Faculty of Business Economics

Master of Management

Master's thesis

Development and Application of a Hybrid Deep Learning Model for Aspect-Based Sentiment Analysis (ABSA): A Comparative Study of McDonalds Tweets Across Two Time Periods

Jonas Golbazi

Thesis presented in fulfillment of the requirements for the degree of Master of Management, specialization Data Science

SUPERVISOR :

dr. Frank VANHOENSHOVEN

2024
2025

www.uhasselt.be
Universiteit Hasselt
Campus Hasselt:
Martelarenlaan 42 | 3500 Hasselt
Campus Diepenbeek:
Agoralaan Gebouw D | 3590 Diepenbeek

Faculty of Business Economics

Master of Management

Master's thesis

Development and Application of a Hybrid Deep Learning Model for Aspect-Based Sentiment Analysis (ABSA): A Comparative Study of McDonalds Tweets Across Two Time Periods

Jonas Golbazi

Thesis presented in fulfillment of the requirements for the degree of Master of Management, specialization Data Science

SUPERVISOR :

dr. Frank VANHOENSHOVEN

Abstract

While extensive research exists on hybrid deep learning models, a few studies have explored their real-world application in fine-grained sentiment analysis. This thesis presents a BERT-based hybrid model—BERT+BiLSTM+Transformers (BBT)—augmented with SMOTE to perform Aspect Category Sentiment Analysis (ACSA), a subtask of Aspect-Based Sentiment Analysis (ABSA), on large-scale, real-world data. The model was designed to classify sentiment polarity, aspect categories, and broad thematic categories. To evaluate its performance, F1-scores were compared across three model variations: BERT base model, BERT+BiLSTM layer, and the proposed BBT model. The BBT model, supported by SMOTE to address class imbalance, outperformed the others with F1-scores of 0.78 for aspect category classification, 0.85 for broad thematic category classification, and 0.86 for Sentiment classification.

To address the challenge of deep learning model interpretability, three post-hoc explainability techniques—LIME, SHAP, and Attention Visualization—were employed. By effectively identifying the most influential tokens in model's predictions, these methods improved transparency and facilitated bias detection.

Then, the trained model was applied to 200,000 McDonald's-related tweets collected in two timeframes to examine changes in public sentiment across predefined broad thematic categories and aspect categories. The analysis of the results indicated a shift from negative to more neutral sentiment over time, with 'Corporate and Social Responsibility' remaining an area attracting ongoing negative feedback. Concurrently, negative feedback has become more concentrated on tangible aspects of the restaurant experience, including 'Food', 'Products', and 'Customer Service'. Categories receiving the most negative feedback were further validated by BERTopic, which was applied to the subset of negatively classified tweets.

Key words: Aspect-Based Sentiment Analysis (ABSA), Aspect Category Sentiment Analysis (ACSA), Deep Learning, BERT (Bidirectional Encoder Representations from Transformers), SMOTE (Synthetic Minority Oversampling Technique), Explainable Artificial Intelligence (XAI), Natural Language Processing (NLP)

Contents

List of Abbreviations

ABSA	Aspect-Based Sentiment Analysis
ACSA	Aspect Category Sentiment Analysis
AI	Artificial Intelligence
ANNs	Artificial Neural Networks
AOPE	Aspect-Opinion Pair Extraction
BBT	BERT-BiLSTM-Transformers
BERT	Bidirectional Encoder Representations from Transformers
BERTopic	Bidirectional Encoder Representations from Transformers for Topic Modeling
BiLSTM	Bidirectional Long Short-Term Memory
BPTT	Backpropagation Through Time
CNNs	Convolutional Neural Networks
DT	Decision Tree
GPT	Generative Pre-trained Transformer
Grad-CAM	Gradient-weighted Class Activation Mapping
GRU	Gated Recurrent Unit
ICL	In-Context Learning
LIME	Local Interpretable Model-agnostic Explanations
LLMs	Large Language Models
LoRA	Low-Rank Adaptation
LSTMs	Long Short-Term Memory networks
NB	Naïve Bayes
NLP	Natural Language Processing
OTE	Opinion Term Extraction
RNNs	Recurrent Neural Networks
SHAP	SHapley Additive exPlanations
SMOTE	Synthetic Minority Oversampling Technique
SVM	Support Vector Machines
XAI	Explainable Artificial Intelligence

List of Figures

List of Tables

1. Introduction

In recent years, social media platforms such as Twitter, Facebook, and Instagram have fundamentally changed how humans interact and participate online. They have created new patterns of communication (Tarigan et al., 2023). Social media has been intricately woven into the fabric of our daily lives, and traditional interactions, previously constrained by temporal and spatial limitations, have changed dramatically due to the advancement of information technology (Maitri et al., 2023). These platforms have become primary channels for accessing information, entertainment, and building relations (Avalle et al., 2024). They have significantly transformed how people live (Boh et al., 2023).

Once, traditional face-to-face interaction was considered critical in promoting collective understanding (Ransom et al., 2022). However, social media has fundamentally altered these traditional modes of interaction (Azzaakiyyah, 2023) and has drawn significant attention to the multifaceted impact of social media on contemporary society (Hall, 2018; Ausat et al., 2023; Azzaakiyyah, 2023) and social interactions.

The massive amounts of real-time data millions of people generate on social media platforms, exhibiting second-by-second variation (Steinert-Threlkeld, 2018), pose significant challenges and opportunities simultaneously. This data offers businesses and brands a unique approach to engaging with their audience (Seo & Park, 2018) more effectively. The large volume of customer feedback has become a valuable resource for assessing customers' satisfaction with products or services. It has contributed to the extraction of insights into customer sentiment in ways that were not possible before (Ananth kumar et al., 2024). In pursuit of more profound insights into the evolving dynamics of society and markets, businesses increasingly leverage the unstructured data derived from social media platforms.

Customer retention is one of the most critical objectives in customer relationship management (CRM) because of its contribution to enhanced business performance and profitability (Chen & Popovich, 2003). To maintain their competitive position, businesses should adopt a continuous improvement strategy for customer experience by detecting and addressing customer pain points while bridging the "gaps between expectations and experience" of the customers (Meyer & Schwager, 2007, p. 3). Therefore, a well-designed process should be developed to identify and solve customer problems and review them regularly to make sure that solutions still work. This approach also helps to detect new customer concerns early.

To help businesses implement this iterative problem-solving process, much research has been implemented to develop machine learning, particularly deep learning models that identify customers' sentiments towards services and products, which can be used to prevent customer churn as one of the significant challenges to businesses across different sectors. Businesses can use these models to identify and address early customer pain points and negative trends (Hu et al., 2018). Understanding the causes of these declines is important for organizations to detect consumer issues and act early to prevent customer churn. Consequently, sentiment analysis models have emerged in response to the need of businesses by offering a more efficient way to detect specific areas of concern compared to traditional methods like direct contact or passive monitoring (Hu et al., 2018). Therefore, examining customers' sentiments during distinct periods can assist businesses in investigating the impact of their interventions on previously detected customer pain points and identifying newly emerging issues. The primary objective of this study is to conduct a comparative analysis of customer feedback about McDonald's across two consecutive years to analyze how the fast-food giant's interventions associated with identified areas of concern have influenced customer sentiment and to identify any newly emerging pain points.

The selection of textual data sources is crucial in sentiment analysis projects. The wealth of user-generated content makes social media platforms ideal for sentiment analysis. Twitter (rebranded as X) is among the most popular social media platforms worldwide. Over 300 million active users generate over 360 thousand tweets per minute on Twitter (Domo, n.d.). Because of its broad user base, this social media platform reflects a diverse sample of the population, removing the need for large research teams to analyze various segments simultaneously (Steinert-Threkeld, 2018). Twitter has distinguishing features that make it an invaluable resource for examining and understanding customer sentiment and perception. These features include global rich, an extensive user base, and diversity of users from different socioeconomic and cultural backgrounds (Smith & Brener, 2012; Weller et al., 2014). Given the mentioned features, Twitter is selected as the primary source of data for this study due to its remarkable capacity to deliver real-time, large-scale, and diverse public opinion data.

The prevalence of unstructured data, which constitutes most of the information shared on social media networks (Gandomi & Haider, 2015), such as Twitter, complicates the analysis process. On the other hand, motivated by the opportunity for deeper insights into the changing dynamics of society and markets, businesses and researchers aim to leverage the massive amounts of unstructured data users generate on social media platforms (Cano-Marin et al., 2023). They need to leverage specialized techniques to achieve this because manually extracting meaningful insight from such vast amounts of data is complex, time-consuming, and often impractical. As a result of this need for efficient analysis of massive datasets, automated approaches to big data analytics have become increasingly essential for addressing these limitations (Mohamed & Al-Jaroodi, 2014; Cao et al., 2015; Ravi & Kamaruddin, 2017). This shift is supported by rapid growth in AI-based technologies, computational power, and cloud computing, which can be used to analyze large-scale datasets effectively (Moreno & Redondo, 2016; Duan et al., 2019).

Artificial Intelligence (AI) has become a solution, offering specialized techniques for analyzing massive amounts of unstructured data, among many other applications. AI systems are revolutionizing numerous industries and reshaping how humans experience the world (Damirchi & Amini, 2023). In the context of this study, AI can be seen as a technology that utilizes deep learning algorithms to generate human-like output (e.g., text) (Arora et al., 2024). AI is a broad field encompassing numerous subdisciplines, such as Machine Learning, Deep Learning, and Natural Language Processing (NLP) (Ertel, 2024; Ahmad et al., 2023a). Among the various subdisciplines of AI, NLP has proven particularly effective for analyzing textual data. It enables researchers and businesses to process and extract meaningful insights from the large quantities of unstructured data generated by social media platforms such as Twitter. NLP techniques such as sentiment analysis, which utilize AI and machine learning capabilities, can effectively extract valuable insights from diverse and dynamic unstructured data collected from platforms such as Twitter (Aguilar-Mareno, 2024).

Sentiment analysis is used to determine the general sentiments conveyed in text-based data (Yadollahi et al., 2017). Previously considered a subject of academic study, sentiment analysis has become increasingly an essential tool for businesses in recent years. Organizations use it to better understand customer expectations and demands in real-time, supporting more informed decisions. This also provides businesses with the opportunity to predict market shifts and optimize their products and services (Kim et al., 2022; Moudhich & Fennan, 2024; Sathyan et al., 2021). However, the implementation of sentiment analysis presents some challenges. One of the significant challenges is class separability, meaning overlapping sentiments within a single short-form text can mislead the model, reducing its accuracy in identifying the distinct sentiments associated with individual aspects. In real-world interactions, users do not always express overall sentiment; They frequently express their sentiments toward specific aspects. For instance, a user might tweet about McDonald's and express positive sentiments towards the fries but negative sentiments towards the drink price. To address this challenge, Aspect-Based Sentiment Analysis (ABSA) is developed. It overcomes the limitations of traditional sentiment analysis, which only classifies the overall sentiment (positive, negative, or neutral) of textual data like tweets (Perikos & Diamantopoulos, 2024). As a branch of sentiment analysis, ABSA takes a more granular approach to identify sentiments associated with specific aspects or categories within given textual data (Liu et al., 2022). For instance, in a tweet about McDonald's, ABSA can identify sentiments expressed towards categories such as 'Core Restaurant Experience', 'Corporate and Social Responsibility', as well as different aspects such as 'Food', 'Customer Service', and 'Ethical Responsibility' of the fast-food giant. This fine-grained approach contributes to a more detailed understanding of the sentiment of customers towards products or services in textual data, which has proven important in various fields, including social media monitoring, feedback analysis, product reviews, and market research (Rodríguez-Ibánez et al., 2023). To illustrate the level of granularity offered by ABSA, consider the following example. In the context of our study, customers may have a positive perception of McDonald's brand and exhibit brand loyalty. Nevertheless, they may express dissatisfaction with aspects such as customer service or removing certain food items from the menu. By understanding such opinions, businesses can make data-driven decisions and implement improvements guided by meaningful insights (Perikos & Diamantopoulos, 2024). Therefore, ABSA has diverse applications in various domains, such as the competitive fast-food restaurant industry, where businesses can identify customer needs and areas of improvement to enhance overall customer satisfaction. In this data-driven world, with unprecedented and increasing amounts of data on the internet, ABSA enables businesses to analyze large amounts of textual data and gain a competitive advantage (Saadati et l., 2024). Therefore, ABSA will be employed throughout this study to capture both overall sentiments expressed in tweets and, more importantly, the fine-grained sentiment towards specific aspects. Specifically, Aspect Category Sentiment Analysis (ACSA) will be utilized throughout this study. ACSA is a key component of ABSA, which identifies multiple aspect categories within a sentence and determines their associated sentiment (Zhou & Law, 2022).

ABSA has made significant Advancements. However, it still encounters important challenges. One impotent issue is the imbalanced distribution of aspects, where some categories appear more frequently than the others. As a result, the model may struggle to classify rare aspects and may not generalize effectively to real-life data. The concern of sparse data for specific aspects further adds to the problem of the lack of sufficient labeled examples, which are essential for training supervised ABSA models. The sample size is another problem that particularly impacts minority aspects. Another critical consideration in ABSA that requires careful attention is the presence of within-class sub-clusters. A seemingly unified aspect, such as 'Brand Perception', can encompass multiple sub-dimensions (e.g., loyalty, brand competition, and brand image). Finally, the ambiguity of language poses another important challenge to ABSA projects (Cambria et al., 2013; Chifu & Fournier, 2024), especially in the context of social media platforms like Twitter, where the use of slang, idioms, sarcasm, and informal language is prevalent (Farias & Rosso, 2017; Li et al., 2023a; Maynard & Greenwood, 2014). To address these challenges, established NLP approaches such as BERT (Devlin et al., 2019) hybrid models and resampling techniques, specifically addressing class imbalance, will be employed throughout this study to enhance the performance of sentiment analysis model.

1.1 Motivation

The fast-food industry is expanding, with international and local restaurant chains trying to meet the diverse needs of their customers. As consumer preferences become increasingly sophisticated in the competitive hospitality sector, ensuring customer satisfaction has become a critical business priority (Chun & Nyam-Ochir, 2020). Restaurants aim to enhance consumers' positive experiences to increase their likelihood of revisiting (Abdelkafi & Täuscher, 2016; Gupta et al., 2019). Businesses and entrepreneurs have recognized that positive customer feedback is essential for establishing a sustainable long-term business. A better and deeper understanding of customer satisfaction factors enables restaurants to develop and deliver the right products (Chun & Nyam-Ochir, 2020) at the right time. The dynamic and competitive fast-food industry, combined with the rapid growth of digital communication, necessitates using advanced analytical tools to capture consumer sentiment. This study leverages advanced ABSA and resampling techniques as well as explainability methods to identify and interpret consumer sentiments towards specific aspect categories associated with McDonald's.

McDonald's was selected for this study primarily due to the availability of a large volume of relevant and high-quality public data, which fits the needs of ABSA. Numerous tweets posted about McDonald's on Twitter present a rich dataset for examining consumer sentiment. McDonald's is renowned for its significant market share and strong brand recognition (Kee et al., 2021). It is a global fast-food giant that serves approximately 70 million customers daily through its extensive network of around 35,000 locations across more than 100 counties (Mulyo, 2023).

1.2 Goals

The goal of this study is to develop an advanced model to perform Aspect Category Sentiment Analysis (ACSA), a core subtask of ABSA by addressing the aforementioned challenges and enhancing its explainability using state-of-the-art techniques. The model aims to accurately identify and classify fine-grained sentiments, aspect categories, and broad thematic categories within large-scale Twitter data. Upon development and enhancement of explainability, the model will be applied to a real-world dataset of tweets about McDonald's to analyze how customer sentiment has evolved between October 2023-May 2024 and October 2024-May 2025, identifying the aspect categories and broad thematic categories that have contributed most significantly to these changes through ACSA model. These timeframes were selected because online discussions and boycott calls against McDonald’s began in October 2023, making it a relevant point for comparison.

1.2.1 Model Development and Training

Hybrid models have been extensively studied in recent years, with increasing attention given to BERT hybrid models for improving NLP tasks such as sentiment analysis. Xiong et al. (2024) developed a hybrid model named BERT-BiLSTM-CNN, which integrates the power of BERT with BiLSTM and CNN to improve extraction of features. Xin and Zakaria (2024) leveraged the BERT-BiLSTM architecture to detect depression-related content from textual data collected from social media platforms, showing its effectiveness in different datasets including Twitter data. Additionally, Wang et al. (2021) proposed BERT-SAN, an extension of BERT that utilizes a self-attention mechanism to capture aspect-specific information. Many studies have also been conducted in the field of ABSA to handle the challenge of imbalanced aspect and class distribution. For instance, Rozi et al. (2024) addressed aspect imbalance in ABSA by applying SMOTE, which led to significant improvements in evaluation metrics for minority aspect categories.

Motivated by the promising results achieved with the hybrid models (Golbazi, 2024; Wang et al., 2021; Xin & Zakaria, 2024; Xiong et al., 2024), this study uses a hybrid model combining BERT, BiLSTM, and Transformers developed to capture both sequential dependencies and contextual information effectively. As Rozi et al. (2024) proposed, SMOTE is used in the current model to address class imbalance and improve its overall performance. BERT is chosen as the base model for our hybrid model because of its effectiveness in sentiment classification and aspect extraction tasks.

1.2.1.1 Explainability

Perikos and Diamantopoulos (2024) explored the use of transformer models for explainable ABSA. Their study integrated five explainability techniques into their model: LIME (Local Interpretable Model-agnostic Explanation) (Ribeiro et al., 2016), SHAP (SHapley Additive explanation) (Lundberg & Lee, 2017), attention visualization, Integrated Gradients (Sundararajan et al., 2017), and Grad-CAM (Gradient-weighted Class Activation Mapping) (Selvaraju et al., 2017). These methods help build trust among stakeholders by identifying important input features and patterns within the text. Therefore, we implement similar explainability techniques, especially LIME, SHAP, and attention visualization, in our study.

In conclusion, this study aims to utilize an advanced hybrid model (BBT) to perform ACSA and apply explainability techniques to address limited explainability caused by deep learning models, providing transparency into the model's black box nature and its underlying decision-making process.

1.3 Research Questions

To achieve the stated research objectives, the following research questions are addressed:

1. How has customer sentiment towards McDonald's evolved between October 2023-May 2024 and October 2024-May 2025, and which aspect categories have contributed most significantly to these changes, as identified through the proposed ACSA model?
2. How does the integration of resampling techniques and a BERT-based hybrid model improve ACSA performance, particularly in addressing underrepresented aspect categories?
3. How can explainability techniques enhance the transparency of the BERT-based hybrid model for ACSA?

1.4 Contributions

This study offers both theoretical and practical contributions to the field of ABSA, which are outlined in the following subsections.

1.4.1 Theoretical Contributions

- Novel Model Architecture: The study presents a hybrid model that integrates BERT, BiLSTM, and Transformers, enhanced with resampling and explainability techniques to achieve superior performance and enhanced interpretability in ABSA tasks.

1.4.2 Practical Implications

- Informed Business Decision-Making: This study helps better understand the factors and aspects influencing sentiment by providing insights into customer sentiments while incorporating explainability. In turn, this can lead to developing actionable strategies based on transparent decision-making processes.
- Enhanced Customer Understanding: By examining customer feedback, this study offers insights into consumer needs and preferences.

1.5 Structure of the thesis

- Chapter 2: Literature Review
- Chapter 3: Methods
- Chapter 4: Results
- Chapter 5: Discussions and Conclusion
- Chapter 6: Limitations and Future Research

2. Literature Review

This chapter presents a focused literature review that explains the key theories and methods used in this thesis. Section 2.1 introduces the field of Artificial Intelligence (AI) and highlights its importance as the foundation context for the remainder of this study. Section 2.2 introduces basics of machine learning and deep learning, with a particular emphasis on architectures such as BERT, BiLSTM, and Transformers. This section provides the rationale for adopting a hybrid model that integrates BERT with BiLSTM and Transformer layers in this study. In section 2.3, development and evolution of sentiment analysis are discussed, with particular attention to its limitations, to justify the study's focus on Aspect-Based Sentiment Analysis (ABSA). Section 2.4 explains Aspect-Based sentiment Analysis (ABSA), covering its theoretical foundations, practical applications, designs, and recent advances, while identifying challenges and gaps in the field. Section 2.5 delves into explainable AI (XAI) techniques, with particular attention in LIME, SHAP, and attention mechanisms, which are employed in the proposed model to improve transparency. Section 2.6 discusses SMOTE (Synthetic Minority Over-sampling Techniques), a resampling method used to address class imbalance in the training data. Section 2.7 introduces BERTopic, a topic modeling approach designed to extract topics from large textual datasets. Finally, section 2.8 covers evaluation metrics relevant for assessing model performance in ABSA tasks.

2.1 Artificial Intelligence

Since the 1950s, Artificial Intelligence (AI) has rapidly developed, with diverse approaches and goals emerging in the field (Alhosani & Alhashmi, 2024). AI has contributed to major improvements in many industries and has changed the way humans perceive the world (Damirchi & Amini, 2023). They are becoming a crucial necessity for numerous organizations (Ahmad et al., 2023b), offering support in complex tasks and decision-making to humans (Horowitz, 2018). To better understand the foundations of AI, it is helpful to see it as a part of broader field of intelligence sciences. Intelligence science explores both natural intelligence, the study of cognitive abilities demonstrated by living organisms (Estep, 2006), and artificial intelligence, which aims to create intelligent software systems and machines that can handle tasks typically requiring human cognition (Wang, 2019). John McCarthy first introduced the term Artificial Intelligence at the 1956 Dartmouth Conference (Li et al., 2025; Sharma, 2024; Trapple, 1986) and initially defined it as the science of designing machines with intelligence (Ertel, 2017). AI is a broad field with subdisciplines, such as Machine Learning, Deep Learning, and Natural Language Processing (NLP) (Ertel, 2017; Ahmad et al., 2023a).

The past decade's rapid advances in AI algorithms and computational resources have facilitated growth of AI applications in various fields, including "healthcare, autonomous vehicles, criminal justice, human resources, and environmental sciences" (Kale et al., 2023, p. 140). In response to the increasing complexity, volume, and unstructured nature of data as well as the increasing demand for automation, AI has emerged as a critical solution for analyzing and making decisions based on such data (Khomh et al., 2018; Ahmad et al., 2023a). Rapid progress in AI, driven by enhanced computing power and access to large-scale data, has improved its capabilities and practical uses. As a result, it has become a popular complement to software portfolios (Holmquist, 2017; Ahmad et al., 2023a) used by organizations, now widely included in digital products across diverse domains (Holmquist, 2017) such as healthcare (Jiang et al., 2017), science (Li & Du, 2017), virtual assistance (Theosaksomo & Widyantoro, 2019; Mekni et al., 2020), and automotive industry and self-driving vehicles (Schroeder et al., 2015). Using AI in software engineering helps organizations work more efficiently and potentially reduce costs by automating tasks such as initial code generation, debugging, optimization, and performance improvement (Alenezi & Akour, 2025).

Generative AI is a subset of AI system capable of creating content in various formats, such as code, textual output, and images (Lightstone, 2024; Pinaya et al., 2023). Generative AI models have been developed not only to analyze and extract patterns from training data but also create new content that is similar to the original training data (Dakhel et al., 2024). Large Language Models (LLMs), which are a type of generative AI, have demonstrated exceptional performance in language understanding and generation tasks (Min et al., 2023; Zhao et al., 2023; Dakhel et al., 2024).

2.1.1 AI's Impact and Challenges

Machine learning along with its subfield, deep learning, are among the most important components of AI (Sarker, 2021a, 2021b, 2023). They have significantly influenced human decisions made in various fields and industries. Despite their advantages, they also present specific challenges. The limited transparency of complex AI models has led to growing concerns about their interpretability, as well as their ability to provide clear, human-understandable explanations for decision-making processes (Goodman & Flaxman, 2017).

Furthermore, due to the data-dependent nature of AI, the quality of the data plays an important role in shaping their accuracy predictability and transparency (Ahmad et al., 2023c), with model performance and outputs often remining uncertain until the model is trained and evaluated (Belani et al., 2019; Agarwal & Goel, 2014; Khomh et al., 2018).

2.2 Machine Learning

Machine learning aims to explore how computers can mimic human learning processes (Wang et al., 2009), specifically, by learning from data and reproducing outcomes similar to those made by humans. Machine learning mainly aims to develop systems that learn from data, improve automatically from experience, and make informed decisions without the need to write detailed codes and explicit programming (Samuel, 1959; Mahesh, 2020). It lies at the crossroads of computer science and statistics and plays a key role in artificial intelligence and data science (Mann et al., 2021; Wang et al., 2009). It is a "driving force in shaping the fourth industrial revolution" (Sarker, 2021a, p. 1). By identifying patterns within past data, machine learning algorithms enable predictive analytics (Sarker, 2023) and automate repetitive data analysis tasks (Larsen & Becker, 2021). Compared to humans, machine learning models require large volume of data to learn and identify patterns; however, they can process and apply data at speeds far beyond human capability (Kühl et al., 2022). Due to this capability, it has made important contributions to various aspects of human society (Wang et al., 2009). Among its many applications, machine learning offers the greatest benefits to industries facing data-intensive challenges (Jordan & Mitchell, 2015). This technology has brought significant advancements to a wide range of such industries, including healthcare (Alanazi, 2022), finance (Yazdani, 2021), marketing (Chen et al., 2017a), customer service (Jain & Kumar, 2020), manufacturing (Rude et al., 2018), and transportation (Tizghadam et al., 2019).

Machine learning primarily focuses on developing algorithms that enable computers to learn. This involves the identification of statistical regularities and patterns present in the data. Human learning processes inspire these machine learning algorithms and represent aspects of human learning and learning challenges in different environments (Nasteski, 2017).

Machine learning models use mathematical functions, usually represented as f(x), to map input data (x) to output (y). The results may differ depending on the task; as shown in figure 1, it may be a categorical label in classification tasks or a continuous value (real numbers) in regression tasks (Wang, 2016).

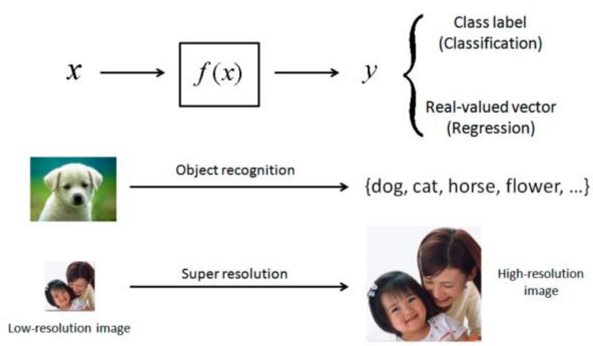

Figure 1. Machine Learning Function. Adapted from Wang et al. (2016, p. 2).

Regardless of the specific task, machine learning models also require optimizing two critical components to generate accurate predictions: fit parameters (Ω) and hyperparameters. While fit parameters are directly and automatically learned from data, hyperparameters that control training are tuned externally (IBM, n.d.).

Machine learning algorithms are divided into the following categories based on their intended outcome: supervised learning, unsupervised learning, semi-supervised learning, reinforcement learning, transaction, and learning to learn. However, they are broadly categorized into two general groups: supervised learning and unsupervised learning (Nasteski, 2017).

The learning process is called supervised learning when each instance is presented with its known label (correct outputs). In contrast, in unsupervised learning, instances do not have labels (Jain et al., 1999; Kühl et al., 2022).

