ROLE OF ARTIFICIAL INTELLIGENCE IN SENTIMENT ANALYSIS AS STATE-OF-THE-ART IN FUTURE BUSINESS SCENARIOS

by

Sanjay Kumar Sahoo

DISSERTATION

Presented to the Swiss School of Business and Management, Geneva

In Partial Fulfillment

Of the Requirements

For the Degree

DOCTOR OF BUSINESS ADMINISTRATION

SWISS SCHOOL OF BUSINESS AND MANAGEMENT GENEVA

MAY, 2025

ROLE OF ARTIFICIAL INTELLIGENCE IN SENTIMENT ANALYSIS AS STATE-OF-THE-ART IN FUTURE BUSINESS SCENARIOS

by

Sanjay Kumar Sahoo

Supervised by

Dr. Dragan Perakovic

APPROVED BY

Dissertation chair

RECEIVED/APPROVED BY:

Adminisions Director

Dedication

This Dissertation is devoted to all upcoming business leaders whose success hinges on whose success hinges on navigating social media feedback and comments.

Acknowledgements

Throughout my doctoral journey, I am deeply grateful to a multitude of individuals who have supported me along the way. Foremost, I express my gratitude to God for granting me the strength and guidance to complete this doctoral degree. Amidst the challenges and demands of my rigorous schedule, I am humbled by the unwavering protection and blessings bestowed upon me, ensuring my safety and well-being.

Secondly, I extend my heartfelt appreciation to my supervisor and dissertation chair, Professor Dragan Perakovic, Ph.D. for his unwavering support, invaluable guidance, constructive feedback, and encouragement throughout the development of this dissertation. Professor, your mentorship has been instrumental in shaping this research journey, and I am profoundly grateful for your contributions. Completing this research work would have been an insurmountable challenge without your expert guidance.

I am also deeply appreciative of SSBM for providing me with the opportunity to pursue my studies at a prestigious business school. I extend my gratitude to all the doctoral students and colleagues whose contributions enriched my academic experience.

Special thanks to Dr. Deepali Bhatnagar for being an integral part of my educational journey. Thank you so much for your guidance and inspiration to me to make this dissertation completed successfully.

Furthermore, I dedicate this doctoral degree to my mom, Annapurna, my dad, Laxmidhar, my spouse, Reshma, my son, Vivaan, and my brothers, Bijaya and Dillip, whose unwavering support has been instrumental throughout my academic journey. Their steadfast commitment and moral encouragement have been indispensable in completing this doctoral research.

Last but not least, I must also thank to the participants in the research for their time and patience with me for those long interviews, follow-up questionnaires and numerous discussions. Without their sincere support, I could never finish the research work. Thank you all again.

ABSTRACT

ROLE OF ARTIFICIAL INTELLIGENCE IN SENTIMENT ANALYSIS AS STATE-OF-THE-ART IN FUTURE BUSINESS SCENARIOS

Sanjay Kumar Sahoo
2025

Dissertation Chair: Aleksandar Erceg, PhD

Any organization, whether it is focused on business or products, must design scalable and efficient sentiment analysis tools in light of the increasing reliance on digital data. Businesses of all sizes must analyze sentiment in order to understand and quickly address customer opinions and comments on a variety of subjects. Businesses depend on feedback to understand customer emotions precisely after improve their processes within marketing and customer service as well as product development. Sentiment analysis using artificial intelligence, namely machine learning, deep learning, and Natural Language Processing (NLP), is essential for text sentiment categorization as positive, negative, or neutral. This study examines the use of Artificial Intelligence specifically in sentiment analysis (SA) with emphasis on Distil BERT and AlBERT, advanced Natural Language Processing models on Flipkart product reviews sentiment classification. The raw data, containing over 200 thousand records of customers' feedback, has fields such as product name, price,

rating, review text, summary, and sentiment. Despite its simple and intuitive nature, the preparation of the given data involves various preprocessing steps like normalization, lemmatization, stop word removal, and data balancing through resampling methods. Both Distil BERT and Albert are trained and tested by using certain benchmarks (accuracy, precision, recall, and F1score) for measuring their performance. Distil BERT had the highest performance with an accuracy of 94.90%, precision 94.91%, recall 94.90%, and F1score of 94.90%, followed by Albert with an accuracy of 93.33% across the evaluation metrics. These models were also characterized by good percentage linger around the ROC curves and nice error matrix, which confirmed the good classification capability between positive, neutral, and negative sentiments. In contrast, traditional ML models like Random Forest and SVM yielded significantly lower accuracy scores of 75.52% and 75.80%, respectively. The research demonstrates how transformer-based models may be used in future business situations for efficient sentiment analysis and product review assessment, allowing companies to use consumer input to improve customer satisfaction and decision-making

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

Abbreviations	Full Form
NLP	Natural Language Processing
SA	Sentiment Analysis
SVM	Support Vector Machine
WOM	Word Of Mouth
ABSA	Aspect-Based Sentiment Analysis
ML	Machine Learning
DL	Deep Learning
SOM	Self-Organizing Map
NSP	Next Sentence Prediction
NLTK	Natural Language Toolkit
EDA	Exploratory Data Analysis
XAI	Explainable AI
DBN	Deep Belief Network
RNN	Recursive Neural Network
CNN	Convolutional Neural Network
KNN	K-Nearest Neighbors
GRU	Gated Recurrent Unit
LSTM	Long Short-Term Memory

CHAPTER I:

INTRODUCTION

1.1 Introduction

An efficient approach for analyzing reviews and critiques is to collect user opinions, which are crucial for the worldwide growth of firms (Alzahrani et al., 2022a). At the moment, businesses are curious about what their customers think of their products because one opinion can cover a variety of topics, such as the product's quality, price, and color, and it can range from negative to neutral to positive. The quality may be good for the customer, but it may not be as intended. Nevertheless, it is crucial for companies to know how buyers perceive a specific product in order to analyze this data and enhance product quality. This is essential for survival in a highly competitive market, as knowing general opinions about products does not provide enough information to identify the entity's strengths and weaknesses(Cambria, 2016). The worldwide e-commerce business is seeing an upsurge in the amount of online merchants using advanced personalization algorithms. Retailers like yours employ recommendation algorithms, and there are early adopters of e-commerce who have access to sites like Amazon.com. As a result, e-commerce platforms like Amazon, Alibaba, and eBay have revolutionized product purchasing, allowing SMEs to capitalize on the rise of online marketplaces (Almahmood & Tekerek, 2022). The Internet has become the main platform for e-commerce activities. Banks that give their services online and travel firms that sell tickets are just two examples of the many types of businesses that rely on electronic commerce on a daily basis to reach thousands of clients digitally. Numerous e-commerce platforms enable consumers to express gratitude for their goods and services in order to gather valuable information, such as product specs, which are crucial to understand and users' sentiments (Geetha & Karthika Renuka, 2021). E-commerce websites provide many benefits for doing business online, including flexibility, loyalty, cost and time savings, and efficient delivery (Alamdari et al., 2020) .

On the other hand, sentiment analysis (a subfield of contextual text mining) use ML to discover and extract subjective information from many sources; this information helps businesses monitor online comments about their products and services and understand the general public's opinion of them(Park & Woo, 2019). Customer reviews, a subset of sentiment analysis, provide an outlet for consumers to share their experiences with a service or product that have purchased. Reading these reviews can help you get a better idea of what other people think of the products and services on the market. Positive reviews, also called "word of mouth" (WOM) (Al-Natour & Turetken, 2020), are important because it let other people learn about the products and merchants. A growing number of reviews makes it more difficult for potential buyers to go through each one before buying a product or service (Chintalapudi et al., 2021). Customers may utilize the useful information extracted from reviews using a variety of analytical approaches to make well-informed judgements. Specifically, sentiment analysis is utilized to ascertain and comprehend the emotions that customers convey in their text remarks. The goal is to retrieve and provide relevant data that improves online purchasing experiences (Savci & Das, 2023).

Sentiment analysis as a task has been investigated from several angles (Birjali et al., 2021). Nonetheless, the three levels of document, phrase, and aspect are the primary focus of emotion and point of view recognition. The third level becomes more difficult since it requires more thorough investigation, but the first two levels are intriguing and provide substantial obstacles. Thus, ML and DL techniques have been utilized in a variety of domains, like academics and healthcare. Companies may have access to advanced analytic tools, data learning capabilities, and the ability to correct class imbalances by using ML approaches.

1.2 Sentiment Analysis

Automatically identifying emotional tones is the goal of SA, which makes use of ML algorithms, rule-based approaches, and NLP. The three tiers of sentiment analysis extraction are the phrase, document, and feature levels. We automatically identify all relevant subjective aspects after collecting subject-related information. The goal is to determine whether the material that users provide is favourable, neutral, or negative.

Currently, there are 3 methodologies utilized for SA: hybrid, ML, and lexicon-based approaches(KABIR et al., 2020). An early strategy for sentiment categorization was dictionary-based methods, which include both dictionary-and corpus-based approaches. ML-based techniques for sentiment analysis use both conventional and DL methods(L. Yang et al., 2020). Combining lexicon-based methodologies with ML, hybrid systems often heavily include sentiment lexicons. Figure 1.1 displays a classification scheme for Sentiment Analysis methods that rely on deep learning.

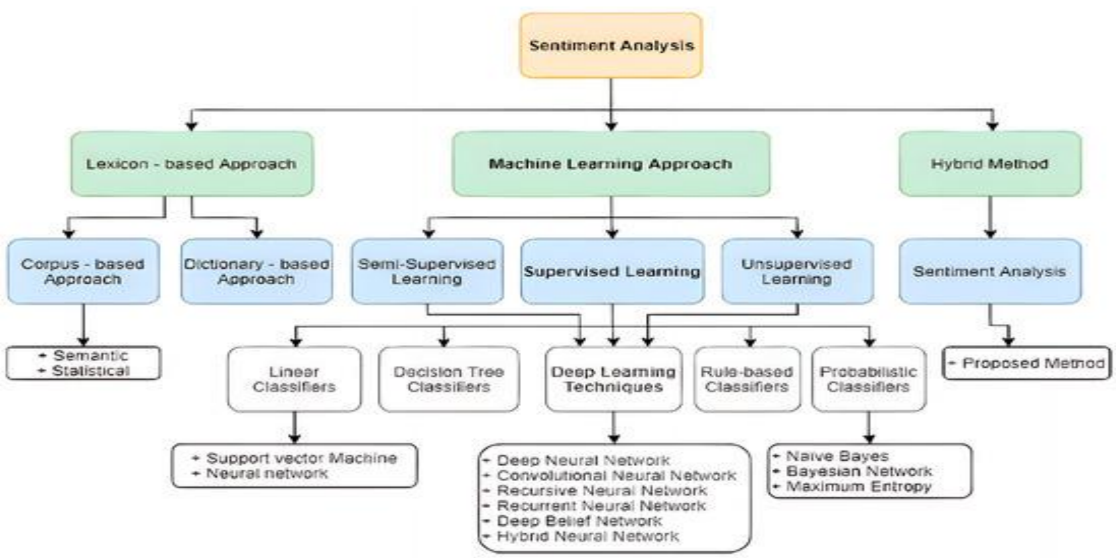


Figure 1.1: Taxonomy of sentiment analysis methods (de Oliveira Carosia et al., 2021)

Types of Sentiment Analysis depending on nature

Depending on the kind and depth of study, sentiment analysis may be classified into several types. These are a few common examples:

- **Binary Sentiment Analysis:** Positive/negative sentiment. Binary sentiment analysis classifies texts as either positively or negatively sentimentated. This method is used in tasks like determining the tone of product evaluations, where the objective is to determine whether the conveyed attitude is positive or negative.
- **Ternary Sentiment Analysis:** Sentiment: positive, negative, neutral. Positive, negative, and neutral are the three categories into which this literary style is separated. When a more nuanced attitude is needed or when several texts reflect a neutral perspective, this is helpful.
- **Multi-Class Sentiment Analysis:** Includes multiple sentiment categories. In contrast to the 3-category method, multi-class SA classifies texts according to many sentiment dimensions. There are four possible emotion categories: very positive, neutral, negative, and extremely negative. More thorough and advanced research on sentiment is possible with this method.

Types of sentiment analysis in different levels

The SA task is carried out on textual data at four different levels as follows:

- Document-level sentiment analysis
- Sentence-level sentiment analysis
- Word-level sentiment analysis
- Aspect-based sentiment analysis

1. **Document-level sentiment analysis**

The goal of Sentiment Analysis is to categorize whole publications as either positively or negatively based on the frequency of sentiment expressions within the text.

When doing Sentiment Analysis on individual documents, it is assumed that every document expresses ideas towards a single entity.

Document level sentiment analysis model gets a set of documents as input, detects number of sentiment words and assigns sentiment label with leading sentiment words category to each document as shown in Figure 1.2 (Ray & Chakrabarti, 2022).

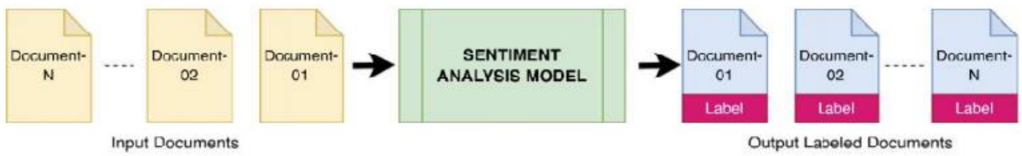


Figure 1.2: Document level sentiment analysis model

2. **Sentence level sentiment analysis**

Figure 1.3 illustrates how the frequency of sentiment phrases is used by phrase-level SA to determine the positivity, negativity, or neutrality of a remark. Sentence level Sentiment Analysis distinguishes between objective statements that convey facts and subjective ones that reflect opinions(Alsayat & Elmitwally, 2020).

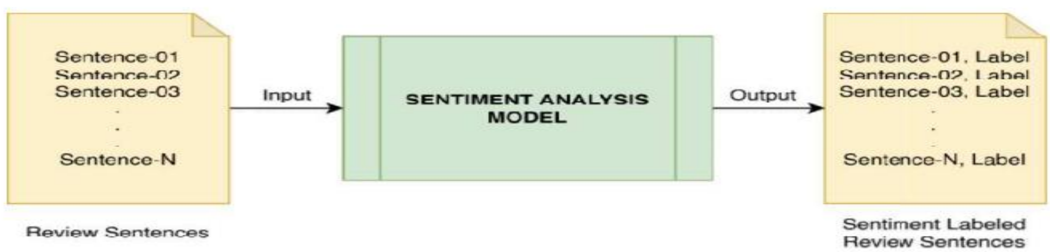


Figure 1.3: Sentence level sentiment analysis model

3. **Word level sentiment analysis**

Each input word is tagged with a positive, negative, or neutral attitude using word-level sentiment analysis, as seen in Figure 1.4. Polarity and an intended recipient are the two main characteristics that differentiate opinions in word-level SA (Ito et al., 2020).

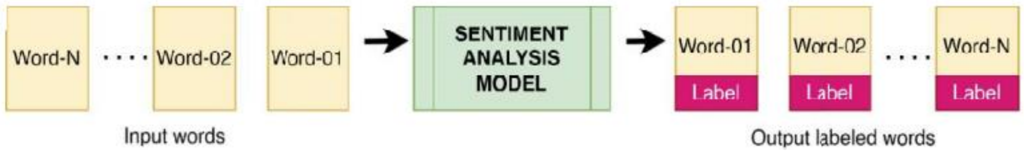


Figure 1.4: Word level sentiment analysis model

4. Aspect based sentiment analysis

In Sentiment Analysis, the emphasis is only on labelling emotions as good, negative, or neutral, rather than on the particular context (or aspect). Finding feelings connected with different features of the input reviews is the method of Aspect Based Sentiment Analysis (ABSA). The method of aspect-based sentiment analysis may be utilized to analyze customer feedback by determining the precise sentiments associated with different parts of a service or product (Figure 1.5)(Chifu & Fournier, 2023).

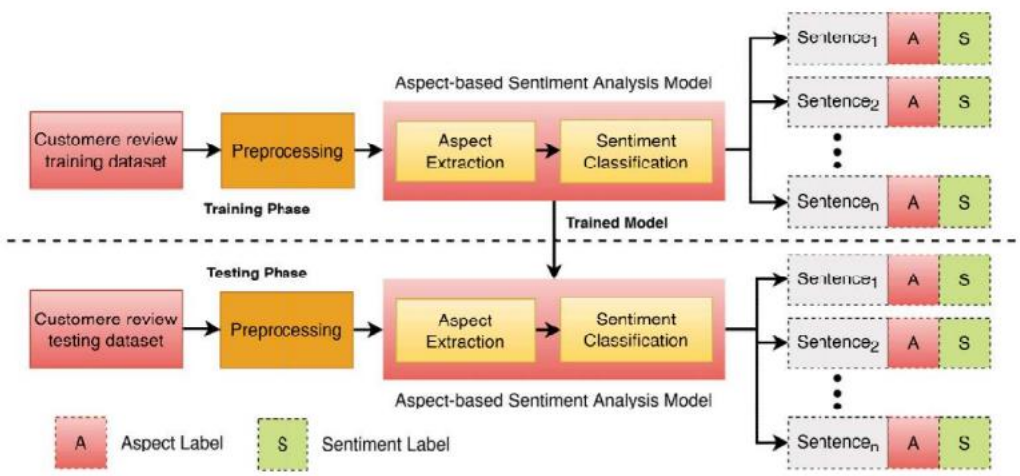


Figure 1.5: Aspect-based Sentiment Analysis Model

Businesses may find aspect-based sentiment analysis useful for automating processes like customer care duties, automatically sorting and analyzing client comments, and obtaining strong insights on the move. Customers appreciate connecting with businesses and leaving good and negative comments. Their insights are consistently varied. It enables businesses to zero in on what makes customers pleased or unhappy with a product or service. If there are unhappy, it helps to fix them in real-time. ABSA can immediately identify the customers who are not satisfied about one particular service or product aspect and it takes the necessary decisions to make customers happy.

Sentiment Analysis Tasks

Sentiment analysis involves several tasks and subtasks. Various classifications of these responsibilities have been proposed in the literature. Subjectivity categorization, sentiment categorization, review usefulness evaluation, and sentiment spam detection are the four types of Sentiment Analysis tasks shown in Figure 1.6 After that, Sentiment Analysis is performed based on these categories utilizing machine learning methods. Subjectivity categorization refers to a process of identifying if a statement contains views, judgements, or evaluations. Although improvements in subjectivity categorization have improved sentiment analysis, the work remains challenging. A goal of sentiment classification is to assess the trend of an opinion. Across languages and domains, sentiment analysis, polarity persistence, and perspective content ambiguity resolution are all linked to sentiment categorization(W. Zhang et al., 2023). Sentiment Analysis is used to accomplish the following goals, with a few examples provided below:

- **Sarcasm Analysis:** The purpose of the somewhat niche field of NLP known as "sarcasm detection" is to identify satirical texts. A context-based feature approach to sarcasm detection using benchmark datasets makes use of DL and the BERT model (Eke et al., 2021). An ensemble-learning and DL-based sarcasm detector (Goel et al., 2022).
- **Aspect extraction and categorization:** A method for extracting multi-domain aspects using BERT (Santos et al., 2021). An employ of a DCNN to extract aspects for opinion mining is another approach (Zadeh et al., 2017)

- **Opinion expression extraction:** In contrast to generic subjective expressions, which may include either the aspect or the expression, aspect-specific expressions maintain both inside the original phrase context while conveying an opinion.
- **Trends topic Detection:** The use of DL with streaming data allows for the detection and prediction of trending topics (Pathak et al., 2021)
- **Product Recommendation:** Recommender systems, which help customers find the best goods and services, have also benefited from deep learning's advancements. Context-aware recommendation system that considers contextual information using deep learning (Jeong & Kim, 2022).
- **Finance prediction:** DL for finance refers to the practice of using neural network methods across several aspects of the financial industry. Companies listed in the digital economy use an upgraded BP neural network for risk prediction in their financial planning(W. Li et al., 2022)
- **Live game prediction:** A training and exercise model for predicting the outcome of sporting events using an attention-based LSTM network. E3 ubiquitin ligase interaction and degrons were systematically predicted using deep learning.
- **Stance Detection:** Social media rumour detection is greatly improved by classifying user comments according to their opinion. Bipartite Graph Neural Networks for Identifying Influencer Attitudes on Twitter. Stack models are employed to extract arguments and determine positions using an integrated LSTM model(Zhou et al., 2023)

Important Notions Used for Sentiment Analysis

It is necessary to differentiate among intuitive and objective statements before conducting an opinion test. The feelings are expressed in natural language. Objective text only contains accurate information. Although it lacks both positive and negative action, subjective phrases may be classified into three portions according to the viewpoint expressed: positive, negative, and neutral ideas and feelings. A content's positive, negative, and object qualities are given polarity according to the context. There is a lot of discussion about whether two or three classes are better, but one thing is certain: considering inactive classes makes you more efficient (Kaur et al., 2017).

Applications of Sentiment Analysis

Sentiment Analysis, sometimes known as opinion mining, has several benefits and is used in many different fields:

1. **Business and Market Intelligence:**
   - **Product Feedback:** By using SA to data gathered from reviews, social media, and surveys, businesses may get valuable insight into how their consumers perceive their products and services.
   - **Competitor Analysis:** Organizations may get insight into their competitors' advantages and disadvantages by evaluating public perception of them via SA.
2. **Customer Support and Engagement:**
   - **Real-time Feedback:** social media and customer care channels may be actively monitored by sentiment analysis technology to discover and resolve client concerns and attitudes.
   - **Chatbots:** chatbots may have a deeper understanding of user emotions and needs when sentiment analysis is included into them.
3. **Brand Reputation Management:**
   - **Crisis Management:** Sentiment research may help firms identify and handle potential public relations crises on news channels and social media.
   - **Brand Monitoring:** This tool is useful for companies who want to monitor the long-term and regional perceptions of their brand.

4. **Product Development:**

- **Feature Prioritization:** companies may prioritize product improvements and innovations based on customer value by analyzing consumer feedback.
- **Innovation:** The use of sentiment analysis in the discovery of emerging trends and customer needs may greatly aid innovation initiatives.

5. **Financial Services:**

- **Stock Market Prediction:** The stock market may be forecasted using sentiment analysis of news stories and online conversations.
- **Risk Management:** Market mood and risk may be estimated by examining the tone of financial reports and news.

Figure 1.6: 5 key business application of sentiment analysis(Matthew McMullen, 2022)

Need for Sentiment Analysis

Examining public opinion is getting more important as internet use continues to rise. As shown in Figure 1.7, SA is necessary for opinion assessment, which is relevant to a wide range of tasks and methods.

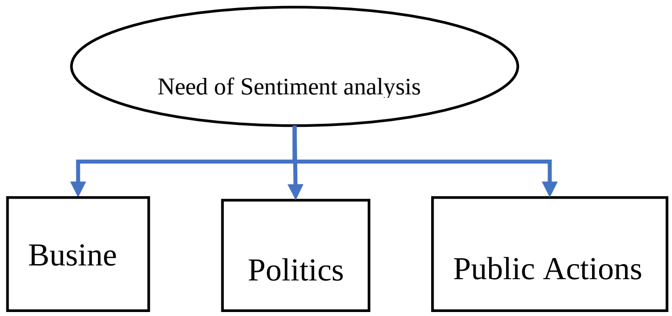


Figure 1.7: Need for Sentiment Analysis (self-created)

- **Business:** The marketing sector makes use of sentiment analysis to inform strategy development, get insight into how customers feel about a product or brand, gauge reaction to ads and new products, and identify non-buying factors.
- **Politics:** The political science community makes use of sentiment analysis to keep tabs on political stances, spot administrative-level contradictions and statements, and predict election results.
- **Public Actions:** Identifying potentially harmful occurrences and gauging the mental health of bloggers are two examples of social phenomena that may be studied and evaluated using opinion categorization (Saini et al., 2019).

Sentiment Analysis in E-Commerce

Financial portfolios are susceptible to sudden shifts or out-of-the-ordinary consequences due to the emphasis on big news items, which may significantly affect investor behaviour and the financial market. Due to the extensive availability and ever-increasing use of the Internet, individuals nowadays want to communicate with one another and share information on healthcare, completed products, and services. These new analytics technologies aim to comprehend and forecast human behaviour and attitude after provide traders with data that can be utilized to plan and organize company processes in advance of investment or risk management choices. Capital of the market is e-commerce sites. No one has to leave the sector because of the places' dependability, which makes them more dependable to the point of being highlighted (Aversa et al., 2018). While waiting for the merchandise to move forward, customers check into online shopping places. Sentiment analysis is a major focus in this area, since many well-known and rumored companies are now using their websites to promote their goods. The end goal is to retain confidence in these platforms. "Sentiment" is "an expressed view or opinion," and "analysis" means "the structure of something," therefore merging the two definitions helps to reveal the underlying emotions.

From many years ago to the present day, there have been several suggestions for and implementations of improvements to Sentiment Analysis methods. With an eye towards providing a thorough overview, this article will examine the most popular methods utilized in retail, with a focus on the e-commerce industry. Several subfields of sentiment classification, which mostly include Machine Learning and Lexicon. Even fewer research have looked at how these two methods might be combined to improve sentiment analysis performance (Wedel & Steenkamp, 1991).

The Internet's meteoric rise in popularity has altered global communication, particularly in the corporate world. Opportunities for the advertising of products and services have expanded due to the expansion of online marketplaces. The pervasiveness of social media platforms allows users to voice their views on a vast array of topics, often drawing on personal experiences and the goods and services that cherish. SA is a fast-expanding topic of NLP that can predict what people will like. Businesses would do well to redirect their attention to SA, which may help them comprehend the desires, needs, when, why, and how of customers, so that can make better decisions based on comments and reviews and avoid repeating past mistakes. The term "online shopping" describes the practice of purchasing goods and services over the internet and is just one facet of electronic commerce. Both large distributors like Amazon and Alibaba and smaller distributors out there had dismal results; a lack of a wide enough selection of products was a major contributor to their sluggish sales. Just like a tech company like Fitbit in 2016, these wholesalers were unable to provide the right items, and buyers retaliated by going elsewhere to buy them. Businesses were primarily motivated to learn about the varied and comprehensive opinion mining on customer evaluations by using sentiment analysis to monitor customer emotions, since consumer knowledge has long been a top priority for distributors. The Internet is a treasure trove of views; any company looking to enhance its offerings would do well to take advantage of the many ways in which customers may voice their ideas(Liu et al., 2020).

Benefits of sentiment analysis in E-Commerce

The various benefits of Sentiment Analysis are as follows:

1. **Predict the future**

SA can tell you what's hot and what's not by looking at how popular things are and the language people use when commenting on them. It can also tell you what's just starting to become popular. Maintaining a competitive edge requires a system that can instantly adjust your ads and sales strategies according to data.

2. **Build a better brand**

Personality, goods, and services may all be fine-tuned with an employ of SA. This establishes a track record of being ahead of the curve, attentive to client demands, and in sync with current trends throughout time.

3. **Valuable business intelligence**

Companies may receive actionable knowledge about their present and potential consumers, as well as information about emerging markets and opportunities, via sentiment analysis data.

4. **Enhanced customer experience**

We may use SA to determine whether a word used on social media is favourable, negative, or neutral, and thus how consumers feel about a service, product, campaign, or anything else. Repeat business comes from customers who are satisfied with the service that received from start to finish.

Text analytics techniques

AI is the study and development of computer systems that can reason and reason like humans. Put another way, it seeks to imbue computers with human-level intelligence by teaching them to reason and reason like humans via the use of computer programs and machinery. AI has the ability to simplify people's lives while simultaneously enhancing the global system via its vast web of interdependent countries, corporations, and governments. Thus, AI's primary objective is to provide machines with cognitive capacities including reasoning, decision-making, seeing their surroundings, and speech recognition. Consequently, creating intelligent automated systems that can keep up with present expectations requires AI-based modelling. The future of almost every sector is being shaped by this new technology, which is upgrading and simplifying several operations.

Challenges and Perspectives in Sentiment Analysis

Obstacles and Viewpoints, SA is a very difficult undertaking when dealing with subjective thoughts and human behaviours. Here are only a few of the obstacles :

- **Recognizing Subjective Parts of The Phrase**

At times, the English language may be difficult to navigate. Words that sound similar and have the same spelling but signify different things are called homophones. Parts of the phrase or sentence that are subjective represent material that is tied to emotion. The phrase's homophones could be either objective or subjective depending on the context. It makes it hard to pick out the part of the statement that is subjective.

For example: 1. The new bulb provided an adequate amount of light for reading. 2. The metal magnesium is quite light. The first half of the sentence uses the term "light" to describe a certain kind of light, while the second portion uses the word to describe how dense something is. On many social media sites, users express their ideas and beliefs online. Studies have shown that older individuals, in contrast to younger ones, are better at expressing their opinions and sharing knowledge.

- **Dependence on The Domains**

Different contexts could give rise to distinct understandings of the same term. As an example, the phrase "unpredictable" conjures up favourable associations in the realms of entertainment and the theatre, but negative associations when used to the topic of car breaks. Nevertheless, accurately determining the domain to which a word is associated remains a tough task. The text is correctly labelled utilizing many domains of pre-trained word embedding corpora, such as the customer reviews dataset and the IMDB movie reviews corpus. Even after all this time, researchers are still unable to resolve this issue to everyone's satisfaction.

- **Detection of Sarcasm in The Phrase**

A sarcastic statement uses positive words in an unusual way to convey an opinion about someone or something unpleasant. A common tactic is to say something completely false in order to make another person seem or feel stupid. Consider this example: "Excellent fragrance. The marinade time is crucial. The statement conveys a negative tone despite the use of solely positive words.

- **Dependence on The Order**

Opinion mining and SA depend on the examination of discourse structures. For instance, saying that A is superior to B expresses the exact opposite viewpoint of saying that B is superior to A. It may be difficult to find SA for these kinds of sentences.

- **Idioms**

A figure of speech is intentionally left out of ML programs' design. Words like "not my cup of tea" might throw off an algorithm that takes literal meanings too seriously. Any time a user leaves a comment or review using an idiom, the computer fails to accurately map the phrase meaning. Having a remark that is bilingual makes the matter much more challenging (Wankhade, Rao and Kulkarni, 2022).

Understanding Various Types of Artificial Intelligence

The fundamental goals of AI include understanding and carrying out intelligent activities, such as reasoning, learning, and adjusting to new situations and problems. Finally, AI is a field of study that aims to copy human intelligence via the use of scientific and technical methods. It is challenging to develop a reliable AI model due to the variety and ever-changing character of real-world data and situations. We examine several forms of artificial intelligence, such as analytical, functional, interactive, textual, and visual AI, in order to address different challenges in the continuing Fourth Industrial Revolution (Figure 1.8). Here, we outline the computational and real-world service boundaries of each category.

- **Analytical AI:** Analytics is usually defined as the study of data with the goals of discovering, understanding, and conveying relevant patterns. Consequently, analytical AI aims to shed light on hidden correlations, patterns, insights, and linkages inside datasets after support data-driven decision-making. Therefore, it is an essential component of AI within the context of modern BI, which, owing to its analytical processing capacity, may provide insights to a company and produce recommendations or ideas. It is possible to build an analytical AI model that targets a particular real-world problem by combining several ML(Sarker, 2021c) and

DL(Sarker, 2021b) approaches. To assess potential danger, a business may employ a data-driven analytical model.

- **Functional AI:** Functional AI, like analytical AI, searches massive databases for patterns and correlations. In contrast, functional AI actually does things rather than just suggesting them. For instance, in robots and Internet of Things applications, a working AI model might be helpful for taking quick decisions.
- **Interactive AI:** As a result of advancements in interactive AI, effective and engaging forms of automated communication have become commonplace in many spheres of human existence, particularly in the corporate sector. The creation of chatbots and AI personal assistants might benefit from an interactive AI model, to provide one example. A variety of techniques may be used to build an AI model that can interact with its environment, including ML, reasoning, AI heuristic search, and frequent pattern mining.
- **Textual AI:** Textual AI often makes use of NLP and text analytics, enabling businesses to use capabilities like Text Recognition, Machine Translation, speech-to-text conversion, and content creation. To facilitate the provision of pertinent services, such as the response of customer enquiries, a business may use textual AI to bolster its internal corporate knowledge store.
- **Visual AI:** Visual AI can often translate visual data like photos and videos into meaningful insights, in addition to recognizing, classifying, and sorting objects. So, visual AI is a subfield of Computer Science that teaches computers to process visual information visually, much like humans. Applications of this kind of AI are common in computer vision and AR.

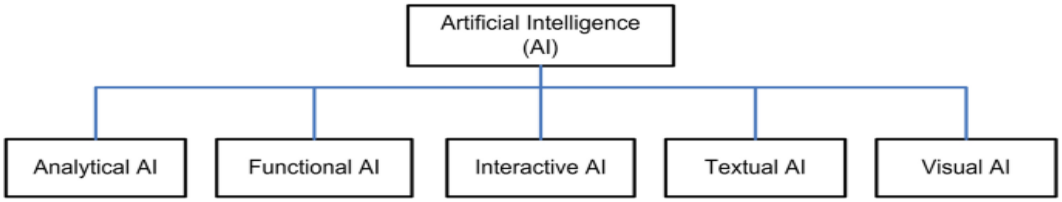


Figure 1.8: Various types of artificial intelligence (AI) considering the variations of real-world issues (Sarker, 2022)

The Relation of AI with ML and DL

These days, when people talk about smart systems or software, that usually mean AI, ML, or DL. Figure 1.9 displayed the current state of AI as it pertains to ML and DL. As can be seen in Figure 1.9, DL is a category that includes both ML and AI. When compared to AI, which imitates human thought and action in digital systems, ML streamlines the creation of analytical models via an employ of data and experience. Commonly referred to as "DL," these data-driven approaches to learning use multi-layer neural networks for computation. In the context of DL, "Deep" means the many layers of processing that data goes through after build a model that is driven by the input(Sarker et al., 2021).

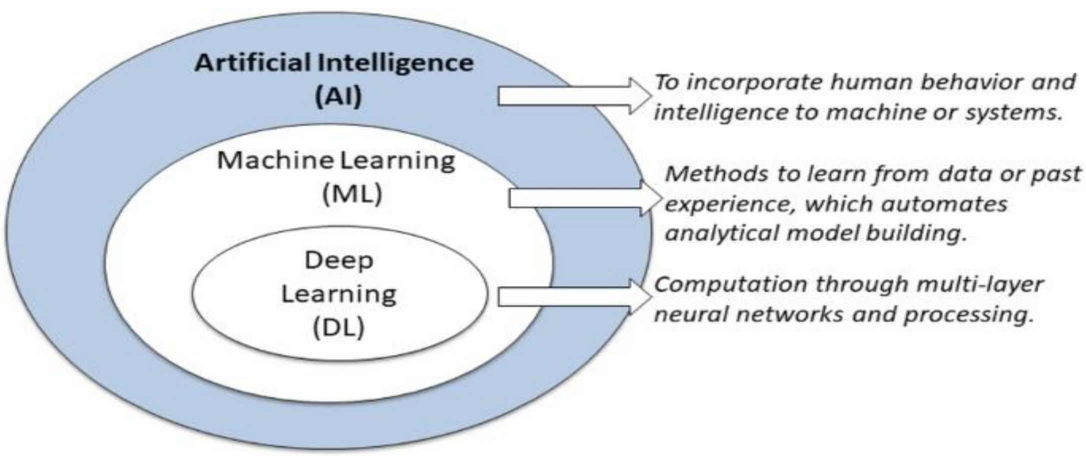


Figure 1.9: An illustration of the position of machine learning (ML) and deep Learning (DL) within the area of artificial intelligence (AI) (Y. Lee, 2023)

Therefore, ML and DL are not only cutting-edge AI technologies that can automate processes and build intelligent systems, but they are also crucial AI technologies. Additionally, it uses data-driven learning to elevate AI to a new level, which is dubbed "Smarter AI." Strong links to "Data Science" exist between ML and DL because of their mutual capacity to learn from data(Sarker, 2021a). Data science, often defined as the whole process of obtaining insights by data in a given issue area, may also benefit greatly from these learning approaches for intelligent decision-making and sophisticated analytics. Collaboratively, ML and DL technologies have tremendous promise for revolutionizing our present-day surroundings, thanks to their strong computational engines and their ability to create technology-driven automated, smart, and intelligent systems.

