite the increasing focus on service recovery, research remains divided on what constitutes an effective recovery strategy. While some scholars emphasize financial compensation as a primary means of restoring customer satisfaction (Tax et al., 1998; Smith et al., 1999), others argue that procedural fairness and interpersonal treatment play a more significant role (Blodgett et al., 1997; Liao, 2007). This divergence suggests that service recovery is not a one-size-fits-all approach; rather, it must be tailored based on the nature of the failure and the individual expectations of affected customers.

Prior studies (e.g., Bitner et al., 1990; Wirtz and Mattila, 2004) have revealed that most consumers can tolerate occasional service failures, provided that recovery efforts are perceived as fair and satisfactory. In contrast, ignorance of complaints and refusal to take corrective actions are often the primary causes of escalated dissatisfaction and long-term customer loss. Notably, some researchers suggest that poor recovery efforts may have a more detrimental impact on customer relationships than the initial failure itself, as they erode trust and signal incompetence or indifference on the part of the service provider (Maxham and Netemeyer, 2002).

Service recovery efforts influence both cognitive evaluations and emotional responses, shaping customers' overall perceptions of a service provider. McCollough (2009) argues that customers are often more 'mindful' of recovery procedures than the initial service itself, as service failures create heightened sensitivity to how companies handle complaints. This means that even a relatively minor service failure can have a lasting negative impact if the recovery process is perceived as inadequate.

For instance, once a telecommunication user has been notified of a billing mistake, the user is likely to closely scrutinize how the company resolves the issue. The speed, transparency, and manner in which the refund is processed will determine whether the customer perceives the company as reliable or untrustworthy. This heightened awareness of recovery efforts aligns with the concept of 'double deviation' (Bitner et al., 1990), where a failed recovery attempt exacerbates the customer's dissatisfaction beyond the level caused by the initial failure. This suggests that service providers must not only focus on resolving issues but also actively manage customer perceptions throughout the recovery process.

Successful service recovery efforts can transform dissatisfied customers into loyal advocates, whereas failed recovery attempts can lead to increased churn and reputational damage (Swanson and Hsu, 2011). Research has shown that well-handled recoveries can increase customer retention, brand trust, and even positive word-of-mouth referrals (Chang et al., 2013). Conversely, service providers that fail to address complaints effectively may experience amplified negative word-of-mouth and diminished customer lifetime value (Miller, Craighead, and Karwan 2000).

Maxham and Netemeyer (2002b) argue that customer loyalty is more likely to be restored if the recovery effort exceeds expectations rather than merely meeting minimum standards. This phenomenon, known as the service recovery paradox, suggests that in some cases, customers who experience a well-executed recovery may exhibit higher levels of satisfaction than those who never encountered a failure at all (Ding, Ho and Lii, 2015). However, not all scholars agree with the validity of the service recovery paradox. Some studies indicate that this effect is conditional and does not apply universally across industries or failure types (Michel and Meuter, 2008). In particular, customers who experience repeated failures from the same provider are unlikely to exhibit increased loyalty, regardless of the quality of the recovery effort (Johnston and Fern, 1999).

Recognizing the long-term implications of service recovery, many businesses have integrated structured recovery programs into their customer relationship management strategies. Research has demonstrated that organisations with proactive complaint-handling mechanisms tend to outperform competitors in customer retention and brand equity. (Gursoy et al., 2007). Bendall-Lyon and Powers (2001) suggest that one of the most effective ways to foster long-term customer commitment is to incorporate service recovery as an integral part of a company's operational framework rather than treating it as a reactive measure.

Despite the growing emphasis on structured recovery processes, there remains a gap between theoretical recommendations and real-world implementation. Many organisations still rely on standardized recovery scripts that fail to address individual customer needs, leading to impersonal and ineffective interactions (Tax et al., 1998). Moreover, studies indicate that companies often underestimate the long-term financial impact of poor service recovery, treating complaint resolution as a cost centre rather than as an opportunity to enhance customer lifetime value (Blodgett et al., 1997).

Effective service recovery is not merely about correcting a mistake but about restoring customer trust and strengthening long-term relationships. While financial compensation, procedural fairness, and interpersonal treatment all play important roles, their relative effectiveness depends on the customer's individual expectations and past experiences. Poorly handled recoveries can exacerbate dissatisfaction, leading to heightened complaints, increased churn, and reputational damage.

Understanding the nuances of service recovery is critical, as it determines whether dissatisfied customers remain engaged with the service provider or choose to sever ties entirely. The following section explores the mechanisms through which organisations manage customer complaints and implement recovery strategies to mitigate the negative consequences of service failures.

2.7 Complaint Handling Practices

Given that complaining (i.e., complaining to the service provider, complaining through a third party) has been one of the common coping strategies taken by dissatisfied customers, organisations are suggested to develop robust complaint management programmes to enhance recovery performance for restoring customer satisfaction as well as correcting mistakes within the service process (Grönroos, 2000). A well-structured complaint handling system not only mitigates the negative impact of service failures but also serves as a strategic tool for customer retention and brand reputation management.

Bendall-Lyon and Powers (2001) propose a complaint management procedure with six steps that organizations can use to facilitate effective service recovery:

1) encouraging customer complaining as opportunities for improvement.
2) establishing a complaint management team to handle complaints.

3) resolving complaints in a quick and effective manner.
4) developing a database to store historical complaints.
5) identifying and correcting failure points within the service process to avoid recurring mistakes.
6) conducting analysis for quality improvement based on historical complaints.

However, despite the existence of structured complaint-handling models, many companies fail to implement these frameworks effectively. Research suggests that organisations often view complaints as operational burdens rather than valuable insights into service improvement (Tax et al., 1998). This mindset leads to defensive or dismissive complaint responses, further damaging customer trust and increasing the likelihood of negative word-of-mouth.

2.7.1 Resolving complaints as a core service function

Resolving complaints has been widely accepted as the central component of an effective complaint management process. The realisation of its importance has led to a proliferation of research on individual complaint-handling practices. Extant literature (e.g., Cui et al, 2017) highlights that the knowledge and complaint-handling skills of frontline employees are critical determinants of complaint resolution effectiveness.

Many practical strategies have been presented in prior studies to improve complaint-handling performance, such as: responding with an apology (Bell and Ridge, 1992; Schweikhart et al., 1993; Miller et al., 2000); providing timely, accurate, and empathetic responses (Cho, Im, Hiltz, and Fjermestad, 2002; Cheng and Loi, 2014; Schwab, 2015); and listening actively, acknowledging concerns, and demonstrating proactive resolution efforts (Harrison-Walker, 2001).

However, while these strategies are widely endorsed, their actual effectiveness varies depending on the context and customer expectations. Some researchers argue that apologies alone are insufficient unless accompanied by tangible compensation or corrective actions (Goodwin and Ross, 1992). Customers who perceive a recovery effort as insincere or tokenistic may react more negatively than those who receive no response at all. This highlights the importance of customizing complaint-handling strategies to align with customer perceptions of fairness and resolution adequacy.

2.7.2 Organisational response strategies to complaints

In complaint handling context, organisational response strategies refer to how companies react to customer complaints and whether they adopt a proactive or defensive stance. Marcus and Goodman (1991) classify response strategies into accommodative, defensive, and non-responsive approaches.

1) Accommodative strategies: Organisations acknowledge service failures, take responsibility, and offer tangible compensation (Marcus and Goodman, 1991). These strategies have been found to positively influence brand reputation and customer trust (Lee and Song, 2010).

2) Defensive strategies: Companies refuse to accept blame, shift responsibility to external factors, or downplay the issue (Marcus and Goodman, 1991; Lee and Song, 2010). These approaches often provoke customer frustration and lead to increased negative word-of-mouth.

3) Non-responsive strategies: Organisations ignore complaints or provide generic, automated responses (Lee and Song, 2010). Research shows that this approach is particularly damaging in digital complaint settings, where lack of engagement can escalate consumer dissatisfaction.

Lee and Song (2010) conducted an experimental study comparing these three strategies and found that accommodative responses significantly improved brand perception, whereas defensive or non-responsive approaches led to increased customer resentment. These findings highlight the critical role of perceived fairness in shaping post-complaint attitudes and future engagement with the brand.

2.7.3 The risks of mishandled complaints

Chang and colleagues (2015) examined Lee and Song (2010)'s findings on the impact of response strategies on negative word-of-mouth. Their study confirmed that accommodative responses - where organisations acknowledge the failure and take corrective action - can reduce negative word-of-mouth by lowering customers' attributions of blame and perceived control over the situation. However, some scholars argue that while accommodative responses may be effective in mitigating short-term dissatisfaction, they do not always foster long-term customer loyalty or prevent future complaints (Mattila and Ro, 2008). This suggests that focusing solely on immediate appeasement may not be a sustainable approach, as customers may still harbour doubts about the reliability of the service provider.

In contrast, studies on online complaint handling reveal that automated responses or a complete lack of response can create a sense of customer neglect. Mattila et al. (2013) refer to this phenomenon as cyber-ostracism, where customers who receive automated replies - or no response at all - develop heightened negative emotions, reduced satisfaction, and an increased likelihood of retaliatory behaviours. While automation can improve efficiency, excessive reliance on generic responses without meaningful human interaction may intensify customer frustration rather than alleviate it. This highlights a critical dilemma for businesses: balancing the need for efficiency with the necessity of personalised engagement in digital complaint resolution.

Within accommodative strategies, previous research (Bell and Ridge, 1992; Schweikhart et al., 1993) have identified two key forms of service recovery: psychological and tangible compensation. Psychological recovery efforts focus on alleviating negative emotions through demonstrations of empathy and concern, whereas tangible compensation provides monetary or material redress for the inconvenience caused. Miller, Craighead, and Karwan (2000) validated these findings, showing that empathy and apologies are effective and inexpensive techniques for managing service complaints. However, despite their low cost, apologies alone are often perceived as insufficient unless accompanied by tangible restitution (Goodwin and Ross, 1992). When customers experience financial loss or significant inconvenience, a mere apology may be viewed as a symbolic gesture rather than a genuine recovery effort (Blodgett et al., 1997).

2.7.4 Integrating complaint handling into organisational culture

In addition to the nature of response strategies, the delivery method of service responses also plays a critical role in shaping customer perceptions. Cho and others (2002) argue that organisations should respond to complaints in an accurate, prompt, and empathetic manner, as delayed responses can heighten frustration and worsen the original complaint. Schwab (2015), through a multinomial regression analysis, found a strong correlation between post-recovery satisfaction and the presence of clear, well-communicated resolutions. His findings highlight that beyond merely acknowledging complaints, companies must proactively demonstrate that corrective actions have been implemented.

However, over-reliance on structured response protocols may also backfire. While organisations are advised to implement personalised follow-up services (Harrison-Walker, 2001), excessive follow-ups without meaningful resolutions can be perceived as redundant or insincere. This suggests that complaint management strategies should be adaptable rather than strictly procedural, ensuring that follow-up interactions add value rather than becoming an additional source of frustration.

2.7.5 Summary

Complaint handling plays a pivotal role in shaping customer perceptions, determining whether dissatisfaction is escalated or effectively managed. While accommodative strategies that acknowledge failures and provide tangible remedies can restore trust, defensive and non-responsive approaches often exacerbate dissatisfaction and lead to long-term reputational harm.

Beyond individual complaint resolutions, integrating complaint management into organisational culture enhances service recovery effectiveness and prevents recurring failures. The next section explores the broader implications of customer evaluations of service recovery efforts, examining how perceived fairness influences post-recovery satisfaction, word-of-mouth, and loyalty behaviours.

2.8 Service Recovery Evaluation

Customer satisfaction with service recovery (CSSR) has been a central topic in service marketing research. Over the past few decades, studies have explored various aspects of service recovery, including its performance, key features, and effects on customer perceptions. Early research, dating back to the 1980s, primarily relied on the critical incidents technique to analyze customer consumption experiences and service recovery evaluations (Bitner et al., 1990). This approach provided valuable insights but was largely descriptive, focusing on categorizing service failures and remedial actions rather than explaining the underlying psychological mechanisms. More recently, marketing scholars have expanded service recovery evaluation by incorporating theoretical frameworks from other disciplines, including social psychology. These frameworks include justice theory (Tax et al., 1998; Smith et al., 1999), the disconfirmation paradigm (McCollough, Berry, and Yadav, 2000), attribution theory (Swanson and Hsu, 2011), mental accounting theory (Chuang et al., 2012), and equity theory (Wen and Chi, 2013).

Among these theories, justice theory has emerged as one of the most widely applied in service recovery research (Tax et al., 1998). This theory, originally developed in political philosophy by Rawls (1971), is derived from Festinger's (1962) cognitive dissonance theory and Adams's (1963) equity theory. Justice theory postulates that consumers evaluate service recovery efforts based on their perceptions of fairness, which in turn influences their overall satisfaction and subsequent behavioural responses. However, despite its prevalence in service recovery studies, some scholars argue that justice theory alone may not fully capture the complexity of customer evaluations, particularly in cases where emotional responses play a dominant role in shaping perceptions.

2.8.1 The three dimensions of perceived fairness in service recovery

According to justice theory, service failures inherently involve economic and social interactions that lead customers to appraise recovery efforts in terms of three justice dimensions: procedural justice, distributive justice, and interactional justice (Rawls, 1971). Prior studies have found that this three-dimensional framework accounts for over 60% of service recovery evaluations (Siu, Zhang, and Yau, 2013). Perceived justice has consistently been identified as a key determinant of customer satisfaction with service recovery, reinforcing the notion that fairness perceptions significantly shape post-failure reactions.

Tax et al. (1998) conceptualize service recovery evaluation as comprising two key components: outcome recovery and process recovery. Outcome recovery refers to the tangible remedial measures provided to compensate customers, whereas process recovery concerns the manner in which the failure is communicated and resolved. Justice theory posits that the three justice dimensions—procedural, distributive, and interactional—are the primary factors shaping customer perceptions of these recovery components. However, some scholars argue that this model oversimplifies the decision-making process, as customers may prioritize certain justice dimensions over others depending on the context of the failure and their prior experiences with the service provider (de Ruyter and Wetzels, 2000; Liao, 2007).

Distributive fairness focuses on the perceived fairness of financial compensation or tangible remedies offered by the service provider (del Río-Lanza et al., 2009). It is commonly associated with compensatory rewards such as refunds, discounts, replacements, and vouchers. In the context of service recovery, Maxham and Netemeyer (2002) define distributive justice as "the extent to which customers feel they have been treated fairly with respect to the final recovery outcome." Prior research has shown that perceptions of distributive justice significantly influence customers' overall recovery evaluations (Boshoff, 1997; Goodwin and Ross, 1992; Tax et al., 1998; Hoffman et al., 1995; Smith et al., 1999). However, some studies suggest that distributive justice alone is not sufficient to restore customer satisfaction, particularly when service failures evoke strong emotional responses or cause substantial inconvenience (Wirtz and Mattila, 2004; Gelbrich and Roschk, 2011).

Procedural fairness concerns the methods an organisation uses to manage and resolve service failures (del Río-Lanza et al., 2009). Davidow (2003) argues that procedural justice encompasses policy fairness, accessibility, decision control, flexibility, and the speed of resolution. In a service recovery context, procedural justice is not merely about whether a complaint is addressed but also about how efficiently and transparently the process unfolds (Blodgett et al., 1997; Tax et al., 1998; Maxham and Netemeyer, 2002). Customers often perceive bureaucratic recovery procedures, long wait times, or rigid policies as signals of indifference, exacerbating dissatisfaction rather than resolving it (Wallin Andreassen, 2000; Wen and Chi, 2013).

Interactional fairness pertains to the interpersonal aspects of service recovery, including the courtesy, empathy, and professionalism of service staff (Smith et al., 1999; Wirtz and Mattila, 2004). Recent studies suggest that interactional justice consists of two sub-dimensions: interpersonal justice and informational justice (Colquitt, 2001; Colquitt et al., 2001). Interpersonal justice refers to the perceived sincerity and respect demonstrated by service employees during recovery efforts, whereas informational justice relates to the adequacy and transparency of the explanations provided regarding the failure and the recovery decision (Colquitt, 2001). Studies indicate that interactional justice is particularly critical in service industries where face-to-face interactions play a major role, as customers are more sensitive to non-verbal cues and tone of communication (McColl-Kennedy and Sparks, 2003).

2.8.2 The role of emotions in recovery evaluation

The three justice dimensions collectively provide a multifaceted framework for understanding how customers assess fairness in service recovery efforts. However, research suggests that customers do not weigh these dimensions equally across all recovery scenarios. For instance, when a failure results in financial loss, distributive justice may be the primary determinant of recovery satisfaction. Conversely, for failures involving perceived mistreatment or poor communication, interactional justice may have a stronger influence on customer perceptions (Wirtz and Mattila, 2004).

Despite the robustness of justice theory in explaining service recovery evaluations, some scholars argue that its traditional framework fails to fully account for emotional and psychological factors that influence customer perceptions. For example, cognitive biases, such as loss aversion, may cause customers to overemphasize negative aspects of a recovery effort even when objective fairness criteria are met (Kahneman and Tversky, 1979). Additionally, cultural variations in justice perceptions suggest that fairness evaluations are not universally consistent—customers from collectivist cultures may place greater emphasis on procedural and interactional justice, whereas customers from individualist cultures may prioritize distributive outcomes (Mattila and Patterson, 2004).

2.8.4 Summary

Customers evaluate service recovery based on multiple factors, including perceived fairness, emotional response, and long-term trust. While distributive, procedural, and interactional fairness all contribute to post-recovery satisfaction, their relative importance varies depending on the nature of the failure and the customer's prior experiences. Furthermore, emotions play a crucial role in shaping recovery perceptions, with both negative and positive emotions influencing long-term brand relationships.

However, effective service recovery does not always guarantee stronger customer loyalty. While some customers may appreciate an organisation's effort in addressing a failure, others may remain skeptical, particularly if service failures occur repeatedly. The following section examines the broader implications of service recovery efforts on overall customer experience and explores how firms can integrate recovery strategies into their long-term service management frameworks.

2.9 The relationship between justice dimensions and emotions

Earlier research primarily focused on the role of cognitive assessments in shaping customer satisfaction judgements. However in recent years, scholars have increasingly recognized that emoptions also play a critical role in service evaluation, influencing both immediate reactions and post-consumption behaviours (Bagozzi et al., 1999). Empirical evidence strongly supports this perspective, with studies demonstrating that customers' emotional responses during service interactions can significantly impact their perception of the service process and subsequent behaviours. (Westbrook, 1987; Mano and Oliver, 1993; Dube´ and Morgan, 1998; Liljander and Strandvik, 1997; Mattila and Wirtz, 2000; Phillips and Baumgartner, 2002)

Despite these findings, relatively little is known about the specific role of emotions in service recovery contexts (Smith and Bolton, 2002; Bougie et al., 2003; Menon and Dube´, 2004; Zeelenberg and Pieters, 2004; Schoefer and Ennew, 2005; Rio-Lanza et al., 2009; Jani and Han, 2011). Existing research has primarily examined emotions in routine service interactions, but whether these findings apply directly to service recovery remains an open question. Wen and Chi (2013) argue that the relationship between emotions and Customer Satisfaction with Service Recovery (CSSR) may be analogous to that observed in standard service settings, given the high level of human interaction in service recovery processes. However, this assumption warrants further scrutiny, as service failures often evoke stronger emotional responses than routine service encounters, potentially amplifying the impact of perceived fairness in recovery efforts.

Some studies (Weiss et al., 1999; Dalci and Kosan, 2012) suggest that fair and well-executed service recovery efforts can neutralize negative emotions and even generate positive emotional responses, leading to increased satisfaction. Conversely, perceptions of unfairness in recovery processes can intensify negative emotions, exacerbate dissatisfaction, and further weaken the customer-provider relationship. However, research in this domain remains fragmented, with studies differing in their operationalization of fairness and emotional response.

To explore this relationship, some scholars (e.g., Schoefer and Ennew, 2005; Rio-Lanza et al., 2009) have applied affect control theory and cognitive appraisal theory to examine how perceived justice influences customer emotions in service recovery settings. Their findings suggest that customers' emotional responses are significantly shaped by their perception of fairness in remedial measures, which in turn affects their overall satisfaction with the recovery process (Schoefer and Diamantopoulos, 2008). Nevertheless, while these studies establish a clear link between perceived justice and emotional reactions, they often overlook moderating factors such as prior brand attachment, cultural expectations, and individual differences in emotional sensitivity, which may influence how customers process and respond to service recovery efforts.

Understanding the interplay between justice perceptions and emotional responses in service recovery is crucial for designing effective recovery strategies. While fairness plays a central role in shaping customer satisfaction, it also influences whether customers seek resolution, escalate their dissatisfaction, or disengage from the service provider entirely. When customers perceive a lack of fairness, they may intensify their complaints, explore external dispute mechanisms, or shift their loyalty elsewhere. Conversely, when recovery efforts exceed expectations, they can restore trust and even strengthen long-term relationships.

However, more research is needed to fully capture the complexities of this relationship, particularly in digital service environments where limited interpersonal interactions make it difficult for service providers to gauge customer emotions and adjust recovery efforts accordingly. The next section will further explore how customer responses to recovery efforts translate into broader behavioural patterns, including word-of-mouth, brand perception, and long-term retention.

2.10 The Interplay of Justice Dimensions and Customer Post-Recovery Behaviours

As the justice framework has been widely used in conceptualising customer evaluation of remedial efforts, empirical studies have demonstrated that the three justice dimensions can explain much of customers' decisions on service recovery (e.g., McCollough, Berry, and Yadav, 2000; Wirtz and Mattila, 2004; Siu, Zhang, and Yau, 2013; Wen and Chi, 2013; Ding, Ho, and Lii, 2015). For example, Smith et al. (1999) developed and tested a model of Customer Satisfaction with Service Recovery (CSSR) in the context of hotels and restaurants, finding that the three justice dimensions together accounted for a substantial proportion of customer evaluations regarding recovery efforts. However, while these dimensions collectively shape post-recovery perceptions, the relative influence of each dimension remains a subject of debate, with conflicting conclusions drawn on the most influential antecedent to CSSR and subsequent behaviours.

Many studies suggest that distributive justice is the most significant predictor of post-recovery satisfaction (Tax and Brown, 1998; Smith et al., 1999; Cranage and Mattila, 2006). However, other research ontradicts this claim, arguing that distributive justice is the least influential factor in post-recovery satisfaction (Ok, Back, and Shanklin, 2005). Additionally, while procedural justice is often reported as the least important dimension, some scholars contend that it plays a central role in shaping CSSR (Karatepe, 2006). These discrepancies suggest that the relative importance of justice dimensions may vary depending on the industry, failure context, and individual customer expectations.

Similarly, studies focusing on distributive justice alone present conflicting findings. Noone (2012) found that the magnitude of compensation does not significantly impact post-recovery satisfaction. In contrast, Boshoff (2012) and Chen et al. (2018) argued that excessive may have counterproductive effects, leading to unrealistic expectations for future recoveries. Consequently, some scholars advocate for a moderate compensation strategy to avoid setting undesirable precedents. Yet, other scholars suggest that higher compensation levels can be particularly effective in repairing customer relationships, especially in severe service failures (Bradley and Sparks, 2012; Maxham III, 2001). These contradictory findings highlight the complexity of customer expectations regarding compensation and suggest that the effectiveness of distributive justice may depend on contextual factors such as service failure severity and customer loyalty.

Many researchers have used scenario-based experiments to examine the relationship between justice framework and post-recovery satisfaction. However, their findings remain inconsistent. In the airline industry, McCollough, Berry and Yadav (2000) found that distributive justice and interactional justice exerted the strongest influences on post-recovery satisfaction. Similarly, Ding, Ho, and Lii (2015) and Nikbin et al. (2015) identified distributive justice as the key determinant of post-recovery satisfaction. However, Chang and Chang's (2010) reported conflicting results, showing that distributive justice had no significant impact on either post-recovery satisfaction or repurchase intent. Instead, their findings indicated that interactional and procedural justice were the primary factors considered by air travelers. Such contradictions emphasize the need for further empirical research to establish clearer conclusions regarding the justice dimensions' relative importance across different service contexts.

Conflicting findings also emerge regarding which justice dimension has the strongest impact on repurchase intent and word-of-mouth following service failures. While some scholars argue that all three justice dimensions are positively correlated with post-recovery retention rates (Ghalandari, Babaeinia, and Jogh (2012), others assert that only distributive justice significantly predicts repurchase behaviour (e.g., Blodgett et al., 1997; Elizabeth, 1993; Lin, Wang, and Chang 2011; Kuo and Wu, 2012). These discrepancies indicate that while fairness perceptions influence repurchase decisions, other factors, such as trust restoration and emotional resolution, may also play critical roles.

In the context of word-of-mouth behaviour (WOM), studies present further contradictions. Lin, Wang, and Chang (2011) found that a perceived low level of interactional justice was the strongest predictor of negative WOM in online retail settings. Meanwhile, Wen and Chi (2013) reported that interactional justice significantly influenced positive WOM among airline passengers. However, other researchers (Kim, Kim, and Kim 2009; Nikbin et al. 2012) found that distributive justice exerted the most significant impact on negative WOM, with customer satisfaction mediating this relationship. These findings suggest that while perceived fairness influences customer communication behaviours, the specific justice dimension that drives WOM may depend on the context of the service failure and the communication channels available.

As demonstrated throughout this section, empirical findings on the interplay between justice perceptions and post-recovery behaviours remain inconsistent. Different scholars have reached varying conclusions regarding the most influential justice dimension, with no clear consensus on their relative impact on satisfaction, WOM, and repurchase intent (e.g., Wirtz and Mattila, 2004; Hocutt et al., 2006). These inconsistencies highlight the need for a more holistic understanding of how justice perceptions interact with emotional and cognitive factors in shaping post-recovery outcomes.

2.11 Chapter summary

This chapter has provided an extensive review of customer behaviours following service failures, critically examining theories of customer satisfaction, coping strategies, individual differences, and justice dimensions in service recovery. Findings suggest that while fairness perceptions significantly shape post-recovery responses, they do not operate in isolation. Customers assess recovery efforts not only based on fairness principles but also in relation to how their complaints are handled, whether their concerns are acknowledged, and how effectively the service provider restores trust. When recovery efforts fail to meet expectations, customers may escalate their dissatisfaction, seek external resolution mechanisms, or permanently disengage from the brand. Conversely, well-executed recovery efforts can mitigate negative emotions, rebuild confidence, and reinforce loyalty.

While effective service recovery is crucial in managing post-failure customer relationships, organisations must also adopt a proactive approach to complaint management to prevent dissatisfaction from escalating. With the rapid expansion of online platforms, customer complaints have become increasingly public, data-rich, and complex, requiring businesses to transition from traditional complaint handling methods to more advanced, data-driven approaches. The next chapter explores how organisations are adapting to these challenges, examining the evolution of complaint management strategies in the big data era, the integration of intelligent complaint handling systems, and the role of technologies such as text mining and AI in improving service recovery outcomes.

3. Revolutionising Complaint Management in the Big Data Era

As organisations strive to enhance customer satisfaction and retention, effective complaint management has become a critical component of service strategy. Traditional complaint handling methods have long relied on human intervention, structured feedback systems, and predefined escalation protocols to resolve customer issues. Having well-developed complaint management procedures and practices can lead to successful service recovery for consumers as well as internal improvements in service processes (Grönroos, 2000). For example, Lyon and Powers (2001) propose a complaint management process involving key steps such as developing a complaint database to track trends and gather information for service enhancements. However, despite the structured nature of these approaches, organisations continue to struggle with effectively resolving customer complaints and ensuring customer recovery (Miller et al., 2000).

In recent years, online complaint handling has received significant attention, driven by the emergence of new digital communication channels between customers and companies. Many researchers (Coussement and Van den Poel, 2008; Stoica and Özyirmidokuz, 2015; Cui et al., 2017) argue that the increasing volume of online reviews and complaints has fundamentally reshaped the landscape of complaint management. This shift has led to the adoption of two broad approaches: manual handling, which involves extensive human participation, and intelligent handling, which leverages electronic computation technologies to automate complaint resolution.

Big data has emerged as a transformative force in business and consumer analytics, enabling organisations to process, analyse, and extract insights from vast amounts of structured and unstructured data. It is commonly characterised by the “three Vs”: volume (large-scale datasets), velocity (real-time or near-real-time data processing), and variety (diverse data sources, including text, images, and social media interactions) (McAfee et al., 2012; Kitchin, 2014). In the context of customer service, big data facilitates predictive analytics, sentiment detection, and automated complaint resolution, allowing businesses to move from reactive to proactive service management. However, despite its advantages, big data analytics poses challenges related to data privacy, algorithmic bias, and the ethical implications of automated decision-making (Boyd and Crawford, 2012).

This chapter explores the transition from traditional to data-driven complaint management approaches, focusing on the role of big data analytics, artificial intelligence (AI), and machine learning in improving complaint resolution processes. Unlike previous chapters, which examined customer behavioural responses to service failures and recovery efforts, this chapter shifts the focus towards the technological evolution of complaint management systems. While Chapter 2 analysed complaint behaviour from a consumer psychology and service marketing perspective, this chapter adopts a computational and data-driven approach, investigating how businesses leverage large-scale data processing techniques to enhance service recovery strategies.

Big data has revolutionised complaint management by enabling organisations to extract insights from vast and unstructured customer feedback sources, including social media posts, online reviews, and chatbot interactions. Advanced text mining techniques, sentiment analysis, and predictive modelling have transformed the way companies detect, classify, and respond to complaints. However, despite these advancements, challenges remain, including issues of data accuracy, ethical concerns related to AI decision-making, and the risk of depersonalisation in automated complaint handling systems.

The following sections critically examine how businesses integrate big data-driven strategies into complaint management, comparing their effectiveness with traditional methods. The discussion covers key areas such as automated complaint categorisation, real-time sentiment analysis, and the ethical implications of AI-powered service recovery. Through this analysis, this chapter aims to provide a comprehensive understanding of how technological advancements reshape customer complaint management in the modern era.

3.1 Evolving Strategies in Online Complaint Handling

During the first decade in this century, manual approaches were the dominant methods for online complaint handling. Companies sought to actively monitor customer dissatisfaction and intervene in online complaints through the direct participation of customer service personnel and first-line employees (Fournier and Avery, 2011; Cui et al., 2017). At that time, the Internet was still in its early growth stages, and businesses primarily focused on improving resolution effectiveness. Employee expertise and complaint-handling skills were regarded as the key determinants of successful complaint resolution (Cui et al., 2017).

Several practical strategies were introduced to guide employees in handling complaints across different scenarios. These strategies included responding promptly through online chat platforms (Cho et al., 2002; Cheng and Loi, 2014); actively listening, acknowledging, and responding to complaints (Harrison-Walker, 2001; Lee and Lee, 2006); delaying responses when appropriate (Mattila et al., 2013; Chang et al., 2015); taking legal actions or forming partnerships with customers (Thomas et al.,2012); using Web care for damage control (Fournier and Avery, 2011; Van Noort and Willemsen, 2012); and establishing long-term engagement mechanisms to maintain customer relationships (Pfeffer et al., 2014; Schwab, 2015). By following these intervention strategies, employees were expected to apply the most effective complaint resolution techniques in practice.

However, despite their adaptability, manual approaches had significant limitations. Cui and others (2017) argue that these methods relied heavily on human labor, making them increasingly inefficient as the volume of online complaints grew exponentially. With the expansion of digital platforms and the surge in customer interactions, companies struggled to maintain timely and effective complaint handling. This led to a fundamental shift in complaint management practices - from an effectiveness-oriented approach (focused on the quality of complaint resolution) to a responsiveness-oriented approach (prioritizing speed and efficiency).

Several researchers have highlighted the importance of responsiveness in modern complaint management. Handling complaints in a timely and efficient manner reduces resource waste and enhances customer satisfaction (Sultan, Abidin and Abdullah, 2008; Dey, Hariharan, and Ho, 2009). Dey et al. (2009) suggest that speed and effectiveness are the two key components of a responsive complaint handling system. Failure to provide timely responses has been identified as one of the most common customer grievances in online complaint management (Sultan, Abidin and Abdullah, 2008). This shift in focus has paved the way for the integration of technology-driven complaint resolution mechanisms, which will be explored in subsequent sections.

3.2 Intelligent Complaint Management: A Data-Driven Approach

Encouraging customers to complain, ensuring accessible complaining channels, and providing prompt responses to complaints are the critical elements in constructing a responsive complaint handling system (Coussement and Van den Poel, 2008; Sultan, Abidin and Abdullah, 2008). A well-structured complaint system not only helps businesses resolve individual disputes but also serves as an opportunity to identify systemic service issues and enhance overall service quality. However, as the volume of complaints continues to grow - particularly with the rise of online platforms - traditional complaint-handling methods have struggled to keep pace with customer expectations for speed and accuracy.

In response to these challenges, researchers have explored the integration of automation and artificial intelligence (AI) to improve complaint handling efficiency. Over recent years, efforts have been directed toward enhancing responsiveness by optimizing daily complaint management workflows through automated systems that leverage electronic computation technologies (Banga and Peddireddy, 2023). These intelligent approaches aim to reduce human workload, increase operational efficiency, and ensure consistency in complaint resolution. However, while automation brings significant advantages, it also introduces concerns regarding fairness, transparency, and the depersonalization of customer interactions.

Broadly, intelligent complaint management approaches can be categorized into text mining approaches and non-text mining approaches. Text mining techniques primarily focus on extracting meaningful insights from customer complaints expressed in textual formats, such as emails, online reviews, and social media posts. These methods leverage natural language processing (NLP), sentiment analysis, and topic modelling to classify complaints, detect recurring issues, and prioritize urgent cases. On the other hand, non-text mining techniques utilize structured data sources - such as complaint logs, customer demographics, and service usage patterns - to predict customer dissatisfaction trends and optimize response strategies.

The following sections will further explore these two categories of intelligent complaint management. Section 3.3 examines non-text mining techniques, including predictive analytics and statistical modelling, while Section 3.4 delves into the role of text mining in extracting valuable insights from unstructured complaint data. Through this analysis, this chapter aims to highlight how businesses can leverage both textual and non-textual data to enhance their complaint management strategies.

3.3 Non-Text Mining Techniques in Complaint Management

The non-text mining approach refers to techniques that do not involve direct analysis of textual data in their methodology. In the context of complaint management, these approaches do not rely on processing raw textual complaints but instead leverage structured data and computational applications to reduce human intervention in handling complaints. This method aims to streamline complaint management by automating decision-making processes based on predefined parameters rather than linguistic interpretation.

One example of a non-text mining approach is the development of rule-based automated complaint handling systems. For example, Zirtiloğlu and Yolum (2008) introduced an online complaint platform designed to improve government responsiveness to citizen complaints. Instead of submitting free-text complaints, users were required to enter specific complaint attributes into structured forms. A predefined set of constraints determined the relative weight of attributes in a keyword database, allowing the system to rank complaints based on urgency, priority, and severity. By ensuring that the most critical complaints receive immediate attention, such systems aim to enhance response efficiency. Similar structured complaint management frameworks have been adopted in customer service settings, particularly in large-scale industries where manual complaint triaging is impractical.