2.2.1 Neural Networks

Although modern computers are very fast at numerical and symbolic computations, machine learning models struggle to generalize to new distortions and unseen conditions, particularly in perceptual tasks such as language and image recognition, where humans still show strong and consistent performance (Geirhos et al., 2018). This is because computers heavily rely on precise input and sequential instructions, while the human brain performs tasks and processes information in parallel. Inspired by the architecture of the human brain, Artificial Neural Networks (ANNs) were developed in an attempt to mimic this ability of the human brain (Zou et al., 2009); however, they remain highly simplified versions of biological learning processes. ANNs consist of layers of interconnected processing nodes or artificial neurons. ANNs are defined by three key elements: node characteristics, network structure, and learning rules. Node characteristics show how a node processes signals, including the number of inputs and outputs, the weights of those connections, and the activation function used to transform input into output. Network structure refers to the overall architecture of the model, including the number of the layers, the number of nodes per layer, and how these nodes are arranged and connected within network. These connections control how data is passed from one layer to another in the network. In contrast, learning rules define how weights are initialized and updated during training (Zou et al., 2009).

Although approaches to classification using machine learning and neural networks can classify data, they often struggle when applied to unstructured and semantically complex data present in platforms such as Twitter. Specifically, these approaches often exhibit limitations in understanding context and modeling connections between distant words. Therefore, to gain a more detailed understanding of natural language in context of tweets, this study employs a deep learning-based approach. In the next section we will explain key deep learning architectures, focusing on those integrated into the hybrid model developed for ACSA in this study.

2.2.2 Deep Learning

Deep learning builds on traditional neural network architectures and tends to perform better than classical machine learning models, especially when applied to large datasets with sufficient training data (Alzubaidi et al., 2021). The concept of neural networks has a long history and was a prominent topic in the late 1980s. However, interest in them decreased over time. The downward trend continued until the introduction of deep learning as a branch of machine learning by Hinton et al. (2006). Due to its remarkable success in classification and regression tasks, deep learning significantly boosted interest in neural networks. It is now the subject of intense research for leading corporations such as Microsoft and Google (Sarker, 2021a). Deep learning methods were introduced to address the limitations of machine learning. Machine learning typically relies on statistical techniques to learn from data, whereas deep learning as a subfield of machine learning uses neural networks with multiple hidden layers to analyze extensive datasets and learn hierarchical and abstract representations. Unlike machine learning and shallow neural networks, deep learning architectures can automatically extract intricate features by transforming data through multiple layers of nonlinear computation (Bengio, 2009; LeCun et al., 2015;

Sharifani & Amini, 2023). Machine learning methods have difficulty processing raw natural data effectively. They required extensive feature engineering, where domain experts manually created feature extractors to transform raw data into representations suitable for learning algorithms, often a classifier, which were then used to identify patterns in the input (LeCun et al., 2015). However, manually choosing features can often introduce selection bias, which can cause poor classification between classes (Alzubaidi et al., 2021). This limitation is addressed by representation learning. Deep learning utilizes representation learning to automatically extract relevant and meaningful features from raw data for tasks such as detection and classification. The distinction between machine learning and deep learning is illustrated in Figure 2. While this figure shows that deep learning reduces the need for manual preprocessing, it includes more than just skipping feature selection. Instead, deep learning distinguishes itself by the use of deep (multi-layered) neural architectures that automatically learn hierarchical representations from raw input, unlike machine learning techniques such as decision trees or random forests, which rely on manually engineered features and do not involve multi-level representation learning.

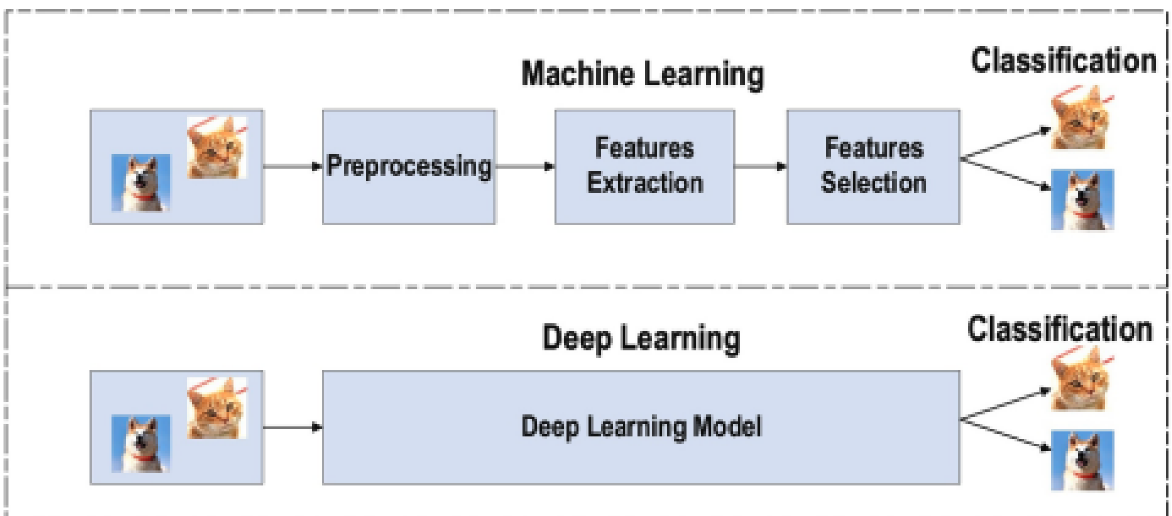

Figure 2. The difference between deep learning and traditional machine learning. Adapted from Alzubaidi et al. (2021, p. 7)

Deep learning builds several hierarchical layers to gradually learn complex features (Sarker, 2021a). While there is no universally agreed-upon threshold distinguishing shallow from deep neural networks, LeCun et al. (2015) define deep learning by its emphasis on neural depth, particularly the use of multiple hidden layers, as a core strategy to improve performance on complex tasks. These layers progressively convert raw inputs (simple things), such as pixel values within an image, into more meaningful representations (more understandable forms). Each layer detects more complex and detailed patterns, allowing the model to better understand the complex features. An important advantage of this process is that layers of features are learned directly for data with minimal manual intervention (Sarker, 2021a), although the difference from machine learning is not always clear-cut (Schmidhuber, 2015). Deep learning has attracted significant attention in different fields such as data science and analytics because of its data-driven learning capabilities (Sarker, 2021a).

It has also contributed substantially to AI advancements by solving complex problems that remained challenging for many years. It is highly effective at finding complex patterns in large datasets, which makes it useful across many fields (LeCun et al., 2015). Although the roots of deep learning lie in early work on deep belief networks (Hinton et al., 2006), the approach began to gain considerable attention around 2011-2012, when it proved successful in tasks like Natural Language Processing (NLP) (Collobert et al., 2011), speech recognition (Hinton et al., 2012), and image classification (Krizhevsky et al., 2012). In addition to its successes in the mentioned fields, it proved effective in diverse fields such as drug discovery (Ma et al., 2015), genetic analysis (Xiong et al., 2015) and particle physics (Ciodaro et al., 2012).

By advancing Natural Language Processing (NLP) capabilities (Collobert et al., 2011), deep learning has contributed to AI's natural language understating and generation (Radford et al., 2019) in tasks such as text classification, including sentiment analysis (Zhang et al., 2018), and question answering (Minaee et al., 2021) as well as language translation (Jean et al., 2014; Sutskever et al., 2014). To better understand the relationship between AI, machine learning, and deep learning, Figure 3 illustrates how each is built upon the other.

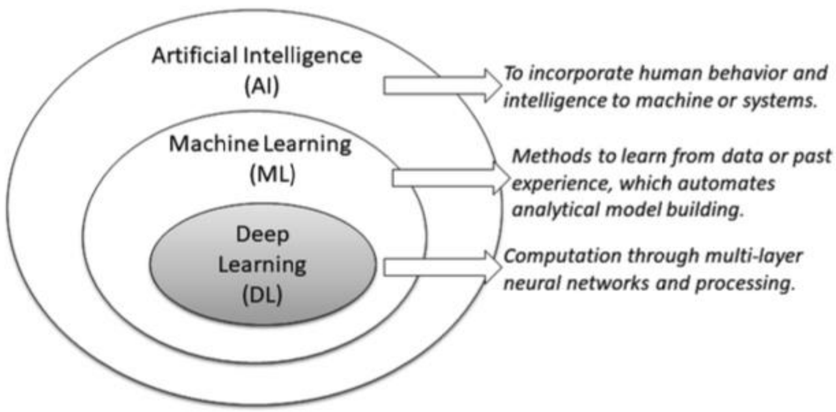

Figure 3. Position of Deep Learning within Machine Learning and AI, adapted from Sarker (2021a, p. 420)

Natural Language Processing (NLP) is a field closely related to computational linguistics (Tsujii, 2021), which allows computers to understand and process human language (Darwish et al., 2021). Since the 1980s, statistical methods, probability, and machine learning have been widely used in NLP tasks. However, thanks to stronger computing power, particularly in GPUs, deep learning has become widespread in NLP tasks. Additionally, the availability of massive datasets has further sped up the use of deep learning in NLP applications (Otter et al., 2021).

Deep learning models have led to remarkable success across different Natural Language Processing (NLP) applications, such as text classification (Asudani et al., 2023; Chai et al., 2020; Radford et al., 2018; Wang et al., 2018; Zhou et al., 2015), machine translation (Edunov et al., 2018; Vaswani et al., 2017; Zhu et al., 2020), natural language understanding (Lan et al., 2019; Sun et al., 2019; Yang et al., 2019) and dialog (Adiwardana et al., 2020; Baheti et al., 2018; Xu et al., 2018). In modern NLP architectures, neural networks are used to convert words into vector representations, which are then further processed by downstream components such as RNNs (Bowman et al., 2015; Socher et al., 2013), LSTMs (Hochreiter & Schmidhuber, 1997), CNNs (Kalchbrenner et al., 2014), and Transformers (Vaswani et al., 2017) to identify and learn complex language patterns (Sun et al., 2021).

2.2.2.1 Recurrent Neural Networks (RNNs)

Recurrent Neural Networks (RNNs) are type of deep learning architectures which are particularly effective for analyzing time-series and sequential data (Weerakody et al., 2021). Therefore, they are well-suited for NLP tasks, speech processing (Batur Dinler & Aydin, 2020; Jagannatha & Yu, 2016), language modeling (Graves, 2013; Mikolov, 2012; Sutskever et al., 2011), word embedding learning (Mikolove et al., 2013a), real-time handwriting recognition (Graves et al., 2008), speech recognition (Graves et al., 2013), time series forecasting, and language translation (Tarwani & Edem, 2017). Google Translate and Siri are practical applications of RNNs (Johri et al., 2021). Their ability to process sequential data allows them to use sequence structure for tasks like understanding the meaning of individual words within a sentence. RNNs operate with a short-term memory mechanism, comprising input (x), output (y), and hidden states (s) layers, as illustrated in the unfolded diagram in Figure 4 (Alzubaidi et al., 2021).

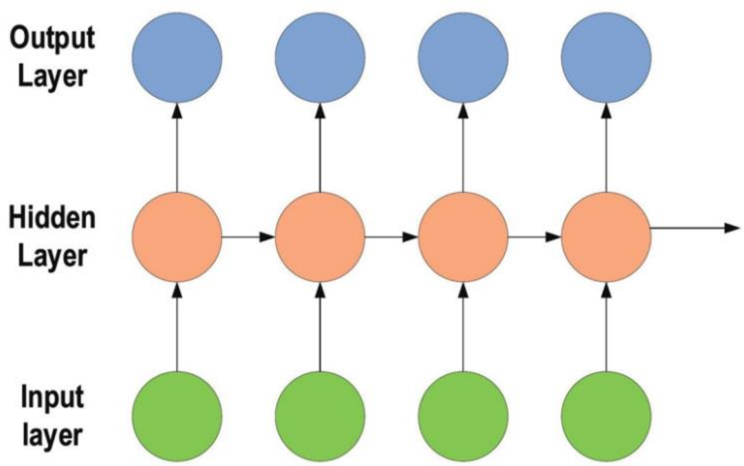

Figure 4. Unfolded RNN Diagram. Adapted from Alzubaidi et al. (2021, p.13)

Pascanu et al. (2013) identified three deep RNN architectures: Hidden-to-Hidden, Hidden-to-Output, and Input-to-Hidden. These connection types help address training challenges in deep RNNs and exploit the performance benefits of deeper RNNs (Alzubaidi et al., 2021). These techniques are more effective because, as Bengio 2009 put it, deep learning is based on the idea that hierarchical models with multiple layers represent complex functions more efficiently compared to shallow networks with fewer layers.

A key limitation in training RNNs is the occurrence of gradient issues, specifically vanishing and exploding gradients (Glorot & Bengio, 2010), which can be addressed using Long Short-Term Memory (LSTM) networks (Gao et al., 2019).

2.2.2.1.1 Long Short-Term Memory (LSTM)

RNNs often struggle to learn long-term dependencies because of vanishing and exploding gradients (Curreri et al., 2021; Siami-Namini et al., 2019). LSTM networks developed by Hochreiter and Schmidhuber (1997) address these challenges by incorporating a memory component. LSTM Networks utilize memory blocks containing memory units that store temporal states and dependencies over extended sequences. They also leverage gated units to control information flow within the network (Greff et al., 2016; Siami-Namini et al., 2019). In RNNs, particularly LSTM networks, as shown in Figure 5, a repeating module with a cell state and gated units, manages how data flows through the network sequentially. This helps the network understand long-term dependencies more effectively. Therefore, LSTM architecture is capable of learning what information is important to be kept and what information can be removed. It can preserve information over long periods (Gao et al., 2019; Siami-Namini et al., 2019). A basic RNN contains a single-layer module, whereas an LSTM network has a more complex module with four interacting layers. The key component of an LSTM is the cell state (C_k), which acts as a memory to store information over time. This cell state is illustrated as a horizontal line in Figure 5. The cell state is controlled by gates that add or remove information from them. These gating mechanisms typically use a sigmoid activation layer, followed by pointwise multiplication, to ensure efficient data flow within the network (Gao et al., 2019).

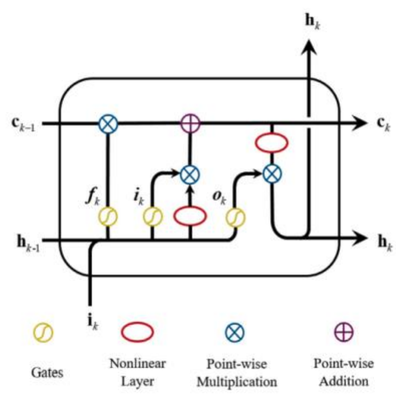

Figure 5. LSTM Structure, adapted from Gao et al. (2019,p.284)

Unlike standard RNNs, which typically use a single, repeating module typically involving a hyperbolic tangent function (tanh), LSTM uses a more advanced architecture that includes multiplicative units. With the assumption that all activation functions in the nonlinear network are tanh, the key components of the LSTM namely forget gate (f_k), input gate (i_k), cell state (c_k), output gate (o_k) and hidden state (h_k) are calculated at each time step k as illustrated below:

$$f_k = \sigma\left(W^f.\left[h_{k-1,}i_k\right] + b^f\right)$$
$$i_k = \sigma\left(W^i.\left[h_{k-1,}i_k\right] + b^i\right)$$
$$c_k = f_k \odot c_{k-1} + i_k \odot \tanh\left(W^c.\left[h_{k-1,}i_k\right] + b^c\right)$$
$$o_k = \sigma\left(W^o.\left[h_{k-1,}i_k\right] + b^o\right)$$
$$h_k = o_k \odot \tanh(c_k)$$

Where the terms W and b present learnable weights and biases. The subscripts '*f*', '*i*', and '*o*' show which gate they belong to within LSTM network: forget, input, and output.

The forget gate plays an important role in enabling Backpropagation Through Time (BPTT) to effectively propagate error signals during training. In other words, it allows the LSTM to learn from its mistakes, particularly when the model needs to learn and retain long-term dependencies or long connections between different parts of sequential data (Gao et al., 2019).

Bidirectional LSTM (BiLSTM) models enhance traditional LSTMs by applying two LSTM layers to input data. One processes the input from beginning to end (forward direction), while the other processes it from end to beginning (backward direction) (Siami-Namini et al., 2019; Wang et al., 2023). In other words, to build a BiLSTM network, the LSTM neurons are divided into two groups. One processes the forward states, and the other processes the backward states (Schuster & Paliwal, 1997).

Figure 6 shows the difference between LSTM and BiLSTM networks. BiLSTM processes input data from both past and future time frames. In contrast, standard LSTM only processes past information. This leads to delays in capturing and understanding future context (Alizadegan et al., 2025; Mahadevaswamy & Swathi, 2023).

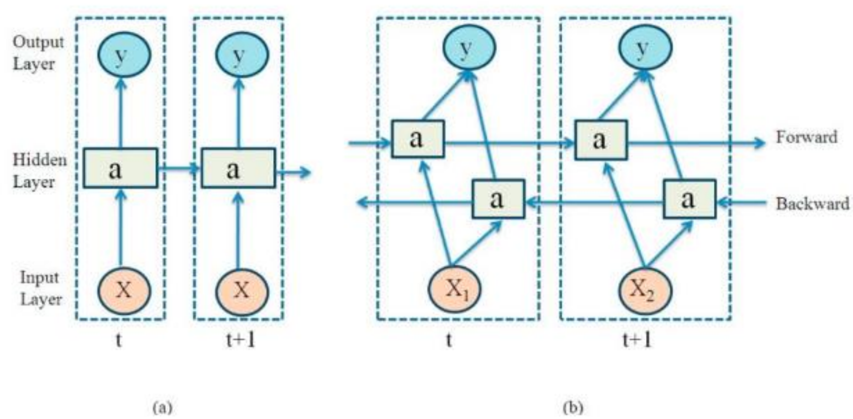

Figure 6. Architectures of (a) LSTM and (b) BiLSTM networks. Adapted from Mahadevaswamy & Swathi (2023, p. 49)

2.2.2.2 Transformer-based Models

RNNs (such as LSTMs and GRUs) process data in order and have a step-by-step structure due to their recurrent nature, which makes them slow to train. To address this challenge, Vaswani et al. (2017) proposed Transformer architecture as illustrated in Figure 7. Transformers, on the other hand, use a technique called 'attention' mechanism to process information all at once through parallel computation. Transformers are effective at learning long-distance relationships in textual data. Attention helps the transformer focus on the most important words. This improves its understanding of the sentence's overall meaning (Gillioz et al., 2020).

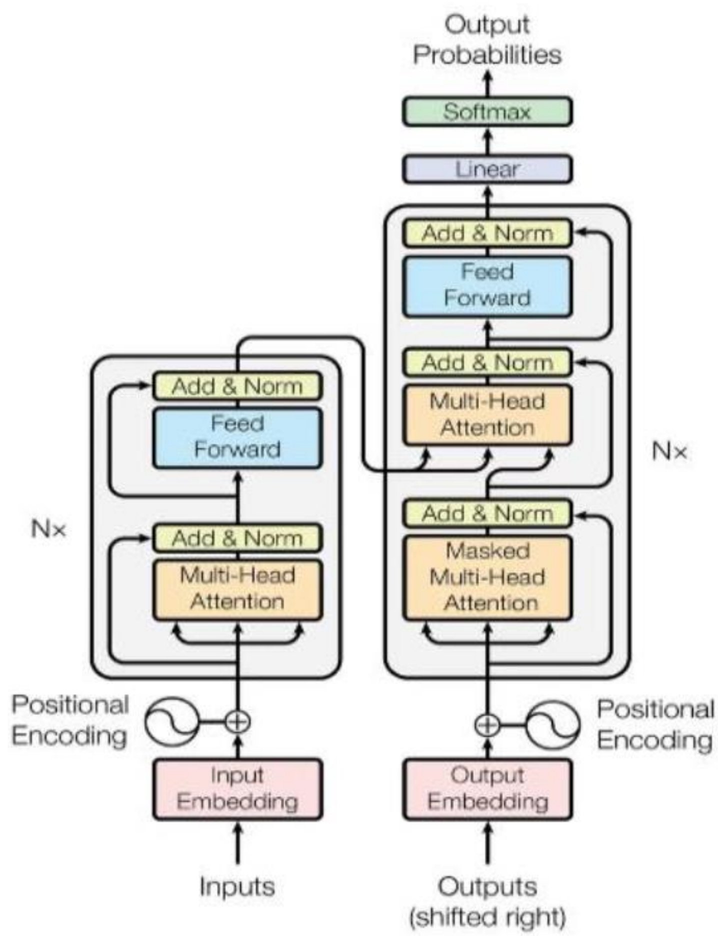

Figure 7. Transformer architecture. Adapted from vaswani et al. (2017)

The training of transformer models is commonly performed in two stages, using unsupervised pretraining (a type of semi-supervised learning). First, the model is exposed to massive amounts of text without labels (unsupervised pretraining). At this stage, it learns general linguistic patterns and structures through pretraining objectives — such as masked-word prediction used in BERT (Devlin et al., 2019) and next-word prediction used in GPT (Radford et al., 2019)— specifically designed to capture syntactic and semantic relationships; these mechanisms will be thoroughly explained in detail in the following subsection on the Bidirectional Encoder Representations from Transformers (BERT) architecture. Then, the pre-trained model is fine-tuned for specific downstream tasks using labeled data (Gillioz et al., 2020). Transformers such as BERT excel at many NLP tasks, including Aspect-Based Sentiment Analysis (ABSA). Fine-tuning these pretrained models allows ABSA systems to learn long-range relationships and contextual nuances better. This process significantly improves sentiment extraction (Xu et al., 2019).

The original transformers used an encoder-decoder structure. The encoder takes tokens (e.g., words or subwords) in a sentence and turns them into a special code known as latent vectors, which are numbers that help the model understand what each token means, and how it is related to other tokens within a sentence. The latent representation contains important information about the sentence, its tokens, and their context. This conceptual representation is similar to how the human brain forms ideas and organizes them (Perlovsky, 2006). The decoder then uses these latent vectors to generate an output sentence, for example, to translate an input sentence into another language. As another example, in computer vision, transformers can take an image that was captured during the day and change it so that it looks like it was taken at night. BERT and Generative pre-trained Transformer (GPT) are recent NLP models that are based on transformer architecture. While BERT mainly uses multiple encoder layers. GPT primarily utilizes multiple decoder layers (Ghojogh & Ghodsi, 2020).

The attention mechanism is the foundation of Transformers' ability to process language. As a key component, single-head attention enables the model to focus on specific input tokens. It involves two key stages: Transformation and aggregation phases.

First, the input sequences are transformed into three vectors: queries (Q), key (K) and value (V). Second, an attention layer calculates attention weights based on the similarity between queries and keys, where n and d represent the length and dimension of the inputs, respectively. The attention mechanism is illustrated in Figure 8. The scaled dot-product attention is illustrated below (Liu et al., 2024).

$$\text{Attention (K, Q, V)} = \text{Softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)\text{V}$$

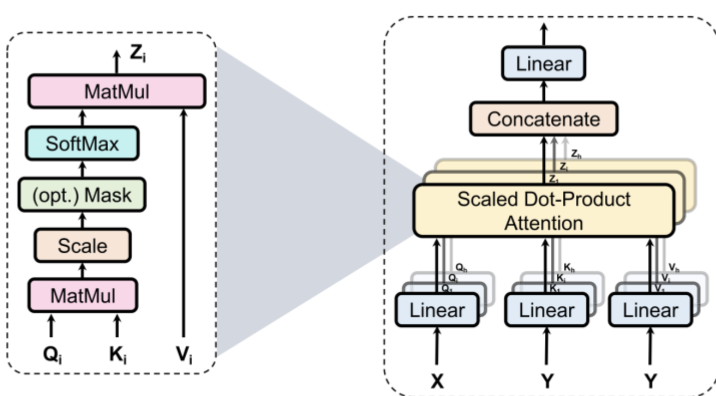

Figure 8. Overview of Attention Layer. Left side: scaled dot-product attention. Right side: multihead attention mechanism. Adapted from Liu et al. (2024, p. 7479)

2.2.2.2.1 Bidirectional Encoder Representations from Transformers (BERT)

BERT (Devlin et al., 2019) is a powerful transformer-based language model that has become one of the most influential models in NLP. It uses multiple layers of Transformer encoders. BERT is known for its strong performance in NLP tasks such as question answering (Qu et al., 2019), natural language understanding (Dong et al., 2019), and ABSA (Song et al., 2020). BERT adopts the technique of Masked Language Modeling to capture contextual relationships between words (tokens) within a sentence (Devlin et al., 2019). Throughout this process, 15% of the words in the input text are randomly hidden (masked). The model then tries to guess the missing words (tokens). As shown in Figure 9, each transformer encoder block receives a masked sentence and tries to guess the missing word. During this stage, the model is trained to predict missing words within a sentence based on the surrounding contextual information. BERT does not require manually labelled data throughout pretraining, as it uses self-supervised learning to predict masked words based on contextual information, with the text itself providing the learning signals (Devlin et al., 2019; Kotei & Thirunavukarasu, 2023). In fact, any word can be hidden, and the model learns from the text itself. This enables BERT to be pre-trained on massive amounts of publicly available textual data from the Internet. To guess a hidden word, BERT uses attention to all the words that come before and after the hidden one. Since BERT understands a word based on its context in a sentence, it generates context-aware word embedding. Therefore, the representation of a word can change depending on the word's context. In contrast, earlier models such as Word2Vec (Mikolov et al., 2013b) and GloVe (Pennington et al., 2014) produced a fixed embedding for each word regardless of its context (Ghojogh & Ghodsi, 2020). For instance, BERT assigns different embeddings for the word "spring" in "Many bulbs bloom in (the) spring." and "I sprang out of bed to answer the door" (Cambridge Dictionary, n.d.).

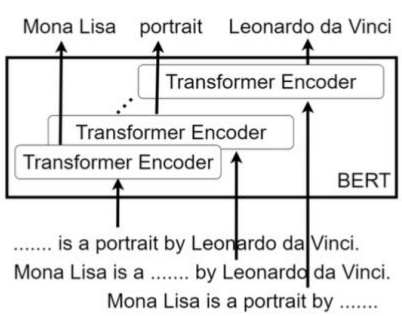

Figure 9. Training BERT using sentence with mask tokens. Adapted from Ghojogh & Ghodsi (2020, p. 10)

One of the useful functions of BERT is providing contextualized embeddings in sentence and word levels (Devlin et al., 2019). Sentence-level embeddings help NLP models handle tasks such as sentiment analysis and spam detection. In addition to word and sentence level embeddings, BERT is also designed to perform the next sentence prediction task. The model decides whether a sentence can logically follow another sentence. Unlike traditional methods, BERT's bidirectional architecture (Tenney et al., 2019) enables text processing in both directions. This contributes to a deeper and more detailed contextual understanding and, as a result, more meaningful vector representations. Since BERT has already been pretrained on an extensive amount of textual data available on the internet, it is not necessary to train the model from the beginning. Instead of implementing new tasks such as sentiment analysis, transfer learning is applied to add some new layers to the BERT model, which has already collected a significant amount of linguistic knowledge throughout the pretraining stage. The added layers are then trained on the target task and adjusted to improve the model's performance. Throughout the training process— following a common transfer learning approach— BERT's weights can remain unchanged while only added layers are trained, or the entire model can be fine-tuned using backpropagation. The encoder architecture of the original Transformer proposed by Vaswant et al., 2017, uses a configuration of 6 layers, 512 hidden layers, and 8 attention heads. In comparison, the standard BERT model (Devlin et al., 2019) is built with 24 layers, 1024 hidden units, and 16 attention heads. This version requires too much memory and computational resources to run. These limitations make using them in embedded systems difficult (Ghojogh & Ghodsi, 2020). To address this limitation, lighter and more efficient variants of BERT such as Small BERT (Tsai et al., 2019), Tiny BERT (Jiao et al., 2019), DistilBERT (Sanh et al., 2019), and RoBERTa-base (Staliūnaitė & Iacobacci, 2020) have been developed.