1.3 Potential AI Techniques

This article offers a high-level overview of the ideas and capabilities of possible AI approaches that might be utilized to create computers that are both clever and perceptive in a variety of practical contexts. Considering different forms of AI, we classify AI approaches into 10 possible groups for this purpose. The following are ten classes of artificial intelligence methods that, depending on the issue at hand, might be crucial to intelligent, clever, or automated computer systems.

Machine Learning approach

Sentiment analysis relies heavily on ML, which enables models to autonomously discover patterns and classify text based on sentiment. It is common practice to use DL models like RNNs and transformers (e.g., BERT) in addition to Supervised Learning techniques like Naïve Bayes and SVM for sentiment classification (Kumar et al., 2020). These models can tell the difference among neutral, positive, and negative feelings after being trained on labelled datasets. Additionally, unsupervised and semi-supervised learning approaches help analyze sentiment without extensive labeled data. Machine learning improves SA by improving accuracy, handling large volumes of text, and adapting to complex linguistic features like sarcasm, context, and informal language, making it an essential tool in opinion mining(Samyuktha et al., 2023). Figure 1.10 shows a process of SA using ML, and Figure 1.11 shows types of machine learning model.

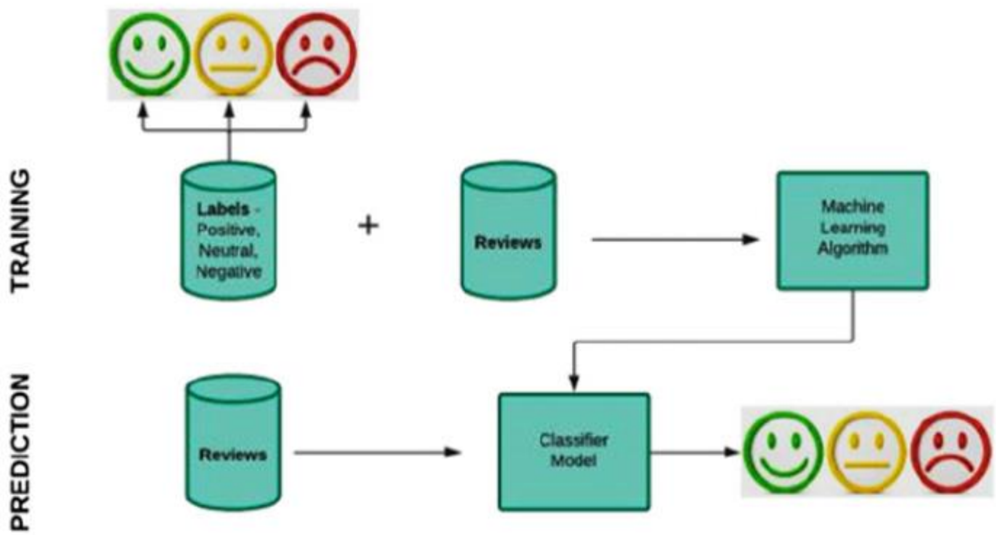


Figure 1.10: Machine learning for sentiment analysis (Kusal et al., 2024)

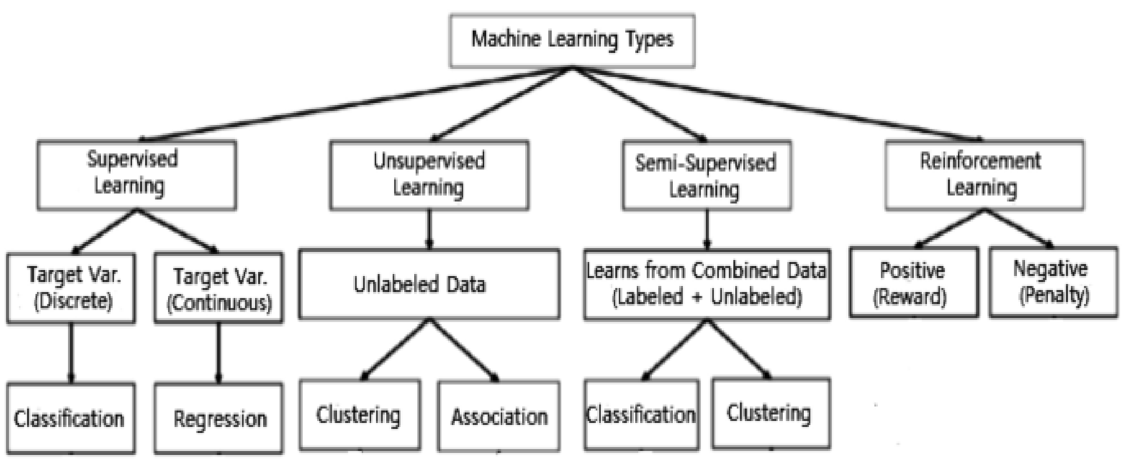


Figure 1.11: Various types of machine learning techniques with examples (Sarker, 2021c)

Supervised: Typically, Supervised Learning is what machine learning is all about. It's the process of learning a function that takes inputs and outputs as examples. An inferred function is obtained by combining tagged training data with a set of training samples. During Supervised Learning, a specific set of inputs is used to achieve predetermined objectives, i.e., a task-driven approach. The "classification" function is utilized for data categorization, while the "regression" function employs data fitting. These are the two most typical supervised jobs. Text categorization is one application of supervised learning. An example of this would be taking a tweet or a product review and predicting the tone or class of the content.

- **Unsupervised:** Unsupervised learning eliminates the need for human intervention by analyzing datasets without labels, i.e., a data-driven process (Han et al., 2012)This is often used for exploratory reasons, finding significant patterns and structures, extracting generative characteristics, and grouping findings. Typical applications of Unsupervised Learning methods include clustering, Dimensionality Reduction, anomaly detection, Feature Learning, discovery of association rules, and density estimation.
- **Semi-supervised:** The semi-supervised learning approach combines supervised and unsupervised learning techniques. It is capable of processing both labelled and unlabelled data (Sarker et al., 2020). Therefore, it is a hybrid of the two extremes of learning styles: "without supervision" and "with supervision". The use of semi-supervised learning becomes apparent in several real-world situations when supervised data is few but unsupervised data is plentiful (K. He et al., 2016). Better prediction results than those obtained from the model using only the labelled data are the final objective of a semi-supervised learning model. Machine translation, text categorization, categorizing data, and fraud detection are some of the domains where semi-supervised learning finds utility.

- **Reinforcement:** Software agents and computer programs may use reinforcement learning, an ML approach, to determine the optimal action to perform in every given scenario automatically (Joshi et al., 2021), i.e., an environment-driven approach. By using the knowledge that environmental activists have gained via this kind of reinforcement learning, the ultimate objective is to take action that will either raise the reward or lessen the hazard. Supply chain logistics, autonomous driving, robotics, and manufacturing are all complicated systems that may benefit greatly from this tool's AI model training capabilities, but fundamental or simple challenges aren't its strong suit(Srinivas et al., 2021).

Consequently, a wide variety of ML techniques could be vital in developing effective models for various use cases. This is because their ability to learn is influenced by the specifics of the material and the intended result. Figure 1.12 shows a high-level perspective of the various machine learning methods, with examples of each. This post will walk you through each ML technique that may improve your data-driven app's intelligence and usefulness. The following are examples of the many kinds of ML models:

- **K-Nearest Neighbours:** The concept of Nearest Neighbour Classification is simple: instances are grouped according to the class of their nearest neighbours. Since it is often helpful to include many neighbours, k-nearest neighbours aid in class definition. Induction is regarded as a sluggish learning technique as it is delayed until run time. Case-Based Classification and Example-focused Classification are two names for it because of its exclusive emphasis on training cases. During runtime, the training examples need to be in memory (Cunningham & Delany, 2021).
- **Naïve Bayes:** The Bayes theorem is the source for the probabilistic models that simplify by assuming solid independence. A core principle of NB is to calculate

the category probabilities for a certain text document using a combined probability evaluation of words and categories. Bayesian network classifiers are often used in supervised classification. It is still a key technique in text classification and categorization, having first been seen in text retrieval. The Bayesian classification approach has two main benefits: first, it provides useful learning algorithms; and second, it combines knowledge about previous data with information about current data (Jing et al., 2008). Bayes theorem for probability has a basic formula as follows:

$$P(y|x) = \frac{P(x|y)}{P(x)} \quad \dots (1.1)$$

- **Random forest:** Regression and classification problems may be resolved with a help of the RF model of supervised ML. Specifically, it's a set of algorithms that rely on decision trees for prediction. It is possible to get a result by training several DT and then combining their predictions. As a result, diverse data samples are used to train an algorithm's DT. Random Subspace is an additional component of RF that boosts and randomizes outputs. Methods for learning generate a large amount of DT while undergoing training. The output is also aggregated from all the trees to get the class mode, which is useful for both classification and regression (Ma et al., 2018).
- **Support Vector Machine:** SVMs are a kind of ML method that can handle challenging regression and classification problems by executing efficient data transformations that establish boundaries among data points according to predetermined labels, classes, or outputs. Several sectors make extensive use of SVMs, including healthcare, signal processing, speech and image recognition, and many more (Guenther & Schonlau, 2016). The SVM formula is as follows:

$$f(x) = \sum_{x_j \in S} y_j K(x_j, x) + b \qquad \text{…. (1.2)}$$

where: xi - training patterns, yi - class labels, S - Set of support vectors

- **Linear Discriminant Analysis (LDA):** LDA is one choice boundary classifier that was constructed employing Bayes' rule and fitted with data using class conditional densities. Reduced computational costs and model complexity are achieved by projecting datasets into a lower-dimensional space, a generalization of Fisher's linear discriminant. Traditional LDA models often fit all classes using a distributional assumption, based on the assumption that their covariance matrices are uniform. A lot of statistical methods look for ways to make a linear combination of independent variables represent a single dependent variable. These approaches include LDA, ANOVA, and regression analysis(L. Xu et al., 2022).
- **Decision tree (DT):** DT are well-known non-parametric supervised learning techniques. We use DT learning algorithms for classification and regression. ID3, C4.5, and CART are 3 of the most well-known DT algorithms. Additionally, both the newly-introduced cybersecurity analytics (Intrude Tree) and user behaviour analytics (BehavDT) by Sarker et al. have fruitful uses. Figure 1.12 shows the tree hierarchy as sorted by class when DT sorts the instances from the root to a few leaf nodes. To sort instances into their respective classes, we start at the very top of the tree (the root node) and work our way down the branches that correspond to the attributes that each node defines. Gini impurity and information gain, formally defined as entropy, are the most often used criterion for splitting.

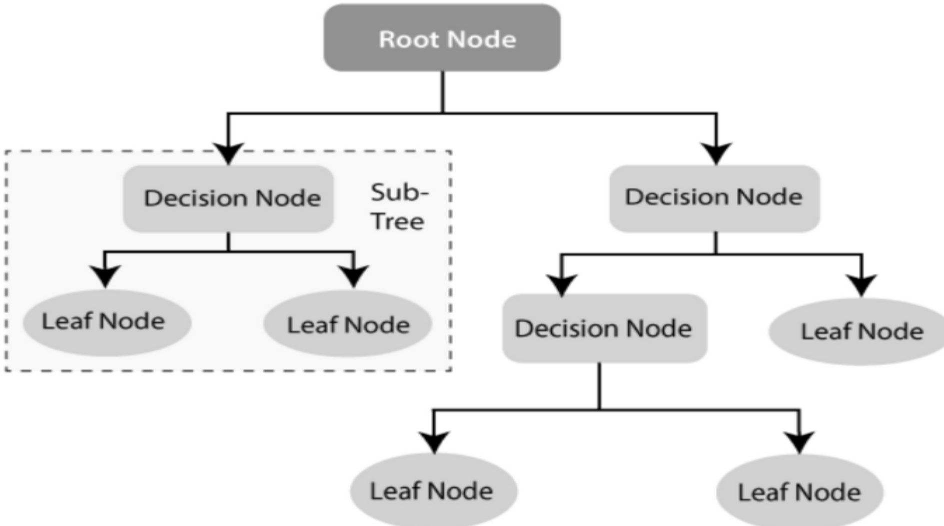


Figure 1.12: An example of a decision tree structure (Alahmadi et al., 2023)

- **Extreme gradient boosting (XGBoost):** Ensemble learning strategies like Gradient Boosting and RF (discussed above) use several models, often DT, to produce a single model. Using the gradient in the same manner as neural networks use gradient descent to change their weights might enhance the loss function. One kind of gradient boosting that considers more precise approximations in finding the optimal model is Extreme Gradient Boosting (XGBoost) (Alghazzawi et al., 2023).

Deep Learning Algorithms

DL is a well-known AI method that uses ANNs or artificial neural networks. Recently, DL's capacity to learn from data in a layer-wise fashion has made it a popular subject in the computer community. There are often several hidden layers in DNNs, including layers for inputs and outputs. Figure 1.13 illustrates the overall structure of a shallow network (hidden layer=1) in comparison to a DNN (hidden layer=N and N 2). There are primarily three broad types of DL approaches (Janiesch et al., 2021).

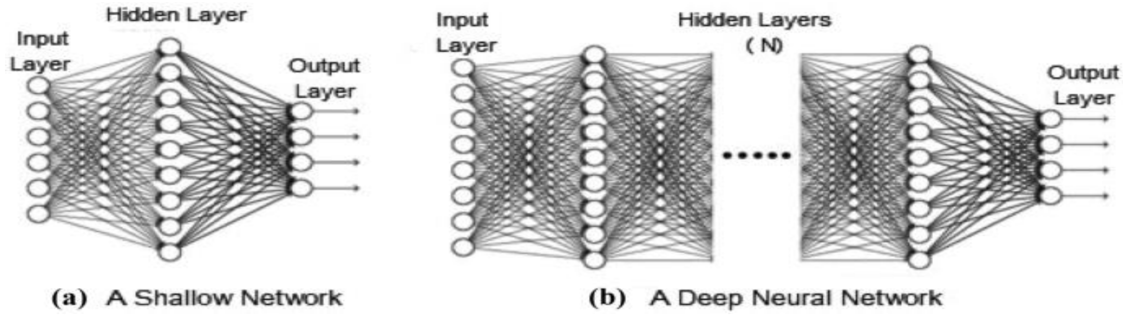


*Figure 1.13: A general architecture of **a** shallow network with one hidden layer and **b** a deep neural network with multiple hidden layers*

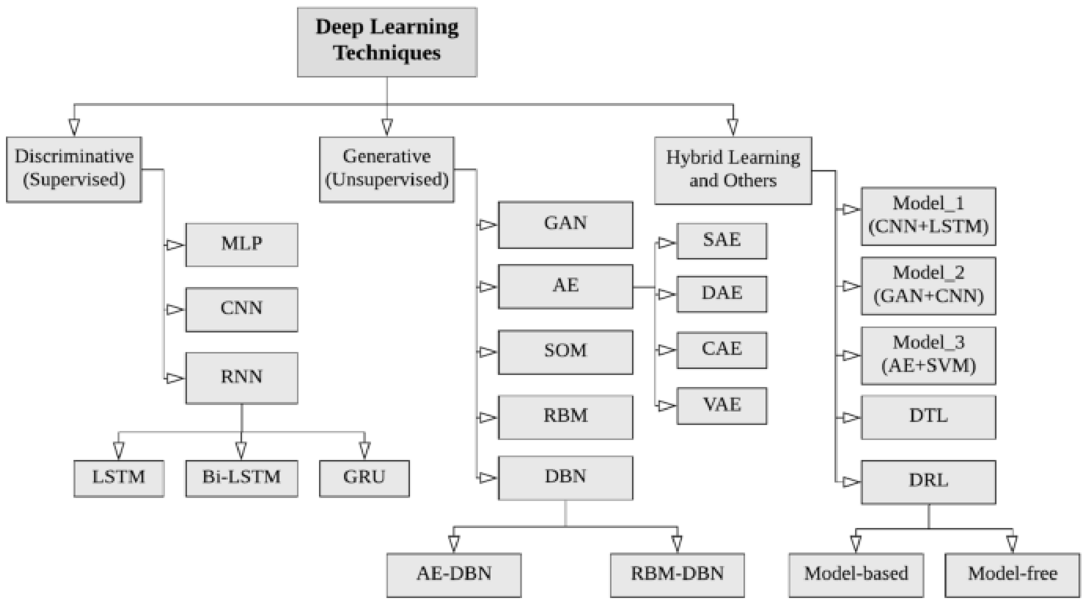


Figure 1.14: A taxonomy of DL techniques (Oliveira & Bollen, 2023)

Generally speaking, there are three main types. There are three types of deep networks:(1) supervised (or discriminative) networks, (2) unsupervised (or generative) networks, and (3) hybrid (or mixed) networks.

- **Deep networks for supervised or discriminative learning**. A discriminative alternative is provided by this DL method for supervised or classification tasks. Pattern categorization discrimination is one typical use case for discriminative deep architectures, which depict the posterior distributions of classes given observable data. For relevant real-world situations, deep discriminative learning models may be built using RNNs, CNNs, or ConvNets, or variants of these.

- **Deep networks for unsupervised or generative learning**. Common applications of this subfield of DL include identifying high-correlation features for pattern synthesis and analysis, and determining the joint distributions of observable data and its classes. One basic principle of generative deep structures is that some supervisory information, like labels for the target class, becomes superfluous as the learning process advances. The majority of generative modeling's applications are in unsupervised learning; however, it also sees regular usage in feature learning and data creation and representation. It guarantees the correctness of discriminative models and is a preprocessing step for supervised learning tasks. To build generative learning models that can solve important real-world issues, you may use: GAN, AE, RBM, Self-organizing (SOM), and DBN, along with its variations.
- **Deep networks for hybrid learning**. Generative models may handle both labelled and unlabeled data, making them very flexible. While generative models excel in supervised tasks, discriminative models fail to learn from unlabeled data. Hybrid networks focus on training deep generative and discriminative models together. The generative or discriminative DL model serves as the foundational model for a hybrid DL model, which consists of two or more additional basic learning models. For example, when dealing with real-world problems, it could be helpful to utilize generative and discriminative models in that order, or to mix the two with a non-deep learning classifier(Deng, 2014).

A few examples of deep learning models

a. Recurrent Neural Networks

The characteristic topology of RNNs is a network of linked neurons arranged in a loop. Due to their inherent memory capabilities, RNNs handle input sequences more efficiently and skillfully than regular feedforward neural networks, making them ideal for applications that need sequential data processing. All elements in a sequence may be consistently processed by an RNN because of its "memory" notion. What this implies is that the RNN will take all of its past computations into account when creating its current output. This means that it keeps track of historical data in a way that is analogous to how humans retain and retrieve information(Tsantekidis et al., 2022).

b. Convolutional Neural Networks

CNNs are a kind of ANN that finds widespread usage in Computer Vision and picture recognition applications. Three primary layers comprise CNNs:

Convolutional layers - Convolutional layers calculate the output of neurons linked to certain input areas by multiplying the weights of the neurons with the input region. • Fully connected layers - Its goal is to use the activations to produce class scores for classification and subclassification. • Pooling layers - A pooling layer decreases the number of parameters for the activation by means of spatial down-sampling. Figure 1.15 shows a CNN architecture

CNNs mimic the human visual system's pattern recognition and learning abilities that operate in the absence of training data (Saxena, 2022). A formula for computing the output of a convolutional operation is provided below:

$$C_{i,j} = \sigma(b_j^l + \sum_{m=1}^{M} W_{j,m}^{l,j} X_{i+m-1}^{l-1}) \quad (1.3)$$

where: l - the layer index, σ - the activation function, b - a feature map's bias term, M - kernel size, W - a feature map's weight

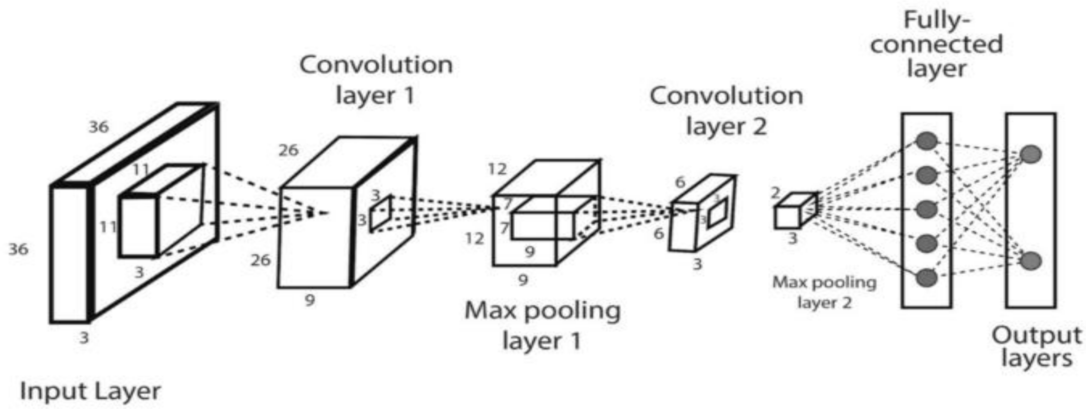


Figure 1.15: CNN architecture (Ayub, 2022)

c. Long Short-Term Memory networks

RNN with LSTM capabilities are among the most sophisticated network types. In their basic unit, which are also called "cells," that build "gates" that ensure a well-constructed structure. With these gates, traditional RNNs are able to avoid gradient bursting and disappearance while simultaneously capturing both short-term and long-term memory as time steps progress. These gates are named as: "forget gate", "input gate", and "output gate"(Hochreiter & Schmidhuber, 1997).

d. Gated Recurrent Units

A specialized RNN architecture that handles sequential data processing, the GRU provides an alternative to LSTMs. There is a significant difference among GRU and LSTM in regard to their internal architecture. Although capturing long-term dependencies is the purpose of both systems, LSTM has 3 components: forget, input, and output. In contrast, GRU is even more simple than the prior structure as it only contains two gates: update and reset. Gating techniques are used to solve the vanishing gradient issue in standard RNNs. While LSTM is useful in most cases, GRU excels in a few. Typically, it employs fewer parameters, leading to reduced computational expenses. However, LSTM excels at detecting complicated connections in other contexts(Van Houdt et al., 2020).

- **MLP:** Deep learning's foundational design is a multilayer perceptron, which is another name for a feed-forward ANN. A typical multi-layer perceptron (MLP) is a fully-connected network with three layers: input, hidden, and output. A particular weight is assigned to each layer's nodes when it links to each layer below them. Built on top of a most "fundamental building block" of neural networks, the "Backpropagation" method, MLP makes internal adjustments to the weight values while the model is being constructed. In order to make computationally expensive models work, MLP lets you tweak a bunch of hyperparameters, including the number of hidden layers, neurons, and iterations, and it's also sensitive to scaling characteristics.
- **Attention based method**

Cognitive attention may be represented using a neural network approach called attention. The concept behind the effect is that the network should pay more attention to that little but important part of the input data, therefore it boosts certain parts while decreases others. Attention takes a look at two phrases, arranges them in a matrix with the first sentence's words in the columns and the second sentence's words in the rows, and then finds relevant context by matching the two (Islam et al., 2021). In machine translation, this is tremendously helpful. That's fantastic, but the translation improves it even more (Bahdanau et al., 2015). To better understand attention, it incorporates into neural networks as a single hidden layer. Using the attention function, the network should be able to calculate the weighted total of all input characteristics and ascertain the importance of each hidden state.

- **Capsule based method**

A Caps Net is a framework for ANNs that may be used to develop sentiment analysis models with hierarchical interconnections. Capsule networks deviate from conventional neural networks in that they use capsules instead of neurons. To create a vector, a capsule must include all of the essential information from a picture. Although neurons just output a scalar value, capsules could remember the feature's orientation. The routing mechanism of capsule neural networks allows them to accept long-range characteristics while simultaneously decreasing training time. Each routing procedure consists of three stages: the main capsule, the secondary capsule, and the squash function(T. Zhao et al., 2019).

Natural Language Processing

NLP is a branch of AI concerned with teaching computers to comprehend, modify, and create meaning from human speech. Computer science and computational linguistics are only two of the many disciplines that natural language processing draws upon to help close the gap among human and machine understanding of language. AI text analysis is the method of extracting meaning from bigger text data sets. Analyzing syntax and meaning are the two primary areas of NLP. Syntactic analysis, sometimes called parsing, is a branch of language study that uses fundamental grammatical principles to decipher written language for patterns of syntax, word order, and word relationships. Semantic analysis uses text capture as its foundation. Our first stop is at the phrase level, looking at its lexical semantics. The potential for NLP to completely automate the process of understanding large-scale customer responses is attractive to businesses. Decisions based on data will be easier for them to make, which will ultimately benefit the company(L. Lee, 2002).

NLP is helpful because it lets computers understand human languages and carry out a range of activities based on those languages (Moreno & Redondo, 2016). For instance, the phrase "natural language processing" refers to the process by which computers can understand spoken language, decipher written material, gauge emotional tone, and prioritize tasks based on their importance. Languages spoken by humans are very varied and complex. A broad range of written and spoken formats have been used by humankind to communicate our thoughts (Ise, 2016). The ability to resolve linguistic ambiguities and provide numerical structure to data is a key component of AI that powers several downstream applications, such as text analytics and voice recognition.

NLP Techniques

Syntactic analytics and semantic analysis are the two main tools utilized in NLP to help computers understand text.

Syntactic Analysis Analyzing text via the lens of basic grammatical rules allows for the detection of sentence structure, the arrangement of words, and their connections (a process known as parsing). Among its primary responsibilities are: • A text may be "tokenized" by breaking it down into smaller bits, such as words or phrases, in order to make material management easier. • A portion of the speech tag specifies the meaning of tokens such as verb, adverb, adjective, substance, and so on. Words' meanings may be inferred from this; for instance, "book" can signify several things when used as either a verb or a substantive. • The goal of lemmatization and stemming is to facilitate analysis by reducing inflected phrases to their essential components. • Stop-word removal is a common technique for getting rid of words like "I," "they," "have," and others that don't provide any semitone value (Skarpathiotaki & Psannis, 2022).

Semantic analysis Text capture is the backbone of semantic analysis. The lexical semantics of each phrase is first examined. The next step is to look at the words and their arrangement in context. Here are the main responsibilities of semantic analysis: Word meaning disambiguation seeks to resolve context-dependent word meaning ambiguities.

Transformer models

An LLM cannot be constructed without the transformer architecture. Neural networks are designed to efficiently process sequential data. Recursion techniques are not used by this design. As an alternative, it finds global input-output interdependence using an attention technique. New model sizes and performance levels have emerged as a consequence, which has greatly improved parallelization and shortened NLP training periods. In addition, it can process input sequences of varied durations and adjust its focus accordingly. Due to its superior efficiency, it quickly became the preferred design in several domains, displacing more complex recurrent or convolutional neural networks. This is where it really shines for LLM programs. Figure 1.16 shows the transformer model's construction. A transformer's design is based on seven main parts. The following is a visual representation of all of the parts. Since it is limited to processing numerical input, ML models train on tokens entered by users. Consequently, "input embeddings"—a numerical format—need to be applied to these textual inputs. Machine learning models may be fed numerical word representations, or input embeddings, for processing. The model is able to understand words and act as a dictionary with the assistance of these embeddings, which group similar phrases together in mathematical space. Words with comparable meanings are represented by vectors of the same size using these embeddings, which the model is taught to create (Raiaan et al., 2024).

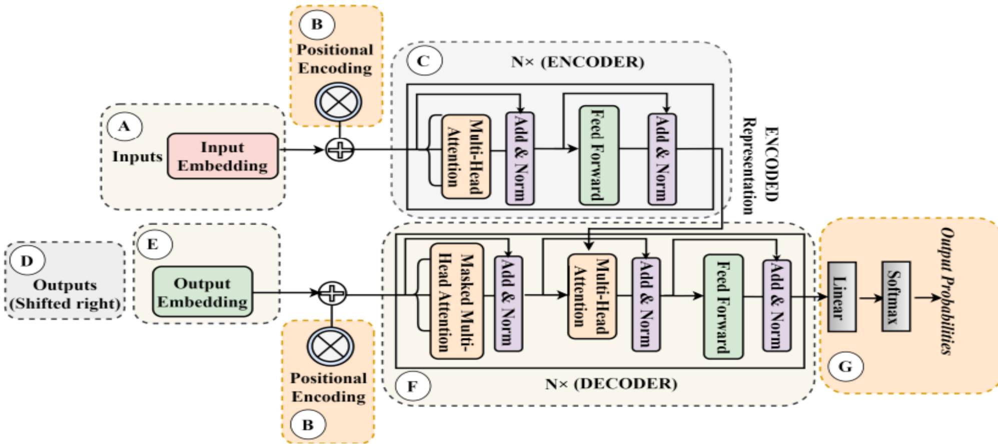


Figure 1.16: Architecture of a Transformer model (Chakraborty et al., 2024)

Types of LLM models

A subset of AI algorithms known as LLMs may carry out a huge variety of NLP tasks. Generating text, analyzing text, translating text, doing sentiment analysis, answering questions, and similar jobs are the most popular. Four of the most well-known transformer-based LLM models trained on large volumes of textual data are GPT-3, GPT-4, PaLM, and LaMDA. Size and depth are two architectural property determinants that exhibit variance in these models. For instance, although GPT-3 generates 175 billion parameters distributed over 96 levels, PaLM generates an even larger number of 540 billion parameters arranged over 106 layers. Different configurations are present in each of these models. The methods for producing output are different in the GPT-3 and PaLM systems. A number of datasets, including those from books, code repositories, social media, and Wikipedia, have been assessed by LLMs. Their ability to properly do a variety of activities has been shown time and time again. A lot of people are interested in LLMs because of all the places that might be used, including as healthcare, media marketing, education, and consumer services. As an example, GPT-3 is famous for its text style generation capabilities, whereas LaMDA excels in providing accurate responses to factual queries. This is just one example of how one LLM software outperforms another. Emerging technology innovations like LLMs have the potential to revolutionize many different industries(Gilhooly, 2024).

Basics of LLM models

The development of language, a miraculous medium of expression and communication, starts in infancy and continues throughout a person's life (Miller et al., 1994). However, advanced AI is required for computers to naturally comprehend and communicate in human language. Consequently, getting computers to read, write, and communicate like humans has been an ongoing scientific problem and goal (Wu et al., 2023). LLMs emerged, however, because to developments in DL techniques, plenty of training data, and powerful computing resources. Using self-supervised learning, this language model class trains neural networks on vast quantities of unlabelled text input, each network having billions of parameters. As far as AI and NLP are concerned, it represents a giant leap forward(Wu et al., 2023). Complex patterns, linguistic nuance, and semantic links may be learnt by these models, which are often pre-trained on huge online datasets. Text synthesis, translation, summarization, question-answering, and SA are just a few of the language-related tasks that have shown competency when combined with DL techniques and large datasets. Furthermore, after being adjusted for specific downstream tasks, these models have shown remarkable potential, attaining top-tier results in several benchmarks (Dogra et al., 2024). The foundation of LLMs may be found in the first stages of building neural networks and language models. Previous efforts to construct a language model relied on statistical methods and n-gram models, but these models failed to capture the nuances of context and long-term dependency. Then, when neural networks improved and greater datasets became available, researchers started to delve into more intricate methods. Development of the RNN was a watershed moment because it made it possible to represent sequential data, which includes language. Unfortunately, RNNs couldn't do much because of long-term dependencies and diminishing gradients. The revolutionary work marked a turning point for LLMs systems with the advent of the transformer design (Vaswani et al., 2017). The self-attention mechanism is the basis of the transformer model, which allows for effective management of long-range dependencies and parallelization. In addition, it was the foundation for models that had great success on a variety of language problems, including the Generative Pre-trained Transformer (GPT) series from open AI and Bidirectional Encoder Representations (Devlin et al., 2019).

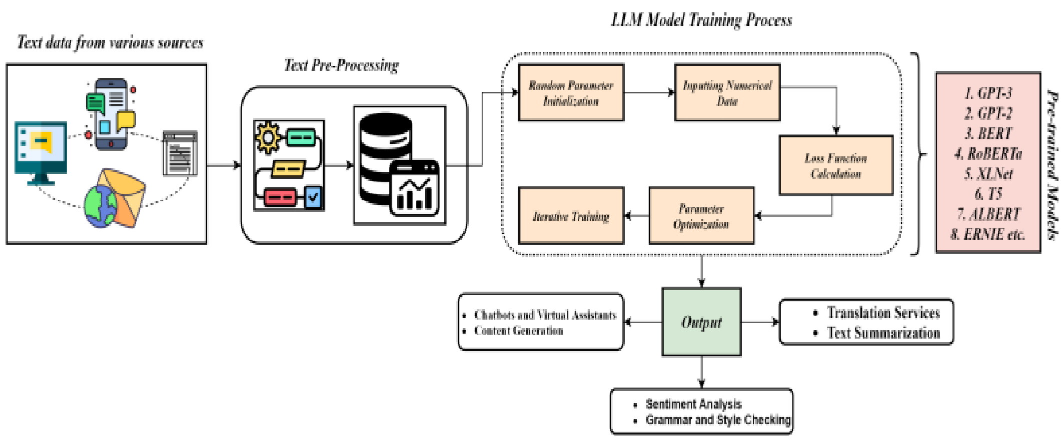


Figure 1.17: Pipeline of the LLM training phase (Raiaan et al., 2024)

Figure 1.17 depicts the pipeline of the fundamental LLM design. Once it receives text data from many sources, the preprocessing stage is triggered. Next, it goes through a sequence of steps to finish training, which includes initializing parameters at random, receiving numerical data, calculating the loss function, optimizing the parameters, and finally, iterative training. As a service after the training process, it provides text summarization, sentiment analysis, translation, and more.

Training of LLMs

To be effective, LLMs need training. Collecting and arranging a large amount of data from many sources is a common first step. Extensive research and evaluation are conducted on the main source of LLM education. Institutions of higher learning often use unsupervised learning. Based on past reports, the model is taught to predict future ones in a certain order. "Language modelling" describes this well-known occurrence. Transformer and other modern neural network topologies build linguistic links between words and sentences. The purpose of the training process is to identify the model's shortcomings in order to improve its chances of being selected for the following phrase within the given context. SGD is a popular tool for doing this. Once the back computation gradients are in the model, backpropagation is utilized to update the model's parameters. Multiple objective feedbacks are included into LLM's training at every level. Initially, the study's participants will undergo LLM training on a variety of unsigned and private items in an unsupervised environment. Stage two, fine-tuning, gives access to basic models with the introduction of more limited data and human input. In order to take the fine-tuned model to the next level, individuals might use ways to transform the LLM into an enhanced model capable of doing the task(Shahzad et al., 2025). The following are some instances of the LLM model:

- **GPT-3**

The GPT-3 language model is among OpenAI's most prominent and cutting-edge creations (J. S. Lee & Hsiang, 2020). This achievement sparked a lot of curiosity, which ultimately led to NLP's successes (Sha et al., 2021). The transformer-based architecture is the basis of GPT-3's capability to analyze complicated networks and structures. By using transformers, GPT-3 becomes feasible. In order for the model to comprehend and condense text at various levels of abstraction, it makes use of multilayer transformers. GPT-3 is one of the most extensive LMs ever made with an incredible 175 billion parameters. A training approach of GPT-3 mostly relies on Unsupervised Learning using a large amount of publicly accessible data. GPT-3's enormous training data and diminutive size give it a broad comprehension of the English language. Articles that span many topics and have human authors are within the capabilities of these systems. This is accomplished by fine-tuning the dimensions of GPT-3(Straka et al., 2016). The Bert model's architecture is shown in Figure 1.18.