Another approach in non-text mining is the implementation of intelligent self-assessment platforms. Galitsky, González and Chesñevar (2009) proposed a complaint processing platform that enables customers to evaluate the validity of their complaints before submission. This system provides structured interactive forms where users characterize complaint-related communicative actions, which are then compared against past labeled complaint records stored in a database. If logical inconsistencies or vague structures are detected, customers are prompted to revise their submissions, thereby reducing the number of unreasonable or invalid complaints. By filtering out non-actionable complaints before they reach service representatives, such platforms help organizations optimize complaint resolution efficiency.

Although non-text mining approaches offer significant advantages in improving complaint handling efficiency, hey also present notable limitations. By excluding textual complaint content, these methods may fail to fully capture the nuances of customer grievances, potentially leading to an incomplete representation of the complaint context (Cui et al., 2017). Structured data alone may not always reflect the emotions, urgency, or implicit meanings conveyed in natural language complaints, resulting in intelligence gaps in automated decision-making. For this reason, text mining techniques - which incorporate linguistic analysis - have gained increasing attention in complaint management research over the past decade. The next section explores how text mining techniques address these limitations by extracting insights from unstructured complaint data.

3.4 Text Mining for Enhanced Complaint Handling

The goal of text mining approaches is to extract valuable information and patterns from textual data by leveraging appropriate text representation techniques. These methods numerically encode text documents through extracted features, making them mathematically computable for further analysis (Liu and Özsu, 2009). Text mining is often combined with machine learning to generate inferences from historical text data, enabling applications such as text clustering, categorization, and information retrieval. In recent years, advances in natural language processing (NLP) and deep learning have significantly expanded the capabilities of text mining, allowing for more nuanced and context-aware analyses of textual data (Mikolov et al., 2013; Devlin et al., 2019).

In the field of complaint management, text-mining-based approaches have been extensively explored to enhance various aspects of handling consumer grievances. Four primary tasks have emerged as focal points in text-driven compliant management strategies:

1) filtering complaints from non-complaints.
2) detecting negative reviews on social media.
3) complaint topic classification.
4) case retrieval for solutions. A discussion on each of these tasks will be provided in the next section.

Each of these tasks presents unique challenges and opportunities, as discussed in the following sections. The next section (3.4.1) explores automated methods for distinguishing complaints from general queries in email-based communication, a crucial first step in improving customer response efficiency.

3.4.1 Differentiating Complaints from Queries: Automated Email Classification

Over recent years, companies have received an increasing volume of emails containing both customer queries and formal complaints. Efficiently distinguishing complaints from general inquiries is crucial for ensuring that customer grievances are addressed promptly, thereby maintaining high responsiveness levels and customer satisfaction (Coussement and Van den Poel, 2008). However, traditional email filtering processes often rely on manual categorization, which is time-consuming and labour-intensive (Chumwatana and Chuaychoo, 2016). To mitigate these challenges, researchers have developed machine learning-based text classification methods to automate this task.

One of the early contributions to automated email classification in complaint management was introduced by Coussement and Van den Poel (2008). Their approach utilized Vector Space Modelling (VSM), a widely adopted technique in text categorization, to numerically represent textual features in documents. In VSM, text documents are transformed into vector spaces based on term frequency, allowing similar documents to be grouped together for classification purposes (Salton and Buckley, 1988). However, a major limitation of conventional VSM is the high dimensionality of generated vectors, which leads to increased computational resource consumption. To address this, Coussement and Van den Poel (2008) integrated Latent Semantic Indexing (LSI), reducing the dimensionality of the term space by grouping semantically similar words into conceptual clusters. Additionally, linguistic features such as verb tense frequencies were introduced to enhance classification accuracy, demonstrating that linguistic attributes can improve machine learning-based complaint detection.

While VSM and LSI have proven effective for English-language complaint classification, their applicability to non-English languages remains a challenge.
Chumwatana and Chuaychoo (2016) highlighted the limitations of conventional vector space models when applied to languages like Thai, where words are not explicitly separated by whitespace. To adapt to this challenge, their study employed n-gram extraction techniques, which segment text into continuous character sequences to form linguistic units. By implementing a 3-gram model, the system was able to effectively classify Thai-language complaints, reducing reliance on manual labour. However, the generalizability of this approach remains questionable due to the small-scale dataset (only 210 emails) used for evaluation.

Recent advancements in deep learning have further improved automated email classification by leveraging neural networks for feature extraction and pattern recognition. For example, transformer-based models such as Bidirectional Encoder Representations from Transformers (BERT) have outperformed traditional text classification techniques by capturing contextual word relationships within complaint texts (Devlin et al., 2019). Pre-trained language models like BERT can be fine-tuned on domain-specific datasets, enabling more precise complaint categorization with reduced reliance on manually engineered linguistic features (Sun, Huang, and Qiu, 2021). Additionally, hybrid deep learning architectures combining convolutional neural networks (CNNs) and recurrent neural networks (RNNs) have shown promise in classifying multilingual complaint datasets with improved accuracy.

Despite these advancements, automated email classification still faces several challenges. First, classifying customer complaints requires balancing accuracy with computational efficiency, as deep learning models often demand substantial processing power. Second, linguistic variations, sarcasm, and implicit complaints remain difficult to detect, highlighting the need for continual improvements in natural language processing (NLP) techniques. Third, privacy concerns regarding the automated analysis of customer emails must be addressed, as regulatory frameworks such as the General Data Protection Regulation (GDPR) impose strict guidelines on data handling in automated decision-making processes.

Overall, the evolution of machine learning in complaint classification has significantly enhanced response efficiency and accuracy. However, further research is required to develop models that are more interpretable, scalable, and ethically responsible. The next section (3.4.2) extends this discussion by examining how text mining techniques are applied to detecting negative reviews on social media, another crucial domain in modern complaint management.

3.4.2 Social Media and Negative Review Detection

The widespread adoption of social media platforms has led to an unprecedented surge in online reviews, posing both opportunities and challenges for businesses in managing customer complaints. Unlike traditional complaint systems, where grievances are filed directly with companies, social media enables customers to voice their dissatisfaction publicly, potentially influencing the perceptions of a vast audience (Yang et al., 2011). As a result, organizations must develop robust mechanisms for detecting and responding to negative reviews in a timely manner to mitigate reputational risks.

However, the informal nature of social media content, characterized by slang, abbreviations, and emotive language, makes automated review classification a complex task (Zimbra et al. 2009).

One of the early challenges in detecting negative reviews was the reliance on supervised machine learning models, which require large-scale labeled datasets for training. While email classification (as discussed in Section 3.4.1) can leverage pre-labeled data from customer service logs, social media reviews often lack structured labels, making fully supervised learning infeasible (Jin, Yan, Yu and Li, 2013). To address this issue, researchers have explored semi-supervised and distant supervision approaches, which infer labels from a small set of manually annotated examples. For example, Yang, Hsu, and Tan (2011) developed a partially supervised classifier that infers class labels for unlabeled reviews by measuring their similarity to labeled examples using the Rocchio algorithm. Similarly, Jin and colleagues (2013) applied similarity-based heuristics, such as Euclidean distance and Cosine coefficient, to expand training datasets from limited manually labeled samples.

Another prominent method in social media complaint detection is the use of lexicon-based approaches combined with domain-specific heuristics. Instead of relying on labeled datasets, these methods leverage pre-defined sentiment lexicons and rule-based classifiers to identify negative expressions. For instance, distant supervision techniques automatically generate training data by matching words in social media posts with dissatisfaction-related terms from domain-specific dictionaries (Fung, Yu, and Lu, 2006). By extracting phrasal relations associated with dissatisfaction—such as negation patterns and complaint-related clauses—these models can label large-scale data without requiring extensive manual annotation. However, lexicon-based methods face limitations in handling sarcasm, contextual polarity shifts, and evolving linguistic trends, necessitating more adaptive machine learning approaches (Manochandar and Punniyamoorthy, 2018).

In recent years, deep learning has significantly enhanced the accuracy of negative review detection on social media. Recurrent neural networks (RNNs) and long short-term memory (LSTM) models have been widely used to capture sequential dependencies in textual data, allowing for better context recognition in sentiment analysis (Zhang, Zhao, and LeCun, 2015). Additionally, transformer-based architectures such as BERT (Bidirectional Encoder Representations from Transformers) have demonstrated superior performance in classifying informal text, as they consider bidirectional word relationships rather than relying on predefined phrase structures (Devlin et al., 2019).

Despite these advancements, automated social media complaint detection still presents unresolved challenges. First, models trained on one platform (e.g., Twitter) may not generalize well to other platforms (e.g., Facebook or Instagram), due to differences in linguistic styles and character limitations. Second, real-time complaint detection requires high computational efficiency, which can be costly for organizations managing large-scale customer feedback streams (Sun, Qiu, and Huang, 2021). Third, ethical concerns surrounding AI-driven review monitoring - including data privacy, bias in sentiment analysis, and the potential suppression of negative feedback—must be addressed to ensure fair and responsible implementation.

As businesses increasingly rely on automated techniques for monitoring social media sentiment, striking a balance between efficiency, accuracy, and ethical considerations remains a critical challenge. The next section (3.4.3) explores the role of automated topic classification in streamlining complaint resolution, providing deeper insights into how businesses can categorize customer grievances more effectively.

3.4.3 Automating Complaint Topic Classification

One of the most essential tasks in complaint handling is the accurate classification of complaints based on their underlying topics. Properly categorizing complaints ensures that cases are routed to the correct departments, enabling a more efficient resolution process and improving customer satisfaction (Yang, Xu, Yang and Chen, 2018; Thomas, 2018). However, manual classification is impractical due to the high volume of daily customer complaints, which often exhibit informal language, abbreviations, and noise. As a result, researchers have explored automated complaint topic classification methods to enhance efficiency and reduce human effort.

Early approaches to complaint topic classification primarily relied on rule-based keyword matching techniques. For instance, Sultan, Abidin and Abdullah (2008) developed a keyword-based system in which predefined dictionaries were used to match complaint texts to specific categories. While this approach reduced manual effort, its effectiveness was constrained by the limited coverage of keyword databases, often failing to recognize variations in linguistic expressions. Additionally, such methods struggled with polysemy—cases where words have multiple meanings depending on context—leading to classification errors (Stoica and Özyirmidokuz, 2015).

To overcome these limitations, researchers have turned to machine learning-based classification models, which learn patterns from labelled complaint datasets and generalize to unseen cases. One such approach is the use of deep learning models, particularly word embeddings, to capture semantic relationships between complaint terms. For instance, Thomas (2018) proposed a deep learning framework utilising Word Embedding methods, where terms and phrases from a corpus are mapped into vector spaces. Neural networks are then trained to classify complaints based on contextual word similarities, demonstrating moderate classification accuracy. However, traditional word embedding models such as Word2Vec and GloVe struggle with capturing the polysemous nature of words, as they assign a single vector representation to each term, disregarding its contextual variations (Mikolov, Chen, Corrado, and Dean, 2013; Pennington, Socher, and Manning, 2014).

Recent advancements in natural language processing (NLP) have further improved complaint topic classification. Transformer-based architectures such as Bidirectional Encoder Representations from Transformers (BERT) offer substantial improvements by considering bidirectional contextual dependencies, allowing for more precise classification of complaints with ambiguous meanings (Devlin, Chang, Lee, and Toutanova, 2019). Additionally, attention mechanisms within transformer models enhance topic recognition by dynamically weighting the importance of different words in complaint texts (Vaswani, Shazeer, Parmar, and Uszkoreit, 2017).

Beyond traditional single-topic classification, researchers have also emphasized the importance of multi-label complaint classification. Dasgupta, Dey, and Verma (2016) highlighted that many customer complaints pertain to multiple aspects of a service, necessitating classification models that assign multiple relevant topics to a single complaint. To address this, Dasgupta and colleagues explored various text representation methods, including Vector Space Modelling, Latent Semantic Analysis, and Pointwise Mutual Information, to reflect linguistic nuances in complaint texts. Their study employed a Fuzzy K-Nearest Neighbours algorithm to compute the probabilities of complaints belonging to multiple categories, enabling a more flexible classification framework. Recent studies have further refined this approach by incorporating graph-based neural networks, which model hierarchical relationships between complaint topics and enhance classification granularity.

Despite these advancements, automated complaint topic classification still faces several challenges. First, domain adaptation remains a critical issue—classification models trained on one industry (e.g., telecommunications) may not generalize well to another (e.g., banking) due to differences in complaint terminology and structure (Howard and Ruder, 2018). Second, informal writing styles, sarcasm, and implicit complaints make it difficult for machine learning models to extract relevant topics, requiring continual refinements in NLP techniques. Third, ethical concerns arise regarding transparency in AI-driven classification, as opaque deep learning models may struggle to provide explainable justifications for their classification decisions.

As complaint classification continues to evolve, businesses must strike a balance between automation, interpretability, and adaptability. The next section (3.4.4) explores how complaint retrieval systems leverage past cases to assist in decision-making, further enhancing service recovery strategies.

3.4.4 Case Retrieval for Solution Reference: Advancements in Complaint Handling

Due to the uncertain characteristics of customer behaviors and the complex nature of service issues, determining appropriate solutions to complaints remains a decision-making process that heavily relies on experienced employees. As noted by Lee, Wang and Trappey (2015), frontline service personnel play a crucial role in evaluating complaints, assessing past cases, and determining the most effective resolution strategies. However, the efficiency of complaint resolution is often constrained by its heavy dependence on human labor, particularly in cases requiring extensive reference to historical complaint records (Cui, Zhang, and Luo, 2017). To address these limitations, researchers have explored automated case-based reasoning (CBR) systems that enable computers to derive solutions for new complaints by leveraging past case knowledge.

Early approaches to automated case retrieval primarily relied on keyword-based search mechanisms. For instance, Trappey, Lee, Chen and Trappey (2010) introduced a keyword-driven system for retrieving reference solutions in restaurant industry. Complainants input specific complaint attributes, which the system then matches against predefined keyword databases to suggest relevant past cases. While this approach facilitated efficient information retrieval, its effectiveness was questioned due to its reliance on surface-level keyword matching, which often failed to capture the deeper semantic relationships between terms (Zhang, et al., 2013). Moreover, keyword-based retrieval systems struggled with cases where the same word carried multiple meanings depending on context, leading to inaccurate recommendations (Cui et al., 2017)

Recognising the limitations of keyword-based retrieval, researchers have increasingly turned to ontology-based methods to enhance case retrieval accuracy. Ontology, as defined by Gruber (1995), is "*an explicit specification of a conceptualisation*", providing structured representations of domain knowledge. By organizing domain-specific terms into hierarchical relationships and defining their interconnections, ontology-based models improve the contextual understanding of complaints (McGuinness and Van Harmelen, 2004). For example, Rodríguez and Egenhofer

(2003) proposed an ontology framework that structures complaint-related terminology into a semantic hierarchy, enabling more accurate case retrieval. Yan and others (2006) further suggested that well-constructed ontologies could enhance the processing, storage, and representation of complaint data, ultimately supporting intelligent decision-making.

Several studies have developed customized ontology-based case retrieval systems tailored to specific service industries. Lee, Wang, and Trappey (2015) and Cui, Zhang, and Luo (2017) constructed domain-specific ontologies for modelling customer complaints in high-end restaurant services. Their approach involved six key steps: (1) splitting text data into words; (2) removing stop words; (3) dividing relational hierarchies where grouping semantically similar terms in a synonym set; (4) defining the relations between upper and lower hierarchies; (5) coding ontology; and (6) integrating with existing generic ontologies (Uschold and Gruninger, 1996). Lee and others (2015) conducted in-depth interviews with domain experts to construct an ontology hierarchy that accurately reflects real-world complaint scenarios.

Once an ontology-based retrieval system is in place, new complaints can be analysed using feature extraction techniques that match complaint attributes with structured ontology terms. Incoming complaints are mapped onto the ontology structure and compared with past cases to retrieve relevant solutions. Gan, Dou and Jiang (2013) proposed that an ontology-based retrieval mechanisms should incorporate semantic similarity measures to improve metrics case-matching accuracy. To achieve this, the semantic similarity between terms is often quantified using hierarchical distance metrics within the ontology tree model (Pedersen, Pakhomov, Patwardhan and Chute, 2007; Wang, Du, Payattakool, Yu and Chen, 2007).

Despite these advancements, case retrieval systems still face challenges. First, maintaining and updating ontologies requires continuous domain expertise, as new complaint types emerge over time. Second, ontology-based approaches may still struggle with handling informal language, misspellings, and sarcasm present in customer complaints, necessitating improvements in NLP preprocessing techniques.

As automated case retrieval continues to evolve, the integration of ontology-based reasoning with deep learning techniques holds significant promise for improving complaint resolution efficiency.

3.5 Chapter summary

This chapter has explored the transformation of complaint management from manual, labour-intensive processes to AI-driven solutions, emphasizing the role of machine learning and text mining in enhancing efficiency and decision-making. These advancements have significantly improved the speed and accuracy of complaint classification and response, reducing reliance on human intervention. However, automation also introduces challenges, including fairness in decision-making, model interpretability, and the scalability of AI-driven systems across different industries.

A critical issue is balancing structured and unstructured data in complaint management. Non-text mining techniques efficiently process structured complaint records but often fail to capture the complexity of customer grievances. In contrast, text mining approaches enhance classification and sentiment analysis by extracting deeper insights from unstructured data, yet they require substantial computational resources and large-scale labelled datasets. Additionally, as AI-driven systems become more prevalent, businesses must address potential risks such as algorithmic bias, over-reliance on historical data, and challenges in adapting models to evolving complaint patterns. Navigating these trade-offs is essential for ensuring that automated complaint management remains both effective and trustworthy.

The next chapter builds upon these advancements by introducing multimodal data fusion, which integrates diverse data sources - particularly textual and structured data - within a unified analytical framework. This approach not only refines complaint management strategies but also strengthens customer churn prediction models, offering a more comprehensive understanding of customer behaviours. By leveraging advanced machine learning and deep learning techniques, Chapter 4 explores how businesses can enhance service recovery and customer retention through a more holistic, data-driven perspective.

4. Enhancing Complaint Management and Churn Prediction through Multimodal Data Fusion

As concluded in the previous chapter, big data technologies have transformed complaint management, shifting from manual methods to intelligent, data-driven approaches. The integration of advanced analytics, particularly text mining, has significantly improved the ability to process and respond to customer complaints efficiently. However, while these techniques enhance individual aspects of complaint handling, they remain largely limited to analysing either structured or unstructured data in isolation.

This chapter introduces multimodal data fusion as a novel approach to overcoming this limitation. By integrating multiple data types - especially textual and structured data—within a unified analytical framework, multimodal learning can provide a more holistic view of customer behaviour and complaint patterns. The application of machine learning and deep learning models in this context not only enhances the predictive power of complaint management systems but also plays a crucial role in customer churn prediction (CCP).

The chapter is structured as follows: First, we examine the role of textual data analysis in complaint management and churn prediction, exploring its strengths and limitations. Next, we introduce multimodal data fusion, outlining its key methodologies and applications in complaint handling. The subsequent sections discuss different fusion strategies, including early, intermediate, and late fusion techniques, highlighting their comparative advantages and challenges.

4.1 Textual Data Analysis in Churn Prediction and Complaint Management

Customer complaints serve as a valuable source of information, offering companies insights into consumer dissatisfaction and potential churn risks. While structured complaint data plays an increasingly significant role in customer churn prediction (CCP), textual complaint data has been widely used in developing automated complaint management systems. However, despite the increasing availability of textual data, research that effectively integrates textual complaints into CCP frameworks remains scarce. This section provides an overview of related studies on textual complaint analysis and its role in both complaint management and churn prediction.

Previous research has explored textual complaint analysis across multiple domain, including complaint identification (Jin and Aletras, 2020), topic categorisation (Forster and Entrup, 2017), and risks or escalation assessment (Yang et al., 2019). These applications predominately rely on supervised machine learning models that extract features from complaint text, such as bag-of-words, topic distributions, or dictionary-based features. Despite extensive research on text representation techniques in natural language processing (NLP), there is no universal agreement on the optimal method, as its suitability varies by application. However, vector space modelling remains one of the most widely used approaches in complaint management due to its interpretability, ease of implementation, and competitive performance (De Caigny et al., 2020; Geiler et al., 2022). This method represents words as independent vectors in a predefined vocabulary space, disregarding word order and syntactic relationships (Xiong et al., 2020).

Despite the demonstrated value of textual data in complaint handling, its integration into CCP framework has historically been limited, with most studies focusing on structured information as the primary data source (De Caigny et al., 2020). As analytics techniques have advanced, researchers have increasingly recognized the potential of incorporating unstructured textual data to enhance predictive accuracy (Benoit and Van den Poel, 2012; Tang, Thomas, Fletcher, Pan, and Marshall, 2014). While textual data from customer reviews, emails, and social media posts is widely available for market research, its use in CRM and CCP remains underdeveloped (Kumar and Ravi, 2016). Given the growing need for comprehensive churn analysis, combining multiple data sources—including text analytics—has become a key research focus (Shirazi and Mohammadi, 2019).

Recent studies have explored novel frameworks for integrating processed textual data into CCP models alongside structured variables.For example, De Caigny and colleagues (2020) developed a method that processes email communications between customers and companies, transforming textual data into a structured format before integrating it with other customer attributes in a CCP model. Their results indicate that incorporating textual data improved churn prediction performance by 2%. However, many CCP studies remain focused on structured variables, overlooking the unique insights that textual complaints can provide. In contrast, the current study analyzes a dataset containing textual complaints from a broad range of industries, rather than relying solely on structured variables. By leveraging textual complaints alongside key structured features, this research aims to characterize retained and churned customers more comprehensively, thereby identifying complaint-related attributes that can enhance service recovery strategies.

Having established the significance of text mining in complaint management, this discussion lays the groundwork for a broader analytical framework. The subsequent sections transition from unimodal textual analysis to multimodal data fusion, where the integration of diverse data modalities offers a more comprehensive and refined approach to churn prediction and complaint resolution. This shift represents a move from isolated text-based insights to a holistic, data-driven strategy that maximizes the potential of big data in customer service intelligence.

4.2 Multimodal Data Fusion in Complaint Analysis

With the increasing prevalence of machine learning and deep learning applications, there is a growing demand for scalable and practical decision-support tools (Kiela et al., 2018). Traditional unimodal approaches, which rely on a single source of information, often fail to capture the complexity of real-world decision-making. Observing human cognitive processes, it is evident that individuals rarely base their decisions on a single data source. This has led to the development of multimodal data fusion techniques, which integrate multiple data modalities—such as text, images, videos, and tabular data—to improve analytical robustness and predictive performance. While each modality provides unique value, it also comes with inherent limitations, making fusion methods essential for extracting more comprehensive insights (Boulahia et al., 2021).

A straightforward example of the necessity for multimodal fusion can be seen in emotion and sentiment analysis. While written text conveys detailed information about an individual's perception, it often struggles to capture abstract concepts such as tone or facial expressions, which can be more effectively analysed through visual or auditory modalities. Similarly, tabular data efficiently represent structured information, such as customer ratings or transactional records, but lacks the contextual depth of unstructured data like customer reviews or complaints. By integrating multiple modalities, multimodal learning enhances analytical reliability and robustness, allowing for a more nuanced understanding of complex interactions.

To fully exploit the potential of multimodal datasets, researchers have increasingly focused on developing effective fusion strategies (e.g., Yang et al., 2019; Kim et al., 2019; Prakash and Madabushi, 2020; Rahman et al., 2020; Gu and Budhkar, 2021; Qu et al., 2022; Borisov et al., 2022). Rather than analyzing textual data in isolation, recent studies have explored how combining textual, visual, and tabular information can enhance predictive performance in tasks such as complaint classification, customer sentiment analysis, and churn prediction. One of the key advantages of multimodal learning is its ability to leverage both complementary and correlational information. Complementary information refers to unique insights provided by each modality that would be absent if only a single data source were considered (Boulahia et al., 2021). Correlational information, on the other hand, enhances model performance by capturing the interdependencies between different data sources, strengthening overall prediction accuracy.

Multimodal learning also aligns with real-world decision-making processes, where humans naturally integrate diverse information sources - such as textual reports, visual cues, and demographic data - to make informed judgments. This parallels the goal of artificial intelligence: developing machines capable of making human-like decisions based on multiple data streams. However, despite its advantages, multimodal learning presents several challenges, including data synchronization, modality alignment, and computational complexity. Effective fusion strategies are required to address these issues and ensure seamless integration of diverse data types.

Deep architectures have played a crucial role in advancing multimodal fusion, enabling flexible and adaptive integration of different data modalities (Ramachandram and Taylor, 2017). Existing fusion methods are commonly categorized based on the stage at which data integration occurs (Boulahia et al., 2021). There are three predominant strategies for multimodal fusion: 1) early fusion; 2) intermediate fusion; and 3) late fusion. Before the rise of deep learning, multimodal fusion was typically divided into feature-level fusion and decision-level fusion (i.e., late fusion). However, with the advent of deep neural networks, feature-level fusion has been further refined into distinct early and intermediate fusion techniques, providing greater flexibility in handling complex multimodal relationships. The following subsection explores these fusion strategies in greater depth, highlighting their respective advantages and limitations in complaint analysis.

4.2.1 Early fusion

Early fusion involves integrating multiple raw data modalities at the input level, forming a unified representation before any learning phase takes place. Ramachandram and Taylor (2017) define early fusion as the process of combining multiple data sources into a single multimodal feature representation, which is then fed into a model as a single input. A key advantage of early fusion is that it requires only a single learning phase, reducing computational overhead by minimizing additional training time, even when multiple modalities are involved (Boulahia et al., 2021).

To incorporate multiple data modalities into a single model, data conversion is often required to ensure compatibility between different feature types. Without appropriate transformation, disparate data formats—such as textual, image-based, or tabular inputs—may not be effectively integrated. For instance, merging textual and visual data presents challenges if fusion is implemented without extracting meaningful features from both sources. In previous studies, Poria et al. (2016) and Zadeh et al (2017) explored the concatenation of higher-level feature vectors from text, vision, and acoustic modalities for multimodal sentiment analysis. While this approach is conceptually simple, early fusion often fails to maximize the complementary information across modalities. A major drawback is the exponential increase in dimensionality and computational complexity caused by concatenation-based fusion, which can lead to inefficiencies in large-scale models. Furthermore, deep learning architectures employing early fusion are generally inflexible to missing data, as the predefined input size requires the presence of all modalities. Since real-world datasets often contain missing modalities, this limitation can hinder model performance, particularly in applications where some data sources may be unavailable. A graphical representation of a concatenation-based fusion method is presented in Figure 4.1.

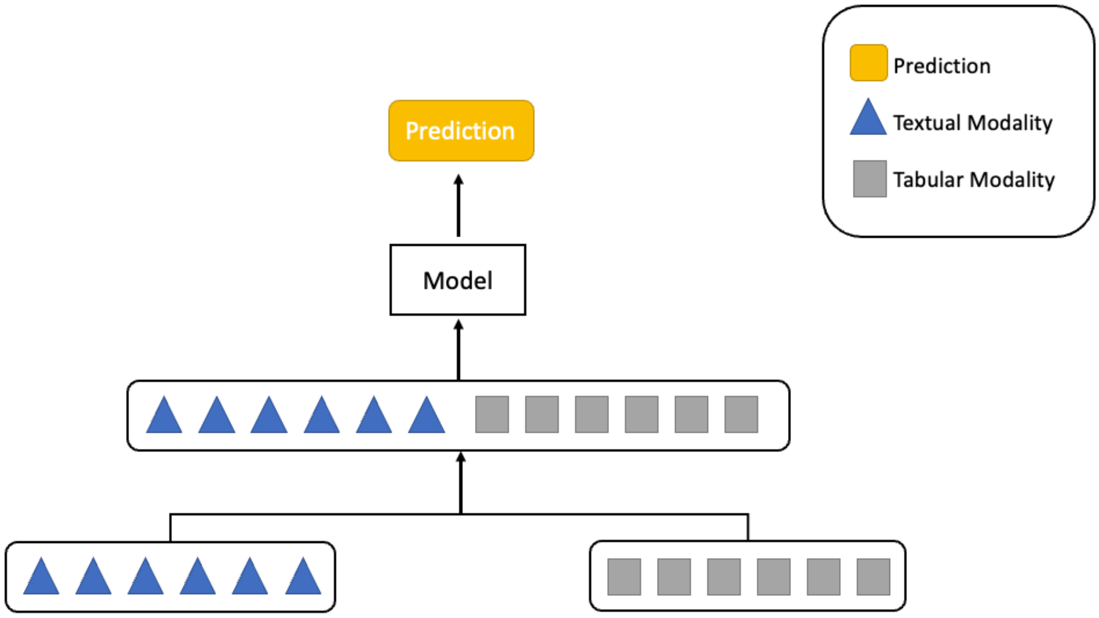

Figure 4.1 A graphical representation of a concatenation-based fusion method.

Another emerging research area explores the integration of textual and tabular data for classification tasks. Gu and Budhkar (2021) proposed an early fusion approach that simplifies the combination of human language and structured information while maintaining competitive performance. This method, referred to as 'Unimodal,' transforms structured data into textual tokens by concatenating tabular values with the corresponding textual data sample. The resulting unified text sequence is then processed using natural language processing (NLP) techniques. This approach not only streamlines the preprocessing of categorical and numerical data but also addresses inefficiencies in traditional fusion methods by making structured information semantically accessible to state-of-the-art NLP models. Despite its advantages, this transformation-based early fusion method may introduce noise if the textual representation of structured data lacks contextual relevance or if the integration process distorts the inherent relationships within tabular features.

4.2.2 Intermediate fusion

With the rise of the deep learning architectures, intermediate fusion has emerged as a powerful approach to multimodal data integration (Ramachandram and Taylor, 2017). Unlike early fusion, which directly combines raw input data, intermediate fusion transforms different modalities into high-level feature representations before merging them within a shared representation layer. This method allows for separate learning of modality-specific features while ensuring compatibility across data sources. Once feature vectors are extracted, they are fused within a dedicated layer of a subsequent learning model to form a joint multimodal representation (Boulahia et al., 2021). Figure 4.2 illustrates a typical intermediate fusion framework, where a shared representation layer connects to multiple modality-specific feature extractors.

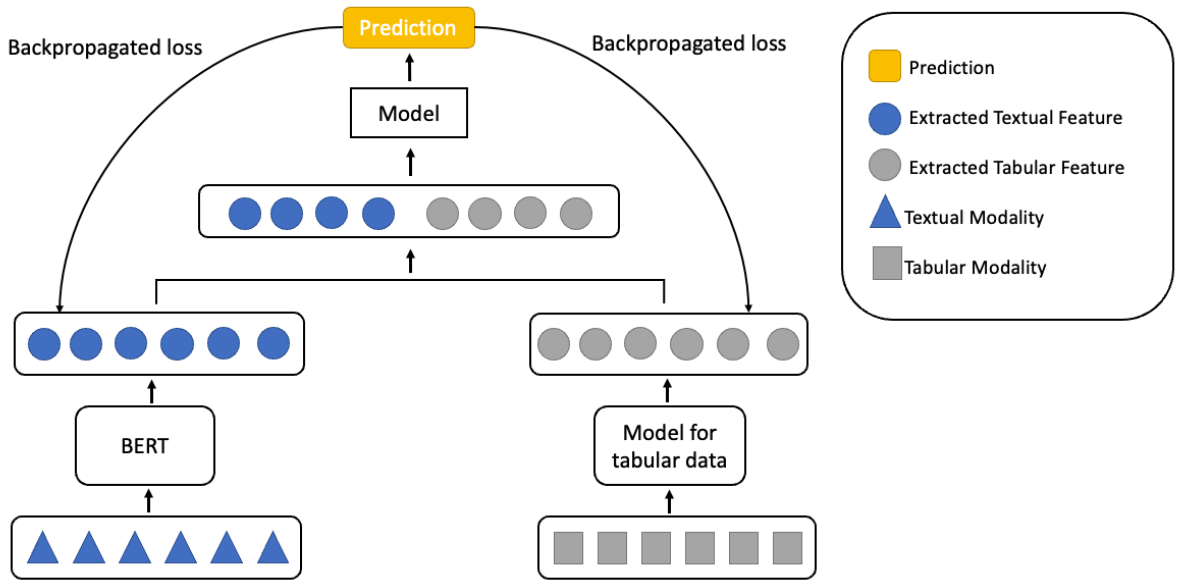

Figure 4.2 Intermediate fusion with a shared representation layer.

One widely studied approach to intermediate fusion is the Multimodal Adaptation Gate (MAG), proposed by Rahman and others (2020). This technique applies a cross-modality gating attention mechanism to Transformer-based architectures, integrating textual, visual, and acoustic modalities at the feature level. Since MAG processes data through pretrained modality-specific models, it enables fine-tuning for downstream tasks while fully leveraging information from each data source. A key advantage of MAG is its ability to handle missing modality, as the gating attention mechanism dynamically adjust weight assignments based on the importance and availability of each input model. This makes intermediate fusion more robust in real-world applications where data completeness cannot always be guaranteed.

Intermediate fusion architectures can also be structured in a stepwise manner, where the output of one model serves as the input for another (Figure 4.3). This hierarchical design allows correlational relationships between different modalities to be captured progressively, improving overall model expressiveness. Moreover, because the learning process is conducted in stages, error signals from later models can be backpropagated to earlier layers, optimizing the entire multimodal framework (Ramachandram and Taylor, 2017). Compared to early fusion, intermediate fusion offers greater flexibility in determining when and how multimodal representations should be integrated. However, this flexibility comes at the cost of increased architectural complexity, requiring meticulous design and additional computational resources.