The earlier sections provided an overview of major deep learning architectures, identifying their strengths and weaknesses in the context of NLP. Based on this comparative analysis, this study developed a hybrid architecture combining BERT with BiLSTM and Transformer layers. With the model foundation established, the next section introduces sentiment analysis and its relevance to this study.

2.3 Sentiment Analysis

The widespread use of digital data driven by the rise of the information revolution has created new types of economies (Serrat, 2017). The popularity of digital communication tools such as the internet and mobile devices has provided unique opportunities for detailed study of human behavior. Different online platforms, such as forums and social media, produce real-time data with moment-to-moment changes (Steinert-Threlkeld, 2018). Due to the Integration of the internet in many people's lives and the growth of social media and e-commerce platforms, the number of online comments has increased dramatically and has become a valuable source for understanding customer satisfaction and opinions about services and products (Jain et al., 2021). Most of the data collected from such platforms is unstructured (Gandomi & Haider, 2015). Social media interactions often include slang and abbreviations, which make it hard for traditional tools to analyze (Shmueli et al., 2017). The fast growth of AI and computational power (Duan et al., 2019) along with cloud-based services and tools, have led to substantial improvements in NLP. In response to the need for a better understanding of customer sentiment and opinion, an NLP technique called sentiment analysis has been increasingly used to analyze textual data (Jain et al., 2021; Liu, 2012). This technique is also called opinion mining (Chaturvedi et al., 2018; Liu & Zhang, 2012; Mao et al., 2024).

In recent years, many research papers have studied sentiment analysis on social media (Al-Tameemi et al., 2022; Gunasekaran, 2023; Kumar et al., 2023). Besides academic research, it has gained significant attention from many businesses as a tool for developing effective marketing strategies (Rodríguez-Ibáñez et al., 2023). Sentiment analysis has different applications, such as evaluating consumers' opinions on products (Geetha & Renuka, 2021) and services (Bensoltane & Zaki, 2022), as well as tracking brand image and identifying market trends through social media (Bonifazi et al., 2022; Patil & Kolhe, 2022). Additionally, understanding human sentiments has helped machines to create more contextually appropriate and emotionally informed intelligent responses. It has significantly improved the performance of super smart question-answering and Large Language Models (LLMs) such as Chat GPT and ERNIR (Huang et al., 2022; Mao et al., 2024; Sudirjo et al., 2023; Susnjak, 2024). While these models may not directly perform sentiment analysis in traditional sense, recent research reveals that they can respond in ways that reflect the emotional tone of the user's input (Broekens et al., 2023; Li et al., 2023b).

2.3.1 Sentiment Analysis Levels

Sentiment analysis can be implemented at different levels: document, phrase, and aspect (Behdenna et al., 2016; Do et al., 2019; Mao et al., 2024).

Document-level sentiment analysis examines the overall sentiment expressed in an entire document. It considers each document as an independent unit and assigns only one sentiment label to it. Therefore, the task focuses on the overall idea, rather than details. Different researchers have studied document-level sentiment analysis. For instance, in 2022, Mao et al. employed attention-based BiLSTM and a 1D CNN, adapted for textual feature extraction, to implement a sentiment analysis on document-level. Wen et al. (2020) introduced a model based on the idea that similar reviews are often written by people who share similar sentiments. Therefore, throughout the study, they exploited document similarity as an enhancement strategy to improve the model's performance in extracting sentiment.

Sentence level is the next level of detail in sentiment analysis. Its object is to identify the sentiment of a single sentence. In most sentiment analysis systems, the sentiment is commonly labeled as either negative, neutral, or positive (Liu, 2012). In the first phase of the task, the sentences are classified into objective and subjective. Objective sentences contain factual content (data) and do not express personal opinions, whereas subjective sentences contain personal feelings and opinions. A sentiment is then associated with subjective sentences (Mao et al., 2024). In their study, Chen et al. (2017b) used sequential-based models that considered word order to identify the sentence's sentiment in relation to its topic. Su et al. (2023) presented a supervised learning approach that followed a sequential (step-to-step) process.

Finally, aspect-level sentiment analysis is a more detailed approach to identify sentiments. Rather than identifying overall sentiment of an entire review, aspect-level sentiment analysis takes a more detailed approach by focusing on specific parts referred to as aspects and identifying exact sentiment target. Therefore, the fundamental objective of these models is to determine how aspect terms, aspect categories, opinion expressions and sentiment labels are related to each other (Wu et al., 2018). The process typically involves two steps: first, identifying aspects and the terms describing them, and then assigning the sentiment labels to each identified aspect (Mao et al., 2024). For instance, in sentence "The hamburger is really delicious", the "hamburger" is aspect terms, and the aspect category is "food", and the sentiment labels associated with the aspect is positive.

2.3.2 Sentiment Analysis Techniques

Sentiment analysis methodologies can be classified into four main categories: lexicon-based approaches, machine learning techniques, deep learning architectures, and hybrid frameworks (Mao et al., 2024; Madhoushi et al., 2015; Sankar & Subramaniyaswamy, 2017; Thakkar & Patel, 2015). The classification suggested by Mao et al. (2024) is illustrated in Figure 10. In addition, to provide a more comprehensive view, Figure 11 further illustrates methodologies used in sentiment analysis, along with their prevalent challenges and various applications.

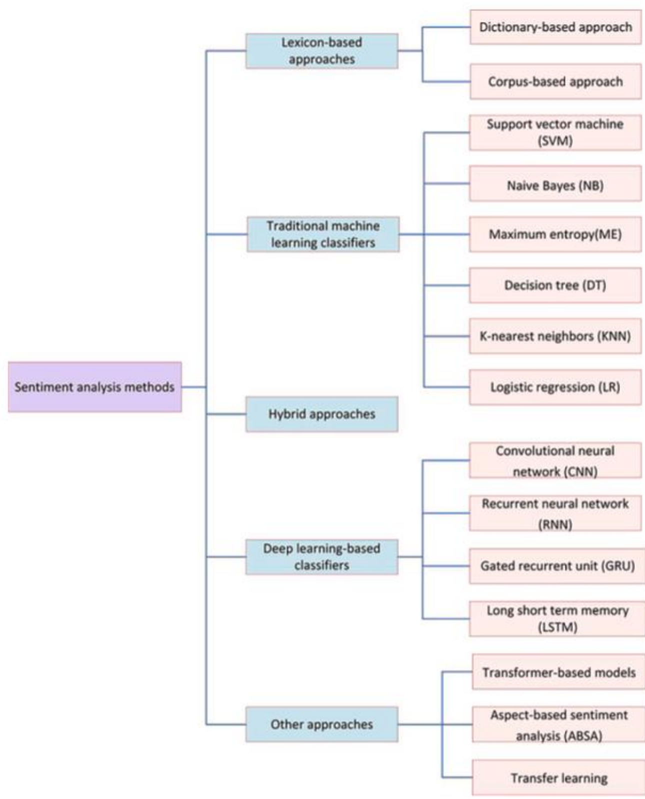

Figure 10. Classification and Learning Techniques Frequently Applied in Sentiment Analysis. Adapted from Mao et al. (2024)

2.3.2.1 Lexicon-based approach

Lexicon-based approaches use sentiment lexicon which is a dictionary of tokens (words or subwords) that have been assigned sentiment scores (Bonta et al., 2019; Taboada et al., 2011). It is an unsupervised method that is highly dependent on domain because sentiments associated with words can vary depending on the context (Mao et al., 2024). For instance, the sentiment associated with the word "sharp" can vary depending on the context as seen in the following sentences: "Her mind was as sharp as razor" and "Emma has a sharp tongue" (Oxford University Press, n.d.). According to Mitra & Mohanty (2020), lexicon-based approaches are typically classified into two categories: dictionary-based and corpus-based methods.

2.3.2.2 Machine Learning Models

Earlier sentiment analysis approaches often relied on machine learning methods, some conventional machine learning classifiers such as Naïve Bayes (NB) (Kang et al., 2012), Support Vector Machines (SVM) (Ahmad et al., 2017), Decision Trees (DT) (Myles et al., 2004) can be used in sentiment analysis systems.

Although machine learning-based models are designed to automatically identify patterns and features, traditional supervised machine learning models such as Naïve Bayes (Webb et al., 2010), Support Vector Machines (Noble, 2006) and artificial neural networks (Agatonvic, 2000) heavily depend on labeled data and extensive training (Phan et al., 2023). In the context of sentiment analysis, these models often require large quantities of manually annotated data (Medhat et al., 2014) and are highly sensitive to the domain they were training on (Phan et al., 2023). This limits their ability to generalize across different domains and extract relevant aspects (Wang et al., 2024) and sentiment information from textual data. The following section explores deep learning-based approaches, which offer greater flexibility and improved performance in sentiment analysis tasks.

2.3.2.3 Deep Learning Techniques for Sentiment Analysis

In recent years, deep learning techniques have led to significant improvements in sentiment analysis tasks by helping models to identify important features and represent them in low-dimensional vectors (Li et al., 2020; Zholshiyeva et al., 2024). This shift has enabled models to automatically learn features without the need for manual work: however, large volumes of annotated data are still typically required. Overall, in addition to removing the need for manual feature engineering, deep learning models bring several other advantages for sentiment analysis tasks, including better contextual understanding, the ability to perform complex tasks such as ABSA, support for transfer learning via pretrained embeddings, and enhanced performance across different domains (Zhang et al., 2018).

2.3.2.4 Hybrid Approach

Sentiment analysis typically employs hybrid methodologies that use a combination of approaches. These typically include lexicon-based, machine learning, and deep learning approaches (Appel et al., 2016). This combination allows the sentiment analysis system to analyze both the linguistic meaning of individual words and their contextual semantics (i.e., their role within a broader textual context) (Mao et al., 2024). Obiedat et al. (2022) developed a hybrid sentiment analysis model integrating SVM, PSO for feature selection and oversampling techniques (SMOTE and ADASYN) to address class imbalance. The model achieved a high F1-score (96.50%), showing its effectiveness.

2.3.2.5 Other Approaches

Sentiment analysis has evolved from initial lexicon-based approaches and machine learning classifiers, such as Naïve Bayes and SVM (Ahmad et al., 2018), to more advanced deep learning models like CNNs and RNNs. However, Transformer-based models, categorized separately in Figure 17 under the label ‘other approaches’, have addressed limitations of earlier models through self-attention mechanisms and bidirectional processing. In sentiment analysis, transformers have proven highly effective at capturing nuanced language and contextual relationships that earlier models often miss (Bashiri & Naderi, 2024).

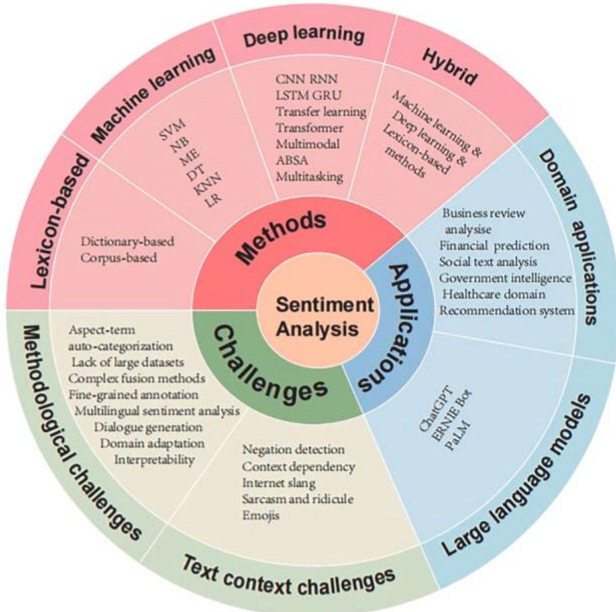

Figure 11. Techniques, challenges and applications of sentiment analysis. Adapted from Mao et al. (2024, p.5)

2.3.3 Applications of Sentiment Analysis

Both academic research and various industry sectors have benefited from sentiment analysis. For example, it helps businesses adjust their marketing plans and offers products and services with better quality (Birjali et al., 2021). Sentiment analysis has also been used to predict trends in financial markets and stock prices (Mao et al., 2024). As demonstrated in a study by Xing et al. (2018), there was a relationship between the sentiment of news (negative or positive) and the downward or upward trends in stock prices. In a similar study, Rognone et al. (2020) examined the effects of new sentiment on the performance of currencies and cryptocurrencies, including Bitcoin. With the growing popularity and use of social media platforms, more researchers focus on reviews and comments posted on social media platforms. For instance, Baker et al. (2023) in their study, analyzed the sentiments on Twitter about the Russian-Ukraine conflict. Similarly, Jihad et al. (2022) used machine learning techniques to explore the consumer’s sentiments about electric vehicles. Sentiment analysis has also been widely used in healthcare to study patient sentiments, adverse medicine reactions, disease outbakes, and trends (Ramírez-Tinoco et al., 2019). Chintalapudi et al. (2021) proposed a system integrating sentiment analysis and text mining to support seafarer monitoring by doctors. In another study, Baker et al. (2022) employed GRU, LSTM, and CNN to examine colon cancer datasets. The findings showed that these models can be helpful in disease prognosis.

2.4 Aspect-Based Sentiment Analysis

Aspect-based sentiment analysis (ABSA) has emerged to address limitations of traditional sentiment analysis that aim to extract overall sentiment of input text in document or sentence levels

(Zhang et al., 2022). ABSA is fine-grained sentiment analysis which aims to identify and understand opinions directed towards specific aspects within a text (Diaz et al., 2020) related to an entity (Tan et al., 2020). The main research areas of ABSA focusing on various layers of sentiment components include: aspect terms (a), aspect categories (c), opinion expressions (o), and sentiment polarity (p) (Mao et al., 2024; Xu et al., 2020). Different layers of sentiment components are shown in Table 1.

In other words, ABSA is divided into several subtasks including aspect extraction (Liu, 2012; Scaria et al., 2023), sentiment classification (Liu, 2012), aspect category identification (Cai et al., 2020), and opinion term extraction (Zhang et al., 2023). Each of these subtasks contributes to a more detailed interpretation and comprehensive understanding of sentiment expression (Nevedistin et al., 2025).

Table 1. Input and output examples for different ABSA tasks. Adapted from Mao et al. (2024, p. 9)

Task	Input	Example Input*	Output	Example Output
Aspect Term Extraction (ATE)	*s*	sentence	*a*	salads,server
Opinion Term Extraction (OTE)	*s*	sentence	*o*	fantastic, unfriendly
Aspect Category Detection (ACD)	*s*	sentence	*c*	food, service
Aspect Sentiment Classification (ASC)	*s,a1* *s,a2*	sentence, salads sentence, server	*p1* *p2*	POS NEG
Aspect-Opinion Pair Extraction (AOPE),	*s*	sentence	*(a,o)*	(salads, fantastic), (server, unfriendly)
Aspect Category Sentiment Analysis (ACSA)	*s*	sentence	*(a,c)*	(salads, food), (server, service)
End-to-End ABSA (E2E-ABSA)	*s*	sentence	*(a,p)*	(salads, POS), (server, NEG)
Aspect Sentiment Triplet Extraction (ASTE)	*s*	sentence	*(a,o,p)*	(salads, fantastic, POS), (server, unfriendly, NEG)
Target Aspect Sentiment Detection (TASD)	*s*	sentence	*(a,c,p)*	(salads, food, POS), (server, service, NEG)
Aspect-Category-Opinion-Sentiment (ACOS) Quadruple Extraction	*s*	sentence	*(a,c,o, p)*	(salads, fantastic, food, POS), (server, unfriendly, service, NEG)

As illustrated in Figure 12, ABSA tasks can be categorized into either single or compound based on the ABSA system output. A single ABSA task predicts a single component (e.g., a, c, o, or p). On the other hand, a compound ABSA task predicts multiple components (Mao et al., 2024; Zhang et al., 2022). For instance, "Opinion Term Extraction" (OTE) is a single task that is implemented to identify opinion expression "o" from input textual data. In contrast, "Aspect-Opinion Pair Extraction" (AOPE) is a compound task that aims to identify aspect-opinion pairs from the text (Mao et al., 2024, p. 8).

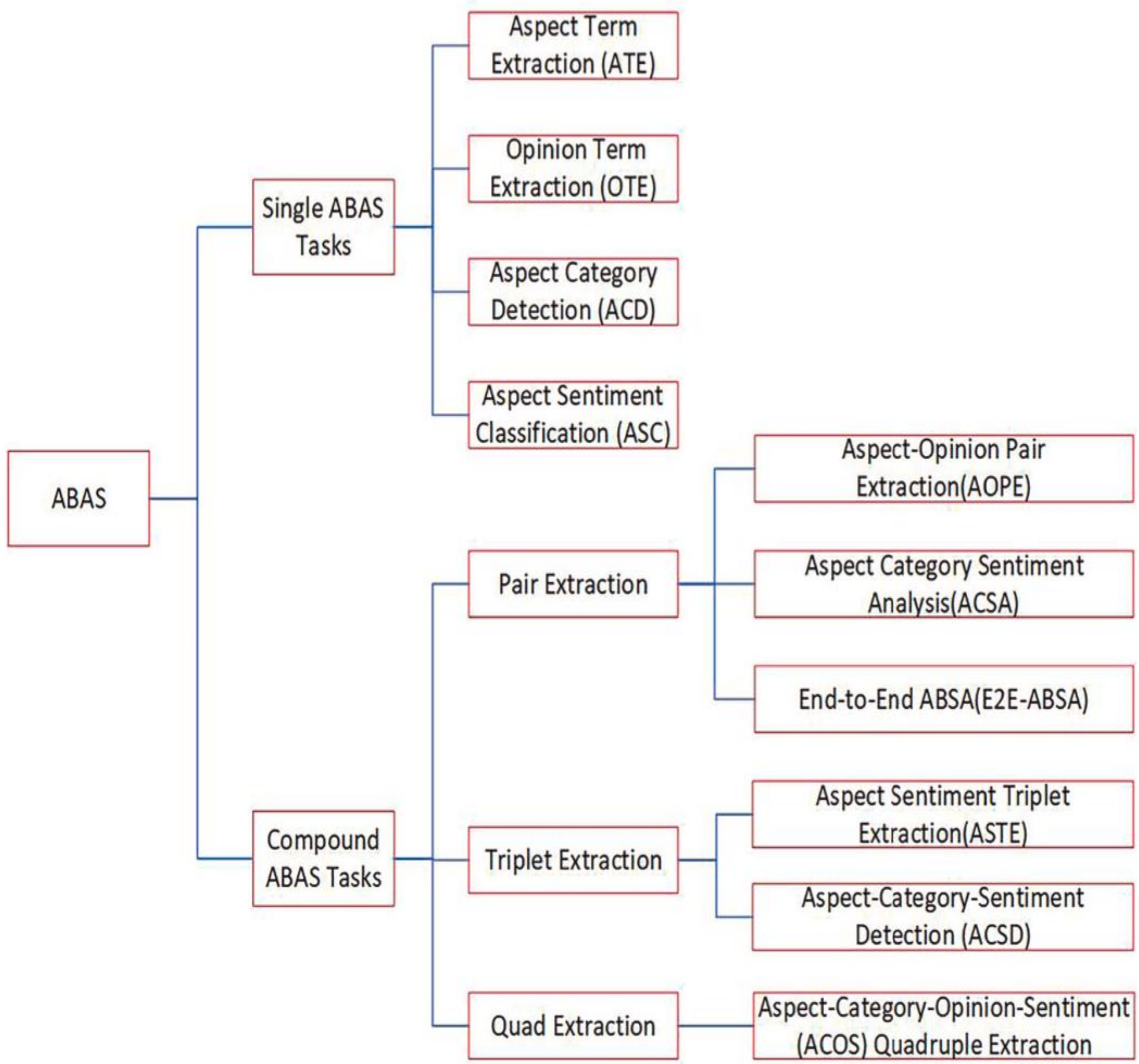

Figure 12. Aspect-Based Sentiment Analysis tasks adapted from Mao et al. (2024, p.8)

Traditional approaches to ABSA used architectures such as bidirectional encoders (Dos Santos et al., 2021; Zhang et al., 2023), RNNs (Xu et al., 2020), graph networks (Zhou et al., 2020; Wang et al., 2024b), sequence-to-sequence models (Ma et al., 2019), and ensemble methods (Yang et al., 2023).

To enhance performance metrics of accuracy, recall, and precision of ABSA tasks, new innovative techniques such as context denoising (Tian et al., 2024), abstract meaning representations (Ma et al., 2023) and global semantic feature integration (Zhou et al., 2024) have been proposed.

In addition to the mentioned techniques for implementing sentiment analysis, more recent studies have employed Large Language Models (LLMs). These models use "in-context learning" (ICL) and lightweight fine-tuning methods to optimize a minimal set of parameters from large pre-trained models. For instance, Low-Rank Adaptation (LoRA), in combination with techniques for decreasing their memory usage (quantization) (Dettmers et al., 2023; Hu et al., 2022), is employed to facilitate the implementation of complex ABSA tasks. These methods allow for the detection of implicit aspects that are not directly mentioned in the text but are understood from context (Nevedistin et al., 2025).

2.5 Explainable Artificial Intelligence

Deep learning has demonstrated remarkable results in various fields, such as NLP tasks (Otter et al., 2021). However, these models are like black boxes due to their complex architecture, which can be difficult for humans to understand (Shams Khoozani et al., 2024; Petch et al., 2022; Buhrmester et al., 2021; Azodi et al., 2020). This has caused concerns among regulators and stakeholders (Tan & Kok, 2024). The lack of transparency in block-box models leads to distrust, particularly when they are used for important decisions that have a strong effect on individuals and business operations (Ribeiro et al., 2016). To solve this issue, Explainable Artificial Intelligence (XAI) emerged to enhance the transparency and interpretability of intricate AI systems (Angelov et al., 2021; Bilgilioğlu et al., 2025). XAI is a critical framework within AI that offers more transparent explanations for reasoning behind AI-generated decisions and outcomes. Unlike black-box deep learning models, XAI techniques help stakeholders understand decision-making process of complex models (Samek & Müller, 2019). This fosters trust among stakeholders (Miller, 2019) and promotes ethical AI practices (Hosain et al., 2024). Figure 13 illustrates the workflow of XAI in deep learning.

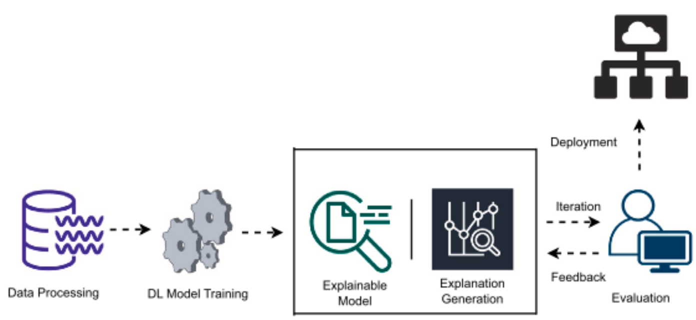

Figure 13. Overview of XAI workflow in deep learning. Adapted from Hosain et al. (2024)

2.5.1 XAI for Deep Learning

XAI approaches, which aim to make deep learning models more transparent and interpretable (Bilgilioğlu et al., 2025), use different methods. Figure 14 illustrates these techniques, which are categorized based on their methodological functions. Among these, three techniques are applied in this study—LIME, SHAP, and sequence and text understanding—which will be described in the following subsections.

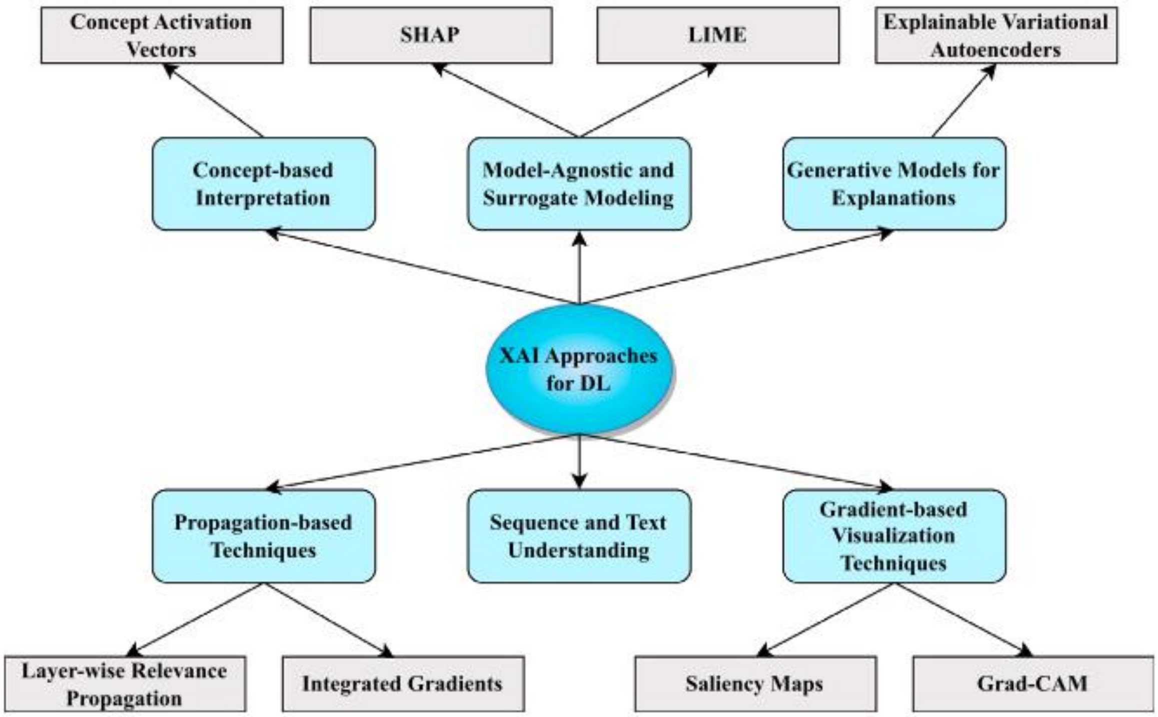

Figure 14. XAI techniques used in deep learning. Adapted from Hosain et al. (2024)

2.5.1.1 Model-agnostic and Surrogate Modeling

Model-agnostic techniques and surrogate modeling are approaches used within deep learning to improve their interpretability. Model agnostic methods explain black-box models without the need to access the internal architecture of the model (Jiarpakdee et al., 2020). They produce local explanations and faithful interpretations that reflect the original model's output. Surrogate modeling, on the other hand, builds a simpler and more interpretable model that behaves like the complex original model, improving the transparency of the decision process (Mariotti et al., 2023).

