Evolving a Framework to Observe and Analyse Customer Experience on the Twitter Platform Using Machine Learning Techniques

Submitted in fulfilment of the requirements for the Degree of

Master of Information and Communications Technology

in the Faculty of

Accounting and Informatics

at The Durban University of Technology

By

Thaneshni Moodley

(21536061)

Date Submitted

13 August 2024

Supervisor

Professor Surendra Thakur

Declaration of Research Work Integrity

I, Thaneshni Moodley, hereby declare that this thesis is the original content of my research, has been written by me and has not been submitted for any previous degree. The experimental work is almost entirely my own work; the collaborative contributions have been indicated clearly and acknowledged with completed references. This study has not been previously submitted in any form to the Durban University of Technology or to any other institution for assessment or for any other purpose.

Student: [redacted]__ 13 August 2024

Thaneshni Moodley Date

Approval for final submission

Supervisor: [redacted] 13 August 2024

Professor Surendra Thakur Date

Acknowledgement

I would like to express my gratitude and appreciation to my supervisors, Professor Surendra Thakur and Dr Alveen Singh, whose guidance, support and encouragement has been invaluable throughout this study. Thank you, Professor Thakur and Dr Singh, for guiding me to this important publication, for the stimulating questions, the meetings and conversations were vital in inspiring me to reach my potential. My biggest gratitude goes to my amazing mum for her continuous love and support that she has given me through this journey.

Table of Contents

Declaration of Research Work Integrity i
Acknowledgement ii
Table of Contents iii
List of Tables vii
List of Figures viii
Abstract ix
List of Acronyms x
Output xi
Chapter 1 Introduction 1
1.1 Introduction 1
1.2 Research Background 1
1.3 Research Problem 2
1.4 Research Aim, Question and Objectives 4
1.5 Structure of the Dissertation 5
Chapter 2 Literature Review 6
2.1 Introduction 6
2.2 A definition of an online consumer complaint 6
2.3 Customer Protection Act and National Consumer Commission 6
2.4 Social media platforms 7
2.4.1 Facebook 7
2.4.2 LinkedIn 8
2.4.3 Twitter 8
2.5 Factors of information credibility on social media platform 9
2.6 The emergence of Twitter as a popular medium for customer complaints 10
2.7 Twitter hashtags, mentions and retweets 11
2.8 Retail complaints on social media 12
2.8.1 Customer complaints 12
2.8.2 Reasons for customers to complain on social media 13
2.8.3 Customer complaints management 14
2.8.4 Business benefit of customer complaint handling 14
2.8.5 Ineffective post-sale services 15

2.8.6 Benefits of effective post sale services and why this research is needed 16
2.9 Social analytic tools 19
2.10 Performance Metrics 20
2.11 Feature Extraction 20
2.12 Exploratory data analysis 21
2.13 Natural language Processing 22
2.14 Natural language pre-processing methods 22
2.15 Sentiment analysis 23
2.16 Approaches to sentiment analysis 25
2.17 The use of sentiment lexicon 26
2.18 Reasons for sentiment analysis application in social media 27
2.19 Supervised machine learning techniques 28
2.20 Unsupervised and weakly supervised machine learning techniques 29
2.20.1 Unsupervised learning 29
2.20.2 Weakly supervised learning 30
2.21 Ensemble machine learning techniques 31
2.22 Deep learning 31
2.23 Sentiment Analysis ML Approach 33
2.24 Ensemble Machine Learning for Sentiment Analysis 33
2.25 Current challenges with sentiment analysis 39
2.26 Identifying barriers to data analytics adoption in retail 42
2.27 Conclusion 43
Chapter 3 Research Methodology 44
3.1 Introduction 44
3.2 Research paradigm 44
3.3 Research methodology and time horizon 45
3.4 Data Source 45
3.4.1 Data Collection 45
3.4.2 Data Preparation 49
3.5 Conceptual Framework 51
3.5.1 Data Extraction 53
3.5.2 Pre-processing Techniques for NLP 53
3.5.3 Feature Extraction 59

3.5.4 Sentiment Analysis 61
3.5.5 Model Evaluation and Visualisation 74
3.5.5 Selection of ML technique 76
3.6 Describing the Experiment Design 77
3.6.5 Sentiment analysis tools 77
3.6.6 Mathematical Description of Algorithms and Techniques 78
3.6.7 Evaluation of Results 81
3.6.8 Application of Algorithms to Twitter Customer Experience Data 82
3.7 Conclusion 82
Chapter 4 Results and Discussion 83
4.1 Introduction 83
4.2 Experiment's Hardware and Software 83
4.3 Data Preparation 84
4.3.1 Exploratory Data Analysis and Missing values 84
4.4 Conceptual Framework 97
4.4.1 Data Extraction 97
4.4.2 Pre-processing Techniques for NLP 97
4.4.3 Feature Extraction 98
4.4.4 Sentiment Analysis 99
4.4.5 Model evaluation and Visualisation 104
4.5 Selection of ML technique 106
4.6 Conclusion 107
Chapter 5 Conclusion and Recommendation 108
5.1 Introduction 108
5.2 Summary of the Study 108
5.3 Research Objective 109
5.3.1 Research Objective 1: To investigate the characteristics for customer complaints on Twitter 109
5.3.2 Research Objective 2: To conduct a comparative performance analysis of ML algorithms used for customer sentiment analysis 109
5.3.3 Research Objective 3: To develop an ML-based customer sentiment analysis model incorporating the characteristic of customer complaints on Twitter to determine best performing ML algorithms 110
5.4 Research Question: What ML technique can be used for accurate identification of customer complaints in Twitter data? 110

5.5 Limitation 111
5.6 Recommendation for Future Research 111
5.7 Conclusion 111
References 112
Appendix 128
Data Extraction 128
Pre-processing 129
Open-source sentiment analysis tool 129
Performance Metrics for Model Evaluation 130
Visualisation 131

List of Tables

Table 2.1: The Effectiveness of Post-Sale Services 18
Table 2.2: Recent ML techniques used in sentiment analysis 37
Table 2.3: Summary of studies that implemented supervised, unsupervised and deep learning algorithms 40
Table 3.1: Description of the dataset's columns 47
Table 3.2: Amount purchased and the usage of data 48
Table 3.3: Example of output for Sentiment Analysis 64
Table 3.4: Architectural and Hyperparameter Configurations of the Supervised Deep Machine Learning Models 66
Table 3.5: Architectural and Hyperparameter Configurations of the Unsupervised Machine Learning Models 68
Table 3.6: Summary of technological tools for the methodology 70
Table 3.7: Select ML algorithm for each ML technique 74
Table 4.1: Range retail complaints per province 97
Table 4.2: Analysis on File Size and Data processing after Data has been pre-processing 98
Table 4.3: Sentiment Analysis Results 99
Table 4.4: Model evaluation results 105

List of Figures

Figure 2.1: Annotated taxonomy of a tweet 12
Figure 2.2: Sentiment Analysis ML Algorithms 33
Figure 2.3: Performance of ML Algorithms (Sadhasivam and Kalivaradhan 2019) 35
Figure 2.4: Performance of Ensemble Models (Sadhasivam and Kalivaradhan 2019) 36
Figure 2.5: Comparison of ML Algorithms and Ensemble Models (Sadhasivam and Kalivaradhan 2019) 36
Figure 3.1: Selected EDA Methods 49
Figure 3.2: Proposed Conceptual Framework 52
Figure 3.3: Pre-processing Techniques for NLP 54
Figure 3.4: Feature Extraction 60
Figure 3.5: The ML techniques required to be trained for sentiment analysis 65
Figure 4.1: Shape of the dataset 84
Figure 4.2: Columns of the dataset 84
Figure 4.3: Null Values 85
Figure 4.4: Updated Table 86
Figure 4.5: Bivariate analysis visualisation on Limpopo 87
Figure 4.6: Bivariate analysis visualisation on Mpumalanga 88
Figure 4.7: Bivariate analysis visualisation on Free State 89
Figure 4.8: Bivariate analysis visualisation on Eastern Cape 90
Figure 4.9: Bivariate analysis visualisation on KZN 91
Figure 4.10: Bivariate analysis visualisation on Gauteng 92
Figure 4.11: Bivariate analysis visualisation on Western Cape 93
Figure 4.12: Bivariate analysis visualisation on Northern Cape 94
Figure 4.13: Bivariate analysis visualisation on North West 95
Figure 4.14: Retail complaints/comment per province bar graph 96
Figure 4.15: Word2Vec Results 98
Figure 4.16: Accuracy Result for Supervised Learning 100
Figure 4.17: Accuracy Result for Deep Learning 101
Figure 4.18: SVM Model that has been built 101
Figure 4.19: Artificial Neural Network model that has been built 102
Figure 4.20: Decision Tree model that has been built 103
Figure 4.21: K-Means model that was built 103
Figure 4.22: Fuzzy C Model that was built 104
Figure 4.23: Multilayer Perceptron Neural Network Model that was built 104
Figure 4.24: Model Evaluation Results 105
Figure 4.25: Visualisation Model Evaluation Diagram Displayed in PyCharm 106

Abstract

Retailers have become more focused on retaining and turning existing customers into long-term clients because retailers have become more competitive, customers more demanding, and competitors more aggressive. The 2020 COVID-19 pandemic has forced a transformation for retailers. Within months, a revolution has taken place, constituting major changes to how consumers view cash, how they shop online and what they expect from retailers as part of a positive buying experience. Consumers increasingly expect retailers to create a seamless customer experience. This often means leaning on digital capabilities to create a seamless, omni-channel experience by linking different aspects of the customer shopping experience. The usage of big data analytics has primarily been implemented outside of South Africa to better understand customer connections and experiences, highlighting a noticeable research gap in South Africa. It has been proven to be an effective tool for retailers in predicting customer behaviour. There is a need to reduce the complexities in understanding which are the most appropriate machine learning techniques for sentiment analysis of online customer experience and to capitalise on development. Thereafter, online retailers are better equipped to tailor machine learning tools to craft analytical tools. Given the massive migration to online transactions, this work presents a rigorous analysis of social media posts, which is paramount for modern-era retailers. Businesses can use sentiment analysis to determine how well their brand is performing in the marketplace, learn more about the attitudes of their customers and determine whether their items receive more positive or negative feedback. A longitudinal study was adopted to analyse a dataset of retail-related tweets for the identification of customer complaints using a sentiment analysis hybrid approach, which is a combination of lexicon and machine learning approaches. A conceptual framework was developed to observe and analyse customer experiences on the Twitter platform using machine learning techniques. The framework encompasses components such as data preparation, natural language processing pre-processing techniques, calculating sentiment using sentiment lexicon and ML techniques, and thereafter a selection of the best-performing machine learning technique for sentiment analysis within the developed conceptual framework. The extracted dataset contains 240 000 tweets posted between 01 January 2017 and 31 January 2019, out of which 27 233 tweets were selected for the study. Natural language pre-processing techniques were applied to the dataset, including tokenisation, stemming, lemmatisation, part-of-speech tagging, and name-of-entity recognition. Supervised and deep machine learning gave the best results of 61.75 and 60.25. This study has identified deep learning as a good technique for sentiment analysis when NLP pre-processing methods are done in a certain order. A study on analysing retail complaints posted on the Twitter platform using a sentiment analytic framework has not been done in South Africa before. This study has proven that the sentiment analysis hybrid approach is highly capable of analysing social media data.