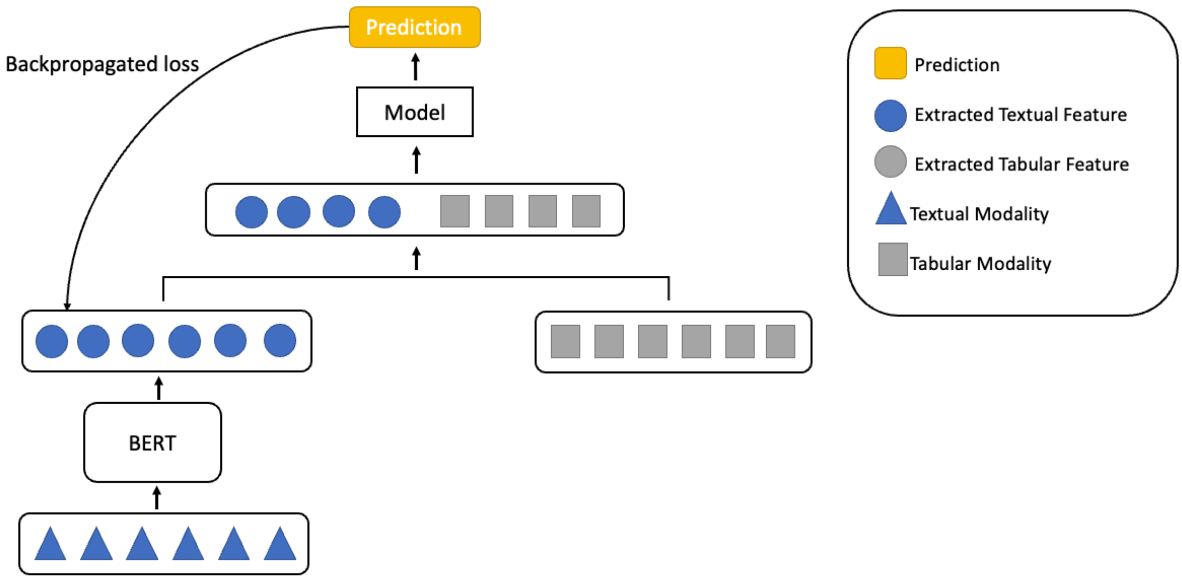

Figure 4.3 Intermediate fusion with a stepwise structure.

A practical application of intermediate fusion in customer analytics is demonstrated in the work of De Caigny and others (2020), who developed a multimodal customer churn prediction system using a stepwise structure. Their approach first applied convolutional feature extraction to textual data, generating highly expressive representations, which were then fed into a feed-forward neural network for classification. Their findings suggest that, within a churn prediction context, intermediate fusion outperforms traditional text-based techniques such as term frequency-inverse document frequency (TF-IDF), particularly when dealing with complex, high-dimensional data.

Despite its advantages, intermediate fusion presents several challenges. The need for careful architectural design makes it less straightforward to implement than early fusion, particularly for tasks requiring real-time processing. Additionally, while the stepwise structure enhances model adaptability, it also increases computational overhead and may require larger training datasets to achieve optimal performance. As research in multimodal learning progresses, finding efficient and scalable intermediate fusion strategies remains a key area of exploration, particularly in domains such as complaint analysis and customer retention modelling.

4.2.3 Late fusion

Late fusion, also known as decision-level fusion, integrates the outputs of multiple independent classifiers, aggregating their decisions to produce a final prediction (Figure 4.4). Unlike early and intermediate fusion, where different modalities are combined at the feature or representation level, late fusion processes each modality separately through its own learning pipeline. The number of classifiers in this framework typically corresponds to the number of modalities, with each classifier making an independent prediction based on its respective data source (Ramachandram and Taylor, 2017). Conceptually, late fusion shares similarities with ensemble learning, as both strategies combine multiple models to enhance prediction accuracy. However, a key distinction is that ensemble learning typically operates on structured data from a single modality, whereas late fusion is designed to aggregate multimodal information at the decision level.

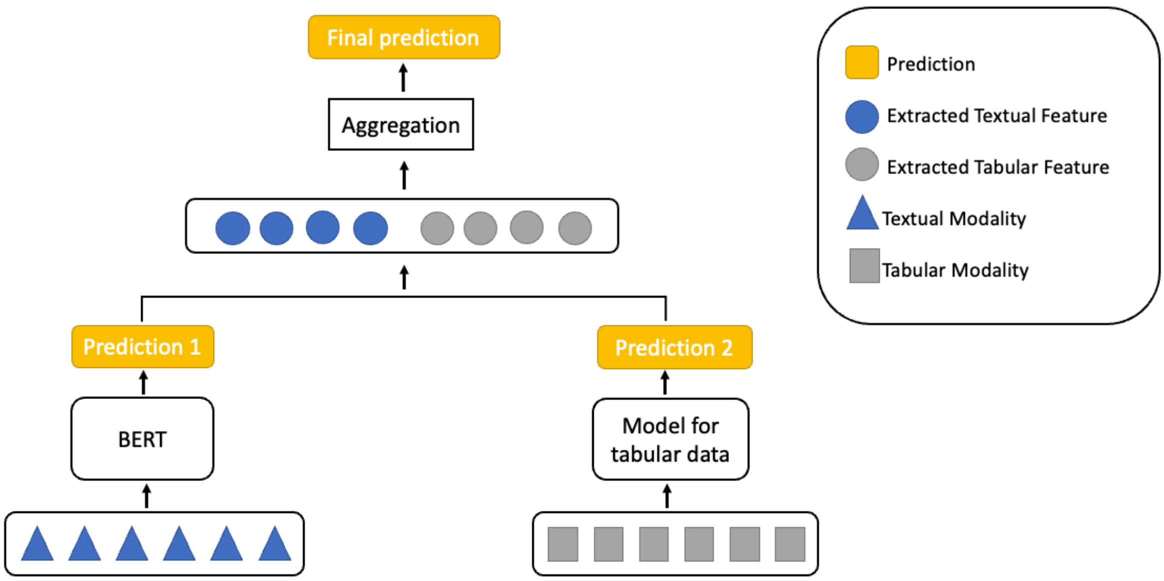

Figure 4.4 A graphical representation of late fusion with two modalities of data

Existing research on late fusion, such as the work by Tsai et al. (2019), primarily employs rule-based aggregation methods rather than deep learning-based decision fusion. Common strategies include majority voting, averaged-scores voting, weighted-scores voting, and Bayes' rule-based fusion (Boulahia et al., 2021). Since late fusion treats different modalities independently, it is often the preferred approach when dealing with a large number of data sources, as it allows for greater modularity and flexibility in model design.

One of the primary advantage of late fusion is its computational efficiency. Because cross-modal interaction occurs only at the final decision stage, this approach significantly reduces processing costs compared to early and intermediate fusion methods. Furthermore, the modular structure of late fusion enables models to be trained separately for each modality, allowing for greater adaptability to different types of data. This is particularly beneficial in real-world applications where data from various sources may not always be available simultaneously, as late fusion can still function effectively even when some modalities are missing.

However, the major limitation of late fusion lies in its restricted ability to capture interactions between modalities during the learning process. Unlike early and intermediate fusion, where multimodal features are integrated during model training, late fusion postpones cross-modal integration until the final prediction step. This may result in the model failing to fully exploit the complementary relationships between different data sources, potentially leading to suboptimal performance in tasks that rely heavily on inter-modal dependencies. For example, in sentiment analysis, textual and visual cues often interact in a complex manner, and late fusion may not capture these nuanced relationships as effectively as intermediate fusion approaches.

Despite its limitations, late fusion remains a widely used strategy in multimodal learning due to its scalability and computational efficiency. As research continues to evolve, hybrid fusion approaches that combine elements of early, intermediate, and late fusion may offer a promising direction for improving multimodal learning frameworks, particularly in applications such as complaint management and customer behavior prediction.

4.3 Chapter summary

This section summarizes the key considerations in selecting an appropriate multimodal fusion strategy, which depend on factors such as the types and number of modalities in the dataset, the downstream task, and available computational resources. Table 4.1 presents a comparative overview between different fusion strategies, highlighting their respective advantages and limitations.

Table 4.1 Attributes of multimodal fusion strategies

Attribute	**Early**	**Intermediate**	**Late**
Multiple learning pipelines required	×	✓	✓
A stepwise structure involved	×	✓	×
Interaction effects considered	✓	✓	×
Flexible to missing modality	×	✓	✓
Computationally expensive for memory	✓	✓	×
Time consuming	×	✓	✓

Building on the foundation established in the previous chapters, this chapter introduces multimodal data fusion as a pivotal advancement in complaint management and customer churn prediction. It explores both theoretical and practical applications of combining diverse data types through advanced computational techniques, demonstrating how these methods enhance the accuracy and efficiency of predictive models.

Overall, the first four chapters systematically address the intersection of customer service, data analytics, and predictive modelling. They trace the evolution from traditional complaint handling methods to sophisticated, data-driven approaches, culminating in the application of multimodal data fusion as a transformative tool in modern complaint management. This progression underscores the increasing role of big data technologies in optimizing service recovery and customer retention strategies.

The next chapter shifts focus to advanced textual data analysis techniques, delving deeper into specific methods for representing, processing, and extracting insights from textual complaints. By exploring both traditional and state-of-the-art text representation models, the upcoming discussion aims to bridge the gap between theoretical data fusion concepts and practical text-driven analytical frameworks.

5. Advanced Techniques in Textual Data Analysis

As the research focus shifts from multimodal data fusion to textual data analysis, this chapter explores advanced techniques for representing and processing textual data. While the previous chapter examined the integration of multiple data modalities, this chapter delves into text-specific methodologies that enhance complaint management and customer churn prediction. By moving from broad data integration concepts to more specialized text analysis techniques, this chapter aims to bridge the gap between theoretical data fusion frameworks and the practical applications in natural language processing (NLP).

The ability to effectively process and extract meaningful insights from textual complaints is crucial for service recovery and customer retention. Traditional text representation methods, such as Vector Space Models, have been widely used in early-stage text analytics but face limitations in capturing semantic and contextual information. Recent advancements, including word embeddings and deep learning-based representations, offer more sophisticated ways to encode textual data for predictive modelling. This chapter systematically examines these techniques, comparing their strengths and weaknesses in the context of automated complaint handling.

By evaluating different text representation strategies and their applications in predictive analytics, this chapter provides a foundation for developing more intelligent, data-driven complaint management systems. The following sections introduce and compare key text representation methods, discuss advanced feature extraction techniques, and assess their effectiveness in improving predictive modelling outcomes.

5.1 Analysis of Text Representation Methods

Textual data, by nature, is unstructured and cannot be directly processed by computational models. To enable mathematical modelling, raw text must be transformed into structured numerical representations, which presents a fundamental challenge in natural language processing (NLP) (Kowsari et al., 2019). The key difficulty lies in determining an optimal machine-readable representation that effectively captures latent semantic and contextual information embedded within human language (Wolfram and Zhang, 2008).

Text representation plays a pivotal role in text analysis frameworks, as it encodes textual content into numerical feature vectors, which can then be leveraged by machine learning algorithms for predictive modelling (Jindal, Malhotra and Jain, 2015). The effectiveness of such algorithms is highly dependent on the choice of text representation, as different approaches encode linguistic features in fundamentally different ways (Bengio et al., 2013).

Two major approaches to text representation dominate the literature: vector space modelling and word embedding vectors. Vector space models, such as TF-IDF, represent documents as high-dimensional sparse vectors, where each dimension corresponds to a unique term in the corpus. This method, though simple and computationally efficient, struggles with capturing semantic relationships between words. Conversely, word embedding methods encode words as dense vectors in a continuous space, capturing contextual similarities and semantic meanings through unsupervised learning on large text corpora. The key difference lies in their ability to model relationships between words and their computational efficiency. Figure 5.1 provides an overview of these two approaches.

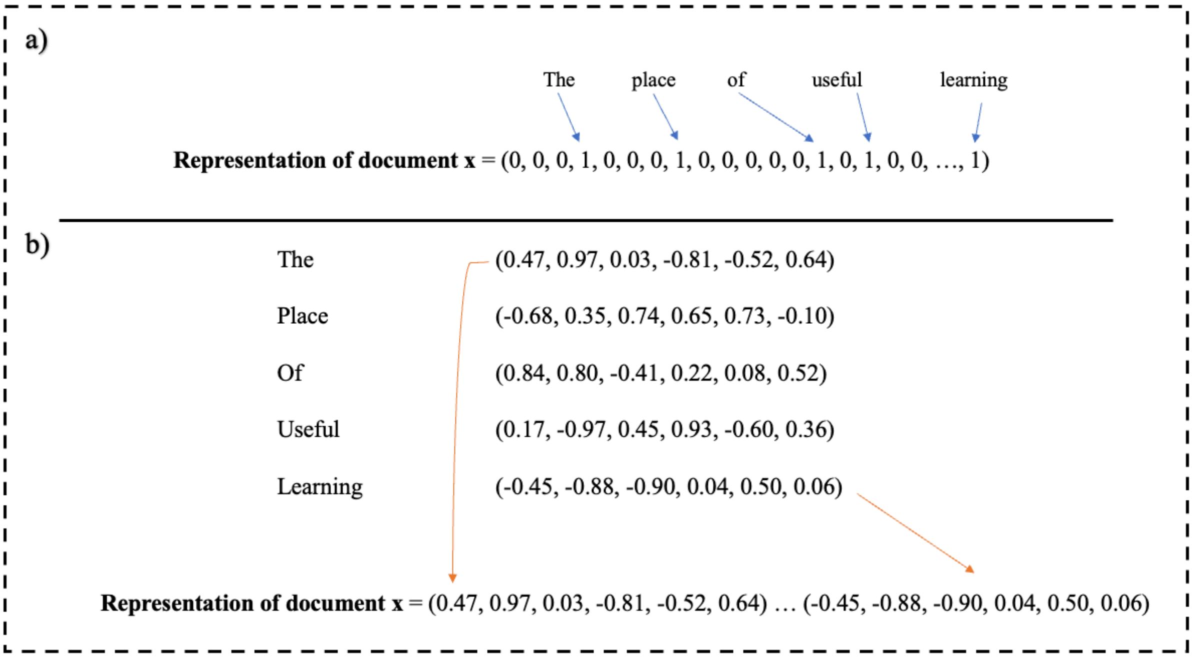

Figure 5.1 Graphical representation of the two text representation approaches

5.1.1 Vector Space Modelling

Vector space modelling represents textual data using sparse, high-dimensional vectors, where each document is transformed into a structured format that facilitates computational analysis (Mirończuk and Protasiewicz, 2018). This method typically involves multiple pre-processing steps to refine textual data before numerical representation, including: 1) cleansing for removing special characters that often appear at the beginning or at the end of a piece of text, 2) tokenisation for splitting the raw text documents into single words, 3) case conversion for turning all words into lower case, 4) filtering for removing words that are non-informative or occur just once or twice in the entire corpus, and 5) stemming for extracting the corresponding stem of each word in order to substantially reduce the number of words to be considered (Shahmirzadi, Lugowski, and Younge, 2019).

After pre-processing, a numerical representation of vector can be retrieved with the use of vector space modelling. A fundamental approach within vector space modelling is One-Hot encoding, where each document is represented as a binary vector indicating the presence or absence of words. However, this approach generates highly sparse vectors and fails to account for term importance. To address this, weighting schemes such as Term Frequency (TF) and Inverse Document Frequency (IDF) were introduced (Guzella and Caminhas, 2009). TF-IDF, one of the most widely used weighting techniques, adjusts word importance by penalizing commonly occurring terms while emphasizing rare but meaningful words (Jones, 1972). Empirical studies have demonstrated that TF-IDF improves text classification accuracy by reducing the dominance of high-frequency, low-information words (Salton et al., 1975).

Specifically, the assignment of feature values (also known as feature weights) to a representation vector is understood as an accessible way to describe the ‘importance’ of words to the text document. That is, the words carrying more information regarding the underlying context should be weighted higher than those that are less informative (Hand, Mannila and Smyth, 2001; Manning, Raghavan and Schütze, 2008). Indeed, not all words are equally informative. For example, a definite article word such as ‘the’, would not provide the same amount of information in most situations as an adjective word such as ‘great’ does. Hence, a vector space model with feature weights, as proposed by Salton et al. (1975), indicate such relative importance of words for a better extraction of the context, the feature weights can affect the calculations when a feature vector is fed to a data mining algorithm. Mirończuk and Protasiewicz (2018) described vector space modelling as an approach that represents a text document as a vector with feature importance where the words in the document form the features.

One early proposed method for assigning feature weights to a sparce vector is the Term Frequency (TF) also known as the bag-of-words model (BOW). With this method, the feature weights are determined based on the statistics of word frequency (Guzella and Caminhas, 2009), deriving the semantic similarity between two pieces of text at the document level then becomes intuitive because the spatial distance between the two vectors can be calculated using the Euclidian distance measure. The rationale behind the term frequency is rather simple, however its performance was later empirically found to be biased due to the existence of some common words such as ‘is’ or ‘the’, this is because these words tend to appear too frequently in almost every sentence and thus are incapable of assisting in linguistically distinguishing different pieces of text. To address this, Jones (1972) proposed the Inverse Document Frequency (IDF) for lowering the feature weights of those words that occur too frequently in the entire corpus of interest and increasing the feature weights of those that appear rarely, this strengthens the ability of the text representation in reflecting how important a word is to a single text document considering its frequency statistics in the whole corpus. Empirical studies have proved that the integrated TF-IDF method performs well on different datasets.

Despite its strengths in document-level similarity computation and interpretability, vector space modelling has notable limitations. The method assumes word independence, disregarding contextual relationships between terms. This results in an inability to capture word semantics and phrase structures, making it less effective for tasks requiring deeper linguistic understanding. Additionally, as vocabulary size grows, vector dimensions increase exponentially, leading to high memory and computational costs. While dimensionality reduction techniques, such as Latent Semantic Analysis (LSA), can mitigate this issue, they still struggle with capturing nuanced word meanings in varying contexts (Manning, Raghavan, and Schütze, 2008).

5.1.2 Word embedding vector

Word embedding methods represent words as dense vectors in a continuous space, enabling models to capture semantic relationships between terms based on their contextual usage in large text corpora (Goldberg, 2016). Unlike vector space modelling, where words are assigned discrete positions, word embeddings map words to a low-dimensional space where similar terms are positioned closer together (De Caigny et al., 2020). In the word embedding approach every word in a text document is embedded in a d dimensional continuous space and represented as a feature vector f in that space. Likewise, a text document can be represented as a d by n matrix $x_{textual}$ where n denotes the total number of words in the document. In other words, a document representation matrix $x_{textual}$ is formed by concatenating the feature vectors $f_1 \dots f_n$ corresponding to the n words in the document. In general, the dimensionality d is defined as a much smaller number than the vocabulary size, often in the range from 100 to 300.

The pre-trained word embeddings are the side product of training language models on huge corpus datasets. Researchers develop language models for a variety of objectives, such as predicting the next word on its left context, predicting the masked words from the context, or predicting the next sentence based on the previous one, and so on. In case of using masked words prediction as the optimisation objective, a word's representation is repeatedly trained and updated based on both linear and nonlinear calculation of the embedding values of its surrounding words, in return the resulting representation vector encodes information of its context, this method is based on the "distributional hypothesis" proposed by Harris (1954) in that the meaning of a word is believed to be formed by its surrounding words. Thus, training a language model can result in semantically similar words sharing similar embedding vectors based on similar contexts, and the dimensions in this embedding space represent shared latent concepts (Goldberg, 2016; De Caigny et al., 2020). It is yet noteworthy that practitioners are advised to draw on pre-trained word embeddings from available resources and then finetune them on domain specific datasets for other use due to the incredibly high computational cost of training embeddings for the whole vocabulary from scratch.

Obtaining dense vector representation from pre-trained embedding resources involves two key stages, pre-processing raw textual data and retrieving feature vectors. Yet only a few elementary steps of pre-processing are necessary, therefore avoiding more advanced manipulation such as removal of uninformative words and stemming. In the stage of representation retrieval, as shown in Figure 5.2, a pre-processed text document X with n words will be first one-hot encoded into a n-by-vocabulary matrix $x_{one-hot\ input}$ where the vocabulary refers to as the ordered word list of the pre-trained word embedding. Then $x_{one-hot\ input}$ is multiplied by a vocabulary-by-d embedding matrix $W_{vocabulary}$ to derive the document representation $x_{document}$ where d represents the predefined dimensionality of the feature vector of a word. As previously mentioned, the word embedding matrix $W_{vocabulary}$, as part of the model parameters, can be obtained from training of a neural language model and afterwards being transferred to a downstream task and updated on another neural network function $F(x)$.

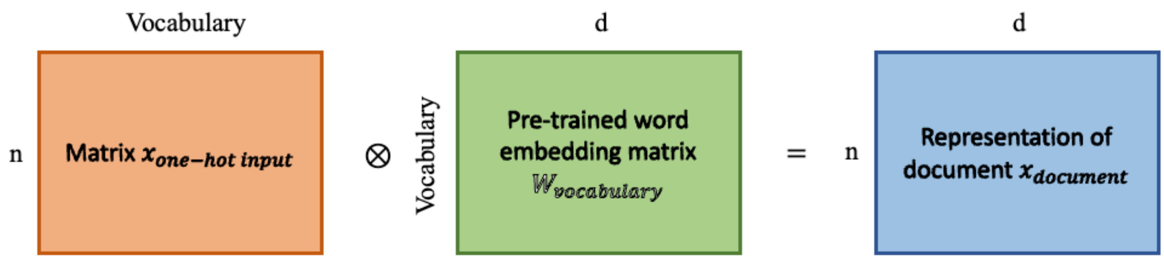

Figure 5.2 The steps of representation retrieval for a text document X

Pre-trained word embeddings, such as Word2Vec, GloVe, and FastText, are obtained by training neural language models on extensive text datasets. These embeddings are widely used due to their ability to generalize across various NLP tasks and their computational efficiency compared to sparse vector representations (Mikolov et al., 2013; Pennington, Socher, and Manning, 2014). Unlike vector space modelling, which treats each word as an independent entity, word embeddings capture rich linguistic relationships, allowing for analogical reasoning and context-aware text processing.

The primary advantage of word embeddings lies in their ability to encode contextual relationships between words, thereby improving performance in sentiment analysis, machine translation, and topic modelling. However, these methods are not without limitations. Pre-trained embeddings may not always be well-suited for domain-specific applications, requiring fine-tuning on industry-specific datasets. Additionally, embeddings can suffer from unintended biases, reflecting social and cultural biases present in training data. Moreover, deep learning models utilizing word embeddings demand significantly more computational resources compared to traditional vector space methods.

5.1.3 Comparative analysis between vector space modelling and word embeddings

Both vector space modelling and word embeddings offer distinct advantages depending on the nature of the task. Vector space modelling provides a simple and interpretable approach for text analysis, particularly in keyword-based applications and document similarity computations. However, it lacks the ability to capture deep semantic relationships and suffers from scalability issues. Word embeddings, in contrast, excel at representing contextual information, making them ideal for tasks involving nuanced language understanding. Nevertheless, they require large-scale training data and substantial computational resources.

The choice between these two methods depends on several factors, including data availability, computational constraints, and the specific NLP task at hand. While traditional vector space models remain relevant in resource-constrained environments, the increasing adoption of deep learning has made word embeddings the preferred approach for modern text analysis applications.

5.2 Extracting Advanced Textual Features for Predictive Modelling

As the field of text representation has evolved, the focus has shifted beyond basic word embeddings to more advanced feature extraction techniques that leverage local contextual indicators and hierarchical representations. These methods enhance textual data processing by capturing nuanced linguistic structures, improving predictive modelling in applications such as sentiment analysis and customer churn classification. This section explores two dominant strategies in advanced textual feature extraction: convolutional neural networks (CNNs) for capturing local textual indicators and Transformer-based architectures for deriving contextualized representations.

5.2.1 Extracting local indicators with convolutional and pooling architecture

While traditional word embedding models such as Word2Vec and GloVe provide rich numerical representations of words, they fail to account for word order and local dependencies, which are critical for sentiment classification and other NLP tasks. For instance, consider the two sentences: “Although the food was good, I was not satisfied with the restaurant overall” and “Although the food was not good, I was satisfied with the restaurant overall.” Despite having nearly identical word compositions, the difference in local word pairings completely alters their meanings. Word2Vec would assign both sentences identical embeddings, leading to misleading sentiment classifications.

To address this limitation, convolutional neural networks (CNNs) have been adapted from computer vision to NLP tasks, offering a solution by extracting local text patterns regardless of their absolute position in a sentence (Collobert et al., 2011). CNNs operate through convolutional filters that slide over text sequences, capturing n-gram features (e.g., word pairs or triplets) that serve as strong predictors in classification tasks. These filters apply non-linear transformations to produce feature maps that highlight key patterns within the text. Figure 5.3 below illustrates a convolutional neural network architecture for sentence classification (Kim, 2014).

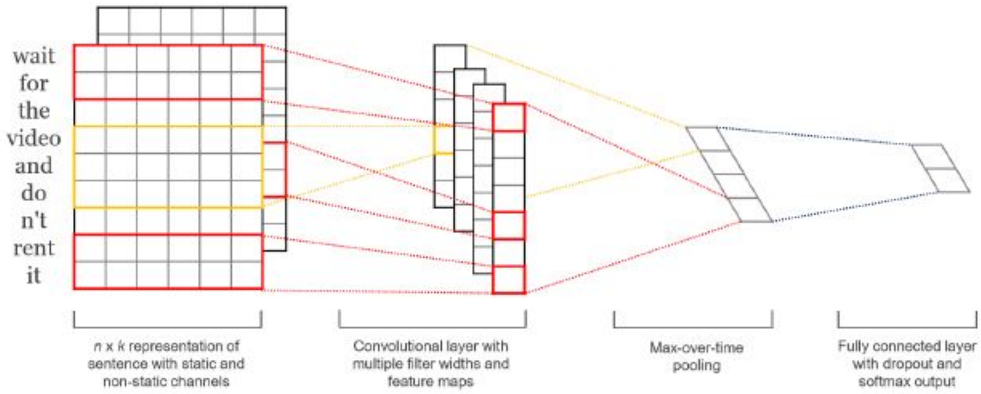

Figure 5.3 the architecture of a sample CNN model for text classification (Kim, 2014)

The rationale behind the CNN architecture for handling textual input is to apply a non-linear learnable function over a text sequence where this function behaves as a window of k-word size (also known as filter) sliding along the pre-trained word embedding representation of n words and compute a $n - k$ dimensional output vector (Kim, 2014). For instance, given a window of size 2 sliding over the sentence “the place of useful learning”, the filter runs over the text sequence and processes every bigram embedding vectors according to the order of word (“the place”, “place of”, “of useful”, and so on), all computed values are concatenated to form an output vector (also known as a feature map) to store the extracted features.

The primary advantage of CNNs in text processing lies in their ability to detect recurring patterns such as negations (“not satisfied”) or positive reinforcements (“highly recommended”) without being affected by their position in the sentence. By stacking multiple convolutional and pooling layers, CNNs aggregate hierarchical feature representations, refining the extracted information for downstream classification models. This approach is particularly effective in short-text classification and tasks where local context plays a crucial role. However, CNNs are inherently limited in capturing long-range dependencies across sentences, making them less suitable for tasks requiring deep contextual understanding (Kowsari et al., 2019).

5.2.2 Extracting contextualised representation with Transformer-based models

The quality of text representation is generally evaluated by its ability to capture the implicit linguistic patterns and common-sense knowledge contained in textual data. Bengio and colleagues (2013) believe that a good representation should encode semantics, syntactic structures and word senses (i.e., polysemy).

Consider learning a dense vector representation from a large text corpus, basic language models such as Word2vec or Glove aim to learn a unique global vector representation for every word in the vocabulary of the corpus. Nonetheless, this type of word embedding is subject to a main limitation, that is ignoring the dynamic contextual information where a word may appear. Similarly, While CNNs excel at capturing localized patterns, they struggle to encode broader semantic relationships and word disambiguation, particularly in polysemous words.

Taking two sentences “There is a football match tonight at eight.” and “This team was no match for its opponent despite a lot of effort.” as example, the word senses of ‘match’ are different in these two sentences in terms of the diverse contexts. A word can have totally different meanings in various contexts, thereby the issue of global vector representation is that the embedding of a word is static and never changes despite its context, only a single representation being obtained for each word results in the failure in modelling complex syntactic information and polysemous words.

Different from the non-contextual embeddings, another class of language models accepts this reasoning and moves beyond the "global" concept by proposing contextual dense vector representation instead. The contextualised approach is designed to distinguish the lexical meanings of words in different contexts to address the issue of polysemy. This type of approach learns contextual word embedding for a word through associating it with the context in the form that its representation is a function of the whole text sequence. These dynamic representations are better suited to extract subtle semantic and syntactic properties of words in diverse contexts than is the case for static representations, while in practice it should be noticed that each input word is generally first mapped to its non-contextualised representation before implementing the contextualisation function.

Just as the theories and concepts in text representation have been progressing, so too have the feature extraction techniques that can materialise those innovations. Vaswani and colleagues (2017) from Google Brain propose 'Transformer' architecture for language modelling with the 'attention' mechanism. The core idea of 'attention' is to compute an approximation of the association between a word and the other words in a text sequence, where such association is believed to be able to reflect the semantic relation and importance between different words to a certain extent, then they can be used as weights to adjust the representation of a word. In case a word appears in different contexts, its contextualised representation can be dynamically updated by applying the 'attention' mechanism to its static word embedding, thus this approach can derive a semantically more robust representation.

Among the various Transformer architectures, Bidirectional Encoder Representations from Transformers (BERT) has gained widespread adoption due to its state-of-the-art performance in multiple NLP tasks (Devlin et al., 2018). Unlike traditional sequence-based models, BERT employs bidirectional training, meaning it learns representations by considering both left and right contexts simultaneously. This bidirectional approach enhances its ability to understand sentence relationships, making it particularly useful for applications such as document classification, named entity recognition, and sentiment analysis.

A distinguishing feature of BERT is its next-sentence-prediction (NSP) task, which allows it to model sentence-level dependencies by predicting whether two given sentences appear in sequence. This additional training objective improves the model's understanding of discourse structure, making it more effective in tasks requiring coherence detection (Young et al., 2018).

In terms of implementation, a pre-trained BERT embedding layer is applied to initialise the embedding layer, where every word in a text document is mapped into a pre-defined 768-dimensional continuous space. The BERT embedding layer provides a non-contextual representation of words that was trained on the English Wikipedia and some other large corpus, when in use for a downstream task these embedding values are transferable and updated according to the surrounding contexts under the BERT architecture. For instance, in the current study, a pre-trained BERT embedding layer is applied to initialize word vectors, which are then fine-tuned to optimize feature extraction for churn classification. The use of pre-trained embeddings significantly reduces computational overhead while leveraging vast linguistic knowledge encoded in large-scale corpora.

5.2.3 Comparative analysis between CNNs and Transformer-based models

Both Convolutional Neural Networks (CNNs) and Transformer-based models have significantly contributed to the advancement of textual feature extraction, yet they cater to different analytical needs. CNNs are particularly effective in capturing local word patterns and short-range dependencies. By applying convolutional filters over text sequences, CNNs can extract key linguistic features such as negations and word pairs, which are essential for sentiment classification and similar NLP tasks. This ability to recognize influential local indicators, regardless of their global position in a sentence, enabling CNNs highly suitable for short-text classification and tasks where phrase-level context holds greater importance. Furthermore, CNNs are computationally efficient, requiring fewer resources compared to deep Transformer-based architectures. Their relatively low computational cost and high-speed inference make them an attractive choice for real-time applications and large-scale text classification tasks. However, a fundamental limitation of CNNs lies in their inability to capture long-term dependencies across sentences, as they primarily focus on extracting local features without considering broader contextual relationships. Additionally, CNNs struggle with polysemy, as the same word in different contexts will be treated identically, leading to potential misinterpretations in tasks requiring semantic understanding.

On the other hand, Transformer-based models address many of these limitations by introducing a self-attention mechanism that allows words to be interpreted in the context of an entire sentence or document. Unlike CNNs, which rely on fixed-length filters, Transformers dynamically adjust word representations based on their relationship with surrounding words. This bidirectional approach enables Transformers to disambiguate words with multiple meanings and capture long-range dependencies, making them particularly effective in complex NLP tasks such as document classification, machine translation, and named entity recognition. Among the various Transformer models, BERT has emerged as a dominant framework due to its ability to incorporate contextualized embeddings and analyze sentence-level relationships. However, despite their advantages, Transformer models are significantly more computationally expensive than CNNs, requiring extensive processing power and large-scale training data. Their high memory consumption and slower inference time can pose challenges in real-time applications, making them less practical in scenarios where efficiency is prioritized over deep contextual understanding.

In summary, the choice between CNNs and Transformer-based models depends largely on the specific requirements of the task. CNNs are well-suited for applications that demand high efficiency and focus on localized textual features, while Transformer models provide a more comprehensive understanding of language by considering long-range dependencies and contextual meanings. As NLP continues to evolve, hybrid approaches that combine the strengths of both CNNs and Transformers may offer optimal solutions, balancing computational efficiency with advanced contextual analysis.

5.3 Evaluating Text Representation Techniques

The effectiveness of different text representation methods in predictive modelling remains an active area of research. While earlier studies in NLP demonstrated that simple unigram models could outperform more complex approaches in certain tasks (Somasundaran and Wiebe, 2010), more advanced linguistic features have since been recognized as crucial for improving classification performance. Specifically, in the domain of complaint management and churn prediction, literature suggests that the way individuals express complaints can provide valuable signals about their likelihood to churn. This makes the selection of an appropriate text representation approach a critical factor in churn classification.

Beyond traditional count-based features, other linguistic patterns, such as local indicators, long-range dependencies, and contextual cues, play an essential role in understanding textual meaning. Word pairs, for example, serve as strong local linguistic markers in sentiment analysis and opinion mining, which are highly relevant for churn prediction. CNNs are particularly effective in capturing these local patterns within short text sequences. However, they struggle with identifying relationships between key phrases that appear far apart in a document. While recurrent neural networks (RNNs) are theoretically capable of maintaining long-term dependencies through their memory cells, they suffer from slow training speeds due to their sequential nature. This limitation makes them computationally expensive for large-scale applications. Given these constraints, RNN-based models were not selected for benchmarking in this study.

The emergence of BERT, built on the Transformer architecture, has revolutionized NLP by offering a universal framework capable of handling various tasks with improved performance. Recent studies (e.g., Alaparthi and Mishra, 2021) highlight BERT’s effectiveness in sentiment analysis, demonstrating substantial improvements in classification accuracy and F1-score on e-commerce datasets. BERT’s superiority lies in its ability to dynamically adjust word embeddings based on sentence context, addressing issues such as polysemy and semantic ambiguity that traditional models struggle with. Unlike previous architectures, which suffer from computational inefficiencies as sentence length increases, the Transformer framework overcomes this bottleneck through its self-attention mechanism. By assigning “attention scores” to each word in a sequence, Transformers efficiently capture relationships between distant words while allowing for parallel computation, significantly reducing training time on large datasets (Minaee et al., 2021).

In summary, while vector space models provide a simple and interpretable baseline for churn prediction, CNNs offer improved feature extraction capabilities by identifying critical local patterns. However, for complex NLP tasks requiring deep contextual understanding and long-range dependencies, Transformer-based models such as BERT have emerged as the most effective solution. The choice of text representation technique ultimately depends on the specific requirements of the predictive task, balancing computational efficiency with the ability to capture nuanced linguistic information.

5.4 Chapter summary

This chapter has explored advanced techniques in textual data analysis, emphasizing their role in enhancing complaint management and churn prediction. Beginning with an examination of text representation methods, the chapter highlighted the trade-offs between traditional vector space models and modern word embedding approaches. While vector space models offer interpretability and computational efficiency, they struggle with capturing deep semantic relationships. In contrast, word embeddings, particularly those derived from deep learning architectures, effectively model contextual information but require substantial computational resources.