2.5.1.1.1 SHAP (SHapley Additive exPlanations)

SHAP is an explainable AI technique based on cooperative game theory, which explains how each feature contributes to a model's prediction. It assigns Shapely values to each feature to calculate its contribution to the model's decision (Antwarg et al., 2021). SHAP has been applied in different fields, such as financial services (Nguyen et al., 2023a), healthcare (Guleria et al., 2023), and image processing and recognition (Walia et al., 2022) to foster trust in the decisions of black-box models by determining the importance of each feature in the model's prediction. SHAP values can be mathematically calculated using the following equation (Hosain et al., 2024), where *S* represents a subset of all input features *(N)* excluding *i*. *f(S)* is the model's output when only the features in the subset S are present. *f(S ∪ {i})* indicates the prediction when feature *i* is added to that subset *(S)*.

$$\phi_i(f) = \sum_{S \subseteq N\{i\}} \frac{|S|!\,(|N| - |S| - 1)!}{|N|!} [f(S \cup \{i\}) - f(S)]$$

2.5.1.1.2 LIME (Local Interpretable Model-agnostic Explanations)

LIME (Local Interpretable Model-agnostic Explanations) helps end-users understand the prediction of complex deep learning models on a local level. In other words, it provides explanations for only a single data instance or prediction at a time. The process begins by selecting a specific data instance and generating a series of perturbed (slightly modified) versions of it. These variations are then provided to the original complex model to obtain corresponding predictions. By analyzing how these small modifications influence the original model's predictions, Lime fits a simpler and interpretable model (e.g., decision trees) that stimulates the complex model's behavior in the local region. That is to say, the simpler model only makes accurate predictions for the modified data instances. The surrogate model helps end-users understand which features (pieces of information) contributed most to the original model's prediction for that specific instance (Hosain et al., 2024; Palatnik de Sousa et al., 2019).

LIME is widely used in different fields such as image processing and recognition (Zafar & Khan, 2019), NLP (Luo et al., 2024), and healthcare (Kumarakulasinghe et al., 2020). It enhances interpretability by providing explanations for individual predictions or outputs of the model. The method can be mathematically expressed as shown in the equation below (Hosain et al., 2024).

$$\hat{f}(x) = arg \min_{g \in G} L(f, g, \pi_x) + \Omega(g)$$

2.5.1.2 Sequence and Text Understanding

Sequence and text understanding in the context of explainable AI are techniques that aim to increase the interpretability of models developed for analyzing sequential data, particularly textual inputs. These models typically employ attention mechanisms, RNNs, or transformer architecture to capture contextual relationships within the sequence (Chen et al., 2021). These approaches are used to identify key tokens influential in the model's decisions. This enhances the transparency of the model's predictions (Amjad et al., 2023; Hosain et al., 2024).

2.6 SMOTE

Many NLP tasks, such as classification and prediction have significantly improved thanks to deep learning models. But the model's architecture is not the only factor that influences its performance. Other important factors include quality, structure, and balance of data. In particular, imbalanced data poses a considerable challenge to the performance of classifiers (Pradipta et al., 2021). Dealing with imbalanced data and how to improve model performance on such datasets with skewed class distributions has been one of the challenges of machine and deep learning models (Branco et al., 2016; Cieslak et al., 2012; LemaÃŽtre et al., 2017; Khan et al., 2018; Krawczuk, 2016). An imbalance dataset means that the number of instances belonging to one class is much smaller than the number of instances of the other classes. The less frequent class is called minority (or positive class), while the more frequent class is called majority (or negative) class (Pradipta et al., 2021). Since machine learning models typically demonstrate higher specificity or local accuracy on the majority class, compared to the minority class, the issue of imbalanced data has received considerable research attention (Fernández, 2018). Research into imbalanced data is gaining increasing attention as deep learning applications expand into real-world domains such as face recognition, social media analytics, and medical diagnostics. These advancements introduce new challenges (Fernández, 2018; Haixiang et al., 2017; Bach et al., 2017). In machine learning projects analyzing imbalance data, a key challenge is to improve prediction for the minority class while minimizing false positives (avoiding incorrect classification of the majority class). Common strategies to address this issue include manipulating the dataset using sampling techniques—namely, undersampling and oversampling—using specialized learning algorithms, or as shown by Li et al. (2021), optimizing the loss function to allocate more weight to the minority class. However, these techniques come with certain challenges. For example, undersampling might remove useful information from majority class which can adversely affect the model's performance (Haixiang et al., 2017). On the other hand, oversampling might cause model to overfit and not perform well on unseen data (Mujahid et al., 2024). To address these challenges, several methods have been introduced. One of the most widely used techniques is SMOTE (Synthetic Minority Oversampling Techniques), which was proposed by Chawla et al. in 2002. SMOTE avoids the issue of overfitting, which arises when the data from the minority class is simply replicated. It creates synthetic samples within the local neighborhood of existing minority class instances. By focusing on feature space, SMOTE improves the classifier's capacity to generalize by operating within the "feature space" and considering relationships between features rather than viewing each data point separately (Fernández et al., 2018, p. 866). Several researchers have studied the application of SMOTE in sentiment analysis tasks. For instance, Rozi et al., applied data imbalance handling techniques, including SMOTE, in an ABSA task focused on radio station reviews. Their findings revealed that SMOTE significantly improved model performance for minority aspect categories, particularly enhanced recall and F1-score. In another study, Saputra & Setianwan (2023) combined TF-IDF, FastText, and SMOTE with an RNN model to perform ABSA on Twitter data. Based on the study's findings, the mentioned combination significantly improved the performance of the ABSA.

2.7 BERTopic for Topic Modeling

The efficient extraction of features from large text datasets has led to the development of numerous text mining techniques (Li et al., 2019). Topic modeling is among the most used methods (Hong & Davidson, 2010). A topic model is a statistical modeling approach in machine learning and NLP that aims to identify underlying topical patterns within a collection of documents (Guo et al., 2017).

BERTopic is a modern topic modeling technique that uses developments in contextual embedding techniques and clustering methods. The process begins by creating document embeddings using a pre-trained language model, followed by reducing the data dimensions to enhance clustering performance. Finally, documents that share similar content are grouped, and a class-based variation of TF-IDF (term frequency—inverse document frequency) is used to find the main topics in each group, improving topic coherence (Grootendorst, 2022).

2.8 Evaluation Metrics

Assessing model's performance is one of the most important steps in building machine learning systems. Selecting appropriate evaluation metrics help researchers assess the robustness and reliability of their models (Rainio et al., 2024). Evaluation metrics can be broadly classified into two groups: those that function independently of the probability threshold, such as ROC-AUC and precision-recall curves (Hernández-Orallo et al., 2012; Powers, 2020), and those that depend on the threshold, including accuracy and confusion matrix-based measures (Powers, 2020). The most used evaluation metrics include accuracy, precision, recall, and F1 score.

Accuracy calculates the ratio of correctly predicted cases to the total number of cases. While being useful, it can be misleading in the context of imbalance datasets (Kok et al., 2024; Owusu-Adjei et al., 2023). In such scenarios, a model that always predicts the majority class may achieve high accuracy despite struggling to predict the minority class.

Other evaluation metrics, like precision, recall, and F1 score, are utilized to address this challenge. Precision, recall, and F1 score, which balance both measures, help identify the types of errors the model makes and assess how well the model avoids false positives and false negatives (Kasana & Rathore, 2024; Powers, 2020).

Precision is the ratio of correctly identified positive instances divided by all instances that were predicted as positive. It is crucial in scenarios where false positives (false alarms) are costly (Powers, 2020).

$$Precision = \frac{True\ Positive}{True\ Positive\ +\ False\ Positive}$$

Recall on the other hand, shows how many actual positive cases the model correctly identifies. It is defined as the proportion of true positive instances divided by the total number of actual positives (Zhou et al., 2025). This metric is particularly important in fields like medical diagnosis, where missing a positive case or not detecting a disease (false negative) leads to serious problems (Piao et al., 2015).

$$Recall = \frac{True\ Positive}{True\ Positive\ +\ False\ Negative}$$

The F1 score uses precision and recall to compute their harmonic mean. It offers a balanced metric for evaluating a model's effectiveness when both error types are important. It is particularly relevant in applications where false positives and false negatives have unequal consequences, such as in medical diagnosis, where false negative may lead to severe consequences, or in financial fraud detection, where too many false positives (false alarms) can lead to customer dissatisfaction (Owusu-Adjei et al., 2023; Qiu et al., 2024).

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

3. Methods

This chapter provides an overview of the data, the model development process, text analysis methodologies, and the evaluation strategies used to interpret the findings. Additionally, the study outlines the use of three explainability techniques, LIME, SHAP, and Attention Visualization, to provide insights into the process by which the model makes decisions and predictions. The chapter concludes by exploring the use of BERTopic for post-deployment topic modeling, which was applied to negatively classified tweets to uncover topic shifts over time.

3.1 Data Collection and Characteristics

A total of 252,410 tweets mentioning McDonald's were collected using a Google Chrome extension developed by the University of Southhampton (Lancaster, n.d.). The data was gathered across two distinct periods: from October 2023 to May 2024 (Year 1), and from October 2024 to May 2025 (Year 2), with approximately 125,000 tweets collected for each period. To ensure a rounded and consistent sample size for comparative analysis, a subset of 100,000 tweets was randomly chosen from each year's set. This sampling strategy was employed to avoid bias caused by the unequal number of tweets collected in each period and to ensure a fair analysis of sentiment trends across predefined aspect categories and broad thematic categories over the two-year period.

3.1.1 Labelled Data

The performance of machine learning, and especially deep learning, is heavily dependent on high-quality labeled datasets (Emam et al., 2021). For the purpose of this study, 1000 tweets were randomly selected and manually annotated by two independent annotators to assign sentiment polarity and classify each instance into a fine-grained aspect category and a corresponding broader thematic category. The annotation process consisted of three sequential stages: (1) specification of topical regions within each tweet; (2) category assignment, where each region was labeled with an aspect-level category and its associated broader thematic category, according to established criteria (as illustrated in table 2); and (3) sentiment polarity annotation. The increase in the final annotated dataset is due to segmenting some tweets into multiple independently labeled regions.

3.1.1.1 Business-centric Logic for Category Selection

The aspect and broad thematic categories used for data annotation in this study were adapted from the categories proposed in Golbazi (2024). These categories were selected due to their connection with fundamental strategic elements within the restaurant industry. For example, Baumann et al. (2019) highlights that competitive productivity (CP) at the meso-level is determined by customer evaluations of a company's performance across key competitive areas such as brand management, corporate culture, and resource management. These areas are reflected in broad thematic categories like 'Brand Perception and Loyalty', 'Corporate and Social Responsibility', and 'Core Restaurant Experience'. In addition, Grant (1991) highlights the role of intangible assets, such as brand equity, organizational values, and internal capability, in establishing strategic advantages. This reinforces inclusion of broad thematic categories such as 'Brand Perception and Loyalty', 'Corporate and Social Responsibility', and

‘Promotions and Marketing’. These categories not only capture consumer-facing issues but also reflect broader organizational priorities essential for achieving long-term competitive advantage.

Table 2. Overview of Aspect Categories and Their Corresponding High-Level Themes for Aspect-Category Sentiment Analysis. Adapted from Golbazi (2024)

No.	Broad Thematic Category	Aspect Category
1	Core Restaurant Experience	Food
		Customer Service
		Products
2	Brand Perception and Loyalty	Brand Perception
		Loyalty
		Brand Competition
3	Corporate and Social Responsibility	Ethical Responsibility
		Public Health Impact
4	Value for Money	Price
5	Promotions and Marketing	Marketing Strategy
		Sponsorships and Events
		Promotions
6	Other	General

3.1.1.2 Illustrative Examples of annotated Tweets

Tables 3 to 9 illustrate selected tweet examples labeled according to their aspect category, broad thematic category, and corresponding sentiment polarity. In example 3, the tweet “Oh my god I want mcdonalds in a dangerous way” is labeled under the broad thematic category ‘Brand Perception and Loyalty’ with the aspect ‘Loyalty’ and assigned a positive sentiment polarity.

Table 3.Sample Annotated Tweet – Example 1

Example 1			
Mcdonalds cherry piea. Estonia crazy for that			
Region	**Category**	**Aspect**	**Classification**
[{"start":0,"end":45,"text":"mcdonalds cherry piea. estonia crazy for that"}]	Core Restaurant Experience	Food	Positive

Table 4. Sample Annotated Tweet – Example 2

Example 2			
$2.39 for a hash brown at McDonalds is wicked			
Region	**Category**	**Aspect**	**Classification**
[{"start":0,"end":46,"text":"$2.39 for a hash brown at McDonaldas is wicked"}]	Value for Money	Price	Negative

Table 5. Sample Annotated Tweet – Example 3

Example 3			
Oh my god I want mcdonalds in a dangerous way			
Region	**Category**	**Aspect**	**Classification**
[{"start":0,"end":46,"text":"oh my god i want mcdonalds in a dangerous way"}]	Brand Perception and Loyalty	Loyalty	Positive

Table 6. Sample Annotated Tweet – Example 4

Example 4			
I am craving McDonalds ion even eat that s**t			
Region	**Category**	**Aspect**	**Classification**
[{"start":0,"end":21,"text":"I am craving McDonalds"}]	Core Restaurant Experience	Food	Positive
[{"start":21,"end":44,"text":" ion even eat that s**t"}]	Brand Perception and Loyalty	Brand Perception	Negative

Table 7. Sample Annotated Tweet – Example 5

Example 5			
I still canat believe hashbrowns are like $3 something at McDonaldas. Iam about to go to Aldi and get that big pack. Iykyk			
Region	**Category**	**Aspect**	**Classification**
[{"start":0,"end":69,"text":"I still canat believe hashbrowns are like $3 something at McDonaldads."}]	Value for Money	Price	Negative
[{"start":69,"end":122,"text":" I am about to go to Aldi and get that big pack. Iykyk","labels"}]	Brand Perception and Loyalty	Brand Competition	Negative

Table 8. Sample Annotated Tweet – Example 6

Example 6			
i always want mcdonalds			
Region	**Category**	**Aspect**	**Classification**
[{"start":0,"end":24,"text":"i always want mcdonalds"}]	Brand Perception and Loyalty	Loyalty	Positive

Table 9. Sample Annotated Tweet – Example 7

Example 7			
I be wanting a Chick-fil-A sandwich with McDonalds fries			
Region	**Category**	**Aspect**	**Classification**
[{"start":0,"end":56,"text":"I be wanting a Chick-fil-A sandwich with McDonalds fries"}]	Core Restaurant Experience	Products	Positive

3.1.1.3 Overview of Labeled Tweet Distribution

Figures 15-17 illustrate the distribution of tweets in the labeled dataset based on sentiment, broad thematic category, and aspect category.

As shown in Figure 15, negative sentiment was most prevalent among the tweets (41%), while neutral (37%) and positive (22%) were observed less frequently.

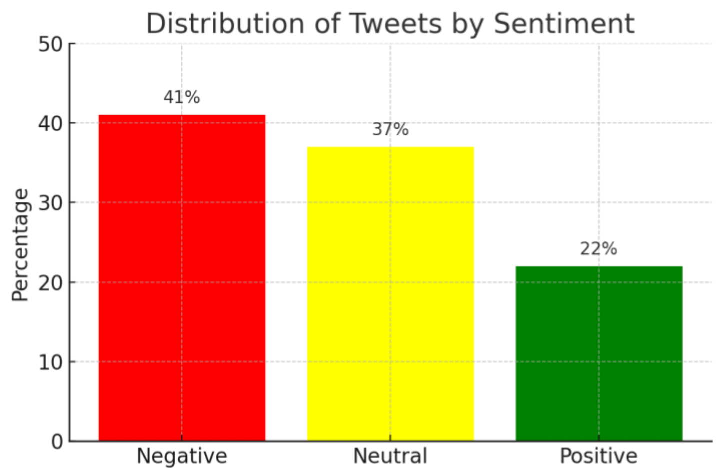

Figure 15. Distribution of Labeled Tweets by Sentiment

Figure 16 represents the distribution of tweets across the broad thematic categories. 'Core Restaurant Experience' is the most represented category, whereas 'Value for Money' and 'Promotions and Marketing' are the least represented.

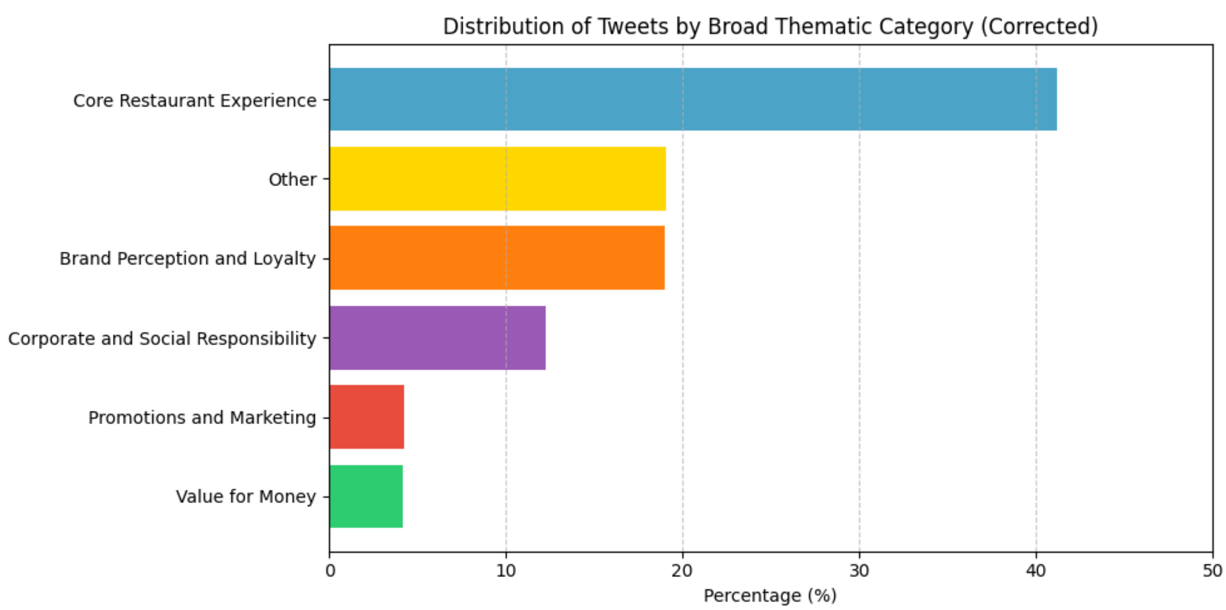

Figure 16. Distribution of Labeled Tweets by Broad Thematic Category

As shown in Figure 17, most tweets were related to 'General', 'Products', 'Customer Service', and 'Ethical Responsibility' aspect categories, while 'Sponsorship and Events', 'Public Health Impact', and 'Promotions' aspect categories attracted the least attention.

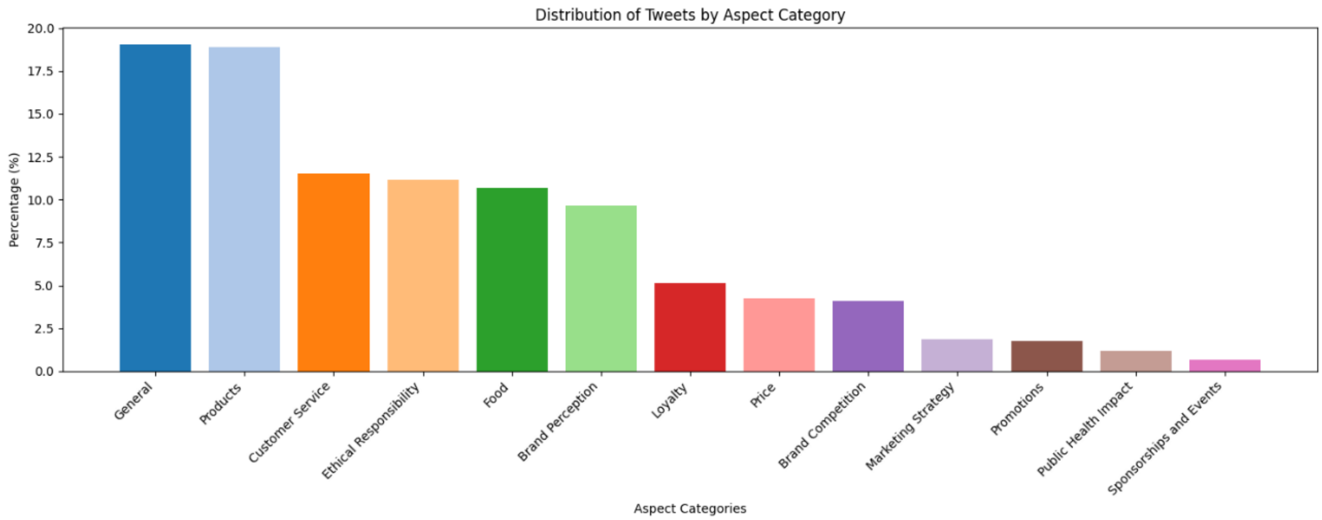

Figure 17. Distribution of Labeled Tweets by Aspect Category

3.1.1.4 Assessment of Inter-Annotator Agreement (IAA)

Inter-Annotator Agreement (IAA) was evaluated using Cohen's Kappa score. The results show a strong alignment between annotators across sentiment (K=0.89), broad thematic category (K=0.84), and aspect category (K=0.76) labeling tasks.

3.2 Model Selection

This study introduces an enhanced BERT-based classification framework that combines pre-trained BERT embeddings with BiLSTM and Transformer encoder layers. In the BBT model (BERT—BiLSTM—Transformer), the BiLSTM layers are utilized to capture sequential relationships, while the Transformer encoders model detailed inter-token relationships.

3.2.1 Preprocessing Pipeline and SMOTE Implementation

The initial dataset was imported from a CSV file, where sentiment, broad thematic category, and aspect category labels were converted into integers based on predefined mapping.

3.2.1.1 Tokenization and Normalization Steps

To preprocess the text data before tokenization, the input was validated, standardized, and cleaned to prevent processing errors during tokenization. The preprocessed text was then tokenized using the BertTokenizerFast from the Hugging Face Transformers library (Wolf et al., 2020). This automatically lowercases inputs, applies WordPiece tokenization for rare and out-of-vocabulary tokens, and ensures consistent sequence length via padding and truncation.

3.2.1.2 Addressing Class Imbalance Using SMOTE

The Synthetic Minority Over-sampling (SMOTE) was utilized to handle class imbalance by creating synthetic instances for underrepresented classes. Before its application, the dataset was split into a training set (80%) and a test set (20%). This is done to ensure SMOTE is exclusively applied to the training data. SMOTE was applied separately to each classification task (sentiment, aspect category, broad thematic category) in a stepwise and task-specific manner. Each SMOTE process oversampled the training data for one of the three target labels to match the size of the largest class within that label type (e.g., 205 for each aspect category, 443 for each broad thematic category, and 436 for each sentiment label), enabling balanced model training. Because SMOTE operates on the features and a single label type at a time, the same tweet could contribute to synthetic samples in multiple balancing steps. As a result, three task specific balanced datasets were generated, rather than a unified dataset containing all synthetic instances. A custom function was developed to assess and compare the number of instances for each label before and after applying SMOTE. The comparative results are shown in Tables 10-12.

Table 10. Instances added by SMOTE - Aspect Category

Aspect Category Label	Before SMOTE	After SMOTE	Added By SMOTE	Percentage Increase
General	205	205	0	0.00
Brand Perception	104	205	101	97.12
Food	115	205	90	78.26
Brand Competition	44	205	161	365.91
Ethical Responsibility	120	205	85	70.83
Price	46	205	159	345.65
Customer Service	124	205	81	65.32
Products	203	205	2	0.99
Loyalty	55	205	150	272.73
Public Health Impact	13	205	192	1476.92
Marketing Strategy	20	205	185	925.00
Sponsorships and Events	7	205	198	2828.57
Promotions	19	205	186	978.95

Table 11. Instances added by SMOTE - Broad Thematic Category

Broad Thematic Category Label	Before SMOTE	After SMOTE	Added by SMOTE	Percentage Increase
Core Restaurant Experience	443	443	0	0.00
Other	205	443	238	116.10
Brand Perception and Loyalty	204	443	239	117.16
Corporate and Social Responsibility	132	443	311	235.61
Promotions and Marketing	46	443	397	863.04
Value for Money	45	443	398	884.44

Table 12. Instances added by SMOTE - Sentiment

Sentiment Label	Before SMOTE	After SMOTE	Added by SMOTE	Percentage Increase
Negative	242	436	194	80.17
Neutral	436	436	0	0.00
Positive	397	436	39	9.82

3.2.2 Contextual Representation Enhancement with BiLSTM and Transformer Encoder

After BERT creates embeddings, the model first applies BiLSTM and then a Transformer encoder to improve them. By processing sequences bidirectionally, the BiLSTM layer gathers contextual information from past and future tokens, facilitating a deeper understanding of the input text.

After BiLSTM layer, a Transformer encoder is used to enhance the contextual representations further. Multi-head self-attention allows the Transformer to analyze different parts of the input simultaneously to understand complex inter-token relationships.

This two-stage contextualization process enhances the learned representations and consequently improves the model's performance in accurate multi-task classification. An overview of the complete model architecture is provided in Table 13.

Table 13. Overview of the BBT Model Architecture

Layer (Type)	Output Shape	Description
Input Layer	(batch_size, seq_len)	Tokenized input tweets
BERT (Pretrained, Frozen)	(batch_size, seq_len, 768)	Generates contextualized token embeddings
BiLSTM Layer	(batch_size, seq_len, 256)	Captures bidirectional sequence context (128 units in each direction)
Transformer Encoder Layer	(batch_size, seq_len, 256)	Applies self-attention to refine contextual relationships
Global Average Pooling	(batch_size, 256)	Aggregates token-level outputs into a fixed-length vector
Dropout Layer (p=0.3)	(batch_size, 256)	Prevents overfitting during training
Dense Layer—Sentiment	(batch_size, 3)	Predicts sentiment polarity (positive, neutral, negative)
Dense Layer—Aspect Category	(batch_size, 14)	Predicts one of 14 fine-grained aspect categories
Dense Layer—Broad Thematic Category	(batch_size, 6)	Predicts one of 6 broad thematic categories

3.2.3 Final Representation and Multi-Task Output Layers

After obtaining the hidden state of the [CLS] token from the final Transformer encoder, the model sends this representation into three parallel classification layers for the prediction of the aspect category, broad thematic category, and sentiment polarity. Each classification layer first transforms the input using a linear function and then uses a softmax activation function to produce class probability distributions.

3.2.4 Training Workflow and Performance Evaluation

The model was trained over eight epochs, utilizing AdamW optimizer at a learning rate of 1e-5 and a weight decay of 0.01 to mitigate overfitting. A linear learning rate scheduler with warm-up was implemented to optimize the learning rate during training. Within each training batch of 16 samples, and the total loss was computed by adding up the cross-entropy losses from the three tasks: aspect category classification, broad thematic category classification, and sentiment classification. Gradient clipping (limit = 1.0) was applied to ensure training stability. The model's training performance was monitored by calculating the average loss for each epoch.

Two versions of the BBT model were trained to investigate the influence of SMOTE on model performance, one with SMOTE-augmented data and another with the original imbalanced dataset. To ensure a direct performance comparison, both models shared exactly the same architecture and hyperparameters. Additionally, separate training processes were implemented for three different architectures: BERT only, BERT combined with BiLSTM, BERT combined with BiLSTM and Transformer layers. Following an assessment of the evaluation metrics, the BBT model trained on SMOTE-augmented data was selected for further analysis and final application.

After training, the model's performance was assessed using F1-score for each classification task. Finally, visual tools were used to facilitate the interpretation of the model's performance.