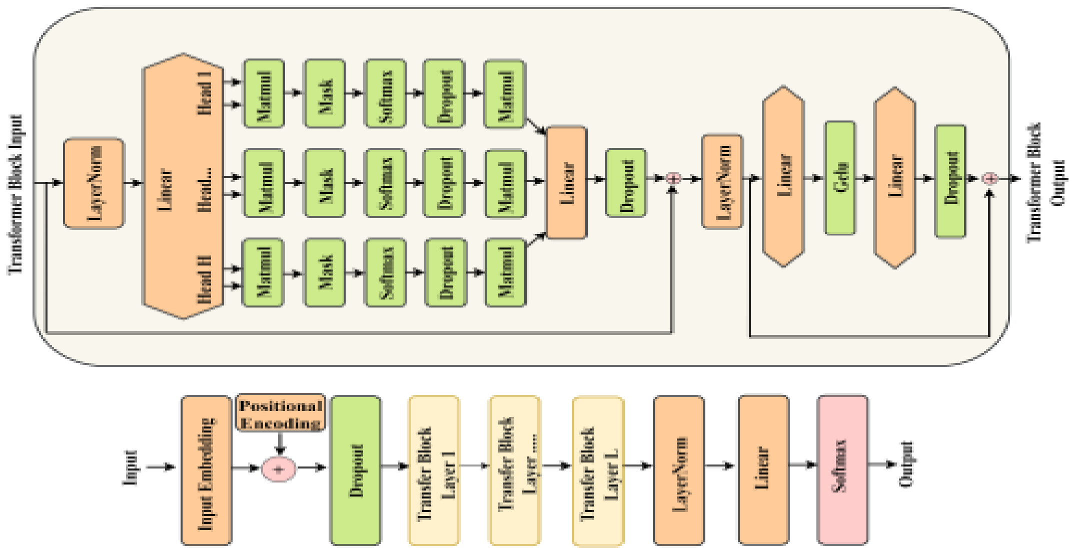


Figure 1.18: GPT-3 architecture (Zheng et al., 2021)

- **Bidirectional**

Encoder Representation of Transformer, or BERT for short, is a well-liked LLM paradigm for handling difficult NLP issues. Training, pre-training, and post-training are the three separate stages that make up the whole process (Praveen & Vajrobol, 2023). To fill out the representation in pre-learning, BERT finds a lot of redundant terms. Activities such as NSP and MLM are used in this research. Employing MLM to elucidate a collection of input tokens and thereafter training a model to anticipate the initial tokens allows one to get a deeper comprehension of both old and new material (W. Zhao et al., 2023). BERT analyses whether the second phrase follows the first expression in order to affect NSP technology to increase relationship comprehension. down order to zero down on certain tasks later on in the first training session, the BERT system makes use of data annotations. For tasks like sentiment analysis and website identification, BERT has trained agents that are tailor-made. The model's characteristics are altered via inverse processing and gradient descent. For BERT training, you'll need an FPGA, TPU, and GPU. A normalization layer, residual connection, and sound input make up the BERT Transformer's design. What this means is that the system is able to explain components' connections and interactions on a big scale(Irfan et al., 2022). Figure 1.19 shows a Bert model architecture

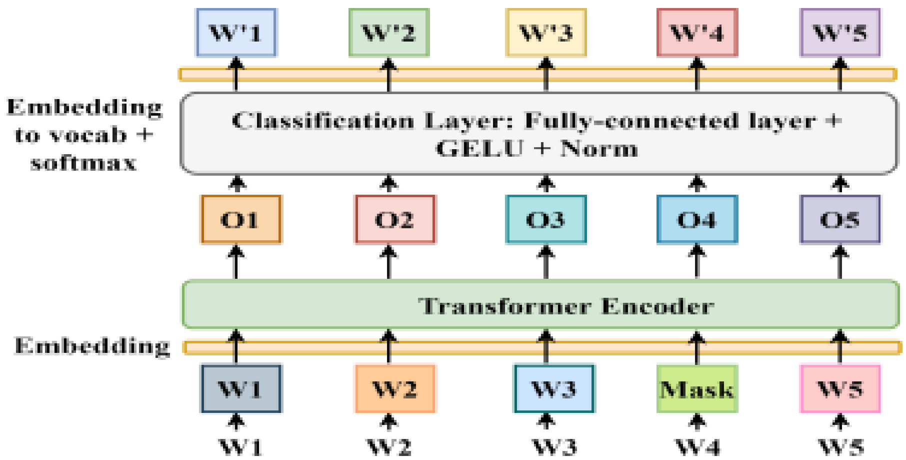


Figure 1.19: Bert architecture (Devlin et al., 2019)

- **T5**

Google created and trained T5-LM, so it's very simple to teach. T5 released a plethora of information via website and releases during the training. Making it easy to learn words with several meanings is the main objective. The beauty of T5 is that it is aware of all the tedious tasks involved in making text. A change has been made to the process of format conversion in order to make text-to-text conversion easier. Text classification, question droughting, interpretation, and analysis are all examples of such jobs. For instance, S5 is under no need to provide detailed responses. However, in order to get a full response, users need make use of all the information that is accessible to them. During its first training, T5 was trained using a variant of the Transformer architecture that was not used before. In order to guarantee the reliability of the connection between components, it is helpful to precisely characterize the connection points at the entrance using the

Transformer model in T5. Training T5 to recognize relevant text information from input is the main goal of maximum possibility estimate pre-training. T5 will undergo further adjustments in future tasks after first training. There is a lack of comprehensive documentation for the various apps, which is a problem in T5 training. The amount and quality of the data impact how well this model works (Pham & Nguyen, 2020). For precise correction, it is essential to collect information particular to the work. It may be necessary to allocate a substantial amount of time and money for T5 training. There are a lot of pre-learning parameters and transformers in the model, which makes it more computationally demanding. Figure 1.20 shows a T5 architecture.

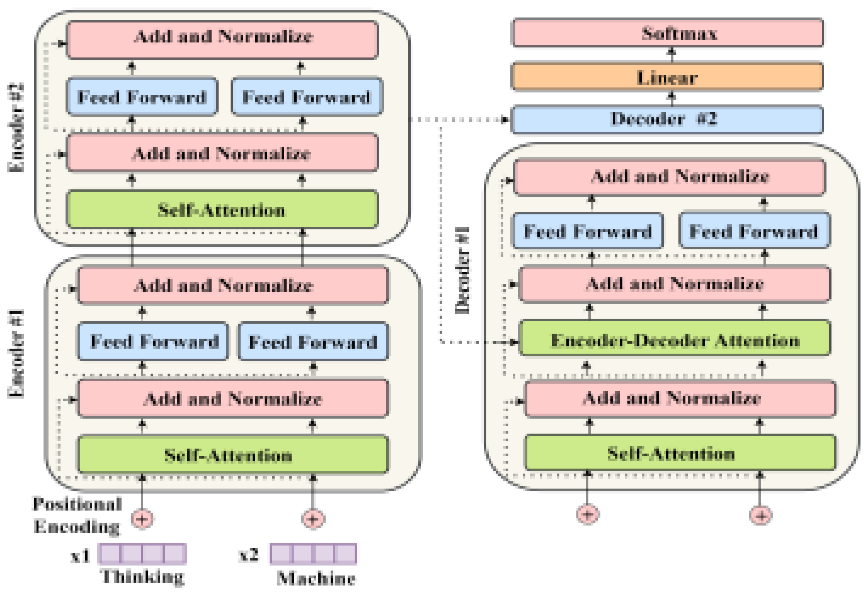


Figure 1.20; T5 Architecture (Ni et al., 2022)

1.4 Motivation

With the exponential growth of e-commerce and online reviews, businesses rely heavily on customer feedback for decision-making, product improvement, and market strategies. Nevertheless, it is still difficult to get useful insights from large volumes of unstructured text. Sarcasm, irony, and contextual variations are language intricacies that conventional sentiment analysis tools often fail to capture. The motivation behind this study is to leverage AI to enhance sentiment analysis accuracy, enabling businesses to make informed decisions. Additionally, ethical concerns regarding AI-driven sentiment analysis, including fairness, transparency, and privacy, highlight the need for responsible AI implementation in business applications.

1.5 Aim and Objectives

The aim of the study is to explore the role of AI in enhancing business through sentiment analysis by improving decision-making, customer engagement, and market insights while ensuring ethical, transparent, and accurate sentiment interpretation.

Objectives:

- To evaluate the efficacy of AI-driven transformer models for SA in online product reviews in comparison to traditional machine learning classifiers.
- To make use of text preprocessing and alter the e-commerce dataset in order to classify sentiment, making sure the text is clean, standardized, and prepared to be fed into the model.
- To assess how data balancing approaches enhance sentiment analysis models' performance and fairness, especially when addressing class imbalances.
- To use cutting-edge NLP techniques to enhance the performance of pre-trained transformer models on SA tasks.
- To compare transformer models to more conventional ML models and to evaluate their performance using a variety of measures, such as re-call, accuracy, precision, and F1score.
- To show how sentiment analysis models may be utilized in real-world scenarios to analyze customer comments and improve company decision-making and customer experience.

1.6 Research questions

This study aims to address the following research questions:

- **RQ1:** How effective are AI-based advanced models in performing sentiment analysis on e-commerce product reviews compared to traditional ML models?
- **RQ2:** What impact do advance text preprocessing techniques and data balancing have on the performance of sentiment classification models?
- **RQ3:** How can AI-driven sentiment analysis contribute to improved customer experience and strategic decision-making in future business scenarios?

1.7 Scope

This study's scope includes the use of AI for SA in corporate settings. Specifically, it focuses on analyzing Flipkart product reviews using AI-driven sentiment classification models. The study includes data collection, preprocessing of textual data, model evaluation, and assessment of class imbalance impact. It further explores advanced emotion recognition and ethical considerations like fairness, transparency, and privacy protection in AI-powered SA. The study's overarching goal is to help businesses make better decisions by enhancing sentiment categorization accuracy and using insights produced by AI.

1.8 Dissertation structure

The dissertation structure are as follows:

Chapter 2: Literature Review – Reviews the literature on sentiment analysis, AI methods, and their commercial uses; it addresses knowledge gaps in the field.

Chapter 3: Methodology – Specifies the study's methodology, including its design, data gathering steps, AI model selection, preprocessing methods, and assessment criteria. It goes on to discuss methods for better sentiment categorization and dealing with class imbalance.

Chapter 4: Results and Analysis– Presents the experimental setup with hardware and software requirements. Dataset analysis and visualization graphs, performance measures and experimental results of proposed models with deep discussion.

Chapter 5: Discussion: provide a contraction of current and suggested models for product review SA. Each response of research questions for this work.

Chapter 6: Conclusion and Future Recommendation – Provides a concise overview of the study's main points, emphasizes its contributions, addresses its shortcomings, and proposes avenues for further research on AI-driven Sentiment Analysis for corporate use.

CHAPTER II:

REVIEW OF LITERATURE

2.1 Background

There are benefits for both consumers and sellers in an online marketplace thanks to the recommendation system. Personalized product recommendations and diverse consumer interests are made possible with the help of recommendation systems, which evaluate user preferences and buying habits across a large catalogue of products to provide tailored suggestions. Collaboration or content-based approaches are used by most recommendation systems (Thomas & Jeba, 2024).

The development of efficient and extensible sentiment analysis systems is critical for any organization, whether it is focused on products or services, due to the increasing reliance on digital data. For companies of any size to understand and quickly react to customer comments or views on a variety of subjects, sentiment analysis is very important. Its ability to assist organizations in precisely gauging customer mood via feedback is crucial in many areas, including marketing, customer service, and product development (Chaturvedi et al., 2017). Professionals in many sectors (e.g., business, marketing, sports, politics, product assessment, etc.) are able to use social media platforms to access and utilize existing methods that may analyze user attitudes about certain topics (Bonifazi et al., 2023). This aids in improved decision-making, as well as monitoring and moulding a business image via internet platforms. Effective use of sentiment research methodologies helps firms to anticipate and handle possible problems beforehand, therefore creating a favourable brand image (Ghatora et al., 2024). A main focus of this research is on how SA might benefit product-based companies. It stresses the need of continuing to engage with customers via product reviews and feedback, which may help in making the right adjustments to product plans and improvements.

A goal of Sentiment Analysis (SA) is to analyze texts for subjective information by using ML and NLP techniques. Reading the tone of product reviews may give businesses a good idea of how satisfied customers are with their purchases. SA controls the arrogance of a speaker or writer in relation to the polarity of the setting, a subject, an event, a conversation, etc. Finding the text's polarity on a document or phrase level is an essential SA job. Everyone is now able to express themselves via a variety of channels, thanks to the proliferation of Internet usage. SA is useful for analyzing such subjective data and drawing important conclusions that could guide decision-making for others. There are many different types of data generated by social media sites, including evaluations of goods, services, films, healthcare, hotels, and news and articles (Elangovan & Subedha, 2023)

There are a number of ways to classify sentiment analysis methodologies, such as by methodology, dataset structure, rating level, etc. These classifications are further subdivided as follows:

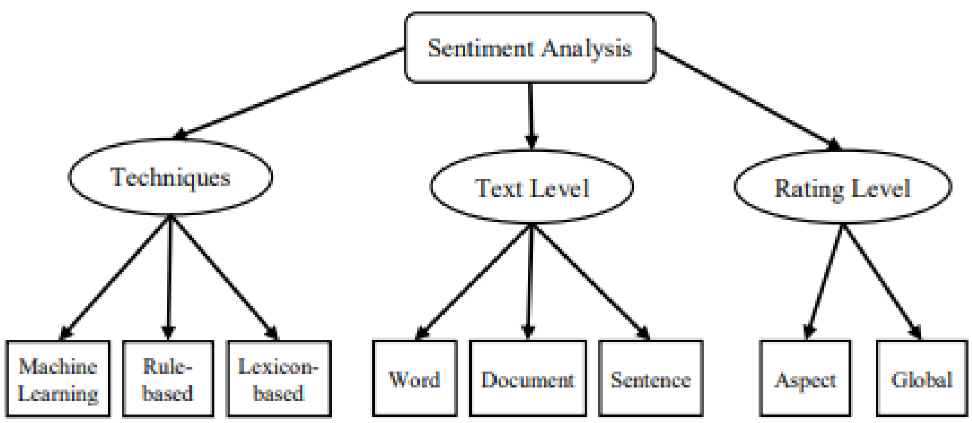


Figure 2.1: Categorization of Sentiment Analysis (P et al., 2020)

SA could theoretically be performed by either

1. **Machine Learning:** Datasets must first be trained. Standard machine algorithms are used to identify polarities (Collomb et al., 2014).

2. **Rule based:** The program attempts to evaluate data extracted from a dataset based on the polarity of words. A variety of restrictions apply, including those pertaining to idioms, dictionary polarity, emoticons, and words that negate one another (Maas et al., 2014).
3. **Lexicon-based:** By analyzing the subjective character of a review or comment and the weight people give to different opinions, it determines if the sentiment is positive or negative. Aspect, document, and sentence level SA classifications are all possible, depending on the dataset or text level structure (Sultana et al., 2023).

An opinion or perspective may be categorized as "positive" or "negative" using document-level SA. It broke down the whole document into its component parts and processed them individually. To categorize the expression of opinion or emotion in every phrase is the goal of sentence-level SA. The only real distinction between document-level and sentence-level categorization is the length of the sentence; papers are much longer. So, they have determined the "positive" or "negative" sentiment polarity of the phrase or text using a document-level Approach. There are two ways to rate a product: globally and at the aspect level. Here is an additional SA classification (Maas et al., 2014). The majority of online marketplaces and movie review sites gauge public opinion on a worldwide scale.

2.2 Related work

Previous work on e-commerce SA utilizing different ML and DL methods is reviewed and analyzed in this section.

Sentiment Analysis in E-Commerce Business Using Machine Learning

In this research Ghatora et al. (2024) explored the possibility of automating the sentiment analysis of reviews for items bought using pre-trained LLMs in conjunction with traditional ML methods. A more complex comprehension of customer attitudes is needed to facilitate data-informed business choices, which is the driving force behind this endeavor. In this work, SA was carried out employing the GPT-4 model in conjunction with other ML-based classifiers including RF, NB, and SVM. SVM is more effective than traditional models in classifying the emotions of brief comments, but traditional models perform better when processing brief, compact text. In contrast, GPT-4 performed better with more comprehensive texts, identifying mixed feelings that the simpler models were unable to detect with better precision, recall, and F1 scores. Ultimately, this work demonstrates that LLMs provide more informative and accurate sentiment classifications than conventional models in context-rich sentiment analysis. Thanks to these findings, LLMs are now a better tool for customer-centric organizations, allowing them to extract useful insights by any textual data.

In this study, Ghaffari et al. (2024) paid special attention to the scenarios involving the creation of product descriptions, the analysis of review sentiment, and the tagging and classification of products. They provide sample LLM-based solutions of these applications by using a variety of rapid engineering approaches. Finally, the paper delves into the potential dangers and difficulties that may arise from using generative AI in these specific settings.

In this research Bellar et al. (2024) Provide a comprehensive benchmark evaluation of several DL models, like RNNs, CNNs, and Bi-LSTMs. Several word embedding approaches, like Fast Text, Word2Vec, and BERT, are utilized to evaluate these models. Two configurations are included in the evaluation: 5-class and 3-class. After anticipate how consumers would feel about a product, this study compares and contrasts the performance metrics of neural network-based models using a dataset of product evaluations written by customers of an online women's clothing store.

In this study Siddique (2024) set out to use ML algorithms to sift through customer reviews posted on Flipkart in search of sentiment. The researchers in this study used NLP and ML methods to look into how to categorize and analyze product reviews based on their sentiment. Data preparation, n-gram analysis, and feature extraction from text data using TF-IDF vectorization are all part of the research. This study employs the SMOTE to tackle problems associated with unequal class representation. They used a database of evaluations written by customers. They developed a SA model and determined its superior performance by employing the following algorithms: LR, DT, KNN, and NB. By using 10-fold cross-validation, they determined how well these classifiers performed. They continued to tweak the Hyperparameters of the working classifiers. Used a number of measures, like F1score, recall, accuracy, and precision, to assess their performance.

In this research, Henderi, Henderi & Siddique (2024) seeks to evaluate several methods for sentiment classification, including LR, SVC, RF, and GB, for their relative efficacy. The dataset includes 205,052 product reviews from different categories, all sourced from Flipkart. Before feature extraction could occur, the data had to be properly preprocessed by dealing with missing values, eliminating duplicates, standardizing text, and using TF-IDF vectorization. The hyperparameters for each method were tweaked and implemented using grid search and randomized search. To guarantee robust model assessment, the data was split 80/20 between the training and testing sets, and cross-validation methods were used. Some measures were utilized to evaluate a performance of every model: recall, accuracy, precision, F1score, and ROC AUC. With an F1score of 0.8736 and a ROC AUC score of 0.9105, the outcomes showed that LR attained an accuracy of 0.8995, precision of 0.8773, recall of 0.8995, A little better accuracy (0.8997), precision (0.8619), re-call (0.8997), and F1score (0.8738) were shown by the SVC model. The RF model, however robust, has a ROC AUC score of 0.9037 but a poorer accuracy (0.7953), precision (0.6326), recall (0.7953), and F1score of 0.7047. The results of Gradient Boosting were comparable to those of LR, with ROC AUC of 0.9098, F1score of

0.8735, accuracy of 0.8993, precision of 0.8512, and recall of 0.8993. The results showed that when it came to balancing accuracy and computing economy, the two best performers were SVC and Logistic Regression.

In this study Daza et al. (2024) proposed PRISMA system, with inclusion and exclusion criteria considered. A next step is to compile all of the relevant data and draw conclusions about: sentiment analysis's practical applications; ways to evaluate the model's efficacy; relevant tactics, metrics, and programming languages; and finally, to suggest areas for further research and development. The most common model testing method was cross-validation, the most efficient technique was SVM, and LSTM had better accuracy. The most prevalent programming language for constructing models was Python, and the most significant measure was F1Score.The most common products were those involving different items. They provide a holistic perspective on ML and DL algorithms used to analyze e-commerce product reviews for sentiment, pointing out current problems and future directions to make these methods more efficient; as a result, they help make sentiment analysis more accessible and effective. Additionally, they emphasize how this study might provide businesses an edge by helping them enhance the quality of their products.

In this study (C. Wang et al., 2023) proposed that respondents from e-commerce companies take part in an online poll. The data was analyzed using the PLS Smart algorithm. They found that the widely used TAM was a good hypothetical model to explore how e-commerce users perceive and utilize AI. This study's results demonstrate that PU and PEU are favorably affected by Subjective Norms, that trust positively affects PEU, and that PEU positively affects PU and attitudes towards usage. Attitudes towards usage and intention to use are also positively impacted by PU. Additionally, there is little evidence that trust has an effect on PU and attitudes on behavioural intention to use.

Finally, behavior-based intention to use had a beneficial effect on AI technology utilization. Research like this helps fill in the gaps in our understanding of the TAM model and how it might be applied to the e-commerce industry. Incorporating the TAM model into a company's operations allows for more effective and appropriate usage of AI.

In this study Kaur & Sharma (2023) proposed doing SA employing a combination of methods. Feature extraction, sentiment categorization, and pre-processing are all components of the process. Using NLP methods, the pre-processing stage discards irrelevant information from the incoming text reviews. A new hybrid approach has been developed for efficient feature extraction. This technique builds a hybrid feature vector specific to each review by combining aspects-related and review-related characteristics. A DL classifier called LSTM is used to categorize the sentiment. In order to put the idea to the test, they used three separate study datasets. The model attains an average F1score of 92.81%, an average recall of 91.63%, and an average precision of 94.46%.

In this study Loukili et al. (2023) examined feedback from customers, marketing initiatives, and product reviews. It aids online retailers in comprehending the sentiments and opinions of their consumers about a product or service. Decisions on future goods and services, marketing initiatives, or customer service concerns may be based on this data. Systems that could analyze customer thoughts and comments on e-commerce platforms might be created and implemented using AI methods including sentiment analysis, NLP, and ML. The main objective of this research is to compare several supervised ML models for consumer SA and choose the one that works best. The dataset is centered around customer comments on various products sold by an online clothing store for women.

In this study H. Huang et al. (2023) researched the methods of SA now used by e-commerce platforms and the potential future developments in this area. Reviewing the current systematic literature revealed that not a single study adequately addressed the research topics. Researchers in the area of SA may benefit from this study's results by gaining a more thorough grasp of the methods and platforms now used, and by learning about potential avenues for future research. After locating 271 research articles using targeted keywords, they narrowed the field down to 54 experimental trials to evaluate. Of them, 24 (44. %) have looked at employing DL methods to resolve SA, 4 (7. %) have used ML and DL techniques in a hybrid strategy, and 26 (48. %) have simply employed ML techniques. It was also found in their analysis that academics preferred Twitter and Amazon as data sources. Potential areas for further research include developing more all-encompassing language models, aspect-based SA, sarcasm detection, implicit aspect extraction and identification, granular sentiment analysis, and more.

In this study, Chath & Chhatrala (2023) might benefit both e-commerce platforms and the academics who work there by using and analyzing the sentiment of consumers' product assessments. One of the most important parts of managing product quality in internet commerce is SA of reviews. Polarity analysis of customer evaluations is an emerging area of sentiment analysis. Both good and negative reviews help businesses learn more about their consumers' experiences with various items and to identify the most profitable ones. After reviewing the literature, this study builds a model called MDLNNM to automatically identify and categorize the polarity of sentiment in the Product Review Dataset. This system has the potential to categorize user reviews according to product review characteristics, utilizing review data. Companies can now track their consumers in real-time and get advice on how to understand them better, according to the study. As a result, they could enhance their product offerings and provide consumers trustworthy assistance after the purchase.

In this research, Khin Sandar Kyaw et al. (2023) uncovered that Sentiment Analysis technology, which is part of business intelligence, could be a useful tool for understanding consumer behaviour through opinion mining. This process involves analyzing customer information and feedback using tools like biometrics, computational linguistics, NLP, and text analysis. This research looked at SA in e-commerce systems from a technological perspective and found that gathering business information is important for effective digital marketing on e-commerce platforms.

In this study, Gupta et al. (2023) shown a method that prioritizes the item's emotional qualities. They have presented and assessed consumer reviews on IMDB and Amazon. They retrieved the data set from the repository at UCI, where the opinion ratings for each study are first recorded. The system eliminates noise and valuable information from datasets by performing pre-processing processes such tokenization, punctuation, whitespace, special character, and stop-word removal. By employing feature selection methods such as TF-IDF, the pre-processed data is accurately shown. Using classifiers such as NB, RF, KNN, and SVM, the classification procedure entails combining consumer evaluations from three datasets: Amazon, Yelp, and IMDB. Finally, they mentioned some plans for future text categorization projects.

In this study, Benarafa et al. (2023) developed a strategy to improve KNN for handling the IAI problem. Their solution to the IAI issue is an enhancement to KNN distance calculation that makes advantage of WordNet semantic relations. To reach a definitive empirical assessment, experiments are carried out on two datasets containing reviews of restaurants and electronic products. They examine and assess their method's efficacy with respect to three metrics: K, the KNN distance utilized to determine similarity, and KNN behaviour with regard to Overfitting and Underfitting. Their strategy increases KNN's performance and handles Overfitting and Underfitting for IAI better, as shown by the experimental results.

In this study Marwat et al. (2022) provide a Senti Deceptive algorithm that can automatically sort user evaluations into three categories: unfavourable, positive, and neutral. It can also detect misleading data from crowd ratings on social media. Annotating 17,781 comments found on Amazon and Flipkart using content analysis and the VADER library is the basis of the method. The sentiment analysis technique trains ML algorithms, eliminates unnecessary data, uses resampling, and extracts features. The automated technology helps with decision-making by graphing consumer sentiment. For positive sentiment categorization, the method produced an average precision of 94.01%, recall of 93.69%, and F-measure value of 93.81%.

In this study, Alzahrani et al. (2022b) analyzed the sentiment of e-commerce reviews employing CNN-LSTM models, which mix DL with LSTM. The system was evaluated in real-time using reviews of various electronic devices sold on Amazon, including cameras, laptops, Mobile Phones, tablets, televisions, and Video Surveillance items. For data cleaning, preprocessing techniques were used, including processing in lowercase, removing stop words and punctuation, and tokenizing. To classify the consumers' positive or negative sentiment, the clean data was analyzed employing the CNN-LSTM and LSTM models. The CNN-LSTM algorithm achieved 91% accuracy, and the LSTM algorithm 94% accuracy. They conclude that the DL techniques used here provide the most accurate results for categorizing consumer sentiment towards the products.

In this study, Chamekh et al. (2022) extracted sentiment from product reviews, ratings, and suggestions by combining several deep learning techniques with a sentiment lexical corpus. To learn how people feel about the merchandise, they looked at an online shopfront in this project. The stats used in this research come from 41,778 French product reviews for smartphones that were gathered from Amazon.com. Their method for review categorization was the LSTM. With a 95% accuracy rate, the findings demonstrated that the LSTM deep learning system produced satisfying results.

In this research Bilal Chandio Asadullah Shaikh (2022) offered a Roman Urdu Stemmer-powered SVM that has been fine-tuned. They suggest starting by cleaning the corpus data to get rid of any text oddities. Every user review is stemmed after the first pre-processing. The bag-of-word model takes the input text and turns it into a feature vector. After that, they apply the SVM to identify and categorize user sentiment. The Roman Urdu stemmer that they advocated for is based on a dictionary. Aiming to standardize the text in order to minimize complexity, the Roman Urdu stemmer is created. To obtain better performance, they empirically test their suggested model using various experimental setups to find the optimal hyperparameters. In addition, they run a battery of tests comparing the suggested model's performance against that of several DL and ML models. The Roman Urdu e-commerce dataset (RUECD), which includes over 26,000 evaluations annotated by a team of professionals, is the biggest dataset on Roman Urdu, and it was also introduced by them. As the biggest dataset of Roman Urdu currently accessible, the RUECD presents a formidable challenge. According to the results, Roman Urdu sentiment analysis is going to need a lot of help from other scholars since the newly created dataset is so difficult.

In this study Ray & Chakrabarti (2022) suggested a DL method for analyzing user sentiment related to text characteristics for feature extraction. Opinionated utterances are sorted using a seven-layer deep CNN. The sentiment score and aspect extraction methods have been improved by combining a DL approach with a set of rule-based procedures. Furthermore, they evaluated their proposed method against existing standards and explored methods to enhance the present rule-based approach to aspect extraction via the use of clustering and a predetermined set of aspect categories for aspect classification. They surpassed current advance methods with an overall accuracy of 0.87, compared to 0.75 for the modified rule-based technique and 0.80 for CNN. They found that their proposed approach had an overall accuracy that was 7-12% higher than state-of-the-art techniques.

In this research Singh et al. (2022) analyzed the Amazon Dataset employing a range of DL and voice component setups. Identifying whether a statement is "Positive," "Neutral," "Negative," or "Indifferent" is the primary focus of the scheduled module. After going over the numbers, it assigns a positive and negative label to the "better" and "worst" assumptions, respectively. Consumers now have access to a plethora of items within the same domain, due to the proliferation of e-commerce websites and the internet, and NLP is crucial in categorizing products according to reviews. NLP and ML can distinguish between paid and unpaid reviews. Several ML algorithms were tested in this study to see how well they could forecast the tone of customer evaluations of online stores.

In this study H. He et al. (2022) planned to use a fusion Sentiment Analysis strategy that integrated textual analysis with ML algorithms in order to glean online product experience. The three basic stages make up the technique. To start, they take use of the sentiment dictionary's sensitivity to emotional data by extracting sentiment characteristics from it. Afterwards, the SVM approach is used to ascertain the polarity of the reviews' sentiment. The LDA model then uses this information to derive themes related to reviews' sentiment. The exclusion of emotive details is further avoided by expanding the vocabulary according to semantic similarity. In the meanwhile, this study considers the fact that words in reviews contribute differently to emotion, a point that has been overlooked in previous research. To quantify the sentiment contribution, they specifically provide the weighting approach. Lastly, an analysis of Amazon customers' online book reading experiences has confirmed the method's viability and accuracy. The outcomes shows that a method accurately detects the subjective nature of reviews and the elements impacting readers' perceptions. Overall, the study offers a useful method for tracking consumer requests and mining online product experience, which will greatly aid in future product development and marketing plan optimization.

In this research Gope et al. (2022) classified the textual expressions of good and negative emotions. Reviewers on e-commerce platforms like Amazon.com might provide a star rating. Their intention in doing this study was to construct Sentiment Analysis using Amazon's data in conjunction with product ratings and text reviews. Several ML techniques were used, including LR, MNB, BNB, RF, and Linear SVM. Using the RFC, they achieved a 91.90% accuracy rate. Their study also achieved maximum accuracy (97.52%) using a DL strategy that combined RNN with LSTM. The RNN-LSTM method is perfect for their model.

In this study, Mykhalchuk et al. (2021) uncovered the theoretical underpinnings of online shopping as a distinct economic phenomenon, the issues plaguing the online marketplace at the moment as an outcomes of the COVID-19epidemic, customer mistrust, and a lack of practicality in apparel unit selection. Improving the efficiency and utility of online commerce is the overarching goal of this article's expert system development, testing, and deployment efforts. Research methods encompass a huge range of tools and techniques, including mathematical apparatus for classification problem solving, statistical methods and tools for data processing and analysis, data visualization tools, artificial neural networks, decision trees, and discrete methods. The development of an expert system to enhance e-commerce functionality was accomplished via the utilization of R and Python programming languages, Data Processing, and analysis. The study has practical implications in three areas: finding and employing an effective algorithm for text SA; improving the technique of collaborative filtering; and developing an information system to provide suggestions to customers. Consequently, e-commerce became more efficient and functional. The obtained research results could serve as a methodological road map for software engineers developing recommendation systems.

In this study Demircan et al. (2021) with the goal of extracting sentiment from social media messages using ML techniques. Initial investigation has shown that online product evaluations and ratings provide the most compelling example of a situation where words and feelings are congruent. A table including product reviews and review ratings obtained from an online store has been prepared for use in sentiment analysis models that rely on machine learning. They utilized the review scores to sort the reviews into three piles: positive, negative, and neutral. It was with this assertion in mind that models for Turkish sentiment analysis were built, using KNN, LR, DT, and SVM. The SVM and RF-based Sentiment Analysis models performed better than the competitors were tested on many sets of test data from the same online retailer using cross-validation. To be more precise, there is no strict order to the scores produced by SVM- and RF-based prediction models; nonetheless, when contrasted with scores from DT- and LR-based models as well as KNN-based models, the results from SVM- and RF-based models are often excellent or, at the very least, similar. Within reasonable error limits, SVM and RF seem to be practical approaches for categorizing product evaluations as either positive, negative, or neutral.

In this research Akter et al. (2021) put forth a model that uses ML to predict if a user would have a good, neutral, or negative attitude towards a Bangla text review. They used five ML algorithms on a dataset that had been manually compiled from the Daraz e-commerce platform in Bangladesh. Utilizing their dataset, they conducted experiments with several algorithms such as XGBoost, RF, LR, SVM, and KNN. When compared to the other four algorithms, KNN takes first place across the board for accuracy, precision, recall, and f1score. At 96.25 percent accuracy, KNN also has a recall of 0.96 and a f1score of 0.96.

In this study Wassan et al. (2021) introduced a fresh method that makes advantage of emotional factors centering on the item's attributes. Amazon customer reviews are real and verified. Opinion rates are first found in each study in the dataset that they obtained from the Data World Centre. The technology removes stop-words, stone coating, tokenization, and boxing from datasets before extracting significant information like positivity or negative. Researchers hope that by looking at the data at the aspect level, marketers would be able to better understand customer preferences and adjust their behaviour appropriately. At last, they reveal a little about their next text classification project.

In this research N. C. Dang et al. (2020) provided us with useful data. Social media platforms such as Facebook and Twitter have many practical uses for sentiment analysis. The difficulties in NLP are limiting Sentiment Analysis’s effectiveness and accuracy. A new line of inquiry suggests that DL models may be the key to solving NLP's problems. This report summarizes current DL-based research on Sentiment Analysis challenges, such as sentiment polarity. Word embedding and TF-IDF models have been applied to several datasets by them. Experiment results were finally compared across all models and input attributes.

In this study, Geetha & Karthika Renuka (2021) LSTM, SVM, and NBC were utilized to categorize reviews from different classification models. A majority of the existing SA methods used to train on this kind of customer-generated product review text data are inaccurate and time-consuming. This research introduces BERT Base Uncased, a powerful Deep Learning Model, to help understand Sentiment Analysis better. In the experimental assessment, the BERT model outperformed other ML techniques, demonstrating great prediction and high accuracy.

In this research, Shah et al. (2021) Analyzed the feelings in this technology era about internet product reviews, the vast majority of individuals use online reviews. They provide their opinions, and based on that, things are suggested for sale or purchase. Customers have the option to provide product reviews on several popular e-commerce sites. This includes Amazon, Flipkart, Myntra, and many more. The buyer was expected to have a thorough comprehension of the product and how it functions before making a purchase. It was seen as a very straightforward product with three distinct states: positive, neutral, and negative. It is possible that they will do this experiment employing ML techniques. In sentiment analysis, participants are asked to rate their level of awareness of a product's impact on them. The data utilized is sourced from a Kaggle of product reviews on Amazon. When it comes to feedback classification, they utilize a variety of NB, RF, and LR approaches to get the greatest precision. They found that the RF method outperforms all the others when it comes to ML.