List of Acronyms

CNN	Convolutional Neural Network
AI	Artificial Intelligence
ANN	Artificial Neural Network
ASA	Advertising Standard Authority
BoW	Bag of Word
CNN	Convolutional Neural Network
COVID-19	Coronavirus disease 2019
CPA	Customer Protection Act
DNN	Deep Neural Network
EDA	Exploratory Data Analysis
GUI	Graphical User Interface
ICT	Information and Communication Technology
IDA	Initial Data Analysis
LSTM	Long Short-Term Memory
ML	Machine learning
MLP	Multilayer Perceptron
NLP	Natural Language Processing
NCC	National Consumer Commission
NCT	National Consumer Tribunal
NCT	National Consumer Tribunal
NER	Named Entity Recognition
NLTK	Natural Language Toolkit
POS-Tag	Part-of-speech tagging
RNN	Recurrent Neural Network
SDL	Service Dominant Logic
SVM	Support Vector Machine
TF-IDF	Term Frequency and Inverse Document Frequency

Output

Journal papers submitted arising from this study (Under review)

[Enter]

Conferences arising from this study

Moodley, T. and Thakur, S. 2023. Evolving a framework to observe and analyse customer experience on the Twitter platform using Machine Learning Techniques. In: *2023 National Electronic Media Institute of South Africa (NEMISA)*. South Africa, 15-17 Febuary 2023.

Presentations arising from this study

[Enter]

Recognition arising from this study

[Enter]

Newspaper articles arising from this study

[Enter]

Chapter 1 Introduction

1.1 Introduction

This study involves the development of a conceptual framework to improve the identification of customer complaints on Twitter using a hybrid approach that combines lexicon-based and machine learning (ML) techniques for sentiment analysis. The conceptual framework begins with extracting the required data from Twitter, followed by pre-processing methods to prepare the data for sentiment analysis using first the lexicon-based approach and then ML techniques. Furthermore, the conceptual framework involves the evaluation of the best-performing ML technique for sentiment analysis.

This chapter begins with a brief introduction of the dissertation (Section 1.1), followed by the relationship between the data analytic framework and ML (Section 1.2) and then the current issues with sentiment analysis (Section 1.3). Thereafter, the explanation of the research problem (Section 1.4), the research aim and objectives (Section 1.5) and then the research question (1.6). Finally, Chapter 1 ends with the structure of the dissertation (1.7).

1.2 Research Background

Data analytics, or data value chain, refers to the process of analysing data to extract usable information and insights. Data analytics include all the processes and technologies required for knowledge discovery, such as data extraction, transformation, loading, and analysis, as well as strategies and methods for delivering results to decision-makers. Data analytics, often called data analysis, is the process of cleaning, examining, modelling, and changing data in order to extract useful information, draw conclusions, and improve decision-making (Carvalho 2018; Blum, Hopcroft and Kannan 2020).

Machine learning is the process of extracting data, learning from the data, and forecasting future trends for a certain topic using algorithms. An ML algorithm learns from data and applies what it has learned without the need for human involvement. Machine learning is a branch of artificial intelligence in which computers are trained to learn on their own. These computers can learn in the same manner that people do, upgrading their skills as new data is introduced (Mahesh 2020; Kundu *et al.* 2022).

Machine learning focuses on constructing and training algorithms through data so that ML can function independently, whereas data analytics focuses on using data to generate insights. Large businesses can benefit from ML techniques for data collection, analysis, and integration. It can be used for data labelling and segmentation, data analytics, and scenario simulation, among other aspects of big data operations (Lu *et al.* 2022; Srivastava, Bharti and Verma 2022).

Customer interaction with a brand or retail establishment is referred to as the customer experience, which covers service provision from product selection to post-service assistance. The result of every interaction a consumer has with a company, be it via the website, chatting with customer support agents, or getting the products or services they request, is known as the customer experience. At its base, digital retail also known as online retail, is a collection of interconnected web or app experiences that allow customers to engage with a brand, learn more about it, study its products and, depending on what it sells, make purchases (Huang *et al.* 2021; Zaid and PATWAYATI 2021; Moore, Bulmer and Elms 2022).

The phrase "online customer experience" refers to the description of all interactions, thoughts and feelings between a consumer and a brand as perceived from the customer's perspective (Molinillo *et al.* 2022; Moore, Bulmer and Elms 2022). Customers' experiences have a significant influence on their purchasing decisions, which suggests that customers would likely determine not to make a purchase from a business, whether it be an online or brick-and-mortar store that offers a bad experience (Silva e Sousa, Pinho and Simões 2022).

Sentiment analysis is a natural language processing (NLP) technique for determining whether data is perceived as positive, negative, or Neural. It is sometimes referred to as emotional analysis or opinion mining (Aggarwal 2022). Sentiment analysis is frequently used by businesses to track how their brands and products are perceived by consumers in online reviews and to better understand their target market. The growth of a brand for an online retailer can benefit from the application of sentiment analysis. Businesses can use sentiment analysis to assess the market performance of their brand, learn more about customer attitudes, and evaluate whether their products receive more positive or negative comments (Aggarwal 2022; Kumar *et al.* 2022).

The primary challenges with sentiment analysis approaches include the ability to function well across domains, the lack of labelled data, which results in inaccurate sentiment analysis, and the inability to handle complex phrases which call for far more than sentiment terms (Alsaeedi and Khan 2019). Since machines need to be trained to analyse and understand emotions in the same manner that the human brain does, sentiment analysis can be difficult in NLP. This is in addition to understanding the differences between other languages. Tools for sentiment analysis will be more effective at addressing these issues as data science develops (Basarslan and Kayaalp 2020; Carrera-Rivera, Larrinaga and Lasa 2022).

1.3 Research Problem

Retail has become competitive, customers have become aware and demanding, competitors have become aggressive, and engagement rules have been more stringent (Lahti 2018; Jumaryadi 2019). All of the outcomes mentioned have compelled online retailers to focus more on retaining and turning existing consumers into long-term clientele (Blum, Hopcroft and Kannan 2020). Successfully dealing with customer complaints improves customer retention rates and nurtures customer loyalty. Furthermore, client complaints are among the most useful sources of information regarding customer experience. The area of concern is the current approaches' incapacity to deliver credible interpretations of customer opinion expressed on social media (COVID, Team and Hay 2020; Da Veiga and Ophoff 2021).

South African online retailers are having a problem identifying a use case to justify the investment in big data analytics. The reasons are that locating and attracting employees with the appropriate abilities to take advantage of large datasets is a challenge for retail organisations, resulting in South African retailers slow adoption of the usage of big data analytics (Donders *et al.* 2020; Ridge, Johnston and O'Donovan 2021).

The use of big data analytics to better understand customer connections and experiences has primarily been implemented outside of South Africa, creating a research gap in the country (Akter and Wamba 2016; Kauffmann *et al., 2*020; Faghani *et al., 2*022). It has been shown to be an effective tool for retailers in predicting customer behaviour (Pantano, Giglio and Dennis 2019). There is a need to reduce the complexities in understanding which are the most appropriate ML algorithms for sentiment analysis of online customer experience and to capitalise on development. Thereafter, online retailers are better equipped to tailor ML tools to craft analytical tools. This dissertation positions effective analysis of social media posts, which are neither temporally nor geographically limited, as paramount for modern-era retailers, given the massive migration to online transactions.

On the usage of text mining and NLP, sentiment analysis uncovers and extracts subjective information from the text. However, there are numerous issues with the sentiment analysis and evaluation process that make it challenging to correctly understand sentiments and select the appropriate sentiment polarity. Sarcasm, irony, informal writing style, and linguistic issues are some of the elements that make sentiment analysis and NLP difficult (Birjali, Kasri and Beni-Hssane 2021; Jindal and Aron 2021; Wankhade, Rao and Kulkarni 2022).

Depending on the context, many words across several languages have different meanings and orientations. As a result, accessibility of tools and information for all languages is limited. Irony and sarcasm are two of the most important issues that have recently captured the interest of researchers. The ability to recognise irony and sarcasm in writing has advanced significantly. According to research, sentiment analysis methods such as lexicon-based, ML and transformer learning techniques have improved in the analysis of irony and sarcasm text (Chen, Lee and Chen 2020; Birjali, Kasri and Beni-Hssane 2021; Gowri *et al.* 2021; Jindal and Aron 2021).

Lexicon-Based Approach, ML Approach, and Hybrid Approach are the three most often used approaches for sentiment analysis. The lexicon-based approach has the advantage of being considered an unsupervised method and not requiring any training data. Lexicon-based approaches have the flaw of being domain-specific, meaning that concepts from one domain cannot be used in another (Mohamad Sham and Mohamed 2022). The Lexicon – based approach can use SentiWordNet for sentiment analysis which is a lexical resource that requires engagement with WordNet database as its operating system which measures the positivity, negativity, or Neutrality needed for sentiment analysis (Maudslay *et al.* 2023; Sridevi and Velmurugan 2023; Sudirman and Nugraha 2023).

ML approach for sentiment analysis begins with a pretraining phase where it learns from labelled data. A model is built using supervised or weakly supervised learning ML with an extensive training resource, with supervised models being the most common. The capacity of ML approaches to represent several features while simultaneously capturing context, the adaptability to changing input, and the ability to quantify the level of uncertainty involved in categorisation all contribute to their increasing popularity (Balaji, Annavarapu and Bablani 2021; Bibi *et al.* 2022).

Hybrid sentiment analysis refers to a combination of lexicon-based and ML approaches. The widely used hybrid approach combines the two and is based on sentiment lexicons, which are found in most systems (Birjali, Kasri and Beni-Hssane 2021). This study has selected the hybrid approach for sentiment analysis.

There are two ML techniques used for sentiment analysis which are supervised ML technique and unsupervised ML technique. Both supervised and unsupervised ML techniques can be applied to a hybrid approach. Unsupervised approaches for sentiment analysis are possible when using knowledge-based ontologies, databases, and lexicons that contain selected comprehensive knowledge. Supervised learning techniques provides the ability to assess a wide range of topics and effectiveness in identifying the issue of subjectively. Unsupervised learning technique does not require trained data, does not require human manipulation and provides productive results in the presence of ambiguity (Beigi and Moattar 2018; Srivastava, Bharti and Verma 2022).

ML has evolved, and deep learning is one of the evolutions. It links together algorithms to simulate how the human brain functions, also known as an artificial neural network, and has made it possible for many useful ML applications. Deep learning, which uses many algorithms in a sequential chain of events to solve complicated problems, is a hierarchical ML that enables the processing of enormous volumes of data accurately and with a minimum of human input (Dashtipour *et al.* 2021; Soong, Ayyasamy and Akbar 2021).

Sentiment analysis is now able to work with large amount of unstructured data that is in existence within social media and provide real – time analysis for business processes(Adak, Pradhan and Shukla 2022). Deep learning has emerged with a revolutionary paradigm and an extensive data driven representation learning architectures to work with unstructured data (Rodrigues *et al.* 2022). Deep learning has recently been introduced for sentiment analysis; therefore, this research has selected deep learning technique to test the sentiment analysis results and compare to frequently used ML techniques which are supervised and unsupervised ML techniques.

1.4 Research Aim, Question and Objectives

This research aims to determine the most suitable ML technique for sentiment analysis of Twitter data. This is to enhance the analysis and interpretation of online customer complaints on Twitter.

The main research question is:

How a ML technique can be used for accurate identification of customer complaints in Twitter data?

To answer the research question, the following research objectives were set:

Research objective 1: To investigate the characteristics of customer complaints on Twitter.

Research objective 2: To conduct a comparative performance analysis of ML algorithms used for customer sentiment analysis.

Research objective 3: To develop and evaluate an ML-based customer sentiment analysis model incorporating the characteristics of customer complaints on Twitter to determine the best-performing ML algorithms.

1.5 Structure of the Dissertation

The dissertation is organised as follows:

Chapter 1 presents the research background, research problem, aim, questions and objectives.

Chapter 2 provides a comprehensive overview of customer complaints and the various social media platforms, including a breakdown of Twitter hashtags, mentions and retweets, followed by the business benefits of customer complaint handling and a discussion on social analytic tools that can be used in the methodology. The chapter ends with the identification of suitable ML techniques to use in sentiment analysis.

Chapter 3 gives a thorough description of the research design and methodology. It contains a thorough explanation of the data sources and the methods used to manage the data for the study. It further details how the conceptual framework's eight stages were applied to this study.