The chapter then delved into advanced feature extraction techniques, showcasing how convolutional neural networks (CNNs) and Transformer-based models, such as BERT, improve textual analysis. CNNs efficiently capture local linguistic patterns, making them suitable for short-text classification. However, they lack the capability to model long-range dependencies. Transformer models address this limitation by leveraging self-attention mechanisms to generate contextualized word representations, enabling more accurate text classification and sentiment analysis. Despite their advantages, Transformers demand significant computational power, which may limit their practicality in resource-constrained environments.

A comparative evaluation of these methods underscored the importance of selecting appropriate text representation techniques based on specific predictive modelling requirements. While simpler models may suffice for basic classification tasks, more complex approaches provide substantial performance gains in understanding nuanced customer complaints and predicting churn behaviour. This chapter sets the foundation for integrating these techniques into real-world complaint management systems, paving the way for more intelligent, data-driven decision-making.

Building upon the exploration of advanced textual data analysis techniques, the next chapter shifts focus to the methodological framework that underpins this research. Chapter 6 outlines the research design, data sources, and analytical techniques employed to investigate complaint management and churn prediction. It provides a systematic approach to integrating the discussed text representation and predictive modelling techniques into a structured research methodology. This chapter also addresses key considerations in data collection, preprocessing, and model evaluation, ensuring the study’s robustness, validity, and applicability in real-world scenarios.

6. Methodological Approach for Multimodal Churn Prediction

6.1 Chapter introduction

This chapter outlines the methodological approach employed in this study for multimodal customer churn prediction. Given the increasing reliance on Natural Language Processing (NLP) techniques and machine learning models for predictive analytics, this research aims to integrate textual complaint data with structured variables to enhance churn prediction accuracy.

The chapter begins by restating the research aim and objectives, emphasizing the importance of multimodal data fusion in churn modelling. Following this, the research philosophy is established, justifying the adoption of positivism due to the study's reliance on empirical data, predictive modelling, and quantitative analysis. The research approach is then discussed, detailing the deductive reasoning process and the application of machine learning-based pattern recognition rather than traditional hypothesis testing.

Subsequently, the research design is presented, explaining the structured methodology used to develop, validate, and compare different predictive models. The data collection process is described, including the nature of the dataset obtained from a third-party complaint intermediary and the steps taken to ensure data reliability and ethical compliance. The chapter also elaborates on the proposed multimodal predictive framework, outlining its design, input selection criteria, and intended application in real-world service recovery decision-making.

Further sections cover feature selection and data preprocessing techniques for both textual and structured data, highlighting the methods used to extract meaningful insights from customer complaints. Various machine learning models and fusion strategies are explored, evaluating their effectiveness in integrating multimodal data for churn prediction. Finally, the chapter addresses research ethics, ensuring that the study aligns with established ethical standards.

By establishing a rigorous methodological foundation, this chapter provides a structured framework for understanding how NLP and multimodal data fusion contribute to customer churn prediction.

6.2 Research aim and objectives

The literature reviewed in the previous chapters has highlighted a gap in understanding how Natural Language Processing (NLP) techniques and multimodal data fusion can enhance customer churn prediction through the analysis of textual complaint data. While previous studies have explored the predictive power of structured data in churn analysis, the role of customer complaint texts and their integration with structured variables remains underexplored.

In the big data era, businesses are increasingly relying on advanced analytics to enhance service recovery strategies, ensuring timely and personalized responses to customer complaints. As discussed in Chapters 2 and 3, effective service recovery is crucial for maintaining customer loyalty and preventing churn, yet traditional approaches often fail to leverage the full potential of unstructured complaint data. Addressing this knowledge gap, this research investigates whether advanced NLP techniques, in combination with structured data, can improve churn prediction accuracy and interpretability.

To achieve this, the study focuses on developing a predictive framework that integrates text representation techniques, machine learning models, and multimodal data fusion strategies to assess how textual complaint data contributes to churn prediction. Based on this scope, the research aim is determined as follows:

"*To investigate the application of Natural Language Processing (NLP) techniques in customer complaint analysis for churn prediction and to develop a predictive framework that integrates textual and structured data for enhanced churn modelling*."

The following research objectives are established to achieve this aim:

1) To investigate the role of textual complaint data in churn prediction and compare its effectiveness with structured data-based models.
2) To develop and evaluate machine learning models that leverage different text representation techniques (e.g., TF-IDF, word embeddings, transformers) and multimodal data fusion approaches for customer churn prediction.
3) To identify and analyse key structured variables that contribute to churn prediction and assess their interaction with textual features.

As outlined above, this study is structured around three key research objectives, which collectively contribute to achieving the research aim. The first objective examines the extent to which textual complaint data improves churn prediction compared to structured data alone. By investigating different text representation techniques and machine learning architectures, the second objective focuses on optimizing churn prediction models and exploring the role of multimodal data fusion. Finally, the third objective seeks to identify which structured variables significantly impact churn prediction and how they interact with textual complaint features.

Based on the set of research objectives, the following five research questions are derived.

RQ1: What are the most effective text representation and feature extraction methods for analysing textual complaint data in customer churn prediction?

The way textual complaint data is represented and transformed into structured features is crucial for churn prediction performance. This research explores multiple text representation methods, including TF-IDF, CNN-based word embeddings, and BERT pre-trained embeddings, to determine which approach best captures the semantic and sentiment cues indicative of customer dissatisfaction and potential churn. Identifying the most effective feature extraction techniques will provide insights into optimizing text-based churn prediction models.

RQ2: How do different machine learning models compare in terms of predictive performance across various classes of churn behaviour?

Different machine learning and deep learning models may exhibit varying performance levels in predicting customer churn, particularly across different churn risk categories. This study systematically compares the Multi-Layer Perceptron (MLP), Convolutional Neural Network (CNN), and BERT Transformer models to evaluate their predictive capabilities. The goal is to identify which model best differentiates between high-risk and low-risk churn customers and to understand the trade-offs between complexity, interpretability, and accuracy.

RQ3: What is the impact of incorporating multimodal data (textual complaints and structured variables) on the performance of churn prediction models?

Traditional churn prediction models primarily rely on structured data, such as transactional records and service usage history. However, textual complaints contain rich emotional and behavioural signals that may provide additional predictive power. This research investigates whether combining textual complaints with structured variables enhances model performance and how textual data contributes to improving the interpretability and accuracy of churn predictions.

RQ4: How effective are different data fusion techniques (early, intermediate, and late fusion) in improving churn prediction accuracy?

The choice of data fusion strategy significantly influences the performance of multimodal machine learning models. This study evaluates the effectiveness of early fusion, intermediate fusion, and late fusion techniques in integrating structured and textual data. By comparing different fusion strategies, the research aims to determine the optimal approach for maximizing predictive performance while maintaining computational efficiency.

RQ5: Which structured variables contribute most significantly to churn prediction, and how do they complement textual features in predictive modelling?

In multimodal churn prediction frameworks, structured variables play a crucial role alongside textual features. This research examines which structured features - such as complaint frequency, prior customer experience, and compensation type - are the most influential predictors of churn. Additionally, it investigates how structured and textual data interact to provide a more comprehensive understanding of customer churn behaviours.

The research objectives and questions outlined above have significant implications for the chosen research methodology. Specifically, these objectives emphasize the necessity of adopting a quantitative approach to evaluate the performance of different machine learning models and data fusion techniques. Furthermore, the emphasis on comparative analysis and model evaluation suggests the need for rigorous data-driven methods to assess predictive accuracy. However, before proceeding to describe the research design and data collection strategies, it is essential to establish the philosophical foundations that underpin this study.

6.3 Research philosophy

A research paradigm provides a philosophical and methodological foundation that guides the approach to knowledge acquisition and scientific inquiry. It establishes the ontological and epistemological assumptions that influence how research is designed and conducted (Saunders, Lewis, and Thornhill, 2019). The selection of an appropriate research philosophy is critical in ensuring that the study aligns with its objectives and methodological approach. Given the data-driven and predictive nature of this study, which focuses on applying Natural Language Processing (NLP) techniques and multimodal data fusion for churn prediction, a positivist paradigm is adopted as the most suitable.

Positivism is a well-established research philosophy that assumes an objective reality independent of human perception (Comte, 1853; Ryan, 2018). It is based on the premise that knowledge can be derived from empirical observations, measurements, and systematic analysis. Within a positivist paradigm, reality is considered external, measurable, and explainable through scientific methods, which aligns with the core principles of this study (Bryman, 2016). The research relies on NLP and machine learning models to analyse customer complaint texts and predict churn probability. The underlying assumptions of objectivity, quantifiability, and predictive modelling correspond directly with positivist principles (Saunders et al., 2019).

A key reason for adopting positivism in this study is the nature of the data and research methodology. This study assumes that customer churn is influenced by identifiable factors within textual complaints and structured data, which can be systematically measured and modelled. It follows an empirical and data-driven approach, leveraging large-scale customer complaint data to generate reliable insights (Creswell and Creswell, 2018). The study aims to establish cause-and-effect relationships between textual complaint features and churn probability through statistical and machine learning models (Hair et al., 2020). The use of structured methodologies ensures that findings are replicable and generalizable, supporting broader applications in customer relationship management (Bell, Bryman, and Harley, 2019).

Alternative research paradigms, such as critical realism and pragmatism, were considered but deemed unsuitable for this study. Critical realism, as proposed by Bhaskar (1978), acknowledges the existence of an objective reality but argues that knowledge is shaped by social, historical, and contextual factors (Danermark et al., 2005). While this approach is useful in understanding complex social phenomena, it is less applicable to data-driven research that focuses on quantifiable predictions.

One major limitation of critical realism in this study is its focus on underlying social mechanisms rather than empirical data and predictive models (Easton, 2010). Critical realism is often applied in qualitative research, where researchers seek to explain hidden structures and mechanisms rather than establish generalizable predictive models (Fletcher, 2017). However, this study does not aim to explore the subjective motivations behind customer complaints, but rather to use textual and structured data to predict churn through machine learning models. Furthermore, critical realism does not emphasize predictive modelling, whereas this study is fundamentally focused on developing an effective churn prediction framework using NLP. Since minimal subjective interpretation is involved in this research, positivism remains the most appropriate philosophical stance.

Similarly, pragmatism was not adopted as the guiding research philosophy for this study. Pragmatism, as proposed by Dewey (1931) and later expanded by Morgan (2014), prioritizes practical solutions over strict adherence to a specific ontological or epistemological stance. It is frequently associated with mixed-methods research, where both quantitative and qualitative approaches are integrated (Tashakkori and Teddlie, 2010). While pragmatism allows for a context-driven research approach, it is not suitable for this study because it does not follow a mixed-methods design (Creswell, 2014).

Pragmatism is particularly useful when research involves subjective interpretation or exploratory analysis, neither of which aligns with this study's structured, theory-driven approach (Shannon-Baker, 2016). This research follows an empirical, quantitative methodology, using established NLP and machine learning techniques to analyse customer complaints and predict churn. Given that pragmatism often involves methodological flexibility rather than strict adherence to a structured process, it does not align well with the objectives of this study (Feilzer, 2010).

In conclusion, positivism provides the most appropriate philosophical foundation for this research, as it aligns with the study's emphasis on data-driven, systematic, and empirical analysis (Ryan, 2018). This study relies on structured methodologies, statistical modelling, and predictive analytics to derive insights into customer churn prediction. The alternative paradigms of critical realism and pragmatism do not sufficiently support the research focus, as they either emphasize subjective interpretation or lack the structured, empirical approach required for predictive modelling. By adopting a positivist stance, this study ensures that customer complaint texts and structured data are systematically analysed to develop a robust, objective, and replicable predictive framework for churn modelling.

As the underlying positivist philosophical position informs the methodological approach, a deductive and quantitative research approach has been adopted to ensure systematic empirical analysis. The next section will further elaborate on this by discussing the approach to theory application and model evaluation.

6.4 Research approach

The research approach refers to the overall strategy employed to fulfil the previously stated research aim and objectives. This study follows a deductive reasoning process, aligning with the positivist research philosophy underpinning the study. The research is structured and systematic, aiming to evaluate different Natural Language Processing (NLP) and machine learning techniques for customer churn prediction through empirical experimentation, pattern recognition, and model evaluation.

Unlike traditional deductive research, which often involves hypothesis testing in a statistical sense, machine learning-based research focuses on optimizing predictive performance through data-driven pattern recognition rather than validating predefined hypotheses (Shmueli, 2010). This study does not seek to test explicit statistical relationships but instead leverages machine learning models to detect underlying patterns in customer complaint data that are indicative of churn. By systematically comparing different text representation techniques and machine learning models, this study aims to identify the most effective methods for churn prediction.

A deductive approach is appropriate in this study because it applies existing theories and methodologies from customer churn prediction, NLP, and machine learning research to develop a predictive framework (Bryman, 2016). Rather than discovering entirely new theoretical concepts, this study aims to refine and optimize existing approaches by systematically evaluating model performance using quantitative metrics such as accuracy, F1-score, and AUC-ROC. The evaluation process relies on iterative experimentation and pattern recognition, where different models and techniques are tested, validated, and compared to determine the most effective approach.

By employing a structured, data-driven experimental design, this study ensures that findings are replicable and generalizable across different datasets and business contexts. The next section will discuss the research design and methodological considerations guiding the implementation of the study.

6.5 Research design

The purpose of a research design is to explain and justify the data collection and analysis process, ensuring that the methodological choices align with the overall research aim and objectives (Easterby-Smith, Thorpe & Jackson, 2008). Research designs can vary significantly depending on whether the study follows a quantitative or qualitative approach. In quantitative research, the research design is typically structured and predetermined, ensuring replicability and objectivity (Corbetta, 2003). In contrast, qualitative research often adopts a flexible and evolving design, allowing for data-driven insights to shape the research process. Given that this study follows a positivist paradigm and employs a deductive, quantitative approach, the research design is highly structured, with clearly defined data collection, processing, and analysis stages.

This study is designed as an experimental, data-driven research project, focusing on the development and validation of a multimodal churn prediction framework that integrates both textual complaint data and structured customer information. The primary objective is to determine whether combining unstructured textual data and structured customer attributes can enhance churn prediction accuracy. This framework serves as the foundation for all subsequent experimentation and optimization efforts. Rather than testing individual machine learning models in isolation, the research aims to validate the effectiveness of the multimodal integration itself by systematically evaluating different text-processing techniques, structured data attributes, and predictive modelling strategies.

The study begins with an extensive review of existing literature on service recovery, customer churn prediction, Natural Language Processing (NLP), machine learning models, and multimodal data fusion. This literature review helped refine the research aim and objectives by identifying gaps in existing research and establishing the need for a holistic churn prediction framework that incorporates both structured and textual data sources. Based on insights gained from prior studies, this research proposes and implements a multimodal predictive framework, serving as the central experimental foundation. This framework is designed to accommodate various text representation techniques and machine learning models, allowing for structured comparisons across multiple configurations.

The study utilizes secondary data, consisting of customer complaint texts and structured customer information from an enterprise database. To ensure that the data is suitable for computational modelling, extensive preprocessing techniques are applied. Textual data undergoes tokenization, stop-word removal, stemming, lemmatization, and vectorization (TF-IDF, Word2Vec, BERT-based embeddings) to transform it into structured input for machine learning algorithms. Similarly, structured customer data is cleaned, normalized, and transformed to ensure consistency and compatibility with the textual features. These data sources are then fused within the proposed multimodal framework, ensuring that both structured attributes and textual information contribute to the predictive process.

Following the preprocessing phase, the study employs a comparative experimental approach to evaluate the performance of different churn prediction models within the proposed multimodal framework. Several models are trained and tested (e.g. Random Forest and Transformer-based architectures) to determine the most effective approach for churn prediction. The core focus is not only on selecting the best-performing model but also on evaluating how different levels of textual and structured data fusion impact predictive performance. To achieve this, the study implements multiple fusion strategies, including early fusion (combining features before modelling), intermediate fusion (merging hidden representations), and late fusion (aggregating model outputs).

These different configurations allow for a robust assessment of how multimodal integration improves churn prediction compared to single-modality approaches.

The trained models are assessed using standard classification metrics, including accuracy, precision, recall, F1-score, and AUC-ROC, to provide a comprehensive evaluation of performance. To ensure robustness, cross-validation techniques are employed, and hyperparameter tuning is conducted to optimize model performance. The results of these experiments are then analysed to determine whether the multimodal framework provides a significant improvement over traditional single-modality churn prediction models. The study critically examines the effectiveness of integrating textual complaint data with structured attributes and discusses the potential business implications of adopting such a predictive framework.

Although the research design follows a structured and linear process, some phases, particularly data preprocessing and model tuning, involve iterative adjustments to improve performance. The development of the multimodal churn prediction framework serves as the foundational element of this research, ensuring that all subsequent experimental comparisons are conducted within a unified architecture. By systematically evaluating different NLP techniques, machine learning models, and fusion strategies, the study ensures that findings are replicable, generalizable, and contribute to the ongoing development of data-driven churn prediction methodologies. The next section will provide further details on the methodological choices and data analysis techniques employed in this study.

6.6 Dataset Collection and Characteristics

The dataset utilized in this study was sourced from a third-party complaint intermediary based in the United Kingdom. This marks a departure from conventional Customer Churn Prediction (CCP) studies, which predominantly rely on structured customer data collected over an extended observation period. Instead, the dataset for this study consists of multimodal case files, incorporating both textual complaint data and structured attributes collected at three distinct time points along the service journey.

The first phase of data collection occurs when complainants submit a case file on the intermediary platform. During this stage, an initial survey captures both textual and structured information, including the nature of the service failure, the description of the service failure, the complainant's emotional response, the time of the incident, and their requested resolution, and so forth. The second phase of data collection takes place during the service recovery process. This phase records all interactions between the service provider and the complainant, capturing the full correspondence where available. Since the resolution period varies across cases, the length of recorded interactions differs accordingly. The third phase of data collection consists of responses to a structured feedback survey, administered three months after the case is either resolved or remains unresolved. This survey gathers information on the resolution status, the type of compensation received (if any), and the complainant's intention to churn, and so forth. This post-resolution survey provides valuable insights into whether service recovery efforts were successful in mitigating customer dissatisfaction and preventing churn.

The final dataset comprises 55761 case files, each containing 29 attributes (including structured and textual data). The dataset was provided in JSON format by the intermediary company, ensuring standardized data representation for further processing. All textual inputs were written in English, and personally identifiable information (PII) was removed to ensure data anonymity and compliance with ethical research standards. The full list of questions used in both the initial and the feedback surveys is provided in the Appendix for reference.

The following research methods section describes the data analysis process.

6.7 Research methods

Research methods refer to the tools and techniques employed for data collection and analysis (Bell et al., 2018). They determine both the nature of the data obtained and the approach used for its analysis and interpretation. The rigorous selection and application of research methods are essential for ensuring the validity and reliability of research findings (Jonker and Pennink, 2010). This section will outline and discuss the research methods adopted in this study.

6.7.1 Data cleaning

To ensure the quality and reliability of the dataset, a rigorous data cleaning process was conducted before further preprocessing and modelling. This process addressed missing data, duplicate records, outliers, categorical inconsistencies, and uninformative textual content, aligning with best practices in machine learning-based research (Kotsiantis et al., 2006).

The dataset was processed and cleaned using Python-based libraries, ensuring a structured and standardized format for subsequent analysis. Although no missing values were detected in the structured attributes, the accuracy of data entries remained a concern, as they relied on user-provided information. Given the potential for human error in manual data entry, additional validation steps were considered but ultimately deemed unnecessary based on exploratory data analysis.

The textual component of the dataset was initially composed of three sections: 1) a mandatory description of the failure event, which forms the core of the complaint; 2) an optional section detailing the impact of the event on the complainant; and 3) an optional section outlining the complainant's desired resolution. To streamline textual analysis and enhance feature extraction, three sections were merged into a single textual column. Complaints with missing mandatory fields were removed to maintain data integrity. Additionally, duplicate records were identified and eliminated through a combination of exact match detection and similarity-based filtering techniques, ensuring that duplicate complaints were accounted for.

Furthermore, non-informative content, including templated responses and automated customer service replies, particularly prevalent in the airline industry, was remove to enhance data relevance. All textual data were standardized to UTF-8 encoding, preventing character encoding inconsistencies.

The cleaned dataset consisted of 32,279 online complaint cases, of which approximately 59% exhibited churn-related behaviours, either through explicit churn intention or confirmed account termination within the follow-up period. To align with the study's objectives, cases where customers indicated churn intention were merged with actual churn cases, resulting in a binary dependent variable that categorizes customers as either churned or retained.

Although the dataset exhibits a slight class imbalance, prior research suggests that imbalance becomes problematic for machine learning models only at extreme ratios (e.g., 1:4 to 1:5000 or beyond) (Krawczyk, 2016). Consequently, the dataset was deemed sufficiently balanced, and no additional resampling techniques were applied. The cleaned dataset was then used for subsequent feature engineering and model development.

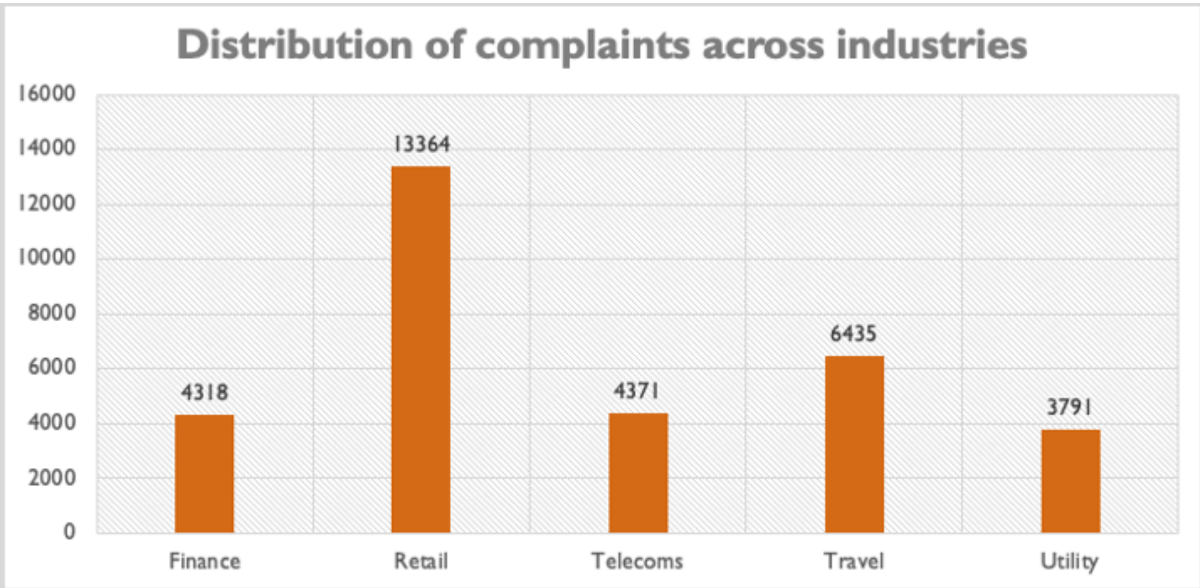

Figure 6.1 Distribution of complaints across industries

The cleaned dataset spans a diverse array of industries, including finance, retail, telecommunications, travel, and utilities. Notably, the retail sector constitutes the largest proportion of complaints, indicating its significant customer interaction volume. Figure 6.1 illustrates the distribution of complaints across various industries included in the study.

Despite this, our analytical approach does not segment the complaint data by industry. Instead, we aim to discern overarching patterns within the complaints that surpass industry-specific nuances, thereby identifying general predictors of churn applicable across various sectors. This holistic method seeks to extract universal insights from the complaint narratives, facilitating the development of a robust churn prediction model that captures the essence of customer dissatisfaction regardless of the industry context.

6.7.2 Proposing a Multimodal Analytical Framework for CCP

The introduction of research aim and objectives in section 6.2 establishes the necessity of employing a multimodal analytical framework in this study. Traditional churn prediction models often rely on unimodal structured data, which may overlook valuable insights embedded in textual complaints. Given the dataset's composition of textual and non-textual modalities, a multimodal approach is essential for integrating these diverse data sources effectively. The proposed framework aims to bridge this gap by leveraging both textual and non-textual information to enhance predictive performance. Inspired by De Caigny and others (2020), this multimodal framework is designed to extract and synergize insights across different modalities, providing a more comprehensive perspective on customer churn prediction. Figure 6.1 illustrates an intermediate fusion scheme as an example implementation of the proposed approach.

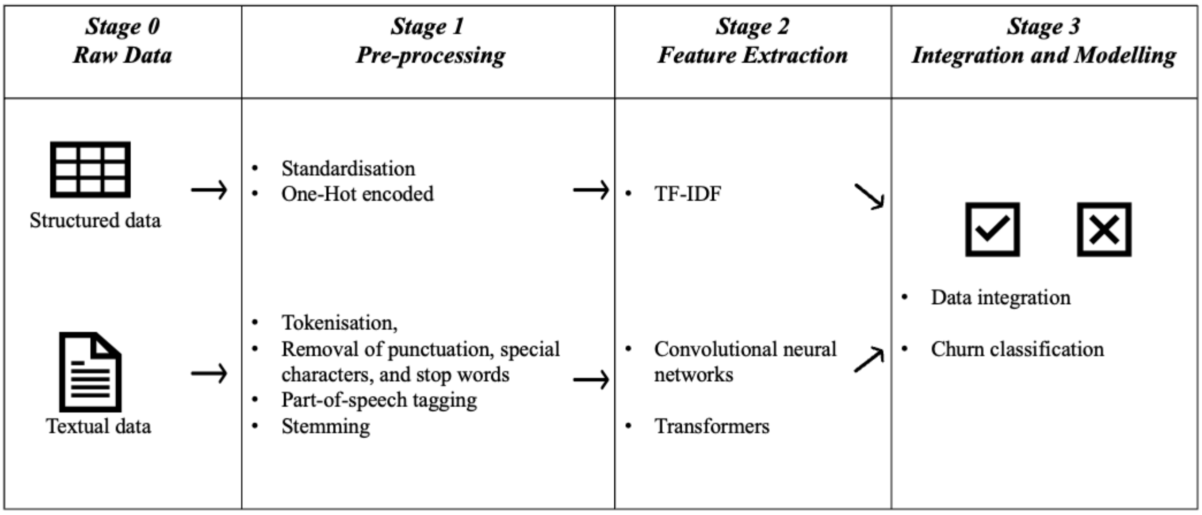

Figure 6.1 Proposed Multimodal Analytical Framework for CCP

A critical aspect of this research design is the time-sensitive nature of the input data selection, as the proposed framework is not only designed for churn prediction but also for recommending optimal service recovery strategies. The goal is to develop a model that allows companies to assess churn risk at the time of complaint submission, while also enabling them to simulate the impact of different remedial actions, such as compensation type and resolution time, on churn outcomes. Given this objective, the model's input variables are carefully selected to reflect both the initial complaint context and subsequent remedial actions. Specifically, the input features include: 1) the textual content of the complaint, as submitted by the customer; 2) structured data related to the complaint submission, such as the emotions, whether previously complained, the requested compensation type (if available), and any other attributes known at the point of submission; 3) remedial action data; 4) the dependent variable (churn intention), which is observed at the end of the follow-up period.

Unlike traditional churn prediction models that focus solely on customer attributes and transaction history, this framework explicitly integrates remedial action variables, allowing businesses to test different recovery strategies before implementing them. In practice, when a new complaint is received, a company can simulate different compensation offers and resolution timeframes as input values in the model. By observing the predicted churn probability under different scenarios, decision-makers can optimize their service recovery strategy to minimize churn risk.

However, before this framework can be effectively applied in practice, it is first necessary to evaluate whether incorporating different data types - such as textual complaints, structured attributes, and remedial action variables—meaningfully enhances the model's ability to predict customer churn. This study therefore investigates whether combining multiple data modalities improves prediction accuracy compared to using only a subset of these features. The empirical validation of these data sources is a crucial step in determining the effectiveness of the proposed framework and ensuring that its recommendations are based on robust predictive insights rather than arbitrary assumptions.

This data selection approach has significant implications for the feature selection process. Since the model must capture the interaction between textual complaint characteristics and remedial actions, the study focuses on extracting meaningful linguistic patterns from complaint texts, identifying structured variables that contribute to the effectiveness of different compensatory strategies, and modelling the impact of compensation type and resolution time on churn likelihood

By incorporating both pre-resolution and post-resolution attributes, this framework extends beyond conventional churn prediction to offer a decision-support system for customer complaint management. This methodology ensures that the proposed framework is not only theoretically sound but also practically implementable in real-world business environments where dynamic recovery decision-making is essential.

In the preprocessing stage, structured and textual data undergo separate preparation procedures tailored to their inherent characteristics. For structured data, transformations such as normalization and encoding are applied to ensure compatibility with machine learning models. For textual data, preprocessing involves tokenization, stop-word removal, and vectorization, converting free-text complaints into numerical representations suitable for predictive modelling. The complexity of textual data preprocessing varies depending on the chosen representation technique, which plays a critical role in model performance. Detailed explanations of text representation techniques and feature selections are provided in Sections 6.7.3 and 6.7.4.

In the integration phase, the framework fuses numerical representations of textual data with structured attributes into a unified input vector z by concatenation. For instance, the intermediate fusion strategy (as illustrated in Figure 6.1) concatenates these feature representations before feeding them into a classification model. In the final modelling stage, the framework employs a deep learning-based classification approach to predict customer churn.

The input vector z is fed into a multi-layer neural network (MLP), which consists of fully connected layers with non-linear activation functions. This architecture is chosen for its ability to capture complex feature interactions and its flexibility in handling high-dimensional data. The model $f(z)$ produces a binary prediction vector, O, representing the likelihood of customer churn.

The selection of an optimal classification algorithm remains task-dependent, as no single classifier consistently outperforms others across all scenarios (Verbeke et al., 2012). While deep learning models offer superior feature learning capabilities, Logistic regression remains a widely adopted choice in business applications due to its simplicity, interpretability, and robustness. Although Logistic Regression has limitations in capturing complex interactions, its high explainability makes it a preferred option in regulatory settings where transparency and interpretability are crucial.

6.7.3 Pre-processing textual data

The pre-processing of textual data is a critical step that directly influences the quality and interpretability of predictive models. Unlike structured data, textual data is unstructured, high-dimensional, and often noisy, requiring transformation into a structured, machine-readable format. This study employs tokenization, stop-word removal, and feature extraction to enhance the efficiency and accuracy of churn prediction models. Given the varied nature of textual representation techniques, different preprocessing strategies are applied to vector space models (VSM), convolutional neural networks (CNN), and Transformer-based architectures, each of which demands tailored preprocessing steps.

6.7.3.1 Preparing Textual data for Vector Space Models

The vector space modelling (VSM) approach necessitates a structured preprocessing pipeline to transform raw text into numerical representations. The process begins with removal of punctuation and special characters, eliminating noise that does not contribute to semantic meaning. Following this, tokenisation is performed, segmenting text into its discrete tokens - typically words. Subsequently, all words are converted to lowercase to ensure uniformity.

Further refinement involves removing stop words, which are frequent words that contribute little to the contextual meaning (Frakes, 1992). Additionally, part-of-speech (POS) tagging is employed to categorize words based on their syntactic roles (e.g., noun, adjective, adverb), aiding in the stemming process. Stemming reduces words to their root forms (e.g., “catching” into “catch”), mitigating variations in word inflection while preserving meaning. These pre-processing steps collectively reduce data dimensionality, enhance model generalization, and improve pattern recognition efficiency (Bell and Jones, 1979).

Tokenisation extracted an initial vocabulary size of approximately 20,000 unique tokens from the research corpus. After stop-word removal, the vocabulary size was reduced to 9,258, significantly improving computational efficiency. The part-of-speech tagger used in this study was trained on the Penn Treebank corpus (Marcus, Santorini, and Marcinkiewicz, 1993), ensuring linguistic accuracy. Finally, stemming further reduced the vocabulary to below 1,000 unique word stems, optimizing the dataset for downstream vectorization. The pre-processed textual data were aggregated in a word-by-customer matrix.

Feature extraction is the next crucial step where meaningful attributes are derived from the text. In the process of textual data analysis, the choice of feature extraction technique is pivotal. We employ Term Frequency- Inverse Document Frequency (TF-IDF) to weigh the frequency of each word against its inverse frequency across all documents, which amplifies the signal of more informative words. While TF-IDF is chosen for vector space modelling, alternative methods such as Bag of Words (BoW) and n-grams were considered. While BoW is a simpler model that convert documents into word collections, disregarding syntax and sequence, which often leads to high-dimensional data with sparse matrices, making it computationally inefficient and less informative. The n-gram model extends the BoW approach by analysing consecutive word sequences, which provides context that BoW lacks but at the cost of significantly increasing the feature space, risking model overfitting and worsen the issues of sparsity.

The TF-IDF representation was selected for this study due to its ability to mitigate the sparsity issues associated with traditional Bag of Words (BoW) models. Unlike BoW, which disregards word sequence and assigns equal weight to all words, TF-IDF assigns greater importance to terms that are more informative within a given document, thereby improving feature relevance. Additionally, while n-gram models capture word co-occurrence patterns and preserve contextual information, they often lead to exponential feature space expansion, increasing the risk of overfitting and computational inefficiency. This characteristic of TF-IDF can be particularly beneficial for churn prediction models, where discerning unique customer complaints and issues is important. By emphasising words that are more likely to contribute to the prediction of churn, TF-IDF facilitates a more nuanced and focused analysis, leading to more accurate and interpretable results.

6.7.3.2 Pre-processing for Dense Embedding Vectors

In contrast to sparse vector representations, dense word embeddings encode semantic relationships in a continuous vector space, requiring minimal text preprocessing. In this study, tokenization, special character removal, and case normalization were performed before embedding.

The maximum sequence length for textual complaints was at 400 tokens, following statistical analysis of text length distributions. As illustrated in Figure 6.2, over 95%

of complaints contained fewer than 400 tokens, ensuring that essential textual information is preserved while maintaining computational efficiency. Longer sequences were truncated, while shorter sequences were zero-padded to ensure uniform input dimensions across models. This standardization allows for fair comparison of different architectures, ensuring that variations in model performance arise from differences in feature extraction rather than input sequence length.