3.2.5 Computational Environment and Implementation Platform

All model development, training, and evaluation tasks were conducted on Google Colab pro+, specifically utilizing a runtime environment powered by an NVIDIA A100 Tensor Core GPU for accelerated cloud-based computation. The implementation was done in Python 3, with PyTorch serving as the main deep learning framework. The Hugging Face Transformers library was employed to manage BERT-based models and perform tokenization.

3.2.6 Explainability and Interpretability Methods

Despite achieving significant performance results, deep learning models, such as our BBT model, inherently behave as a black box system that make it difficult to understand the rationale behind their outputs and predictions. Although relatively lightweight in size and training time, the BBT model still shows this typical lack of transparency.

To deal with this issue, explainability techniques were used in our study. After training and testing the model, we used three methods, LIME, SHAP and Attention Visualization, to better understand how it makes decisions.

3.3 Application of the Trained Model to New Datasets

After training and evaluation, the saved BBT model is utilized to analyze two separate datasets of tweets concerning McDonald's collected from two distinct timeframes. This approach enables a comparative analysis of changes in customer sentiment across aspect categories and broader thematic categories over time. Subsequently, the predicted sentiments are systematically recorded to visualize emerging trends and pinpoint recurring problems.

3.4 Post-Deployment Topic Modeling with BERTopic

After classifying 200,000 tweets using the fine-tuned BERT-BiLSTM-Transformer model, the negatively classified tweets were further analyzed to identify underlying topics. To achieve this, BERTopic method was used. BERTopic was applied separately to tweets from each of the two timeframes, allowing for comparison of topic shifts over time. Additionally, it was used to identify specific keywords that contributed to the negative sentiment expressed in the tweets, supporting a more nuanced understanding of the model's output.

4. Results

This chapter outlines the evaluation results of the proposed BBT model and its application to 200,000 tweets sampled from two distinct time periods. It begins with a comparative evaluation of three models: the baseline model (BERT architecture), a BERT-based model with BiLSTM layers, and the proposed BBT model. All models were trained on the manually annotated dataset. After model evaluation, the best performing model was applied to 200,000 tweets (randomly and equally sampled from each period) to perform large-scale ACSA, demonstrating the effectiveness of the proposed model. The chapter also includes the results of explainability techniques (LIME, SHAP, and attention visualization) as well as post-deployment topic modeling using BERTopic.

4.1 Performance Analysis of the Proposed Model Architecture

Table 14 summarizes the evaluation results of the BERT-only model, BERT-based model with BiLSTM layers, and the proposed BBT architecture. All models were trained on the SMOTE-augmented training set and evaluated on the 20% test data, on which SMOTE was not applied, in order to prevent data leakage. Overall, all three models performed well across the three classification tasks. However, the BBT model exhibited the most balanced and consistently strong performance across all three tasks. It achieved the highest score for aspect category classification (0.78), which is typically considered the most challenging task in ABSA frameworks. For broad thematic category classification, both the BBT model and the BERT-based model with BiLSTM layers achieved comparable results, with F1-scores close to 0.85. In sentiment classification, BERT-only model performed best (0.89); however, the BBT model followed closely (0.86) while maintaining a more balanced and consistently strong performance across all tasks.

Table 14. Model Performance Comparison with SMOTE

Task	BERT-only Model	BERT-based Model with BiLSTM Layers	BBT Model
Aspect Category F1-Score	0.71	0.68	0.78
Broad Thematic Category F1-Score	0.84	0.85	0.85
Sentiment F1-Score	0.89	0.83	0.86

Tables 15-17 present a comparative analysis of F1-scores for the proposed hybrid model, with and without SMOTE, across aspect category, broad thematic category, and sentiment classification tasks. The results clearly indicate that integrating SMOTE leads to significant performance improvements across all labels.

As shown in Table 15, using SMOTE proved effective in improving the model's classification performance across all aspect categories. For example, the F1-scores for 'Sponsorship and Events' and 'Brand Perception' increased from 0.29 to 0.92 and from 0.07 to 0.69, respectively. Similarly, 'Customer Service' and 'Marketing Strategy' showed considerable performance gains following the application of SMOTE.

Table 15. Impact of SMOTE on Aspect Category Classification Task

Aspect Category	SMOTE	No SMOTE
General	0.88	0.53
Brand Perception	0.69	0.07
Food	0.68	0.37
Brand Competition	0.76	0.68
Ethical Responsibility	0.88	0.85
Price	0.81	0.47
Customer Service	0.81	0.23
Products	0.83	0.61
Loyalty	0.57	0.43
Public Health Impact	0.85	0.64
Marketing Strategy	0.71	0.23
Sponsorship and Events	0.92	0.29
Promotions	0.81	0.46

As shown in Table 16, applying SMOTE positively impacted the model's classification performance across all broad thematic categories. Most significantly, the F1-score for 'Brand Perception and Loyalty', 'Other', and 'Value for Money' improved from 0.24 to 0.79, 0.43 to 0.90, and 0.60 to 0.82, respectively.

Table 16. Impact of SMOTE on Broad Thematic Category Classification Task

Broad Thematic Category	SMOTE	No SMOTE
Core Restaurant Experience	0.86	0.71
Other	0.90	0.43
Brand Perception and Loyalty	0.79	0.24
Corporate and Social Responsibility	0.88	0.76
Promotions and Marketing	0.83	0.73
Value for Money	0.82	0.60

The results shown in Table 17 suggest that, besides enhancing the model's classification performance across aspect categories and broad thematic categories, the application of SMOTE significantly improved its ability to classify sentiment polarity across the three sentiment classes: positive, negative, and neutral. Notably, the F1-score for 'Negative' sentiment increased from 0.64 to 0.90. 'Positive' and 'Neutral' sentiments also showed significant enhancements, improving from 0.57 to 0.78 and 0.65 to 0.87, respectively.

Table 17. Impact of SMOTE on Sentiment Classification Task

Sentiment	SMOTE	NO SMOTE
Positive	0.78	0.57
Negative	0.90	0.64
Neutral	0.87	0.65

4.2 Exploring Model Interpretability through Explainability Techniques

In this section, we present example outputs from three post-hoc explainability techniques used in this study: LIME, SHAP, and attention visualization. As shown in the Figures 18-22, these methods provide insights into how the hybrid model assigns sentiment polarities at the aspect level. For instance, as illustrated by the LIME explanation in Figure 18, the model identifies 'Weird' as the highest-weighted word contributing to the negative sentiment prediction, followed by 'McDonalds'. This suggests that even a neutral brand mention can be interpreted negatively depending on the surrounding context. In another LIME visualization (Figure 19), the word 'want', carries the highest positive weight, influencing the model's decision to assign a positive sentiment to the tweet.

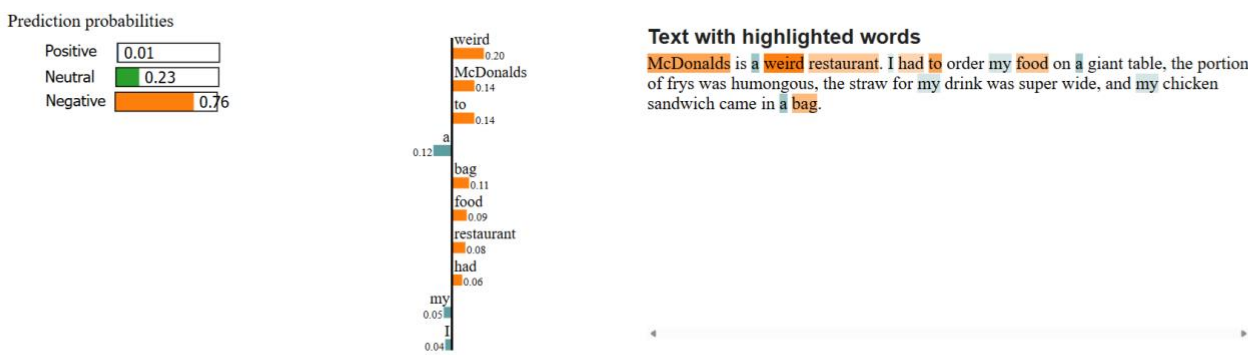

Figure 18. LIME Explanation for Sentiment Prediction – Example 1

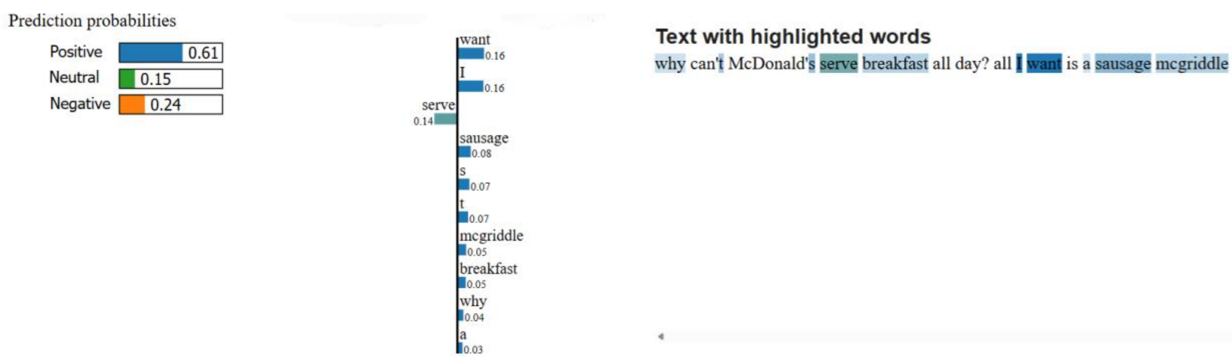

Figure 19. LIME Explanation for Sentiment Prediction – Example 2

Figure 20 illustrates a SHAP explanation where the word 'dislike' has the highest impact on a negative sentiment toward fast food, while terms like 'breakfast' and 'pass' contribute to a slightly positive tone. Figure 21 presents another SHAP explanation for a tweet classified as positive. The model assigns strong positive weight to the word 'Love' and slightly positive weight to 'getting', 'gestures', and 'McDonald'. These examples indicate that the model successfully identifies sentiment-relevant tokens and accurately classifies sentiment associated with specific aspects.

Figure 20. SHAP Visualization of Feature Impact – Example 1

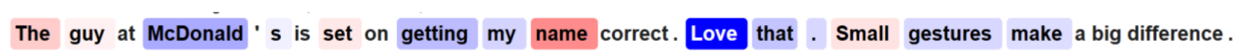

Figure 21. SHAP Visualization of Feature Impact – Example 2

Figures 29 presents attention visualizations that highlight the tokens that the model focused on most when making predictions. Words such as 'dislike', 'breakfast', and 'mcdonald' receive the highest attention weights, indicating their role in influencing the model's sentiment prediction.

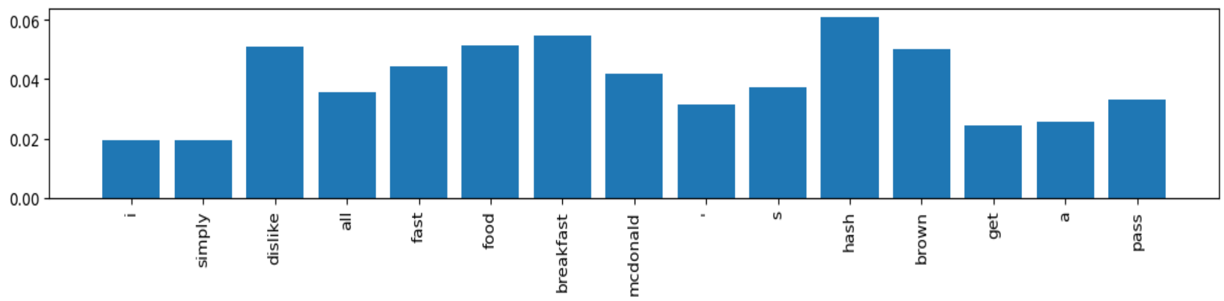

Figure 22. Attention Weights for Input Text – Example 1

4.3 Results of Model Application on the Twitter Dataset

Based on the comprehensive comparative analysis of two distinct Twitter datasets related to McDonald's, collected during two successive periods: October 2023-May 2024 and October 2024-May 2025, the following results offer insights into sentiment trends across both aspect categories and broad thematic categories.

4.3.1 Overall Sentiment Trends

A comparison of sentiment polarity across the two-year period reveals a clear change in public sentiment toward McDonald's. As illustrated in Figure 23, in the first year, negative sentiment was prevalent in majority of the tweets (40%), followed by neutral sentiment (39%) and positive sentiment representing only 21%. In the second year, however, negative sentiment dropped to 30%; in contrast, neutral sentiment experienced a considerable increase to 50%. Unlike the other two, positive sentiment remained relatively stable and decreased only slightly, falling to 19%.

4.3.2 Aspect Category Trends Over Time

As shown in Figure 24, further analysis of negative sentiment by aspect category shows that 'Ethical Responsibility' remained the most prevalent negative dimension in both observed timeframes. However, among all tweets expressing negative sentiment, its proportion declined significantly from 30% to 21%. Meanwhile, negative sentiment toward certain aspect categories, such as 'Food', 'Products', and 'Customer Service' increased slightly in Year 2. In addition, the results revealed slight increases in negative sentiment toward 'Brand Competition', 'Marketing Strategy', 'Loyalty', and 'Public Health Impact'. This suggests that several lower-profile aspects started getting more critical public attention in the second year.

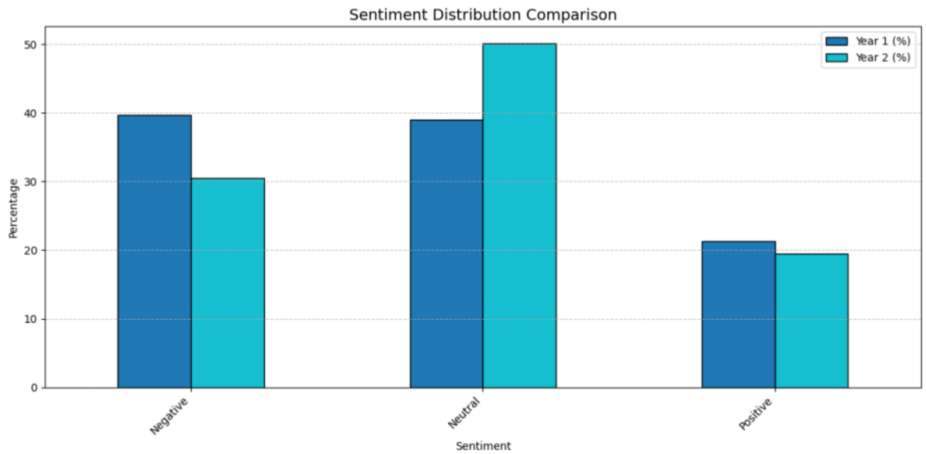

Figure 23. Sentiment Distribution Comparison Across Two Years

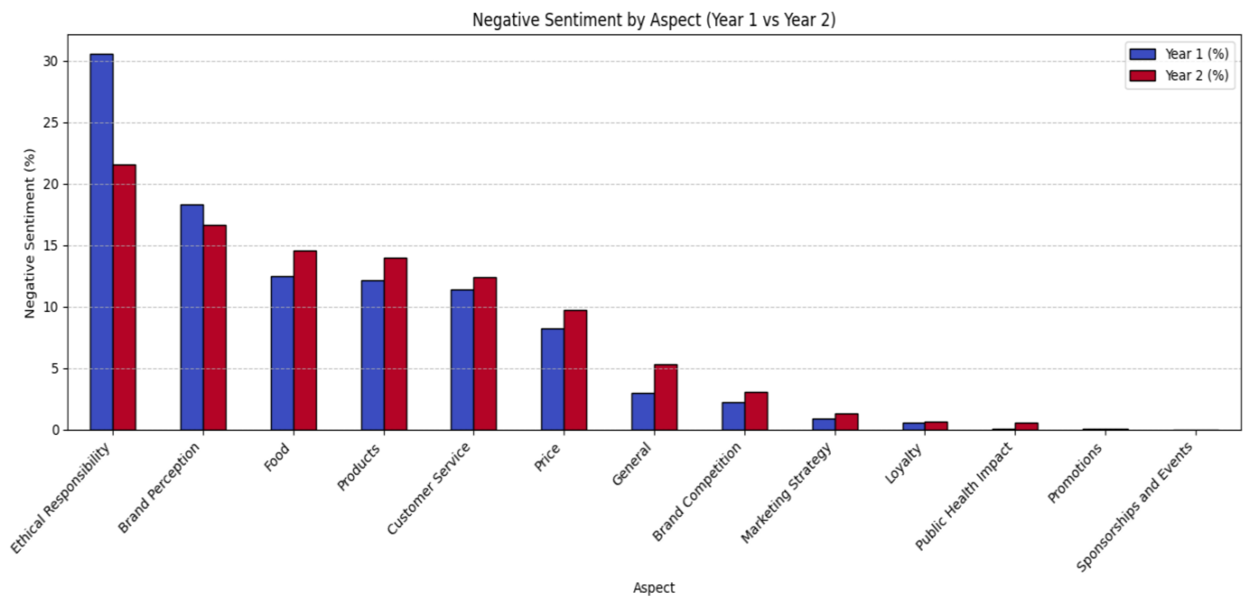

Figure 24. Aspect Category Negative Sentiment Comparison (Year 1 vs Year 2)

4.3.3 Broad Thematic Category Trends

As illustrated in Figure 25, among all tweets expressing negative sentiment, the proportion related to 'Core Restaurant Experience' increased from 36% to 42% at the broad thematic category level. This change positioned it as the category with the highest level of negative sentiment, while 'Corporate and Social Responsibility' declined significantly from 30% to 21%. This shift further supports the idea that consumer conversations in Year 2 focused more on operational aspects than value-driven criticism. Other broad thematic categories, including 'Brand Perception and Loyalty', 'Value for Money' and 'Promotions and Marketing' remained relatively stable.

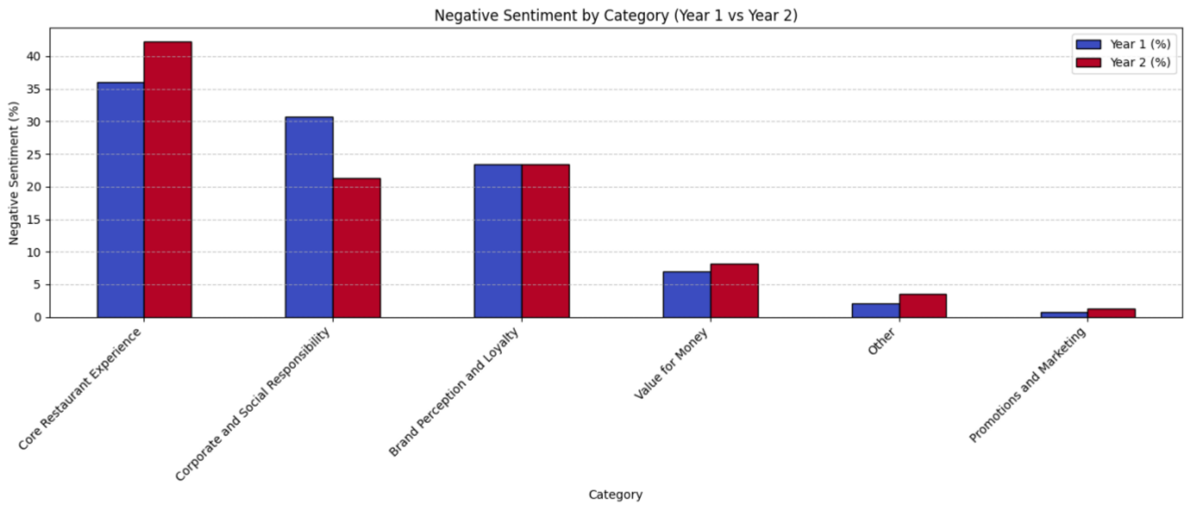

Figure 25. Broad Thematic Category Negative Sentiment Comparison (Year 1 vs Year 2)

4.4 Topic Modeling Insights Using BERTopic

To analyze evolving concerns, BERTopic was conducted on tweets identified as negative by the BBT model after applying it to 200,000 tweets. BERTopic revealed a clear shift in customer concerns between the two timeframes. As illustrated in Figures 26-37, in Year 1, many key topics focused on corporate social responsibility topics, with terms such as 'boycott', 'Palestine', 'Isreal', 'genocide', and 'supporting' appearing frequently in topics 0, 1, 9, 11, 16, and 18. This was the most frequent topic in Year 1, appearing in the highest-ranked topics: 0 and 1. However, in Year 2, although these CSR-related topics continued to appear in topics 1, 4, 21, 46, the highest-ranked topic in Year 2, was clearly focused on customer service dissatisfaction, shown by words, like 'service', 'worst', 'terrible', and 'bad' (Figure 27). Year 1's Topic 33 also expressed dissatisfaction with service; however, it ranked considerably lower in prominence compared to the dominant service-related topic in Year 2.

In Year 2, everyday customer issues such as product quality, pricing and machine and order problems received more attention. For example, keywords such as 'order', 'wrong', 'machine', 'broke' and 'working', 'service', 'price', 'bring back', 'menu' and 'expensive' were found in year 2 Topics 0, 5, 7, 8, 9, 54.

Figure 26. Comparison of Topic "Boycott" in Year 1 and Year 2 Based on BERTopic Output

Figure 27. Comparison of Topic "Customer Dissatisfaction and Service Criticism" in Year 1 and Year 2 Based on BERTopic Output

Figure 28. Comparison of Topic "Brand Boycotts" in Year 1 and Year 2 Based on BERTopic Output

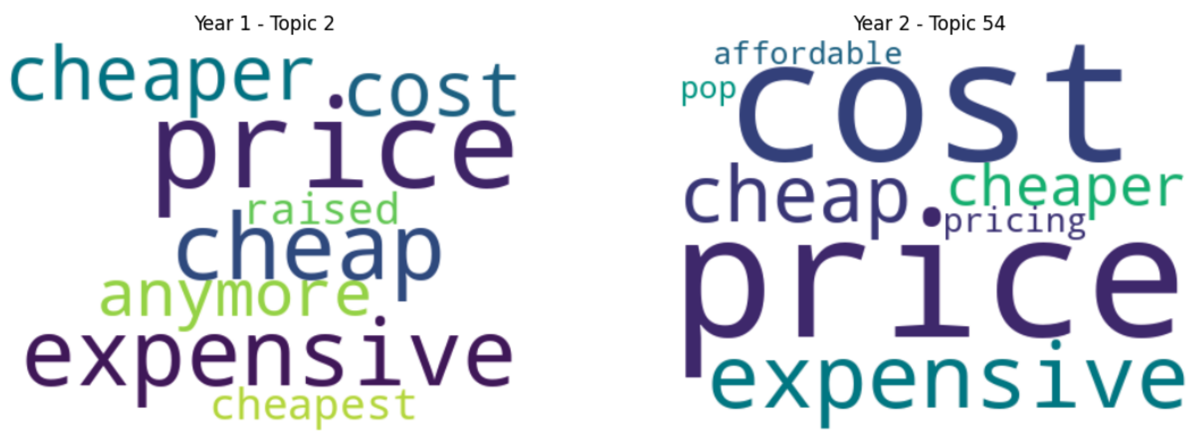

Figure 29. Comparison of Topic "Pricing and Affordability" in Year 1 and Year 2 Based on BERTopic Output

Figure 30. Comparison of Topic "Equipment and Machine Issues" in Year 1 and Year 2 Based on BERTopic Output

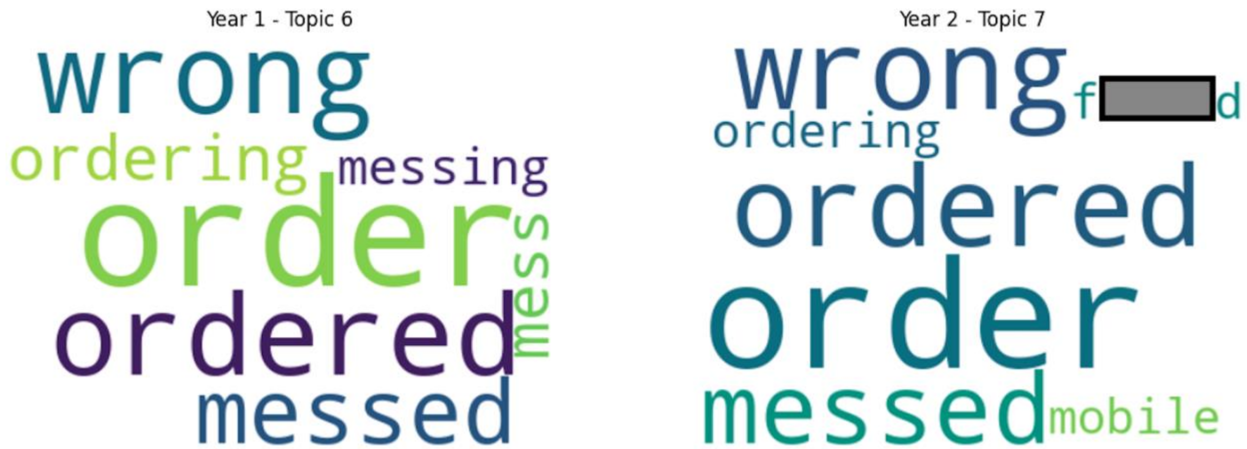

Figure 31. Comparison of Topic "Order Mistakes and Issues" in Year 1 and Year 2 Based on BERTopic Output

Figure 32. Comparison of Topic "Boycotting Brands" in Year 1 and Year 2 Based on BERTopic Output

Figure 33. Comparison of Topic "Palestine-Israel Conflict and Boycott" in Year 1 and Year 2 Based on BERTopic Output

Figure 34. Comparison of Topic "Menu Item Nostalgia and Return Requests" in Year 1 and Year 2 Based on BERTopic Output

Figure 35. Comparison of Topic "Israel-Palestine Conflict and Brand Boycotts" in Year 1 and Year 2 Based on BERTopic Output

Figure 36. Comparison of Topic "Ethical Concerns" in Year 1 and Year 2 Based on BERTopic Output

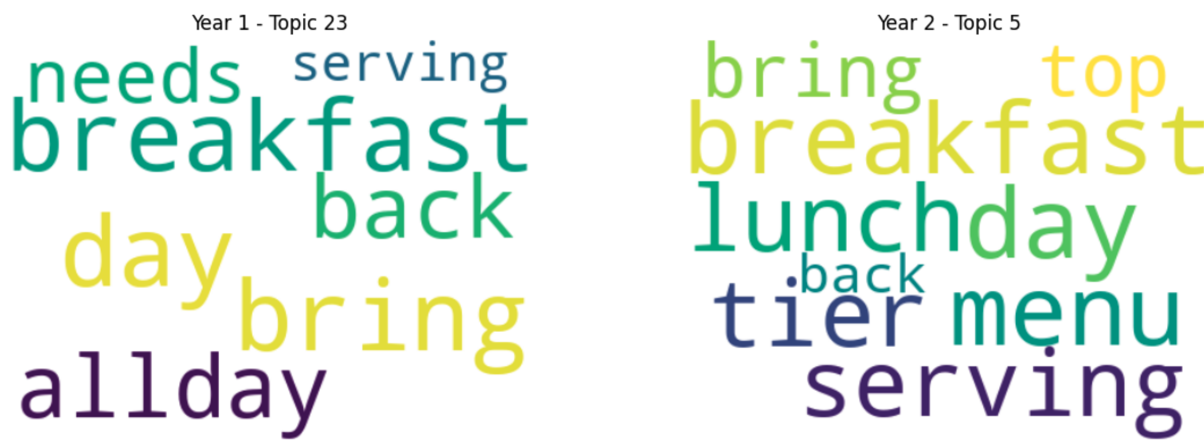

Figure 37. Comparison of Topic "Breakfast Menu Availability" in Year 1 and Year 2 Based on BERTopic Output

5. Discussions

Unlike much of the existing research that focuses primarily on the theoretical dimensions of ABSA, this study distinguishes itself by taking a practical step forward and applying an advanced hybrid deep learning model, further enhanced with SMOTE and supported by effective explainability techniques, to real-world, large-scale social media data. The following sections address the study’s three research questions:

1. How has customer sentiment towards McDonald's evolved between October 2023-May 2024 and October 2024-May 2025, and which aspect categories have contributed most significantly to these changes, as identified through the proposed ACSA model?
2. How does the integration of resampling techniques and a BERT-based hybrid model improve ACSA performance, particularly in addressing underrepresented aspect categories?
3. How can explainability techniques enhance the transparency of the BERT-based hybrid model for ACSA?