In this study, C. N. Dang et al. (2021) demonstrated and assessed a strategy for making recommendations that incorporates Sentiment Analysis with collaborative filtering techniques. The suggested recommender system's adaptive architecture is built on top of DL models based on Sentiment Analysis and new approaches for feature extraction. Collaborative Filtering methods and sentiment-based DL models may substantially improve the recommender system's efficacy, according to the empirical study that employed two shared datasets.

In this study, Mukherjee et al. (2021) provided an innovative approach to negation handling using end-to-end sentiment analysis, including features such as negation scope marking and negation identification. They use a variety of ML techniques, including ANN, RNN, SVM, and NB, to test out Sentiment Analysis on Amazon mobile phone evaluations. Additionally, they showcase an approach for explicit negation detection that is customized to negation marking. When tested on Sentiment Analysis tasks, RNNs with negation marking processing outperformed those without it, reaching an accuracy of 95.67%. This was in comparison to the RNNs' accuracy levels when negative phrases were not identified. Their method also showed a considerable increase in general accuracy when applied to a different dataset of Amazon reviews.

In this research Nasiri & Shokouhyar (2021) Conducted research on the CSDs pertaining to refurbished cellphones and offered a model for customer satisfaction with regard to reviews found on e-commerce platforms. The most concerning aspects of the product, according to the results, are its function, look, and battery health. The primary drivers of purchase are the reduced costs and resemblance to new items. Perceived motivation, quality, utility, and risk all influence how much refurbished cellphones are thought to be worth. Product development and marketing tactics might benefit from solutions that enhance consumer perceptions and lessen misunderstandings.

In this study, Behera et al. (2021), offered a method for sentiment categorization of reviews submitted across several domains that combines 2 DL architectures: CNN and LSTM (RNN with memory). While LSTM recurrent networks are great at sequentially analyzing lengthy texts, deep convolutional networks have been shown to be very effective at local feature selection. There are two primary goals in sentiment analysis that the suggested Co-LSTM model attempts to accomplish. To start with, it can easily adjust to scalability concerns when analyzing massive social data, and secondly, it doesn't limit itself to any one topic, unlike traditional machine learning methods. The model, which can handle all types of dependencies that often occur in a post, was trained using four review datasets from various disciplines. The proposed ensemble model outperforms the previous ML approaches in the experiments.

In this study S. M. AlQahtani (2021) Examined the Amazon reviews dataset and performed research on sentiment categorization using various ML approaches. Several approaches were used to transform the reviews into vector format as the initial stage. A bag-of-words, Tf-Idf, glove etc. Many ML algorithms were subsequently taught. The following terms are used: logistic regression, BERT, bi-LSTM, naïve bias, and RF. Then, they used the cross-entropy loss function, recall, accuracy, precision, and f1score to assess the model. Before delving into its sentiment categorization, they examined the top performing model. They retrained the top-performing model for binary classifications after discovering it had done well on multiclass classifications.

In this research H. Zhao et al. (2021) developed a novel ML technique for analyzing the tone of online product evaluations; it is known as the LSIBA-ENN. Data collection, data cleaning, Feature Extraction, Feature Selection, and sentiment classification make up the four steps of the algorithm. The Web Scrapping Tool is utilized to get the data by online stores. Two hybrid mutation-based algorithms, LTF-MICF and HMEWA, are used to apply TW and FS processing on the pre-processed data. Customer reviews may be categorized as good, negative, or neutral by the LSIBA-ENN. A Sentiment Analysis benchmark test found that the LSIBA-ENN outperformed other cutting-edge algorithms.

In this research, Kabir et al. (2021) used popular ML techniques that are examined and contrasted with one another. The next stage is to test these strategies using online user evaluations from different sectors to see how they perform. The study makes use of data sets sourced from many websites, including Amazon, Yelp, and IMDb. SVM, DT, Bagging, Boosting, RF, and Maximum Entropy are some of the well-known approaches used in the trials. The results of the experiments show that users can get useful information out of review datasets for BI and improved product sales production, and that the two ML algorithms Boosting and Maximum Entropy are the best at identifying user sentiment in online reviews.

In this research Liu et al. (2020) presented the Bert- BiGRU- SoftMax DL models, which incorporate the Sentiment Bert model for extracting multi-dimensional E-commerce reviews, the Bidirectional GRU model for obtaining semantic codes and calculating review representation weights, and the SoftMax with attention mechanism for sentiment tendency classification as the output layer. They used a large dataset with over 500,000 reviews to test several learning models. Experiments show that the proposed models outperform RNN, BiGRU, and Bert-BiLSTM according to accuracy and loss, and they reach an impressive 95.5% accuracy rate on the E-commerce reviews.

In this study, Kumar et al. (2020) examined an effects of age and gender on Sentiment Analysis, since this may help e-commerce businesses target certain demographics with their product marketing. The dataset is created by asking Facebook users to complete a questionnaire that asks about their book tastes, age groupings, and gender in order to gather the reviews of books. The study next looks for emotions in the segregated data according to each gender and age group. Lastly, sentiment analysis is conducted using several ML approaches, including Maximum Entropy, SVM, CNN, and LSTM, to study the effect of age and gender on user assessments. Researchers have used experimental methods to learn how gender and age influence Sentiment Analysis.

In this study, Jagdale et al. (2019) pulled data from an Amazon database that includes evaluations written by customers on a broad variety of electronic devices, including TVs, security cameras, tablets, laptops, and computers. After preprocessing, reviews are categorized as good or unfavourable using ML approaches. This study's results demonstrate that ML Techniques are the most effective for categorizing product reviews. The SVM obtained 93.54 percent accuracy for Camera Reviews, whereas Naïve Bayes reached 98.17 percent.

In this study H. Nguyen et al. (2018) assessed models for text sentiment categorization via supervised ML and lexicon-based approaches, with a focus on TF-IDF vectorization. The paper evaluates and contrasts six algorithms: LR, SVM, Gradient Boosting, Pattern, VADER, Senti Word Net, and Valence Aware Dictionary and Sentiment Reasoner. Customers' reviews on Amazon serve as the foundational dataset. All metrics indicate that the three ML models perform better than the lexicon-based models. Models like as SVM, Gradient Boosting, and LR provide better outcomes according to accuracy, precision, recall, and F1 scores compared to Pattern, VADER, and Senti Word Net. While the lexicon-based findings are poor, the machine learning outcomes are somewhat better than the results of prior studies.

In this research Bayhaqy et al. (2018) compared the Sentiment Analysis labels of Twitter messages made by e-commerce customers using data mining methods. Tokopedia and Bukalapak Tweets on E-Commerce make up the dataset. Text mining methods such as classification, tokenization, stemming, and transform are used to construct sentiment analysis and classification models. Utilizing the DT, K-NN, and NBC techniques, Rapidminer can help with sentiment analysis and comparison, utilizing three alternative dataset classifications to obtain the greatest accuracy. With a recall of 64%, a precision of 88.50%, and an accuracy of 77%, the NB technique achieved the best result in this investigation.

In this research Khedkar & Dhande (2018) noticed that reviews must be categorized as either positive or negative. One area of computer research that seeks to glean subjective information from text is called Sentiment Analysis. Using Sentiment Analysis, more than 4,000 evaluations were categorized as favourable or negative for the planned research. The reviews have been categorized using three of the available classification models: Decision Tree, NB, and SVM. Models are evaluated by means of 10-fold cross-validation.

Sentiment Analysis in E-Commerce Business Using Deep Learning

In this study Moore (2025) investigated the methodology behind sentiment analysis utilizing ML techniques and a dataset consisting of Amazon Product Reviews. In this study, the sentiment of Amazon product evaluations is analyzed using several ML algorithms, including GB, LR, NB, and RNNMS. Punctuation removal, stop word filtering, and text tokenization are the first steps in preprocessing the dataset. After that, techniques like BoW are used to extract features. The models are evaluated employing the F1score, re-call, accuracy, and precision with the data separated into training and testing sets. Gradient Boosting maintains a steady 82% across all criteria, showcasing its exceptional classification capabilities, surpassing all other evaluated models. Despite GB's top performance, the results suggest that more complex models and methodologies might be investigated in future research to enhance sentiment classification accuracy over a wider range of product types.

In this research Thomas & Jeba (2024) provided a fresh approach to product suggestion using SA and CF as its foundation. LSTM model was used to conduct the SA. Built on top of CF were two separate recommendation algorithms. The top Recommendation System was used in conjunction with the suggested SA model to improve the suggestions. The experiment results proved that the suggested strategy outperformed the previous product suggestion techniques. The results showed that e-commerce systems might benefit from a combination of CF and SA in terms of customer happiness and product suggestion.

In this study, Wasif & Pearl (2024) investigated the effect of time on user behaviour via the integration of three datasets. Online shopping and customer happiness are on the rise, and they've seen that more and more consumers are taking the time to offer feedback. Using a time series analysis, they find that customer star ratings change over time, creating a column connection. They then use the ARIMA approach to model this relationship and forecast future trends. In addition, they look at how reviews affect consumer reactions and discover that bad reviews travel like a virus via social media. They show that bad reviews significantly affect consumption patterns by using the SIR virus model, which is a model for text-based network spread. Last but not least, they analyze the sentiment in customer reviews employing sophisticated DL models for NLP, including BERT, LLM, in addition to GPT. Optimism and pessimism are distinguished by using word vectors using classification techniques. Their study provides a more thorough knowledge of customer sentiment and product perception by exploring the relationship among star ratings and these feelings.

In this research, X. Zhang et al. (2023) demonstrated a variety of ML and DL techniques, like DT, RF, SVM, Generative Adversarial Networks, and CNN. The next section provides a brief overview of the key points, which include areas such as recommendation systems, sentiment analysis, picture recognition, product categorization, sales forecast, customer churn prediction, fraud detection, and false review detection. Finally, they go over the most important issues and current developments concerning unbalanced data, generalization and over-fitting, multi-modal learning, interpretability, customization, chatbots, and virtual assistants. The present and future of ML and DL as they pertain to online commerce is briefly summarized in this study. The ever-changing e-commerce industry brought new possibilities and threats, necessitating more research and development.

In this study, Hicham et al. (2023) analyzed the sentiment in Arabic using CNN, LSTM, GRU, BiGRU, BiLSTM, CNN-BiGRU, CNN-GRU, CNN-LSTM, and CNN-biLSTM. Two large datasets, the HARD and the BRAD datasets, are used to test the proposed model. The outcomes proved that the given model could decipher the Arabic expressions of emotion. First, the Arabert model characteristics are extracted as part of the suggested approach. Next, they created and trained a total of nine DL models: CNN, LSTM, GRU, BiGRU, BiLSTM, CNN-BiGRU, CNN-GRU, CNN-LSTM, and CNN biLSTM. The Fast Text and GLOVE word embedding models are combined. Their method was superior than the two most popular DL approaches by a margin of 0.9112.

In this research, Das et al. (2023) looked examined how well different ML models—including hybrid and deep learning ones—classified English and Bangla texts. A popular Bengali e-commerce site called "DARAZ" has reviews posted in both Bangla and English; this study aims to analyze the sentiment of these remarks. Results from comparing several models, with a special emphasis on Sentiment Analysis, will be the primary contribution of this study. The research method employs a total of seven ML/DL models, including LSTM, Bi-LSTM, Conv1D, and a hybrid of the two. Applying preprocessing approaches to an updated text collection enhances the model's accuracy. SVM models were shown to be the most effective in the investigation. Using the Porter stemming approach, the SVM models were able to obtain an accuracy of 82.56% for English Text Sentiment Analysis and 86.43% for Bangla Text Sentiment Analysis. A Bi-LSTM-based model also performs better than the other DL models. Using Porter stemming, it attains an acc-uracy of 78.10 percent for English Text and 83.72 percent for Bangla Text. By improving tet categorization models and procedures, this work marks a substantial advancement in Bangla-specific natural language processing research. This study's findings provide important new information for sentiment analysis and its future studies as well as practical applications.

In this study, Elangovan & Subedha (2023) identified and extracted sentiments and views from text using techniques for ML and NLP. It is challenging for ML models to correctly categorize feelings in online reviews due to the prevalence of informal language, slang, and accents. Furthermore, the analysis might be further complicated if poor grammar or misspelt words are used. Sentiment analysis might benefit from the most current advancements in Deep Learning (DL) models. This study introduces a technique for effectively classifying feelings in product evaluations using an APGWO-DLSA. To start, the product reviews were pre-processed using data using the word2vec embedding procedure to make them better. The suggested technique utilized a Deep Belief Network (DBN) model for sentiment categorization. Before finishing, the APGWO algorithm was used to tune the DBN's hyperparameters. Results from a comprehensive experimental study showed that APGWO-DLSA outperformed competing approaches, with a maximum acc-uracy of 94.77% and 85.31% on the CPAA and AP datasets, respectively, demonstrating the superiority of the algorithm.

In this research Mutinda et al. (2023) demonstrated Lebert, a model for sentiment classification that combines BERT, CNN, N-grams, and a sentiment lexicon. Vectorization of words from a subset of the input text utilizing sentiment lexicon, N-grams, and BERT is the first step in the model implementation. CNNs are used to classify output sentiment as well as for feature mapping. They put the suggested model through its paces using three publicly available datasets: reviews of restaurants on Yelp, films on IMDB, and Amazon.com. There are three metrics used to evaluate the model's performance: precision, accuracy, and F-measure. Compared to the current methods, the proposed LeBERT model outperforms them in binary sentiment categorization. An F-measure score of 88.73% is achieved.

In this research, Chath & Chhatrala (2023) CNNs and RNNs, two suggested DL models, have been used to analyze text sentiment with rather impressive outcomes. This work describes a CNN and RNN architecture that is joined to analyze sentiment in short texts by using the long-distance dependencies learnt by RNN and the coarse-grained local characteristics produced by CNN. With accuracy of 82.28%, 51.50%, and 89.95% on three benchmark corpora—MR, SST1, and SST2—experimental findings clearly outperform the state-of-the-art.

In this study Sayeed et al. (2023) investigated and assessed the Sentiment Analysis uses of the BERT model, an NLP method, in several domains. Studies including a number of languages, restaurant chains, farms, Automated Essay Scoring (AES), Twitter, and Google Play have made use of the approach. An important aspect of refining the BERT model is using pre-trained BERT to finish different language comprehension tasks. The data is cleaned up and converted to numbers by text pre-processing before it is sent into BERT, which creates vectors for every input character. In comparison to other broad language comprehension algorithms, BERT fared better on many tests measuring linguistic acceptability, sentiment analysis, paraphrase identification, and question answering. The model's accuracy is affected by two issues: locating unbiased reviews and determining if the dataset contains any false reviews. The lengthy training time is caused by the enormous size and the abundance of weights that need updating. In order to train the algorithm to detect neutral reviews more effectively and build a model for false review categorization, more research could be undertaken to improve the BERT model's performance.

In this study, Esfahanian & Lee (2022) assessed the effectiveness of the shipment according on the feedback provided by the recipients. Using SA, one may ascertain whether reviews are positive or unfavourable. In addition, the suggested approach finds packaging-related reviews by drawing on an internal dictionary (Pack-List) that contains terms connected to packaging. Therefore, this technique offers a structured way to find issues by studying consumer evaluations as feedback on a packaged product that is flawed. They may analyze the packing performance throughout delivery by calculating the proportion of negative and positive evaluations employing the findings of SA with Pack-List. The impact of time on packing failure is then determined by looking at the failure rate throughout different months and years. The most often cited problem is then shown in a word cloud of negative statements. The suggested approach would allow package designers to monitor the performance of their product and spot issues early on.

In this research W. Huang et al. (2022) suggested the ERNIE-BiLSTM-Att (EBLA) model for sentiment analysis. In this case, the BiLSTM is utilized to extract text characteristics from the dynamic word vector that was created using the ERNIE word embedding model. After that, the concealed layer's weight is optimized using the Attention Mechanism (Att). This sentiment categorization system concludes with SoftMax as its output layer. With F1 values above 0.87, precision exceeding 0.87, and recall exceeding 0.87, the proposed model surpasses traditional DL models that previous researchers have suggested using the JD.com Chinese e-commerce product review dataset. This suggests that it could be useful for SA in this context.

In this research Kathuria et al. (2022) used the ML model and NLP concepts to assess sentiment in reviews and evaluations of fashion e-commerce products. Understanding customer behaviour in a digital environment is the broad focus of the study. Consumer sentiment and purchase likelihood are examined in relation to eWOM and product reviews. They have gone a step further by showing that ratings, reviews, and suggested products all work together. An exploratory analysis supported by sound logic is included in the study. Applying ML classification models like LR, ADAboost, SVM, NB, and RF to customer evaluations is something they have done. For sentiment analysis, they have also used Vader and text blob technology.

In this research B. Nguyen et al. (2021) presented a strategy for gleaning consumer feedback and sentiment from a dataset consisting of 236,867 Vietnamese reviews posted on diadiemanuong.com and foody.vn between 2011 and 2020. The best ML model was then selected after its application and evaluation. Results: The outcomes of the experimental investigation displays that the suggested method can achieve an accuracy of up to 91.5%. Conclusions: Insight into consumer satisfaction with goods or services and understanding of their sentiments might be provided by the study findings, which could aid company managers and service providers in making changes and right business choices. And it aids food e-commerce managers in making sure their services are designed and delivered better for e-commerce.

Luc Phan et al. (2021) provided a step-by-step guide on developing a Vietnamese social listening system using aspect-based sentiment analysis, beginning with the collection of relevant data and ending with a working application. The Vietnamese Smartphone Feedback Dataset (UIT-ViSFD) is their first product. It is built on a strict annotation methodology for evaluating aspect-based SA and is freely available for researchers to use. There are 11,122 user-annotated comments in the dataset that pertain to mobile e-commerce. Additionally, they lay out the necessary steps to apply the Bi-LSTM architecture that incorporates fast Text word embeddings to the aspect-based sentiment job in Vietnamese. Based on their trials, their technique outperforms various traditional ML and DL systems, achieving an F1-score of 84.48% on the aspect task and 63.06% on the sentiment task. Finally, they used the highest performing model on their dataset to construct SA2SL, a social listening system that will hopefully lead to the development of additional such systems.

In this research work, Karthik & Ganapathy (2021) presented a novel product suggestion system that utilizes fuzzy logic to automatically determine, depending on users' present preferences, which goods would be most suitable for online buying. An innovative method for calculating a product's emotional score with respect to its target market is introduced in this study. In conclusion, the suggested fuzzy rules and ontology-based Recommendation System make better use of ontology alignment to produce smarter decisions and provide personalized, ever-changing predictions depending on the search environment. The suggested Recommendation System outperforms the present methods in two areas: the time required to deliver such recommendations and the forecast accuracy of the relevant items for target clients.

In this work Oktaviani et al. (2021) intended to assess and enhance the future product's quality. A tool like sentiment analysis might help with that by determining if the review was good or negative. Data extraction can be possible with the use of Sentiment Analysis. Those in need may be able to glean helpful information from the recorded data. The following steps are included in some phases of sentiment analysis: gathering sentiment data, preprocessing the data, using TF-IDF word weighting, scoring sentiment, classifying review data using the NBC algorithm, and text association. Ten 10-fold cross-validations were used to assess the model. The Confusion Matrix was utilized for measurements. Overall, the MNB achieved 91.20 percent accuracy and 59.56 percent kappa accuracy when applied to the reviews provided by Traveloka customers on Google Play. The classification model's effectiveness improves as the total accuracy value and kappa accuracy rise.

Rong et al. (2021) proposed using a Sentiment Analysis model that is based on DL to analyze data collected from internet product reviews. In their model, you may find both favourable and bad remarks. Upon text segmentation, word vectors and word frequencies are inputted into the NN for training purposes. CNNs are trained to identify emotions by tapping into the strong correlations between various data sets and these emotions. Based on the results of the experiments, the model is a good tool for evaluating internet reviews as it can accurately extract product attributes and identify sentiment.

In this research Wedjdane et al. (2021) aided consumers by providing a roadmap for them to follow while making a choice. They may examine previous customer reviews using sentiment analysis and create a chart showing both positive and negative feedback. The optimal outcome was achieved by using five classifiers: Deep Learning, Naïve Bayes, KNN, SVM, and Decision Tree. This review analysis utilizes Python and RapidMiner to sift through customer feedback found on Sephora.com. A website's 10,000 reviews are compiled here. Classifiers in SA using DL and DT achieved an accuracy of over 80% and an F1 measurement of 60%, according to the results. A test's accuracy may be measured using the F1 measure. Additional data might be gathered for future improvements; no data imbalance was introduced.

In this research, Priyadarshini & Cotton (2021) developed a new model for SA using deep neural networks based on grid search and LSTM and CNN. The research takes into account baseline methods that have been tested on various datasets using metrics like F-1 score, accuracy, precision, sensitivity, and neural networks, such as CNN-LSTM, LSTM-CNN, CNN, and K-nearest neighbour. With an overall accuracy better than 96%, their findings demonstrate that the hyperparameter optimization-based model they suggested surpasses alternative baseline models.

A framework is presented by Onan (2021) that offers a method for analyzing product evaluations' sentiment using data collected from Twitter using deep learning. This approach combines CNN-LSTM with weighted Glove word embedding using TF-IDF. The CNN-LSTM architecture consists of five layers: a convolution layer with 1-g, 2-g, and

3-g convolutions; an LSTM layer; a dense layer; a max-pooling layer; and a weighted embedding layer. Combining conventional DNN architectures with several word embedding schemes, such as word2vec, fast Text, Glove, LDA2vec, and DOC2vec, has allowed for an empirical evaluation of their predictive performance. The weighting functions include IDF, TF-IDF, and smoothed inverse document frequency function. Experimental findings demonstrate that compared to the more traditional DL approaches, the suggested DL architecture performs better.

The present study investigates by Sarisa et al. (2020) proposed a methodology for analyzing a dataset of online purchases in order to make predictions about future sales for the purpose of strategic management. RF, DT, and XGBoost were the three models that were created using 70% of the data for training and 30% for testing. Models were assessed employing performance metrics like R^2 and RMSE. Therefore, with a R^2 score of 96.3%, the XGBoost model surpassed all others in terms of marketing's predictive power for online store sales. Results like these show that the XGBoost algorithm can outperform competing models when it comes to predicting monthly sales for online retailers, which might help with inventory management and lead to faster, more informed choices in the explosively expanding online retail industry. The results highlight the significance of using sophisticated analytics to improve E-commerce industry performance and consumer experience.

Ahuja et al. (2019) examined the impact of two attributes on the SS-Tweet dataset for sentiment analysis: TF-IDF word level and N-Gram. Utilizing performance criteria such as F-score, Accuracy, Precision, and Re-call, they examined sentiment using six distinct classification algorithms (DT, SVM, KNN, RF, LR, and NB) and found that TF-IDF word level outperformed N-gram features by a margin of 3-4%.

In this study, X. Li et al. (2019) investigated the effect of internet reviews on companies. It takes a look at the correlation between numerical and textual evaluations and how well products sell. To extract the subjects and related feelings from review texts, they employ a Joint Sentiment-Topic model. Additionally, they argue that numerical rating acts as a mediator between textual feelings and their consequences. The results not only advance our understanding of how eWOM affects product sales, but they also show how text reviews and numerical ratings interact to influence product sales. By concentrating on more pertinent elements that eventually increase sales, the results assist online merchants in strategically planning business analytics activities.

G. Xu et al. (2019) suggested incorporating sentiment data into the construction of weighted word vectors as a means to enhance the conventional TF-IDF technique. BiLSTM receives the weighted word vectors and better represents the comment vectors while efficiently capturing the context information. A classifier based on feedforward neural networks determines the comment's sentiment trend. In the same context, RNN, CNN, LSTM, and NB models are compared to the proposed sentiment analysis method. In comparison to the state-of-the-art, the proposed sentiment analysis method achieves better results experimentally in terms of F1score, recall, and precision. The technique is working since the comments are quite accurate.

Poria et al. (2016) focus on exhibited the first method for extracting aspects in opinion mining using deep learning. As a subtask of SA, aspect extraction seeks to discover which product or service attributes the opinion holder finds favourable or unfavourable by analyzing opinionated language. For each word in opinionated statements, they used a 7-layer deep CNN to classify it as either trait or non-aspect. They integrated the neural network with a collection of language patterns they generated for the same objective. The strategy achieved far higher accuracy than advance approaches by combining a word-embedding model for SA with the resultant ensemble classifier.

In this work, Qi et al. (2016) suggested a model for automated filtering that takes into account the viewpoint of the product creator after forecast an usefulness of online reviews. Then, in a novel way, online reviews are analyzed using the KANO approach, which is based on the conventional conjoint analysis model, in order to formulate suitable strategies for product development. In addition, they used data gathered from JD.com, one of the biggest Chinese online marketplaces, to perform an empirical case study using the novel technique. It is clear from the case study that the suggested method is both effective and sturdy. Big data commerce may find success with a mix of big data and traditional management techniques, according to their study.

In this study, Ouyang et al. (2015) suggested a system that combines Word2vec with CNN. The first step is to generate word vector representations using the Google-proposed word2vec. These vector representations will then be fed into the CNN. To get the word vector representation and reflect the word distance, word2vec is used. In the end, it led to CNN parameter initialization, which effectively improved the nets' performance on this task. Second, they specifically tailored the CNN architecture to the sentiment analysis job. This design makes use of three sets of convolutional and pooling layers. They believe this to be the first instance of using a 7-layers architectural model to sentence sentiment analysis utilizing word2vec and CNN. To make their model even more accurate and applicable to a wider range of situations, they use technologies like the PReLU, normalization, and dropout. They put their model to the test using a publicly available dataset, which is a collection of snippets from movie reviews labelled as either favourable, neutral, slightly positive, or somewhat negative. Compared to other neural network models, such as the Recursive Neural Network (RNN) and the Matrix-Vector Recursive

Neural Network (MV-RNN), their network's 45.4% test accuracy on this dataset is superior.

2.3 Research Gap

A number of research gaps remain in spite of the significant progress made in sentiment analysis in e-commerce via the use of ML and DL methods. Most existing studies focus predominantly on accuracy improvements through varied model architectures, such as CNNs, RNNs, and BERT-based models, yet there is a noticeable lack of attention to domain adaptation, interpretability, and multilingual sentiment classification in real-world e-commerce contexts. While hybrid and ensemble models have been introduced, their applicability across diverse product domains and cultural nuances remains underexplored. Furthermore, challenges like handling imbalanced datasets, detecting sarcasm, and recognizing implicit aspects in reviews are still inadequately addressed. Additionally, few studies deeply investigate the integration of SA into Recommender Systems in a dynamic and context-aware manner. Thus, there remains a critical need for more robust, explainable, and adaptable SA frameworks that could operate effectively across various languages, domains, and review structures. Moreover, the real-time deployment and scalability of these models in high-traffic e-commerce environments remain a challenge, with limited research on optimizing computational efficiency without sacrificing performance. There is also insufficient exploration into user behavior dynamics over time and how evolving sentiment trends impact product success. Lastly, the ethical considerations surrounding data privacy and biased sentiment predictions have yet to be adequately addressed in existing literature, necessitating more responsible AI practices in future research.

CHAPTER III:

RESEARCH METHODOLOGY

3.1 Proposed Methodology

The AI-powered approach to SA in the future of business uses cutting-edge transformer-based models, namely Distil BERT and Albert, to categorize customer sentiment from product evaluations on Flipkart. The process begins with the installation of Python libraries like Pandas, NumPy, Seaborn, Matplotlib, Label Encoder, stop words, word tokenize, Word Cloud, Counter, Plotly, and Porter Stemmer. All necessary NLTK resources are downloaded to support natural language processing tasks. The dataset is then loaded, containing over 200,000 records and six attributes, including product name, price, rating, review, summary, and sentiment label. Initial data exploration includes checking for and removing null values, followed by normalization of textual features (review and summary) by removing non-alphanumeric characters and expanding contractions. Lemmatization using Word Net Lemmatizer is used to break words down into their component parts, and stop words are eliminated to make the data clearer.

Word clouds and bar graphs are used in EDA to display the distribution of attitudes and the most common words. Label encoding is applied to convert categorical sentiment values (Positive, Negative, Neutral) into the numerical format. The Distil BERT and Albert models are then tokenized using their respective tokenizers, with padding applied to ensure uniform input length. The model is fed a single input that combines the review and summary attributes. To handle class imbalance, a hybrid resampling technique is used, combining under sampling of majority classes and oversampling of minority classes. The balanced dataset is then split into training and testing sets in an 80:20 ratio.

Each model is fine-tuned using the Hugging Face Transformers library, utilizing GPU acceleration where available. Training parameters include a batch size of 16, multiple epochs (3 for Distil BERT, 5 for Albert), and regular logging steps. Once trained, the models are saved and later loaded for inference and evaluation. Several evaluation metrics, including as Accuracy, Precision, Recall, and F1Score, are used to verify the model's dependability and resilience. Additionally, a Confusion Matrix is generated to visualize the classification outcomes, and the ROC curve is plotted to evaluate model discriminative power. Lastly, these measurements are used to compare the two models and find out which one is better for sentiment-aware commercial applications in the real world. The following is a flow diagram of the model's implementation, as seen in Figure 3.1:

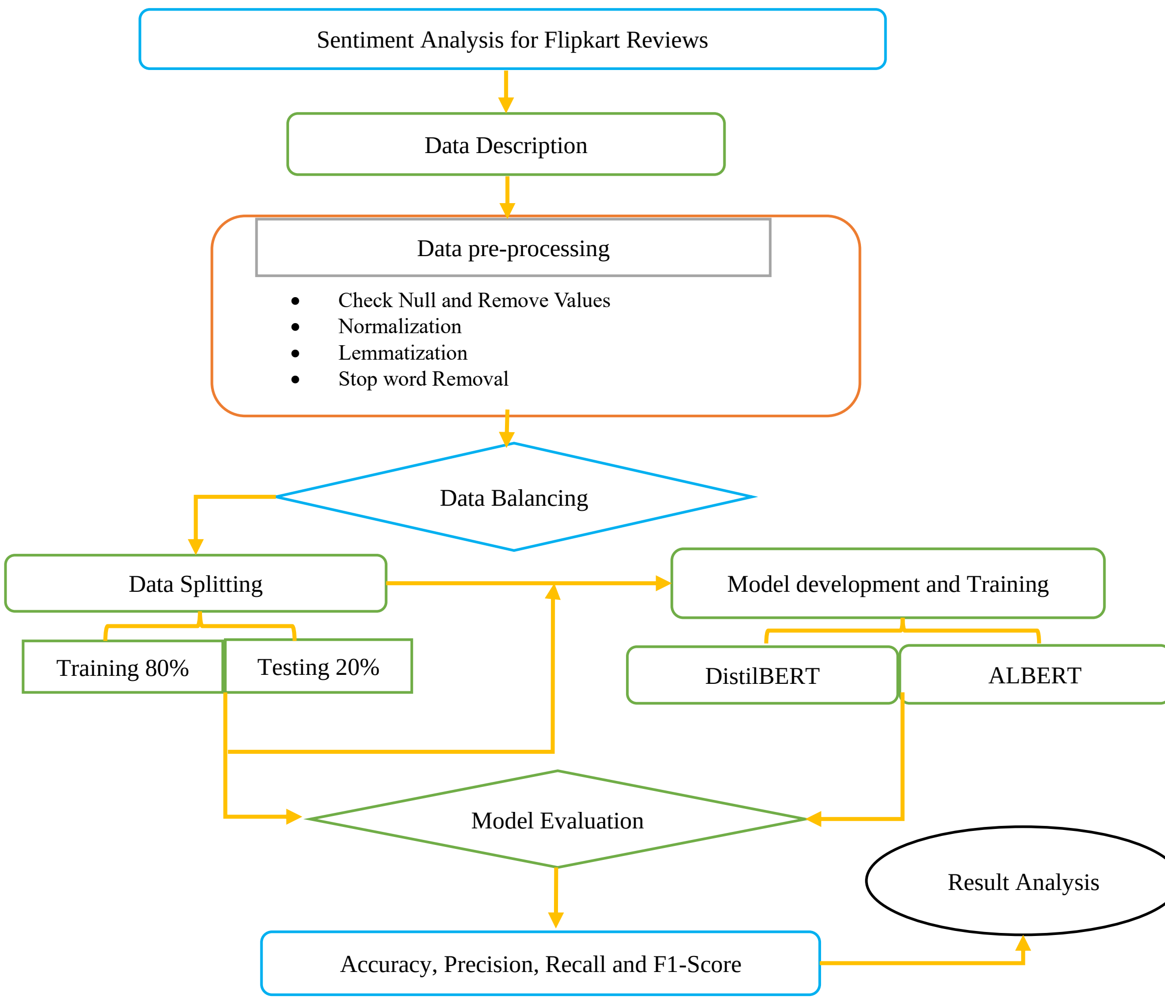


Figure 3.1: Flowchart of Proposed methodology for Sentiment Analysis

Each step of proposed methodology and flowchart discussed below in more detailed.

Dataset Description

The **sentiment analysis for flipchart reviews dataset**, sourced from Kaggle, contains over 205,000 records and six attributes including product name, price, rating, review, summary, and sentiment label (Positive, Negative, Neutral). The EDA includes visualizations like word clouds and frequency plots, along with an analysis of sentiment distribution. Label encoding was applied to transform sentiments into numeric form (negative: 0, neutral: 1, positive: 2). To address class imbalance, a hybrid resampling method combining oversampling and under sampling was used. The cleaned features, including reviews and summaries, were tokenized and padded for model training.

Data information

A dataset utilized for SA comprises 205,053 customer reviews from Flipkart and includes six attributes: `product name`, `product price`, `rate`, `review`, `summary`, and `sentiment`. The dataset, approximately 3.5-4 MB in size, contains textual and numerical data reflecting customer feedback on various products. The `sentiment` column serves as the target variable and is labelled as Positive, Negative, or Neutral, derived based on the review summary. NLP models may be trained and tested on this dataset specifically for sentiment categorization tasks.

Text Pre-processing

Data preparation is a crucial step in analyzing text data. The complexity of text data may increase due to repetitions and redundancies in textual material such as tweets, blogs, reviews, and more. In order to normalize data, a filtering operation known as data pre-processing is used (Faria et al., 2025). Data pre-processing includes tasks such as normalization, lemmatization, deleting stop words, checking null values, removal of null values, and normalization. This task's execution of many tasks resulted in the data being in the required format.

Check and removed Null Values

Null values were detected in key textual attributes, specifically the Review and Summary fields. These null values indicate the complete absence of data and should not be confused with zero or blank entries. Such occurrences are common and may result from various factors(Bello et al., 2023), including technical issues, incomplete data collection, user privacy preferences, and human error. To ensure data integrity and enhance the efficiency of the pre-processing phase, all records containing null values in these critical fields were removed from the dataset. In this review dataset first step in preprocessing involved checking the dataset for missing or null values. The elimination of records containing missing entries from essential columns Review, Summary and Sentiment took place due to their detrimental impact on both model performance and data consistency. The training process used complete and trustworthy data by ensuring all required information existed in each record.