Chapter 4 presents the discussion of the experiment.

Finally, Chapter 5 provides a summary of how the research has achieved its objectives, identifies gaps, limitations, and future areas for research before concluding the study.

Chapter 2 Literature Review

2.1 Introduction

This chapter begins with a brief definition of online consumer complaints (Section 2.2), followed by the Customer Protection Act (CPA) and National Consumer Commissions (Section 2.3), together with their relevance to this study. This is then followed by a description of social media platforms (Section 2.4), which includes Facebook, LinkedIn and Twitter. Thereafter, factors on information credibility on social media platforms (Section 2.5) and the emergence of Twitter as a popular medium for customer complaints (Section 2.6), followed by a short description of hashtags, mentions and retweets and why the tweets are of concern to this study, are presented. Twitter hashtags, mentions and retweets (Section 2.7), followed by retail complaints on social media (Section 2.8), are discussed in further detail. This will be followed by a comparison of social analytic tools (Section 2.9); thereafter, a discussion on NLP is presented (Section 2.10). Section 2.11 introduces and defines sentiment analysis; Section 2.12 elaborates on the approaches to sentiment analysis and Section 2.13 explores use cases for sentiment analysis. This chapter ends with an evaluation of candidate ML algorithms to identify post-sale complaints found on Twitter (Section 2.15) and identifies barriers to data science adoption in retail (Section 2.18).

2.2 A definition of an online consumer complaint

Consumer complaints are the subject of contradictions between what a business promises in terms of a product or service and what the client obtains. Customers' perceptions of the brand are not in sync with the degree of customer care they actually receive (Shin and Larson 2020). Customer complaints can come in a variety of forms and causes, such as ineffective communication, inefficient internal processes, or poor service quality (Jumaryadi 2019).

In the context of this dissertation, a customer complaint is one that appears on a social media platform voicing the experience of interaction, quality, or after-sales provisioning having paid for a service or product. The complaint has unrestricted viewing access and the ability to trigger further customer experiences from other social media users (Adak, Pradhan and Shukla 2022).

2.3 Customer Protection Act and National Consumer Commission

The Customer Protection Act (CPA) outlines the consumer rights that South African consumers are entitled to, as its name suggests. As an example, consumers have the right to safe and high-quality goods, to be fully informed about a product or service, to not be misled by misleading information or inaccurate marketing, and to reasonable terms and conditions for acquiring goods and services (Goga and Paelo 2019). If a product or service is substandard or defective, customers have the right to seek restitution in the form of a refund, repair, or replacement (Da Veiga and Ophoff 2021).

Consumers frequently believe that filing a complaint with a well-known regulatory body, such as the National Consumer Commission (NCC), is the best option to remedy a problem. While the NCC is unquestionably a strong tool, there is a procedure that must be followed and other avenues that must be explored before contacting any tribunal or ombudsman (Goga and Paelo 2019). The NCC helps consumers have their complaints resolved when suppliers breach their rights. Consumer complaints regarding goods and services are handled by the NCC, not credit agreements (Struwig 2018).

The NCC will examine complaints and recommend cases to the National Consumer Tribunal (NCT) for adjudication based on all available information and evidence (Struwig 2018). The NCC encourages concerns to be resolved amicably rather than intervening actively in disagreements. The Consumer Tribunal has the power to find the CPA to be in violation, compel practice adjustments, levy administrative fines, and prohibit future prohibited practices (Da Veiga and Ophoff 2021).

In South Africa, social media influencers and reviewers do not have precise guidelines on how to post on social media as a brand ambassador or when evaluating a product (Goga and Paelo 2019). Advertising must be fair and reasonable, according to the CPA. Marketing cannot imply a deceptive or fraudulent depiction of the items or services. Being an influencer and reviewing a product or experience requires truthfulness and honesty, according to the Advertising Standard Authority (ASA) Code (Struwig 2018).

2.4 Social media platforms

Social media can be defined as online tools, media and platforms that encourage communication, teamwork and content sharing (Shearer and Mitchell 2021). Social networks are generally used for peer group or community contact and the exchange of ideas that are of interest (Zayani 2021).

Social media is an instrument for both business communication and personal profile creation for millions of people. Social media enables businesses to communicate with their customers. Research has shown that most businesses prefer to direct their customers to their social media pages more than their traditional web pages, as this enables them to reach the majority of their customers (Duffy and Chan 2019; Munthali *et al.* 2021). A social media network allows a user to build personal websites accessible to other users in exchange for personal content and communication. (Shearer and Mitchell 2021) estimated that about 1.7 billion people use social media to receive or send messages daily.

Currently, different kinds of social media networking services have emerged, and there are several social media sites. The most popular social networking platforms that are widely used for business are Facebook, Twitter and LinkedIn (Duffy and Chan 2019; Shearer and Mitchell 2021).

2.4.1 Facebook

Facebook is a social networking site that offers users a platform for communication, the ability to share photos and connect with friends who also use the site, and access to the internet. Over 800 million people are reportedly active Facebook users (Wilson, Gosling and Graham 2021). Facebook's paid campaigns and advertisements can be used to target certain audiences. The website collects a lot of data on its users, which is helpful for focusing on advertisements. Another effective marketing tactic for small businesses is creating a Facebook business page (Klassen *et al.* 2018).

2.4.2 LinkedIn

Florenthal (2015) explained that LinkedIn was created as a platform for professionals to share knowledge and insight in more than one million LinkedIn groups. On LinkedIn, companies have access to a wealth of information that is mostly provided by users through their profile data, which includes the company name, job title and size of the company. The profile data is used for advertising targeted towards LinkedIn members or affiliation groups on LinkedIn. Through this, the company has been able to increase their brand awareness among target market segments (Brewer 2018; Davis *et al.* 2020).

LinkedIn is a social networking site that allows users to interact with co-workers, clients, potential customers, strategic partners, and other professionals. LinkedIn is mostly used for business-to-business networking and increasing website traffic (Geyik, Ambler and Kenthapadi 2019). Consumers can choose to follow retailers and learn everything retailers have to say. Retailers can obtain ideas, discuss implementation tactics for using LinkedIn for business promotion, and strategise by seeking advice from reputable individuals and business role models (Brewer 2018; Davis *et al.* 2020).

2.4.3 Twitter

Twitter is considered a miniature blog where information is posted to keep people informed. Unlike Facebook, where one can have friends to share different things, with Twitter, one must get connected to the latest information on what consumers find interesting (Kursuncu *et al.* 2019). One must find a public stream that interests them and follow conversations. A person can still follow tweets regardless of whether that person does not tweet at all, and there is also no limit as to how many tweets one can send within a given day (Allem *et al.* 2018; Murthy 2018).

Twitter has helped lift brands, enhance customer relationship marketing and improve direct sales. This is because Twitter allows customers to quickly share their sentiments about products that they have purchased. An organisation that adopts Twitter should have the ability to respond quickly to customers' needs and complaints (Alsaeedi and Khan 2019; Kühl *et al.* 2019).

A methodology for categorising well-known Twitter threads as real or false news has been presented in earlier studies. This model used four procedures, including picking which characteristics to utilise, aligning the three datasets, predicting accuracy through "featuring," and assessing the final models (Pershad et al., 2018; Karami et al., 2020). Four different feature types—structural, user, content, and temporal—were applied to predict the classification accuracy. Properties like tweet volume and activity distributions (e.g., percentages of retweets or media shares) are examples of Twitter's structural features. Twitter verified status, friend/follower numbers, account ages, and tweet author(s) are examples of user features.

Measures of content include things like the polarity, subjectivity, and agreement of tweet text. Finally, temporal characteristics record changes over time in the preceding features, such as the average author age or the downward trend of the number of tweets (Karami *et al.* 2020).

However, the challenge of the model is the structural differences that affect model transfer. This shows that discrepancies in feature sets or cross-context performance could be attributed to structural problems rather than actual model capabilities if the underlying distributions that generated the samples diverge significantly (Karami *et al.* 2020).

The Twitter API enables businesses to carry out complex searches like retrieving all recent tweets about a specific topic or extracting users' non-retweeted tweets. This is a straightforward application in the analysis of how the public perceives the company. This might also be used to map the regions of the world where the company has received the most mentions (Wojcik and Hughes 2019). There were also public companies that scraped Twitter according to user supplied information. However, Karami et al. (2020) advised examining Twitter's data policy.

2.5 Factors of information credibility on social media platform

Research has indicated that several factors that can be observed in the social media platforms that are useful to establish information credibility include :

I. The responses that particular topics elicit and the emotions shown by users when it is debated, for instance, if they employ opinion phrases to indicate positive or negative thoughts about the subject.

II. The degree to which users are certain about the information that is circulating, for instance, whether or not users question the information that is provided to them.

III. The external sources cited, such as if the user cites a specific URL with the information they are propagating and whether or not that source is well-known.

IV. User characteristics that help spread the information, such as how many followers each user has on the site (Alsaeedi and Khan 2019; Kursuncu *et al.* 2019; Hswen *et al.* 2021).

These characteristics were utilised to identify features, which were then used to model subjects and the information cascades connected to them. Although most of these features are rather generic and may be used on other social media platforms, some of them are unique to the Twitter platform. According to Duffy and Chan (2019), there are four broad categories that these traits fall under.

I. **Message-based features:** These features, which are independent of Twitter or dependent on it, take message attributes into consideration. The length of a message, whether the text contains exclamation marks or question marks, and the quantity of positive or negative emotion words in a message are all elements that are independent of Twitter. Features that depend on Twitter include things like whether a tweet contains a hashtag and whether it is being retweeted.

II. **User-based features:** These consider information about the users who submit messages, such as their age at registration, how many people they follow (or "friends" in Twitter), and how many tweets they have previously written.

III. **Topic-based features:** These aggregates, which are calculated from the first two feature sets, include the percentage of tweets with URLs, the percentage of tweets with hashtags, and the percentage of tweets with positive and negative sentiment in a set.

IV. **Propagation-based features:** These take into account traits connected to the propagation tree that can be created from a message's retweets. These include elements like the depth of the retweet tree or the quantity of the topic's first tweets.

2.6 The emergence of Twitter as a popular medium for customer complaints

Twitter is used with regards to customer service by retrieving concerns addressed and questions answered, and on the other hand, it is also about being thankful for excellent service or receiving assistance with a critical issue (Wedel, Palk and Voß 2021). Twitter transformed the nature of customer service as well as enabling companies and customers to engage quickly (Okazaki *et al.* 2020b). There is no evidence to suggest that customers post only in a specific negative or positive manner as customers share their experiences on social media (Singh *et al.* 2021).

The retail industry's online customer base mostly consists of millennials. Millennials are between the ages of 18 and 34 who entered adulthood around the year 2000. Therefore, there is a link between the social median age and retail in South Africa (Ibrahim and Wang 2019).

Users believe Twitter is the best place to address customer care issues with brands. Not just with brands, but also with other customers who may be having similar problems. It is an illustration of how Twitter is changing customer service from a one-to-one relationship to a one-to-many relationship interaction (Singh *et al.* 2021).

Firms from many industries utilise Twitter for customer service, and that is what makes it so popular, from finance to retail, tourism, and telecommunications. Some industries, on the other hand, are more active than others (Wedel, Palk and Voß 2021). Users are most likely to utilise Twitter as a customer support channel for retail and travel. According to research, 40% of individuals who had recently utilised the platform for customer support had done so for retail, 33% for travel, and 28% for telecommunications (Okazaki *et al.* 2020b).

Users want brands to respond promptly when it comes to customer service. According to research, 24% of users consider response speed as the most critical attribute for customer service on Twitter (Ibrahim and Wang 2019). In particular, 71% of Twitter users want a business to answer to their question within an hour of sending it, which for many businesses, is a problem. According to research, 63% of users who tweeted a company concerning customer service received a response within two days. Overall, 60% of individuals who used Twitter for customer service received a response, and 40% of those who used Twitter for customer service were satisfied with their experience (Singh *et al.* 2021).