5.1 Changes in Customer Sentiment and Aspect Category Focus Over Time

To address Research Question 1, a comparative analysis was conducted using the proposed ABSA model—specifically, an ACSA approach—on 200,000 McDonald’s-related tweets collected across two consecutive timeframes, revealing notable patterns in customer sentiment and aspect category focus. In addition to ACSA classification, a BERTopic model was applied specifically to the subset of tweets identified as negative, enabling a deeper exploration of the evolving topics and lexical patterns within customer feedback. The following subsections interpret the outcomes of each analytical method.

5.1.1 Interpretation of Results from BBT Model Deployment

The results from the ACSA classification generated by the BBT model across two timeframes show a clear change in public sentiment about McDonald’s. The analysis indicates that the company was able to manage negative feedback, possibly with the help of improved digital marketing and strategic brand communication. Although there was a decline in negative sentiment, the sentiment trend moved mainly toward a neutral stance. This suggests that while McDonald’s has successfully reduced negative feedback, it did not generate stronger positive engagement and customer satisfaction.

In conducting the broad thematic and aspect category analysis, ‘Corporate and Social Responsibility’, particularly its subcategory, Ethical Responsibility’, remained a persistent source of criticism and negative sentiment, revealing a gap between public expectations and perceived corporate behavior. However, their proportion dropped sharply in the second year. This may be due to a decrease in public conversations about ethical issues such as company boycotts or involvement in geopolitical matters.

Interestingly, there was a noticeable shift in negative customer feedback toward day-to-day core restaurant experiences, particularly in aspects such as ‘Food’, ‘Products’, and ‘Customer Service’. The increased criticism toward core restaurant experiences and price could mean that consumer commentary became more focused on direct experience with the actual McDonald’s products and services instead of broader ideological issues. Sentiment towards ‘Loyalty’ remained relatively steady, while negative feedback related to ‘Brand Perception’ showed only a slight increase. This supports previous research suggesting that while the brand image may be affected by negative perceptions, customer loyalty remains a distinct phenomenon influenced by other factors (Andreassen & Lindestad, 1998).

5.1.2 Interpretation of Topics in Negative Tweets Using BERTopic

To supplement the sentiment and aspect-based insights generated by the ACSA model, BERTopic was applied to tweets classified as negative across both time periods. This approach provides deeper insights into the topics and keywords found in negative tweets.

The topic modeling results from BERTopic further supported the findings of the ACSA model by highlighting key topics and frequently used words in tweets classified as negative across both time periods. The results also revealed that boycott-related conversations remained a dominant and recurring topic in both periods but evolved in focus. As illustrated in Figures 26, 28, 32, 33, 35, and 36, boycott-related topics in Year 1 (e.g., Topics 0, 1, 9, 11, 16, and 18) were primarily focused on general sentiments about the act of boycotting and referenced other brands such as, Disney and Starbucks. In contrast, Year 2 boycott topics (e.g., Topics 1, 4, 21, and 46), as shown in Figures 26, 28, 33, and 36, reflected a more geopolitically charged tone, with increased reference to Isreal, Gaza, Palestine, alongside targeted criticism toward brands perceived to be involved, such as Starbucks, Zara, and Walmart. These findings confirm the results of the ACSA model, which indicated that while overall mentions of ‘Corporate and Social Responsibility’ decreased, the remaining discussions became more emotionally focused on political and broader global issues and conflicts.

In addition, the BERTopic results showed increased attention to core operational issues, such as product quality, pricing, and problems related to machines and orders. This was reflected by keywords like ‘order’, ‘wrong’, ‘machine’, ‘broke’, and ‘working’, ‘service’, ‘price’, ‘bring back’, ‘menu’ and ‘expensive’, which frequently appeared in Year 2 Topics 0, 5, 7, 8, 9, and 54 as illustrated in Figures 27, 29, 30, 31, 34, and 37. This qualitative trend supports the observed rise in negative sentiment related to ‘Food’, ‘Products’ and, ‘Customer Service’ as consumers increasingly expressed dissatisfaction with everyday dining experiences. Similarly, price-related complaints persisted across both years—for instance, in Year 1 (Topic 2) and Year 2 (Topic 54), as illustrated in Figure 29. These product- and service-related topics indicate that negative consumer feedback was not only tied to service quality and CSR, but also to issues such as product availability, pricing, equipment malfunctions, etc. Each of these factors influences customer satisfaction and brand engagement.

5.1.3 Strategic Implications of Customer Sentiment and Topic Trends for McDonald’s

Collecting and interpreting customer feedback allows businesses to understand new trends, address pain points, strengthen brand-customer relationships, and improve service and product quality (Reichheld & Schefter, 2000). Unlike traditional methods, AI-driven approaches such as NLP enable rapid and in-depth analysis of large-scale textual data to uncover customer sentiments and preferences (Chatterjee et al., 2022; Ranjan et al., 2024). Therefore, utilizing AI systems allows businesses to access customer feedback in real-time and gain deeper insights into emerging trends and market changes which facilitates faster responses to such shifts, leading to improved customer experience and more informed strategic decision-making (Okeke et al., 2024; Rane, 2023).

For instance, when negative comments or boycott calls on platforms such as Twitter increase, companies can use AI tools to quickly track and detect negative trends early. Informed by such analysis, companies can respond fast and clear up misunderstandings using strategic hashtags, and work with trusted influencers to guide public opinion and boost positive feedback through electronic word of mouth (e-WOM) (Pradhipta et al., 2024). However, negative content spreads fast across online platforms, making reputation management and strategic communication crucial (Coombs, 2007). In these situations, quality of communication plays a key role in building trust between brands and consumers (Yannopoulou, 2011).

Therefore, companies like McDonald's should not only address aspect categories related to day-to-day restaurant experiences—such as pricing strategies, which contribute to customer satisfaction and repeat business (Wantrara & Tambrin, 2019), and customer services, which directly influences perceived product quality and competitive advantage (Goffin & Price, 1996)—but also take deliberate steps to enhance customer's cognitive and emotional perceptions that go beyond the product's physical characteristics and help shape brand image (Saxena & Dhar, 2017). This is particularly important because customer food choices are influenced by a range of factors, including economic conditions, product quality, cultural and religious values, and broader global issues (Reswara et al., 2024).

In politically or religiously sensitive environments, companies should communicate openly, clearly state their values, respect cultural differences and demonstrate ethical behavior. Boycotts motivated by religious or moral beliefs are remarkably persistent, requiring both immediate crisis management and long-term strategies for ethical branding, community trust and proactive management of misinformation across digital platforms (Dekhil et al., 2017; Muhamad et al., 2018; Samudra et al., 2024). In these situations, combining AI monitoring with strategic communication and ethical branding is key to maintaining customer trust and protecting the brand (Pradhipta et al., 2024).

Additionally, given the growing negative feedback directed at McDonald's social responsibility efforts and the increasing consumer preference for brands that truly help society (Ha et al., 2023), the company's communication strategy should go beyond image protection. It should be rooted in ethical principles that foster transparency, accountability and social impact (Coombs & Holladay, 2002; Domschat et al., 2023; Sturges, 1994). Like many organizations that increasingly rely on digital and social media to communicate their initiatives and build support among stakeholders (Dwivedi et al., 2015; Grover et al, 2019), McDonald's should leverage its digital marketing capabilities to foster stronger customer engagement and communication.

In conclusion, McDonald's should not only address categories related to daily operations, such as service quality but also focus on broader categories like corporate social responsibility, which shape the emotional perceptions of customers toward the brand. To build and maintain trust—especially during sensitive periods like those examined in this study—the company must adopt a strategic communication approach that emphasizes transparency and high-quality messaging. Specifically, McDonald's should utilize AI tools to go beyond reactive responses by continuously tracking and detecting negative sentiment trends early. This enables the company to respond swiftly by using strategic hashtags, collaborating with credible influencers, and promoting positive engagement through e-WOM. Finally, combining AI-driven sentiment analysis, culturally sensitive and clear communication, ethical branding, and digital stakeholder engagement will allow McDonald's to manage its reputation more effectively while fostering long-term customer trust and loyalty.

5.2 Impact of Model Architecture and Sampling on Performance

To address research question 2, this section discusses the evaluation metrics used to measure performance of the proposed hybrid model and its integration with SMOTE. The findings revealed the strong performance of the BBT model—a BERT-based hybrid architecture enhanced with SMOTE—achieving F1-scores of 0.78 for aspect category classification, 0.85 for broad thematic category classification, and 0.86 for Sentiment classification. Its strong performance in aspect category classification—a task widely recognized as the most challenging within ABSA—further reinforces robustness of the BBT model. These findings align with previous research (Rozi et al., 2024; Wang et al., 2021; Xin & Zakaria, 2024; Xiong et al., 2024). Although BERT effectively captures contextual information, the BiLSTM layer enhances it by providing a deeper understanding of sequential dependencies, while the Transformer layer builds on this foundation by employing a self-attention mechanism to identify and emphasize the most relevant parts of the text.

In addition, the use of SMOTE improved overall model performance by addressing class imbalance. This was particularly evident in the prediction of aspect categories with limited instances, which the model had difficulty predicting previously. Interestingly, even though the sentiment classes were not highly imbalanced, applying SMOTE still improved the F1-scores, particularly for the 'Negative' sentiment. This improvement may be due to the SMOTE's ability to better separate the classes and reduce overlap in the feature space, allowing the model to generalize more effectively—even under relatively balanced conditions.

5.3 Impact of Explainability on Model Transparency in Sentiment Classification

To answer research question 3, this study used three post-hoc explainability techniques—LIME, SHAP, and attention visualization—to make the model's predictions easier to understand. While no quantitative improvement in predictive performance was measured, these methods added qualitative value by highlighting specific tokens that most influenced model's sentiment classifications. For instance, LIME explanations revealed the terms that contributed most strongly to the model's predicted sentiment labels. SHAP also indicated the degree of positivity and negativity associated with each token in the tweets. However, certain limitations were observed. Attention weights were less explanatory, as they merely identified words which received high attention without clarifying whether these carried a negative or positive tone—suggesting a need for caution when interpreting attention as explanation. Additionally, since interpretability was not evaluated using quantitative metrics, the contribution of these techniques remains qualitative and illustrative.

Although this approach does not improve numerical performance, it addresses a key challenge in AI deployment: understanding and interpreting the decision-making of deep learning models. By providing visual explanations, these techniques help build stakeholder trust (Miller, 2019), support a balance between accuracy and interpretability, and make ethical evaluation more feasible and encouraged (Talaat et al., 2024). While explainability tools can also reveal potential biases—such as the model's repeated emphasis on brand mentions like 'McDonald's'—further analysis is required to determine whether such patterns are meaningful or misleading.

These observations not only demonstrate the value of explainability within this study but also are in line with the previous research (Perikos & Diamantopoulos, 2024). Using the model's predictions and internal representations, post-hoc techniques, particularly LIME, and SHAP can highlight the most influential tokens contributing to classification outcomes. Such techniques are particularly useful in real-world situations, where understanding the reasoning behind AI decisions is essential for ensuring transparency and accountability (Islam, 2022), especially in high-stake fields where AI decisions can carry significant consequences (Hosain et al., 2024). However, to ensure that explainability techniques contribute meaningfully to such domains, they should be combined with complementary evaluation strategies—both quantitative and qualitative.

6. Limitations and Future Research

While the proposed model performed well during empirical evaluation and real-world deployment, some limitations should be acknowledged.

1. While 1,000 labeled tweets were sufficient for initial model fine-tuning, it may limit the model's ability to generalize across unseen and changing language patters.
2. This study focused only on English-language tweets. Therefore, these findings may not generalize to other languages or social media platforms. Future research could address this limitation by including data from other platforms and in different languages.
3. The annotation process in this study focused on labeling aspect and broad thematic categories with their associated sentiments, but did not explicitly identify opinion words. Labeling these words could improve the interpretability of techniques such as SHAP and LIME by more clearly highlighting the parts of the text that influence model decisions. However, these techniques provide only an estimation—rather than a complete explanation—of the model's decision-making process.
4. A limitation of the applied explainability techniques is the absence of quantitative evaluation or accuracy—interpretability trade-off analysis, leaving their contribution largely illustrative.
5. A limitation of this study is relying on F1-score, which may not provide a complete picture for majority classes, as the metric is highly affected by low precision. To address this, future research could include additional evaluation metrics such as ROC-AUC or precision-recall curves to provide a more balanced assessment of performance, especially in the context of imbalanced datasets.
6. Based on the results generated by the deployed BBT model to large-scale data, although the significant decline in tweets related to ethical concerns may reflect a genuine shift in public focus, it could also be simply because of reduced discussion about these issues during the second period. Future studies could examine external influences like news events or global developments, to better contextualize these shifts.

Despite the above-mentioned limitations, this study presents a scalable, affordable and transparent deep learning model with practical applications for businesses aiming to analyze and act on large-scale, fine-grained customer sentiment.

References

Abdelkafi, N., & Täuscher, K. (2016). Business models for sustainability from a system dynamics perspective. *Organization & Environment*, *29*(1), 74-96.

Adiwardana, D., Luong, M. T., So, D. R., Hall, J., Fiedel, N., Thoppilan, R., ... & Le, Q. V. (2020). Towards a human-like open-domain chatbot. *arXiv preprint arXiv:2001.09977*.

Agarwal, M., & Goel, S. (2014, May). Expert system and it's requirement engineering process. In *International Conference on Recent Advances and Innovations in Engineering (ICRAIE-2014)* (pp. 1-4). IEEE.

Agatonovic-Kustrin, S., & Beresford, R. (2000). Basic concepts of artificial neural network (ANN) modeling and its application in pharmaceutical research. *Journal of pharmaceutical and biomedical analysis*, *22*(5), 717-727.

Aguilar-Moreno, J. A., Palos-Sanchez, P. R., & Pozo-Barajas, R. D. (2024). Sentiment analysis to support business decision-making. A bibliometric study. *AIMS Mathematics*, *9*(2), 4337-4375.

Ahmad, M., Aftab, S., & Ali, I. (2017). Sentiment analysis of tweets using svm. *Int. J. Comput. Appl*, *177*(5), 25-29.

Ahmad, M., Aftab, S., Bashir, M. S., Hameed, N., Ali, I., & Nawaz, Z. (2018). SVM optimization for sentiment analysis. *Int. J. Adv. Comput. Sci. Appl*, *9*(4), 393-398.

Ahmad, K., Abdelrazek, M., Arora, C., Bano, M., & Grundy, J. (2023a). Requirements engineering for artificial intelligence systems: A systematic mapping study. *Information and Software Technology*, *158*, 107176.

Ahmad, K., Abdelrazek, M., Arora, C., Baniya, A. A., Bano, M., & Grundy, J. (2023b). Requirements engineering framework for human-centered artificial intelligence software systems. *Applied Soft Computing*, *143*, 110455.

Ahmad, K., Abdelrazek, M., Arora, C., Grundy, J., & Bano, M. (2023c). Requirements elicitation and modelling of artificial intelligence systems: an empirical study. *arXiv preprint arXiv:2302.06034*.

Alanazi, A. (2022). Using machine learning for healthcare challenges and opportunities. *Informatics in Medicine Unlocked*, *30*, 100924.

Alenezi, M., & Akour, M. (2025). AI-Driven Innovations in Software Engineering: A Review of Current Practices and Future Directions. *Applied Sciences*, *15*(3), 1344.

Alhosani, K., & Alhashmi, S. M. (2024). Opportunities, challenges, and benefits of AI innovation in government services: a review. *Discover Artificial Intelligence*, *4*(1), 18.

Alizadegan, H., Rashidi Malki, B., Radmehr, A., Karimi, H., & Ilani, M. A. (2025). Comparative study of long short-term memory (LSTM), bidirectional LSTM, and traditional machine learning approaches for energy consumption prediction. *Energy Exploration & Exploitation*, *43*(1), 281-301.

Al-Tameemi, I. K. S., Feizi-Derakhshi, M. R., Pashazadeh, S., & Asadpour, M. (2024). A comprehensive review of visual–textual sentiment analysis from social media networks. *Journal of Computational Social Science*, *7*(3), 2767-2838.

Alzubaidi, L., Zhang, J., Humaidi, A. J., Al-Dujaili, A., Duan, Y., Al-Shamma, O., ... & Farhan, L. (2021). Review of deep learning: concepts, CNN architectures, challenges, applications, future directions. *Journal of big Data*, *8*, 1-74.

Amjad, H., Ashraf, M. S., Sherazi, S. Z. A., Khan, S., Fraz, M. M., Hameed, T., & Bukhari, S. A. C. (2023). Attention-Based Explainability Approaches in Healthcare Natural Language Processing. *HEALTHINF*, 689-696.

Ananth kumar, T., J., Z., Kanimozhi, P., R., R., Ananth, C., & Ajagbe, S. A. (2024). Machine Learning and Sentiment Analysis: Analysing Customer Feedback. In R. Masengu, O. Chiwaridzo, M. Dube, & B. Ruzive (Eds.), *AI-Driven Marketing Research and Data Analytics* (pp. 245-262). IGI Global Scientific Publishing. https://doi.org/10.4018/979-8-3693-2165-2.ch014.

Andreassen, T. W., & Lindestad, B. (1998). Customer loyalty and complex services: The impact of corporate image on quality, customer satisfaction and loyalty for customers with varying degrees of service expertise. *International Journal of service Industry management*, *9*(1), 7-23.

Angelov, P. P., Soares, E. A., Jiang, R., Arnold, N. I., & Atkinson, P. M. (2021). Explainable artificial intelligence: an analytical review. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *11*(5), e1424.

Antwarg, L., Miller, R. M., Shapira, B., & Rokach, L. (2021). Explaining anomalies detected by autoencoders using Shapley Additive Explanations. *Expert systems with applications*, *186*, 115736.

Appel, O., Chiclana, F., Carter, J., & Fujita, H. (2016). A hybrid approach to the sentiment analysis problem at the sentence level. *Knowledge-Based Systems*, *108*, 110-124.

Arora, C., Grundy, J., & Abdelrazek, M. (2024). Advancing requirements engineering through generative ai: Assessing the role of llms. In *Generative AI for Effective Software Development* (pp. 129-148). Cham: Springer Nature Switzerland.

Asudani, D. S., Nagwani, N. K., & Singh, P. (2023). Impact of word embedding models on text analytics in deep learning environment: a review. *Artificial intelligence review*, *56*(9), 10345-10425.

Ausat, A. M. A. (2023). The role of social media in shaping public opinion and its influence on economic decisions. *Technology and Society Perspectives (TACIT)*, *1*(1), 35-44.

Avalle, M., Di Marco, N., Etta, G., Sangiorgio, E., Alipour, S., Bonetti, A., ... & Quattrociocchi, W. (2024). Persistent interaction patterns across social media platforms and over time. *Nature*, *628*(8008), 582-589.

Azodi, C. B., Tang, J., & Shiu, S. H. (2020). Opening the black box: interpretable machine learning for geneticists. *Trends in genetics*, *36*(6), 442-455.

Azzaakiyyah, H. K. (2023). The impact of social media use on social interaction in contemporary society. *Technology and Society Perspectives (TACIT)*, *1*(1), 1-9.

Bach, M., Werner, A., Żywiec, J., & Pluskiewicz, W. (2017). The study of under-and over-sampling methods' utility in analysis of highly imbalanced data on osteoporosis. *Information Sciences*, *384*, 174-190.

Baheti, A., Ritter, A., Li, J., & Dolan, B. (2018). Generating more interesting responses in neural conversation models with distributional constraints. *arXiv preprint arXiv:1809.01215*.

Baker, M. R., Mohammed, E. Z., & Jihad, K. H. (2022, December). Prediction of colon cancer related tweets using deep learning models. In *International Conference on Intelligent Systems Design and Applications* (pp. 522-532). Cham: Springer Nature Switzerland.

Baker, M. R., Taher, Y. N., & Jihad, K. H. (2023). PREDICTION OF PEOPLE SENTIMENTS ON TWITTER USING MACHINE LEARNING CLASSIFIERS DURING RUSSIAN AGGRESSION IN UKRAINE. *Jordanian Journal of Computers & Information Technology*, *9*(3).

Bashiri, H., & Naderi, H. (2024). Comprehensive review and comparative analysis of transformer models in sentiment analysis. *Knowledge and Information Systems*, *66*(12), 7305-7361.

BATUR DİNLER, Ö., & Aydin, N. (2020). An optimal feature parameter set based on gated recurrent unit recurrent neural networks for speech segment detection. *Applied Sciences*, *10*(4), 1273.

Baumann, C., Cherry, M., & Chu, W. (2019). Competitive productivity (CP) at macro–meso–micro levels. *Cross Cultural & Strategic Management*, *26*(2), 118-144.

Behdenna, S., Barigou, F., & Belalem, G. (2016). Sentiment analysis at document level. In *Smart Trends in Information Technology and Computer Communications: First International Conference, SmartCom 2016, Jaipur, India, August 6–7, 2016, Revised Selected Papers 1* (pp. 159-168). Springer Singapore.

Belani, H., Vukovic, M., & Car, Ž. (2019, September). Requirements engineering challenges in building AI-based complex systems. In *2019 IEEE 27th international requirements engineering conference workshops (REW)* (pp. 252-255). IEEE.

Bengio, Y. (2009). Learning deep architectures for AI. *Foundations and trends® in Machine Learning*, *2*(1), 1-127.

Bensoltane, R., & Zaki, T. (2022). Towards Arabic aspect-based sentiment analysis: A transfer learning-based approach. *Social Network Analysis and Mining*, *12*, 1-16.

Bilgilioğlu, S. S., Gezgin, C., Iban, M. C., Bilgilioğlu, H., Gündüz, H. I., & Arslan, Ş. (2025). Explainable Sinkhole Susceptibility Mapping Using Machine-Learning-Based SHAP: Quantifying and Comparing the Effects of Contributing Factors in Konya, Türkiye. *Applied Sciences*, *15*(6), 3139.

Birjali, M., Kasri, M., & Beni-Hssane, A. (2021). A comprehensive survey on sentiment analysis: Approaches, challenges and trends. *Knowledge-Based Systems*, *226*, 107134.

Boh, W., Constantinides, P., Padmanabhan, B., & Viswanathan, S. (2023). Building digital resilience against major shocks. *Mis Quarterly*, *47*(1), 343-360.

Bonifazi, G., Breve, B., Cirillo, S., Corradini, E., & Virgili, L. (2022). Investigating the COVID-19 vaccine discussions on Twitter through a multilayer network-based approach. *Information Processing & Management*, *59*(6), 103095.

Bonta, V., Kumaresh, N., & Janardhan, N. (2019). A comprehensive study on lexicon based approaches for sentiment analysis. *Asian Journal of Computer Science and Technology*, *8*(S2), 1-6.

Bowman, S. R., Angeli, G., Potts, C., & Manning, C. D. (2015). A large annotated corpus for learning natural language inference. *arXiv preprint arXiv:1508.05326*.

Branco, P., Torgo, L., & Ribeiro, R. P. (2016). A survey of predictive modeling on imbalanced domains. *ACM computing surveys (CSUR)*, *49*(2), 1-50.

Broekens, J., Hilpert, B., Verberne, S., Baraka, K., Gebhard, P., & Plaat, A. (2023, September). Fine-grained affective processing capabilities emerging from large language models. In *2023 11th international conference on affective computing and intelligent interaction (ACII)* (pp. 1-8). IEEE.

Buhrmester, V., Münch, D., & Arens, M. (2021). Analysis of explainers of black box deep neural networks for computer vision: A survey. *Machine Learning and Knowledge Extraction*, *3*(4), 966-989.

Cai, H., Tu, Y., Zhou, X., Yu, J., & Xia, R. (2020, December). Aspect-category based sentiment analysis with hierarchical graph convolutional network. In *Proceedings of the 28th international conference on computational linguistics* (pp. 833-843).

Cambria, E., Schuller, B., Xia, Y., & Havasi, C. (2013). New avenues in opinion mining and sentiment analysis. *IEEE Intelligent systems*, *28*(2), 15-21.

Cambridge Dictionary. (n.d.). *Spring*. In *Cambridge English Dictionary*. Cambridge University Press. https://dictionary.cambridge.org/dictionary/english/spring

Cano-Marin, E., Mora-Cantallops, M., & Sánchez-Alonso, S. (2023). Twitter as a predictive system: a systematic literature review. *Journal of Business Research*, *157*, 113561.

Cao, M., Chychyla, R., & Stewart, T. (2015). Big data analytics in financial statement audits. *Accounting horizons*, *29*(2), 423-429.

Chai, D., Wu, W., Han, Q., Wu, F., & Li, J. (2020, November). Description based text classification with reinforcement learning. In *International conference on machine learning* (pp. 1371-1382). PMLR.

Chan, R. H., Kan, K. K., Nikolova, M., & Plemmons, R. J. (2020). A two-stage method for spectral–spatial classification of hyperspectral images. *Journal of Mathematical Imaging and Vision*, *62*(6), 790-807.

Chatterjee, S., Chaudhuri, R., & Vrontis, D. (2022). Big data analytics in strategic sales performance: mediating role of CRM capability and moderating role of leadership support. *EuroMed Journal of Business*, *17*(3), 295-311.

Chaturvedi, I., Cambria, E., Welsch, R. E., & Herrera, F. (2018). Distinguishing between facts and opinions for sentiment analysis: Survey and challenges. *Information Fusion*, *44*, 65-77.

Chavez, M. R., Butler, T. S., Rekawek, P., Heo, H., & Kinzler, W. L. (2023). Chat Generative Pre-trained Transformer: why we should embrace this technology. *American journal of obstetrics and gynecology*, *228*(6), 706-711

Chen, I. J., & Popovich, K. (2003). Understanding customer relationship management (CRM): People, process and technology. *Business process management journal*, *9*(5), 672-688.

Chen, Y., Iyengar, R., & Iyengar, G. (2017a). Modeling multimodal continuous heterogeneity in conjoint analysis—a sparse learning approach. *Marketing Science*, *36*(1), 140-156.

Chen, T., Xu, R., He, Y., & Wang, X. (2017b). Improving sentiment analysis via sentence type classification using BiLSTM-CRF and CNN. Expert Systems with Applications, 72, 221–230. https://doi.org/10.1016/j.eswa.2016.10.056

Chen, C., Wu, T., Guo, Z., & Cheng, J. (2021). Combination of deep neural network with attention mechanism enhances the explainability of protein contact prediction. *Proteins: Structure, Function, and Bioinformatics*, *89*(6), 697-707.

Chifu, A. G., & Fournier, S. (2024). Linguistic features for sentence difficulty prediction in ABSA. *arXiv preprint arXiv:2402.03163*.