Normalization

Text normalization is the process of making all text uniformly lowercase and then eliminating extra spaces, special characters, and numerals to make the information easier to read and understand. Following that comes content reduction, which comprises deleting irrelevant features like hashtags (Suhaeni & Yong, 2024), URLs, and user handles, as well as excessively short words (those less than one character), and so on. Normalization was applied to the textual features, specifically the Review and Summary columns. This process involved removing non-alphanumeric characters, unnecessary punctuation, and special symbols while retaining meaningful textual content. For consistency's sake, we also enlarged contractions to their full forms (such "don't" to "do not"). This step prepared the data for further linguistic processing.

Lemmatization

The term "lemmatization" describes the process of simplifying a word. The stemming technique boils down to tracing the origin of an infectious phrase to its most basic form (Obiedat et al., 2022) . Lemmatization is a NLP pre-processing technique that finds similarities by reducing words to their root words (Almuayqil et al., 2022). The process begins with recognizing lemmas(Ranjan et al., 2023), the simplest forms of words. In this work lemmatization was carried out using NLTK’s Word Net Lemmatizer, which reduced words to their root form while preserving their actual meaning. Unlike stemming, which can distort words, lemmatization converts words like “running” to “run” based on their grammatical context. This helped in reducing the dimensionality of the dataset without losing semantic information.

Stop Word Removal

The method entails skimming the text for terms that do not contribute to the understanding of the emotional content. Stop words are language-specific, functional terms that change depending on the text's language and don't provide any information. Words that are set to not be included in the index or removed are called stop words (Özmen & Gündüz, 2025). It has minimal analytical utility and provides a lot of duplicate information; hence it lacks considerable centrality. In this work used stop word removal on lemmatized_normalized_review and lemmatized_normalized_summary for removing non-words characters. Common English stop words like “is,” “the,” “and,” and “was” were removed from the lemmatized text using NLTK’s predefined stop word list. These words, although frequent, add little value to sentiment analysis. Their removal allowed the model to better concentrate on words that conveyed meaning and emotion, leading to improved classification accuracy.

Tokenization and Padding

The cleaned and preprocessed textual data was then tokenized using transformer-specific tokenizers from Hugging Face—Distil BERT Tokenizer and Albert Tokenizer. Tokenization converted the text into integer sequences understood by the models. Tokenization is a method for reducing large text streams to smaller ones, often phrases. "Tokens" are fragments of writing. Simplifying complex textual themes is the aim of this strategy. Lexical evaluation is highly dependent on tokenization, and it also helps with sentiment analysis and semantics. Tokenization is a crucial step in the pipeline for NLP. In order to facilitate batch training in transformer designs, padding was used to guarantee that each sequence was of uniform length.

Label Encoding

ML algorithms may make use of numerical values extracted from category data via a process known as label encoding. The initial sentiment category which contained the textual values positive, neutral, and negative underwent conversion to numeric categories using Label Encoding. The conversion enabled ML and DL models to train efficiently through numerical assignment where 0 indicated negative and 1 indicated neutral and 2 indicated positive.

Data Balancing

A hybrid data balancing strategy was used to address the problem of unequal class representation in the sentiment labels. This method involved a sequential combination of both under sampling the majority class and oversampling the minority classes. The research utilized under sampling on overrepresented sentiment classes but employed oversampling on underrepresented categories to create an equal distribution across sentiment categories.

The same balancing strategy worked identically on all datasets created for both Distil BERT and Albert models to provide equal participation for positive, negative, and neutral classes during training process. The implemented hybrid technique succeeded in reducing bias by achieving balanced class distributions which were both verified post-processing.

Data Splitting

An 80:20 proportion served as the basis to split the data where testing received 20% of the data and training obtained the remaining 80%. This same split process applied to both Distil BERT and Albert models helped maintain consistency during evaluation. A class-distributed split from the dataset preserved the outcome achieved through hybrid sampling to deliver steady performance evaluation results.

Model Development and Training

For model development and training, both Distil BERT and Albert were fine-tuned employing Hugging Face with GPU acceleration. A model’s performance is given below:

Distil BERT

Distil BERT is an optimized, lightweight variant of the BERT model that offers improved speed and efficiency without sacrificing performance. One method for training it is known as knowledge distillation, and it involves taking a bigger "teacher" model (BERT) and teaching a smaller "student" model (Distil BERT) (Sanh et al., 2019) to mimic its behaviour. Unlike BERT, which has 12 transformer layers, Distil BERT has only 6, and it omits the next sentence prediction (NSP) objective, reducing its size and computational requirements. For text classification tasks, the process begins by tokenizing the input text using the BERT tokenizer(Adel et al., 2022), which adds special tokens like (CLS) at the beginning and (SEP) at the end of the input sequence. These tokens are then converted into embeddings that combine token, position, and segment information.

The transformer layers take the embedded tokens and use feed-forward neural networks and multi-head self-attention to collect contextual links. An equation including equation (3.1) is used by the model to determine attention scores inside the self-attention mechanism:

$$Attention(Q, K, V) = softmax\left(\frac{Qk^T}{\sqrt{dk_K}}\right) V \dots (3.1)$$

where the input embeddings provide the query matrix (Q), key matrix (K), and value matrix (V), furthermore, layer normalization and residual connections are also included in each transformer block. It is possible to get the hidden state associated with the (CLS) token as a representation of the whole input sequence with a fixed size after the last transformer layer. Some kind of classification head, usually a fully connected layer, is used to process this CLS vector. For binary classification, a sigmoid activation is used to predict probabilities(Y. Yang et al., 2021), whereas for multi-class classification, a SoftMax activation is applied. A model is trained using a loss function like cross-entropy loss. For multi-class tasks, the loss is calculated using equation (3.2):

$$\mathcal{L} = -y \log(\hat{y}) - (1 - y) \log(1 - y) \dots . (3.2)$$

where $-y$ is the true label and $(\hat{y})$ is the forecasted probability, through training, the model learns to minimize this loss by adjusting its internal parameters. Overall, DistilBERT provides an efficient and effective architecture for performing text classification tasks in resource-constrained environments without sacrificing much accuracy.

Albert

Albert represents a stripped-down version of BERT, which reduces processing needs while maintaining performance quality. It reaches high performance while maintaining speed through interlayer parameter sharing and factorized embeddings methods. The primary objective of Albert is to make transformer models more efficient and easier to deploy in real-world applications(X. Wang et al., 2021). As mentioned earlier, the core equation (same as 3.1) used in Albert is usually associated with the transformer architecture self-attention mechanism. For compilation, Albert optimizes the training procedure through consisting of several techniques. One of the strategies is “Cross-layer parameter sharing” by which the weights in each of the transformer encoder layers are set to be equal to the weights in other layers of the transformer encoder. This decrease the number of parameters significantly(Quintero Vanegas, 2023), allowing the model to scale effectively without a linear increase in memory and computation costs. The other significant improvement in Albert is the use of ‘factorized embeddings’, which helps in reducing the size of the embedding layer and yet retains representational power. Another important equation used in the transformer model that Albert builds upon is the position-wise feedforward network equation, which is used after the self-attention step in equation (3.3):

$$FFN(x) = \max(0, xW_1 + b_1)\, W_2 \; + b_2 \, \ldots .\, (3.3)$$

Here, x is an input to a feed-forward Network, which is the output from the previous layer or the attention mechanism. The weight matrices W_1 and W_2 are learned during training and are responsible for transforming the input into a higher-dimensional space and then reducing it back to the original dimension. The bias terms b_1 and b_2 help shift the activation function and the output of the layer to enhance the model’s ability to fit the data. An activation function that is used is the ReLU (Nicolae et al., 2022), which is represented as max(0, ·) and creates non-linearity by making all negative values equal to zero while keeping positive values intact. This enhances the model's versatility by enabling it to understand complex patterns within the data.

Model Training

The Distil BERT and Albert models were implemented using the Transformers library from Hugging Face and fine-tuned during this process when facilitated by GPU support for improved performance. The following training parameters are:

- The fine-tuning of both Distil BERT and Albert models was done with the help of the Hugging Face Transformers library, which is a high-level and versatile library that offers easy access to powerful NLP models. This training pipeline assumed utilization of GPU resources where available, considerably enhancing the model training and evaluation steps. These benefits of using GPU enhance training both in terms of time and enable training with higher batch sizes and complex transformer model architectures. The compatibility of Hugging Face with frameworks like PyTorch and TensorFlow allows users to offload the computation to a GPU for improved performance and concessions in real-time business approaches to sentiment analysis.
- For training and evaluating both the Distil BERT and Albert models, specific hyperparameters were set to ensure optimal performance. Pricing the training and evaluation batch size at 16 enabled efficient memory utilization and steady gradient updates during the backpropagation process for both models. GPU acceleration performance benefits from this particular batch size because it keeps an optimal equilibrium between training speed and convergence quality.
- The setting of the number of epochs varies between models based on their complexity. This configuration dictates how many times the whole training dataset is passed through the model. Distil BERT was trained for 3 epochs, as it is a lighter and faster model optimized for speed and efficiency, often requiring fewer epochs

to converge. The training duration for Albert was extended to 5 epochs because its complex architectural features need an extended time to reach optimized results.

- The logging steps parameter was set to 10 for both models. The configuration enables automatic logging of training progress together with evaluation metrics throughout each training period. Performance tracking at regular intervals of 10 steps through the training process enables improved ability to spot trends and detect overfitting, and allows better determination about early stopping and hyperparameter optimization.

3.2 Model Evaluation

The evaluation process required an analysis of the performance capabilities between Distil BERT and Albert models based on extensive classification metrics. The model evaluation used Accuracy, Precision, Recall and F1Score to determine the total sentiment classification effectiveness across training and adjustment of the filtered and balanced dataset. Each sentiment class appeared in the Confusion Matrix to show TP, FP, TN and FN results which helped understand misclassifications. Also, the ROC curve was used to compare the models' discriminative ability across the threshold values of the diagnostic tests for sensitivity and specificity, and various other measures of accuracy. The evaluation was comprehensive and this style made the advantages and the problems of each model when operating on actual sentiment data in reference to the other emerge.

3.3 Conclusion

In conclusion, this study successfully demonstrates the application of AI, particularly transformer-based models Distil BERT and Albert, in sentiment analysis for future business scenarios using Flipkart customer reviews. To further clean the data normalized the text data, lemmatized the data, removed stop words and the data was resampled using both oversampling and under sampling to get the best outcome.

CHAPTER IV:

RESULTS ANALYSIS

4.1 Simulation Experiment Setup

For the development and evaluation of the proposed Distil BERT and ALBERT transformer-based models for sentiment classification on the Flipkart product reviews dataset, a robust computational environment was employed. The following hardware equipment's are used in this research work:

- The experiments were executed on an HP workstation designed to handle large-scale text processing and deep learning workloads. This workstation provided the foundational hardware infrastructure necessary for efficient experimentation with transformer-based models like Distil BERT and Albert.
- The system was equipped with 32GB of RAM, allowing it to handle extensive data preprocessing, large batch sizes, and memory-intensive model training without lag or memory overflow. This high memory capacity ensured smooth operation during the tokenization, vectorization, and model fitting stages.
- For data storage, the workstation included a 1TB hard drive, offering sufficient space to store the dataset, trained model files, and intermediate results such as embeddings and checkpoints. The large disk space played a crucial role in managing the complete training pipeline efficiently.
- It ran on Windows 10 as the operating system, providing a stable and user-friendly environment for integrating various tools and libraries needed for machine learning tasks.
- At its core, the machine was powered by an Intel i7 processor, which provided the multi-core computing power required for parallel processing during data loading, model evaluation, and auxiliary tasks.

- Most importantly, the workstation featured a 24GB NVIDIA GPU, which significantly accelerated the training and fine-tuning of transformer models. The high GPU memory allowed for faster computation, larger input sequences, and efficient backpropagation, reducing training time while maintaining high accuracy.

Normalizing, lemmatizing, and removing stop words from text, as well as training and evaluating models, were all made possible by this system, which are fundamental in NLP. Python was used in the implementation inside a Jupyter Notebook environment together with important libraries like NLTK, Scikit-learn, TensorFlow, and Keras. Both Distil BERT and ALBERT models were fine-tuned on the dataset, benefiting from the hardware acceleration and, when needed, additional GPU resources provided by Google Colab. This integrated hardware-software ecosystem ensured fast training times, effective fine-tuning, and high accuracy in classifying product sentiments.

- Python is a popular language for scientific and deep learning tasks. A wide variety of DL methods may be implemented using its huge library collection (Iqbal et al., 2022). Python packages (libraries) simplify a number of important processes, such as textual information, machine learning model building, image processing, online data retrieval, and data analysis and visualization. We found the following libraries to be useful("Top Popular Python Libraries in Research," 2022):
- An open-source, sophisticated, flexible, and user-friendly tool for data analysis and manipulation, Pandas is built on top of Python. Pandas is a popular option for data wrangling and analysis because it provides convenient techniques for cleaning, manipulating, and transforming data.
- NumPy is an open-source, free, and scalable software that comes with multi-dimensional arrays. Using the NumPy ND array, data may be stored in an object with homogenous "n" dimensions.

- Data exploration and visualization are made easy using matplotlib. You may use it to create basic graphical representations of data, such as line graphs, scatter plots, bar graphs, pie charts, and histograms. As an extension to Python, NumPy allows users to work with numerical data.
- Seaborn relies on Matplotlib for its foundation. Makes use of visual and instructive statistics visualizations. Data visualization expert Seaborn has extensive experience with Pandas Data Frame and is particularly well-versed in the use of category variables. To aid in the discovery of previously unseen patterns in data, Seaborn provides a colour palette choosing option.
- The open-source Python Plotly package is based on the Plotly JavaScript library (plotly.js). As a robust and user-friendly library that can communicate with visualizations, Plotly is undeniably a necessary tool for development of visualizations.
- Keras is a DL API built on top of the ML framework TensorFlow. A key factor in its creation was the possibility of conducting trials quickly. Building NN and other types of DL models is where Keras really shines. Keras has a wide range of possible uses, including CV, NLP, and providing effective and reliable ML solutions for practical use.
- Python's Natural Language Toolkit (NLTK) is an essential resource for developing NLP modules. Essential text preparation tasks, including stemming, lemmatisation, and tokenisation, are supported, which are needed before NLP models are trained. NLTK offers a wide range of built-in methods and functions that can be directly applied to datasets to efficiently process and understand the text language.

4.2 Dataset Visualization Results

The Flipkart product reviews dataset used in this study comprises 205,053 instances and 6 key attributes: product name, product price, rate, review, summary, and sentiment. It is approximately 3.5–4 MB in size and serves as a rich source for SA. The sentiment label—Positive, Negative, or Neutral—is derived from the summary of each review. During exploratory data analysis (EDA), null values were identified and removed to ensure data quality. Non-numeric features, such as review and summary, underwent normalization by removing unwanted characters and expanding contractions. After that, we lemmatized the words using Word Net Lemmatizer to get them to their root form, and last, we stopped word removal to get rid of any unnecessary terms. After the cleaning process, word clouds and other visualizations were created for the review and summary sections to show how the text was distributed and which terms were used most often. Next, we used label encoding to transform the categorical labels into a numerical format that is compatible with the models, and we visualized the dataset's sentiment label distribution. We then initialized the sentiment feature as the target variable. These preprocessing and visualization steps provided a strong foundation for building an accurate sentiment classification model.

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 205052 entries, 0 to 205051
Data columns (total 6 columns):
 #   Column         Non-Null Count   Dtype
---  ------         --------------   -----
 0   product_name   205052 non-null  object
 1   product_price  205052 non-null  object
 2   Rate           205052 non-null  object
 3   Review         180388 non-null  object
 4   Summary        205041 non-null  object
 5   Sentiment      205052 non-null  object
dtypes: object(6)
memory usage: 9.4+ MB
```

Figure 4.1: Data information of Flipkart product reviews

Figure 4.1 presents a summary of the Flipkart product reviews dataset, revealing that it consists of 205,052 entries indexed from 0 to 205,051 and comprises a total of 6 columns. Each column's name, the count of non-null values, and its data type are listed. Notably, the 'product_name', 'product_price', 'Rate', and 'Sentiment' columns contain no missing values, each having 205,052 non-null entries and are stored as object data types. The 'Summary' column has a minor number of missing values with 205,041 non-null entries, also of object type. The 'Review' column shows a more significant number of missing entries, with 180,388 non-null values, and is also an object type. Overall, the dataset primarily contains categorical or textual data, as indicated by the 'object' data type for all columns, and occupies approximately 9.4+ MB of memory. The presence of missing values in the 'Review' and 'Summary' columns suggests a need for appropriate data cleaning or imputation strategies before further analysis.

```
            Review normalized_review  \
0           super!             super
1          awesome           awesome
2             fair              fair
3  useless product   useless product
4             fair              fair

                                             Summary  \
0  great cooler excellent air flow and for this p...
1             best budget 2 fit cooler nice cooling
2  the quality is good but the power of air is de...
3                 very bad product its a only a fan
4                                     ok ok product

                                  normalized_summary
0  great cooler excellent air flow and for this p...
1             best budget 2 fit cooler nice cooling
2  the quality is good but the power of air is de...
3                 very bad product its a only a fan
4                                     ok ok product
```

Figure 4.2: Check feature values after normalization

Figure 4.2 displays a sample of the 'Review', 'normalized review', 'Summary', and 'normalized summary' columns after the normalization preprocessing step. By comparing the original 'Review' column with the 'normalized review' column, it's evident that the normalization process, in this specific sample, has primarily converted the text to lowercase. For instance, "super!" has become "super", and "awesome" remains "awesome" as it was already in lowercase. Similarly, the 'Summary' column and the 'normalized summary' column show that the normalization has also applied lowercase conversion to the text within the summaries. This step is important for ensuring consistency in the text data, which helps in subsequent text processing steps like tokenization and feature extraction by treating words with different casing as the same.

```
           Review lemmatized_normalized_review  \
0          super!                        super
1         awesome                      awesome
2            fair                         fair
3  useless product             useless product
4            fair                         fair

                                             Summary  \
0  great cooler excellent air flow and for this p...
1             best budget 2 fit cooler nice cooling
2  the quality is good but the power of air is de...
3                 very bad product its a only a fan
4                                      ok ok product

                         lemmatized_normalized_summary
0  great cooler excellent air flow and for this p...
1                 best budget 2 fit cooler nice cool
2  the quality be good but the power of air be de...
3                  very bad product it a only a fan
4                                      ok ok product
```

Figure 4.3: Check feature values after lemmatization

Figure 4.3 presents a comparison of the 'Review' and 'Summary' columns with their lemmatized and normalized counterparts, 'lemmatized_normalized_review' and 'lemmatized_normalized_summary'. Observing the 'lemmatized_normalized_review', it appears that for this particular sample, the lemmatization process hasn't resulted in significant changes compared to the normalized review shown in the previous figure. Words like "super", "awesome", "fair", and "useless product" remain the same. However, when examining the 'lemmatized_normalized_summary', some changes due to lemmatization are noticeable. For instance, "nice cooling" has become "nice cool", "is" has been transformed to "be", "its" to "it", and "a fan" remains "a fan". The goal of lemmatization is to reduce words to their dictionary or base form, which may assist in standardizing the text and improve performance on downstream tasks by considering diverse word inflections as the same. The example shows that while some words are already in their base form, others undergo transformation to their lemma.

```
             Review stop_words_removed_lemmatized_normalized_review  \
0            super!                                           super
1           awesome                                         awesome
2              fair                                            fair
3   useless product                                 useless product
4              fair                                            fair

                                             Summary  \
0  great cooler excellent air flow and for this p...
1                best budget 2 fit cooler nice cooling
2  the quality is good but the power of air is de...
3                    very bad product its a only a fan
4                                        ok ok product

    stop_words_removed_lemmatized_normalized_summary
0  great cooler excellent air flow price amaze un...
1                   best budget 2 fit cooler nice cool
2                        quality good power air decent
3                                      bad product fan
4                                        ok ok product
```

Figure 4.4: Check feature values Stop Word Removal

Figure 4.4 displays an effect of stop word removal on the 'Review' and 'Summary' columns, comparing them to the 'stop_words_removed_lemmatized_normalized_review' and 'stop_words_removed_lemmatized_normalized_summary' columns. Examining the 'Review' and its processed version, for this specific sample, no stop words appear to have been present in the initial reviews, as both columns remain identical. However, when looking at the 'Summary' and its processed counterpart, significant changes are evident. Stop words such as "and", "for", "this", "the", "but", "of", "is", "a", and "only" have been removed. For example, "great cooler excellent air flow and for this p..." has become "great cooler excellent air flow price amaze un...". Similarly, "the quality is good but the power of air is de..." is now "quality good power air decent". The purpose of deleting less informative words is to make SA models work better by lowering the amount of noise in the text data and drawing attention to the key phrases that make up the sentiment.

Figure 4.5: Word cloud for review feature

A word cloud visualization of the most common terms occurring in the 'Review' feature of the preprocessed Flipkart product reviews dataset is shown in Figure 4.5. With bigger terms signifying greater occurrence, the size of every word in the cloud is proportionate to its frequency in the reviews. Prominent words such as "product", "best", "buy", "worth", "every", "perfect", "great", "good", "excellent", "must", "fair", "awesome", "terrific", "simply", and "penny" stand out, suggesting that these terms are commonly used by customers when describing their experiences with the products. The presence of both positive words like "perfect", "great", "good", "excellent", "awesome", "brilliant", "delightful", "fabulous", "wonderful", and negative words such as "terrible", "useless", "waste", "rubbish", "unsatisfactory", "worthless", and "disappointed" indicates a diverse range of sentiments expressed in the reviews. Additionally, terms like "market", "quality", "purchase", and "recommend" provide context about the nature of the reviews, often discussing product quality, the act of purchasing, and whether the product is recommended.

This word cloud provides a visual representation of the most important terms and feelings expressed in the evaluations.

Figure 4.6: Word cloud for summary feature

The most common terms used in the 'Summary' section of the Flipkart Product Reviews dataset are displayed in Figure 4.6, a word cloud visualization. A magnitude of every word in the summary represents how often it appears, just as in a review word cloud. Dominant terms include "good", "product", "nice", "quality", "awesome", "excellent", "best", "sound", "use", "money", and "well", indicating that these words are commonly used by customers to provide brief overviews of their opinions. The presence of positive adjectives like "good", "nice", "awesome", "excellent", "best", and "well" suggests that many summaries convey positive feedback. However, the appearance of "bad" also indicates the presence of negative summaries. Phrases like "sound quality", "product good", and "nice product" are also prominent, suggesting common themes in the summarized reviews. Other words such as "buy", "price", "value", "battery", "speaker", and mentions of "Flipkart" provide additional context about the aspects frequently highlighted in the product summaries. This word cloud offers a quick understanding of the key terms and overall sentiment expressed in the concise summaries provided by customers.

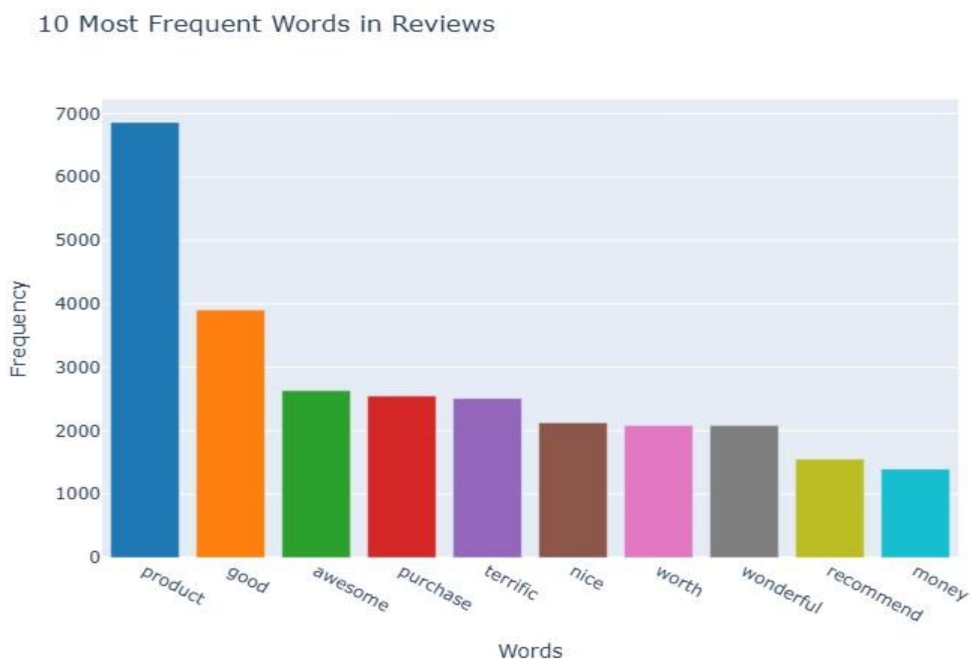


Figure 4.7: Bar graph of 10 most frequent words in reviews

In Figure 4.7, we can see a bar graph that shows the top ten terms from the Flipkart Product Reviews dataset's 'Review' feature. An x-axis displays a word, while a y-axis represents their frequency of occurrence. A word "product" appears most frequently, with a count nearing 7000. Following "product", "good" is the next most frequent word, appearing close to 4000 times. The words "awesome", "purchase", and "terrific" show similar frequencies, each occurring approximately 2500 times. The terms "nice", "worth", and "wonderful" also appear frequently, with counts around 2000. Lastly, "recommend" and "money" are among the top 10 but have lower frequencies compared to the preceding words, with "recommend" appearing roughly 1500 times and "money" slightly above 1000 times. This visualization highlights the key terms used by customers in their reviews, with "product" being the most common subject of discussion, often accompanied by positive adjectives like "good", "awesome", "terrific", "nice", and "wonderful", and considerations of "purchase", "worth", "recommendation", and "money".

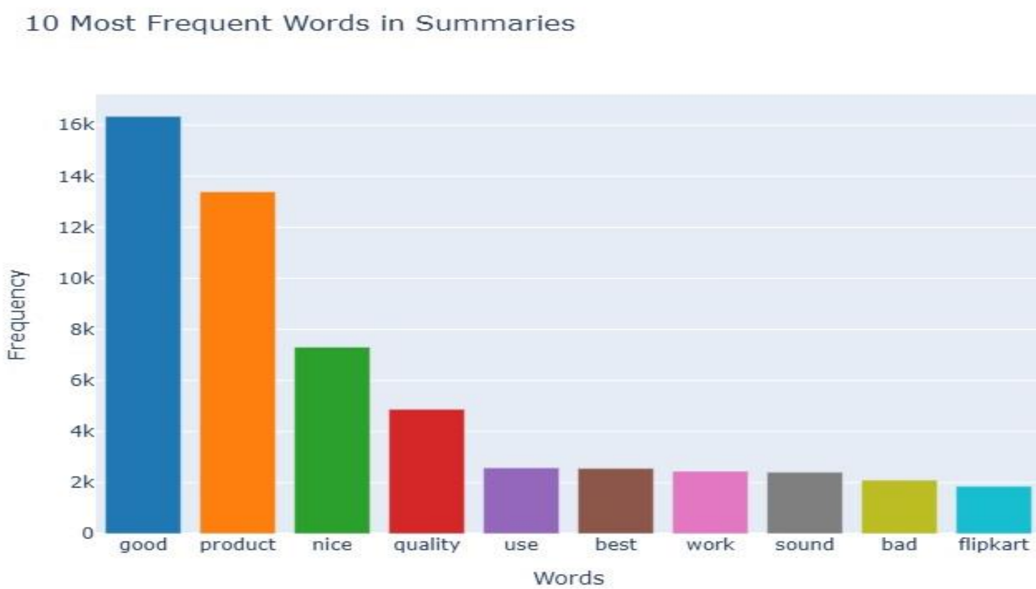


Figure 4.8:Bar graph of the 10 most frequent words in summaries

The top ten terms used in the 'Summary' section of the Flipkart Product Reviews dataset are shown in Figure 4.8, a bar graph. An x-axis displays words, and a y-axis shows their frequency of occurrence, scaled in thousands (k). The word "good" is by far the most frequent, appearing over 16,000 times. Following "good", "product" is the next most frequent, with a count exceeding 13,000. The words "nice" and "quality" also appear frequently, with "nice" occurring over 7,000 times and "quality" close to 5,000 times. The terms "use", "best", and "work" have similar frequencies, each appearing around 2,500 times. The word "sound" occurs slightly less frequently, just above 2,000 times. Finally, "bad" and "Flipkart" are among the top 10 but have lower frequencies compared to the preceding words, with "bad" appearing approximately 2,000 times and "Flipkart" just under 2,000 times. This visualization highlights that positive descriptor like "good" and "nice", along with the subject "product" and the attribute "quality", are the most common terms used in the concise summaries provided by customers.

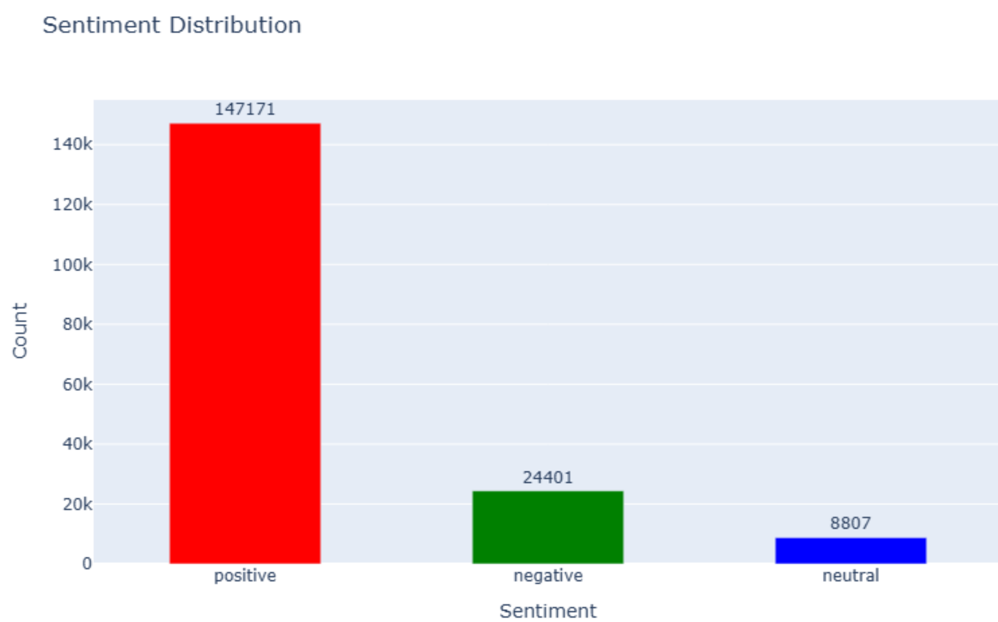


Figure 4.9: Bar graph of sentiments features class distributions

Figure 4.9 is a bar graph that shows how the Flipkart Product Reviews dataset is distributed across different sentiment groups. An x-axis represents the 3 sentiment categories: positive (red bar), negative (green bar), and neutral (blue bar), while a y-axis displays the count of reviews belonging to every category. The graph clearly shows a significant class imbalance. There are 147,171 reviews in the good sentiment category, which is the greatest number of reviews overall. The negative sentiment category has considerably fewer reviews, totaling 24,401. Of the three types of evaluations, 8,807 fall under the neutral sentiment group, which is the least popular. This distribution shows that the dataset is heavily skewed towards positive ratings, while neutral and negative evaluations are noticeably under-represented. Such class imbalance can potentially bias sentiment analysis models towards the majority class (positive) and may require specific techniques, such as resampling or weighted loss functions, to address during model training to ensure fair and accurate classification across all sentiment categories.

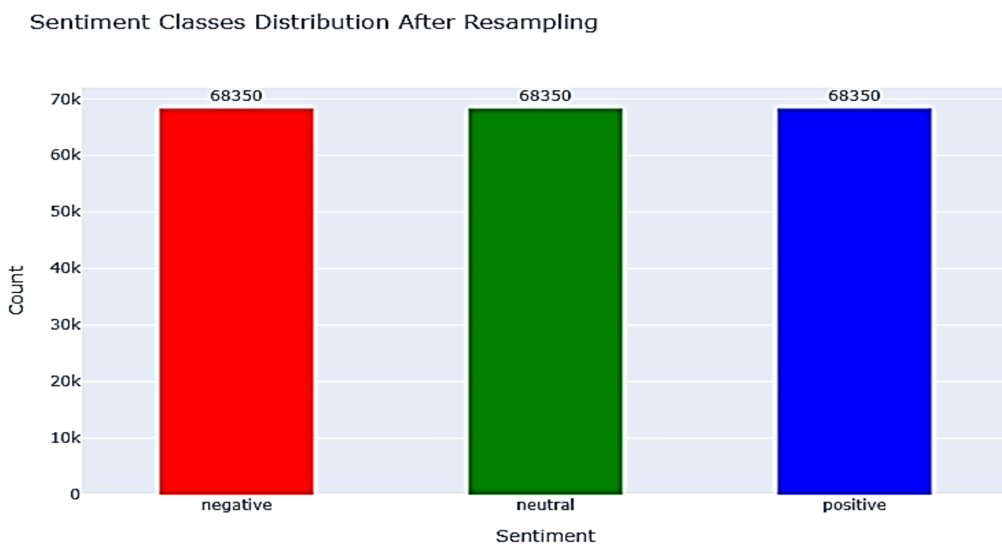


Figure 4.10: Bar graph of sentiments features class distributions after data balanced

Figure 4.10 displays a bar graph showing a distribution of sentiment classes in the Flipkart product reviews dataset after a data balancing or resampling technique has been applied. An x-axis displays a 3-sentiment category: negative (red bar), neutral (green bar), and positive (blue bar), while a y-axis indicates a count of reviews in every category. In contrast to the previous distribution, this graph shows a significantly more balanced dataset. Each sentiment category now contains an equal number of reviews, specifically 68,350. This indicates that the class imbalance observed earlier (where positive reviews heavily outnumbered negative and neutral ones) has been successfully addressed through the chosen resampling method. By equalizing the number of samples across all sentiment classes, the dataset is now better suited for training SA models, as it reduces the potential for bias towards the previously dominant positive class and should allow the models to learn more effectively from the minority classes.

4.3 Performance Evaluation parameters

Several metrics are used to assess classification algorithms, including recall, F-score, accuracy, and precision. These factors seem to be quite useful for assessing supervised ML algorithms. To evaluate the model's accuracy, a confusion matrix—sometimes called a contingency table—is used. This matrix gives a complete analysis of the model's TP, TN, FP, and FN. The following performance measures used for this work discussed as:

1) Accuracy

Accuracy is the gold standard for measuring performance. The computation and identification of it are simple and easy. No matter how good or bad a predictor is, accuracy measures how well it can identify all samples (eq.4.1):

$$\text{Accuracy} = \frac{TP + TN}{\text{n}} \dots \dots \dots (4.1)$$

2) Sensitivity/Recall

It is possible to determine sensitivity using the recall or TPR. It is possible to get the actual positive percentage by following certain easy steps. As sensitivity increases, the number of false negatives decreases, and vice versa for lesser sensitivity. There are cases when the accuracy decreases as the sensitivity rises (eq.4.2).

$$\text{Recall} = \frac{TP}{\text{TP} + \text{FN}} \dots \dots \dots (4.2)$$

3) Precision

The accuracy of the classifier is shown by the precision. There will be fewer false positives and lesser accuracy with high precision or low accuracy. Reducing sensitivity is the result of improving accuracy, which is inversely proportional to sensitivity (eq.4.3).

$$\text{Precision} = \frac{TP}{\text{TP} + \text{FP}} \dots\dots\dots (4.3)$$

4) F1 score

F1-Measure combines accuracy with sensitivity. For pinpoint accuracy, this is the weighted harmonic approach. As seen in equation 4.4, the F1 measurement is just as effective as precision.

$$\text{F1}\_Score = 2\frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \dots\dots\dots (4.4)$$

5) AUC-ROC

AUC-ROC: The area under the ROC curve is represented by the AUC-ROC score. The AUC is a numerical value between 0 and 1, similar to the F1score; higher values imply superior performance. The connection between 1-specificity (FPR) and sensitivity (TPR) is graphically represented by the ROC curve.

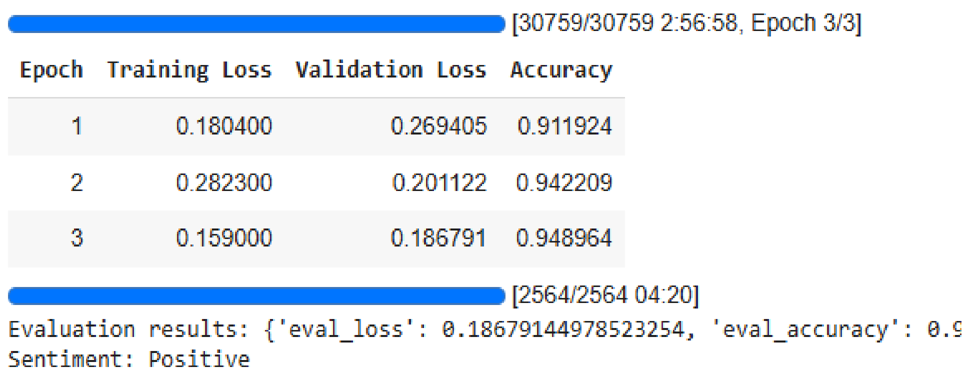
4.4 Experimental Results of proposed Models

Presented and discussed in this part are the outcomes of the experiments. A chosen Flipkart Product Reviews dataset is assessed using two DL methods, including Distil BERT and Albert transformer, focusing on key performance metrics like accuracy, precision, recall, F1score, and ROC curves. Each method for distinguishing positive, negative, and neutral sentiment classes was highlighted in the models' evaluations using confusion matrices, which analyzed accurate and wrong classifications.

Table 4.1: Performance of proposed both models for sentiment analysis

Parameters	**Distil BERT Model**	**Albert Model**
Accuracy	0.9490	0.9333
Precision	0.9491	0.9333
Recall	0.9490	0.9333
F1 Score	0.9490	0.9333

Table 4.1 provides a concise numerical comparison of a performance of a proposed Distil BERT and ALBERT models for SA of Flipkart product reviews. A table clearly indicates that the Distil BERT model outperforms the ALBERT model across all evaluated metrics. Specifically, Distil BERT achieves an accuracy of 0.9490, a precision of 0.9491, a recall of 0.9490, and an F1score of 0.9490. In contrast, the ALBERT model demonstrates slightly lower performance with an accuracy of 0.9333, a precision of 0.9333, a recall of 0.9333, and an F1score of 0.9333. These results suggest that for this particular task of Sentiment Analysis on Flipkart product reviews, the Distil BERT model is more effective in correctly classifying the sentiment of the reviews, exhibiting both higher accuracy and a better balance among precision and recall as reflected in the F1score.


[30759/30759 2:56:58, Epoch 3/3]

Epoch	Training Loss	Validation Loss	Accuracy
1	0.180400	0.269405	0.911924
2	0.282300	0.201122	0.942209
3	0.159000	0.186791	0.948964

[2564/2564 04:20]

```
Evaluation results: {'eval_loss': 0.18679144978523254, 'eval_accuracy': 0.9
Sentiment: Positive
```

Figure 4.11: Performance parameters of proposed Distil BERT model

Figure 4.11 displays the training and evaluation performance of a proposed Distil BERT model over three epochs. During training, the training loss fluctuates, starting at 0.180400 in the first epoch, increasing to 0.282300 in the second, and then decreasing to 0.159000 in the final epoch. Concurrently, the validation loss shows a decrease from 0.269405 in the first epoch to 0.201122 in the second, followed by a slight decrease to 0.186791 in the third epoch. The accuracy on the validation set consistently improves with each epoch, reaching 0.911924 in the first, 0.942209 in the second, and culminating at 0.948964 in the third epoch. The evaluation results after training indicate a final evaluation loss of approximately 0.1868 and an evaluation accuracy of 0.9490. It seems that the Distil BERT model has learnt well and is able to generalize to unknown data for sentiment classification, as shown by its high evaluation accuracy and decreasing validation loss trend throughout epochs. As a result, the model eventually predicted that the reviewed text had a positive sentiment.