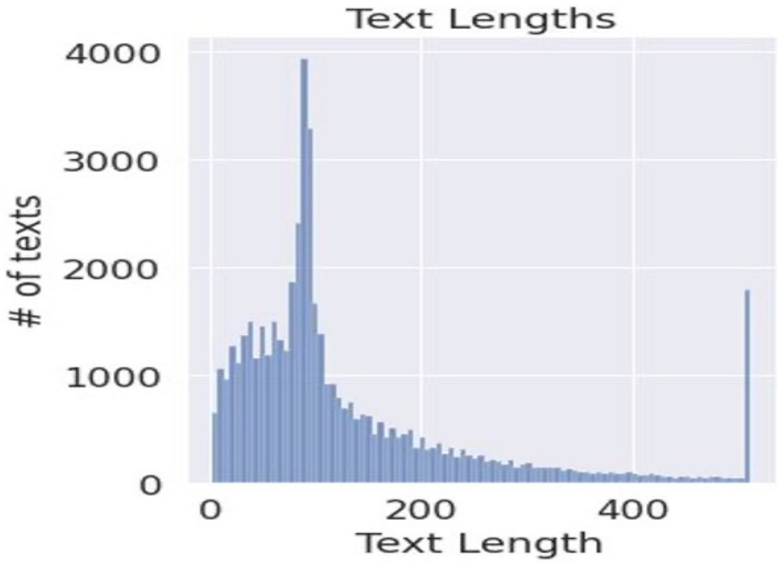

Figure 6.2 Distribution of text length

Moreover, applying this 400-token limit consistently across different models, including the embedding layer of the CNN-based model, allows for a controlled comparison of model performance. By standardizing the embedding size, any variations in performance can be attributed to the model's feature extraction capabilities rather than differences in input length. This consistency ensures that the model evaluation focuses on the efficacy of textual feature extraction in the context of churn prediction.

6.7.4 Structured data feature selection and pre-processing

To evaluate the efficacy of data fusion techniques in the context of the complaint-driven churn prediction, it is essential to incorporate structured data alongside textual complaints. Existing research in service recovery has extensively examined the factors influencing customer perceptions and behaviours post-recovery. Constructs such as customer satisfaction, post-recovery satisfaction, and repurchase intent are widely recognized as key determinants of a customer's decision to maintain or terminate their relationship with a company following a service failure and subsequent recovery efforts.

Accordingly, this study integrates structured variables that capture these influential factors into the research dataset. These constructs, which have been shown to correlate significantly with customer perceptions and behaviours in prior literature, provide an empirical basis for investigating the interaction between service recovery efforts and customer churn. By embedding these constructs into the dataset, the model is expected to achieve improved predictive accuracy and offer deeper insights into how different dimensions of service recovery influence customer retention.

Based on a systematic review of service recovery literature, a set of constructs was selected according to two primary criteria: relevance and availability, that is, the variable should have been identified in prior studies as an influential predictor of customer perceptions or churn behaviours and it must be accessible through the data collection process. The set of constructs includes: 1) emotions; 2) previous experience with service provider; 3) remedial actions; and 4) remedial procedure.

1. **Emotions**: prior research have pointed out that customers' affective states can have impact on their satisfaction judgement (e.g., Westbrook, 1980; Westbrook and Oliver, 1991; Stephens and Gwinner, 1998; Maute and Dube´s, 1999), while in more recent studies negative emotions have been said to partly account for one's decision on how to behave in case of a service failure (Yi and Baumgartner, 2004; Zeelenberg and Pieters, 2004, Mattila and Ro, 2008). Furthermore, it has also been observed that emotions might not only affect an individual's immediate or short-term attitudes and behaviours but could also consequently impact on one's evaluation of the recovery efforts (e.g., Watkins and Liu, 1996; Stephens and Gwinner, 1998; Mattila, 2004; Schoefer and Diamantopoulos, 2008). Recent studies (e.g. Alaparthi and Mishra, 2021) have

shown that BERT demonstrates higher predictive accuracy in sentiment analysis, which could imply its potential effectiveness in predicting customer churn if sentiment is a significant factor in churn.

In this study, the 'Emotions' variable was derived from the initial survey when the complaint case file was created by customers, where they rated their overall emotional response on a scale from 1(very negative) to 7 (very positive).

2. **Previous experience with service provider**: prior research (e.g., Stephens and Gwinner, 1998; Gregoire and Fisher, 2008; Kim et al., 2010) have found that customers are likely to take their previous experience with the service provider into account during satisfaction evaluation. Mattila (2004) indicates that a history of good relationship may lead customers to be more tolerant and forgiving towards service providers in a service failure event. However, other studies (Bitner et al., 1990; Bolton, 1998) suggest that customers with higher cumulative satisfaction from previous consumptions can be, in effect, more demanding and fractious to poor service performance.

   In this study, the 'Previous complained' variable was derived from the initial survey when the complaint case file was created by customers, where they pointed out whether a previous complaint had been formally made on the same service issue.

3. **Remedial actions**: a few studies have pointed out that financial compensation would be one of the most influential factors to post-recovery satisfaction (Tax and Brown, 1998; Smith, Bolton, and Wagner, 1999; Cranage and Mattila, 2006), despite their findings seem contradictory. Some scholars (e.g., Maxham III, 2001; Bradley and Sparks, 2012) found that providing a higher level of compensation can be more effective in amending the relationship with customers. Yet, Noone (2012) reported that the magnitude of compensation is shown to not have a significant impact on post-recovery satisfaction. Also focusing on financial compensation, Boshoff (2012) and Chen and others

(2018) found that overcompensation can lead to counterproductive effects on customer satisfaction, and thus compensation for complainants is suggested to be offered appropriately at a more moderate level.

In this study, the 'Remedial action' variable was derived from the feedback survey, where they indicated the type of compensation received from the service provider.

4. **Remedial procedure**: The remedial procedure has shown to significantly impact on one's post-recovery satisfaction among airline travellers. Ghalandari and others (2012) confirmed that the remedial procedure perceived by customers is positively correlated with retention rate.

   In this study, the 'Remedial procedure' variable was derived from the feedback survey, where they indicated the length of time taken by the service provider for resolving the complaint.

Apart from the set of constructs suggested by prior literature, we also incorporate into the dataset a handcrafted count-based feature – it is conceivable that statistics such as the length of a complaint may benefit the prediction of churn, the rationale behind is that the more aggressive complainants may tend to write longer complaints to express their grievance, thereby such information can be insightful to the evaluation of their repurchase intent. Thus, the structured variables selected for use in this study is listed below in Table 6.1.

Table 6.1 The list of structured variables used in current study.

Variable	Modality	Group
Sentiment score	Numeric	Emotion
Previously complained	Binary	Previous experience
Types of compensation	Categorical	Remedial action
Length of time for resolution	Numeric	Remedial procedure

Word count of textual complaint	Numeric	Handcrafted count-based feature

To ensure compatibility with deep learning neural network models, structured tabular data underwent preliminary preprocessing to facilitate seamless integration with textual embeddings. This involved distinct processing strategies for numerical and categorical variables to standardize feature representations across modalities.

For numerical variables, standardization was applied, transforming each variable to have a mean of zero and a standard deviation of one. This process ensures that numerical inputs are on a consistent scale, which is crucial for the stable and efficient training of neural networks. Standardization prevents certain features from dominating others due to differences in scale, thereby improving the convergence properties of gradient-based optimization methods in deep learning models.

For categorical variables, One-Hot Encoding (OHE) was employed. This method converts a categorical variable with n categories into $n-1$ binary variables, ensuring that categorical data is represented in a format that is interpretable by neural networks. Compared to label encoding, which assigns arbitrary numerical values to categories, One-Hot Encoding prevents the model from interpreting categorical relationships as ordinal, thereby preserving the non-hierarchical nature of categorical attributes. Additionally, this encoding approach is particularly well-suited for Transformer-based architectures, as it facilitates efficient feature integration.

The final step in structured data processing involved concatenating numerical and categorical features with textual representations. Specifically, the output vector from a BERT model - which encapsulates the semantic representation of complaint texts - was concatenated with One-Hot encoded categorical variables and standardized numerical features. This multimodal data fusion approach allows the model to leverage both structured attributes and textual patterns simultaneously, without requiring further complex preprocessing steps. By streamlining the data preparation pipeline, this method enhances computational efficiency while preserving the rich information embedded in both structured and unstructured data sources.

6.7.5 Evaluation of Feature Extraction Techniques in Predictive Models

To systematically assess the effectiveness of different text representation techniques in churn prediction, this study compares three distinct models: A Multi-Layer Perceptron (MLP) using TF-IDF features, a Convolutional Neural Network (CNN) leveraging pre-trained word embeddings, and a BERT model incorporating contextual embeddings. Each model is selected based on its ability to handle specific linguistic properties of customer complaints, allowing for a comprehensive evaluation of how different feature extraction techniques contribute to predictive performance.

The MLP model, based on Term Frequency-Inverse Document Frequency (TF-IDF), serves as a baseline model due to its simplicity, interpretability, and effectiveness in high-dimensional sparse text representations. TF-IDF quantifies word importance by considering term frequency within a document relative to its occurrence across the dataset, providing a fundamental statistical approach to text representation. The MLP architecture consists of a single fully connected layer with 128 hidden units, followed by a Softmax classifier for binary classification.

Unlike MLP, the CNN model employs a spatially aware approach to text feature extraction, leveraging convolutional filters to detect local patterns and contextual dependencies within complaint narratives. CNNs are particularly adept at identifying n-gram structures, which are crucial for capturing key phrases indicative of customer dissatisfaction.

The word embeddings for CNN are initialized using a Continuous Bag-of-Words (CBoW) model trained on the English Wikipedia corpus (Mikolov et al., 2013), ensuring that the model benefits from pre-trained semantic representations. Each word is embedded in a 200-dimensional vector space, following prior research recommendations (Fauconnier, 2015). The CNN architecture includes 1D convolutional layers with max pooling, as suggested by De Caigny and others (2020), followed by a fully connected layer (Kim, 2014). The final Softmax classifier assigns churn probability to each complaint case.

The BERT model represents an advanced approach to contextual feature extraction, utilizing deep bidirectional attention mechanisms to capture long-range dependencies and intricate linguistic structures. Given that customer complaints often contain nuanced expressions of dissatisfaction, BERT's contextual embeddings provide a richer and more accurate representation of textual meaning compared to traditional methods.

For implementation, we utilize Hugging Face PyTorch BERT library (Wolf et al., 2019), initialising the model with default hyperparameters as proposed by Devlin et al. (2018). A few different sizes of BERT versions are provided by the library, while the selection of model is restricted by limited computing resources. Thus here in this study, the default version of BERT - $BERT_{base}$ is employed, comprising 12 layers, 12 self-attention heads, and 110M parameters – a more computationally efficient alternative to $BERT_{large}$, which has 340 million parameters.

The final feature representation produced by BERT is a 768-dimensional vector, which is concatenated with structured data features via multimodal fusion techniques. To ensure effective classification, fully connected layers with ReLU activation are applied before the final Softmax classifier. The model is finetuned for 4 epochs using Adam optimiser (Kingma and Ba, 2015) with a learning rate of $2e^{-5}$, batch size of 32, a maximum input sequence length of 400 tokens, and a dropout rate of 0.1 for all layers.

To illustrate how these models handle customer complaints differently, we consider the following example of a real-world anonymized complaint: "*I have been a loyal customer for years, but after multiple service failures and unhelpful support, I am seriously considering cancelling my subscription.*" Each model processes this complaint using a distinct text representation technique.

With use of MLP with TF-IDF, the complaint is transformed into a numerical vector based on word frequencies. For example, terms such as “loyal,” “failures,” and “cancelling” may have high term frequencies, while words like “I” or “for” contribute little information. However, TF-IDF lacks context awareness and cannot distinguish between a positive statement about loyalty and a negative shift in sentiment due to poor service. The resulting churn probability might be moderate, as it captures key complaint-related terms but fails to understand the transition in sentiment.

With use of CNN with pre-trained word embeddings, the complaint is tokenized and mapped into 200-dimensional word vectors, with convolutional layers extracting local n-gram patterns. CNN might recognize that phrases like “multiple service failures” and “unhelpful support” frequently co-occur in churn-related complaints. However, due to its reliance on localized feature extraction, CNN might miss the overall progression of sentiment across the sentence. The predicted churn probability could be higher, as CNN better captures negative phrases but lacks a full understanding of sentence-level context.

With use of BERT with contextual embeddings, the complaint is processed with transformer-based self-attention, allowing the model to understand the evolving sentiment within the text. BERT recognizes that “loyal customer for years” initially conveys a positive history, but “multiple service failures and unhelpful support” signals a deterioration in sentiment, leading to churn intent expressed in “seriously considering cancelling my subscription.” The model captures the contextual relationships between words and phrases, leading to a high churn probability.

The selection of these three models is driven by their complementary strengths in text feature extraction and their relevance to the complaint-driven churn prediction task. The MLP model, leveraging TF-IDF representations, provides a statistical baseline that allows for a straightforward, interpretable, and computationally efficient approach to text classification. Its primary advantage lies in its ability to capture word frequency-based signals without requiring extensive computational resources.

The CNN model is chosen for its ability to recognize spatially local patterns in text, making it particularly effective in identifying recurrent phrase structures and contextual dependencies within complaints. By applying convolutional filters, CNNs can efficiently extract localized semantic features, offering a more flexible and dynamic representation of textual information compared to TF-IDF.

BERT is integrated into this study due to its superior contextual understanding and ability to capture complex linguistic structures. Unlike TF-IDF and CNN, which rely on fixed-length representations, BERT dynamically encodes word meanings based on surrounding context, making it especially well-suited for analyzing customer complaints, where sentiment, intent, and linguistic subtleties play a crucial role in churn prediction.

The comparison between these models is expected to provide valuable insights into the informative characteristics of complaint texts, assess the feasibility of different computational approaches, and identify the most effective strategies for customer churn prediction and service recovery optimization.

6.7.6 Analysis of Fusion Methods in Multimodal Data Integration

In the field of churn prediction, the fusion of textual and structured data is a decisive factor. This study distinguishes two predominant fusion frameworks: early and intermediate fusion. Our selection of fusion strategy is informed by reviewing extant literature, identifying affordable and accessible methods that may synergise effectively with the collected multimodal dataset.

A list of fusion methods to be evaluated in this paper is presented in Table 6.2, here the output of the fusion module is denoted as m, the textual input is represented as x, the categorical input is denoted as c, and the numeric input is represented as n, concatenation is denoted as $\|$. The following sections provide details of the selected fusion methods.

Table 6.2 The fusion methods for integration of multimodal data.

Integration method	Equation	Fusion
‘Unimodal’	$m = BERT(x\|c\|n)$	Early
Simple concatenation	$m = BERT(x)\|c\|n$	Intermediate
Individual MLPs	$m = BERT(x)\|MLP(c)\|MLP(n)$	Intermediate
Single MLP	$m = BERT(x)\|MLP(c\|n)$	Intermediate
MAG mechanism	$m = x + \alpha h$ $h = g_c \odot (W_c c) + g_n \odot (W_n n) + b_h$ $\alpha = min\left(\frac{\|x\|_2}{\|h\|_2} * \beta, 1\right)$ $g_i = R\left(W_{gi}[i\|x] + b_i\right)$ Where β is a hyperparameter and R is an activation function.	Intermediate
BERT + Random Forests	$m = RF(BERT(x)\|c\|n)$	Intermediate

6.7.6.1 Early fusion methods

Early fusion, or feature-level fusion, integrates raw data from multiple modalities at the outset. Textual data undergoes initial transformation into vectors via a pre-trained BERT model, thereby completing a form of early feature extraction. A practical early fusion approach could involve appending structured data as textual tokens, creating a combined text sequence.

Gu and Budhkar (2021) proposed this rather intuitive fusion method ‘Unimodal’, through which only textual modality is present after the integration. The ‘Unimodal’ fusion functions as an attached module prior to the input layer of BERT, it concatenates the structured data inputs of an instance to the end of the associated text sequence to form a combined text sequence – certainly this should comply with our truncation strategy for meeting the requirement of a maximum 400 tokens in a text sequence. The combined text sequence would then be fed into the BERT model and flows along the normal learning procedure. However, this method, while intuitive, may not fully capture the distinct nature of structured data and could potentially oversimplify the complex relationships within the data.

6.7.6.2 Intermediate fusion methods

Intermediate fusion, or representation-level fusion, unites outputs from modality-specific models at a shared layer. It ranges from simple concatenation of BERT outputs with structured data to more complex structures like MAG mechanism. The intermediate fusion offers a balance between the raw data integration of early fusion and the decision-level combination of late fusion. The following provides principles of each method.

1. **Simple concatenation** involves concatenation of the output of BERT with categorical and numerical inputs. The dynamics of a stepwise intermediate fusion model is presented in Algorithm 1.
2. **Individual MLPs** applies individual multi-layer perceptron (fully connected layers) for independently processing categorical and numeric inputs before being concatenated with the output of BERT.
3. **Single MLP** applies a single MLP to the concatenated categorical and numeric inputs.
4. **MAG mechanism** implements an attention mechanism proposed by Rahman and others (2020). A unified representation is computed through a gated summation of BERT output, categorical and numeric inputs.
5. **BERT + Random Forests** first concatenates the output of BERT with categorical and numeric inputs, then fed to a Random Forests classifier for prediction. The dynamics of a stepwise intermediate fusion model is presented in Algorithm 1.

Algorithm 1 Intermediate fusion with a stepwise structure

Input: *textual data x, tabular data c and n, and pretrained BERT model*

Output: *m, an output of a stepwise intermediate fusion method*

For *sample in textual data* **do**

    Feed forwards the sample in BERT, compute the loss and backpropagate to update BERT parameters;

End
Save the BERT parameters for later use;
For *sample in textual data and tabular data* ***do***
    Loading the saved parameter then freeze the BERT model;
    Feed forwards to obtain the BERT output;
End
Concatenate BERT (x), c, *and* n *to form a unified representation vector;*
Use the unified vector in a stepwise intermediate fusion method to obtain m.

Each method carries trade-offs between computational efficiency, model interpretability, and the ability to capture inter-modal dynamics. Early fusion methods may lose out on the potential benefits of deeper interaction between modalities captured in later fusion stages. Meanwhile, intermediate strategies provide a compromise, allowing for higher level of interaction modelling without deferring all integration to the final decision-making stage. The choice of strategy is thus a critical consideration, the chosen methods for this analysis aim to explore these trade-offs and identify which fusion strategy best captures the intricacies of customer churn behaviours.

6.7.7 Selection of Evaluation Metric

The effectiveness of a predictive model is measured by its ability to accurately classify unseen data, which is critical in the context of customer churn prediction (Alpaydin, 2020). Given the asymmetric costs and consequences of false positives and false negatives in churn prediction, precision, recall, and the F1 score are particularly informative evaluation metrics.

To quantify these performance measures, a confusion matrix is constructed for each model, designating the ‘churn’ category as the positive class. The confusion matrix is a 2*2 table that categorises predictions into four outcomes (as illustrated in Figure 6.3):

- true positives (TP), where predictions accurately identify the churn class,
- true negatives (TN), where predictions correctly recognise the non-churn class,
- false positives (FP), where non-churn cases are mistakenly identified as churn,

- false negatives (FN), where churn cases are incorrectly labelled as non-churn.

From this matrix, several performance metrics are derived: accuracy, precision, recall, and F1 score. Each of these plays a distinct role in model evaluation.

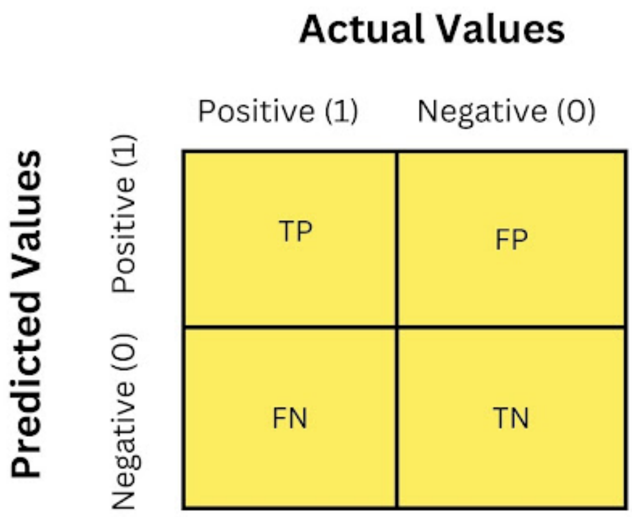

Figure 6.3 An illustration of a confusion matrix

Precision is defined as the ratio of correctly predicted churn cases (TP) to the total predicted churn cases (TP + FP), A high precision indicates that when the model predicts a customer will churn, it is highly likely to be correct. This metric is particularly relevant when the cost of false positives is high, such as in targeted retention campaigns, where incorrectly classifying non-churners as churners may lead to unnecessary customer incentives and financial losses (Sokolova and Lapalme, 2009).

$$Precision = \frac{TP}{TP + FP}$$

Recall measures the proportion of actual churn cases that were correctly identified. This metric is crucial in scenarios where failing to detect true churners carries greater business risk, such as revenue loss due to undetected at-risk customers (Davis and Goadrich, 2006). A high recall ensures that most actual churners are flagged, allowing proactive intervention strategies. However, models optimized for high recall often suffer from lower precision, leading to more false positives.

$$Recall = \frac{TP}{TP + FN}$$

The F1 score provides a balanced measure of precision and recall, particularly useful in cases where class distributions are slightly imbalanced. In churn prediction, while precision and recall are individually important, the F1 score helps assess the trade-off between them, ensuring that neither false positives nor false negatives are disproportionately high (Powers, 2011).

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

Accuracy, defined as the proportion of correctly classified instances among all instances, is calculated as:

$$Accuracy = \frac{TP + TN}{PTP + TN + FP + FN}$$

Although accuracy is a commonly used metric, it can be misleading in imbalanced datasets where one class dominates the other. For example, if 80% of customers are retained and only 20% churn, a model that predicts all customers as “non-churn” would still achieve 80% accuracy, despite being completely ineffective for identifying actual churners.

Churn prediction involves an inherent trade-off between false positives and false negatives, making precision and recall more informative than raw accuracy. A model optimized for high precision ensures that only customers with a high probability of churning are targeted, minimizing unnecessary intervention costs. Conversely, a model emphasizing high recall ensures that most at-risk customers are identified, reducing revenue loss due to undetected churners. The F1 score is selected as a comprehensive performance metric to balance precision and recall, Given the practical implications of churn prediction in business strategy and decision-making, the F1 score provides a robust measure for evaluating model effectiveness.

6.8 Research ethics

According to Saunders et al. (2019), research ethics refers to the standards of conduct that govern the research process, ensuring that the rights of research participants and other affected parties are respected. Ethical considerations must be proactively identified and addressed throughout the research process to maintain integrity, transparency, and accountability.

This study relies on secondary data provided by a client company, meaning that the researcher did not have direct involvement or control over the data collection process. As such, the ethical responsibility for obtaining informed consent, ensuring data privacy, and addressing any potential risks to participants rests with the original data collectors. Furthermore, all personally identifiable information (PII) within the dataset was anonymized before being accessed for this research, ensuring compliance with data protection regulations and minimizing ethical concerns related to confidentiality and participant privacy.

By adhering to these ethical standards, this study ensures that the research process remains responsible, transparent, and compliant with established ethical guidelines by the University of Strathclyde.

6.9 Summary of the chapter

This chapter has outlined the methodological framework employed in this study for customer churn prediction using multimodal data fusion. It began by establishing the research philosophy, where positivism was justified due to the study's reliance on empirical data, machine learning models, and quantitative evaluation. The research approach was then discussed, highlighting the use of deductive reasoning and pattern recognition techniques rather than traditional hypothesis testing.

The research design was structured to systematically develop, validate, and compare different predictive models. The dataset, obtained from a third-party complaint intermediary, was described in detail, emphasizing data integrity, preprocessing, and ethical considerations. The chapter introduced the proposed multimodal predictive framework, designed to integrate textual complaint data and structured variables for improved churn prediction accuracy. A key focus was placed on input selection criteria, ensuring that the model captures not only the initial customer complaint but also remedial actions such as compensation type and resolution time to inform service recovery decision-making.

The feature selection and data preprocessing process was elaborated for both structured and textual data, detailing how numerical, categorical, and textual attributes were prepared for deep learning models. Various machine learning and fusion techniques were explored to assess their effectiveness in integrating multimodal data for churn prediction.

Finally, ethical considerations related to data privacy and research integrity were addressed, ensuring compliance with established ethical standards. By establishing a rigorous methodological foundation, this chapter lays the groundwork for the subsequent experimental analysis, where the proposed approach will be empirically validated and its performance assessed.

7. Experimental results

7.1 Chapter introduction

This chapter presents the experimental evaluation designed to systematically address the five research questions (RQs) formulated in this study. Each section in this chapter corresponds to a specific research question, outlining the experimental setup, model performance, and key findings. This study evaluates model effectiveness through empirical performance metrics, such as F1-score, precision, and recall. These metrics provide an objective measure of how well different feature extraction techniques, machine learning architectures, and data fusion strategies contribute to churn prediction.

Each experiment follows a structured evaluation approach, comparing multiple models to determine the most effective methods for churn prediction. The results are assessed based on predefined success criteria, ensuring a rigorous and reproducible evaluation. The experiments in this study were conducted using Google Colab as the computational platform. To ensure the robustness and generalizability of the churn prediction models, a systematic data partitioning strategy was implemented. The dataset was divided into three subsets: training, validation, and test sets, following best practices in machine learning research.

The training set, comprising 80% of the data, was used to develop and train the models, enabling them to learn complex patterns indicative of customer churn. A validation set, consisting of 10% of the data, served as an intermediary evaluation tool to fine-tune model parameters and prevent overfitting - an essential step in optimizing model performance. The remaining 10% of the data was allocated to the test set, which was completely withheld during both training and validation. This unseen dataset was used to rigorously assess the final model's predictive accuracy, simulating real-world deployment and providing an unbiased evaluation of its generalization capability.

This structured data partitioning approach is critical for validating the churn prediction models, ensuring reproducibility and reliability in the findings. By adopting this methodology, the study adheres to established best practices in data-driven research, strengthening the credibility and practical applicability of the proposed predictive framework.

The chapter is structured as follows: section 7.2 investigates RQ1, focusing on the impact of different text representation techniques (TF-IDF, CNN embeddings, and BERT) on churn prediction; section 7.3 addresses RQ2, comparing the performance of various machine learning models, including MLP, CNN, and BERT-based architectures; section 7.4 explores RQ3, analysing the impact of incorporating structured variables alongside textual complaints; section 7.5 evaluates RQ4, examining the effectiveness of different data fusion techniques (early, intermediate, and late fusion) in multimodal learning; and section 7.6 investigates RQ5, identifying the structured variables that contribute most significantly to churn prediction.

By systematically analysing these experimental results, this chapter provides empirical insights into the relative effectiveness of different machine learning approaches for churn prediction. The findings presented here form the basis for the discussion in Chapter 8, where the implications, limitations, and potential applications of these results are explored in greater depth.

7.2 Evaluating Text Representation Techniques for Churn Prediction

This section presents an experimental evaluation addressing Research Question 1 (RQ1):

'*What are the most effective text representation and feature extraction methods for analysing textual complaint data in customer churn prediction?*'

To systematically assess the impact of different text representation techniques, we compare three approaches: TF-IDF, convolutional and pooling architectures (CNN), and Transformer-based representations (BERT). These techniques are integrated into three respective models:

1) TF-IDF + MLP: A multi-layer perceptron (MLP) model trained on TF-IDF encoded features.

2) CNN Embeddings: A convolutional neural network (CNN) utilizing pre-trained word embeddings.
3) BERT: A Transformer-based model leveraging pre-trained contextual embeddings.

The evaluation aims to determine which method offers the highest predictive accuracy for customer churn classification based on textual complaints. Given that customer churn prediction involves class imbalance (churn cases being relatively less frequent than retained cases), we prioritize metrics that reflect a model's ability to correctly identify churners, rather than solely relying on overall accuracy.

The derived classification performance metrics for the three models are presented in Table 7.1. From these results, we observe that BERT consistently outperforms the other models across key performance metrics. The Transformer-based model achieves the highest recall (0.7689) and F1 score (0.7340), indicating its effectiveness in identifying churners. This suggests that contextualized word representations capture more meaningful semantic information in complaint texts, enhancing predictive power.

Table 7.1 Performance metrics for RQ1

Model	**Accuracy**	**Precision**	**Recall**	**F1 Score**
model_tfidf	0.6268	0.7257	0.6060	0.6605
model_cnn	0.6491	0.6895	0.7534	0.7200
model_bert	0.6662	0.7021	0.7689	0.7340

Conversely, the TF-IDF model (MLP) achieves the highest precision (0.7257), meaning it is better at minimizing false positives. This can be attributed to its ability to identify retained customers with greater confidence, as further demonstrated in the confusion matrix analysis (Figure 7.1).

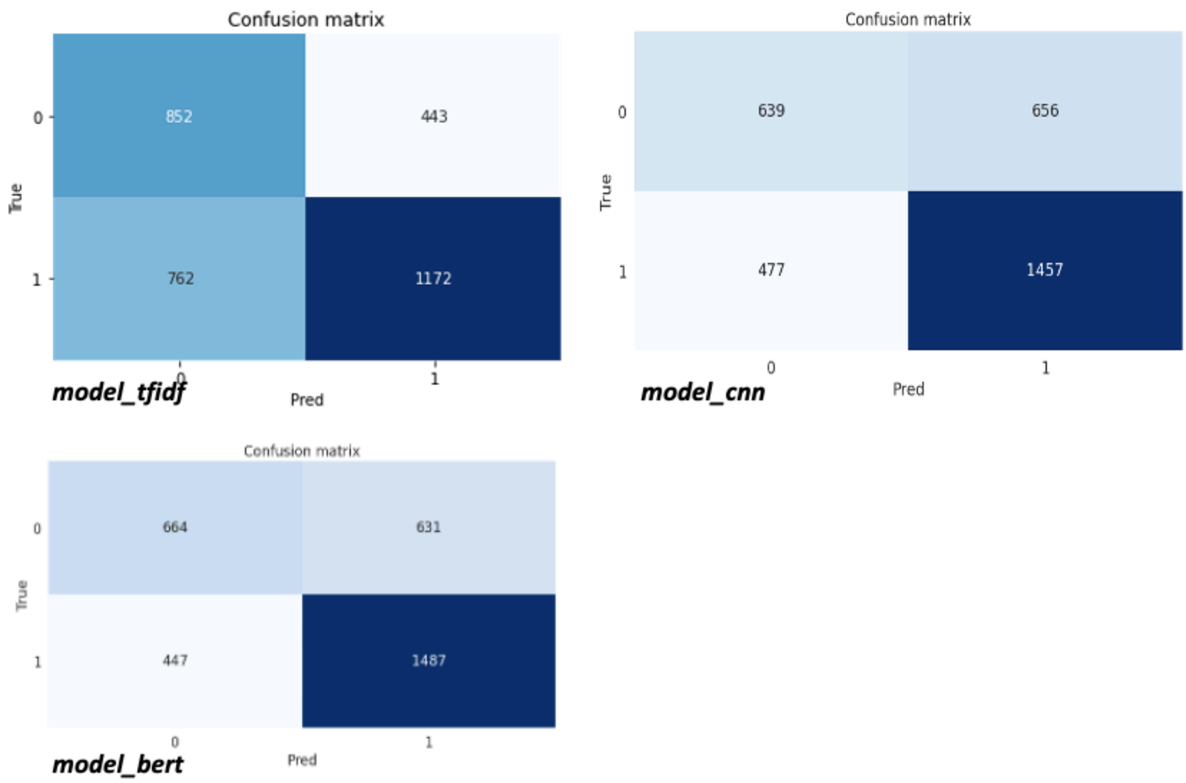

Figure 7.1 Confusion matrices for the three tested models for RQ1

The confusion matrices further illustrate the models' performance differences. While BERT and CNN exhibit higher recall for churn cases - indicating a better ability to detect at-risk customers, TF-IDF has a lower false positive rate, resulting in higher precision but lower recall for churners. Since the primary goal of this study is to develop a proactive churn prediction model that enables companies to identify and retain at-risk customers, recall is a more critical metric than precision. A model with high recall ensures that a larger proportion of churners are correctly flagged, allowing businesses to intervene before customer attrition occurs.

To assess the success of different text representation methods, we define the following success criterion: *A text representation technique is considered superior if it increases the F1-score by at least 1% compared to the baseline (TF-IDF + MLP).*

Despite a longer running needed on BERT models (16 minutes more compared to TF-IDF + MLP), BERT achieves a 7.3% improvement in F1-score over TF-IDF based on the predefined criterion, confirming its effectiveness in churn prediction. This experiment validates the importance of contextualized word representations in analyzing complaint texts and supports the integration of Transformer-based architectures in churn prediction models.

These findings provide empirical evidence addressing RQ1, demonstrating that BERT-based text representations outperform traditional TF-IDF and CNN embeddings in churn classification tasks. The next section explores RQ2, examining how different machine learning architectures compare in predictive performance across various churn behaviours.

7.3 Assessing Ensemble Learning for Churn Prediction Performance

This section presents an experimental evaluation addressing Research Question 2 (RQ2):

'*How do different machine learning models compare in terms of predictive performance across various classes of churn behaviour?*'

Building on the empirical findings from Section 7.1, we explore the potential of ensemble learning to enhance churn prediction accuracy. The prior analysis revealed that count-based text representations (TF-IDF) performed better in identifying retained customers, while contextualized dense embeddings (BERT) were more effective in predicting churners. Given these complementary strengths, we investigate whether a hybrid ensemble model - integrating both feature types - can achieve a more balanced and robust performance.

To leverage the strengths of both count-based and contextualized representations, we construct an ensemble framework that integrates TF-IDF and BERT features via an intermediate fusion strategy. This approach concatenates the output vectors from two models: the 128-dimensional hidden layer output from the TF-IDF-based MLP model, and the 768-dimensional pooled output vector from BERT's embedding layer.

The resulting 896-dimensional feature vector is then fed into a fully connected neural network for final classification. The classification performance of the ensemble model is compared with its individual components (TF-IDF and BERT) and summarized in Table 7.2.

Table 7.2 Performance metrics for RQ2

Model	Accuracy	Precision	Recall	F1 Score
model_tfidf	0.6268	**0.7257**	0.6060	0.6605
model_bert	0.6662	0.7021	**0.7689**	**0.7340**
model_tfidf_bert	**0.6686**	0.7093	0.7570	0.7324

The ensemble model (‘*model_tfidf_bert’*) exhibits a comparable F1-score to BERT while showing a modest improvement in overall accuracy. Although its recall for churn prediction decreases slightly compared to BERT, the ensemble model leverages count-based information to enhance recall for retained customers, leading to a more balanced predictive performance.

To evaluate the effectiveness of ensemble learning, we define the following success criterion: ‘*An ensemble model is considered beneficial if it improves overall accuracy by at least 1% over the best-performing single model while maintaining competitive recall for churn prediction*.’ The rationale behind is that the BERT model serves as the baseline, achieving the highest F1-score (0.7340) and demonstrating strong performance in churn recall. However, BERT’s precision is lower than that of the TF-IDF model, indicating a higher false positive rate - misclassifying retained customers as churners. While the goal of the ensemble model is to leverage the complementary strengths of both feature extraction techniques, balancing the trade-offs between precision and recall improving overall classification performance, rather than optimizing for a single class. Therefore, overall accuracy, as a holistic metric, reflects the proportion of correctly classified instances across all classes. If the ensemble model achieves a higher accuracy, it suggests a more globally stable predictive model that performs consistently well across both churn and retained classes, rather than disproportionately enhancing either recall or precision alone

The ensemble model achieves a 0.24% increase in accuracy over BERT, demonstrating marginal yet positive improvements in predictive performance. However, the trade-off between recall and precision highlights the need for further hyperparameter tuning to optimize feature integration.