It has been suggested by Fox and Cowley (2018) that retailers should find value in responding to complaints found on Twitter so followers can see the way that the brand addresses things. Most of the time, the worst thing retailers will be able to do is ignore it. Retailers that fail to respond to legitimate complaints or do so in an inauthentic or dismissive way only make things worse for themselves (Okazaki *et al.* 2020a).

2.7 Twitter hashtags, mentions and retweets

A hashtag is created by adding a "#" to the beginning of an unbroken word or phrase on Twitter (Murthy 2018). When including a hashtag in a Tweet, it is connected to all other Tweets that contain it. A hashtag adds context to a Tweet and can help a conversation last longer. Consider hashtags as a technique to link social media content to a particular topic, event, theme, or conversation. Since hashtags combine all social media information with the same hashtag, it is also easier to find postings on those specific themes (Greenhalgh *et al.* 2020).

A mention is a Tweet that includes the username of another person anywhere in the body of the message (Murthy 2018). These communications, as well as all of the responses, are saved in the Notifications tab. If including multiple usernames in the Tweet, it will appear on the Notifications page for all users (Hayon *et al.* 2019).

Customers interact with brands on Twitter mostly through mentions. Customers use them to share pleasant experiences with brands or to express disappointment with less-than-good ones (Greenhalgh *et al.* 2020).

A Retweet is a Tweet that has been re-posted. The Retweet function on Twitter allows the user and others to rapidly share a Tweet with all of their followers. A retweet, in its most basic form, is a technique to swiftly spread information while crediting the original source (Hayon *et al.* 2019). Retweeting is encouraged since it reaches a larger audience by sending tweets first to the user's own followers and then to the followers of anyone who retweets the user (Murthy 2018). Figure 2.1 breaks down the taxonomy of a tweet.

Figure 2.1: Annotated taxonomy of a tweet

Figure 2.1 presents the taxonomy of a tweet which shows that Edgars is Twitter-aware right up to the CEO level. The tweet above holds three hashtags, which resulted in 21 retweets and 16 likes because hashtags increased the visibility of the tweet.

Figure 2.1 shows that the tweet mentions not only EdgarsHelp and EdgarsFashion but also the CEO of Edgars, Grant Pattison. By mentioning a user in a tweet, this grabs the user's attention, resulting in Grant Pattison replying to the tweet.

The tweeter targets the specific users in their tweet by mentioning them. A Twitter mention is a technique for attracting the attention of another user in order to initiate a direct conversation with them on the network. It is shown that Grant Pattison responds to the tweet because he was mentioned in the tweet (Takashima *et al.* 2019).

2.8 Retail complaints on social media

This section provides an explanation of retail complaints on social media, starting with the description of customer complaints (Section 2.8.1), followed by the reasons customers complain on social media (Section 2.8.2). The section ends with customer complaint handling (Section 2.8.3) and the business benefits of customer complaint handling (Section 2.8.4).

2.8.1 Customer complaints

Traditional markets have discovered that factors such as a customer's specific characteristics, their views of the causes of their unhappiness, their expectations for the outcome, the type of product, and the expenses associated with complaining all have an impact on customer complaints (Mei, Bagaas and Relling 2019). Customers gather information about potential products or services during the information phase, search for suppliers, and inquire about prices and terms. The agreement phase establishes a strong bond between the supplier and the consumer, and the contract's details (such as product specs, payment, and delivery) are worked out. The settlement phase is when the product or service is delivered, after which guarantee claims are filed or help desk services are needed (TEHCİ and Ersoy 2020).

Customer complaints should not be viewed as a post-sale action but rather as an unpleasant service experience, according to a Service Dominant Logic (SDL) perspective (Mei, Bagaas and Relling 2019). Customers frequently have a preferred experience in mind, and when their expectations are not met, customer complaint behaviour occurs. Mei, Bagaas and Relling (2019) created a dynamic model based on this premise. The model depicts three types of customer complaint behaviour. Convey the complaint verbally or in writing, communicate the complaint verbally or in writing; and communicate the complaint verbally or in writing (Mei, Bagaas and Relling 2019).

The decision of action implies that a customer's decision is influenced by the cost of filing a complaint from the customer's perspective. If the customer believes that the chances of a successful compensation request outweigh the expense of filing a complaint, the consumer will choose to file a complaint (Jumaryadi 2019).

As a result of the recent expansion of social media, customers now have the ability to communicate with hundreds, if not thousands, of other consumers all over the world in a very short amount of time (Gunarathne, Rui and Seidmann 2017). Reasons for customers to complain on social media (Section 2.8.2) are explained in the next section. Customer complaints on social media can present retailers with both obstacles and possibilities. Customer complaints are generally regarded as beneficial by academics and practitioners alike (Preotiuc-Pietro, Gaman and Aletras 2019).

2.8.2 Reasons for customers to complain on social media

Platforms like Twitter and Facebook are increasingly being seen as ideal platforms for customers to escalate issues. Social media platforms are a simple tool for buyers to express their complaints because they are inexpensive, convenient, and public (Balaji, Jha and Royne 2019). Another reason is that users use platforms to amplify complaints when no other avenue seems to work (Gunarathne, Rui and Seidmann 2019).

According to Abraham et al. (2019), 79% of people who complain about a brand on Twitter do so in the hopes that their "friends will see it". Only 36% expect the brand to "notice it and handle the problem," while 43% hope "the company would see it."

The majority of customers use Twitter and other social media platforms to make sure that their complaints are seen by their friends, rather than to genuinely fix their concerns. The impact of social media referrals and endorsements on businesses is described by the idea of social proof. However, just as social media amplifies favourable suggestions, it can also elevate negative mentions (Grégoire, Salle and Tripp 2019).

The major reasons for customers to complain on social media platforms are to vent frustration, share negative experiences, a have desire to be heard, understood, and respected, and seek retribution by harming the retailer's reputation while also providing the store with an opportunity to improve (Gunarathne, Rui and Seidmann 2019). Customers who frequently voice their complaints on social media typically have two main motives: aim to caution others about their negative experiences or seek retribution or compensation, as they choose to share their grievances after the post-sale phase (Duica, Florea and Dobrescu 2019).

2.8.3 Customer complaints management

This section explains customer complaint handling, which leads to the business benefit of customer complaint handling. Many customers are dissatisfied with how frontline personnel handle complaints, and the majority of retailers fail to take advantage of the learning opportunities that complaints provide. A double or triple deviation is plainly the result of this. Double deviation causes annoyance and ambiguity in the scenario, prompting customers to express their dissatisfaction by sharing their negative experiences on social media (Mei, Bagaas and Relling 2019).

If a customer chooses to file a complaint after having a bad experience, it is critical that the complaint be handled properly. Otherwise, it will result in a second unsuccessful service recovery and customer discontent (Gunarathne, Rui and Seidmann 2017, 2019). A service process that has gone wrong for the second time, suggests that a failed recovery will often worsen the situation. It is also possible that upon receiving a second complaint, the service provider does not effectively address the issue, resulting in another unsuccessful attempt at service recovery. This causes triple variance, and the customer may decide to leave the retailer permanently (Duica, Florea and Dobrescu 2019).

2.8.4 Business benefit of customer complaint handling

Businesses benefit from customer complaint handling since customers aid in the identification of flaws, provide possibilities for improvement and furthermore increase consumer loyalty and improve their offerings. Businesses are more worried than ever about customer complaints because of increased competition in the retail industry as well as changes in customer behaviour and information communication technology (Mei, Bagaas and Relling 2019).

A customer complaint indicates a problem, whether it is a fault with the product, staff, or internal processes, and the business can investigate and adjust to prevent future complaints by hearing these difficulties directly from their customers (Yilmaz, Varnali and Kasnakoglu 2019). By responding to client complaints and concerns on social media, businesses demonstrate to other consumers that they are committed to keeping them happy. There are a variety of ways to use social media for customer service, including strategically using hashtags and creating a separate handle for customer assistance (Taylor *et al.* 2022).

Sentiment analysis aims to provide a method for analysing and evaluating client feedback found on websites, blogs, Twitter and Instagram (Yilmaz, Varnali and Kasnakoglu 2019). Customer reviews have had a tremendous impact on business development and enticing potential clients in recent years, due to the proliferation of online platforms (Kao, Wu and Syu 2018).

User-generated materials, such as customer experiences, user comments, and product evaluations, have increasingly become the key information source for both consumers and businesses on various social media platforms and websites. Customer reviews show the customer's experience with the firm, which is critical in comprehending the customer's thinking (Kao, Wu and Syu 2018). These reviews have a significant impact on other customers' decisions and serve as a foundation for business growth. The quantity of reviews has risen in recent years and paying attention to hidden characteristics in these reviews will undoubtedly improve business performance. In other words, while customers use these reviews to help them make decisions, firms use this information to help them build their businesses (Yilmaz, Varnali and Kasnakoglu 2019).

Information from customer complaints helps identify areas that need to be improved. Additionally, reading bad evaluations can inspire new, creative ideas for enhancing the company's offerings. Most companies have a set of rules and guidelines in place to make them run more smoothly (Mei, Bagaas and Relling 2019). Negative reviews usually bring out flaws in internal procedures and give hints as to what customers are not understanding or what is not working. By looking at this data, the company will be able to identify which operations need to be enhanced and which may simply be ignored because they are inconvenient or unneeded (Wei *et al.* 2020).

Customer feedback serves as a channel of communication between the business and its clients. Customers are also more likely to remain loyal to a business and spread the word if they feel that there is an open line of communication and that their input is respected (Yilmaz, Varnali and Kasnakoglu 2016). The reputation of the brand is enhanced by maintaining an open channel of contact so customers can voice complaints about the service or bad experiences. It improves the company's reputation by making it seem dependable and caring. It only helps to spread the word and enhance the reputation of the business when consumers are happy since they are more inclined to tell their coworkers, friends, and family about their pleasant experience (Wei *et al.* 2020). In the next section, the contextualisation of post-sale in the retail industry is discussed.

2.8.5 Ineffective post-sale services

According to research, ineffective post-sale services may result in online clients retaliating. Order cancellation without an issue and simple return and refund processes have become essential task-related expectations and indications of service performance. For online businesses, understanding the impact of product returns on consumer loyalty is critical (Pilkington 2021). However, there is a lack of evidence about whether and how product return experiences affect customer repurchase intent or loyalty. As a result, internet purchasing policies concerning product returns are increasingly being scrutinised (Penu and Nkanbia-Davies 2019; Savitha 2021).

Aside from product returns, the ability of customers to swap a purchased item has become a growing concern for e-commerce managers (Jaller and Pahwa 2020). The implications of a product exchange are different from those of a cash return because the latter allows for less consumer involvement. Researchers also discovered that a retailer's post-purchase practices, such as product exchange, have a beneficial impact on customer retention (Penu and Nkanbia-Davies 2019). However, the impact of online retailers' product exchange services on the customer-retailer relationship and customer purchasing behaviour has been studied in an empirical model as a single component (Penu and Nkanbia-Davies 2019; Pilkington 2021).

Furthermore, especially for individuals who shop online, repairs and upkeep are becoming a major problem. Home appliances and technological devices purchased online, for example, may need to be repaired throughout the guarantee term (Jaller and Pahwa 2020; Savitha 2021). Furthermore, Agarwal et al. (2019) stated that items have become more sophisticated in recent years, raising the buyer's after-sale risks. As a result of their more advanced technology, maintenance and repair of such products is becoming an increasingly important feature (e.g., electronics) (Savitha 2021).

2.8.6 Benefits of effective post sale services and why this research is needed

Complaining customers may become loyal if complaints are appropriately managed. It is possible to win a customer for life if the sales team can perceive the complaint as an opportunity and use it to its advantage by managing it correctly, because if the consumer is handled badly, he or she will never forget (Duica, Florea and Dobrescu 2019). This provides retailers with the opportunity to build a relationship and learn something new, to broaden their knowledge, to better understand behaviours, gain new experiences, and receive feedback (Mei, Bagaas and Relling 2019; Shin and Larson 2020).