```
Classification Report:
              precision    recall  f1-score   support

    Negative       0.96      0.95      0.96     13782
     Neutral       0.93      0.96      0.94     13567
    Positive       0.95      0.94      0.95     13661

    accuracy                           0.95     41010
   macro avg       0.95      0.95      0.95     41010
weighted avg       0.95      0.95      0.95     41010
```

Figure 4.12: Classification Report of proposed Distil BERT model

In Figure 4.12, we can see the suggested Distil BERT model's classification report, which describes how well it performed for the three sentiment classes (negative, neutral, and positive). For the Negative Class, the model achieves a precision of 0.96, a recall of 0.95, and an F1score of 0.96, based on a support of 13782 instances. The Neutral class shows a recall of 0.96, a precision of 0.93, and an F1score of 0.94, with a support of 13567 instances. For the Positive class, the model obtains a precision of 0.95, a recall of 0.94, and an F1score of 0.95, supported by 13661 instances. A model's overall accuracy across all classes is 0.95, which is rather good. The Distil BERT model performs well in identifying sentiment across all 3 categories in the dataset of 41010 samples, as seen by the constant 0.95 macro average and weighted average for precision, recall, and F1score.

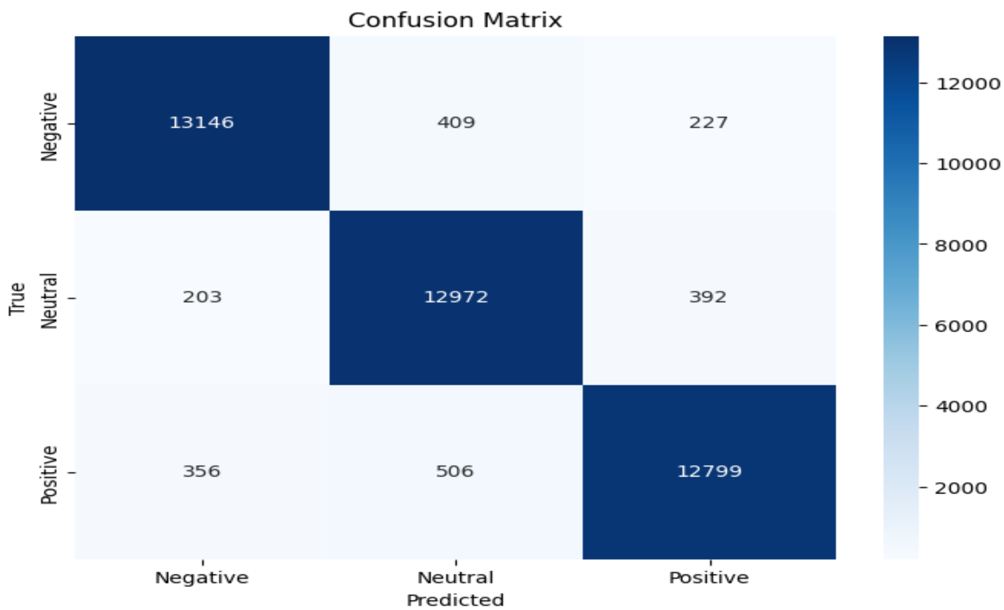


Figure 4.13: Confusion matrix of proposed Distil BERT model

Figure 4.13 shows the suggested Distil BERT model's confusion matrix, which gives a thorough look at how well it classified negative, neutral, and positive sentiments. A matrix reveals that for a Negative sentiment, out of the actual 13782 instances, 13146 were correctly classified as Negative, while 409 were incorrectly forecasted as Neutral and 227 as Positive. For the Neutral sentiment, of the 13567 real instances, 12972 were correctly classified, with 203 misclassified as Negative and 392 as Positive. In the case of the Positive sentiment, out of 13661 actual instances, 12799 were correctly identified, with 356 wrongly predicted as Negative and 506 as Neutral. Overall, the confusion matrix indicates a strong performance with high numbers along the diagonal, representing correct classifications, and relatively lower numbers in the off-diagonal cells, signifying misclassifications. The model's general high number of right predictions across all classes adds to the high accuracy given in the classification report, however there is a modest propensity for it to conflate Neutral and Positive feelings more than Negative and Neutral attitudes.

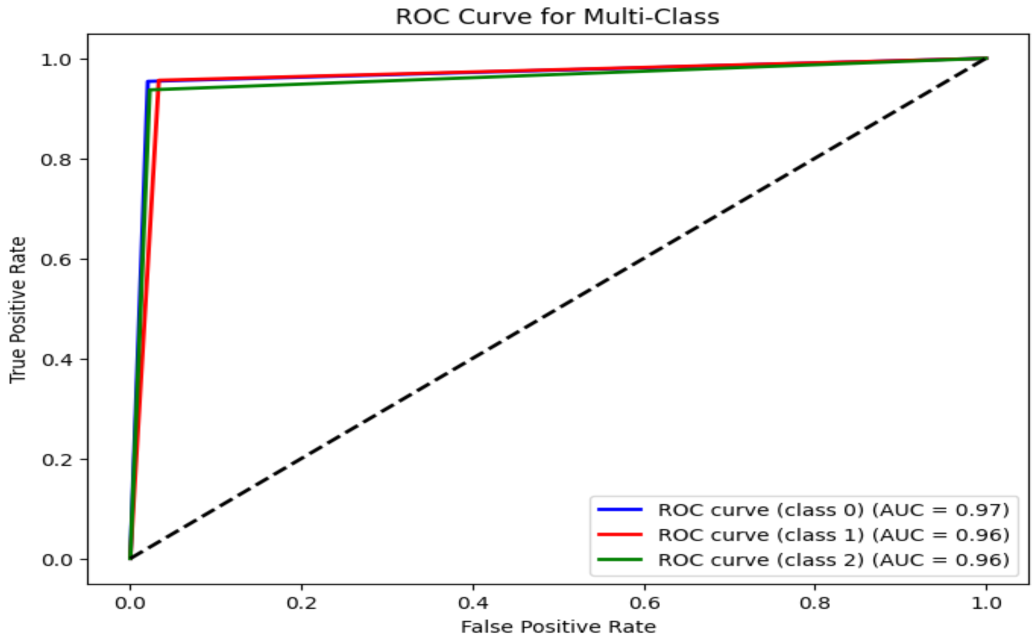


Figure 4.14: ROC curve of proposed Distil BERT model

Figure 4.14 displays the ROC curves for the multi-class sentiment classification performed by the proposed Distil BERT model. Each curve represents one of the 3 sentiment classes: Negative (class0, blue), Neutral (class1, red), and Positive (class2, green). The area under each ROC curve (AUC) is a measure of the model's ability to distinguish between the positive and negative categories for each emotion. The AUC for the Negative class is 0.97, indicating an excellent ability to differentiate negative reviews from the other sentiments. The Neutral class has an AUC of 0.96, also demonstrating strong discriminatory power. Similarly, the Positive class achieves an AUC of 0.96, signifying a high level of performance in distinguishing positive reviews. The Distil BERT model outperforms chance in sentiment classification across all three categories, since all three ROC curves are placed well above the diagonal dashed line, which indicates a random classifier. The strong AUC values across all classes demonstrate the model's effectiveness in SA and corroborate the positive performance metrics seen in the Classification Report and Confusion Matrix.

[51265/51265 10:20:06, Epoch 5/5]

Epoch	Training Loss	Validation Loss	Accuracy
1	0.436200	0.443955	0.837430
2	0.449800	0.362979	0.877323
3	0.415900	0.298417	0.901414
4	0.163500	0.273148	0.919044
5	0.153700	0.245539	0.933285

[2564/2564 09:54]