These findings provide empirical insights into RQ2, suggesting that ensemble models can offer incremental benefits in churn prediction but may require additional refinements to fully capitalize on the complementary strengths of different text representation methods. The next section investigates RQ3, examining the impact of incorporating structured variables alongside textual complaints.

7.4 Comparing Multimodal and Unimodal Approaches in Churn Prediction

This section addresses Research Question 3 (RQ3):
'*What is the impact of incorporating multimodal data (textual complaints and structured variables) on the performance of churn prediction models?*'

The primary objective of this experiment is to assess whether integrating both textual and structured data enhances churn prediction performance compared to unimodal approaches that rely on a single data type. Understanding the contribution of multimodal architectures is critical for developing more robust and generalizable churn prediction models that leverage the full spectrum of available customer data.

To systematically evaluate this, we compare four models:

1) Unimodal Text Model ('*model_bert*') – A BERT-based model trained exclusively on textual complaint data.
2) Unimodal Structured Model ('*model_rf*') – A Random Forest classifier trained solely on structured tabular data.
3) Early Fusion Model ('*model_unimodal'*) – A model that integrates structured data as textual tokens into the BERT processing pipeline.

4) Intermediate Fusion Model ('*model_concat'*) – A model that concatenates structured data with the BERT-generated feature vectors before final classification.

The BERT model underwent finetuning for 4 epochs, using Adam optimiser (Kingma and Ba, 2015) with a learning rate of $2e^{-5}$, a batch size of 32, and a maximum input sequence length of 400 tokens, and a dropout rate of 0.1 applied across all layers. The Random Forests classifier was configured with 10 trees in the ensemble using default hyperparameters.

Table 7.3 Performance metric for RQ3.

Model	Accuracy	Precision	Recall	F1 Score
model_bert	0.6662	0.7021	0.7689	0.7340
model_rf	0.6312	0.6934	0.6887	0.6911
model_unimodal	**0.6711**	0.7039	**0.7782**	**0.7392**
model_concat	0.6662	**0.7056**	0.7596	0.7316

The results suggest that fusion-based models generally outperform unimodal models, reinforcing the hypothesis that integrating structured and textual data improves predictive accuracy. Notably, the early fusion model ('*model_unimodal'*) achieves the highest recall (0.7782) and F1-score (0.7392), outperforming both unimodal models ('*model_bert*' and '*model_rf'*) and the intermediate fusion model ('*model_concat*').

However, '*model_concat*' performs comparably to '*model_bert*', indicating that concatenation alone does not significantly enhance predictive power. Meanwhile, the structured-only model ('*model_rf*') underperforms across all metrics, with an F1-score of only 0.6911, suggesting that structured data alone is insufficient for effective churn prediction.

To evaluate whether multimodal models provide a meaningful improvement, we define the following success criterion: '*A multimodal model is considered superior if it achieves at least a 1% improvement in F1-score compared to the best-performing unimodal model*.' In this experiment, '*model_unimodal*' achieves a 0.72% increase in F1-score over '*model_bert*'. While this falls slightly below the predefined success threshold, the observed improvements in recall (+1.3%) and overall accuracy (+0.49%) suggest that structured data does contribute useful information, particularly when integrated via early fusion.

The findings provide empirical evidence addressing RQ3, supporting the notion that multimodal architectures are generally more effective than unimodal models, though the extent of improvement depends on how structured and textual data are integrated. The next section (RQ4) investigates the effectiveness of different data fusion strategies in churn prediction.

7.5 Investigating the Impact of Fusion Strategies on Churn Prediction

This section addresses Research Question 4 (RQ4):
'*How effective are different data fusion techniques (early, intermediate, and late fusion) in improving churn prediction accuracy?*'

The objective of this experiment is threefold:

1) To determine which fusion method delivers the best predictive performance when applied to multimodal customer complaint data.
2) To evaluate whether the simple early fusion model ('*model_unimodal*') can compete with more complex fusion architectures, assessing if the additional computational complexity of advanced fusion techniques translates into tangible performance gains.
3) To compare stepwise intermediate fusion models ('*model_ind_mlp*' and '*model_sin_mlp*') against methods using a shared representation layer, investigating whether structured data should be processed independently or integrated into a common latent space.

To achieve these objectives, six fusion-based models were tested using the Multimodal Toolkit (Gu and Budhkar, 2021), ensuring a standardized implementation of fusion strategies.

Before assessing predictive performance, we compare the computational costs associated with each fusion method. Table 7.4 outlines the running time required for each tested fusion frameworks to complete 1 epoch of training. All models are based on the BERT architecture, which necessitates a minimum expectation of approximately 40 minutes per training round. The '*model_mag*' typically requires an additional 10 minutes compared to simpler architectures, attributable to its more complex attention mechanism. Meanwhile, the '*model_bert_rf*' takes about 48 minutes per training epoch, a result of its fusion strategy that incorporates a sequential, two-step modelling process.

Table 7.4 Running time per epoch for fusion techniques

Model	**Running Time**
model_unimodal	39 mins
model_concat	40 mins
model_ind_mlp	41 mins
model_sin_mlp	40 mins
model_mag	**51 mins**
model_bert_rf	48 mins

The experimental results, as shown in Table 7.5, provide critical insights into the effectiveness of different fusion strategies in churn prediction. The attention-based fusion model ('*model_mag*') demonstrates the highest predictive performance, achieving an F1-score of 0.7470 and an accuracy of 0.6823. This result suggests that integrating structured data through an attention-based mechanism enables more effective feature interaction between textual and tabular data, thereby enhancing predictive accuracy. In contrast, the '*model_unimodal*' approach, which encodes structured data as textual tokens and integrates them into the BERT processing pipeline, performs competitively, achieving an F1-score of 0.7392. This performance surpasses that of several more complex fusion models, indicating that early fusion strategies based on direct textual incorporation of structured variables can be a viable alternative to dedicated fusion architectures.

The intermediate fusion methods incorporating MLP-based transformations of structured data (‘*model_ind_mlp*’ and ‘*model_sin_mlp*’) demonstrate a moderate improvement over the simple concatenation approach (‘*model_concat*’). This suggests that structured data benefits from additional transformations before integration, rather than being directly concatenated with textual representations. The ensemble model ‘*model_bert_rf*’, which combines the predictive outputs of BERT and Random Forest, exhibits only a marginal improvement over the unimodal BERT model. This finding highlights the limited independent predictive power of structured data, reinforcing the notion that tabular features contribute meaningfully to churn prediction only when effectively fused with textual data. Overall, while certain fusion strategies yield meaningful improvements in predictive performance, others, particularly simple concatenation-based approaches, fail to provide substantial gains over unimodal models.

Table 7.5 Performance metric for RQ4

Model	**Accuracy**	**Precision**	**Recall**	**F1 Score**
model_unimodal	0.6711	0.7039	0.7782	0.7392
model_concat	0.6662	0.7056	0.7596	0.7316
model_ind_mlp	0.6640	0.6954	0.7813	0.7358
model_sin_mlp	0.6702	0.7050	0.7725	0.7372
model_mag	0.6823	0.7139	0.7834	0.7470
model_bert_rf	0.6692	0.7072	0.7642	0.7346

To establish a rigorous benchmark for evaluating the effectiveness of different fusion methods, a performance improvement threshold is defined: ‘*A fusion method is considered effective if it achieves at least a 1% improvement in F1-score compared to the best-performing unimodal model (‘model_bert’ with F1 = 0.7340).*’ This threshold ensures that any observed performance gains are both statistically and practically meaningful, rather than resulting from minor variations in model training.

Based on this criterion, only the attention-based fusion model ('*model_mag*') meets the predefined success threshold, improving the F1-score by 1.3% over '*model_bert*'. This finding suggests that attention-based mechanisms facilitate a more effective interaction between textual and structured data, leading to a meaningful enhancement in predictive performance. In contrast, all other fusion models, including '*model_concat*' and '*model_bert_rf*', fail to exceed this benchmark, indicating that not all fusion techniques yield substantial benefits over unimodal architectures. The results underscore the importance of selecting fusion strategies that enable meaningful cross-modal interactions, rather than relying on simplistic feature concatenation.

These findings provide empirical support for Research Question 4 (RQ4), demonstrating that while multimodal fusion has the potential to improve churn prediction, its effectiveness is contingent on the specific fusion strategy employed. The next section investigates Research Question 5 (RQ5), focusing on the individual contribution of structured variables to churn prediction and assessing their complementary role in predictive modelling.

7.6 Examining the Contribution of Structured Variables to Churn Prediction

This section addresses Research Question 5 (RQ5):
'*Which structured variables contribute most significantly to churn prediction, and how do they complement textual features in predictive modelling?*'

The objective of this experiment is to assess the individual contribution of structured variables to churn prediction and determine whether specific structured features enhance model performance when integrated with textual complaint data. The analysis focuses on five structured variables identified in Section 6.7.4, including:

1) Emotions expressed in the complaint submission,

2) Previous experience with the service provider (e.g., whether the complainant had previously lodged complaints),
3) Remedial actions taken by the company in response to the complaint,
4) Remedial procedure (i.e., time taken to resolve the complaint),
5) A handcrafted word-count-based feature derived from textual complaint to enhance interpretability.

Given the findings from previous experiments, the 'Unimodal' fusion method is employed in this analysis due to its reliable performance and computational efficiency. While '*model_mag*' has demonstrated superior performance in prior sections, the use of '*model_unimodal*' in this experiment is justified by its ability to integrate structured data into BERT's processing pipeline without explicit contextual associations, thereby enabling a more direct assessment of structured variables within the textual representation space. To isolate the impact of each structured variable, five separate models were constructed, with each model incorporating only one of the structured variables alongside textual features. The results of these models are presented in Table 7.6.

Table 7.6 Performance metrics for RQ5

Model	Accuracy	Precision	Recall	F1 Score
model_bert	0.6662	0.7021	0.7689	0.7340
model_emo	0.6631	0.6999	0.7658	0.7314
model_exp	0.6624	0.6998	0.7642	0.7306
model_action	**0.6702**	0.7049	0.7730	**0.7374**
model_time	0.6692	0.7033	0.7746	0.7372
model_length	0.6674	0.7028	0.7704	0.7351

The results reveal not all structured variables contribute positively to churn prediction. Notably, incorporating 'emotions' and 'previous experience with the service provider' leads to a slight decrease in performance compared to the unimodal BERT model. This suggests that these features may introduce noise rather than useful predictive signals, potentially due to subjective variability in emotional expression or inconsistent reporting of past complaints.

In contrast, ‘remedial actions’, ‘remedial procedure’, and ‘length of complaints’ demonstrates meaningful improvements in predictive performance. The ‘*model_action*’ (F1-score = 0.7374) and ‘*model_time*’ (F1-score = 0.7372) show that service recovery strategies, particularly compensation type and resolution speed, significantly influence churn likelihood. These findings are consistent with prior research in service recovery literature, which highlights financial compensation and resolution time as key drivers of post-service satisfaction and repurchase intent.

To evaluate whether individual structured variables contribute meaningfully to churn prediction, we define the following success criterion: ‘*A structured variable is considered significant if incorporating it into the model improves the F1-score by at least 0.5% over the unimodal BERT model (‘model_bert’ with F1 = 0.7340).*’

Based on this criterion, it is found out that all individual structured variables fail to improve model performance, indicating these attributes may not generalize well enough. These findings provide empirical support for Research Question 5 (RQ5), demonstrating that while structured data can complement textual features in churn prediction, its effectiveness is dependent on the specific variable used.

7.7 Chapter summary

This chapter systematically evaluated the research questions (RQs) by conducting a series of comparative experiments to assess different text representation methods, machine learning models, multimodal architectures, and fusion strategies for churn prediction. The findings highlight the relative effectiveness of various methodologies and provide empirical evidence supporting the research objectives.

The results from RQ1 (Section 7.2) demonstrate that Transformer-based representations (BERT) significantly outperform traditional text feature extraction techniques, such as TF-IDF and CNN embeddings, particularly in recall and F1-score metrics. This underscores the advantage of contextual embeddings in capturing nuanced customer sentiments and complaint patterns.

In RQ2 (Section 7.3), the integration of ensemble models combining TF-IDF and BERT embeddings showed only marginal improvements in accuracy compared to standalone BERT models. This suggests that while ensemble learning can help balance precision and recall trade-offs, it does not necessarily yield substantial gains over contextualized feature representations.

The evaluation of multimodal vs. unimodal architectures in RQ3 (Section 7.4) reveals that combining structured and textual data enhances predictive performance, though the extent of improvement depends on the fusion strategy employed. While early fusion approaches (such as '*model_unimodal*') proved effective, more complex concatenation-based methods exhibited diminishing returns.

For RQ4 (Section 7.5), an investigation into fusion strategies indicated that attention-based mechanisms ('*model_mag*') yielded the most substantial gains in performance, surpassing other integration techniques, including simple concatenation and stepwise feature fusion. This finding suggests that structured and textual data interactions require sophisticated modelling techniques to fully harness their predictive potential.

Finally, RQ5 (Section 7.6) examined the role of individual structured variables in churn prediction. The results indicate that remedial actions and resolution time are the most impactful structured variables, while features related to customer emotions and prior experiences show limited predictive value. These insights emphasize that structured variables can complement textual features, but their utility varies depending on the specific feature set.

These empirical findings provide a strong foundation for the subsequent discussion in Chapter 8, which will critically analyse the broader implications of these results. Chapter 8 will explore how these insights contribute to existing literature, their practical applications for churn management, and the limitations that may inform future research directions.

8. Discussion and Conclusion

8.1 Chapter introduction

This chapter presents a critical discussion of the findings derived from the empirical analysis in Chapter 7. It systematically addresses the research questions (RQs) by interpreting the results in relation to prior literature, theoretical frameworks, and industry practices. Additionally, it evaluates the implications of the study for both academia and managerial practice, highlighting the study's contributions to customer churn prediction, service recovery, and multimodal learning frameworks. The chapter also outlines the study's limitations and provides recommendations for future research. Finally, the chapter concludes with a summary of key insights and their broader implications.

The research was designed to answer five key research questions:

1. What are the most effective text representation and feature extraction methods for analysing textual complaint data in customer churn prediction?
2. How do different machine learning models compare in terms of predictive performance across various classes of churn behaviour?
3. What is the impact of incorporating multimodal data (textual complaints and structured variables) on the performance of churn prediction models?
4. How effective are different data fusion techniques (early, intermediate, and late fusion) in improving churn prediction accuracy?
5. Which structured variables contribute most significantly to churn prediction, and how do they complement textual features in predictive modelling?

By analysing these questions through a multimodal framework, this research provides a comprehensive assessment of textual complaint data in churn prediction, demonstrating the interplay between text representations, predictive models, data fusion techniques, and structured variables.

8.2 Discussion of key findings

8.2.1 Text representation and feature extraction for churn prediction (RQ1)

The first research question aimed to evaluate the effectiveness of different text representation and feature extraction methods in customer churn prediction. The results in Section 7.2 demonstrate that contextualized word embeddings using BERT significantly outperform traditional representations such as TF-IDF and CNN-based encodings, particularly in identifying churn-prone complainants.

These findings reinforce the critical role of text representation techniques in predictive modelling, particularly for customer-generated complaint narratives. Traditional service recovery research has primarily focused on structured variables such as customer demographics, past interactions, and complaint resolution outcomes (Tax et al., 1998; Smith et al., 1999). However, recent advances in text analytics and sentiment-driven churn prediction suggest that deep learning-based text representations can uncover nuanced linguistic patterns linked to customer dissatisfaction and attrition risk (Grégoire and Mattila, 2021; Voorhees et al., 2020). The results of this study confirm this argument, showing that BERT-based embeddings provide a richer understanding of textual complaints, capturing latent semantic structures that are otherwise lost in traditional frequency-based representations like TF-IDF.

However, TF-IDF achieves higher precision than BERT, particularly in correctly identifying retained customers. This aligns with findings from Luo, Liberatore, Niu, and Lotero (2021), who demonstrated that TF-IDF is effective at identifying stable, non-fluctuating customer segments due to its reliance on term frequency rather than contextual semantics. The implication of this finding is that service recovery models may need to incorporate hybrid text representation approaches, combining both frequency-based and contextualized embeddings to optimize predictive accuracy across different customer groups.

8.2.2 Predictive performance of different machine learning models (RQ2)

The second research question sought to compare the predictive performance of different machine learning models for churn classification. The findings in Section 7.3 show that ensemble models, which integrate count-based (TF-IDF) and contextual (BERT) features, offer marginal improvements over unimodal models. However, their increased computational costs raise concerns about practical deployment in real-time business applications.

This result aligns with prior research that has emphasized the trade-offs between deep learning's high predictive power and its operational complexity (Lemmens & Croux, 2006). Studies in customer churn analytics have consistently found that deep learning models, such as CNNs and Transformer-based architectures, outperform traditional classifiers such as logistic regression and decision trees in predictive tasks (Verbeke et al., 2012; Wuest et al., 2020). However, as noted by Nguyen and Sidorova (2018), interpretable models remain a priority in commercial settings, where model transparency is essential for managerial decision-making.

The finding that BERT-based models improve recall for churn prediction but suffer from lower precision aligns with research by Voorhees et al. (2020), who identified that Transformer-based text classifiers tend to overfit sentiment-heavy text sequences, leading to a higher false positive rate. This highlights a crucial limitation of deep learning in business intelligence applications, where companies may prioritize minimizing false churn predictions to avoid unnecessary customer retention costs.

From a managerial perspective, these results suggest that firms must weigh the trade-offs between interpretability, computational efficiency, and prediction accuracy when selecting machine learning architectures for churn modeling. Businesses seeking a balance between predictive performance and practical deployment may benefit from using ensemble strategies that selectively apply deep learning models to high-risk complaints while using simpler classifiers for routine cases.

8.2.3 The impact of multimodal data integration (RQ3)

The third research question examined the impact of multimodal data integration (textual + structured features) in churn prediction models. The findings in Section 7.4 indicate that structured variables alone do not significantly enhance churn prediction performance, particularly when used independently of textual complaint narratives.

This result challenges traditional assumptions in churn modeling, which have historically emphasized the predictive power of customer demographic data, transaction histories, and past interactions (Lemmens & Croux, 2006). Instead, it aligns with recent studies suggesting that customer complaints contain rich sentiment-driven cues that serve as strong standalone predictors of churn (Grégoire & Mattila, 2021).

A key insight from this study is that structured features, when embedded as additional tokens within textual sequences, enhance predictive performance. This supports findings from Baltrušaitis, Ahuja, and Morency (2019), who emphasized that feature alignment in multimodal models is critical for maximizing predictive performance. This result suggests that firms should rethink how they integrate structured data into churn prediction models, shifting from separate feature concatenation approaches to embedded multimodal representations.

8.2.4 Effectiveness of fusion techniques in churn prediction (RQ4)

The fourth research question focused on evaluating different data fusion techniques in multimodal churn prediction. Findings in Section 7.5 reveal that advanced fusion methods such as Modality Attention Gating (MAG) outperform simple concatenation-based approaches.

These results align with prior research demonstrating that simple feature concatenation often leads to suboptimal performance due to feature redundancy and dilution of signal strength (Xu et al., 2015). Instead, attention-based fusion mechanisms, which dynamically weigh the importance of different modalities, are more effective in capturing relevant cross-modal relationships (Tsai et al., 2019).

However, the practical implications of this finding suggest that businesses must assess whether the accuracy gains from complex fusion techniques justify their increased computational costs. While deep fusion architectures offer improved predictive performance, they may not always be viable for real-time implementation in customer service operations.

8.2.5 Contribution of structured variables to churn prediction (RQ5)

The final research question examined the role of structured variables in churn prediction. Findings in Section 7.6 reveal that not all structured attributes contribute equally to predictive performance.

The study identifies remedial actions and procedural fairness as significant predictors of churn, aligning with prior research in service recovery theory (Tax & Brown, 1998). However, customer emotions and past experiences were found to have limited predictive value, suggesting that the complaint text itself encapsulates most of the sentiment-related information.

This challenges traditional service recovery models, which often assume that customer emotion indicators significantly impact post-recovery churn decisions (Grégoire and Mattila, 2021). Instead, the results suggest that compensation type and resolution speed are stronger determinants of churn risk, emphasizing the need for businesses to optimize remedial strategies rather than focusing solely on sentiment analysis.

8.3 Theoretical contributions

This research makes significant theoretical contributions in the areas of service recovery, customer churn prediction, and multimodal machine learning by offering new insights into how textual complaint data can be leveraged for predictive analytics. The findings challenge existing assumptions about customer behaviour modelling, advance the application of deep learning in service recovery research, and provide empirical validation of multimodal learning strategies.

One of the primary theoretical contributions of this study lies in its advancement of text representation methods in customer complaint analytics. Prior research in service recovery and complaint management has often relied on conventional sentiment analysis techniques, which primarily focus on lexicon-based approaches or simple bag-of-words models. However, this research empirically demonstrates that contextualized text representations using deep learning models such as BERT significantly improve the identification of churn-prone complainants. Unlike traditional approaches that treat text as isolated tokens without considering their surrounding context, this study shows that BERT's ability to capture semantic nuances and sentence-level dependencies makes it superior for churn prediction based on textual complaints. This finding supports and extends previous work in natural language processing for customer feedback analysis while reinforcing the importance of deep learning in service-related predictive modelling.

Additionally, this research advances multimodal learning in customer churn prediction by demonstrating the impact of integrating textual complaints with structured customer attributes. In traditional churn modelling, structured variables such as customer demographics, previous service interactions, and transaction history are often considered primary predictors. However, this study finds that textual complaint data alone can achieve comparable, if not superior, performance compared to structured data models. This highlights the latent richness of complaint narratives, where customers often disclose dissatisfaction levels, resolution expectations, and underlying reasons for dissatisfaction. By proving that structured variables do not always enhance predictive performance, this research challenges established assumptions in churn modelling and suggests that future predictive frameworks should focus more on textual analytics rather than heavily relying on predefined customer attributes.

The study also contributes to the empirical validation of fusion strategies for multimodal data integration in churn prediction. A key finding is that naïve concatenation approaches to multimodal fusion do not necessarily yield superior results. Instead, the study demonstrates that structured attributes, when represented as additional tokens within textual sequences, can improve churn prediction accuracy by providing contextual signals within the primary mode of analysis. This insight contributes to the growing literature on representation learning and feature fusion in deep learning, suggesting that the way structured variables are integrated into text-based models fundamentally influences their predictive power.

Furthermore, this research bridges a gap between service recovery theories and machine learning applications. Traditional service recovery frameworks have focused on customer satisfaction and justice theory, emphasizing how firms should manage service failures to retain customers. However, these frameworks lack an empirical, data-driven approach to predicting churn risk based on real-time customer interactions. By introducing a predictive framework that enables businesses to assess churn probability immediately upon receiving a complaint, this study provides a practical extension to service recovery theories, integrating machine learning into the decision-making process.

Finally, this study extends the understanding of feature selection and model interpretability in churn prediction. Existing research in churn modelling often struggles with the explainability of deep learning models, making it difficult for practitioners to trust or deploy these models in real-world settings. By systematically comparing multiple feature extraction techniques, this study not only identifies the most effective text representation methods but also sheds light on the interpretability trade-offs between different machine learning architectures. This work lays a foundation for future research on explainable AI in service recovery and churn prediction, emphasizing the need for models that are both accurate and interpretable for business decision-makers.

8.4 Managerial implications

The findings of this research hold important implications for customer relationship management (CRM), service recovery strategies, and the deployment of AI-driven churn prediction systems. The results emphasize the necessity of rethinking how businesses handle customer complaints, leveraging predictive analytics to proactively address churn risks rather than reacting to customer dissatisfaction after the fact.

One key managerial implication is that companies can prioritize customer retention efforts by integrating text analytics into their CRM systems. The study shows that textual complaints contain rich behavioural and sentiment information that is highly predictive of churn. This suggests that organizations should invest in automated complaint classification systems, where machine learning models can assess complaint severity, urgency, and churn risk in real time. By implementing such a system, firms can assign high-priority cases to experienced service agents or specialized recovery teams, ensuring that customers at risk of churning receive immediate attention.

Another crucial implication relates to resource allocation in service recovery. Many organizations rely on standardized compensation policies and resolution procedures, treating all complaints similarly. However, this study highlights that not all complaints have the same churn risk. By leveraging predictive models, companies can identify which complainants are most likely to churn and allocate higher-value recovery efforts to those individuals. For example, rather than offering generic goodwill gestures, firms could provide targeted retention incentives - such as personalized discounts, account upgrades, or direct managerial intervention - to customers identified as high churn risks.

Additionally, this research provides guidance on optimizing multimodal AI adoption in business settings. While deep learning models such as BERT provide superior accuracy, their computational demands make them challenging for real-time deployment. This study suggests that businesses can achieve near-optimal results using hybrid models, where simpler text-based classifiers handle routine complaints, while BERT-based architectures are reserved for complex, high-risk cases. This tiered approach to AI deployment ensures that businesses balance computational efficiency with predictive performance, making AI-driven CRM solutions scalable and cost-effective.

Moreover, the study's findings challenge existing customer data collection practices. Many organizations over-rely on structured demographic and transactional data for churn prediction, but this research demonstrates that textual complaints alone can be a powerful predictor of churn. This suggests that companies should shift from collecting vast amounts of structured data to investing in more sophisticated NLP-driven customer interaction monitoring.

From a strategic perspective, these insights reinforce the importance of early intervention in service recovery. Predictive churn analytics can help businesses identify service failures before they escalate, allowing firms to proactively engage dissatisfied customers before they decide to leave. This aligns with best practices in customer retention, where the cost of preventing churn is significantly lower than reacquiring lost customers.

8.5 Limitations and future research directions

Despite its contributions, this study has several limitations that should be acknowledged. One key limitation is the computational constraints associated with deep learning models, particularly BERT-based architectures. While these models offer superior predictive accuracy, they require substantial processing power, making them less accessible for small businesses or organizations with limited AI infrastructure. Future research could explore more efficient, lightweight alternatives that retain predictive power while reducing computational costs.

Another limitation concerns the generalizability of the dataset. This study relies on complaint data from a single third-party complaint intermediary, which may not fully represent broader consumer behaviour across industries. Although the dataset spans multiple sectors, future studies should consider expanding the dataset to include company-internal complaints, call centre interactions, and social media complaints to enhance external validity.

The interpretability of deep learning models remains a challenge. While this research compares different feature extraction techniques, the black-box nature of deep learning models remains an obstacle to deployment in regulatory-sensitive industries such as finance and healthcare. Future work should explore explainable AI approaches, such as attention visualization, to improve model transparency.

Finally, while this study focused on multimodal fusion strategies, it did not exhaustively examine time-series dynamics in churn prediction. Customer churn is an evolving process, influenced by cumulative interactions over time. Future research could integrate longitudinal customer interaction data, exploring how customer sentiment shifts over multiple complaint instances and whether churn risk fluctuates dynamically.

By addressing these limitations, future research can further refine AI-driven service recovery models, improving their practical utility for businesses while expanding their theoretical foundations in predictive analytics and customer retention.

Appendix A Initial Complaint Survey

1. **What is the name of the company?**

2. **What is the sector of the company?**

3. **What is the service type?**

4. **What is the issue type?**

5. **What is the issue name**

6. **What is your feeling at the time of complaining on a scale from 1 (very unhappy) to 7 (very happy)?**

7. **Please describe what happened?**

8. **Please describe how do you want the issue to be resolved?**

9. **Please describe the impact of the issue on you.**

10. **Please indicate whether you have previously complained on the same issue before.**

Appendix B Feedback Survey

1. **Please indicate whether you have discontinued or intend to discontinue your service.**

2. **Please indicate how the issue was resolved.**

3. **Please indicate the compensation amount if applicable.**

4. **What is your feeling at the time of being resolved on a scale from 1 (very unhappy) to 7 (very happy)?**

5. **On a scale from 1 (very unlikely) to 7 (very likely), how likely you would recommend the company to s a friend or colleague?**

6. **On a scale of 1 (very difficult) to 7 (very easy), how easy was it to resolve your issue?**

7. **On a scale of 1 (very dissatisfied) to 7 (very satisfied), how would you rate your overall experience with our product/service?**

Appendix C Modelling Results

Modelling Performance of 'model_ifidf'

```
Accuracy: 0.63
Detail:
              precision    recall  f1-score   support

           0     0.5279    0.6579    0.5858      1295
           1     0.7257    0.6060    0.6605      1934

    accuracy                         0.6268      3229
   macro avg     0.6268    0.6320    0.6231      3229
weighted avg     0.6464    0.6268    0.6305      3229
```

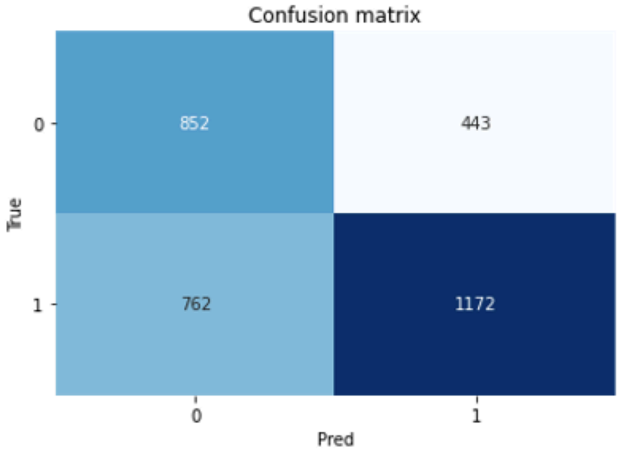

Figure A.1 Performance stats of 'model_tfidf'

Modelling Performance of 'model_cnn'

```
Accuracy: 0.649
Detail:
              precision    recall  f1-score   support

           0     0.5726    0.4934    0.5301      1295
           1     0.6895    0.7534    0.7200      1934

    accuracy                         0.6491      3229
   macro avg     0.6311    0.6234    0.6251      3229
weighted avg     0.6426    0.6491    0.6439      3229
```

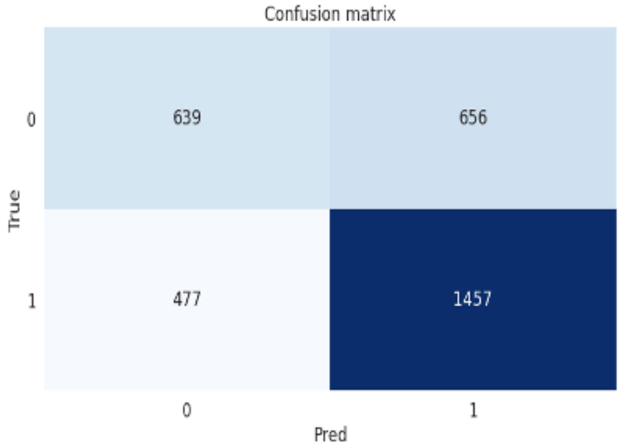

Figure A.2 Performance stats of 'model_cnn'

Modelling Performance of 'model_bert'

```
Accuracy: 0.67
Auc: 0.64
Detail:
              precision    recall  f1-score   support

           0     0.5977    0.5127    0.5520      1295
           1     0.7021    0.7689    0.7340      1934

    accuracy                         0.6662      3229
   macro avg     0.6499    0.6408    0.6430      3229
weighted avg     0.6602    0.6662    0.6610      3229
```

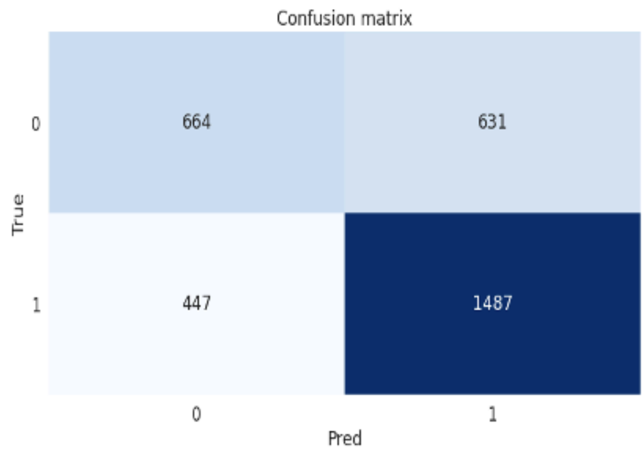

Figure A.3 Performance stats of 'model_bert'

Modelling Performance of 'model_bert_tfidf'

```
Accuracy: 0.67
Auc: 0.65
Detail:
              precision    recall  f1-score   support

           0     0.5966    0.5367    0.5650      1295
           1     0.7093    0.7570    0.7324      1934

    accuracy                         0.6686      3229
   macro avg     0.6529    0.6468    0.6487      3229
weighted avg     0.6641    0.6686    0.6653      3229
```

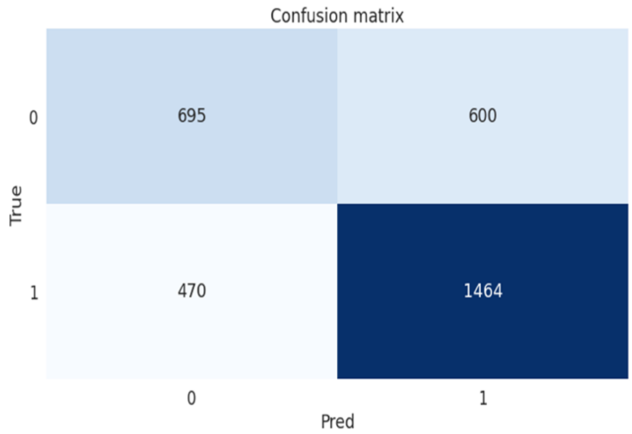

Figure A.4 Performance stats of 'model_bert_tfidf'

Modelling Performance of 'model_rf'

```
Accuracy: 0.631
Detail:
              precision    recall  f1-score   support

           0     0.5398    0.5452    0.5425      1295
           1     0.6934    0.6887    0.6911      1934

    accuracy                         0.6312      3229
   macro avg     0.6166    0.6170    0.6168      3229
weighted avg     0.6318    0.6312    0.6315      3229
```

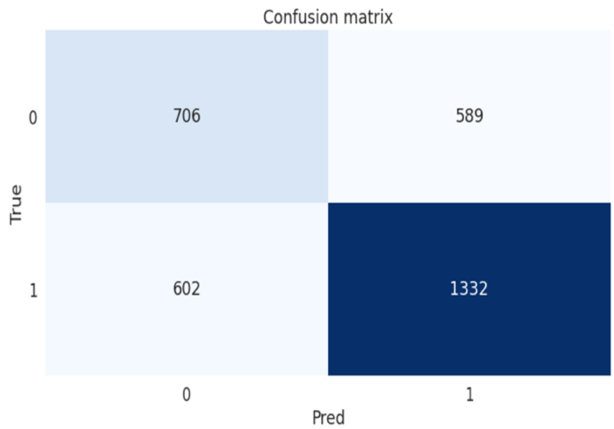

Figure A.5 Performance stats of 'model_rf'