Chintalapudi, N., Battineni, G., Di Canio, M., Sagaro, G. G., & Amenta, F. (2021). Text mining with sentiment analysis on seafarers' medical documents. *International Journal of Information Management Data Insights*, *1*(1), 100005.

Chun, S. H., & Nyam-Ochir, A. (2020). The effects of fast food restaurant attributes on customer satisfaction, revisit intention, and recommendation using DINESERV scale. *Sustainability*, *12*(18), 7435.

Cieslak, D. A., Hoens, T. R., Chawla, N. V., & Kegelmeyer, W. P. (2012). Hellinger distance decision trees are robust and skew-insensitive. *Data Mining and Knowledge Discovery*, *24*, 136-158.

Ciodaro, T., Deva, D., De Seixas, J. M., & Damazio, D. (2012, June). Online particle detection with neural networks based on topological calorimetry information. In *Journal of physics: conference series* (Vol. 368, No. 1, p. 012030). IOP Publishing.

Collobert, R., Weston, J., Bottou, L., Karlen, M., Kavukcuoglu, K., & Kuksa, P. (2011). Natural language processing (almost) from scratch.

Coombs, W. T., & Holladay, S. J. (2002). Helping crisis managers protect reputational assets: Initial tests of the situational crisis communication theory. *Management communication quarterly*, *16*(2), 165-186.

Curreri, F., Patanè, L., & Xibilia, M. G. (2021). RNN-and LSTM-based soft sensors transferability for an industrial process. *Sensors*, *21*(3), 823.

Dakhel, A. M., Nikanjam, A., Khomh, F., Desmarais, M. C., & Washizaki, H. (2024). An overview on large language models. *Generative AI for Effective Software Development*, 3-21.

Damirchi, R., & Amini, A. (2023, November). Non-Functional Requirement Extracting Methods for AI-based Systems: A Survey. In *2023 13th International Conference on Computer and Knowledge Engineering (ICCKE)* (pp. 535-539). IEEE.

Darwish, K., Habash, N., Abbas, M., Al-Khalifa, H., Al-Natsheh, H. T., Bouamor, H., ... & Mubarak, H. (2021). A panoramic survey of natural language processing in the Arab world. *Communications of the ACM*, *64*(4), 72-81.

Dekhil, F., Jridi, H., & Farhat, H. (2017). Effect of religiosity on the decision to participate in a boycott: The moderating effect of brand loyalty–the case of Coca-Cola. *Journal of Islamic Marketing*, *8*(2), 309-328.

Dettmers, T., Pagnoni, A., Holtzman, A., & Zettlemoyer, L. (2023). Qlora: Efficient finetuning of quantized llms. *Advances in neural information processing systems*, *36*, 10088-10115.

Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019, June). Bert: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 conference of the North American chapter of the association for computational linguistics: human language technologies, volume 1 (long and short papers)* (pp. 4171-4186).

Diaz, G. O., Zhang, X., & Ng, V. (2020, May). Aspect-based sentiment analysis as fine-grained opinion mining. In *Proceedings of the Twelfth Language Resources and Evaluation Conference* (pp. 6804-6811).

Do, H. H., Prasad, P. W., Maag, A., & Alsadoon, A. (2019). Deep learning for aspect-based sentiment analysis: a comparative review. *Expert systems with applications*, *118*, 272-299.

Domschat, L., Stephens, M., & Saeed, H. (2023). Communication to protect brand image during a terrorism crisis: Application of situational crisis communication theory. *Journal of Contingencies and Crisis Management*, *31*(4), 913-927.

Dong, L., Yang, N., Wang, W., Wei, F., Liu, X., Wang, Y., ... & Hon, H. W. (2019). Unified language model pre-training for natural language understanding and generation. *Advances in neural information processing systems*, *32*.

Dos Santos, B. N., Marcacini, R. M., & Rezende, S. O. (2021). Multi-domain aspect extraction using bidirectional encoder representations from transformers. *IEEE Access*, *9*, 91604-91613.

Duan, Y., Edwards, J. S., & Dwivedi, Y. K. (2019). Artificial intelligence for decision making in the era of Big Data–evolution, challenges and research agenda. *International journal of information management*, *48*, 63-71.

Dwivedi, Y. K., Kapoor, K. K., & Chen, H. (2015). Social media marketing and advertising. *The Marketing Review*, *15*(3), 289-309.

Dwivedi, Y. K., Ismagilova, E., Hughes, D. L., Carlson, J., Filieri, R., Jacobson, J., ... & Wang, Y. (2021). Setting the future of digital and social media marketing research: Perspectives and research propositions. *International journal of information management*, *59*, 102168.

Edunov, S., Ott, M., Auli, M., & Grangier, D. (2018). Understanding back-translation at scale. *arXiv preprint arXiv:1808.09381*.

Emam, Z., Kondrich, A., Harrison, S., Lau, F., Wang, Y., Kim, A., & Branson, E. (2021). On the state of data in computer vision: Human annotations remain indispensable for developing deep learning models. *arXiv preprint arXiv:2108.00114*.

Ertel, W., & Ertel, W. (2017). Machine learning and data mining. *Introduction to Artificial Intelligence*, 175-243.

Ertel, W. (2024). *Introduction to artificial intelligence*. Springer Nature.

Estep, M. (2006). *Self-organizing natural intelligence: Issues of knowing, meaning, and complexity*. Springer Science & Business Media.

Farias, D. H., & Rosso, P. (2017). Irony, sarcasm, and sentiment analysis. In *Sentiment analysis in social networks* (pp. 113-128). Morgan Kaufmann.

Fernández, A., Garcia, S., Herrera, F., & Chawla, N. V. (2018). SMOTE for learning from imbalanced data: progress and challenges, marking the 15-year anniversary. *Journal of artificial intelligence research*, *61*, 863-905.

Gandomi, A., & Haider, M. (2015). Beyond the hype: Big data concepts, methods, and analytics. *International journal of information management*, *35*(2), 137-144.

Gao, C., Yan, J., Zhou, S., Varshney, P. K., & Liu, H. (2019). Long short-term memory-based deep recurrent neural networks for target tracking. *Information Sciences*, *502*, 279-296.

Geetha, M. P., & Renuka, D. K. (2021). Improving the performance of aspect based sentiment analysis using fine-tuned Bert Base Uncased model. *International Journal of Intelligent Networks*, *2*, 64-69.

Geirhos, R., Temme, C. R., Rauber, J., Schütt, H. H., Bethge, M., & Wichmann, F. A. (2018). Generalisation in humans and deep neural networks. *Advances in neural information processing systems*, *31*.

Ghojogh, B., & Ghodsi, A. (2020). Attention mechanism, transformers, BERT, and GPT: tutorial and survey.

Gillioz, A., Casas, J., Mugellini, E., & Abou Khaled, O. (2020, September). Overview of the Transformer-based Models for NLP Tasks. In *2020 15th Conference on computer science and information systems (FedCSIS)* (pp. 179-183). IEEE.

Glorot, X., & Bengio, Y. (2010, March). Understanding the difficulty of training deep feedforward neural networks. In *Proceedings of the thirteenth international conference on artificial intelligence and statistics* (pp. 249-256). JMLR Workshop and Conference Proceedings.

Goffin, K., & Price, D. (1996). Service documentation and the biomedical engineer: Results of a survey.

Golbazi, J. (2024). *Analysis of User Sentiment Towards McDonald's on Twitter Using Aspect-Based Sentiment Analysis (ABSA).*

Goodman, B., & Flaxman, S. (2017). European Union regulations on algorithmic decision-making and a "right to explanation". *AI magazine*, *38*(3), 50-57.

Grant, R. M. (1991). The resource-based theory of competitive advantage: implications for strategy formulation. *California management review*, *33*(3), 114-135.

Graves, A., Liwicki, M., Fernández, S., Bertolami, R., Bunke, H., & Schmidhuber, J. (2008). A novel connectionist system for unconstrained handwriting recognition. *IEEE transactions on pattern analysis and machine intelligence*, *31*(5), 855-868.

Graves, A. (2013). Generating sequences with recurrent neural networks. *arXiv preprint arXiv:1308.0850*.

Graves, A., Mohamed, A. R., & Hinton, G. (2013, May). Speech recognition with deep recurrent neural networks. In *2013 IEEE international conference on acoustics, speech and signal processing* (pp. 6645-6649). Ieee.

Greff, K., Srivastava, R. K., Koutník, J., Steunebrink, B. R., & Schmidhuber, J. (2016). LSTM: A search space odyssey. *IEEE transactions on neural networks and learning systems*, *28*(10), 2222-2232.

Grootendorst, M. (2022). BERTopic: Neural topic modeling with a class-based TF-IDF procedure. *arXiv preprint arXiv:2203.05794*.

Grover, P., Kar, A. K., & Janssen, M. (2019). Diffusion of blockchain technology: Insights from academic literature and social media analytics. *Journal of Enterprise Information Management*, *32*(5), 735-757.

Guleria, P., Srinivasu, P. N., & Hassaballah, M. (2024). Diabetes prediction using Shapley additive explanations and DSaaS over machine learning classifiers: a novel healthcare paradigm. *Multimedia Tools and Applications*, *83*(14), 40677-40712.

Gunasekaran, K. P. (2023). Exploring sentiment analysis techniques in natural language processing: A comprehensive review. *arXiv preprint arXiv:2305.14842*.

Guo, Y., Barnes, S. J., & Jia, Q. (2017). Mining meaning from online ratings and reviews: Tourist satisfaction analysis using latent dirichlet allocation. *Tourism management*, *59*, 467-483.

Gupta, S. A. C. H. I. N., McLaughlin, E. D. W. A. R. D., & Gomez, M. (2019). Guest satisfaction and restaurant performance. *The Next Frontier of Restaurant Management: Harnessing Data to Improve Guest Service and Enhance the Employee Experience*, 33-53.

Ha, N. M., Thang, T. K. V., Tam, H. L., Dang, T. H., & Hoa, N. T. M. (2023). Mediating role of corporate reputation in the relationship between corporate social responsibility and repurchase intention: A case study of McDonald's fast food restaurant chain. *HO CHI MINH CITY OPEN UNIVERSITY JOURNAL OF SCIENCE-ECONOMICS AND BUSINESS ADMINISTRATION*, *13*(2), 19-38.

Haixiang, G., Yijing, L., Shang, J., Mingyun, G., Yuanyue, H., & Bing, G. (2017). Learning from class-imbalanced data: Review of methods and applications. *Expert systems with applications*, *73*, 220-239.

Hall, J. A. (2018). When is social media use social interaction? Defining mediated social interaction. *New media & society*, *20*(1), 162-179.

Hernández-Orallo, J., Flach, P., & Ferri, C. (2012). A unified view of performance metrics: Translating threshold choice into expected classification loss. *The Journal of Machine Learning Research*, *13*(1), 2813-2869.

Hinton, G. E., Osindero, S., & Teh, Y. W. (2006). A fast learning algorithm for deep belief nets. *Neural computation*, *18*(7), 1527-1554.

Hinton, G., Deng, L., Yu, D., Dahl, G. E., Mohamed, A. R., Jaitly, N., ... & Kingsbury, B. (2012). Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups. *IEEE Signal processing magazine*, *29*(6), 82-97.

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural computation*, *9*(8), 1735-1780.

Holmquist, L. E. (2017). Intelligence on tap: artificial intelligence as a new design material. *interactions*, *24*(4), 28-33.

Hong, L., & Davison, B. D. (2010, July). Empirical study of topic modeling in twitter. In *Proceedings of the first workshop on social media analytics* (pp. 80-88).

Horowitz, M. C. (2018). Artificial intelligence, international competition, and the balance of power (May 2018).

Hosain, M. T., Jim, J. R., Mridha, M. F., & Kabir, M. M. (2024). Explainable AI approaches in deep learning: Advancements, applications and challenges. *Computers and electrical engineering*, *117*, 109246.

Hu, K., Li, Z., Liu, Y., Cheng, L., Yang, Q., & Li, Y. (2018). A Framework in CRM Customer Lifecycle: Identify Downward Trend and Potential Issues Detection. *arXiv preprint arXiv:1802.08974*.

Hu, E. J., Shen, Y., Wallis, P., Allen-Zhu, Z., Li, Y., Wang, S., ... & Chen, W. (2022). Lora: Low-rank adaptation of large language models. *ICLR*, *1*(2), 3.

Huang, W., Lin, M., & Wang, Y. (2022). Sentiment analysis of Chinese e-commerce product reviews using ERNIE word embedding and attention mechanism. *Applied Sciences*, *12*(14), 7182.

IBM. (n.d.). Supervised machine learning: Regression. Coursera. https://www.coursera.org/learn/supervised-machine-learning-regression

IBM. (n.d.). Supervised machine learning: Classification. Coursera. https://www.coursera.org/learn/supervised-machine-learning-classification

Islam, M. R., Ahmed, M. U., Barua, S., & Begum, S. (2022). A systematic review of explainable artificial intelligence in terms of different application domains and tasks. *Applied Sciences*, *12*(3), 1353.

Jagannatha, A. N., & Yu, H. (2016, November). Structured prediction models for RNN based sequence labeling in clinical text. In *Proceedings of the conference on empirical methods in natural language processing. conference on empirical methods in natural language processing* (Vol. 2016, p. 856).

Jain, A. K., Murty, M. N., & Flynn, P. J. (1999). Data clustering: a review. *ACM computing surveys (CSUR)*, *31*(3), 264-323.

Jain, V. K., & Kumar, S. (2020). Predictive analysis of emotions for improving customer services. In *Natural language processing: concepts, methodologies, tools, and applications* (pp. 808-817). IGI Global.

Jain, P. K., Pamula, R., & Srivastava, G. (2021). A systematic literature review on machine learning applications for consumer sentiment analysis using online reviews. *Computer science review*, *41*, 100413.

Jean, S., Cho, K., Memisevic, R., & Bengio, Y. (2014). On using very large target vocabulary for neural machine translation. *arXiv preprint arXiv:1412.2007*.

Jiang, F., Jiang, Y., Zhi, H., Dong, Y., Li, H., Ma, S., ... & Wang, Y. (2017). Artificial intelligence in healthcare: past, present and future. *Stroke and vascular neurology*, *2*(4).

Jiao, X., Yin, Y., Shang, L., Jiang, X., Chen, X., Li, L., ... & Liu, Q. (2019). Tinybert: Distilling bert for natural language understanding. *arXiv preprint arXiv:1909.10351*.

Jiarpakdee, J., Tantithamthavorn, C. K., Dam, H. K., & Grundy, J. (2020). An empirical study of model-agnostic techniques for defect prediction models. *IEEE Transactions on Software Engineering*, *48*(1), 166-185.

Jihad, K. H., Baker, M. R., Farhat, M., & Frikha, M. (2022, December). Machine learning-based social media text analysis: Impact of the rising fuel prices on electric vehicles. In *International conference on hybrid intelligent systems* (pp. 625-635). Cham: Springer Nature Switzerland.

Johri, P., Khatri, S. K., Al-Taani, A. T., Sabharwal, M., Suvanov, S., & Kumar, A. (2021). Natural language processing: History, evolution, application, and future work. In *Proceedings of 3rd International Conference on Computing Informatics and Networks: ICCIN 2020* (pp. 365-375). Springer Singapore.

Jordan, M. I., & Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and prospects. *Science*, *349*(6245), 255-260.

Kalchbrenner, N., Grefenstette, E., & Blunsom, P. (2014). A convolutional neural network for modelling sentences. *arXiv preprint arXiv:1404.2188*.

Kale, A., Nguyen, T., Harris Jr, F. C., Li, C., Zhang, J., & Ma, X. (2023). Provenance documentation to enable explainable and trustworthy AI: A literature review. *Data Intelligence*, *5*(1), 139-162.

Kang, H., Yoo, S. J., & Han, D. (2012). Senti-lexicon and improved Naïve Bayes algorithms for sentiment analysis of restaurant reviews. *Expert Systems with Applications*, *39*(5), 6000-6010.

Kasana, S. S., & Rathore, A. S. (2024). Attention-Based Light Weight Deep Learning Models for Early Potato Disease Detection. *Applied Sciences*, *14*(17), 8038.

Kee, D. M. H., Ho, S. L., Ho, Y. S., Lee, T. W., Ma, H., & Yin, Y. (2021). Critical success factors in the fast food industry: a case of McDonald's. *International Journal of Tourism and Hospitality in Asia Pasific*, *4*(2), 124-143.

Khan, S. H., Hayat, M., Bennamoun, M., Sohel, F. A., & Togneri, R. (2017). Cost-sensitive learning of deep feature representations from imbalanced data. *IEEE transactions on neural networks and learning systems*, *29*(8), 3573-3587.

Khomh, F., Adams, B., Cheng, J., Fokaefs, M., & Antoniol, G. (2018). Software engineering for machine-learning applications: The road ahead. *IEEE Software*, *35*(5), 81-84.

Kim, K., Lee, S. Y. T., & Assar, S. (2022). The dynamics of cryptocurrency market behavior: sentiment analysis using Markov chains. *Industrial Management & Data Systems*, *122*(2), 365-395.

Kok, C. L., Ho, C. K., Tan, F. K., & Koh, Y. Y. (2024). Machine learning-based feature extraction and classification of emg signals for intuitive prosthetic control. *Applied Sciences*, *14*(13), 5784.

Kotei, E., & Thirunavukarasu, R. (2023). A systematic review of transformer-based pre-trained language models through self-supervised learning. *Information*, *14*(3), 187.

Krawczyk, B., Galar, M., Jeleń, Ł., & Herrera, F. (2016). Evolutionary undersampling boosting for imbalanced classification of breast cancer malignancy. *Applied Soft Computing*, *38*, 714-726.

Kühl, N., Goutier, M., Baier, L., Wolff, C., & Martin, D. (2022). Human vs. supervised machine learning: Who learns patterns faster?. *Cognitive Systems Research*, *76*, 78-92.

Kumar, S., Roy, P. P., Dogra, D. P., & Kim, B. G. (2023). A comprehensive review on sentiment analysis: Tasks, approaches and applications. *arXiv preprint arXiv:2311.11250*.

Kumarakulasinghe, N. B., Blomberg, T., Liu, J., Leao, A. S., & Papapetrou, P. (2020, July). Evaluating local interpretable model-agnostic explanations on clinical machine learning classification models. In *2020 IEEE 33rd international symposium on computer-based medical systems (CBMS)* (pp. 7-12). IEEE.

Lan, Z., Chen, M., Goodman, S., Gimpel, K., Sharma, P., & Soricut, R. (2019). Albert: A lite bert for self-supervised learning of language representations. *arXiv preprint arXiv:1909.11942*.

Lancaster, L. (n.d.). *Web data research assistant*. University of Southampton. Retrieved May 20, 2025, from https://www.southampton.ac.uk/~lac/WebDataResearchAssistant/

Larsen, K. R., & Becker, D. S. (2021). *Automated machine learning for business*. Oxford University Press.

LeCun, Yann, Yoshua Bengio, and Geoffrey Hinton. "Deep learning." *nature* 521, no. 7553 (2015): 436-444.

LemaÃŽtre, G., Nogueira, F., & Aridas, C. K. (2017). Imbalanced-learn: A python toolbox to tackle the curse of imbalanced datasets in machine learning. *Journal of machine learning research*, *18*(17), 1-5.

Li, D., & Du, Y. (2017). *Artificial intelligence with uncertainty*. CRC press.

Li, Q., Li, S., Zhang, S., Hu, J., & Hu, J. (2019). A review of text corpus-based tourism big data mining. *Applied Sciences*, *9*(16), 3300.

Li, M., Zhang, X., Thrampoulidis, C., Chen, J., & Oymak, S. (2021). Autobalance: Optimized loss functions for imbalanced data. *Advances in Neural Information Processing Systems*, *34*, 3163-3177.

Li, F., Larimo, J., & Leonidou, L. C. (2023a). Social media in marketing research: Theoretical bases, methodological aspects, and thematic focus. *Psychology & Marketing*, *40*(1), 124-145.

Li, C., Wang, J., Zhang, Y., Zhu, K., Hou, W., Lian, J., ... & Xie, X. (2023b). Large language models understand and can be enhanced by emotional stimuli. *arXiv preprint arXiv:2307.11760*.

Li, Y., Chen, H., Yu, P., & Yang, L. (2025). A Review of Artificial Intelligence Applications in Architectural Design: Energy-Saving Renovations and Adaptive Building Envelopes. *Energies*, *18*(4), 918.

Lightstone, S. C. (2024). Train or Restrain? Using International Perspectives to Inform the American Fair Use Analysis of Copyright in Generative Artificial Intelligence Training. *Nw. J. Int'l L. & Bus.*, *44*, 471.

Liu, B., & Zhang, L. (2012). A survey of opinion mining and sentiment analysis. *Mining text data*, 415-463.

Liu, B. (2022). *Sentiment analysis and opinion mining*. Springer Nature.

Liu, R., Cao, J., Sun, N., & Jiang, L. (2022, July). Aspect feature distillation and enhancement network for aspect-based sentiment analysis. In *Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval* (pp. 1577-1587).

Liu, Y., Zhang, Y., Wang, Y., Hou, F., Yuan, J., Tian, J., ... & He, Z. (2023). A survey of visual transformers. *IEEE Transactions on Neural Networks and Learning Systems*.

Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. *Advances in neural information processing systems*, *30*.

Luo, S., Ivison, H., Han, S. C., & Poon, J. (2024). Local interpretations for explainable natural language processing: A survey. *ACM Computing Surveys*, *56*(9), 1-36.

Ma, J., Sheridan, R. P., Liaw, A., Dahl, G. E., & Svetnik, V. (2015). Deep neural nets as a method for quantitative structure–activity relationships. *Journal of chemical information and modeling*, *55*(2), 263-274.

Ma, D., Li, S., Wu, F., Xie, X., & Wang, H. (2019, July). Exploring sequence-to-sequence learning in aspect term extraction. In *Proceedings of the 57th annual meeting of the association for computational linguistics* (pp. 3538-3547).

Ma, F., Hu, X., Liu, A., Yang, Y., Li, S. A., Yu, P. S., & Wen, L. (2023, July). AMR-based network for aspect-based sentiment analysis. In *Proceedings of the 61st annual meeting of the association for computational linguistics (volume 1: long papers)* (pp. 322-337).

Madhoushi, Z., Hamdan, A. R., & Zainudin, S. (2015, July). Sentiment analysis techniques in recent works. In *2015 science and information conference (SAI)* (pp. 288-291). IEEE.

Mahadevaswamy, U. B., & Swathi, P. (2023). Sentiment analysis using bidirectional LSTM network. *Procedia Computer Science*, *218*, 45-56.

Mahesh, B. (2020). Machine learning algorithms-a review. *International Journal of Science and Research (IJSR).[Internet]*, *9*(1), 381-386.

Maitri, W. S., Suherlan, S., Prakosos, R. D. Y., Subagja, A. D., & Ausat, A. M. A. (2023). Recent trends in social media marketing strategy. *Jurnal Minfo Polgan*, *12*(1), 842-850.

Mann, M., Kumar, C., Zeng, W. F., & Strauss, M. T. (2021). Artificial intelligence for proteomics and biomarker discovery. *Cell systems*, *12*(8), 759-770.

Mao, Y., Zhang, Y., Jiao, L., & Zhang, H. (2022). Document-level sentiment analysis using attention-based bi-directional long short-term memory network and two-dimensional convolutional neural network. *Electronics*, *11*(12), 1906.

Mao, Y., Liu, Q., & Zhang, Y. (2024). Sentiment analysis methods, applications, and challenges: A systematic literature review. *Journal of King Saud University-Computer and Information Sciences*, 102048.

Mariotti, E., Sivaprasad, A., & Moral, J. M. A. (2023, July). Beyond prediction similarity: ShapGAP for evaluating faithful surrogate models in XAI. In *World conference on explainable artificial intelligence* (pp. 160-173). Cham: Springer Nature Switzerland.

Martínez-Fernández, S., Bogner, J., Franch, X., Oriol, M., Siebert, J., Trendowicz, A., ... & Wagner, S. (2022). Software engineering for AI-based systems: a survey. *ACM Transactions on Software Engineering and Methodology (TOSEM)*, *31*(2), 1-59.

Maynard, D. G., & Greenwood, M. A. (2014, March). Who cares about sarcastic tweets? investigating the impact of sarcasm on sentiment analysis. In *Lrec 2014 proceedings*. ELRA.

McCarthy, J., Minsky, M. L., Rochester, N., & Shannon, C. E. (2006). A proposal for the dartmouth summer research project on artificial intelligence, august 31, 1955. *AI magazine*, *27*(4), 12-12.

Medhat, W., Hassan, A., & Korashy, H. (2014). Sentiment analysis algorithms and applications: A survey. *Ain Shams engineering journal*, *5*(4), 1093-1113.

Mekni, M., Baani, Z., & Sulieman, D. (2020, January). A smart virtual assistant for students. In *Proceedings of the 3rd international conference on applications of intelligent systems* (pp. 1-6).

Meyer, C., & Schwager, A. (2007). Understanding customer experience. *Harvard business review*, *85*(2), 116.

Mikolov, T. (2012). Statistical language models based on neural networks. *Presentation at Google, Mountain View, 2nd April*, *80*(26).

Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., & Dean, J. (2013a). Distributed representations of words and phrases and their compositionality. *Advances in neural information processing systems*, *26*.

Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013b). Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*.

Miller, T. (2019). Explanation in artificial intelligence: Insights from the social sciences. *Artificial intelligence*, *267*, 1-38.

Min, B., Ross, H., Sulem, E., Veyseh, A. P. B., Nguyen, T. H., Sainz, O., ... & Roth, D. (2023). Recent advances in natural language processing via large pre-trained language models: A survey. *ACM Computing Surveys*, *56*(2), 1-40.

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.

Mitra, A., & Mohanty, S. (2020). Sentiment analysis using machine learning approaches. *Emerging Technologies in Data Mining and Information Security: Proceedings of IEMIS*, *2*, 63-68.

Mohamed, N., & Al-Jaroodi, J. (2014, July). Real-time big data analytics: Applications and challenges. In *2014 international conference on high performance computing & simulation (HPCS)* (pp. 305-310). IEEE

Moreno, A., & Redondo, T. (2016). Text analytics: the convergence of big data and artificial intelligence. *IJIMAI*, *3*(6), 57-64.

Moudhich, I., & Fennan, A. (2024). Graph embedding approach to analyze sentiments on cryptocurrency. *International Journal of Electrical & Computer Engineering (2088-8708)*, *14*(1).

Muhamad, N., Khamarudin, M., & Fauzi, W. I. M. (2018). The role of religious motivation in an international consumer boycott. *British Food Journal*, *121*(1), 199-217.

Mujahid, M., Kına, E. R. O. L., Rustam, F., Villar, M. G., Alvarado, E. S., De La Torre Diez, I., & Ashraf, I. (2024). Data oversampling and imbalanced datasets: An investigation of performance for machine learning and feature engineering. *Journal of Big Data*, *11*(1), 87.

Mulyo, I. A. (2023). Value chain analysis of McDonald. *Jurnal Ekonomi*, *12*(02), 908-913.

Myles, A. J., Feudale, R. N., Liu, Y., Woody, N. A., & Brown, S. D. (2004). An introduction to decision tree modeling. *Journal of Chemometrics: A Journal of the Chemometrics Society*, *18*(6), 275-285.

Nasteski, V. (2017). An overview of the supervised machine learning methods. *Horizons. b*, *4*(51-62), 56.