Examples of post-sales support include assistance with warranty service, training, or repairs and upgrades (Buratti, Parola and Satta 2018). There are times when a company's whole marketing strategy includes post-sales service. Due to its after-sales service, a business's products could draw in some customers (Mishra and Singh 2020). Examples of post-sales services include businesses that help with the installation process (such as computer software), maintain products through free or inexpensive service (oil changes included with the purchase of a new car or via a paid service plan), or have a clear exchange and refund policy (Kaur and Singh ; Mishra and Singh 2020).

Post-sales support may be included with the purchase of an item or provided separately as part of a more complete service plan (Mishra and Singh 2020). Real-time online assistance can take the form of email, chat, forums, and a social media interface (and monitoring) that helps in responding to public complaints and criticisms. This can require handling returns or repairs (Buratti, Parola and Satta 2018). The research has shown how important digitalised technology has become in all aspects of life in times of the COVID-19 pandemic during the last few weeks and months (McKibbin and Fernando 2020). Particularly in the workplace, whether at home, in schools, workshops, or manufacturing plants.

During COVID-19's production halt, several manufacturers and operators realised that the production business was insufficient as a sole source of revenue (Ozili and Arun 2020). Finding new revenue streams, revising processes, and expanding the service business are all important. According to surveys conducted by the Fraunhofer Institute and the VDMA, post-sales service, which has been grossly undervalued to date, accounts for 20% of total sales (Ashfaq 2019).

However, there are no real indicators of digitisation, particularly in the service sector. Retail services are still analogue. The process of digitisation begins with a shift in mentality, making it easier for the service professionals to do their jobs by moving away from paper catalogues and handwritten notes to digital online and offline solutions (Vega Bernal 2019). Service 4.0, inspired by "Industry 4.0," is concerned with the forward-thinking technical possibilities opened by digitalisation and networking to innovate the traditional service company. Moving away from time-consuming, expensive, and inefficient on-site customer service and towards perfect, personalised customer care and intelligent spare parts management at a distance, utilising all digitalisation capabilities (Vega Bernal 2017, 2019; Korzer *et al.* 2020; Serravalle 2020).

The focus of post-sales is shifting away from the previous supply-oriented model and towards a new demand-oriented approach, driven by customer needs (Froot *et al.* 2017, 2019) . The focus of Service 4.0 is no longer on feasibility and practicability through technology or the manufacturer but on acute customer needs (Serravalle 2020). Focusing on consumer needs while also individualising the service offering are significant concerns for the coming years, and will be implemented in large, irreversible steps (Serravalle 2020). Digitisation creates enormous prospects and, in addition to large efficiency advantages, provides a competitive advantage that cannot be overstated: delighted customers (Korzer *et al.* 2020).

Table 2.1 describes a few examples of post-sale services that are currently being provided and the impact that it has on the consumers.

Table 2.1: The Effectiveness of Post-Sale Services

Reference	Post-sale Services	Effectiveness
(Kaur and Singh ; Buratti, Parola and Satta 2018; Mishra and Singh 2020)	Support for warranty service, training, or repair and upgrades are typical examples of post-sales service. Post-sales service is sometimes included in a company's overall marketing plan. Some customers may be attracted to a company's products because of its post-sales support. Companies that assist with the installation process (such as computer software), maintain products through free or cheap service (oil changes included with the purchase of a new automobile or via a paid-for service plan), or have a clear exchange and refund policy are examples of post-sales services.	Warranty services increase customer confidence in a product, it protects the business reputation and builds customer loyalty
(Kaur and Singh ; Buratti, Parola and Satta 2018; Mishra and Singh 2020)	Technical Support and Help Desk Services Post-sales support may be included with the purchase of an item or provided separately as part of a more complete service plan. Technical assistance for personal computers, mobile phones, software, machines, and a range of other products may be provided through a help or support desk.	There is no control over the quality of service provided. Many technical support and help desk services did not have in-house experience working on a specific product. There is no direct feedback from customers. Feedback from customers will go through several sources before reaching the company directly.
(Kaur and Singh ; Buratti, Parola and Satta 2018; Mishra and Singh 2020)	Real-time online support Email, chat, forums, and a social media interface (and monitoring) that assists in reacting to public complaints and critiques are all examples of real-time online assistance. This could entail taking care of returns or repairs.	The user experience of live chat is poor; 38% of customers and 43% of businesses feel that the user experience is inadequate. Scripted responses are inconvenient: 29% of customers and 38% of enterprises think that scripted responses are the most inconvenient. Waiting for an agent is aggravating: 24% of customers believe the most aggravating aspect of their experience is extended wait periods. Customers are most frustrated by long wait times, according to 19 percent of firms.

Reference	Post-sale Services	Effectiveness
		50/50 chance of having a positive live chat experience: In the last month, 47% of consumers have experienced a negative live chat experience.

2.9 Social analytic tools

This section compares three social analytic tools to find the best one to use. Talkwalker, Keyhole and TrackMyHashTag were selected based on their reputation in the literature (Jindal and Aron 2021; Kausar, Soosaimanickam and Nasar 2021; Moore, Bulmer and Elms 2022). Organisations now deploy Talkwalker to identify or locate significant rumours and misleading facts. Talkwalker has a dashboard as it gathers, analyses, and categorises data based on keywords from social media handles. Talkwalker is a commercial product and could be deployed for data collection. In order to compile all of the publicly viewable instances of a set of keywords found on the Internet, it makes use of ML and AI. Talkwalker allows users to further filter, contextualise, export, and analyse huge data sets (Caldwella et al., 2020; Lohiniva et al., 2022).

The software also has a tool for classifying data as false information. By compiling data on the total reach, participation, and demographics of those taking part in these discussions, it increases the understanding of the posts that are being tweeted. Due to privacy access limitations, the tool's understandable limitation is that it is unable to access discussions on Facebook, Instagram, and WhatsApp. There is evidence that men and young adults dominate Talkwalker exchanges and posts, omitting the voices of women, children and seniors (Caldwella et al., 2020; Lohiniva et al., 2022).

The second tool, called Keyhole, was created specifically for Twitter, Facebook, and Instagram. Keyhole is used to compute or establish the effects of a brand or trend on Twitter, Facebook, and Instagram. One may create a shareable institutional dashboard to support real-time access to records while also tracking operational, promotional, and campaigning indicators in real-time. The scheduling content option was challenging to utilise, according to Sokolov et al. (2020). However, it does take a long time to understand the tool because it is not user-friendly (Li and Lu 2018; Sokolov et al., 2020; Pyeon et al., 2021).

TrackMyHashTag focuses on monitoring social media campaigns, gathering data, and making various deductions from the data acquired. TrackMyHashTag enables users to choose or specify the desired data. Users can also select the region, year, and hashtags from which the data must be retrieved. The rate will be higher the more recent the data or tweets are. The information is in real time. TrackMyHashTag provides a seamless and expert service. The order, the invoice, and the delivery of the dataset all happened without issue. This tool also provides a tool for necessary post-sale support (Teh, Low and Si 2020; Zaver *et al.* 2021; Stracqualursi and Agati 2022).

2.10 Performance Metrics

A performance metric is used to evaluate the performance of the ML techniques, which includes four statistical measures, including reliability, accuracy, specificity and the F-measure. PER represents the probability of correctly identifying the True Positive (TP) class, Y means 'Yes', while specificity represents the probability of correctly identifying the True Negative (TN) class, and here Y means 'No'. If the model predicts a class as negative while the actual class is positive, the study defines it as False Negative (FN). On the contrary, if the model predicts a class as positive while the actual class is negative, the study defines it as False Positive (FP). Overall, accuracy measures the probability of detecting the true class. The data sets are divided into two categories: test and training, in order to establish the model with classifier algorithms and then to evaluate the performance of the models. The performance of the classification models for a certain set of test data is evaluated using a matrix called the confusion matrix.

Reliability value is shown as follows:

$$\text{Reliability} = \frac{TP+TN}{TP+FN+FP+FN} \qquad \text{Equation 2.1}$$

Accuracy value is shown as follows:

$$\text{Accuracy} = \frac{TP}{TP+FN} \qquad \text{Equation 2.2}$$

Specificity value is shown as follows:

$$\text{Specificity} = \frac{TP}{TP+FN+FP} \qquad \text{Equation 2.3}$$

F – Measure value is shown as follows:

$$\text{F – Measure} = \frac{2*Reliability*Specificity}{Reliability+Specificity} \qquad \text{Equation 2.4}$$

2.11 Feature Extraction

Feature extraction is a key job in natural language processing (NLP) that includes turning raw text input into a format that machine learning algorithms can easily analyse. In NLP, there are several strategies for feature extraction, each with its own set of advantages and disadvantages. Feature extraction is one of the most crucial phases in NLP for gaining a better knowledge of the context of the data. After the initial text has been cleaned, it must be transformed into characteristics that may be applied to modelling (Li et al. 2022). Document data must be transformed into numerical data, such as a vector space model since it cannot be computed. Feature extraction of document data is the term used most frequently to describe this processing activity. Feature extraction is also known as text representation, text extraction, and text vectorisation (Benyahia, Meftah and Lézoray 2022; Li, Tang and Jiao 2023).

One-Hot Encoding, Bag of Words (BoW), n-grams, Tf-idf, custom features and Word2Vec also known as word embedding, are the techniques found in feature extraction. One-hot encoding entails transforming your document's words into a V-dimension vector. This technique is straightforward and easy to code. This is one of the benefits of One-hot encoding. It is one of the most often used text vectorisation methods (Li *et al.* 2022). A bag-of-words is a text representation that specifies the appearance of words in a document. Used specifically in the text classification task, Scikit-learn's CountVectoriser class can be used directly. It is one of the most often used text vectorisation methods. A bag-of-words is a text representation that specifies the appearance of words in a document. Used specifically in the Text Classification task. Scikit-learn's CountVectoriser class can be used directly. A text document is represented by a bag-of-n-grams model as an unordered collection of its n-grams (Benyahia, Meftah and Lézoray 2022; Zhang *et al.* 2022a).

Tf-Idf is an abbreviation for term frequency and inverse document frequency. The term frequency-inverse document frequency (TF-IDF) statistic determines how relevant a word is to a document in a collection of documents. In NLP, word embedding is a term used for the representation of words for text analysis, typically in the form of a real-valued vector that encodes the meaning of the word in such a way that the words that are closer in the vector space are expected to be similar in meaning. Word embedding types are based on frequency (count of the frequency of words), BOW, Tf-idf, GloVe (using matric factorisation) (Zhang *et al.* 2022a; Li, Tang and Jiao 2023).

2.12 Exploratory data analysis

Exploratory data analysis (EDA) is basically a method for figuring out what the data may indicate, besides what can be gathered from formal modelling or hypothesis testing jobs. The form of the distribution, the existence of outliers, and measures of central tendency (which include the mean, the mode, and the median) are the four main criteria that EDA focuses on in order to analyse the data sets and emphasise their statistical properties (Majumder, Gupta and Paul 2022; Wang *et al.* 2023).

It is a method for condensing the data by highlighting its key features and representing it appropriately. EDA is more strictly focused on addressing missing values, transforming variables as necessary, and validating the assumptions necessary for model fitting and hypothesis testing. IDA (Initial Data Analysis) is a part of EDA and provides a brief description of the data sets' number of rows and columns, missing data, data types, and preview and deals with missing data, corrupted data, and inaccurate information types. The following are the techniques found in EDA (Milo and Somech 2020; Wang *et al.* 2023).

I. **Data exploration**: It is the initial step in the analysis of the data. Data exploration helps to explain the data set's content and characteristics in more detail. The size of the data is disclosed, and it provides the location of missing values in the data. It also Identifies any potential connections between the data (Milo and Somech 2020; Majumder, Gupta and Paul 2022).