Modelling Performance of 'model_unimodal'

```
Accuracy: 0.67
Auc: 0.64
Detail:
              precision    recall  f1-score   support

           0     0.6068    0.5112    0.5549      1295
           1     0.7039    0.7782    0.7392      1934

    accuracy                         0.6711      3229
   macro avg     0.6554    0.6447    0.6470      3229
weighted avg     0.6650    0.6711    0.6653      3229
```

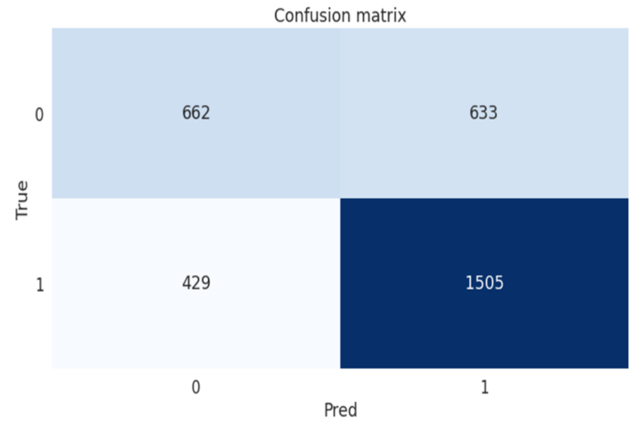

Figure A.6 Performance stats of 'model_unimodal'

Modelling Performance of 'model_concat'

```
Accuracy: 0.67
Detail:
              precision    recall  f1-score   support

           0     0.5946    0.5266    0.5586      1295
           1     0.7056    0.7596    0.7316      1934

    accuracy                         0.6662      3229
   macro avg     0.6501    0.6431    0.6451      3229
weighted avg     0.6611    0.6662    0.6622      3229
```

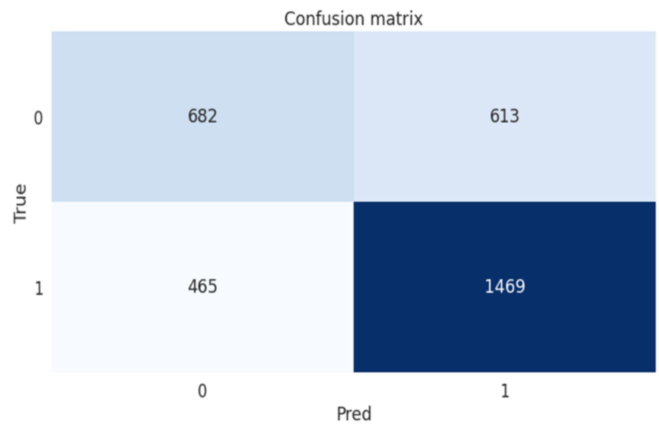

Figure A.7 Performance stats of 'model_concat'

Modelling Performance of 'model_ind_mlp'

```
Accuracy: 0.66
Auc: 0.64
Detail:
              precision    recall  f1-score   support

           0     0.5994    0.4888    0.5385      1295
           1     0.6954    0.7813    0.7358      1934

    accuracy                         0.6640      3229
   macro avg     0.6474    0.6350    0.6372      3229
weighted avg     0.6569    0.6640    0.6567      3229
```

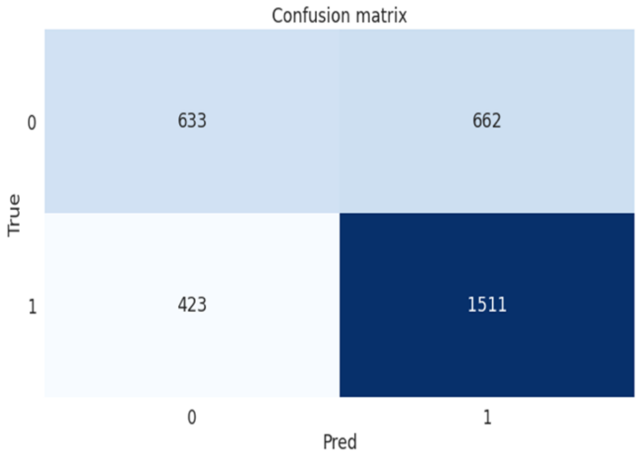

Figure A.8 Performance stats of 'model_ind_mlp'

Modelling Performance of 'model_sin_mlp'

Accuracy: 0.67
Auc: 0.64
Detail:

	precision	recall	f1-score	support
0	0.6036	0.5174	0.5572	1295
1	0.7050	0.7725	0.7372	1934
accuracy			0.6702	3229
macro avg	0.6543	0.6449	0.6472	3229
weighted avg	0.6644	0.6702	0.6650	3229

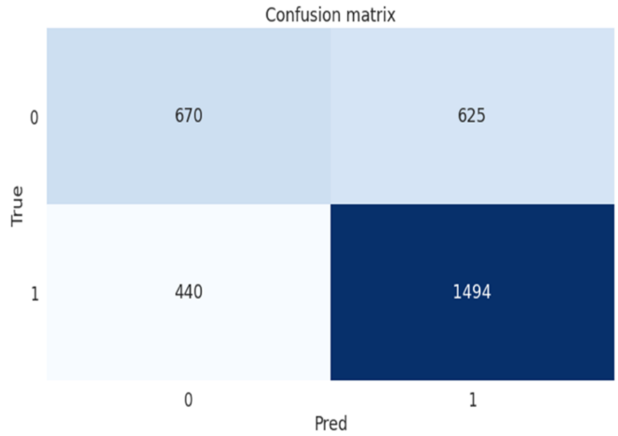

Figure A.9 Performance stats of 'model_sin_mlp'

Modelling Performance of 'model_mag'

Accuracy: 0.68
Auc: 0.66
Detail:

	precision	recall	f1-score	support
0	0.6215	0.5313	0.5729	1295
1	0.7139	0.7834	0.7470	1934
accuracy			0.6823	3229
macro avg	0.6677	0.6573	0.6599	3229
weighted avg	0.6769	0.6823	0.6772	3229

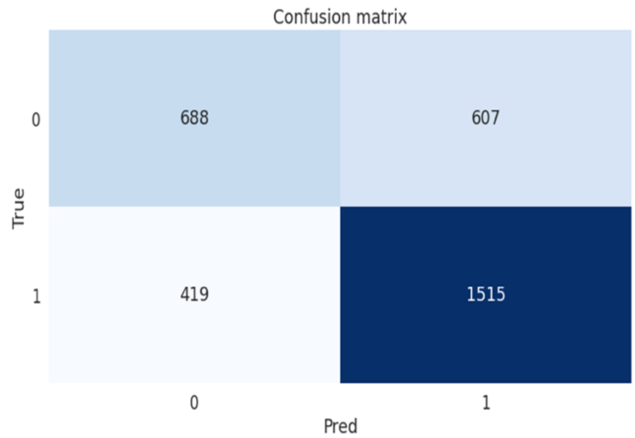

Figure A.10 Performance stats of 'model_mag'

Modelling Performance of 'model_bert_rf'

```
Accuracy: 0.669
Detail:
              precision    recall  f1-score   support

           0     0.5996    0.5274    0.5612      1295
           1     0.7072    0.7642    0.7346      1934

    accuracy                         0.6692      3229
   macro avg     0.6534    0.6458    0.6479      3229
weighted avg     0.6641    0.6692    0.6651      3229
```

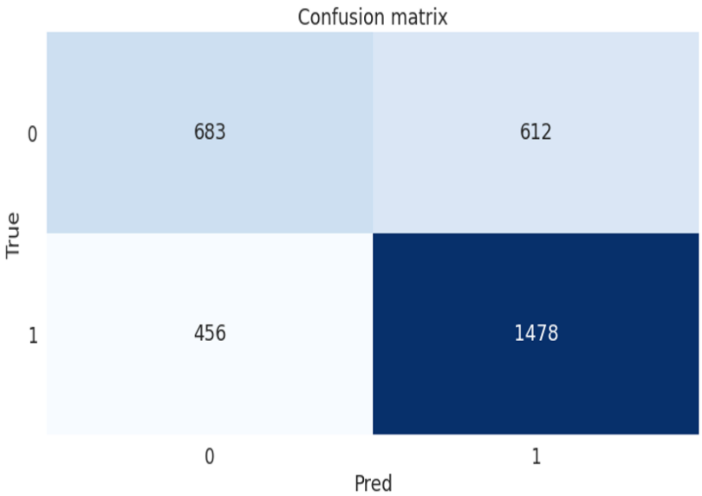

Figure A.11 Performance stats of 'model_bert_rf'

Bibliography

Adams, J. S. (1963). Towards an understanding of inequity. *The Journal of Abnormal and Social Psychology*, *67*(5), 422.

Alaparthi, S., & Mishra, M. (2021). BERT: A sentiment analysis odyssey. *Journal of Marketing Analytics*, *9*(2), 118–126.

Alpaydin, E. (2020). *Introduction to machine learning*. MIT press.

Anderson, E. W., Fornell, C., & Lehmann, D. R. (1994). Customer satisfaction, market share, and profitability: Findings from Sweden. *Journal of Marketing*, *58*(3), 53–66.

Anderson, E. W., Fornell, C., & Rust, R. T. (1997). Customer satisfaction, productivity, and profitability: Differences between goods and services. *Marketing Science*, *16*(2), 129–145.

Average customer retention rate by industry, CustomerGauge. (2024b) Available at: *https://customergauge.com/blog/average-customer-retention-rate-by-industry/*

Averill, J. R. (1999). Individual differences in emotional creativity: Structure and correlates. *Journal of Personality*, *67*(2), 331–371.

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.

Banga, D., & Peddireddy, K. (2023). Artificial intelligence for customer complaint management. *International Journal of Computer Trends and Technology*, 71(3), 1-6.

Bateson, J. E. (1999). the Service Experience. *Handbook of Services Marketing and Management*, 127.

Bearden, W. O., & Mason, J. B. (1984). An investigation of influences on consumer complaint reports. *ACR North American Advances*.

Bearden, W. O., & Teel, J. E. (1983). Selected determinants of consumer satisfaction and complaint reports. *Journal of Marketing Research*, *20*(1), 21–28.

Becker, L., & Jaakkola, E. (2020). Customer experience: Fundamental premises and implications for research. *Journal of the Academy of Marketing Science*, 48, 630-648.

Bell, C., & Jones, K. P. (1979). Towards everyday language information retrieval systems via minicomputers. *Journal of the American Society for Information Science*, *30*(6), 334–339.

Bell, C. R., & Ridge, K. (1992). Service for recovery trainers. *Training & Development*, *46*(5), 58–63.

Bell, C. R., & Zemke, R. E. (1987). Service breakdown: the road to recovery. *Management Review*, *76*(10), 32.

Bell, E., Bryman, A., & Harley, B. (2019). Business research methods. *Oxford University Press.*

Bendall-Lyon, D., & Powers, T. L. (2001). The role of complaint management in the service recovery process. *The Joint Commission Journal on Quality Improvement*, *27*(5), 278–286.

Bengio, Y., Courville, A., & Vincent, P. (2013). Representation learning: A review and new perspectives. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *35*(8), 1798–1828.

Benoit, D. F., & Van den Poel, D. (2012). Improving customer retention in financial services using kinship network information. *Expert Systems with Applications*, *39*(13), 11435–11442.

Bhaskar, R. (1978). A Realist Theory of Science. *Harvester Press.*

Bhat, S., & Culotta, A. (2017). Identifying leading indicators of product recalls from online reviews using positive unlabeled learning and domain adaptation. *Proceedings of the International AAAI Conference on Web and Social Media*, *11*(1), 480–483.

Bitner, M. J. (1990). Evaluating service encounters: the effects of physical surroundings and employee responses. *Journal of Marketing*, *54*(2), 69–82.

Bitner, M. J., Booms, B. H., & Tetreault, M. S. (1990). The service encounter: diagnosing favorable and unfavorable incidents. *Journal of Marketing*, *54*(1), 71–84.

Blodgett, J. G., Granbois, D. H., & Walters, R. G. (1993). The effects of perceived justice on complainants' negative word-of-mouth behavior and repatronage intentions. *Journal of Retailing*, *69*(4), 399–428.

Blodgett, J. G., Hill, D., & Bakir, A. (2006). Cross-cultural complaining behavior? An alternative explanation. *Journal of Consumer Satisfaction, Dissatisfaction and Complaining Behavior*, *19*, 103–117.

Blodgett, J. G., Hill, D. J., & Tax, S. S. (1997). The effects of distributive, procedural, and interactional justice on postcomplaint behavior. *Journal of Retailing*, *73*(2), 185–210.

Blodgett, J. G., Wakefield, K. L., & Barnes, J. H. (1995). The effects of customer service on consumer complaining behavior. *Journal of Services Marketing*, *9*(4), 31–42.

Blume, E. R. (1988). Customer service: giving companies the competitive edge. *Training & Development Journal*, *42*(9), 24–31.

Bodey, K., & Grace, D. (2007). Contrasting "complainers" with "non-complainers" on attitude toward complaining, propensity to complain, and key personality characteristics: A nomological look. *Psychology & Marketing*, *24*(7), 579–594.

Bolton, R. N. (1998). A dynamic model of the duration of the customer's relationship with a continuous service provider: The role of satisfaction. *Marketing Science*, *17*(1), 45–65.

Bolton, R. N., & Lemon, K. N. (1999). A dynamic model of customers' usage of services: Usage as an antecedent and consequence of satisfaction. *Journal of Marketing Research*, *36*(2), 171–186.

Borisov, V., Leemann, T., Seßler, K., Haug, J., Pawelczyk, M., & Kasneci, G. (2022). Deep neural networks and tabular data: A survey. *IEEE Transactions on Neural Networks and Learning Systems*.

Boshoff, C. (1997). An experimental study of service recovery options. *International Journal of Service Industry Management*, *8*(2), 110–130.

Boshoff, C. (2012). Can service firms overdo service recovery? An assessment of non-linearity in service recovery satisfaction. *South African Journal of Business Management*, *43*(3), 1–12.

Bougie, R., Pieters, R., & Zeelenberg, M. (2003). Angry customers don't come back, they get back: The experience and behavioral implications of anger and dissatisfaction in services. *Journal of the Academy of Marketing Science*, *31*(4), 377–393.

Boulahia, S. Y., Amamra, A., Madi, M. R., & Daikh, S. (2021). Early, intermediate and late fusion strategies for robust deep learning-based multimodal action recognition. *Machine Vision and Applications*, *32*(6), 121.

Boyd, D., & Crawford, K. (2012). Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon. *Information, communication & society*, 15(5), 662-679.

Bradley, G., & Sparks, B. (2012). Explanations: if, when, and how they aid service recovery. *Journal of Services Marketing*, *26*(1), 41–51.

Brashers, D. E. (2001). Communication and uncertainty management. *Journal of Communication*, *51*(3), 477–497.

Broadbridge, A., & Marshall, J. (1995). Consumer complaint behaviour: the case of electrical goods. *International Journal of Retail & Distribution Management*, *23*(9), 8–18.

Brown, S. W., & Swartz, T. A. (1989). A gap analysis of professional service quality. *Journal of Marketing*, *53*(2), 92–98.

Bryman, A. (2016). Social Research Methods. *Oxford University Press*.

Cabanac, M. (2002). What is emotion? *Behavioural Processes*, *60*(2), 69–83.

Caruana, A. (2002). Service loyalty: The effects of service quality and the mediating role of customer satisfaction. *European Journal of Marketing*, *36*(7/8), 811–828.

Chang, H. H., Tsai, Y.-C., Wong, K. H., Wang, J. W., & Cho, F. J. (2015). The effects of response strategies and severity of failure on consumer attribution with regard to negative word-of-mouth. *Decision Support Systems*, *71*, 48–61.

Chang, Y.-W., & Chang, Y.-H. (2010). Does service recovery affect satisfaction and customer loyalty? An empirical study of airline services. *Journal of Air Transport Management*, *16*(6), 340–342.

Chebat, J.-C., Davidow, M., & Codjovi, I. (2005). Silent voices: why some dissatisfied consumers fail to complain. *Journal of Service Research*, *7*(4), 328–342.

Chebat, J.-C., & Slusarczyk, W. (2005). How emotions mediate the effects of perceived justice on loyalty in service recovery situations: an empirical study. *Journal of Business Research*, *58*(5), 664–673.

Chen, T., Ma, K., Bian, X., Zheng, C., & Devlin, J. (2018). Is high recovery more effective than expected recovery in addressing service failure?—A moral judgment perspective. *Journal of Business Research*, *82*, 1–9.

Cheng, V. T. P., & Loi, M. K. (2014). Handling negative online customer reviews: the effects of elaboration likelihood model and distributive justice. *Journal of Travel & Tourism Marketing*, *31*(1), 1–15.

Cho, Y., Im, I., Hiltz, R., & Fjermestad, J. (2002). An analysis of online customer complaints: implications for web complaint management. *Proceedings of the 35th Annual Hawaii International Conference on System Sciences*, 2308–2317.

Chuang, S.-C., Cheng, Y.-H., Chang, C.-J., & Yang, S.-W. (2012). The effect of service failure types and service recovery on customer satisfaction: a mental accounting perspective. *The Service Industries Journal*, *32*(2), 257–271.

Chumwatana, T., & Chuaychoo, I. (2016). Automatic filtering non-English complaint emails in tourism industry using N-gram extraction and classification techniques. *2016 4th International Symposium on Computational and Business Intelligence (ISCBI)*, 216–220.

Collobert, R., Weston, J., Bottou, L., Karlen, M., Kavukcuoglu, K., & Kuksa, P. (2011). Natural language processing (almost) from scratch. *Journal of Machine Learning Research*, *12*(ARTICLE), 2493–2537.

Colquitt, J. A. (2001). On the dimensionality of organizational justice: a construct validation of a measure. *Journal of Applied Psychology*, *86*(3), 386.

Colquitt, J. A., Conlon, D. E., Wesson, M. J., Porter, C. O. L. H., & Ng, K. Y. (2001). Justice at the millennium: a meta-analytic review of 25 years of organizational justice research. *Journal of Applied Psychology*, *86*(3), 425.

Comte, A. (1853). The Positive Philosophy of Auguste Comte. *Calvin Blanchard*.

Coussement, K., & Van den Poel, D. (2008a). Churn prediction in subscription services: An application of support vector machines while comparing two parameter-selection techniques. *Expert Systems with Applications*, *34*(1), 313–327.

Coussement, K., & Van den Poel, D. (2008b). Improving customer complaint management by automatic email classification using linguistic style features as predictors. *Decision Support Systems*, *44*(4), 870–882.

Cranage, D. A., & Mattila, A. S. (2006). Service recovery and pre-emptive strategies for service failure: Both lead to customer satisfaction and loyalty, but for different reasons. *Journal of Hospitality & Leisure Marketing*, *13*(3–4), 161–181.

Creswell, J. W. (2014). Research Design: Qualitative, Quantitative, and Mixed Methods Approaches. *Sage publications*.

Creswell, J. W., & Creswell, J. D. (2018). Research Design: Qualitative, Quantitative, and Mixed Methods Approaches. *Sage publications*.

Crosby, L. A., & Stephens, N. (1987). Effects of relationship marketing on satisfaction, retention, and prices in the life insurance industry. *Journal of Marketing Research*, *24*(4), 404–411.

CUI, X., CAI, S., & FENG, J. (2017). Ontology-based similarity measurement for online complaint cases of communications services. *Systems Engineering—Theory & Practice*, *37*(6), 6.

Cui, X., Cai, S., & Qin, Y. (2017). Similarity-based approach for accurately retrieving similar cases to intelligently handle online complaints. *Kybernetes*, *46*(7), 1223–1244.

Customer service: Churn rate by industry U.S. | Statista. (2024) Available at: *https://www.statista.com/statistics/816735/customer-churn-rate-by-industry-us/*

Dalci, I., & Kosan, L. (2012). Theory of constraints thinking-process tools facilitate goal achievement for hotel management: A case study of improving customer satisfaction. *Journal of Hospitality Marketing & Management*, *21*(5), 541–568.

Danaher, P. J., & Mattsson, J. (1994). Cumulative encounter satisfaction in the hotel conference process. *International Journal of Service Industry Management*, *5*(4), 69–80.

Danermark, B., Ekstrom, M., Jakobsen, L., & Karlsson, J. C. (2005). Explaining Society: Critical Realism in the Social Sciences. *Routledge*.

Dasgupta, T., Dey, L., & Verma, I. (2016). Fuzzy multi-label classification of customer complaint logs under noisy environment. *Rough Sets: International Joint Conference, IJCRS 2016, Santiago de Chile, Chile, October 7–11, 2016, Proceedings*, 376–385.

Davidow, M. (2003). Have you heard the word? The effect of word of mouth on perceived justice, satisfaction and repurchase intentions following complaint handling. *Journal of Consumer Satisfaction, Dissatisfaction and Complaining Behavior*, *16*, 67–80.

Davis, J., & Goadrich, M. (2006). The relationship between Precision-Recall and ROC curves. *Proceedings of the 23rd International Conference on Machine Learning*, 233–240.

Day, R. L. (1984). Modelling choices among alternative responses to dissatisfaction. *ACR North American Advances*.

Day, R. L., Grabicke, K., Schaetzle, T., & Staubach, F. (1981). The hidden agenda of consumer complaining. *Journal of Retailing*.

Day, R. L., & Landon, E. L. (1977). Toward a theory of consumer complaining behavior. *Consumer and Industrial Buying Behavior*, *95*(1), 425–437.

De Caigny, A., Coussement, K., De Bock, K. W., & Lessmann, S. (2020). Incorporating textual information in customer churn prediction models based on a convolutional neural network. *International Journal of Forecasting*, *36*(4), 1563–1578.

De Keyser, A., Verleye, K., Lemon, K. N., Keiningham, T. L., & Klaus, P. (2020). Moving the Customer Experience Field Forward: Introducing the Touchpoints, Context, Qualities (TCQ) Nomenclature. *Journal of Service Research*, 23(4), 433-455. https://doi.org/10.1177/1094670520928390

De Ruyter, K., & Wetzels, M. (2000). Customer equity considerations in service recovery: a cross-industry perspective. *International Journal of Service Industry Management*, *11*(1), 91–108.

del Río-Lanza, A. B., Vázquez-Casielles, R., & Díaz-Martín, A. M. (2009). Satisfaction with service recovery: Perceived justice and emotional responses. *Journal of Business Research*, *62*(8), 775–781.

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*.

Dewey, J. (1931). Philosophy and Civilization. *Minton, Balch & Company*.

Dey, P. K., Hariharan, S., & Ho, W. (2009). Innovation in healthcare services: a customer-focused approach. *International Journal of Innovation and Learning*, *6*(4), 387–405.

Ding, M.-C., Ho, C.-W., & Lii, Y.-S. (2015). Is corporate reputation a double-edged sword? Relative effects of perceived justice in airline service recovery. *International Journal of Economics and Business Research*, *10*(1), 1–17.

Dubé, L., & Morgan, M. S. (1998). Capturing the dynamics of in-process consumption emotions and satisfaction in extended service transactions. *International Journal of Research in Marketing*, *15*(4), 309–320.

Easton, G. (2010). 'Critical realism in case study research', *Industrial Marketing Management*, 39(1), pp. 118-128.

Eiglier, P., & Langeard, E. (1987). *Servuction, Strategie et management*. McGraw-Hill, Paris.

Elizabeth, C. (1993). An investigation into the relationship of fairness and customer satisfaction with service. *Fairness in the Workplace: Approaching Fairness in Human Resource Management*.

Etzel, M. J., & Silverman, B. I. (1981). A managerial perspective on directions for retail customer dissatisfaction research. *Journal of Retailing*.

Fauconnier, J.-P. (2015). French word embeddings. *URL: Http://Fauconnier. Github. Io*.

Feilzer, M. Y. (2010). 'Doing mixed methods research pragmatically: Implications for the rediscovery of pragmatism as a research paradigm', *Journal of Mixed Methods Research*, 4(1), pp. 6-16.

Festinger, L. (1962). Cognitive dissonance. *Scientific American*, *207*(4), 93–106.

Fisher, J. E., Garrett, D. E., Arnold, M. J., & Ferris, M. E. (1999). Dissatisfied consumers who complain to the Better Business Bureau. *Journal of Consumer Marketing*, *16*(6), 576–589.

Fletcher, A. J. (2017). 'Applying critical realism in qualitative research: Methodology meets method', *International Journal of Social Research Methodology*, 20(2), pp. 181-194.

Folkes, V. S. (1984). Consumer reactions to product failure: An attributional approach. *Journal of Consumer Research*, *10*(4), 398–409.

Folkes, V. S., Koletsky, S., & Graham, J. L. (1987). A field study of causal inferences and consumer reaction: the view from the airport. *Journal of Consumer Research*, *13*(4), 534–539.

Fornell, C. (1992). A national customer satisfaction barometer: The Swedish experience. *Journal of Marketing*, *56*(1), 6–21.

Fornell, C., & Wernerfelt, B. (1987). Defensive marketing strategy by customer complaint management: a theoretical analysis. *Journal of Marketing Research*, *24*(4), 337–346.

Forster, J., & Entrup, B. (2017). A cognitive computing approach for classification of complaints in the insurance industry. *IOP Conference Series: Materials Science and Engineering*, *261*(1), 012016.

Fournier, S., & Avery, J. (2011). Putting the relationship back into CRM. *MIT Sloan Management Review*, *52*(3), 63.

Fox, G. L. (2008). Getting good complaining without bad complaining. *Journal of Consumer Satisfaction, Dissatisfaction and Complaining Behavior*, *21*, 27–40.

Frakes, W. (1992). Introduction to information storage and retrieval systems. *Space*, *14*(10).

Fung, G. P. C., Yu, J. X., Lu, H., & Yu, P. S. (2005). Text classification without negative examples revisit. *IEEE Transactions on Knowledge and Data Engineering*, *18*(1), 6–20.

Galitsky, B. A., González, M. P., & Chesñevar, C. I. (2009). A novel approach for classifying customer complaints through graphs similarities in argumentative dialogues. *Decision Support Systems*, *46*(3), 717–729.

Gan, M., Dou, X., & Jiang, R. (2013). From ontology to semantic similarity: calculation of ontology-based semantic similarity. *The Scientific World Journal*, *2013*.

Ganesh, J., Arnold, M. J., & Reynolds, K. E. (2000). Understanding the customer base of service providers: an examination of the differences between switchers and stayers. *Journal of Marketing*, *64*(3), 65–87.

Geiler, L., Affeldt, S., & Nadif, M. (2022). An effective strategy for churn prediction and customer profiling. *Data & Knowledge Engineering*, *142*, 102100.

Gelbrich, K., & Roschk, H. (2011). Do complainants appreciate overcompensation? A meta-analysis on the effect of simple compensation vs. overcompensation on post-complaint satisfaction. *Marketing Letters*, *22*, 31–47.

George, G., Osinga, E. C., Lavie, D., & Scott, B. A. (2016). Big data and data science methods for management research. In *Academy of Management Journal* (Vol. 59, Issue 5, pp. 1493–1507). Academy of Management Briarcliff Manor, NY.

Ghalandari, K., Babaeinia, L., & Jogh, M. G. G. (2012). Investigation of the effect of perceived justice on post-recovery overall satisfaction, post-recovery revisit intention and post-recovery word-of-mouth intention from airline industry in Iran: The role of corporate image. *World Applied Sciences Journal*, *18*(7), 957–970.

Goldberg, Y. (2016). A primer on neural network models for natural language processing. *Journal of Artificial Intelligence Research*, *57*, 345–420.

Goodwin, C., & Ross, I. (1992). Consumer responses to service failures: Influence of procedural and interactional fairness perceptions. *Journal of Business Research*, *25*(2), 149–163.

Grégoire, Y., & Fisher, R. J. (2008). Customer betrayal and retaliation: when your best customers become your worst enemies. *Journal of the Academy of Marketing Science*, *36*, 247–261.

Grönroos, C. (1984). A service quality model and its marketing implications. *European Journal of Marketing*, *18*(4), 36–44.

Gronroos, C. (1994). From marketing mix to relationship marketing: towards a paradigm shift in marketing. *Asia-Australia Marketing Journal*, *2*(1), 9–29.

Grönroos, C. (2000). Creating a relationship dialogue: communication, interaction and value. *The Marketing Review*, *1*(1), 5–14.

Gruber, T. R. (1995). Toward principles for the design of ontologies used for knowledge sharing? *International Journal of Human-Computer Studies*, *43*(5–6), 907–928.

Gu, K., & Budhkar, A. (2021). A package for learning on tabular and text data with transformers. *Proceedings of the Third Workshop on Multimodal Artificial Intelligence*, 69–73.

Gunawan, D., Siregar, R. P., Rahmat, R. F., & Amalia, A. (2018). Building automatic customer complaints filtering application based on Twitter in Bahasa Indonesia. *Journal of Physics: Conference Series*, *978*(1), 012119.

Gursoy, D., Ekiz, E. H., & Chi, C. G. (2007). Impacts of organizational responses on complainants' justice perceptions and post-purchase behaviors. *Journal of Quality Assurance in Hospitality & Tourism*, *8*(1), 1–25.

Guzella, T. S., & Caminhas, W. M. (2009). A review of machine learning approaches to spam filtering. *Expert Systems with Applications*, *36*(7), 10206–10222.

Gyung Kim, M., Wang, C., & Mattila, A. S. (2010). The relationship between consumer complaining behavior and service recovery: An integrative review. *International Journal of Contemporary Hospitality Management*, *22*(7), 975–991.

HaCohen-Kerner, Y., Dilmon, R., Hone, M., & Ben-Basan, M. A. (2019). Automatic classification of complaint letters according to service provider categories. *Information Processing & Management*, *56*(6), 102102.

Hair, J. F., Page, M., & Brunsveld, N. (2020). Essentials of Business Research Methods. *Routledge*.

Halstead, D. (2002). Negative word of mouth: substitute for or supplement to consumer complaints? *Journal of Consumer Satisfaction, Dissatisfaction and Complaining Behavior*, *15*, 1–12.

Hand, D. J., Mannila, H., & Smyth, P. (2001). *Retrieval by Content*.

Harris, Z. S. (1954). Distributional structure. *Word*, *10*(2–3), 146–162.

Hennig-Thurau, T., Gwinner, K. P., Walsh, G., & Gremler, D. D. (2004). Electronic word-of-mouth via consumer-opinion platforms: what motivates consumers to articulate themselves on the internet? *Journal of Interactive Marketing*, *18*(1), 38–52.

Heung, V. C. S., & Lam, T. (2003). Customer complaint behaviour towards hotel restaurant services. *International Journal of Contemporary Hospitality Management*, *15*(5), 283–289.

Hirschman, A. O. (1972). *Exit, voice, and loyalty: Responses to decline in firms, organizations, and states*. Harvard university press.

Hocutt, M. A., Bowers, M. R., & Todd Donavan, D. (2006). The art of service recovery: fact or fiction? *Journal of Services Marketing*, *20*(3), 199–207.

Hoffman, K. D., Kelley, S. W., & Rotalsky, H. M. (1995). Tracking service failures and employee recovery efforts. *Journal of Services Marketing*, *9*(2), 49–61.

Hogg, G., & Gabbott, M. (1998). *Service industries marketing: new approaches* (Vol. 18). Psychology Press.

Holloway, B. B., & Beatty, S. E. (2003). Service failure in online retailing: A recovery opportunity. *Journal of Service Research*, *6*(1), 92–105.

Homans, G. C. (1961). The humanities and the social sciences. *American Behavioral Scientist*, *4*(8), 3–6.

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.

Humphreys, A. (2021). Automated text analysis. In *Handbook of market research* (pp. 633–664). Springer.

Humphreys, A., & Wang, R. J.-H. (2018). Automated text analysis for consumer research. *Journal of Consumer Research*, *44*(6), 1274–1306.

Hung, C.-J., Chang, H. H., Eng, C. J., & Wong, K. H. (2013). Service quality and perceived value of technology-based service encounters: evaluation of clinical staff satisfaction in Taiwan. *Health Information Management Journal*, *42*(1), 29–36.

Hunt, H. K. (1991). Consumer satisfaction, dissatisfaction, and complaining behavior. *Journal of Social Issues*, *47*(1), 107–117.

Huppertz, J. W. (2014). The effort model of consumer complaining behavior: An update and new research directions. *Journal of Consumer Satisfaction, Dissatisfaction and Complaining Behavior*, *27*, 2–5.

Jaakkola, E., & Terho, H. (2021). Service journey quality: conceptualization, measurement and customer outcomes. *Journal of Service Management*, 32(6), 1-27.

Jacoby, J., & Jaccard, J. J. (1981). The sources, meaning, and validity of consumer complaint behavior: A psychological analysis. *Journal of Retailing*.

Jani, D., & Han, H. (2011). Investigating the key factors affecting behavioral intentions: Evidence from a full-service restaurant setting. *International Journal of Contemporary Hospitality Management*, *23*(7), 1000–1018.

Jean Harrison-Walker, L. (2001). E-complaining: a content analysis of an Internet complaint forum. *Journal of Services Marketing*, *15*(5), 397–412.

Jin, J., Yan, X., Yu, Y., & Li, Y. (2013). *Service failure complaints identification in social media: A text classification approach*.

Jin, M., & Aletras, N. (2020). Complaint identification in social media with transformer networks. *ArXiv Preprint ArXiv:2010.10910*.

Jindal, R., Malhotra, R., & Jain, A. (2015). Techniques for text classification: Literature review and current trends. *Webology*, *12*(2).

Jonker, J., & Pennink, B. (2010). The essence of research methodology: A concise guide for master and PhD students in management science. *Springer Science & Business Media*.

Karatepe, O. M. (2006). Customer complaints and organizational responses: the effects of complainants' perceptions of justice on satisfaction and loyalty. *International Journal of Hospitality Management*, *25*(1), 69–90.

Kiela, D., Grave, E., Joulin, A., & Mikolov, T. (2018). Efficient large-scale multi-modal classification. *Proceedings of the AAAI Conference on Artificial Intelligence*, *32*(1).

Khodadadi, A., Hosseini, S. A., Pajouheshgar, E., Mansouri, F., & Rabiee, H. R. (2019). ChOracle: A Unified Statistical Framework for Churn Prediction. *arXiv preprint arXiv:1909.06868*.

Kim, C., Kim, S., Im, S., & Shin, C. (2003). The effect of attitude and perception on consumer complaint intentions. *Journal of Consumer Marketing*, *20*(4), 352–371.

Kim, J., Amplayo, R. K., Lee, K., Sung, S., Seo, M., & Hwang, S. (2019). Categorical metadata representation for customized text classification. *Transactions of the Association for Computational Linguistics*, *7*, 201–215.