```
Evaluation results: {'eval_loss': 0.24553878605365753, 'eval_accuracy': 0.93
Sentiment: Positive
```

Figure 4.15: Performance parameters of proposed Albert model

Figure 4.15 displays the performance parameters of a proposed Albert model over five training epochs on the Flipkart product review sentiment analysis task. A steady decline in training and validation losses pointed to efficient generalization and learning. Starting with an accuracy of 83.74% in the 1epoch, the model showed steady improvement across subsequent epochs, reaching its peak accuracy of 93.33% by the fifth epoch. A validation loss also reduced from 0.443955 to 0.245539, supporting a model's improved performance. A final evaluation outcome confirmed an evaluation accuracy of 93.33% with a low evaluation loss of approximately 0.2455, and a correctly predicted sentiment label of "Positive," demonstrating the robustness and reliability of the fine-tuned Albert model in sentiment classification.

Classification Report:

	precision	recall	f1-score	support
Negative	0.94	0.94	0.94	13561
Neutral	0.93	0.93	0.93	13796
Positive	0.93	0.93	0.93	13653
accuracy			0.93	41010
macro avg	0.93	0.93	0.93	41010
weighted avg	0.93	0.93	0.93	41010

Figure 4.16: Classification Report of proposed Albert model

Figure 4.16 shows the suggested ALBERT model's classification report, which summarizes its performance for the Positive, Neutral, and Negative sentiment categories. For the Negative Class, the model achieves a recall of 0.94, a precision of 0.94, and an

F1score of 0.94, based on a support of 13561 instances. The Neutral class shows a precision of 0.93, a recall of 0.93, and an F1score of 0.93, with a support of 13796 instances. Similarly, for the Positive class, the model obtains a precision of 0.93, a recall of 0.93, and an F1score of 0.93, supported by 13653 instances. The ALBERT model achieves a class-wide accuracy of 0.93 on average. Precision, recall, and F1score all average out to 0.93 on the 41010-sample dataset, suggesting that the three metrics are well-balanced when it comes to sentiment classification. Compared to the Distil BERT model's classification report (Figure 4.7), the ALBERT model shows slightly lower precision, recall, and F1 scores for all sentiment classes, as well as a lower overall accuracy.

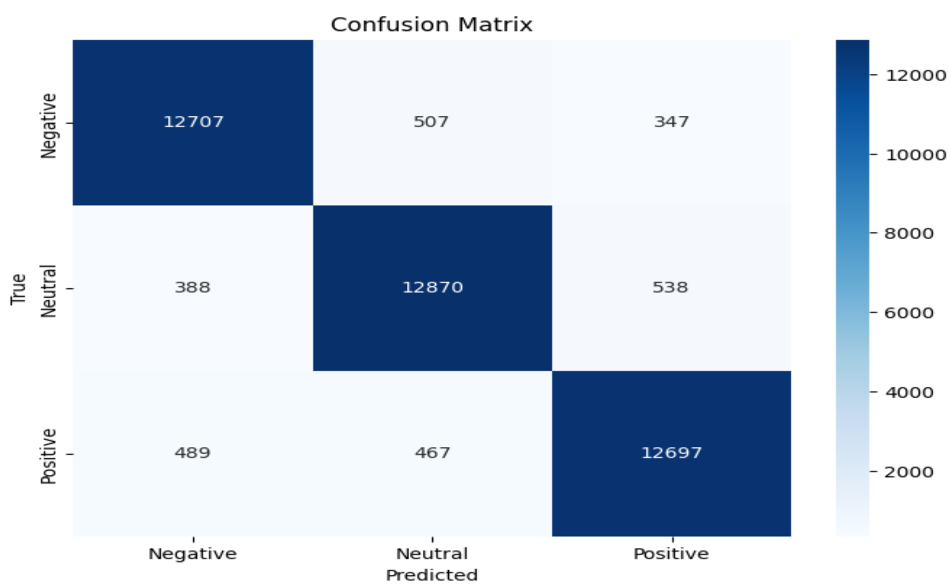


Figure 4.17: Confusion matrix of proposed Albert model

Figure 4.17 displayed the suggested ALBERT model's confusion matrix, which shows how well it classified negative, neutral, and positive sentiments. From the total of 13561 cases, 12707 were accurately categorized as Negative attitude, while 507 were mistakenly projected as Neutral and 347 as Positive. There were 13796 real occurrences of the neutral attitude; 12870 were accurately identified, 388 were incorrectly labelled as negative, and 538 were incorrectly classed as positive. Of the 13653 real occurrences of positive emotion, 12697 were classified correctly, while 489 were incorrectly forecasted as negative and 467 as neutral. A high percentage of diagonally-positioned accurate guesses for each emotion indicates solid performance in the Confusion Matrix. However, compared to the Distil BERT model's confusion matrix (Figure 4.8), the ALBERT model exhibits slightly more misclassifications across the sentiment categories, particularly between Negative and Neutral, and between Neutral and Positive sentiments, which aligns with the lower precision and recall scores observed in its classification report.

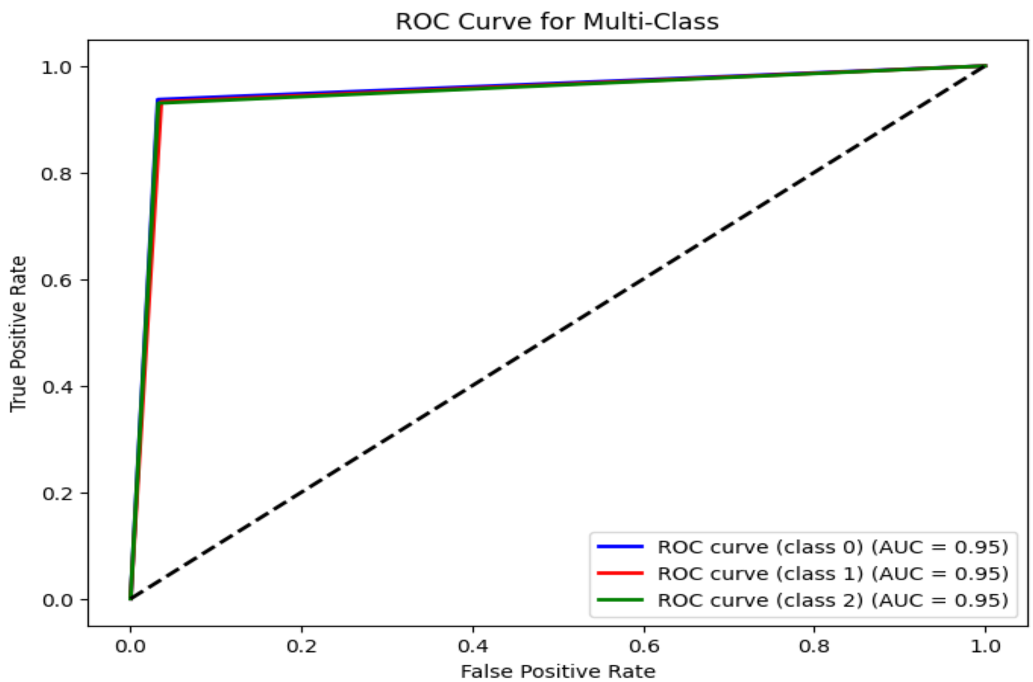


Figure 4.18: ROC of proposed Albert model

Figure 4.18 displays the ROC curves for the multi-class sentiment classification performed by the proposed ALBERT model. The AUC for the Negative class is 0.95, demonstrating a strong ability to differentiate negative reviews from other sentiments. Similarly, the Neutral class has an AUC of 0.95, indicating good discriminatory power. The Positive class also achieves an AUC of 0.95, signifying a high level of performance in distinguishing positive reviews. The ALBERT model outperforms chance in sentiment classification across all three categories, since all three ROC curves are situated high above the diagonal dashed line, which indicates a random classifier. The ALBERT model's strong performance metrics in the Classification Report and Confusion Matrix are supported by its high AUC values across all classes. However, when compared to the Distil BERT model, the ALBERT model's ability in SA is slightly lower, although it is still effective.

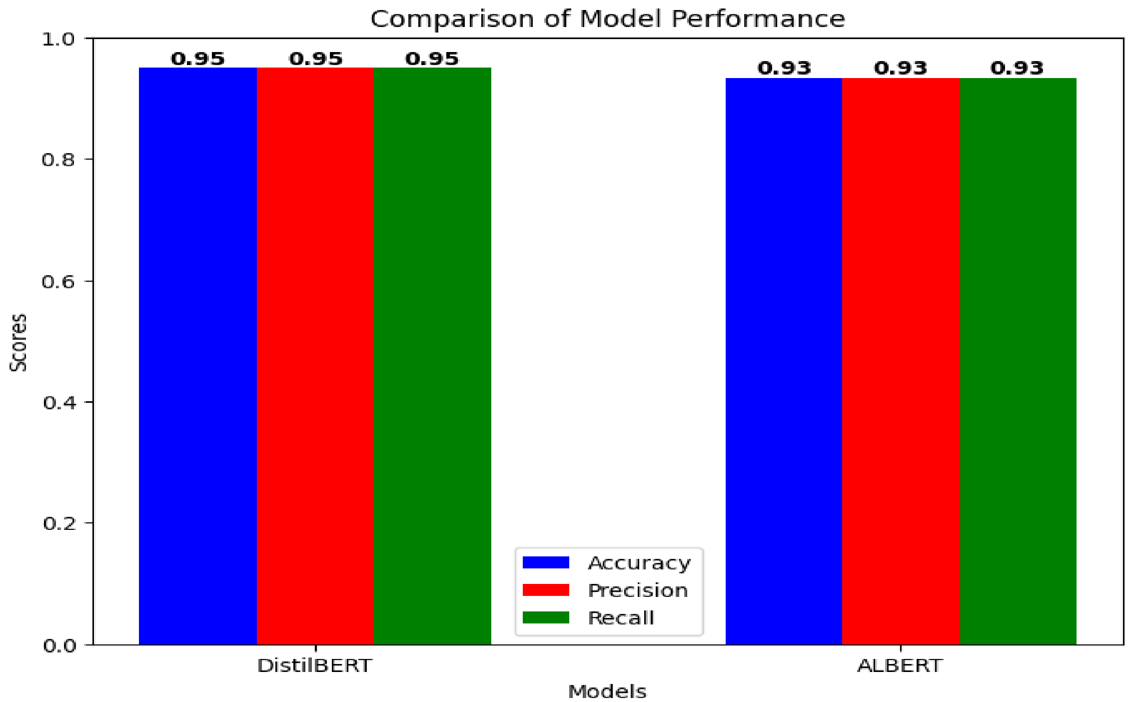


Figure 4.19: Comparison between Albert and Distil BERT Models

The Distil BERT and ALBERT models' performance indicators in sentiment analysis are directly compared in Figure 4.19. The bar chart displays the accuracy (blue), precision (red), and recall (green) achieved by each model. For Distil BERT, all three metrics show a value of 0.95, indicating a balanced and high level of performance across the entire dataset. In contrast, the ALBERT model exhibits slightly lower scores across the board, with an accuracy, precision, and recall of 0.93 for each metric. By consistently outperforming the ALBERT model in terms of accuracy, precision, and recall, as shown visually, the Distil BERT model definitely outperformed it in this SA test.

4.5 Conclusion

The proposed sentiment analysis model was developed using the Flipkart product reviews dataset. After thorough preprocessing—null value removal, normalization, lemmatization, and stop word removal—visualizations such as word clouds and sentiment distributions were generated. The sentiment labels were encoded, and the text features were combined and tokenized using Distil BERT and Albert tokenizers. Both models were trained and evaluated after balancing the dataset using a hybrid resampling approach. Distil BERT achieved higher performance with an accuracy, precision, recall, and F1score of 0.9490, while Albert followed closely with a consistent score of 0.9333 across all metrics. Analysis revealed Distil BERT's superior performance in sentiment classification.

CHAPTER V:

DISCUSSION

5.1 Discussion of Results

This research work of analyzing gradient boosting classifiers and text classification applied on the Flipkart product review dataset proved that AI is capable of understanding customer opinions in natural language by using Distil BERT and Albert Models. Two Hugging Face transformer models received fine-tuning following normalization then lemmatization and stop word elimination and hybrid data balancing methods during extensive preprocessing. The Distil BERT model outperformed Albert with 94.90% accuracy along with 93.33% precision, recall and F1score. Hence, these outcomes underscore once again the potential of the transformer models to learn deep contextual dependencies within textual data with Distil BERT though having relatively fewer parameters and relatively less training time. The confusion matrix for Distil BERT showed fewer misclassifications, indicating a better understanding of subtle differences between positive, neutral, and negative sentiments.

5.2 Comparison between existing and proposed models

Table 5.1 displays a Comparative Analysis among existing ML models (RF and SVM) and the proposed deep learning-based models (Distil BERT and ALBERT) for sentiment analysis on Flipkart Product Reviews. The results show that the suggested models outperformed the competition on every single criterion. While the existing models achieved moderate accuracy levels (0.7552 for Random Forest and 0.7580 for SVM), the proposed models significantly outperformed them, with Distil BERT achieving the highest accuracy of 0.9490 and ALBERT closely following at 0.9333. A similar trend is observed in precision, re-call, and F1score, where the proposed models consistently maintain values above 0.93, indicating their robustness and effectiveness in capturing sentiment nuances.

In contrast, the existing models exhibit lower precision and F1 scores, particularly SVM with an F1 score of 0.7354, reflecting limitations in handling complex linguistic patterns. Overall, the integration of transformer-based architectures in the proposed models significantly enhances the performance of Sentiment Analysis tasks compared to traditional machine learning approaches.

Table 5.1: Parameters comparison between Existing and proposed models for sentiment analysis on Flipkart product reviews

Parameters	Existing models		Proposed Models	
	Random Forest	SVM	Distil BERT	Albert
Accuracy	0.7552	0.7580	0.9490	0.9333
Precision	0.7586	0.7942	0.9491	0.9333
Recall	0.7552	0.7580	0.9490	0.9333
F1 Score	0.7529	0.7354	0.9490	0.9333

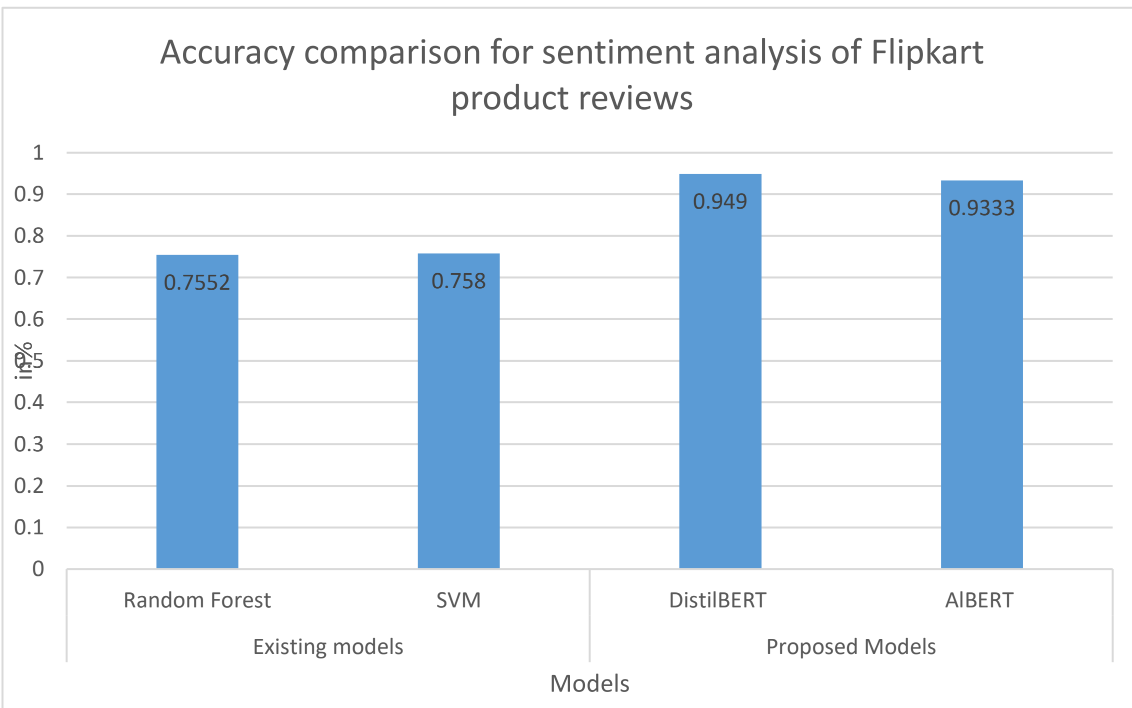


Figure 5.1: Comparison of accuracy measure between existing and proposed models

Figure 5.1 presents a comparison of accuracy measures for Sentiment Analysis of Flipkart Product Reviews using existing models (Random Forest and SVM) and proposed models (Distil BERT and AIBERT). A y-axis shows accuracy in percentage, while an x-axis indicates the models. According to the findings, Distil BERT and AIBERT outperform the current models in terms of accuracy. Distil BERT shows an accuracy of 0.949 or 94.9%, and AIBERT achieves an accuracy of 0.9333 or 93.33%. In contrast, the existing models, RF and SVM, exhibit lower accuracy levels, with RF at 0.7552 or 75.52% and SVM at 0.758 or 75.8%. This figure also clearly indicates that the proposed Distil BERT and the AIBERT models performs better than the Random Forest and the SVM models to accurately predict the overall sentiment of Flipkart Product Reviews.

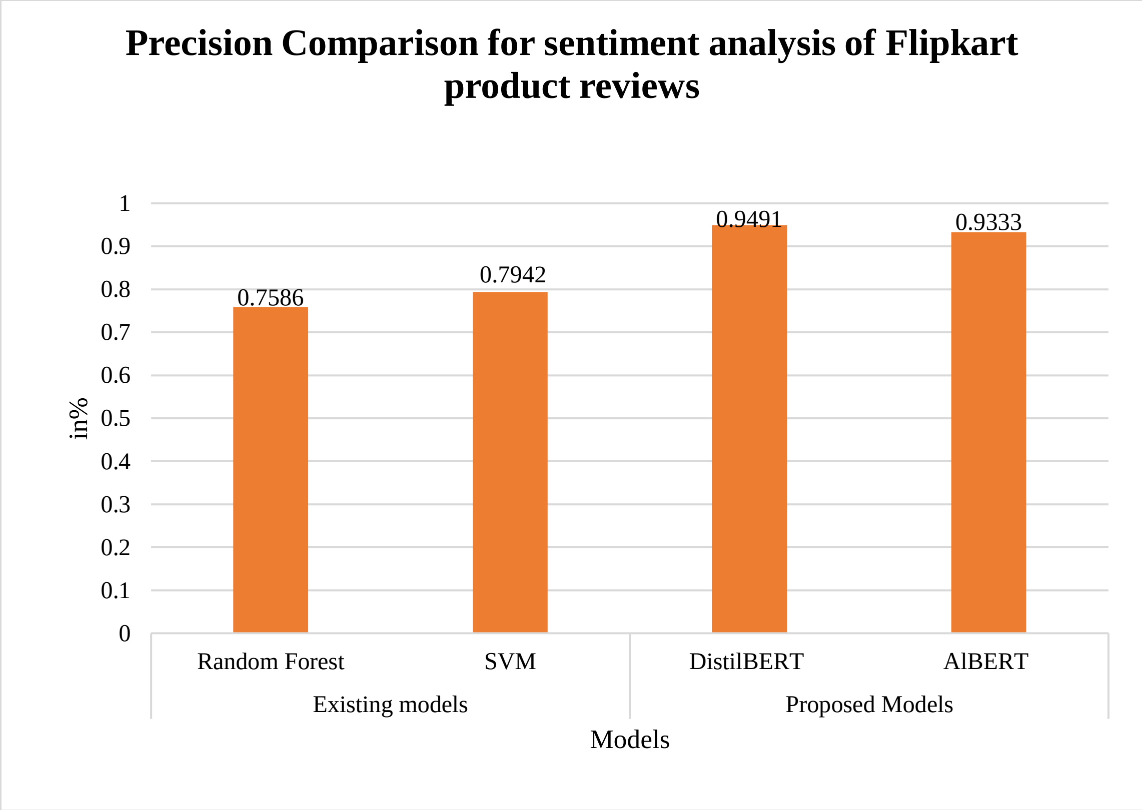


Figure 5.2: Comparison of precision measure between existing and proposed models

Figure 5.2 illustrates the precision comparison for Sentiment Analysis of Flipkart Product Reviews using both existing models (Random Forest and SVM) and proposed models (Distil BERT and AIBERT). A y-axis displays the precision values, while an x-axis identifies the respective models. The outcomes shows that the proposed models demonstrate higher precision compared to the existing ones. Specifically, Distil BERT achieves a precision of 0.9491, and AIBERT shows a precision of 0.9333. In contrast, the existing models, Random Forest and SVM, exhibit lower precision values, with Random Forest at 0.7586 and SVM at 0.7942. This comparison suggests that the proposed Distil BERT and AIBERT models are better at correctly identifying positive sentiment among the reviews that are classified as positive, with fewer false positives compared to the traditional RF and SVM models.

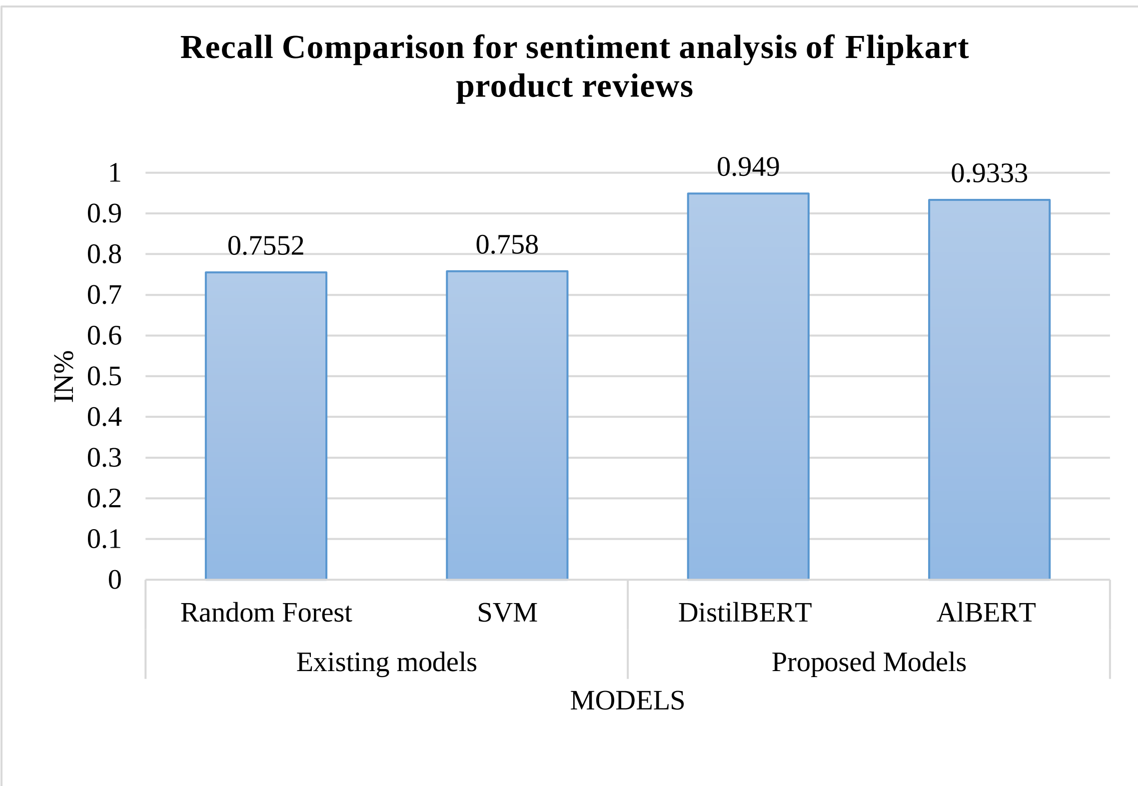


Figure 5.3: Comparison of recall measure between existing and proposed models

Figure 5.3 presents a comparison of the recall measure for Sentiment Analysis of Flipkart Product Reviews using existing models (Random Forest and SVM) and proposed models (Distil BERT and AIBERT). The recall values appear on a y-axis while the model types are displayed on an x-axis. The study findings demonstrate that both Distil BERT and AIBERT deliver higher recall performance when compared to previous models. Distil BERT demonstrates a recall of 0.949, and AIBERT achieves a recall of 0.9333. In contrast, the existing models exhibit lower recall values, with Random Forest at 0.7552 and SVM at 0.758. The proposed Distil BERT and AIBERT models outperform traditional Random Forest and SVM models in detecting positive reviews by the Flipkart dataset leading to decreased false negative results.

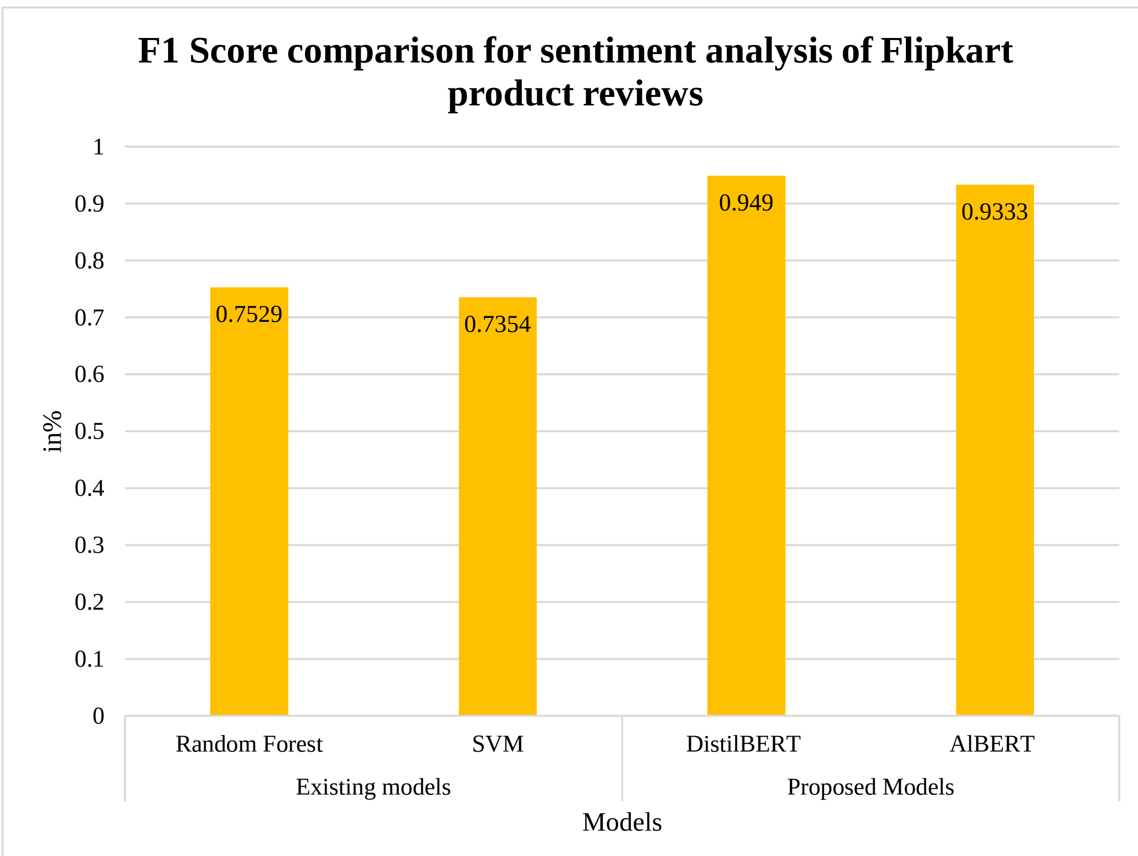


Figure 5.4: Comparison of f1-score measure between existing and proposed models

Figure 5.4 presents a comparison of the F1score for Sentiment Analysis of Flipkart Product Reviews using existing models (Random Forest and SVM) and proposed models (Distil BERT and ALBERT). A y-axis indicates the F1-score values, while a x-axis indicates the models. An experimental outcome display that the Distil BERT and ALBERT models reach higher F1-scores than all existing models tested. Distil BERT shows an F1score of 0.949, and ALBERT achieves an F1score of 0.9333. The current models possess reduced F1-scores which stand at 0.7529 for Random Forest and 0.7354 for SVM. These findings show that the suggested Distil BERT and ALBERT models outperform the conventional RF and SVM models according to achieving a harmonic mean of re-call and precision when it comes to Sentiment Analysis of Flipkart Product Reviews.

5.3 Discussion of Each Research Questions

In this section, discuss the findings related to each research question posed in this study:

RQ 1: How effective are AI-based advanced models in performing sentiment analysis on e-commerce product reviews compared to traditional machine learning models?

Discussion: Various experiments of sentiment analysis indicate that the state-of-art AI-based advanced models like Distil BERT and Albert performed better than traditional ML models such as RF and SVM. The design of transformer models enables them to identify intricate linguistic patterns alongside context automatically from large datasets that results in exceptional performance during review analysis. In contrast to previous models that can require quite a lot of feature engineering and may use only basic BoW, transformers employ DL mechanisms to learn the texts. An analysis of the e-commerce product reviews' sentiments show that transformer models offer relatively better accuracy, precision, re-call, and F1 scores compared to the traditional models. This enhancement in performance signifies suitability of AI models to deal with huge volumes of untidy text data to achieve more refined sentiment categorization.

RQ 2: What impact do advance text preprocessing techniques and data balancing have on the performance of sentiment classification models?

Discussion: To make the sentiment classification algorithms effective, data balance and advanced text preparation methods play an important role. Tokenization and normalization helps to detect characteristics of the text input for further analysis while lemmatization and stop word removal helps to filter out the noise. As an example, lemmatization aids in reducing words to their simplest form, which increases the model's generalizability to diverse word forms. The common class imbalance problem affecting sentiment analysis datasets becomes manageable through data balancing strategies such as under sampling dominant cases together with oversampling minority cases. The classification accuracy improves when feelings are distributed evenly because this prevents the model from preferring the most common class. Overall, these preprocessing and balancing methods boost the reliability and accuracy of SA models.

RQ 3: How can AI-driven sentiment analysis contribute to improved customer experience and strategic decision-making in future business scenarios?

Discussion: Mobile application of artificial intelligence in form of SA enriches customer experience and advanced business decisions due to improved knowledge about clients' opinions. The analysis of large-scale customer reviews reveals recurring customer problems and positive sentiments and unfulfilled needs which allows businesses to take preventive actions toward meeting their customers' needs. Such a technique is useful in customization of customers by targeting their needs and expectations in products and services as well as in marketing *strategies*. Moreover, there is a critical role of using AI in driving sentiments, analyses of brand reputation, assessing competitors, or evaluating marketing campaign performances. Software technology will assist businesses to manage extensive consumer input through data analysis thereby helping organizations make better choices that bring about better customer satisfaction and sustained expansion.

5.4 Conclusion

The analysis establishes that transformer-models Distil BERT and Albert deliver superior results than standard ML models RF and SVM for analyzing product review sentiment in online retail environments. These models are more refined by other methods such as text preprocessing and data balancing making the classification of sentiments more accurate and efficient. Business organizations can also benefit from the decisions and overall organizational strategies that are based on AI sentiment analysis that can be used to point business towards the right direction in relation to the consumers and to further improve the quality of the customers experience.

CHAPTER VI:

SUMMARY, IMPLICATIONS, AND RECOMMENDATIONS

6.1 Summary

Customer sentiment analysis from product reviews in e-commerce platforms stands as a crucial problem for businesses who want to enhance customer experiences through data-based decisions. Natural language processing proves difficult for classic machine learning models operating on vast review database text. This study addresses the challenge by employing two state-of-the-art transformer models, Distil BERT and Albert, for SA on Flipkart product reviews. The models shared equivalent preprocessing methods that integrated tokenization, padding with additional dimensions and feature unification and applied a hybrid resampling method for data equality. Distil BERT produced superior performance than Albert in the sentiment analysis task through higher accuracy measurements (94.90%) with precision (94.91%) and recall (94.90%) and F1 score (94.90%), whereas Albert reached 93.33% for all metrics. Both transformer models demonstrated robust performance although their measurement metrics varied slightly because of their different processing capabilities of large-scale sentiment analysis tasks in business applications and customer feedback analysis.

6.2 Research contribution /Novelty and Justification

The research advances sentiment analysis by utilizing Distil BERT and Albert transformers to analyze genuine product reviews available on Flipkart. The main contribution of this research appears through its combination of advanced preprocessing methods including lemmatization and stop-word removal with hybrid resampling for balancing data. The preprocessing strategies together with the strong transformer models enhance the model's capabilities to detect sentiments with accuracy. The study implements advanced technological methods to solve the difficulties that accompany analysis of extensive unorganized e-commerce data. The choice to use transformer-based models including Distil BERT and Albert results from their demonstrated better results compared to standard NLP machine learning models. This research shows these models achieve precise outcomes in addition to effective operation which makes them ideal for real-time SA work. These latest models have become ideal for the classification of sentiments that help organizations to derive great benefits from information gathered from a large number of customers. Through the assessment of the research, it has been established that the application of transformer models can help in the achievement of business-oriented decisions based on sentiments hence improve on their adoption to fit customer's needs.

Based on the study on sentiment analysis using Distil BERT and Albert transformer models for Flipkart product reviews, here are eight research contributions:

- **Transformer Model Performance in Sentiment Analysis:** The research demonstrates that Distil BERT achieves superior performance than Albert in sentiment analysis while delivering greater accuracy and precision together with better recall and F1 scores. This study shows that Distil BERT has a significant capacity to perform in different tasks of sentiment classification.
- **Hybrid Resampling Approach for Data Balancing:** Hybrid resampling techniques assist to bring the variance of sentiment classes to the same level during the training process thus improving performance on overall sentiments classification activities.
- **Effectiveness of Pretrained Transformer Models in Business Applications:** The paper providing the empirical evidence on the benefits of Distil BERT and Albert transformer architectures in the large-scale analysis of product reviews messages with the aim to extract essential information from the unstructured business text.

- **Impact of Tokenization and Padding on Model Performance:** This paper also explains how data preprocessing steps like tokenization and padding enhance the input data because these steps influence Distil Bert's and Albert's performance.
- **Analysis of Model Evaluation Metrics:** Specifically for the goal of choosing a model for use in other applications, the accuracy, precision, recall, and F1score are used to evaluate the models' strengths and shortcomings with respect to the transformer-based models.
- **Fine-Tuning Pretrained Models for Specific Domains:** Product review sentiment analysis can be performed efficiently after specialized training on the predefined transformer models like Distil BERT and Albert.
- **Comparison with Traditional Machine Learning Models:** Thus, compared to other conventional ML algorithms like the currently available RF and SVM, transformer models score better on accuracy metrics, making them a great choice for SA in the future.
- **Scalability of Transformer Models for Large-Scale Sentiment Analysis:** The transformer models incorporated by Distil BERT and Albert and it shows high dataset processing rate and accuracy as well to make it more convenient to apply in practical industries approach especially the e-commerce and product development field using sentiment analysis.

These contributions enhance the importance of current transformers in handling other challenging activities in intelligent sentiment analysis and their escalating importance in the business context.

6.3 Implications/Limitations

This research demonstrates how companies can leverage Artificial Intelligence to improve business performance, reviews, processing and decision making in real time. This model only functions for analyzing Flipkart review data yet its computational needs and linguistic complexities alongside dataset preconceptions could influence its performance. The following limitations of this work as:

1) **Enhanced Business Intelligence through AI**

The approach demonstrated that how the two models Distil BERT and Albert have the potential to make a difference in giving valuable insights from the customers' feedback. Saves a lot of the business time which makes it easier for it to attend to the clients' needs, makes better market decisions to make and be able to offer personalized recommendations. This enhances customers' satisfaction, loyalty, and brand loyalty thus providing a competitive advantage in the current market.

2) **Scalability and Real-Time Application**

The high accuracy obtained (94.90% for Distil BERT and 93.33% for Albert) gives this work a stable and reliable platform for employing sentiment analysis across large datasets. The concepts described above can be applied to other e-commerce systems or personalized to different industries, and the information provided helps respond to the immediate detection of public sentiment and trends as well as potential crises for more effective data-driven decision-making.

3) **Domain-Specific Dataset Limitation**

The model faces a major drawback because its training occurs with Flipkart reviews leading to limited suitability outside the platform's domain and language boundaries. Performance maintenance requires the use of domain adaptation methods or dataset retraining since vocabulary, style, sentiment expression differs greatly across business domains and regions.

4) **Computational Resource Dependency**

The accuracy achievements of transformer models come at a high computational expense. Running Transformer models demands both powerful GPUs or TPUs for training functions which many small companies together with academic organizations cannot afford. This restricts the widespread deployment of such models unless optimized alternatives or cloud-based solutions are used.

5) **Challenges with Language Nuances**

The model struggles to correctly interpret linguistics elements that include sarcasm and slang together with idioms and culturally sensitive expressions. The model struggles to translate essential sentiment information found in linguistic nuances which results in possible misdiagnosis. Future development requires enhanced ability from models to understand linguistic elements effectively.

6) **Dataset Quality and Label Bias**

The learning process of the model gets affected when noisy labels and user-generated biases exist in the dataset. The dataset predictions can become inaccurate when reviews lack proper labeling or when particular sentiments create unbalanced data distribution. This can be avoided by using highly accurate labeling or through techniques such as active learning and human validation on the labels.

6.4 Recommendations for Future Work

The following research Recommendations for Future work:

1) **Cross-Domain and Multilingual Model Adaptation**

Further studies regarding extending existing models of SA with using the method to analyze multilingual data and working with data from different domains could be chosen for further investigation. This would enhance versatility of the specific formulated model and make it suitable for application in diverse fields, languages, and cultures, which would increase its applicability at the international level.

2) **Handling Linguistic Nuances**

Possible future improvements for enhancing an effectiveness and enhancing the accuracy of the SA include considering the models that can operate with sarcasm, idioms, slang terms, or other forms of cultural sensitivity. Eliminating these styles of language would enable models to better perceive detailed and nuanced sentiments present in customer reviews.

3) **Model Efficiency and Lightweight Architectures**

As a result of the computational complexity of transformer-based models, future research can aim to either design efficient transformers or employ methods that reduce the complexity of these models, making them suitable to be used in small businesses or mobile apps.

4) **Incorporating Explainable AI (XAI)**

The combination of Explainable AI (XAI) in the sentiment analysis models means that businesses can further understand the reasons behind decisions made in the process. Additional research may focus on explaining how these AI models work to classify sentiments so that there is better acceptance of AI insights within business organizations.

5) **Improved Data Labeling and Bias Mitigation**

The researchers can extend their work by fine-tuning data labeling and gaining awareness of the sources of bias in order to get better training datasets. This would help to overcome such problems associated with label noise or bias, thus reducing potential distortions in the output of SA.

6.5 Conclusion

In this work demonstrates an effectiveness of transformer models, specifically Distil BERT and Albert, for SA of Flipkart Product Reviews. Distil BERT achieves better scores than Albert for accuracy, precision, recall and F1 metrics which demonstrates its effectiveness for handling big-scale sentiment classification operations. Both Albert and Distil BERT demonstrate powerful sentiment detection abilities from textual content which confirms transformer models as highly effective tools for sentiment analysis in business fields particularly retail e-commerce. Pretrained transformer models demonstrate the ability to improve business customer feedback analysis by generating valuable insights which support better decision-making and enhance customer satisfaction.

REFERENCES

Adel, H., Dahou, A., Mabrouk, A., Elaziz, M. A., Kayed, M., El-Henawy, I. M., Alshathri, S., & Ali, A. A. (2022). Improving Crisis Events Detection Using DistilBERT with Hunger Games Search Algorithm. *Mathematics*, *10*(3), 1–22. https://doi.org/10.3390/math10030447

Ahuja, R., Chug, A., Kohli, S., Gupta, S., & Ahuja, P. (2019). The impact of features extraction on the sentiment analysis. *Procedia Computer Science*. https://doi.org/10.1016/j.procs.2019.05.008

Akter, M. T., Begum, M., & Mustafa, R. (2021). Bengali Sentiment Analysis of E-commerce Product Reviews using K-Nearest Neighbors. *2021 International Conference on Information and Communication Technology for Sustainable Development (ICICT4SD),* 40–44. https://doi.org/10.1109/ICICT4SD50815.2021.9396910

Al-Natour, S., & Turetken, O. (2020). A comparative assessment of sentiment analysis and star ratings for consumer reviews. *International Journal of Information Management*. https://doi.org/10.1016/j.ijinfomgt.2020.102132

Alahmadi, A. A., Aljabri, M., Alhaidari, F., Alharthi, D. J., Rayani, G. E., Marghalani, L. A., Alotaibi, O. B., & Bajandouh, S. A. (2023). DDoS Attack Detection in IoT-Based Networks Using Machine Learning Models: A Survey and Research Directions. In *Electronics (Switzerland)*. https://doi.org/10.3390/electronics12143103

Alamdari, P. M., Navimipour, N. J., Hosseinzadeh, M., Safaei, A. A., & Darwesh, A. (2020). A Systematic Study on the Recommender Systems in the E-Commerce. *IEEE Access*. https://doi.org/10.1109/ACCESS.2020.3002803

Alghazzawi, D. M., Alquraishee, A. G. A., Badri, S. K., & Hasan, S. H. (2023). ERF-XGB: Ensemble Random Forest-Based XG Boost for Accurate Prediction and Classification of E-Commerce Product Review. *Sustainability (Switzerland)*. https://doi.org/10.3390/su15097076

Almahmood, R. J. K., & Tekerek, A. (2022). Issues and Solutions in Deep Learning-Enabled Recommendation Systems within the E-Commerce Field. In *Applied Sciences (Switzerland)*. https://doi.org/10.3390/app122111256

Almuayqil, S. N., Humayun, M., Jhanjhi, N. Z., Almufareh, M. F., & Javed, D. (2022). Framework for Improved Sentiment Analysis via Random Minority Oversampling for User Tweet Review Classification. *Electronics (Switzerland)*. https://doi.org/10.3390/electronics11193058

Alsayat, A., & Elmitwally, N. (2020). A comprehensive study for Arabic Sentiment Analysis (Challenges and Applications). *Egyptian Informatics Journal*. https://doi.org/10.1016/j.eij.2019.06.001

Alzahrani, M. E., Aldhyani, T. H. H., Alsubari, S. N., Althobaiti, M. M., & Fahad, A. (2022a). Developing an Intelligent System with Deep Learning Algorithms for Sentiment Analysis of E-Commerce Product Reviews. In *Computational Intelligence and Neuroscience*. https://doi.org/10.1155/2022/3840071

Alzahrani, M. E., Aldhyani, T. H. H., Alsubari, S. N., Althobaiti, M. M., & Fahad, A. (2022b). Developing an Intelligent System with Deep Learning Algorithms for Sentiment Analysis of E-Commerce Product Reviews. *Computational Intelligence and Neuroscience*, *2022*(1), 3840071. https://doi.org/https://doi.org/10.1155/2022/3840071

Aversa, J., Doherty, S., & Hernandez, T. (2018). Big Data Analytics: The New Boundaries of Retail Location Decision Making. *Papers in Applied Geography*. https://doi.org/10.1080/23754931.2018.1527720

Ayub, U. (2022). *In-Depth Review of Convolutional Neural Networks (CNN's)*.

https://umairayub110.medium.com/convolutional-neural-networks-cnns-1ed0f1f8e176

Bahdanau, D., Cho, K. H., & Bengio, Y. (2015). Neural machine translation by jointly learning to align and translate. *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings*.

Bayhaqy, A., Sfenrianto, S., Nainggolan, K., & Kaburuan, E. R. (2018). Sentiment Analysis about E-Commerce from Tweets Using Decision Tree, K-Nearest Neighbor, and Naïve Bayes. *2018 International Conference on Orange Technologies (ICOT)*, 1–6. https://doi.org/10.1109/ICOT.2018.8705796

Behera, R. K., Jena, M., Rath, S. K., & Misra, S. (2021). Co-LSTM: Convolutional LSTM model for sentiment analysis in social big data. *Information Processing and Management*. https://doi.org/10.1016/j.ipm.2020.102435

Bellar, O., Baina, A., & Ballafkih, M. (2024). Sentiment Analysis: Predicting Product Reviews for E-Commerce Recommendations Using Deep Learning and Transformers. *Mathematics*, *12*(15). https://doi.org/10.3390/math12152403

Bello, A., Ng, S. C., & Leung, M. F. (2023). A BERT Framework to Sentiment Analysis of Tweets. *Sensors*, *23*(1). https://doi.org/10.3390/s23010506

Benarafa, H., Benkhalifa, M., & Akhloufi, M. (2023). WordNet Semantic Relations Based Enhancement of KNN Model for Implicit Aspect Identification in Sentiment Analysis. *International Journal of Computational Intelligence Systems*. https://doi.org/10.1007/s44196-022-00164-8

Bilal Chandio Asadullah Shaikh, M. B. M. A. J. B. A. S. A. R. W. N. (2022). Sentiment Analysis of Roman Urdu on E-Commerce Reviews Using Machine Learning. *Computer Modeling in Engineering \& Sciences*, *131*(3), 1263–1287. https://doi.org/10.32604/cmes.2022.019535

Birjali, M., Kasri, M., & Beni-Hssane, A. (2021). A comprehensive survey on sentiment analysis: Approaches, challenges and trends. *Knowledge-Based Systems*. https://doi.org/10.1016/j.knosys.2021.107134

Bonifazi, G., Cauteruccio, F., Corradini, E., Marchetti, M., Terracina, G., Ursino, D., & Virgili, L. (2023). A framework for investigating the dynamics of user and community sentiments in a social platform. *Data & Knowledge Engineering*, *146*, 102183. https://doi.org/https://doi.org/10.1016/j.datak.2023.102183

Cambria, E. (2016). Affective Computing and Sentiment Analysis. *IEEE Intelligent Systems*. https://doi.org/10.1109/MIS.2016.31

Chakraborty, J., Xia, W., Majumder, A., Ma, D., Chaabene, W., & Janvekar, N. (2024). *DetoxBench: Benchmarking Large Language Models for Multitask Fraud & Abuse Detection*. https://doi.org/10.48550/arXiv.2409.06072

Chamekh, A., Mahfoudh, M., & Forestier, G. (2022). Sentiment Analysis Based on Deep Learning in E-Commerce. In G. Memmi, B. Yang, L. Kong, T. Zhang, & M. Qiu (Eds.), *Knowledge Science, Engineering and Management* (pp. 498–507). Springer International Publishing.

Chath, H. R., & Chhatrala, B. K. (2023). Sentiment Analysis of Online Product Reviews Using Modified Deep Learning Model. *International Journal of Darshan Institute on Engineering Research and Emerging Technologies*, *12*(1), 13–20. https://doi.org/10.32692/ijdi-eret/12.1.2023.2303

Chaturvedi, S., Mishra, V., & Mishra, N. (2017). Sentiment analysis using machine learning for business intelligence. *2017 IEEE International Conference on Power, Control, Signals and Instrumentation Engineering (ICPCSI)*, 2162–2166. https://doi.org/10.1109/ICPCSI.2017.8392100

Chifu, A. G., & Fournier, S. (2023). Sentiment Difficulty in Aspect-Based Sentiment

Analysis. *Mathematics*. https://doi.org/10.3390/math11224647

Chintalapudi, N., Battineni, G., Canio, M. Di, Sagaro, G. G., & Amenta, F. (2021). Text mining with sentiment analysis on seafarers' medical documents. *International Journal of Information Management Data Insights*. https://doi.org/10.1016/j.jjimei.2020.100005

Collomb, A. A., Joyeux, D., Hasan, O., & Brunie, L. (2014). A Study and Comparison of Sentiment Analysis Methods for Reputation Evaluation. *Rapport de Recherche RR-LIRIS-2014-002*, 1–10.

Cunningham, P., & Delany, S. J. (2021). K-Nearest Neighbour Classifiers-A Tutorial. In *ACM Computing Surveys*. https://doi.org/10.1145/3459665

Dang, C. N., Moreno-García, M. N., & De la Prieta, F. (2021). An approach to integrating sentiment analysis into recommender systems. *Sensors*. https://doi.org/10.3390/s21165666

Dang, N. C., Moreno-García, M. N., & De la Prieta, F. (2020). Sentiment analysis based on deep learning: A comparative study. *Electronics (Switzerland)*. https://doi.org/10.3390/electronics9030483

Das, R. K., Islam, M., Hasan, M. M., Razia, S., Hassan, M., & Khushbu, S. A. (2023). Sentiment analysis in multilingual context: Comparative analysis of machine learning and hybrid deep learning models. *Heliyon, 9*(9). https://doi.org/10.1016/j.heliyon.2023.e20281

Daza, A., González Rueda, N. D., Aguilar Sánchez, M. S., Robles Espíritu, W. F., & Chauca Quiñones, M. E. (2024). Sentiment Analysis on E-Commerce Product Reviews Using Machine Learning and Deep Learning Algorithms: A Bibliometric Analysis, Systematic Literature Review, Challenges and Future Works. *International Journal of Information Management Data Insights, 4*(2), 100267.

https://doi.org/https://doi.org/10.1016/j.jjimei.2024.100267

de Oliveira Carosia, A. E., Coelho, G. P., & da Silva, A. E. A. (2021). Investment strategies applied to the Brazilian stock market: A methodology based on Sentiment Analysis with deep learning. *Expert Systems with Applications*. https://doi.org/10.1016/j.eswa.2021.115470

Demircan, M., Seller, A., Abut, F., & Akay, M. F. (2021). Developing Turkish sentiment analysis models using machine learning and e-commerce data. *International Journal of Cognitive Computing in Engineering*, *2*, 202–207. https://doi.org/https://doi.org/10.1016/j.ijcce.2021.11.003

Deng, L. (2014). A tutorial survey of architectures, algorithms, and applications for deep learning. In *APSIPA Transactions on Signal and Information Processing*. https://doi.org/10.1017/atsip.2013.9

Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. *NAACL HLT 2019 - 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies - Proceedings of the Conference*.

Dogra, V., Verma, S., Kavita, Wozniak, M., Shafi, J., & Ijaz, M. F. (2024). Shortcut Learning Explanations for Deep Natural Language Processing: A Survey on Dataset Biases. *IEEE Access*. https://doi.org/10.1109/ACCESS.2024.3360306

Eke, C. I., Norman, A. A., & Shuib, L. (2021). Context-Based Feature Technique for Sarcasm Identification in Benchmark Datasets Using Deep Learning and BERT Model. *IEEE Access*. https://doi.org/10.1109/ACCESS.2021.3068323

Elangovan, D., & Subedha, V. (2023). Adaptive Particle Grey Wolf Optimizer with Deep Learning-based Sentiment Analysis on Online Product Reviews. *Engineering, Technology & Applied Science Research*, *13*(3), 10989–10993.

https://doi.org/10.48084/etasr.5787

Esfahanian, S., & Lee, E. (2022). A novel packaging evaluation method using sentiment analysis of customer reviews. *Packaging Technology and Science*. https://doi.org/10.1002/pts.2686

Faria, F. T. J., Baniata, L. H., Baniata, M. H., Khair, M. A., Bani Ata, A. I., Bunterngchit, C., & Kang, S. (2025). SentimentFormer: A Transformer-Based Multimodal Fusion Framework for Enhanced Sentiment Analysis of Memes in Under-Resourced Bangla Language. *Electronics (Switzerland)*, *14*(4), 1–33. https://doi.org/10.3390/electronics14040799

Geetha, M. P., & Karthika Renuka, D. (2021). Improving the performance of aspect based sentiment analysis using fine-tuned Bert Base Uncased model. *International Journal of Intelligent Networks*. https://doi.org/10.1016/j.ijin.2021.06.005

Ghaffari, S., Yousefimehr, B., & Ghatee, M. (2024). Generative-AI in E-Commerce: Use-Cases and Implementations. *2024 20th CSI International Symposium on Artificial Intelligence and Signal Processing (AISP)*, 1–5. https://doi.org/10.1109/AISP61396.2024.10475266

Ghatora, P. S., Hosseini, S. E., Pervez, S., Iqbal, M. J., & Shaukat, N. (2024). Sentiment Analysis of Product Reviews Using Machine Learning and Pre-Trained LLM. *Big Data and Cognitive Computing*, *8*(12). https://doi.org/10.3390/bdcc8120199

Gilhooly, K. (2024). AI vs humans in the AUT: Simulations to LLMs. *Journal of Creativity*. https://doi.org/10.1016/j.yjoc.2023.100071

Goel, P., Jain, R., Nayyar, A., Singhal, S., & Srivastava, M. (2022). Sarcasm detection using deep learning and ensemble learning. *Multimedia Tools and Applications*. https://doi.org/10.1007/s11042-022-12930-z

Gope, J. C., Tabassum, T., Mabrur, M. M., Yu, K., & Arifuzzaman, M. (2022). Sentiment

Analysis of Amazon Product Reviews Using Machine Learning and Deep Learning Models. *2022 International Conference on Advancement in Electrical and Electronic Engineering (ICAEEE)*, 1–6. https://doi.org/10.1109/ICAEEE54957.2022.9836420

Guenther, N., & Schonlau, M. (2016). Support vector machines. *Stata Journal*. https://doi.org/10.1177/1536867x1601600407

Gupta, K., Jiwani, N., & Afreen, N. (2023). A Combined Approach of Sentimental Analysis Using Machine Learning Techniques. *Revue d'Intelligence Artificielle*, *37*(1), 1–6. https://doi.org/10.18280/ria.370101

Han, J., Kamber, M., & Pei, J. (2012). Data Mining: Concepts and Techniques. In *Data Mining: Concepts and Techniques*. https://doi.org/10.1016/C2009-0-61819-5

He, H., Zhou, G., & Zhao, S. (2022). Exploring E-Commerce Product Experience Based on Fusion Sentiment Analysis Method. *IEEE Access*. https://doi.org/10.1109/ACCESS.2022.3214752

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. https://doi.org/10.1109/CVPR.2016.90

Henderi, Henderi & Siddique, Q. (2024). Comparative Analysis of Sentiment Classification Techniques on Flipkart Product Reviews : A Study Using Logistic Regression , SVC , Random Forest , and Gradient Boosting. *Journal of Digital Market and Digital Currency*, *1*(1), 21–42. https://doi.org/10.47738/jdmdc.v1i1.4

Hicham, N., Habbat, N., & Karim, S. (2023). Enhancing Arabic E-Commerce Review Sentiment Analysis Using a hybrid Deep Learning Model and FastText word embedding. *EAI Endorsed Transactions on Internet of Things*. https://doi.org/10.4108/eetiot.4601

Hochreiter, S., & Schmidhuber, J. (1997). Long Short-Term Memory. *Neural*

Computation. https://doi.org/10.1162/neco.1997.9.8.1735

Huang, H., Zavareh, A. A., & Mustafa, M. B. (2023). Sentiment Analysis in E-Commerce Platforms: A Review of Current Techniques and Future Directions. *IEEE Access*, *11*(July), 90367–90382. https://doi.org/10.1109/ACCESS.2023.3307308

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). https://doi.org/10.3390/app12147182

Iqbal, A., Amin, R., Iqbal, J., Alroobaea, R., Binmahfoudh, A., & Hussain, M. (2022). Sentiment Analysis of Consumer Reviews Using Deep Learning. *Sustainability (Switzerland)*. https://doi.org/10.3390/su141710844

Irfan, M., Sanka, A. I., Ullah, Z., & Cheung, R. C. C. (2022). Reconfigurable content-addressable memory (CAM) on FPGAs: A tutorial and survey. *Future Generation Computer Systems*. https://doi.org/10.1016/j.future.2021.09.037

Ise, O. A. (2016). Integration and Analysis of Unstructured Data for Decision Making: Text Analytics Approach. *International Journal of Open Information Technologies*.

Islam, M. S., Sultana, M. S., Kumar, M. U., Mahmud, J. Al, & Islam, S. J. (2021). HARC-New Hybrid Method with Hierarchical Attention Based Bidirectional Recurrent Neural Network with Dilated Convolutional Neural Network to Recognize Multilabel Emotions from Text. *Jurnal Ilmiah Teknik Elektro Komputer Dan Informatika*. https://doi.org/10.26555/jiteki.v7i1.20550

Ito, T., Tsubouchi, K., Sakaji, H., Yamashita, T., & Izumi, K. (2020). Word-level contextual sentiment analysis with interpretability. *AAAI 2020 - 34th AAAI Conference on Artificial Intelligence*. https://doi.org/10.1609/aaai.v34i04.5845

Jagdale, R. S., Shirsat, V. S., & Deshmukh, S. N. (2019). Sentiment analysis on product reviews using machine learning techniques. *Advances in Intelligent Systems and*

Computing. https://doi.org/10.1007/978-981-13-0617-4_61

Janiesch, C., Zschech, P., & Heinrich, K. (2021). Machine learning and deep learning. *Electronic Markets*. https://doi.org/10.1007/s12525-021-00475-2

Jeong, S. Y., & Kim, Y. K. (2022). Deep Learning-Based Context-Aware Recommender System Considering Contextual Features. *Applied Sciences (Switzerland)*. https://doi.org/10.3390/app12010045

Jing, Y., Pavlović, V., & Rehg, J. M. (2008). Boosted Bayesian network classifiers. *Machine Learning*. https://doi.org/10.1007/s10994-008-5065-7

Joshi, D. J., Kale, I., Gandewar, S., Korate, O., Patwari, D., & Patil, S. (2021). Reinforcement Learning: A Survey. *Advances in Intelligent Systems and Computing*. https://doi.org/10.1007/978-981-33-4859-2_29

KABIR, A. I., AHMED, K., & KARIM, R. (2020). Word Cloud and Sentiment Analysis of Amazon Earphones Reviews with R Programming Language. *Informatica Economica*. https://doi.org/10.24818/issn14531305/24.4.2020.05

Kabir, M., Kabir, M. M. J., Xu, S., & Badhon, B. (2021). An empirical research on sentiment analysis using machine learning approaches. *International Journal of Computers and Applications*. https://doi.org/10.1080/1206212X.2019.1643584

Karthik, R. V, & Ganapathy, S. (2021). A fuzzy recommendation system for predicting the customers interests using sentiment analysis and ontology in e-commerce. *Applied Soft Computing*, *108*, 107396. https://doi.org/https://doi.org/10.1016/j.asoc.2021.107396

Kathuria, P., Sethi, P., & Negi, R. (2022). Sentiment analysis on E-commerce reviews and ratings using ML & NLP models to understand consumer behavior. *2022 International Conference on Recent Trends in Microelectronics, Automation, Computing and Communications Systems (ICMACC)*, 1–5.

https://doi.org/10.1109/ICMACC54824.2022.10093674

Kaur, G., & Sharma, A. (2023). A deep learning-based model using hybrid feature extraction approach for consumer sentiment analysis. *Journal of Big Data*. https://doi.org/10.1186/s40537-022-00680-6

Khedkar, B., & Dhande, N. (2018). *Sentiment Analysis of Customer Product Reviews Classification Techniques: A Review*. *4*(1), 178–188.

Khin Sandar Kyaw, Tepsongkroh, P., Thongkamkaew, C., & Sasha, F. (2023). Business Intelligent Framework Using Sentiment Analysis for Smart Digital Marketing in the E-Commerce Era. *Asia Social Issues*, *16*(3), e252965. https://doi.org/10.48048/asi.2023.252965

Kumar, S., Gahalawat, M., Roy, P. P., Dogra, D. P., & Kim, B. G. (2020). Exploring impact of age and gender on sentiment analysis using machine learning. *Electronics (Switzerland)*. https://doi.org/10.3390/electronics9020374

Kusal, S., Patil, S., Gupta, A., Saple, H., Jaiswal, D., Deshpande, V., & Kotecha, K. (2024). Sentiment Analysis of Product Reviews Using Deep Learning and Transformer Models: A Comparative Study. *Lecture Notes in Networks and Systems*. https://doi.org/10.1007/978-981-99-8476-3_15

Lee, J. S., & Hsiang, J. (2020). Patent claim generation by fine-tuning OpenAI GPT-2. *World Patent Information*. https://doi.org/10.1016/j.wpi.2020.101983

Lee, L. (2002). *A non-programming introduction to computer science via NLP, IR, and AI*. https://doi.org/10.3115/1118108.1118113

Lee, Y. (2023). The CNN: The Architecture Behind Artificial Intelligence Development. *Journal of Student Research, 12*. https://doi.org/10.47611/jsrhs.v12i4.5579

Li, W., Shao, W., Ji, S., & Cambria, E. (2022). BiERU: Bidirectional emotional recurrent unit for conversational sentiment analysis. *Neurocomputing*.

https://doi.org/10.1016/j.neucom.2021.09.057

Li, X., Wu, C., & Mai, F. (2019). The effect of online reviews on product sales: A joint sentiment-topic analysis. *Information & Management*, *56*(2), 172–184. https://doi.org/https://doi.org/10.1016/j.im.2018.04.007

Liu, Y., Lu, J., Yang, J., & Mao, F. (2020). Sentiment analysis for e-commerce product reviews by deep learning model of Bert-BiGRU-Softmax. *Mathematical Biosciences and Engineering*. https://doi.org/10.3934/MBE.2020398

Loukili, M., Messaoudi, F., & El Ghazi, M. (2023). Sentiment Analysis of Product Reviews for E- Commerce Recommendation based on Machine Learning. *International Journal of Advances in Soft Computing and Its Applications*, *15*, 1–13. https://doi.org/10.15849/IJASCA.230320.01

Luc Phan, L., Huynh Pham, P., Thi-Thanh Nguyen, K., Khai Huynh, S., Thi Nguyen, T., Thanh Nguyen, L., Van Huynh, T., & Van Nguyen, K. (2021). SA2SL: From Aspect-Based Sentiment Analysis to Social Listening System for Business Intelligence. In H. Qiu, C. Zhang, Z. Fei, M. Qiu, & S.-Y. Kung (Eds.), *Knowledge Science, Engineering and Management* (pp. 647–658). Springer International Publishing.

Ma, C., Luo, G., & Wang, K. (2018). Concatenated and Connected Random Forests with Multiscale Patch Driven Active Contour Model for Automated Brain Tumor Segmentation of MR Images. *IEEE Transactions on Medical Imaging*. https://doi.org/10.1109/TMI.2018.2805821

Maas, A. L., Daly, R. E., Pham, P. T., Huang, D., Ng, A. Y., & Potts, C. (2014). *Learning Word Vectors for Sentiment Analysis*. *January 2011*.

Marwat, M. I., Khan, J. A., Dahman Alshehri, M., Ali, M. A., Hizbullah, Ali, H., & Assam, M. (2022). Sentiment Analysis of Product Reviews to Identify Deceptive Rating Information in Social Media: A SentiDeceptive Approach. *KSII Transactions on*

Internet and Information Systems. https://doi.org/10.3837/tiis.2022.03.005

Matthew McMullen. (2022). *Top 5 Key Business Applications of Sentiment Analysis*. https://www.datasciencecentral.com/top-5-key-business-applications-of-sentiment-analysis/

Miller, N., Oldham, G., & Pinker, S. (1994). The Language Instinct: How the Mind Creates Language. *The Antioch Review*. https://doi.org/10.2307/4613021

Moore, C. S. (2025). Leveraging Big Data Analytics and Machine Learning Techniques for Sentiment Analysis of Amazon Product Reviews in Business Insights. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.5103205

Moreno, A., & Redondo, T. (2016). Text Analytics: the convergence of Big Data and Artificial Intelligence. *International Journal of Interactive Multimedia and Artificial Intelligence*. https://doi.org/10.9781/ijimai.2016.369

Mukherjee, P., Badr, Y., Doppalapudi, S., Srinivasan, S. M., Sangwan, R. S., & Sharma, R. (2021). Effect of Negation in Sentences on Sentiment Analysis and Polarity Detection. *Procedia Computer Science*. https://doi.org/10.1016/j.procs.2021.05.038

Mutinda, J., Mwangi, W., & Okeyo, G. (2023). Sentiment Analysis of Text Reviews Using Lexicon-Enhanced Bert Embedding (LeBERT) Model with Convolutional Neural Network. *Applied Sciences (Switzerland)*. https://doi.org/10.3390/app13031445

Mykhalchuk, T., Zatonatska, T., Dluhopolskyi, O., Zhukovska, A., Dluhopolska, T., & Liakhovych, L. (2021). Development of Recommendation System in e-Commerce using Emotional Analysis and Machine Learning Methods. *2021 11th IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS)*, *1*, 527–535. https://doi.org/10.1109/IDAACS53288.2021.9660854

Nasiri, M. S., & Shokouhyar, S. (2021). Actual consumers' response to purchase refurbished smartphones: Exploring perceived value from product reviews in online retailing. *Journal of Retailing and Consumer Services*. https://doi.org/10.1016/j.jretconser.2021.102652

Nguyen, B., Nguyen, V.-H., & Ho, T. (2021). Sentiment Analysis of Customer Feedback in Online Food Ordering Services. *Business Systems Research Journal*, *12*(2), 46–59. https://doi.org/10.2478/bsrj-2021-0018

Nguyen, H., Veluchamy, A., Diop, M., Al Khaimah Academy, R., Diop, M. L., Iqbal, R., Towers, J., Hisn Rd, A., & Al Khaimah, R. (2018). Comparative Study of Sentiment Analysis with Product Reviews Using Machine Learning and Lexicon-Based Approaches. *SMU Data Science Review*, *1*(4), 1–22. https://scholar.smu.edu/datasciencereviewAvailableat:https://scholar.smu.edu/datasciencereview/vol1/iss4/7http://digitalrepository.smu.edu.

Ni, J., Ábrego, G. H., Constant, N., Ma, J., Hall, K. B., Cer, D., & Yang, Y. (2022). Sentence-T5 (ST5): Scalable Sentence Encoders from Pre-trained Text-to-Text Models. *Proceedings of the Annual Meeting of the Association for Computational Linguistics*. https://doi.org/10.18653/v1/2022.findings-acl.146

Nicolae, D. C., Yadav, R. K., & Tufiş, D. (2022). A Lite Romanian BERT:ALR-BERT. *Computers*. https://doi.org/10.3390/computers11040057

Obiedat, R., Qaddoura, R., Al-Zoubi, A. M., Al-Qaisi, L., Harfoushi, O., Alrefai, M., & Faris, H. (2022). Sentiment Analysis of Customers' Reviews Using a Hybrid Evolutionary SVM-Based Approach in an Imbalanced Data Distribution. *IEEE Access*. https://doi.org/10.1109/ACCESS.2022.3149482

Oktaviani, V., Warsito, B., Yasin, H., Santoso, R., & Suparti. (2021). Sentiment analysis of e-commerce application in Traveloka data review on Google Play site using Naïve Bayes classifier and association method. *Journal of Physics: Conference Series*, *1943*(1), 12147. https://doi.org/10.1088/1742-6596/1943/1/012147

Oliveira, R. A. de, & Bollen, M. H. J. (2023). Deep learning for power quality. In *Electric Power Systems Research*. https://doi.org/10.1016/j.epsr.2022.108887

Ouyang, X., Zhou, P., Li, C. H., & Liu, L. (2015). Sentiment Analysis Using Convolutional Neural Network. *2015 IEEE International Conference on Computer and Information Technology; Ubiquitous Computing and Communications; Dependable, Autonomic and Secure Computing; Pervasive Intelligence and Computing*, 2359–2364. https://doi.org/10.1109/CIT/IUCC/DASC/PICOM.2015.349

Özmen, C. G., & Gündüz, S. (2025). Comparison of Machine Learning Models for Sentiment Analysis of Big Turkish Web-Based Data †. *Applied Sciences (Switzerland)*, *15*(5), 1–20. https://doi.org/10.3390/app15052297

P, S. B., Akilan, R., N, A. K., & Nithis, T. N. (2020). *Sentiment Analysis of Reviews and Ratings in Retail Shop*. 973–977.

Park, S., & Woo, J. (2019). Gender classification using sentiment analysis and deep learning in a health web forum. *Applied Sciences (Switzerland)*. https://doi.org/10.3390/app9061249

Pathak, A. R., Pandey, M., & Rautaray, S. (2021). Deep learning-based detection and prediction of trending topics from streaming data. *International Journal of Reasoning-Based Intelligent Systems*. https://doi.org/10.1504/IJRIS.2021.114631

Pham, N.-L., & Nguyen, V.-V. (2020). Adaptation in Statistical Machine Translation for Low-resource Domains in English-Vietnamese Language. *VNU Journal of Science: Computer Science and Communication Engineering*. https://doi.org/10.25073/2588-1086/vnucsce.231

Poria, S., Cambria, E., & Gelbukh, A. (2016). Aspect extraction for opinion mining with a deep convolutional neural network. *Knowledge-Based Systems*.

https://doi.org/10.1016/j.knosys.2016.06.009

Praveen, S. V., & Vajrobol, V. (2023). Understanding the Perceptions of Healthcare Researchers Regarding ChatGPT: A Study Based on Bidirectional Encoder Representation from Transformers (BERT) Sentiment Analysis and Topic Modeling. In *Annals of Biomedical Engineering*. https://doi.org/10.1007/s10439-023-03222-0

Priyadarshini, I., & Cotton, C. (2021). A novel LSTM–CNN–grid search-based deep neural network for sentiment analysis. *Journal of Supercomputing*. https://doi.org/10.1007/s11227-021-03838-w

Qi, J., Zhang, Z., Jeon, S., & Zhou, Y. (2016). Mining customer requirements from online reviews: A product improvement perspective. *Information & Management*, *53*(8), 951–963. https://doi.org/https://doi.org/10.1016/j.im.2016.06.002

Quintero Vanegas, E. O. (2023). On Albert Problem and Irreducible Modules. *Mathematics*. https://doi.org/10.3390/math11183866

Raiaan, M. A. K., Mukta, M. S. H., Fatema, K., Fahad, N. M., Sakib, S., Mim, M. M. J., Ahmad, J., Ali, M. E., & Azam, S. (2024). A Review on Large Language Models: Architectures, Applications, Taxonomies, Open Issues and Challenges. *IEEE Access*. https://doi.org/10.1109/ACCESS.2024.3365742

Ranjan, M., Tiwari, S., Sattar, A. M., & Tatkar, N. S. (2023). A New Approach for Carrying Out Sentiment Analysis of Social Media Comments Using Natural Language Processing †. *Engineering Proceedings*. https://doi.org/10.3390/engproc2023059181

Ray, P., & Chakrabarti, A. (2022). A Mixed approach of Deep Learning method and Rule-Based method to improve Aspect Level Sentiment Analysis. *Applied Computing and Informatics*. https://doi.org/10.1016/j.aci.2019.02.002

Rong, L., Weibai, Z., & Debo, H. (2021). Sentiment Analysis of Ecommerce Product Review Data Based on Deep Learning. *IMCEC 2021 - IEEE 4th Advanced*

Information Management, Communicates, Electronic and Automation Control Conference. https://doi.org/10.1109/IMCEC51613.2021.9482223

S. M. AlQahtani, A. (2021). Product Sentiment Analysis for Amazon Reviews. *International Journal of Computer Science and Information Technology*. https://doi.org/10.5121/ijcsit.2021.13302

Saini, S., Punhani, R., Bathla, R., & Shukla, V. K. (2019). Sentiment Analysis on Twitter Data using R. *2019 International Conference on Automation, Computational and Technology Management, ICACTM 2019*. https://doi.org/10.1109/ICACTM.2019.8776685

Samyuktha, A., Pallavi, M., Jagan, L., Srinivasulu, D. . Y., & Rakesh, M. M. (2023). Sentiment Analysis Using Machine Learning. *International Journal for Research in Applied Science and Engineering Technology*. https://doi.org/10.22214/ijraset.2023.48706

Sanh, V., Debut, L., Chaumond, J., & Wolf, T. (2019). *DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter*. 2–6.

Santos, B. N. Dos, Marcacini, R. M., & Rezende, S. O. (2021). Multi-Domain Aspect Extraction Using Bidirectional Encoder Representations from Transformers. *IEEE Access*. https://doi.org/10.1109/ACCESS.2021.3089099

Sarisa, M., Boddapati, V. N., Patra, G. K., & Kuraku, C. (2020). *An Effective Predicting E-Commerce Sales & Management System Based on Machine Learning Methods*. 1–11. https://doi.org/10.31586/jaibd.2020.1110

Sarker, I. H. (2021a). Data Science and Analytics: An Overview from Data-Driven Smart Computing, Decision-Making and Applications Perspective. In *SN Computer Science*. https://doi.org/10.1007/s42979-021-00765-8

Sarker, I. H. (2021b). Deep Learning: A Comprehensive Overview on Techniques, Taxonomy, Applications and Research Directions. In *SN Computer Science*. https://doi.org/10.1007/s42979-021-00815-1

Sarker, I. H. (2021c). Machine Learning: Algorithms, Real-World Applications and Research Directions. In *SN Computer Science*. https://doi.org/10.1007/s42979-021-00592-x

Sarker, I. H. (2022). AI-Based Modeling: Techniques, Applications and Research Issues Towards Automation, Intelligent and Smart Systems. *SN Computer Science*. https://doi.org/10.1007/s42979-022-01043-x

Sarker, I. H., Furhad, M. H., & Nowrozy, R. (2021). AI-Driven Cybersecurity: An Overview, Security Intelligence Modeling and Research Directions. In *SN Computer Science*. https://doi.org/10.1007/s42979-021-00557-0

Sarker, I. H., Kayes, A. S. M., Badsha, S., Alqahtani, H., Watters, P., & Ng, A. (2020). Cybersecurity data science: an overview from machine learning perspective. *Journal of Big Data*. https://doi.org/10.1186/s40537-020-00318-5

Savci, P., & Das, B. (2023). Prediction of the customers' interests using sentiment analysis in e-commerce data for comparison of Arabic, English, and Turkish languages. *Journal of King Saud University - Computer and Information Sciences*. https://doi.org/10.1016/j.jksuci.2023.02.017

Saxena, A. (2022). An Introduction to Convolutional Neural Networks. *International Journal for Research in Applied Science and Engineering Technology*. https://doi.org/10.22214/ijraset.2022.47789

Sayeed, M. S., Mohan, V., & Muthu, K. S. (2023). BERT: A Review of Applications in Sentiment Analysis. *HighTech and Innovation Journal*, *4*(2), 453–462. https://doi.org/10.28991/HIJ-2023-04-02-015

Sha, L., Rakovic, M., Whitelock-Wainwright, A., Carroll, D., Yew, V. M., Gasevic, D., &

Chen, G. (2021). Assessing Algorithmic Fairness in Automatic Classifiers of Educational Forum Posts. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. https://doi.org/10.1007/978-3-030-78292-4_31

Shah, B. K., Jaiswal, A. K., Shroff, A., Dixit, A. K., Kushwaha, O. N., & Shah, N. K. (2021). Sentiments Detection for Amazon Product Review. *2021 International Conference on Computer Communication and Informatics, ICCCI 2021*. https://doi.org/10.1109/ICCCI50826.2021.9402414

Shahzad, T., Mazhar, T., Tariq, M. U., Ahmad, W., Ouahada, K., & Hamam, H. (2025). A comprehensive review of large language models: issues and solutions in learning environments. In *Discover Sustainability* (Vol. 6, Issue 1). Springer International Publishing. https://doi.org/10.1007/s43621-025-00815-8

Siddique, Q. (2024). Comparative Analysis of Sentiment Classification Techniques on Flipkart Product Reviews : A Study Using Logistic Regression , SVC , Random Forest , and Gradient Boosting. *Journal of Digital Market and Digital Currency*, *1*(1), 21–42. https://doi.org/10.47738/jdmdc.v1i1.4

Singh, U., Saraswat, A., Azad, H. K., Abhishek, K., & Shitharth, S. (2022). Towards improving e-commerce customer review analysis for sentiment detection. *Scientific Reports*, *12*(1), 21983. https://doi.org/10.1038/s41598-022-26432-3

Skarpathiotaki, C. G., & Psannis, K. E. (2022). Cross-Industry Process Standardization for Text Analytics. *Big Data Research*. https://doi.org/10.1016/j.bdr.2021.100274

Srinivas, M., Sucharitha, G., & Matta, A. (2021). Machine learning algorithms and applications. In *Machine Learning Algorithms and Applications*. https://doi.org/10.1002/9781119769262

Straka, M., Hajič, J., & Straková, J. (2016). UDPipe: Trainable pipeline for processing

CoNLL-U files performing tokenization, morphological analysis, POS tagging and parsing. *Proceedings of the 10th International Conference on Language Resources and Evaluation, LREC 2016*.

Suhaeni, C., & Yong, H. S. (2024). Enhancing Imbalanced Sentiment Analysis: A GPT-3-Based Sentence-by-Sentence Generation Approach. *Applied Sciences (Switzerland)*, *14*(2). https://doi.org/10.3390/app14020622

Sultana, N., Kumar, P., Patra, M. R., Chandra, S., & Alam, S. K. S. (2023). Sentiment Analysis for Product Review. *International Research Journal of Modernization in Engineering Technology and Science*, *6956*(April), 1913–1919. https://doi.org/10.56726/irjmets36776

Thomas, R., & Jeba, J. R. (2024). A novel framework for an intelligent deep learning based product recommendation system using sentiment analysis (SA). *Automatika*, *65*(2), 410–424. https://doi.org/10.1080/00051144.2023.2295148

Top Popular Python Libraries in Research. (2022). *Journal of Robotics and Automation Research*. https://doi.org/10.33140/jrar.03.02.02

Tsantekidis, A., Passalis, N., & Tefas, A. (2022). Recurrent neural networks. In *Deep Learning for Robot Perception and Cognition*. https://doi.org/10.1016/B978-0-32-385787-1.00010-5

Van Houdt, G., Mosquera, C., & Nápoles, G. (2020). A review on the long short-term memory model. *Artificial Intelligence Review*, *53*(8), 5929–5955. https://doi.org/10.1007/s10462-020-09838-1

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. *Advances in Neural Information Processing Systems*.

Wang, C., Ahmad, S. F., Bani Ahmad Ayassrah, A. Y. A., Awwad, E. M., Irshad, M., Ali,

Y. A., Al-Razgan, M., Khan, Y., & Han, H. (2023). An empirical evaluation of technology acceptance model for Artificial Intelligence in E-commerce. *Heliyon*. https://doi.org/10.1016/j.heliyon.2023.e18349

Wang, X., Wang, H., Zhao, G., Liu, Z., & Wu, H. (2021). Albert over match-lstm network for intelligent questions classification in chinese. *Agronomy*. https://doi.org/10.3390/agronomy11081530

Wasif, R., & Pearl, J. (2024). *LLM-Enhanced Sentiment Analysis in E-commerce: A Deep Learning Approach to Customer Feedback*. https://doi.org/10.13140/RG.2.2.22894.42564

Wassan, S., Chen, X., Shen, T., Waqar, M., & Jhanjhi, N. (2021). *Amazon Product Sentiment Analysis using Machine Learning Techniques*. *30*, 695–703. https://doi.org/10.24205/03276716.2020.2065

Wedel, M., & Steenkamp, J.-B. E. M. (1991). A Clusterwise Regression Method for Simultaneous Fuzzy Market Structuring and Benefit Segmentation. *Journal of Marketing Research*. https://doi.org/10.1177/002224379102800401

Wedjdane, N., Khaled, R., & Okba, K. (2021). Better Decision Making with Sentiment Analysis of Amazon reviews. *Proceedings - 2021 International Conference on Information Systems and Advanced Technologies, ICISAT 2021*. https://doi.org/10.1109/ICISAT54145.2021.9678483

Wu, J., Gan, W., Chen, Z., Wan, S., & Yu, P. S. (2023). Multimodal Large Language Models: A Survey. *Proceedings - 2023 IEEE International Conference on Big Data, BigData 2023*. https://doi.org/10.1109/BigData59044.2023.10386743

Xu, G., Meng, Y., Qiu, X., Yu, Z., & Wu, X. (2019). Sentiment analysis of comment texts based on BiLSTM. *IEEE Access*. https://doi.org/10.1109/ACCESS.2019.2909919

Xu, L., Raitoharju, J., Iosifidis, A., & Gabbouj, M. (2022). Saliency-Based Multilabel

Linear Discriminant Analysis. *IEEE Transactions on Cybernetics*. https://doi.org/10.1109/TCYB.2021.3069338

Yang, L., Li, Y., Wang, J., & Sherratt, R. S. (2020). Sentiment Analysis for E-Commerce Product Reviews in Chinese Based on Sentiment Lexicon and Deep Learning. *IEEE Access*. https://doi.org/10.1109/ACCESS.2020.2969854

Yang, Y., Chen, H., Heidari, A. A., & Gandomi, A. H. (2021). Hunger games search: Visions, conception, implementation, deep analysis, perspectives, and towards performance shifts. *Expert Systems with Applications*. https://doi.org/10.1016/j.eswa.2021.114864

Zadeh, A., Chen, M., Cambria, E., Poria, S., & Morency, L. P. (2017). Tensor fusion network for multimodal sentiment analysis. *EMNLP 2017 - Conference on Empirical Methods in Natural Language Processing, Proceedings*. https://doi.org/10.18653/v1/d17-1115

Zhang, W., Li, X., Deng, Y., Bing, L., & Lam, W. (2023). A Survey on Aspect-Based Sentiment Analysis: Tasks, Methods, and Challenges. *IEEE Transactions on Knowledge and Data Engineering*. https://doi.org/10.1109/TKDE.2022.3230975

Zhang, X., Guo, F., Chen, T., Pan, L., Beliakov, G., & Wu, J. (2023). A Brief Survey of Machine Learning and Deep Learning Techniques for E-Commerce Research. *Journal of Theoretical and Applied Electronic Commerce Research*, *18*(4), 2188–2216. https://doi.org/10.3390/jtaer18040110

Zhao, H., Liu, Z., Yao, X., & Yang, Q. (2021). A machine learning-based sentiment analysis of online product reviews with a novel term weighting and feature selection approach. *Information Processing and Management*. https://doi.org/10.1016/j.ipm.2021.102656

Zhao, T., Liu, Y., Huo, G., & Zhu, X. (2019). A Deep Learning Iris Recognition Method

Based on Capsule Network Architecture. *IEEE Access*. https://doi.org/10.1109/ACCESS.2019.2911056

Zhao, W., Hu, H., Zhou, W., Shi, J., & Li, H. (2023). BEST: BERT Pre-training for Sign Language Recognition with Coupling Tokenization. *Proceedings of the 37th AAAI Conference on Artificial Intelligence, AAAI 2023*. https://doi.org/10.1609/aaai.v37i3.25470

Zheng, X., Zhang, C., & Woodland, P. C. (2021). Adapting GPT, GPT-2 and BERT Language Models for Speech Recognition. *2021 IEEE Automatic Speech Recognition and Understanding Workshop, ASRU 2021 - Proceedings*. https://doi.org/10.1109/ASRU51503.2021.9688232

Zhou, N., Yao, N., Li, Q., Zhao, J., & Zhang, Y. (2023). Multi-MCCR: Multiple models regularization for semi-supervised text classification with few labels. *Knowledge-Based Systems*. https://doi.org/10.1016/j.knosys.2023.110588

APPENDIX A:

PROPOSED CODE

#In order to handle imbalanced datasets in machine learning, this command installs the imblearn (imbalanced-learn) library. It offers methods for oversampling and undersampling to balance class distributions, such as SMOTE, ADASYN, and Tomek connections.