II. **Data cleaning**: The process of cleaning data includes removing errors and confirming information. Data cross-checking can assist in identifying and correcting problems.

The problem can be fixed by validating this data (Milo and Somech 2020; Majumder, Gupta and Paul 2022).

The following are the methods used to perform EDA:

I. **Univariate analysis:** this method yields summary statistics for a single variable or for each field in the raw data set, for example, box plots, violin plots, PDFs and CDFs (Arisandy *et al.* 2023).

II. **Bivariate analysis**: this method involves utilising two variables to determine the relationship between them or determining the association between each variable in the dataset and the target variable of interest, for example, box plots, violin plots and scatter plots (Arora *et al.* 2022).

III. **Multivariate analysis:** this method is used to detect interactions between variables that are more than two or to comprehend interactions between various fields in the dataset. For example, a pair plot and a three-dimensional scatter plot (Arora *et al.* 2022).

2.13 Natural language Processing

Human-machine language interaction is the focus of the "field of computer science and artificial intelligence (AI)" known as NLP. It consists of several algorithms and methods, with the primary objective being to teach computers how to comprehend human language. According to Dreisbach et al. (2019), some NLP tasks include sentiment analysis document categorisation, translation, paraphrase detection, text similarity, summarisation, and question answering. Due to the complexities and ambiguities in the structure of human language, developing NLP is difficult. The literal meanings of words change depending on word form, sarcasm, and domain specificity in natural language, which also has high context specificity. Recent research on deep learning techniques has shown multiple successful attempts to achieve high accuracy in NLP tasks.

This study focuses on the NLP model, which uses a similar preprocessing step: (1) the input text is divided into words using tokenisation, and (2) these words are then recreated as vectors, or n-grams. To accurately convey the similarities and contrasts between diverse words, it is crucial to represent words in a low dimension (WANG 2018; Moon *et al.* 2021).

2.14 Natural language pre-processing methods

A set of features can be used by machine learning algorithms to identify patterns, and these patterns can then be used to predict a particular target variable for new, ambiguous data. A mathematical function that correctly translates the values of X (the features) to the unknown value of Y (the target) is what makes up the final trained model (Dreisbach *et al.* 2019; Kanakaraj and Guddeti 2019).

Machine learning algorithms can only function with data that is represented as numbers, as with any mathematical calculations. Additionally, it is crucial that these figures are presented in a form that reflects how the algorithm interprets the data because each algorithm operates under a unique set of restrictions and presumptions. Preprocessing is the process of converting unprocessed features into information that a machine learning algorithm can comprehend and learn from (Dreisbach *et al.* 2019).

The effectiveness of an NLP system is enhanced by text preprocessing. Adding a text preprocessing layer improves accuracy for tasks like sentiment analysis, document categorisation, document retrieval based on user queries, and more. Sentiment analysis is the goal of this research work. Stemming, lemmatisation, and text normalisation stages reduce the vocabulary size and standardise the language across a range of documents that were received from various sources (Pradha, Halgamuge and Vinh 2019; Mao *et al.* 2022).

With this strategy, the relevance between words and documents as well as between words and sentiment classes for sentiment analysis would be improved. An efficient NLP-based strategy is suggested for preparing online unstructured content in order to accomplish this goal. The five NLP phases make up the processing strategy, which aims to eliminate more unhelpful information and handle the expression styles that are frequently used in online informal language (Widiastuti 2019).

The method of tokenisation in NLP involves breaking down the given text into tokens, which are the smallest grammatical units of a phrase. Tokens include letters, numbers and punctuation (Mao *et al.* 2022). Words are reduced to their word stems when they are stemmed. A word stem need not be the same root as a morphological root from a dictionary; it is just equivalent to or a smaller form of the word (Widiastuti 2019). Lemmatisation seeks to reduce inflectional forms to a fundamental form, much like stemming does. In contrast to stemming, lemmatisation does not merely eliminate inflections. Instead, in order to obtain the correct word base forms, it uses lexical knowledge bases. Part-of-speech tagging (POS-Tag) is the process of labelling words in a document according to the kind of word they belong to (noun, adjective, adverb, verb, etc.). NER, or named entity recognition, is a technique used to determine whether a word is a named entity or not. Named entities can be things like places, organisations, people, and temporal expressions (Moon *et al.* 2021).

The pre-processing method is best done in the following order because it prepares the data ideally for the next pre-processing method: tokenisation, stemming, lemmatisation, part-of-speech tagging (POS - Tag) and named entity recognition (NER) (Mao *et al.* 2022).

2.15 Sentiment analysis

Sentiment analysis is a task in NLP that aids in comprehending the emotions of people communicating via text. People use text and words to communicate their feelings and opinions. The study of how to utilise computers to comprehend and alter natural language text or speech for beneficial purposes is known as natural language processing. It is a new technology that powers numerous types of AI (Soong *et al.* 2019; Mishev *et al.* 2020; Yadav and Vishwakarma 2020).

Social media offers a venue for coordinating the nature of sentiment. These feelings can be misclassified when they are not properly categorised. Sentiment analysis is the technique of separating non-trivial, subjective information from a variety of source materials that contain latent information about people's thoughts (Naldi 2019). A growing number of tools, packages, and APIs are available for sentiment analysis. Stanford CoreNLP is a comprehensive framework for handling NLP operations. It contains an analytical tool that makes use of deep learning methods (Khattak *et al.* 2021). The Natural Language Toolkit (NLTK) is a Python framework for creating Python programmes that make use of human language data. Its sentiment analysis technology, which is based on text classification, can categorise a sentence as good, negative, or neutral (Jagdale, Shirsat and Deshmukh 2019).

Three levels of sentiment analysis are studied: document level, sentence level, and aspect/phrase level. At the document level, the whole document is input, and the aim is to identify the document's overall viewpoint. Setting the sentence level involves two steps (Nanda, Dua and Nanda 2018). Determine whether the sentence is objective or subjective first. Next, classify each subjective sentence's viewpoint polarity. It can be challenging to identify these and their opinion orientations in sentence level analysis because sentences may express opinions about several different elements of several items (Iglesias and Sánchez-Rada 2021). Since aspect level operates at the word or phrase level, it does fine-grained opinion analysis. An aspect is a property of an object that users can comment on in reviews. The entity's characteristics are extracted, and the opinion polarity of each retrieved characteristic is determined (Priyavrat 2017; Kauffmann *et al.* 2020).

Aspect-based opinion mining, also known as feature-based opinion mining, is an innovative method that results from this. Aspect level analysis facilitates the discovery of numerous elements of various products from the reviews. Identifying the opinion orientation expressed on each aspect is required for the aspect opinion classification job (Adarsh *et al.* 2019; Bonta and Janardhan 2019). There are three basic methods for classifying aspect of opinions: supervised learning, poorly supervised learning and unsupervised learning (Bonta and Janardhan 2019).

The Valence Aware Dictionary and Sentiment Reasoner (VADER) and the Lexicon-based method are two of the most popular sentiment analysis Python modules (Arora *et al.* 2023; Elinda, Yuliansyah and Latiffi 2023). As opposed to ML techniques, lexicon-based approaches like VADER use lexicons for sentiment analysis. The lexicon approach creates a connection between words and emotion, and a sentence's overall emotion is the culmination of the emotions connected to each term. A Lexicon emotion analysis polarity score runs from -1 to 1, with -1 signifying a very negative emotion and 1 signifying a very positive attitude. A score that is nearly zero indicates neutral sentiment (Elbagir and Yang 2019; Pano and Kashef 2020).

An economical rule-based system for analysing the sentiment of text from social media is called VADER. It examines a range of lexical signals to ascertain the text's polarity (intensity) and mood. The VADER lexicon is a dictionary that assigns a score—from -4 (most extreme negative) to 4 (most extreme positive)—to each property, which may be a word, an acronym, or an emoticon. A valence-based language that can distinguish between the polarity and intensity of sentiments was developed by VADER. This is combined with strong modifiers such as negations, contractions, conjunctions, booster words, degree adverbs, capitalisation, punctuation, and slangs (s) to calculate scores for the input text (Elbagir and Yang 2019; Pano and Kashef 2020).

The Lexicon-based approach and VADER differ significantly in that VADER is focused on social networking. As a result, VADER spends a lot of time deciphering the emotions behind content like emoticons, repeated words, and punctuation that regularly occurs on social media (Elbagir and Yang 2019; Pano and Kashef 2020).

2.16 Approaches to sentiment analysis

Methods used for sentiment analysis are divided into three categories: ML-based approaches, lexicon-based approaches, and hybrid approaches. Tweets can be categorised as positive, negative, or neutral as part of the opinion mining process. Lexicon-based sentiment analysis is based on a word's presence in a document. The lexicon includes elements such as the subjectivity of words, part-of-speech tagging, and sentiment ratings. Using the properties offered by these lexicons, the sentiment analysis of tweets is annotated. Each word has a unique value assigned to it. By averaging the word-level sentiment values, it can be used to determine the overall polarity of the tweet. When employing a lexicon-based technique, sentiment analysis is carried out at the document and sentence levels by looking up words in a specified word list and comparing them to sentences in a predefined dictionary (Erşahin *et al.* 2019; Shinde *et al.* 2021).

Sentiment analysis based on machine learning creates a model by teaching the classifier with labelled instances. First collect a dataset with positive, negative, and neutral classes, then extract the features and words from that dataset. Finally, train the algorithm using the instances from the dataset (Erşahin *et al.* 2019).

Lexicon-based and ML approaches are combined in a hybrid approach. Lexicon-based and machine learning (ML) approaches for sentiment analysis are combined in a hybrid approach. The commonly utilised hybrid approach combines the two and primarily depends on sentiment lexicons, which are available in most systems. Sentiment analysis is a hybrid approach that distinguishes between polarity utilising both knowledge-based and statistical methods. A hybrid ML strategy using SVM and two feature selection strategies utilising the multi-verse optimiser was developed by Hassonah et al. (Birjali, Kasri and Beni-Hssane 2021). This study has selected the hybrid approach, combining both the sentiment lexicon approach and ML techniques.

Require locating data that contain opinion because tweets contain a lot more unrequired information. This information is used for sentiment analysis. This is achieved through feature selection. Before feature selection, look for tweets with adjectives, since the presence of an adjective in a tweet denotes that the tweet expresses an opinion about something in the outside world. The next phase is feature selection to determine which tweets are subjective. Subjective tweets are those that express the author's feelings or opinions about the outside world (Erşahin *et al.* 2019; Shinde *et al.* 2021).

Thereafter, substitute the word position with the polarity value supplied by the lexical dictionary to determine the polarity of tweets by searching the word's occurrences in the dictionary. The total number of times the word "polarity" appears in a tweet is used to calculate its overall polarity. Prior to that, each word had its own polarity assigned. One of the main problems with sentiment analysis is how to handle negatives (Erşahin *et al.* 2019).

Due to the fact that many sentences start with a negative word, which changes the polarity of the phrase, numerous classifiers eliminate negation words by treating them as stop words. This issue can be solved by replacing any negative words with the punctuation mark "!" when discovered in sentences. To do this, add the sign "!" before each word in the lexicon, which merely shifts the polarity of those words. The lexical dictionary's polarity for each sentence can be used as training data. These training data are used to train the machine learning classifier (Erşahin *et al.* 2019; Shinde *et al.* 2021).

2.17 The use of sentiment lexicon

The "words" that convey the viewpoint are the most important ones for sentiment analysis. Positive opinion is the term that should be used to describe an entity, while negative opinion should not be used. A lexicon is a collection of words that have been predetermined and are given polarity scores. It is a simple technique for sentiment classification. This classifier matches words in a lexicon and utilises the results to classify sentences. The quantity of the lexicon affects how efficient this classification method is. Sentiment lexicon is a list of lexical qualities that are typically classified as either positive or negative depending on their semantic orientation (Chiong, Budhi and Dhakal 2021).