Kim, T. T., Kim, W. G., & Kim, H.-B. (2009). The effects of perceived justice on recovery satisfaction, trust, word-of-mouth, and revisit intention in upscale hotels. *Tourism Management*, *30*(1), 51–62.

Kim, W., Ok, C., & Canter, D. D. (2012). Value-driven customer share of visits. *The Service Industries Journal*, *32*(1), 37–58.

Kim, Y. (2014). Convolutional neural networks for sentence classification. *ArXiv Preprint ArXiv:1408.5882*.

Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. *ArXiv Preprint ArXiv:1412.6980*.

Kitchin, R. (2014). Big Data, new epistemologies and paradigm shifts. *Big data & society*, 1(1), 2053951714528481.

Kotsiantis, S. B., Kanellopoulos, D., & Pintelas, P. E. (2006). 'Data preprocessing for supervised learning', *International Journal of Computer Science*, 1(2), pp. 111-117.

Kowalski, R. M. (1996). Complaints and complaining: Functions, antecedents, and consequences. *Psychological Bulletin*, *119*(2), 179.

Kowsari, K., Jafari Meimandi, K., Heidarysafa, M., Mendu, S., Barnes, L., & Brown, D. (2019). Text classification algorithms: A survey. *Information*, *10*(4), 150.

Krawczyk, B. (2016). Learning from imbalanced data: open challenges and future directions. *Progress in Artificial Intelligence*, *5*(4), 221–232.

Kumar, B. S., & Ravi, V. (2016). A survey of the applications of text mining in financial domain. *Knowledge-Based Systems*, *114*, 128–147.

Kuo, Y.-F., & Wu, C.-M. (2012). Satisfaction and post-purchase intentions with service recovery of online shopping websites: Perspectives on perceived justice and emotions. *International Journal of Information Management*, *32*(2), 127–138.

LaBarbera, P. A., & Mazursky, D. (1983). A longitudinal assessment of consumer satisfaction/dissatisfaction: the dynamic aspect of the cognitive process. *Journal of Marketing Research*, *20*(4), 393–404.

Laksana, J., & Purwarianti, A. (2014). Indonesian Twitter text authority classification for government in Bandung. *2014 International Conference of Advanced Informatics: Concept, Theory and Application (ICAICTA)*, 129–134.

Lapidus, R. S., & Schibrowsky, J. A. (1994). Aggregate Complaint Analysis: A Procedure for Developing Customer ServiceSatisfaction. *Journal of Services Marketing*, *8*(4), 50–60.

LaTour, S. A., & Peat, N. C. (1980). The role of situationally-produced expectations, others' experiences, and prior experience in determining consumer satisfaction. *ACR North American Advances*.

Lau, G. T., & Ng, S. (2001). Individual and situational factors influencing negative word-of-mouth behaviour. *Canadian Journal of Administrative Sciences/Revue Canadienne Des Sciences de l'Administration*, *18*(3), 163–178.

Lazarus, R. S. (1966). *Psychological stress and the coping process.*

Lazarus, R. S., & DeLongis, A. (1983). Psychological stress and coping in aging. *American Psychologist*, *38*(3), 245.

Lazarus, R. S., & Folkman, S. (1984). *Stress, appraisal, and coping*. Springer publishing company.

Le Claire, K. A. (1993). Chinese complaints behaviour. *Journal of International Consumer Marketing*, *5*(4), 73–92.

Lee, C.-H., Wang, Y.-H., & Trappey, A. J. C. (2015). Ontology-based reasoning for the intelligent handling of customer complaints. *Computers & Industrial Engineering*, *84*, 144–155.

Lee, H. (2005). Behavioral strategies for dealing with flaming in an online forum. *The Sociological Quarterly*, *46*(2), 385–403.

Lee, S. J., & Lee, Z. (2006). An experimental study of online complaint management in the online feedback forum. *Journal of Organizational Computing and Electronic Commerce*, *16*(1), 65–85.

Lee, Y. L., & Song, S. (2010). An empirical investigation of electronic word-of-mouth: Informational motive and corporate response strategy. *Computers in Human Behavior*, *26*(5), 1073–1080.

Lemmink, J., & Mattsson, J. (2002). Employee behavior, feelings of warmth and customer perception in service encounters. *International Journal of Retail & Distribution Management*, *30*(1), 18–33.

Lemon, K. N., & Verhoef, P. C. (2016). Understanding customer experience throughout the customer journey. *Journal of Marketing* 80(November), 69-96.

Liao, H. (2007). Do it right this time: The role of employee service recovery performance in customer-perceived justice and customer loyalty after service failures. *Journal of Applied Psychology*, *92*(2), 475.

Liljander, V., & Strandvik, T. (1997). Emotions in service satisfaction. *International Journal of Service Industry Management*, *8*(2), 148–169.

Lin, H., Wang, Y., & Chang, L. (2011). Consumer responses to online retailer's service recovery after a service failure: A perspective of justice theory. *Managing Service Quality: An International Journal*, *21*(5), 511–534.

Linder, A., Schmitt, S., & Schmitt, R. (2014). Technical complaint management from a quality perspective. *Total Quality Management & Business Excellence*, *25*(7–8), 865–875.

Liu, L., & Özsu, M. T. (2009). *Encyclopedia of database systems* (Vol. 6). Springer New York, NY, USA:

Liu, R. R., & McClure, P. (2001). Recognizing cross-cultural differences in consumer complaint behavior and intentions: an empirical examination. *Journal of Consumer Marketing*, *18*(1), 54–75.

Maia, P., Carvalho, R., Ladeira, M., Rocha, H., & Mendes, G. (2014). Application of text mining techniques for classification of documents: a study of automation of complaints screening in a Brazilian federal agency. *Solid-State Electronics*, *38*, 1461–1463.

Manning, C., Raghavan, P., & Schutze, H. (2008). Term weighting, and the vector space model. *Introduction to Information Retrieval*, 109–133.

Mano, H., & Oliver, R. L. (1993). Assessing the dimensionality and structure of the consumption experience: evaluation, feeling, and satisfaction. *Journal of Consumer Research*, *20*(3), 451–466.

Manochandar, S., & Punniyamoorthy, M. (2018). Scaling feature selection method for enhancing the classification performance of Support Vector Machines in text mining. *Computers & Industrial Engineering*, *124*, 139–156.

Marcus, A. A., & Goodman, R. S. (1991). Victims and shareholders: The dilemmas of presenting corporate policy during a crisis. *Academy of Management Journal*, *34*(2), 281–305.

Marcus, M., Santorini, B., & Marcinkiewicz, M. A. (1993). *Building a large annotated corpus of English: The Penn Treebank.*

Marquis, M., & Filiatrault, P. (2002). Understanding complaining responses through consumers' self-consciousness disposition. *Psychology & Marketing*, *19*(3), 267–292.

Mattila, A. S. (2004). The impact of service failures on customer loyalty: The moderating role of affective commitment. *International Journal of Service Industry Management*, *15*(2), 134–149.

Mattila, A. S., & Enz, C. A. (2002). The role of emotions in service encounters. *Journal of Service Research*, *4*(4), 268–277.

Mattila, A. S., & Ro, H. (2008). Discrete negative emotions and customer dissatisfaction responses in a casual restaurant setting. *Journal of Hospitality & Tourism Research*, *32*(1), 89–107.

Mattila, A., & Wirtz, J. (2000). The role of preconsumption affect in postpurchase evaluation of services. *Psychology & Marketing*, *17*(7), 587–605.

Mattsson, J., & den Haring, M. J. (1998). Communication dynamics in the service encounter: A linguistic study in a hotel conference department. *International Journal of Service Industry Management*, *9*(5), 416–435.

Maute, M. F., & Dubés, L. (1999). Patterns of emotional responses and behavioural consequences of dissatisfaction. *Applied Psychology*, *48*(3), 349–366.

Maute, M. F., & Forrester Jr, W. R. (1993). The structure and determinants of consumer complaint intentions and behavior. *Journal of Economic Psychology*, *14*(2), 219–247.

Maxham III, J. G. (2001). Service recovery's influence on consumer satisfaction, positive word-of-mouth, and purchase intentions. *Journal of Business Research*, *54*(1), 11–24.

Maxham III, J. G., & Netemeyer, R. G. (2002a). A longitudinal study of complaining customers' evaluations of multiple service failures and recovery efforts. *Journal of Marketing*, *66*(4), 57–71.

Maxham III, J. G., & Netemeyer, R. G. (2002b). Modelling customer perceptions of complaint handling over time: the effects of perceived justice on satisfaction and intent. *Journal of Retailing*, *78*(4), 239–252.

McAfee, A., Brynjolfsson, E., Davenport, T. H., Patil, D. J., & Barton, D. (2012). Big data: the management revolution. *Harvard Business Review*, *90*(10), 60–68.

McAlister, D. T., & Erffmeyer, R. C. (2003). A content analysis of outcomes and responsibilities for consumer complaints to third-party organizations. *Journal of Business Research*, *56*(4), 341–351.

McColl-Kennedy, J. R., & Sparks, B. A. (2003). Application of fairness theory to service failures and service recovery. *Journal of Service Research*, *5*(3), 251–266.

McCollough, M. A. (2009). The recovery paradox: The effect of recovery performance and service failure severity on post-recovery customer satisfaction. *Academy of Marketing Studies Journal*, *13*(1), 89–104.

McCollough, M. A., Berry, L. L., & Yadav, M. S. (2000). An empirical investigation of customer satisfaction after service failure and recovery. *Journal of Service Research*, *3*(2), 121–137.

McGuinness, D. L., & Van Harmelen, F. (2004). OWL web ontology language overview. *W3C Recommendation*, *10*(10), 2004.

McKee, D., Simmers, C. S., & Licata, J. (2006). Customer self-efficacy and response to service. *Journal of Service Research*, *8*(3), 207–220.

Menon, K., & Dube, L. (2004). Service provider responses to anxious and angry customers: different challenges, different payoffs. *Journal of Retailing*, *80*(3), 229–237.

Merson, F., & Mary, R. (2017). A text mining approach to identify and analyse prominent issues from public complaints. *International Journal of Advanced Research in Computer and Communication Engineering*, *6*(3), 54–59.

Meuter, M. L., Ostrom, A. L., Bitner, M. J., & Roundtree, R. (2003). The influence of technology anxiety on consumer use and experiences with self-service technologies. *Journal of Business Research*, *56*(11), 899–906.

Meuter, M. L., Ostrom, A. L., Roundtree, R. I., & Bitner, M. J. (2000). Self-service technologies: understanding customer satisfaction with technology-based service encounters. *Journal of Marketing*, *64*(3), 50–64.

Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient estimation of word representations in vector space. *ArXiv Preprint ArXiv:1301.3781*.

Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., & Dean, J. (2013). Distributed representations of words and phrases and their compositionality. *Advances in Neural Information Processing Systems*, *26*.

Miller, J. L., Craighead, C. W., & Karwan, K. R. (2000). Service recovery: a framework and empirical investigation. *Journal of Operations Management*, *18*(4), 387–400.

Minaee, S., Kalchbrenner, N., Cambria, E., Nikzad, N., Chenaghlu, M., & Gao, J. (2021). Deep learning--based text classification: a comprehensive review. *ACM Computing Surveys (CSUR)*, *54*(3), 1–40.

Mirończuk, M. M., & Protasiewicz, J. (2018). A recent overview of the state-of-the-art elements of text classification. *Expert Systems with Applications*, *106*, 36–54.

Morgan, D. L. (2014). Pragmatism as a Paradigm for Social Research. *Sage Publications*.

Ngai, E. W. T., Heung, V. C. S., Wong, Y. H., & Chan, F. K. Y. (2007). Consumer complaint behaviour of Asians and non-Asians about hotel services: An empirical analysis. *European Journal of Marketing*, *41*(11/12), 1375–1391.

Nikbin, D., Ismail, I., Marimuthu, M., & Salarzehi, H. (2012). The relationship of service failure attributions, service recovery justice and recovery satisfaction in the context of airlines. *Scandinavian Journal of Hospitality and Tourism*, *12*(3), 232–254.

Nikbin, D., Marimuthu, M., Hyun, S. S., & Ismail, I. (2015). Relationships of perceived justice to service recovery, service failure attributions, recovery satisfaction, and loyalty in the context of airline travelers. *Asia Pacific Journal of Tourism Research*, *20*(3), 239–262.

Noone, B. M. (2012). Overcompensating for severe service failure: Perceived fairness and effect on negative word-of-mouth intent. *Journal of Services Marketing*, *26*(5), 342–351.

Ok, C., Back, K.-J., & Shanklin, C. W. (2005). Modelling roles of service recovery strategy: a relationship-focused view. *Journal of Hospitality & Tourism Research*, *29*(4), 484–507.

Oliver, R. L. (1980). A cognitive model of the antecedents and consequences of satisfaction decisions. *Journal of Marketing Research*, *17*(4), 460–469.

Oliver, R. L. (1981). Measurement and evaluation of satisfaction processes in retail settings. *Journal of Retailing*.

Oliver, R. L. (1999). Whence consumer loyalty? *Journal of Marketing*, *63*(4_suppl1), 33–44.

Oliver, R. L., & DeSarbo, W. S. (1988). Response determinants in satisfaction judgments. *Journal of Consumer Research*, *14*(4), 495–507.

Oliver, R. L., & Swan, J. E. (1989). Consumer perceptions of interpersonal equity and satisfaction in transactions: a field survey approach. *Journal of Marketing*, *53*(2), 21–35.

Panksepp, J. (2004). *Affective neuroscience: The foundations of human and animal emotions*. Oxford university press.

Parker, C., & Mathews, B. P. (2001). Customer satisfaction: contrasting academic and consumers' interpretations. *Marketing Intelligence & Planning*, *19*(1), 38–44.

Pedersen, T., Pakhomov, S. V. S., Patwardhan, S., & Chute, C. G. (2007). Measures of semantic similarity and relatedness in the biomedical domain. *Journal of Biomedical Informatics*, *40*(3), 288–299.

Pennington, J., Socher, R., & Manning, C. D. (2014). Glove: Global vectors for word representation. *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 1532–1543.

Pfeffer, J., Zorbach, T., & Carley, K. M. (2014). Understanding online firestorms: Negative word-of-mouth dynamics in social media networks. *Journal of Marketing Communications*, *20*(1–2), 117–128.

Phillips, D. M., & Baumgartner, H. (2002). The role of consumption emotions in the satisfaction response. *Journal of Consumer Psychology*, *12*(3), 243–252.

Pieters, R., Baumgartner, H., & Allen, D. (1995). A means-end chain approach to consumer goal structures. *International Journal of Research in Marketing*, *12*(3), 227–244.

Pinto, M. B., & Mansfield, P. (2012). Facebook as a complaint mechanism: An investigation of millennials. *Journal of Behavioral Studies in Business*, *5*, 1.

Poria, S., Cambria, E., Howard, N., Huang, G.-B., & Hussain, A. (2016). Fusing audio, visual and textual clues for sentiment analysis from multimodal content. *Neurocomputing*, *174*, 50–59.

Prakash, A., & Madabushi, H. T. (2020). Incorporating count-based features into pre-trained models for improved stance detection. *ArXiv Preprint ArXiv:2010.09078*.

Preotiuc-Pietro, D., Gaman, M., & Aletras, N. (2019). Automatically identifying complaints in social media. *ArXiv Preprint ArXiv:1906.03890*.

Prim, I., & Pras, B. (1999). "Friendly" complaining behaviors: Toward a relational approach. *Journal of Market-Focused Management*, *3*, 333–352.

Qu, L., Liu, S., Wang, M., Li, S., Yin, S., Qiao, Q., & Song, Z. (2022). TransFuse: A unified transformer-based image fusion framework using self-supervised learning. *ArXiv Preprint ArXiv:2201.07451*.

Raguseo, E. (2018). Big data technologies: An empirical investigation on their adoption, benefits and risks for companies. *International Journal of Information Management*, *38*(1), 187–195.

Rahman, W., Hasan, M. K., Lee, S., Zadeh, A., Mao, C., Morency, L.-P., & Hoque, E. (2020). Integrating multimodal information in large pretrained transformers. *Proceedings of the Conference. Association for Computational Linguistics. Meeting*, *2020*, 2359.

Ramachandram, D., Lisicki, M., Shields, T. J., Amer, M. R., & Taylor, G. W. (2017). Structure optimization for deep multimodal fusion networks using graph-induced kernels. *ArXiv Preprint ArXiv:1707.00750*.

Rawls, J. (1971). Atheory of justice. *Cambridge (Mass.)*.

Reynolds, K. L., & Harris, L. C. (2006). Deviant customer behavior: An exploration of frontline employee tactics. *Journal of Marketing Theory and Practice*, *14*(2), 95–111.

Richins, M. L. (1983). Negative word-of-mouth by dissatisfied consumers: A pilot study. *Journal of Marketing*, *47*(1), 68–78.

Richins, M. L. (1987). A multivariate analysis of responses to dissatisfaction. *Journal of the Academy of Marketing Science*, *15*(3), 24–31.

Richins, M. L., & Verhage, B. J. (1985). Seeking redress for consumer dissatisfaction: The role of attitudes and situational factors. *Journal of Consumer Policy*, *8*(1), 29–44.

Risselada, H., Verhoef, P. C., & Bijmolt, T. H. A. (2010). Staying power of churn prediction models. *Journal of Interactive Marketing*, *24*(3), 198–208.

Rodriguez, M. A., & Egenhofer, M. J. (2003). Determining semantic similarity among entity classes from different ontologies. *IEEE Transactions on Knowledge and Data Engineering*, *15*(2), 442–456.

Rowley, J., & Dawes, J. (2000). Disloyalty: a closer look at non-loyals. *Journal of Consumer Marketing*, *17*(6), 538–547.

Rust, R. T., Moorman, C., & Dickson, P. R. (2002). Getting return on quality: revenue expansion, cost reduction, or both? *Journal of Marketing*, *66*(4), 7–24.

Ryan, G. (2018). 'Introduction to positivism, interpretivism and critical theory', *Nurse Researcher,* 25(4), pp. 41-49.

S. Mattila, A., Andreau, L., Hanks, L., & E. Kim, E. (2013). The impact of cyberostracism on online complaint handling: Is "automatic reply" any better than "no reply"? *International Journal of Retail & Distribution Management*, *41*(1), 45–60.

Salton, G., & Buckley, C. (1988). Term-weighting approaches in automatic text retrieval. *Information Processing & Management*, *24*(5), 513–523.

Salton, G., Yang, C.-S., & Yu, C. T. (1975). A theory of term importance in automatic text analysis. *Journal of the American Society for Information Science*, *26*(1), 33–44.

Saunders, M., Lewis, P., & Thornhill, A. (2019). Research Methods for Business Students. *Pearson Education*.

Schoefer, K., & Diamantopoulos, A. (2008). The role of emotions in translating perceptions of (in) justice into postcomplaint behavioral responses. *Journal of Service Research*, *11*(1), 91–103.

Schoefer, K., & Ennew, C. (2005). The impact of perceived justice on consumers' emotional responses to service complaint experiences. *Journal of Services Marketing*, *19*(5), 261–270.

Schütze, H., Manning, C. D., & Raghavan, P. (2008). Introduction to information retrieval (Vol. 39, pp. 234-265). *Cambridge: Cambridge University Press*.

Schwab, P.-N. (2015). *Online complaint handling practices: Company strategies and their effects upon post-complaint satisfaction*. Univ. Libre de Bruxelles, Solvay Brussels School of Economics and Management ….

Schweikhart, S. B., Strasser, S., & Kennedy, M. R. (1993). Service recovery in health services organizations. *Journal of Healthcare Management*, *38*(1), 3–21.

Shahmirzadi, O., Lugowski, A., & Younge, K. (2019). Text similarity in vector space models: a comparative study. *2019 18th IEEE International Conference on Machine Learning and Applications (ICMLA)*, 659–666.

Shannon-Baker, P. (2016). 'Making paradigms meaningful in mixed methods research', *Journal of Mixed Methods Research*, 10(4), pp. 319-334.

Shapiro, T., & Nieman-Gonder, J. (2006). Effect of communication mode in justice-based service recovery. *Managing Service Quality: An International Journal*, *16*(2), 124–144.

Shirazi, F., & Mohammadi, M. (2019). A big data analytics model for customer churn prediction in the retiree segment. *International Journal of Information Management*, *48*, 238–253.

Shmueli, G. (2010). 'To explain or to predict?', *Statistical Science*, 25(3), pp. 289-310.

Shostack, L. (1985). Planning the service encounter. *The Service Encounter*, 243–254.

Singh, J. (1988). Consumer complaint intentions and behavior: definitional and taxonomical issues. *Journal of Marketing*, *52*(1), 93–107.

Singh, J. (1989). Determinants of consumers' decisions to seek third party redress: An empirical study of dissatisfied patients. *Journal of Consumer Affairs*, *23*(2), 329–363.

Singh, J. (1990). Voice, exit, and negative word-of-mouth behaviors: An investigation across three service categories. *Journal of the Academy of Marketing Science*, *18*, 1–15.

Singh, J., & Widing, R. E. (1991). What occurs once consumers complain? A theoretical model for understanding satisfaction/dissatisfaction outcomes of complaint responses. *European Journal of Marketing*, *25*(5), 30–46.

Siu, N. Y.-M., Zhang, T. J.-F., & Yau, C.-Y. J. (2013). The roles of justice and customer satisfaction in customer retention: A lesson from service recovery. *Journal of Business Ethics*, *114*, 675–686.

Smith, A. K., & Bolton, R. N. (2002). The effect of customers' emotional responses to service failures on their recovery effort evaluations and satisfaction judgments. *Journal of the Academy of Marketing Science*, *30*, 5–23.

Smith, A. K., Bolton, R. N., & Wagner, J. (1999). A model of customer satisfaction with service encounters involving failure and recovery. *Journal of Marketing Research*, *36*(3), 356–372.

Sokolova, M., & Lapalme, G. (2009). A systematic analysis of performance measures for classification tasks. *Information Processing & Management*, *45*(4), 427–437.

Solomon, M. R., Surprenant, C., Czepiel, J. A., & Gutman, E. G. (1985). A role theory perspective on dyadic interactions: the service encounter. *Journal of Marketing*, *49*(1), 99–111.

Somasundaran, S., & Wiebe, J. (2010). Recognizing stances in ideological on-line debates. *Proceedings of the NAACL HLT 2010 Workshop on Computational Approaches to Analysis and Generation of Emotion in Text*, 116–124.

Spanoudes, P., & Nguyen, T. (2017). Deep Learning in Customer Churn Prediction: Unsupervised Feature Learning on Abstract Company Independent Feature Vectors. *arXiv preprint arXiv:1703.03869*.

Sparck Jones, K. (1972). A statistical interpretation of term specificity and its application in retrieval. *Journal of Documentation*, *28*(1), 11–21.

Sparks, B., & Fredline, L. (2007). Providing an explanation for service failure: Context, content, and customer responses. *Journal of Hospitality & Tourism Research*, 31(2), 241-260.

Spreng, R. A., MacKenzie, S. B., & Olshavsky, R. W. (1996). A reexamination of the determinants of consumer satisfaction. *Journal of Marketing*, *60*(3), 15–32.

Stephens, N., & Gwinner, K. P. (1998). Why don't some people complain? A cognitive-emotive process model of consumer complaint behavior. *Journal of the Academy of Marketing Science*, *26*(3), 172–189.

Stoica, E. A., & Özyirmidokuz, E. K. (2015). Mining customer feedback documents. *International Journal of Knowledge Engineering*, *1*(1), 68–71.

Sultan, A. B., Abidin, K. Z., & Abdullah, M. T. (2008). The implementation of agent-based complaint management system. *Journal of Computer Science*, *8*(5), 205–207.

Sun, C., Qiu, X., Xu, Y., & Huang, X. (2019, October). How to fine-tune bert for text classification?. *In China national conference on Chinese computational linguistics* (pp. 194-206). Cham: Springer International Publishing.

Sundaram, D. S., Mitra, K., & Webster, C. (1998). Word-of-mouth communications: A motivational analysis. *ACR North American Advances*.

Surprenant, C. F., & Solomon, M. R. (1987). Predictability and personalization in the service encounter. *Journal of Marketing*, *51*(2), 86–96.

Susskind, A. M. (2005). A content analysis of consumer complaints, remedies, and repatronage intentions regarding dissatisfying service experiences. *Journal of Hospitality & Tourism Research*, *29*(2), 150–169.

Swan, J. E., & Oliver, R. L. (1991). An applied analysis of buyer equity perceptions and satisfaction with automobile salespeople. *Journal of Personal Selling & Sales Management*, *11*(2), 15–26.

Swanson, S. R., & Hsu, M. K. (2011). The effect of recovery locus attributions and service failure severity on word-of-mouth and repurchase behaviors in the hospitality industry. *Journal of Hospitality & Tourism Research*, *35*(4), 511–529.

Szymanski, D. M., & Henard, D. H. (2001). Customer satisfaction: A meta-analysis of the empirical evidence. *Journal of the Academy of Marketing Science*, *29*(1), 16–35.

Tang, L., Thomas, L., Fletcher, M., Pan, J., & Marshall, A. (2014). Assessing the impact of derived behavior information on customer attrition in the financial service industry. *European Journal of Operational Research*, *236*(2), 624–633.

Tashakkori, A., & Teddlie, C. (2010). Mixed methodology: Combining qualitative and quantitative approaches. *Sage*.

Tax, S. S., Brown, S. W., & Chandrashekaran, M. (1998). Customer evaluations of service complaint experiences: implications for relationship marketing. *Journal of Marketing*, *62*(2), 60–76.

Thoits, P. A. (1986). Social support as coping assistance. *Journal of Consulting and Clinical Psychology*, *54*(4), 416.

Thomas, J. B., Peters, C. O., Howell, E. G., & Robbins, K. (2012). Social media and negative word of mouth: strategies for handing unexpecting comments. *Atlantic Marketing Journal*, *1*(2), 7.

Thomas, N. T. (2018). A LSTM based Tool for Consumer Complaint Classification. *2018 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, 2349–2351.

Tjandra, S., Warsito, A. A. P., & Sugiono, J. P. (2015). Determining citizen complaints to the appropriate government departments using KNN algorithm. *2015 13th International Conference on ICT and Knowledge Engineering (ICT & Knowledge Engineering 2015)*, 1–4.

Trappey, A. J. C., Lee, C.-H., Chen, W.-P., & Trappey, C. V. (2010). A framework of customer complaint handling system. *2010 7th International Conference on Service Systems and Service Management*, 1–6.

Tronvoll, B. (2007). Customer complaint behaviour from the perspective of the service-dominant logic of marketing. *Managing Service Quality*, *17*(6), 601–620.

Tsai, Y.-H. H., Bai, S., Liang, P. P., Kolter, J. Z., Morency, L.-P., & Salakhutdinov, R. (2019). Multimodal transformer for unaligned multimodal language sequences. *Proceedings of the Conference. Association for Computational Linguistics. Meeting*, *2019*, 6558.

Uschold, M., & Gruninger, M. (1996). Ontologies: Principles, methods and applications. *The Knowledge Engineering Review*, *11*(2), 93–136.

Van Noort, G., & Willemsen, L. M. (2012). Online damage control: The effects of proactive versus reactive webcare interventions in consumer-generated and brand-generated platforms. *Journal of Interactive Marketing*, *26*(3), 131–140.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. *Advances in Neural Information Processing Systems*, *30*.

Verbeke, W., Dejaeger, K., Martens, D., Hur, J., & Baesens, B. (2012). New insights into churn prediction in the telecommunication sector: A profit driven data mining approach. *European Journal of Operational Research*, *218*(1), 211–229.

Vitell, S. J., & Muncy, J. (2005). The Muncy–Vitell consumer ethics scale: A modification and application. *Journal of Business Ethics*, *62*, 267–275.

Voorhees, C. M., Brady, M. K., & Horowitz, D. M. (2006). A voice from the silent masses: an exploratory and comparative analysis of noncomplainers. *Journal of the Academy of Marketing Science*, *34*(4), 514–527.

Vos, J. F. J., Huitema, G. B., & de Lange-Ros, E. (2008). How organisations can learn from complaints. *The TQM Journal*, *20*(1), 8–17.

Walker, J. L. (1995). Service encounter satisfaction: conceptualized. *Journal of Services Marketing*, *9*(1), 5–14.

Wallin Andreassen, T. (2000). Antecedents to satisfaction with service recovery. *European Journal of Marketing*, *34*(1/2), 156–175.

Walster, E., Berscheid, E., & Walster, G. W. (1973). New directions in equity research. *Journal of Personality and Social Psychology*, *25*(2), 151.

Wang, J. Z., Du, Z., Payattakool, R., Yu, P. S., & Chen, C.-F. (2007). A new method to measure the semantic similarity of GO terms. *Bioinformatics*, *23*(10), 1274–1281.

Watkins, H. S., & Liu, R. (1996). Collectivism, individualism and in-group membership: implications for consumer complaining behaviors in multicultural contexts. *Journal of International Consumer Marketing*, *8*(3–4), 69–96.

Weiner, B. (1985). An attributional theory of achievement motivation and emotion. *Psychological Review*, *92*(4), 548.

Weiss, H. M., Suckow, K., & Cropanzano, R. (1999). Effects of justice conditions on discrete emotions. *Journal of Applied Psychology*, *84*(5), 786.

Wen, B., & Geng-qing Chi, C. (2013). Examine the cognitive and affective antecedents to service recovery satisfaction: A field study of delayed airline passengers. *International Journal of Contemporary Hospitality Management*, *25*(3), 306–327.

Westbrook, R. A. (1980). Intrapersonal affective influences on consumer satisfaction with products. *Journal of Consumer Research*, *7*(1), 49–54.

Westbrook, R. A. (1981). Sources of consumer satisfaction with retail outlets. *Journal of Retailing*.

Westbrook, R. A. (1987). Product/consumption-based affective responses and postpurchase processes. *Journal of Marketing Research*, *24*(3), 258–270.

Westbrook, R. A., & Oliver, R. L. (1991). The dimensionality of consumption emotion patterns and consumer satisfaction. *Journal of Consumer Research*, *18*(1), 84–91.

What's the average churn rate by industry?, CustomerGauge. (2024a) Available at: *https://customergauge.com/blog/average-churn-rate-by-industry/*

White, L., & Yanamandram, V. (2004). Why customers stay: reasons and consequences of inertia in financial services. *Managing Service Quality: An International Journal*, *14*(2/3), 183–194.

Wirtz, J., & Mattila, A. S. (2004). Consumer responses to compensation, speed of recovery and apology after a service failure. *International Journal of Service Industry Management*, *15*(2), 150–166.

Wolf, T., Debut, L., Sanh, V., Chaumond, J., Delangue, C., Moi, A., Cistac, P., Rault, T., Louf, R., & Funtowicz, M. (2019). Huggingface's transformers: State-of-the-art natural language processing. *ArXiv Preprint ArXiv:1910.03771*.

Wolfram, D., & Zhang, J. (2008). The influence of indexing practices and weighting algorithms on document spaces. *Journal of the American Society for Information Science and Technology*, *59*(1), 3–11.

Woodruff, R. B., Cadotte, E. R., & Jenkins, R. L. (1983). Modelling consumer satisfaction processes using experience-based norms. *Journal of Marketing Research*, *20*(3), 296–304.

Wu, H. (2022). A High-Performance Customer Churn Prediction System based on Self-Attention. *arXiv preprint arXiv:2206.01523*.

Xiang, Z., Schwartz, Z., Gerdes Jr, J. H., & Uysal, M. (2015). What can big data and text analytics tell us about hotel guest experience and satisfaction? *International Journal of Hospitality Management*, *44*, 120–130.

Xiong, Y., Chen, S., Qin, H., Cao, H., Shen, Y., Wang, X., Chen, Q., Yan, J., & Tang, B. (2020). Distributed representation and one-hot representation fusion with gated network for clinical semantic textual similarity. *BMC Medical Informatics and Decision Making*, *20*, 1–7.

Yan, Y., Zhang, J., & Yan, M. (2006). Ontology modelling for contract: Using OWL to express semantic relations. *2006 10th IEEE International Enterprise Distributed Object Computing Conference (EDOC'06)*, 409–412.

Yang, C., Shi, X., Luo, J., & Han, J. (2019). I Know You'll Be Back: Interpretable New User Clustering and Churn Prediction on a Mobile Social Application. arXiv preprint *arXiv:1910.01447*.

Yang, M., Li, Y., & Kiang, M. (2011). *Uncovering social media data for public health surveillance*.

Yang, W., Tan, L., Lu, C., Cui, A., Li, H., Chen, X., Xiong, K., Wang, M., Li, M., & Pei, J. (2019). Detecting customer complaint escalation with recurrent neural networks and manually-engineered features. *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 2 (Industry Papers)*, 56–63.

Yang, Y., Xu, D.-L., Yang, J.-B., & Chen, Y.-W. (2018). An evidential reasoning-based decision support system for handling customer complaints in mobile telecommunications. *Knowledge-Based Systems*, *162*, 202–210.

Yi, S., & Baumgartner, H. (2004). Coping with negative emotions in purchase-related situations. *Journal of Consumer Psychology*, *14*(3), 303–317.

Young, T., Hazarika, D., Poria, S., & Cambria, E. (2018). Recent trends in deep learning based natural language processing. *Ieee Computational IntelligenCe Magazine*, *13*(3), 55–75.

Yuksel, A., Kilinc, U., & Yuksel, F. (2006). Cross-national analysis of hotel customers' attitudes toward complaining and their complaining behaviours. *Tourism Management*, *27*(1), 11–24.

Zaby, C., & Wilde, K. D. (2018). Intelligent business processes in CRM: exemplified by complaint management. *Business & Information Systems Engineering*, *60*, 289–304.

Zadeh, A., Chen, M., Poria, S., Cambria, E., & Morency, L.-P. (2017). Tensor fusion network for multimodal sentiment analysis. *ArXiv Preprint ArXiv:1707.07250*.

Zeelenberg, M., & Pieters, R. (2004). Beyond valence in customer dissatisfaction: A review and new findings on behavioral responses to regret and disappointment in failed services. *Journal of Business Research*, *57*(4), 445–455.

Zeithaml, V. A., Bitner, M. J., & Gremler, D. D. (2018). *Services marketing: Integrating customer focus across the firm*. McGraw-Hill.

Zhang, X., Hou, X., Chen, X., & Zhuang, T. (2013). Ontology-based semantic retrieval for engineering domain knowledge. *Neurocomputing*, *116*, 382–391.

Zhang, X., Zhao, J., & LeCun, Y. (2015). Character-level convolutional networks for text classification. *Advances in neural information processing systems*, 28.

Zimbra, D., Fu, T., & Li, X. (2009). *Assessing public opinions through Web 2.0: a case study on Wal-Mart*.

Zirtiloğlu, H., & Yolum, P. (2008). Ranking semantic information for e-government: complaints management. *Proceedings of the First International Workshop on Ontology-Supported Business Intelligence*, 1–7.