```
pip install imblearn
```

#In order to create word clouds from text data, this script installs the wordcloud library. With larger words denoting higher frequency, word clouds aid in visualizing the frequency of words in a dataset. It is frequently employed in text analysis and NLP jobs.

```
pip install wordcloud
```

#To handle tabular data, import pandas

```
import pandas as pnd
```

#To perform numerical calculations, import NumPy

```
import numpy as nmpy
```

#Bring in PyTorch for jobs involving deep learning

```
import torch
```

#Plotly can be imported to create interactive visuals

```
import plotly.graph_objects as plgo
```

```
#To determine how many times an element appears in a dataset, use a counter from
collections
from collections import Counter

#To create word clouds from text data, import WordCloud
from wordcloud import WordCloud

#Turn off alerts to maintain a clean output
import warnings
warnings.filterwarnings("ignore")

#Create a function that applies particular text preparation procedures to the incoming text
in order to normalize it
def normalize_text(txt):
    """
    Normalize the input text by:
    - Lowercasing
    - Removing punctuation
    - Removing extra whitespace
    - Expanding contractions (optional)
    """
    #For consistency, change the text to lowercase
    txt = txt.lower()

    #To make the text simpler, eliminate any punctuation
```

    txt = txt.translate(str.maketrans('', '', string.punctuation))

 #Once all the modifications have been applied, return the transformed text
    return txt

#'normalized_review' is a new column that is created by applying the `normalize_text` function on the 'Review' column
dfsentimnt_anlysis['normalized_review'] =
dfsentimnt_anlysis['Review'].apply(normalize_text)

#'normalized_summary' is a new column created by applying the `normalize_text` function on the 'Summary' column
dfsentimnt_anlysis['normalized_summary'] =
dfsentimnt_anlysis['Summary'].apply(normalize_text)

#To make sure the changes are accurate, look at the first few rows of the original and normalized columns ('Review', 'normalized_review', 'Summary', and 'normalized_summary')
print(dfsentimnt_anlysis[['Review', 'normalized_review', 'Summary',
'normalized_summary']].head())

#Create new columns (`lemmatized_normalized_review` and `lemmatized_normalized_summary`) to hold the results of using the `apply_lemmatization` function to the 'normalized_review' and 'normalized_summary' columns

```
dfsentimnt_anlysis['lemmatized_normalized_review'] =
dfsentimnt_anlysis['normalized_review'].apply(apply_lemmatization)
dfsentimnt_anlysis['lemmatized_normalized_summary'] =
dfsentimnt_anlysis['normalized_summary'].apply(apply_lemmatization)

#To check the changes, show the first few rows of the lemmatized and original columns
('Review', 'lemmatized_normalized_review', 'Summary', and
'lemmatized_normalized_summary')
print(dfsentimnt_anlysis[['Review', 'lemmatized_normalized_review', 'Summary',
'lemmatized_normalized_summary']].head())

#Create two new columns, 'stop_words_removed_lemmatized_normalized_review' and
'stop_words_removed_lemmatized_normalized_summary',
#by using the `remove_stop_words` function on the 'lemmatized_normalized_review' and
'lemmatized_normalized_summary' columns
dfsentimnt_anlysis['stop_words_removed_lemmatized_normalized_review'] =
dfsentimnt_anlysis['lemmatized_normalized_review'].apply(remove_stop_words)
dfsentimnt_anlysis['stop_words_removed_lemmatized_normalized_summary'] =
dfsentimnt_anlysis['lemmatized_normalized_summary'].apply(remove_stop_words)

#Verify the modifications by looking at the first few rows of the original and stopword-
removed columns ('Review', 'stop_words_removed_lemmatized_normalized_review',
'Summary', and 'stop_words_removed_lemmatized_normalized_summary')
```

```
print(dfsentimnt_anlysis[['Review',
'stop_words_removed_lemmatized_normalized_review', 'Summary',
'stop_words_removed_lemmatized_normalized_summary']].head())

#Using Plotly's Figure and Bar functions, create a bar chart where the x-axis shows the
various sentiments and the y-axis shows their counts.
fig = plgo.Figure(data=[plgo.Bar(
    x=sentimntcounts.index,
    y=sentimntcounts.values,
    marker_color=['red', 'green', 'blue'],  #To give each sentiment a particular color (such
as red, green, or blue), set the marker_color argument.
    width=0.5,  #To make the bars look cleaner, use the width argument to make them
smaller.
    text=sentimntcounts.values,  #Display count on bars
    textposition='outside'  #Using the text parameter, display the number of each sentiment
exactly above the bars. To put the count above each bar, use textposition='outside'
)])

#Designate the variable `ysentimnt_anlysis` as the target variable for sentiment analysis
and assign the 'Sentiment' column from the `dfsentimnt_anlysis` DataFrame to it.
ysentimnt_anlysis = dfsentimnt_anlysis['Sentiment']

#Verify the target variable's size and dimensions by using the `shape` attribute.
print(f"Target Shape (ysentimnt_anlysis): {ysentimnt_anlysis.shape}") #To verify the
target variable's structure and element count, print its shape.
```

#Set up a `LabelEncoder` object to translate numerical labels into categorical values.
labl_encodr = LabelEncoder()

#To convert the sentiments into numerical values, apply the `fit_transform()` method to the target variable `ysentimnt_anlysis`.
yencodedsa = labl_encodr.fit_transform(ysentimnt_anlysis)

#To clearly see the label encoding, print a dictionary that maps the original category sentiments to their matching encoded number values.
print(dict(zip(labl_encodr.classes_, labl_encodr.transform(labl_encodr.classes_))))

#Finding the greater of the two input sequence lengths from X_review_distilbert and X_summary_distilbert is the first step in determining the maximum sequence length.
maxlength = max(X_review_distilbert['input_ids'].shape[1], X_summary_distilbert['input_ids'].shape[1]) #The length of the sequence for both review and summary input IDs is provided by the .shape[1].

#Next, a for loop iterates across the keys of X_review_distilbert. For each key, it determines how much padding is required to attain the maximum length for both the review and summary datasets.
for key in X_review_distilbert:
    pad_size_review = maxlength - X_review_distilbert[key].shape[1] #The pad function is used to add padding to the sequences. For X_review_distilbert, the padding size

(pad_size_review) is computed by subtracting the current length of the sequence from the max_length.

pad_size_summary = maxlength - X_summary_distilbert[key].shape[1] #The identical technique is done on the summary dataset (pad_size_summary).

X_review_distilbert[key] = pad(X_review_distilbert[key], (0, pad_size_review)) #The pad() function is used to apply the padding, which lengthens the sequence along the second axis. This ensures that both the review and summary sequences are uniformly padded to the same maximum length.

X_summary_distilbert[key] = pad(X_summary_distilbert[key], (0, pad_size_summary))

#To apply class balancing to the dataset, call the balance_classes method.
#The resampled features and target labels will be stored in X_combined_distilbert_resampled and y_distilbert_resampled.

X_combined_distilbert_resampled, y_distilbert_resampled = balance_classes(

#Prior to balancing, the input feature set was X_combined_distilbert.
#Prior to balancing, the target labels are yencodedsa.
X_combined_distilbert, yencodedsa,

#total_samples=n: Indicates how many samples there are overall after balancing.
total_samples=n,

```
    #The input dimensionality for balancing is determined by input_dim,
    #which is taken from the shape of one of the feature arrays in X_combined_distilbert.
    input_dim = X_combined_distilbert[

        #next(iter(X_combined_distilbert)): Accesses the matching feature array
        #by retrieving the dictionary's first key.
        next(iter(X_combined_distilbert))

    ].shape[1] #shape[1]: Determines how many characteristics (dimensionality) are
present in the chosen array.
)

from transformers import AlbertForSequenceClassification, AlbertTokenizer

#Importing `AlbertTokenizer` for text preprocessing and
`AlbertForSequenceClassification` to load the saved model.

#The directory in which the tokenizer and trained model are stored is specified by
defining the variable `modlpathalbrt`.
modlpathalbrt = "./albert_sentiment_model"

#To utilize the optimized ALBERT model for inference, load it from the stored path.
modlalbrt = AlbertForSequenceClassification.from_pretrained(modlpathalbrt)

#To guarantee consistent text preprocessing, load the tokenizer from the same directory.
```

```
tokenizralbrt = AlbertTokenizer.from_pretrained(modlpathalbrt)

#Constructing a data loader for the test dataset in order to process data batches during inference in an efficient manner.
tstloadralbrt = DataLoader(tstdatasetalbrt, batch_size=16, shuffle=False)

#Setting `shuffle=False` to preserve the dataset's original order and `batch_size=16` to analyze 16 samples at a time.

#Putting the model in evaluation mode, which guarantees consistent behaviour during inference by turning off dropout layers.
modlalbrt.eval()

#Lists are initialized to hold real labels (`allabels_al`) and model predictions (`alpreds_al`) for future analysis.
alpreds_al = []
allabels_al = []

#Using `torch.no_grad()` to disable gradient computations will speed up inference because backpropagation is not required.
with torch.no_grad():
    #Processing several samples at once by batch iterating across the test dataset.
    for batch in tstloadralbrt:
```

```
    #Labels are moved to the same device to maintain uniformity throughout the assessment process.
    labels = batch['labels'].to(computedev)

    #Letting the model process the supplied data in order to produce predictions.
    modloutputs = modlalbrt(**inputsalbrt)

    #Taking the model output and extracting logits (raw scores prior to applying softmax).
    logitsval = modloutputs.logits

    #Choosing the index with the best likelihood and converting logits to projected class labels.
    predictns = torch.argmax(logitsval, dim=1).cpu().numpy()

    #Putting real and forecasted class labels in different lists for future performance analysis.
    alpreds_al.extend(predictns)
    allabels_al.extend(labels.cpu().numpy())

#Transforming lists of real labels and predictions into NumPy arrays to facilitate analysis and assessment.
alpreds_al = nmpy.array(alpreds_al)
allabels_al = nmpy.array(allabels_al)
```

#Establishing model names and the performance evaluation criteria that go along with them.

combmodls = ["DistilBERT", "ALBERT"]

#Comparing the classification performance of the two models by storing their accuracy ratings.

combaccuracy = [accdistilbert, accalbrt]

#Calculating precision scores to assess each model's capacity to reduce false positives.

combprecision = [precisiondistilbert, precalbrt]

#Storing recall scores to assess each model's capacity to find pertinent examples.

combrecall = [recdistilbert, recalbrt]

#Establishing a range of x-axis locations so that the models may correctly align the bars in the graphic.

x = nmpy.arange(len(combmodls))

#To guarantee consistent spacing and correct alignment of grouped bars, use a fixed bar width.

widthcomb = 0.2

#To guarantee a readable and clear visualization, initialize a figure and axes to a predetermined size.

fig, axs = ptlt.subplots(figsize=(8, 6))

#Plotting bars for recall, accuracy, and precision allows you to see how well both models perform.

bars1comb = axs.bar(x - widthcomb, combaccuracy, widthcomb, label="Accuracy", color="blue")

bars2comb = axs.bar(x, combprecision, widthcomb, label="Precision", color="red")

bars3comb = axs.bar(x + widthcomb, combrecall, widthcomb, label="Recall", color="green")

#Each measure is given a distinct color to help differentiate between recall (green), accuracy (blue), and precision (red).

#Retrieving height values and adding text labels to bars by iterating across each bar group.

for bars in [bars1comb, bars2comb, bars3comb]:

#To make x-axis ticks easier to see, label them with the names of the respective models.

axs.set_xticklabels(combmodls)

#Adding a legend to the bar chart to distinguish between recall, accuracy, and precision.

axs.legend()

#To make sure all metric values fit within the visualization, set the y-axis limit to 1.

axs.set_ylim(0, 1)