SentiWordNet is an extension of WordNet that annotates 147 306 synsets with three number scores for objectivity (neutrality), negativity and positivity. WordNet is a lexical database that records the semantic connections between words in more than 200 different languages (Akçay and Oğuz 2020). Based on semantic connections between words, such as synonyms, hyponyms, and meronyms, WordNet joins words together (Freihat *et al.* 2023; Maudslay *et al.* 2023). The synonyms are divided into synsets along with brief definitions and usage examples. A wide range of terms are covered. Each score has a range of 0.0 to 1.0, and their total for each synset is 1.0. It is a helpful and well-liked vocabulary for a variety of text mining activities. SentiWordNet is used in this study via Python's NLT.

The syn.pos_score () and syn.neg_score () functions are used by the SentiWordNet technique to calculate each word's scores from this lexicon. The positive score (TweetPos) of a tweet is calculated by adding the terms in the tweet that have a greater positive score than a negative score. Similar to this, words in a tweet that have a greater negative score than positive score are added together to determine the tweet's negative score (TweetNeg). All tweets are given these scores. Sentiment polarity was determined by filtering and analysing the data using NLP algorithms based on the emotion terms found in the user tweets. In order to extract emotions for the textual data from each tweet, the dataset is prepared using natural language pre-processing techniques, such as word tokenisation, stemming and lemmatisation, POS tagging, NER, and parsing. When a word in a sentence has a meaning in the current context, the derived algorithm uses WordNet to extract emotional terms and assigns sentiment polarity using the SentiWordNet dictionary and a lexicon-based technique.

This approach in PyCharm makes use of the Natural Language Toolkit (NLTK) and the Python programming language. The development process is facilitated by PyCharm's capability to highlight errors. It also provides access to C#, HTML, XML, Dart, and Rust, among other languages. The PYQT web toolkit is used to construct sizable Graphical User Interface (GUI) applications. The resulting method retrieves emotional terms from phrases that have significance in the current context using WordNet and its POS.

Tokenisation is a process that divides lengthy text strings into tokens, which are smaller units of text. Sentences can be tokenised into words, and words can be tokenised into larger sections of text. Tokenisation was carried out using the split() technique (Akçay and Oğuz 2020). After breaking the given string with the designated separator, it returns a list of strings. Split() splits a string at each space by default.

Stemming is the process of removing affixes from a word in order to retrieve the word stem (suffixes, prefixes, infixes, and circumfixes) (Beysolow II 2018). Utilising LancasterStemmer, a Python-based stemming tool, one may do stemming in PyCharm. It was necessary to import NLTK, a package that does stemming using various classes.

Lemmatisation is similar to stemming but differs in that it can capture canonical forms based on the lemma of a word (Beysolow II 2018).

2.18 Reasons for sentiment analysis application in social media

As the amount of information available on social media expands, sentiment analysis is becoming more essential; it can provide online advice and suggestions for both customers and businesses. By disclosing customer preferences, the data can assist e-commerce platforms in analysing their products and services. The rise of Web 2.0 is reshaping the social media landscape (Gowri et al., 2021). Not only are people using online social media to interact, share information, and express their personal opinions with others, but businesses can also utilise social media to communicate, understand, and improve their products and services. The number of social media users is growing , and it is estimated that in 2023, there will be 2.77 billion social media users globally (Kausar, Soosaimanickam and Nasar 2021; Rahman, Sadat and Siddik 2021).

Monitoring customer loyalty on social media and measuring consumer loyalty can be achieved through monitoring their opinions towards brands or products, assessing the effectiveness of marketing campaigns and communications, and locating and interacting with the top influencers who are most relevant to the brand, product, or campaign (Chauhan, Sharma and Sikka 2021; Jindal and Aron 2021).

Consumers are more likely to believe the opinions of other consumers, particularly those who have had prior experience with a product or service, than corporation marketing. The internet's influence on people's purchasing behaviour, particularly via social networking, has expanded over time (Kauffmann et al., 2020).

Government, business, and now scientific organisations are becoming increasingly interested in gathering public opinion on social issues, political movements, business plans, and marketing initiatives. According to Yoo, Song, and Jeong (2018) and Patel and Passi (2020), 95% of data is categorised as unstructured and is infrequently examined. Sentiment analysis looks at people's attitudes, feelings, and views towards various objects, including goods, people, subjects, organisations, and services. The majority of sentiment analysis studies have concentrated on market and perceived consumer responses to commercial brands. For instance, Van de Kruis (2019) concentrated on brand attitude and sentiment analysis.

Persuasion (such as from a campaign lobbyist) can all influence how people feel. The research interest in sentiment analysis or opinion mining studies is to understand what the sentiment is and why it is a negative sentiment (Gowri et al., 2021; Kausar, Soosaimanickam and Nasar 2021).

2.19 Supervised machine learning techniques

When labelled data is provided for the data analytic model, supervised ML is applied. Training the model and making predictions are the two processes in supervised learning. A model is created by feeding a labelled data set through a classification algorithm during training. The model is then fed the test data, and it makes a category prediction (Han *et al.* 2020).

One of the most often used strategies in supervised ML is classification. It is the process of identifying and distinguishing data classes by using a collection of models. The goal of classification is to use historical data to determine the class of future data objects. A training set is typically used to learn the model, and then the learned information is tested on the test set (Asha Kiranmai and Jaya Laxmi 2018; Lashari et al., 2018). A few examples of supervised learning methods are presented hereafter.

I. Naïve Bayes is a supervised ML algorithm with a high probability of success. It does not take into account the position of a term in the sentence; therefore, each word is treated independently. The probability of each phrase that corresponds to a label is calculated using the Bayes theorem by Naïve Bayes. The prior probability of the label in the dataset is p(label). The prior probability of a feature associated with a label is p(feature|label). The prior probability of a feature occurring is p(feature). SentiWordNet Lexicon with Naïve Bayes was utilised in the study to improve the classification of the Twitter dataset by providing the score of positive and negative tweets (Jiang, Gradus and Rosellini 2020; Haupt *et al.* 2021). The Naïve Bayes classifier cannot establish a semantic association between words since each word is treated independently, whereas the Bayesian network can. The words heavily depend on each other in a Bayesian network. The Bayesian network expresses dependency as an acyclic directed graph, with each node representing a variable and edges representing the relationship between variables (Ren *et al.* 2020; Singh *et al.* 2021).

II. Supper Vector Machine (SVM) is set up to solve binary classification issues. Its goal is to find the optimal hyperplanes, which serve as a separator to characterise the decision boundaries between data points from different classes. A hyperplane should be chosen that can maintain the maximum distance between two support vectors of different classes. The SVM is capable of handling both linear and non-linear classification tasks. For classification, SVM is employed with various weighting techniques such as term frequency-inverse document frequency (TF-IDF), term occurrence, and binary occurrence. The study chose chi-square and noise reduction. The study demonstrated that using chi-square feature selection with SVM improves accuracy with the help of the experiment (Ren, El-Khamy and Lee 2020; Singh *et al.* 2021).

III. Artificial Neural Networks (ANN) mimic the arrangement of neurons in the human brain. Neural networks' fundamental building block is the neuron. An input layer, a hidden layer, and an output layer make up an ANN. The input to the neuron is a vector called "a (i)" that represents the frequency of a word in a document. The function is calculated using a weight "A" that is assigned to each neuron. The linear function used by the neural network is x (i) = A. (a (i)). The class is categorised using the sign of x (i). There are two processes involved in ANN model training: forward propagation and backward propagation. The input is supplied to the input layer of neurons in forward propagation, where it is multiplied by weights that are generated at random. Functions are used to normalise the output value between 0 and 1 (Vanhaeren *et al.* 2020; Singh *et al.* 2021).

IV. The decision tree is a tree-like structure in which the non-terminal nodes represent features and the terminal nodes represent labels. The path is chosen based on a criteria. This is a recursive procedure that will lead to a terminal node that will assign a label to an input. A decision tree is an effective tool for sentiment analysis since it can handle vast amounts of data. The training data is divided into hierarchical groups via a decision tree. A condition based on the attribute value is used to divide the data. The presence or absence of a word determines the condition. The division process is repeated until the terminal nodes represent the limited number of attributes needed for categorisation (Vanhaeren *et al.* 2020; Singh *et al.* 2021).

2.20 Machine learning techniques evaluated for the research aim

2.20.1 Unsupervised learning

Contrary to unsupervised ML, supervised ML requires trained data in order to process it, while having a relatively high efficiency. While unsupervised ML does not need training data, its accuracy may be constrained. Unlabelled data is easier to collect than labelled data. Each category's keyword lists are used to categorise the sentence. The unsupervised approach is more convenient for analysing domain-dependent data (El Rahman, AlOtaibi and AlShehri 2019; Chamarczuk *et al.* 2020; Bibi *et al.* 2022).

A unique method for analysing sentiments expressed on social media websites is to use sentiment analysis algorithms based on clusters. It is a key objective of exploratory ML and a typical ML technique. The formal description of k-means clustering is given as an optimisation problem, where the objective is to assign the objects to the nearest cluster centre while minimising the squared distances from the cluster. Since the optimisation issue is known to be NP-hard, the typical strategy would be to look solely for approximations of the answers. The "k-means algorithm" (also known as Lloyd's algorithm, or simply "algorithm") is a particularly well-known approximate method (Beigi and Moattar 2018; Sebai, Wang and Wang 2020).

In fuzzy logic, which is an extension of deterministic logic, the truth value ranges from 0 to 1, rather than having a binary value. Fuzzy logic theory's main goal is to transform a black-and-white problem into a grey-area problem. The easiest way to describe human knowledge in the context of AI may be to convert it into IF-THEN rule-based natural language phrases. These laws are based on natural language models and representations, which are based on fuzzy sets and fuzzy logic. Fuzzy rule-based systems are effective and well-recognised tools for pattern identification and categorisation. Due to the presence of this fuzziness, fuzzy rule-based systems can handle doubt, ambiguity, or vagueness quite effectively (Vashishtha and Susan 2019; Alharbi and El-kenawy 2021).

I. Clustering: By grouping things into clusters, objects can be organised so that those with the greatest degree of similarity remain in one group while those with the least or no similarity remain in another. When using cluster analysis, data objects are compared for similarities and categorised based on whether they share those similarities or not (Huang *et al.* 2020).

II. Association: An unsupervised learning method for finding relationships between variables in a large database is called an association rule. It identifies the collection of objects that coexist in the dataset. The association rule enhances marketing effectiveness. (Kumar, Glisson and Cho 2020)

2.20.2 Weakly supervised learning

In practice, the supervised methods cannot always be employed since they require labelled corpora, which are not always available. Weakly-supervised and unsupervised approaches, which do not require pre-tagged data, are another option for ML. A huge collection of unlabelled data plus a small set of labelled data makes up weakly supervised learning (Sebai, Wang and Wang 2020).

The learning device is used for the input in the unsupervised technique, and no predicted output values are specified. These algorithms compile sentiment text using a dictionary-based technique. Every term in a dictionary has antonyms and synonyms. This approach is utilised to locate seed emotive words using a dictionary's antonym and synonym arrangement. Initially, a small set of terms with known positive or negative coordination is gathered. This method was repeated until no new words were discovered (Ren, El-Khamy and Lee 2020).

Examples of weakly supervised learning are zero-short learning, distant supervision and transfer learning with weak supervision, to mention just a few. Zero-short learning is the training of models to generalise to classes that were not seen during training (Xian *et al.* 2018). Distant supervision is when noisy data from a related domain is used for the labelling of the training dataset (Smirnova and Cudré-Mauroux 2018). Transfer learning with weak supervision involves the pre-training of models on large labelled datasets and fine-tuning them with weakly labelled datasets in the target domain (Zhang *et al.* 2022b). (Dong *et al.* 2020).