Negnevitsky, M. (2005). *Artificial intelligence: a guide to intelligent systems*. Pearson education.

Nguyen, H. H., Viviani, J. L., & Ben Jabeur, S. (2023). Bankruptcy prediction using machine learning and Shapley additive explanations. *Review of Quantitative Finance and Accounting*, 1-42.

Noble, W. S. (2006). What is a support vector machine?. *Nature biotechnology*, *24*(12), 1565-1567.

Neveditsin, N., Lingras, P., & Mago, V. (2025). From Annotation to Adaptation: Metrics, Synthetic Data, and Aspect Extraction for Aspect-Based Sentiment Analysis with Large Language Models. *arXiv preprint arXiv:2503.20715*.

Obiedat, R., Qaddoura, R., Ala'M, A. Z., Al-Qaisi, L., Harfoushi, O., Alrefai, M. A., & Faris, H. (2022). Sentiment analysis of customers' reviews using a hybrid evolutionary SVM-based approach in an imbalanced data distribution. *IEEE Access*, *10*, 22260-22273

Okeke, N. I., Alabi, O. A., Igwe, A. N., Ofodile, O. C., & Ewim, C. P. M. (2024). AI in customer feedback integration: A data-driven framework for enhancing business strategy. *World J. Advanced Res. Reviews*, *24*(1), 3207-3220.

Otter, D. W., Medina, J. R., & Kalita, J. K. (2020). A survey of the usages of deep learning for natural language processing. *IEEE transactions on neural networks and learning systems*, *32*(2), 604-624.

Owusu-Adjei, M., Ben Hayfron-Acquah, J., Frimpong, T., & Abdul-Salaam, G. (2023). Imbalanced class distribution and performance evaluation metrics: A systematic review of prediction accuracy for determining model performance in healthcare systems. *PLOS Digital Health*, *2*(11), e0000290.

Oxford University Press. (n.d.). Sharp. In Oxford Learner's Dictionaries. Retrieved March 30, 2025, from https://www.oxfordlearnersdictionaries.com/definition/english/sharp_1

Owusu-Adjei, M., Ben Hayfron-Acquah, J., Frimpong, T., & Abdul-Salaam, G. (2023). Imbalanced class distribution and performance evaluation metrics: A systematic review of prediction accuracy for determining model performance in healthcare systems. *PLOS Digital Health*, *2*(11), e0000290.

Palatnik de Sousa, I., Maria Bernardes Rebuzzi Vellasco, M., & Costa da Silva, E. (2019). Local interpretable model-agnostic explanations for classification of lymph node metastases. *Sensors*, *19*(13), 2969.

Pascanu, R., Gulcehre, C., Cho, K., & Bengio, Y. (2013). How to construct deep recurrent neural networks. *arXiv preprint arXiv:1312.6026*.

Patil, R. S., & Kolhe, S. R. (2022). Supervised classifiers with TF-IDF features for sentiment analysis of Marathi tweets. *Social Network Analysis and Mining*, *12*(1), 51.

Pennington, J., Socher, R., & Manning, C. D. (2014, October). Glove: Global vectors for word representation. In *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)* (pp. 1532-1543).

Perikos, I., & Diamantopoulos, A. (2024). Explainable Aspect-Based Sentiment Analysis Using Transformer Models. *Big Data and Cognitive Computing*, *8*(11), 141.

Perlovsky, L. I. (2006). Toward physics of the mind: Concepts, emotions, consciousness, and symbols. *Physics of Life Reviews*, *3*(1), 23-55.

Petch, J., Di, S., & Nelson, W. (2022). Opening the black box: the promise and limitations of explainable machine learning in cardiology. *Canadian Journal of Cardiology*, *38*(2), 204-213.

Phan, H. T., Nguyen, N. T., & Hwang, D. (2023). Aspect-level sentiment analysis: A survey of graph convolutional network methods. *Information Fusion*, *91*, 149-172.

Piao, Y., Piao, M., Jin, C. H., Shon, H. S., Chung, J. M., Hwang, B., & Ryu, K. H. (2015). A new ensemble method with feature space partitioning for high-dimensional data classification. *Mathematical Problems in Engineering*, *2015*(1), 590678.

Pinaya, W. H., Graham, M. S., Kerfoot, E., Tudosiu, P. D., Dafflon, J., Fernandez, V., ... & Cardoso, M. J. (2023). Generative ai for medical imaging: extending the monai framework. *arXiv preprint arXiv:2307.15208*.

Powers, D. M. (2020). Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation. *arXiv preprint arXiv:2010.16061*.

Pradhipta, A. Y., Yonita, L., & Novaria, S. A. P. (2024, December). Crisis Communication Strategy in the Face of Consumer Boycott: Exploring the Boycott Case of Israel-affiliated Companies and Brands. In *6th Open Society Conference (OSC 2024)* (pp. 140-152). Atlantis Press.

Pradipta, G. A., Wardoyo, R., Musdholifah, A., Sanjaya, I. N. H., & Ismail, M. (2021, November). SMOTE for handling imbalanced data problem: A review. In *2021 sixth international conference on informatics and computing (ICIC)* (pp. 1-8). IEEE.

Qiu, Z., Xu, Y., Chen, C., Zhou, W., & Yu, G. (2024). Enhanced Disease Detection for Apple Leaves with Rotating Feature Extraction. *Agronomy*, *14*(11), 2602.

Qu, C., Yang, L., Qiu, M., Croft, W. B., Zhang, Y., & Iyyer, M. (2019, July). BERT with history answer embedding for conversational question answering. In *Proceedings of the 42nd international ACM SIGIR conference on research and development in information retrieval* (pp. 1133-1136).

Radford, A., Narasimhan, K., Salimans, T., & Sutskever, I. (2018). Improving language understanding by generative pre-training.

Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., & Sutskever, I. (2019). Language models are unsupervised multitask learners. *OpenAI blog*, *1*(8), 9.

Rainio, O., Teuho, J., & Klén, R. (2024). Evaluation metrics and statistical tests for machine learning. *Scientific Reports*, *14*(1), 6086.

Ramírez-Tinoco, F. J., Alor-Hernández, G., Sánchez-Cervantes, J. L., Salas-Zárate, M. D. P., & Valencia-García, R. (2019). Use of sentiment analysis techniques in healthcare domain. *Current trends in semantic web technologies: theory and practice*, 189-212.

Rane, N. (2023). Enhancing customer loyalty through Artificial Intelligence (AI), Internet of Things (IoT), and Big Data technologies: improving customer satisfaction, engagement, relationship, and experience. *Internet of Things (IoT), and Big Data Technologies: Improving Customer Satisfaction, Engagement, Relationship, and Experience (October 13, 2023)*.

Ranjan, M., Tiwari, S., Sattar, A. M., & Tatkar, N. S. (2024). A New Approach for Carrying Out Sentiment Analysis of Social Media Comments Using Natural Language Processing. *Engineering Proceedings*, *59*(1), 181.

Ransom, A., LaGrant, B., Spiteri, A., Kushnir, T., Anderson, A. K., & De Rosa, E. (2022). Face-to-face learning enhances the social transmission of information. *Plos one*, *17*(2), e0264250

Ravi, V., & Kamaruddin, S. (2017). Big data analytics enabled smart financial services: opportunities and challenges. In *Big Data Analytics: 5th International Conference, BDA 2017, Hyderabad, India, December 12-15, 2017, Proceedings 5* (pp. 15-39). Springer International Publishing.

Reichheld, F. F., & Schefter, P. (2000). E-loyalty: your secret weapon on the web. *Harvard business review*, *78*(4), 105-113.

Reswara, A. A., Kamiilah, N. S., Aprilia, M., Kusuma, S. O., Mardianto, M. F. F., & Kurniawan, A. (2024). Analysis of Indonesian Public Perception on the Influence of American Food Brands with the Indonesia-America Cooperation Relationship Using SEM-PLS. *Journal of Data Science*, *2024*.

Ribeiro, M. T., Singh, S., & Guestrin, C. (2016, August). " Why should i trust you?" Explaining the predictions of any classifier. In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining* (pp. 1135-1144).

Rodríguez-Ibánez, M., Casánez-Ventura, A., Castejón-Mateos, F., & Cuenca-Jiménez, P. M. (2023). A review on sentiment analysis from social media platforms. *Expert Systems with Applications*, *223*, 119862.

Rognone, L., Hyde, S., & Zhang, S. S. (2020). News sentiment in the cryptocurrency market: An empirical comparison with Forex. *International Review of Financial Analysis*, *69*, 101462.

Rozi, I. F., Arianto, R., Yunianto, D. R., Ananta, A. Y., & Rahmawati, S. (2024, September). Enhancing Aspect-Based Sentiment Analysis for Radio Station Public Opinion: Evaluating Preprocessing Strategies and Imbalanced Data Handling. In *2024 International Conference on Electrical and Information Technology (IEIT)* (pp. 103-108). IEEE.

Rude, D. J., Adams, S., & Beling, P. A. (2018). Task recognition from joint tracking data in an operational manufacturing cell. *Journal of Intelligent Manufacturing*, *29*, 1203-1217.

Saadati, P., Husamaldin, L., & Abdelnour-Nocera, J. (2024). Using aspect-based sentiment analysis as a tool for socio-technical system design.

Samek, W., & Müller, K. R. (2019). Towards explainable artificial intelligence. In *Explainable AI: interpreting, explaining and visualizing deep learning* (pp. 5-22). Cham: Springer International Publishing.

Samudra, F., Zaman, A. R. B., & Mukti, D. A. (2024). Reviving the essence of jihad in the contemporary era: advocating struggle and solidarity for Palestine through boycott of pro-Israel products. *MILRev: Metro Islamic Law Review*, *3*(1), 22-42.

Samuel, A. L. (1959). Some studies in machine learning using the game of checkers. *IBM Journal of research and development*, *3*(3), 210-229.

Sanh, V., Debut, L., Chaumond, J., & Wolf, T. (2019). DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter. *arXiv preprint arXiv:1910.01108*.

Sankar, H., & Subramaniyaswamy, V. (2017, December). Investigating sentiment analysis using machine learning approach. In *2017 International conference on intelligent sustainable systems (ICISS)* (pp. 87-92). IEEE.

Saputra, M. A., & Setiawan, E. B. (2023, August). Aspect Based Sentiment Analysis Using Recurrent Neural Networks (RNN) on Social Media Twitter. In *2023 International Conference on Data Science and Its Applications (ICoDSA)* (pp. 265-270). IEEE.

Sarker, I. H. (2021a). Machine learning: Algorithms, real-world applications and research directions. *SN computer science*, *2*(3), 160.

Sarker, I. H. (2021b). CyberLearning: Effectiveness analysis of machine learning security modeling to detect cyber-anomalies and multi-attacks. *Internet of Things*, *14*, 100393.

Sarker, I. H. (2023). Machine learning for intelligent data analysis and automation in cybersecurity: current and future prospects. *Annals of Data Science*, *10*(6), 1473-1498.

Saxena, N. K., & Dhar, U. (2017). Brand image: A seasoned review. *Asian J. Management*, *8*(1), 103-106.

Scaria, K., Gupta, H., Goyal, S., Sawant, S. A., Mishra, S., & Baral, C. (2023). Instructabsa: Instruction learning for aspect based sentiment analysis. *arXiv preprint arXiv:2302.08624*.

Schmidhuber, J. (2015). Deep learning in neural networks: An overview. *Neural networks*, *61*, 85-117.

Schuster, M., & Paliwal, K. K. (1997). Bidirectional recurrent neural networks. *IEEE transactions on Signal Processing*, *45*(11), 2673-2681.

Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., & Batra, D. (2017). Grad-cam: Visual explanations from deep networks via gradient-based localization. In *Proceedings of the IEEE international conference on computer vision* (pp. 618-626).

Sathyan, R., Parthiban, P., Dhanalakshmi, R., & Minz, A. (2021). A combined big data analytics and Fuzzy DEMATEL technique to improve the responsiveness of automotive supply chains. *Journal of Ambient Intelligence and Humanized Computing*, *12*, 7949-7963.

Schroeder, J., Holzner, D., Berger, C., Hoel, C. J., Laine, L., & Magnusson, A. (2015, May). Design and evaluation of a customizable multi-domain reference architecture on top of product lines of self-driving heavy vehicles-an industrial case study. In *2015 IEEE/ACM 37th IEEE International Conference on Software Engineering* (Vol. 2, pp. 189-198). IEEE.

Seo, E. J., & Park, J. W. (2018). A study on the effects of social media marketing activities on brand equity and customer response in the airline industry. *Journal of Air Transport Management*, *66*, 36-41.

Serrat, O. (2017). *Knowledge solutions: Tools, methods, and approaches to drive organizational performance* (p. 1140). Springer Nature.

Shams Khoozani, Z., Sabri, A. Q. M., Seng, W. C., Seera, M., & Eg, K. Y. (2024). Navigating the landscape of concept-supported XAI: Challenges, innovations, and future directions. *Multimedia Tools and Applications*, *83*(25), 67147-67197.

Sharifani, K., & Amini, M. (2023). Machine learning and deep learning: A review of methods and applications. *World Information Technology and Engineering Journal*, *10*(07), 3897-3904

Sharma, S. (2024). Benefits or concerns of AI: A multistakeholder responsibility. *Futures*, *157*, 103328.

Shmueli, G., Bruce, P. C., Yahav, I., Patel, N. R., & Lichtendahl Jr, K. C. (2017). *Data mining for business analytics: concepts, techniques, and applications in R*. John Wiley & Sons.

Siami-Namini, S., Tavakoli, N., & Namin, A. S. (2019). A comparative analysis of forecasting financial time series using arima, lstm, and bilstm. *arXiv preprint arXiv:1911.09512*.

Smith, A., & Brenner, J. (2012). Twitter use 2012. *Pew internet & American life project*, *4*, 1-12.

Steinert-Threlkeld, Z. C. (2018). *Twitter as data*. Cambridge University Press.

Socher, R., Perelygin, A., Wu, J., Chuang, J., Manning, C. D., Ng, A. Y., & Potts, C. (2013, October). Recursive deep models for semantic compositionality over a sentiment treebank. In *Proceedings of the 2013 conference on empirical methods in natural language processing* (pp. 1631-1642).

Song, Y., Wang, J., Liang, Z., Liu, Z., & Jiang, T. (2020). Utilizing BERT intermediate layers for aspect based sentiment analysis and natural language inference. *arXiv preprint arXiv:2002.04815*.

Staliūnaitė, I., & Iacobacci, I. (2020). Compositional and lexical semantics in RoBERTa, BERT and DistilBERT: A case study on CoQA. *arXiv preprint arXiv:2009.08257*.

Steinert-Threlkeld, Z. C. (2018). *Twitter as data*. Cambridge University Press.

Sturges, D. L. (1994). Communicating through crisis: A strategy for organizational survival. *Management communication quarterly*, *7*(3), 297-316.

Su, J., Chen, Q., Wang, Y., Zhang, L., Pan, W., & Li, Z. (2023). Sentence-level sentiment analysis based on supervised gradual machine learning. *Scientific Reports*, *13*(1), 14500.

Sudirjo, F., Diantoro, K., Al-Gasawneh, J. A., Azzaakiyyah, H. K., & Ausat, A. M. A. (2023). Application of ChatGPT in improving customer sentiment analysis for businesses. *Jurnal Teknologi Dan Sistem Informasi Bisnis*, *5*(3), 283-288.

Sun, Y., Wang, S., Li, Y., Feng, S., Chen, X., Zhang, H., ... & Wu, H. (2019). Ernie: Enhanced representation through knowledge integration. *arXiv preprint arXiv:1904.09223*.

Sun, X., Yang, D., Li, X., Zhang, T., Meng, Y., Qiu, H., ... & Li, J. (2021). Interpreting deep learning models in natural language processing: A review. *arXiv preprint arXiv:2110.10470*.

Sundararajan, M., Taly, A., & Yan, Q. (2017, July). Axiomatic attribution for deep networks. In *International conference on machine learning* (pp. 3319-3328). PMLR.

Susnjak, T. (2024). Applying bert and chatgpt for sentiment analysis of lyme disease in scientific literature. In *Borrelia burgdorferi: methods and protocols* (pp. 173-183). New York, NY: Springer Us.

Sutskever, I., Martens, J., & Hinton, G. E. (2011). Generating text with recurrent neural networks. In *Proceedings of the 28th international conference on machine learning (ICML-11)* (pp. 1017-1024).

Sutskever, I., Vinyals, O., & Le, Q. V. (2014). Sequence to sequence learning with neural networks. *Advances in neural information processing systems*, *27*.

Taboada, M., Brooke, J., Tofiloski, M., Voll, K., & Stede, M. (2011). Lexicon-based methods for sentiment analysis. *Computational linguistics*, *37*(2), 267-307.

Talaat, F. M., Aljadani, A., Badawy, M., & Elhosseini, M. (2024). Toward interpretable credit scoring: integrating explainable artificial intelligence with deep learning for credit card default prediction. *Neural Computing and Applications*, *36*(9), 4847-4865.

Tan, X., Cai, Y., Xu, J., Leung, H. F., Chen, W., & Li, Q. (2020). Improving aspect-based sentiment analysis via aligning aspect embedding. *Neurocomputing*, *383*, 336-347.

Tan, X. W., & Kok, S. (2024). Explainable risk classification in financial reports. *arXiv preprint arXiv:2405.01881*.

Tarigan, I. M., Harahap, M. A. K., Sari, D. M., Sakinah, R. D., & Ausat, A. M. A. (2023). Understanding Social Media: Benefits of Social Media for Individuals. *Jurnal Pendidikan Tambusai*, *7*(1), 2317-2322.

Tarwani, K. M., & Edem, S. (2017). Survey on recurrent neural network in natural language processing. *Int. J. Eng. Trends Technol*, *48*(6), 301-304.

Tenney, I., Das, D., & Pavlick, E. (2019). BERT rediscovers the classical NLP pipeline. *arXiv preprint arXiv:1905.05950*.

Thakkar, H., & Patel, D. (2015). Approaches for sentiment analysis on twitter: A state-of-art study. *arXiv preprint arXiv:1512.01043*.

Theosaksomo, D., & Widyantoro, D. H. (2019, October). Conversational recommender system chatbot based on functional requirement. In *2019 IEEE 13th International Conference on Telecommunication Systems, Services, and Applications (TSSA)* (pp. 154-159). IEEE.

Tian, Y., Liu, C., Song, Y., Xia, F., & Zhang, Y. (2024, June). Aspect-based Sentiment Analysis with Context Denoising. In *Findings of the Association for Computational Linguistics: NAACL 2024* (pp. 3083-3095).

Tizghadam, A., Khazaei, H., Moghaddam, M. H., & Hassan, Y. (2019). Machine Learning in Transportation. *Journal of Advanced Transportation*.

Trappl, R. (1986). *Impacts of artificial intelligence*. North Holland Publishing Co..

Tsai, H., Riesa, J., Johnson, M., Arivazhagan, N., Li, X., & Archer, A. (2019). Small and practical BERT models for sequence labeling. *arXiv preprint arXiv:1909.00100*.

Tsujii, J. (2021). Natural language processing and computational linguistics. *Computational Linguistics*, *47*(4), 707-727.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). Attention is all you need. *Advances in neural information processing systems*, *30*.

Walia, S., Kumar, K., Agarwal, S., & Kim, H. (2022). Using xai for deep learning-based image manipulation detection with shapley additive explanation. *Symmetry*, *14*(8), 1611.

Wang, H., Ma, C., & Zhou, L. (2009, December). A brief review of machine learning and its application. In *2009 international conference on information engineering and computer science* (pp. 1-4). IEEE.

Wang, H., Lei, Z., Zhang, X., Zhou, B., & Peng, J. (2016). Machine learning basics. *Deep learning*, 98-164.

Wang, G., Li, C., Wang, W., Zhang, Y., Shen, D., Zhang, X., ... & Carin, L. (2018). Joint embedding of words and labels for text classification. *arXiv preprint arXiv:1805.04174*.

Wang, L. (2019). From intelligence science to intelligent manufacturing. *Engineering*, *5*(4), 615-618.

Wang, Y., Chen, Q., & Wang, W. (2021, August). Multi-task BERT for aspect-based sentiment analysis. In *2021 IEEE international conference on smart computing (SMARTCOMP)* (pp. 383-385). IEEE.

Wang, Z., Hou, Y., Zeng, C., Zhang, S., & Ye, R. (2023). Multiple learning features–enhanced knowledge tracing based on learner–resource response channels. *Sustainability*, *15*(12), 9427.

Wang, S., Zhou, J., Chen, Q., Zhang, Q., Gui, T., & Huang, X. (2024a). Domain generalization via causal adjustment for cross-domain sentiment analysis. *arXiv preprint arXiv:2402.14536*.

Wang, Z., Zhang, B., Yang, R., Guo, C., & Li, M. (2024b, June). DAGCN: Distance-based and Aspect-oriented Graph Convolutional Network for Aspect-based Sentiment Analysis. Association for Computational Linguistics.

Wantara, P., & Tambrin, M. (2019). The Effect of price and product quality towards customer satisfaction and customer loyalty on madura batik. *International Tourism and Hospitality Journal*, *2*(1), 1-9.

Webb, G. I., Keogh, E., & Miikkulainen, R. (2010). Naïve Bayes. *Encyclopedia of machine learning*, *15*(1), 713-714.

Weerakody, P. B., Wong, K. W., Wang, G., & Ela, W. (2021). A review of irregular time series data handling with gated recurrent neural networks. *Neurocomputing*, *441*, 161-178.

Weller, K., Bruns, A., Burgess, J., Mahrt, M., & Puschmann, C. (2014). Twitter and society: An introduction. *Twitter and society [Digital Formations, Volume 89]*, xxix-xxxviii.

Wen, J., Zhang, G., Zhang, H., Yin, W., & Ma, J. (2020). Speculative text mining for document-level sentiment classification. *Neurocomputing*, *412*, 52-62.

Wolf, T., Debut, L., Sanh, V., Chaumond, J., Delangue, C., Moi, A., ... & Rush, A. M. (2020, October). Transformers: State-of-the-art natural language processing. In *Proceedings of the 2020 conference on empirical methods in natural language processing: system demonstrations* (pp. 38-45).

Wu, L., Morstatter, F., & Liu, H. (2018). SlangSD: building, expanding and using a sentiment dictionary of slang words for short-text sentiment classification. *Language Resources and Evaluation*, *52*, 839-852.

Xin, C., & Zakaria, L. Q. (2024). Integrating Bert with CNN and Bilstm for Explainable Detection of Depression in Social Media Contents. *IEEE Access*.

Xing, F. Z., Cambria, E., & Welsch, R. E. (2018). Natural language based financial forecasting: a survey. *Artificial Intelligence Review*, *50*(1), 49-73.

Xiong, Hui Y., Babak Alipanahi, Leo J. Lee, Hannes Bretschneider, Daniele Merico, Ryan KC Yuen, Yimin Hua et al. "The human splicing code reveals new insights into the genetic determinants of disease." *Science* 347, no. 6218 (2015): 1254806.

Xiong, Y., Chen, G., & Cao, J. (2024). Research on Public Service Request Text Classification Based on BERT-BiLSTM-CNN Feature Fusion. *Applied Sciences (2076-3417)*, *14*(14).

Xu, J., Ren, X., Lin, J., & Sun, X. (2018). Dp-gan: diversity-promoting generative adversarial network for generating informative and diversified text. *arXiv preprint arXiv:1802.01345*.

Xu, H., Liu, B., Shu, L., & Yu, P. S. (2019). BERT post-training for review reading comprehension and aspect-based sentiment analysis. *arXiv preprint arXiv:1904.02232*.

Xu, B., Wang, X., Yang, B., & Kang, Z. (2020, April). Target embedding and position attention with lstm for aspect based sentiment analysis. In *Proceedings of the 2020 5th International Conference on Mathematics and Artificial Intelligence* (pp. 93-97).

Yadollahi, A., Shahraki, A. G., & Zaiane, O. R. (2017). Current state of text sentiment analysis from opinion to emotion mining. *ACM Computing Surveys (CSUR)*, *50*(2), 1-33.

Yazdani, M. (2021). Machine learning, deep learning, and hedonic methods for real estate price prediction. *arXiv preprint arXiv:2110.07151*.

Yang, Z., Dai, Z., Yang, Y., Carbonell, J., Salakhutdinov, R. R., & Le, Q. V. (2019). Xlnet: Generalized autoregressive pretraining for language understanding. *Advances in neural information processing systems*, *32*.

Yang, H., Zhang, C., & Li, K. (2023, October). Pyabsa: A modularized framework for reproducible aspect-based sentiment analysis. In *Proceedings of the 32nd ACM international conference on information and knowledge management* (pp. 5117-5122).

Yannopoulou, N., Koronis, E., & Elliott, R. (2011). Media amplification of a brand crisis and its affect on brand trust. *Journal of Marketing Management*, *27*(5-6), 530-546.

Zafar, M. R., & Khan, N. M. (2019). DLIME: A deterministic local interpretable model-agnostic explanations approach for computer-aided diagnosis systems. *arXiv preprint arXiv:1906.10263*.

Zeberga, M. A., & Gebremariam, T. S. (2022). Imbalanced data handling in multi-label aspect categorization using oversampling and ensemble learning. Computational Intelligence and Neuroscience, 2022, 1–16. https://doi.org/10.1155/2022/9684516

Zhang, L., Wang, S., & Liu, B. (2018). Deep learning for sentiment analysis: A survey. *Wiley interdisciplinary reviews: data mining and knowledge discovery*, *8*(4), e1253.

Zhang, W., Li, X., Deng, Y., Bing, L., & Lam, W. (2022). A survey on aspect-based sentiment analysis: Tasks, methods, and challenges. *IEEE Transactions on Knowledge and Data Engineering*, *35*(11), 11019-11038.

Zhang, M., Zhu, Y., Liu, Z., Bao, Z., Wu, Y., Sun, X., & Xu, L. (2023, July). Span-level aspect-based sentiment analysis via table filling. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)* (pp. 9273-9284).

Zhao, W. X., Zhou, K., Li, J., Tang, T., Wang, X., Hou, Y., ... & Wen, J. R. (2023). A survey of large language models. *arXiv preprint arXiv:2303.18223*, *1*(2).

Zhou, C., Sun, C., Liu, Z., & Lau, F. (2015). A C-LSTM neural network for text classification. *arXiv preprint arXiv:1511.08630*.

Zhou, J., Huang, J. X., Hu, Q. V., & He, L. (2020). Sk-gcn: Modeling syntax and knowledge via graph convolutional network for aspect-level sentiment classification. *Knowledge-Based Systems*, *205*, 106292.

Zhou, T., & Law, K. M. (2022). Semantic relatedness enhanced graph network for aspect category sentiment analysis. *Expert Systems with Applications*, *195*, 116560.

Zhou, L., Tan, J., Fu, J., & Shao, G. (2025). Fast 3D Transmission Tower Detection Based on Virtual Views. *Applied Sciences*, *15*(2), 947.

Zhao, X., Zhou, Y., Xu, X., & Liu, Y. (2024). Extensible Multi-Granularity Fusion Network for Aspect-based Sentiment Analysis. *arXiv preprint arXiv:2402.07787*.

Zou, J., Han, Y., & So, S. S. (2009). Overview of artificial neural networks. *Artificial neural networks: methods and applications*, 14-22.

Zhu, J., Xia, Y., Wu, L., He, D., Qin, T., Zhou, W., ... & Liu, T. Y. (2020). Incorporating bert into neural machine translation. *arXiv preprint arXiv:2002.06823*.