2.21 Ensemble machine learning techniques

By adding unique learning algorithms, the traditional ML methods are evolving and advancing. Hybridisation and ensemble modelling techniques are being used to enhance the computation, accuracy, resilience, and utility of machine learning models. Recently, there have been a lot of hybrid and ensemble ML models released. In predictive modelling, machine learning techniques outperform most statistical and physical techniques in terms of computation cost, computation accuracy, robustness, simplicity, and uncertainty analysis. Due to this, machine learning (ML) techniques have become much more popular in the last several years in a variety of fields, such as energy, hydrology, finance, economics, hazard prediction and computational mechanics.

There are many distinct ML techniques, and academics have recently categorised these techniques in several ways. One of the more common classification techniques is grouping the techniques into the three categories of single approaches, hybrid methods, and ensembles (Rahimi, Eassa and Elrefaei 2020).

It is frequently seen that the hybrid and ensemble methods outperform individual ML algorithms. The two main strategies for developing more precise and dependable ML systems are ensemble and hybrid ML methods. Hybrid ML models are created by combining ML approaches with other ML methods, with other soft computing methods, and/or with optimisation techniques to enhance the method in a variety of ways. While ensemble methods combine multiple ML classifiers using different grouping approaches like bagging or boosting, it is claimed that the development of fresh ensemble and hybridisation methodologies will be crucial to the success of ML in the future (Dong *et al.* 2020).

2.22 Deep learning

The use of deep neural networks to train the nodes for extracting complicated characteristics from the available input with minimal contribution is the most crucial concept in deep learning. These algorithms can automatically learn new complex features because they are flexible and do not require operator input. The disadvantage of employing deep learning techniques is that they need a large amount of data to provide results that are highly efficient. The method used is to extract a set of emotionally charged words from the text that have been professionally nominated, then classify these words using lexical resources like WordNet or vocabularies. Much of the work cited above focuses on identifying the prior polarity of the terms or phrases to use before assigning the sentiment polarity to the word using the WordNet lexical resource. The following lists deep learning algorithms.

Deep Neural Network (DNN) is an extension of traditional Artificial Neural Network (ANN). There are two major differences between DNNs and regular neural networks. One or two hidden layers are sufficient for ordinary neural networks. On the other hand, DNNs have a lot of hidden layers. The Google Brain project, for example, uses a neural network with millions of neurons. Deep learning algorithms come in a variety of models, including DNNs, Convolutional Neural Network (CNN), Recurrent Neural Network (RNN), Long Short Term Memory (LSTM), which is a variant of RNNs and a host of others. Attention-based networks, which focus on specific regions of a deep neural network, have been developed to further extend the capability of RNN for NLP tasks. The following lists deep learning algorithms.

I. A Convolutional Neural Network (CNN) is a type of feedforward neural network that was developed for use in computer vision. Its design is based on the human visual cortex, which is a visual system in the brain of animals. The visual cortex comprises a large number of cells that sense light in small, overlapping sub-regions of the visual fields known as receptive fields. Over the input space, these cells act as local filters. CNN is made up of several convolutional layers, each of which performs the function that the cells in the visual cortex process (Pandian 2021).

II. A Deep Neural Network (DNN) is a collection of layered neural networks, or networks with multiple layers. FF-DNN, also known as multilayer perceptrons (MLPs), are DNNs with more than one hidden layer and only one forward direction, as the name suggests (no loopback). Both categorisation and prediction are possible with these neural networks (Ain *et al.* 2019). This study employs a classification technique for spoken LID. The input and output nodes will match the input features and output classes when the FF-DNN is employed as a classifier. Weights, biases, nonlinear activation, and backpropagation are the most significant ideas in an FF-DNN (Goularas and Kamis 2019).

III. Recurrent Neural Networks (RNNs) in order to forecast a layer's output, (RNNs) operate on the principle of preserving the layer's output and feeding it back into the input. The RNN will standardise the various activation functions, weights, and biases to ensure that each hidden layer has the same properties. It will just create one hidden layer and loop over it as many times as necessary, rather than building several. The information in RNNs cycles through a loop to the middle hidden layer (Dong, Wang and Abbas 2021).

There are four types of Recurrent Neural Networks:

1. One to one
2. One to many
3. Many to one
4. Many to many

In this study, a many to one RNN was used. A single output is produced by this RNN from a number of inputs. An example of this kind of network is sentiment analysis, which determines if a text expresses positive or negative thoughts.

2.23 Sentiment Analysis ML Approach

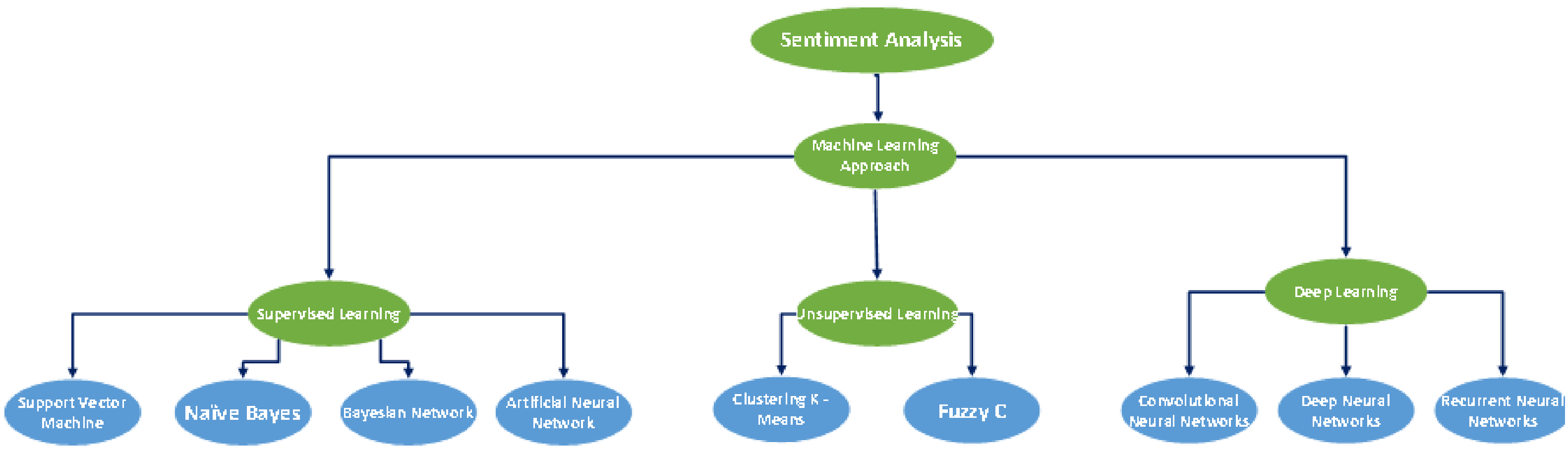


Figure 2.2: Sentiment Analysis ML Algorithms

Figure 2.2 illustrates the best-suited ML techniques for sentiment analysis based on the literature review, which are supervised learning, unsupervised learning and deep learning, along with their ML algorithms, which are support vector machines, Naïve Bayes, Bayesian networks and artificial neural networks for supervised learning, clustering c-means and fuzzy C for unsupervised learning. Deep learning techniques incorporate convolutional neural networks, deep neural networks, and recurrent neural networks. Figure 2.2 will serve as the foundation for the methodology employed in this study.

2.24 Ensemble Machine Learning for Sentiment Analysis

An ensemble of models is a collection of learning models whose individual predictions are integrated so that component models compensate for each other's flaws (Sadhasivam and Kalivaradhan 2019). Although the general ML community is becoming more interested in ensemble learning approaches, their usage in sentiment classification is still limited (Khai Tran and Thi Phan 2019).

This technique is based on the idea that various models have distinct inductive biases. If the errors induced by these biases are uncorrelated, the models in the ensemble should correct for one another's faults, lowering the overall error when the model outputs are aggregated (Khai Tran and Thi Phan 2019). Ensemble approaches have been demonstrated to efficiently use this attribute to reduce variance error without increasing bias error. Ensemble learning approaches aiming at developing more generalisable models are gaining popularity (Sadhasivam and Kalivaradhan 2019).

In their 2020 study on sentiment classification in Arabic customer reviews, Sultana and Islam employed an ensemble method using stacking ensembles with Naïve Bayes, logistic regression, and SVM as meta-learners. In terms of the well-known F-measure, it was found that all ensemble strategies outperformed their individual component models and that stacking with logistic regression produced the best outcomes for subjectivity and sentiment classification. However, difficulties in choosing base learners for a stacking strategy were noted in this study, and such tactics were purposefully avoided. Instead, contrasted ensemble learning methods such as bagging, boosting, random subspace, and vertical partitioning, using as base learners Naïve Bayes, maximum entropy, a decision tree, k closest neighbours, and an SVM. (Sultana and Islam 2020).

In order to classify sentiment at the document level, a study by Kazmaier and Van Vuuren (2022) used ensemble learning. They extracted numerous features for base learners from datasets by recognising the various structures that induce polarity shifting in the text, termed 'surface features.' Other features, known as 'deep features,' are extracted via word embedding and deep learning (Kazmaier and van Vuuren 2022). The proposed system was created, and experiments employing datasets were undertaken to evaluate its performance in Vietnamese and English. The suggested system outperformed even state-of-the-art deep learning models and other ensemble learning systems in comparison with other ML methodologies. The experimental results also suggest that considering 'deep features' for base learners improves the system's effectiveness (Kazmaier and van Vuuren 2022).

Kanakaraj and Guddeti (2019) presented an effective ensemble learning system based on datasets of base learners, which include features derived from linguistic characteristics and the application of a deep learning model. Word embedding and creating a deep learning model were utilised to make the system work better for basic learners.

By introducing generalisability and robustness, classification ensemble methods try to develop the bias of a single machine using various base estimators and methods. Two main categories of approaches can be distinguished based on how the predictions are combined: Averaging approaches, in which numerous estimators are generated separately and the ensemble's outcome is the average of the individual forecasts (Kanakaraj and Guddeti 2019). This area includes tagging systems, randomised tree forests, and other related topics. Boosting approaches, in which the base estimators are constructed progressively and the prediction model output is the result of this sequential arrangement. This category includes AdaBoost, Gradient Tree Boosting, and others (Kanakaraj and Guddeti 2019).

(Sadhasivam and Kalivaradhan 2019) (Figure 2.3) shows a side-by-side comparison of the performance of classic ML algorithms with ensemble approaches. It can be seen that the ensemble technique exceeds the other solo methods in terms of precision. The ensemble technique definitely outperforms SVM in terms of classification performance, as evidenced by the F-score values (Sadhasivam and Kalivaradhan 2019).

(Sadhasivam and Kalivaradhan 2019) compared the performance of the ensemble approach to that of various ML techniques such as SVM, Baseline, Maximum Entropy, and Naïve Bayes.

Experiments were also carried out to examine the results of different boosting and bagging ensemble approaches (Figure 2.4).

Each constituent learning method in the ensemble method of classification differs in terms of the subset of the dataset used for training and the subset of the feature vector utilised for training and testing. These subsets are randomly selected. Extremely Randomised Trees perform better than other ensemble approaches. This is because the way splits are generated in severely randomised trees has a large randomness factor (Sadhasivam and Kalivaradhan 2019).

Further(Sadhasivam and Kalivaradhan 2019) (Figure 2.5) compared the performance of a proposed enhanced bag-of-words model (keywords with synsets and sense) to the simple bag-of-words model (keywords only). On the same dataset, the same set of classification algorithms were tested with different feature vectors, such as plain keywords, keywords with synsets, and keywords with synsets plus word sense values. When employing synsets instead of basic keywords, there is a 3-6 percent improvement. This is because the coverage of related terms in the feature vectors has increased (Sadhasivam and Kalivaradhan 2019).

When employing word senses in feature vectors, there is a modest performance improvement. This is because there is limited room for sense judgement when the training and test data are from the same context. In generalised systems, where the training and test data come from a wide range of domains, the sense value in the feature vector is extremely important (Sadhasivam and Kalivaradhan 2019).

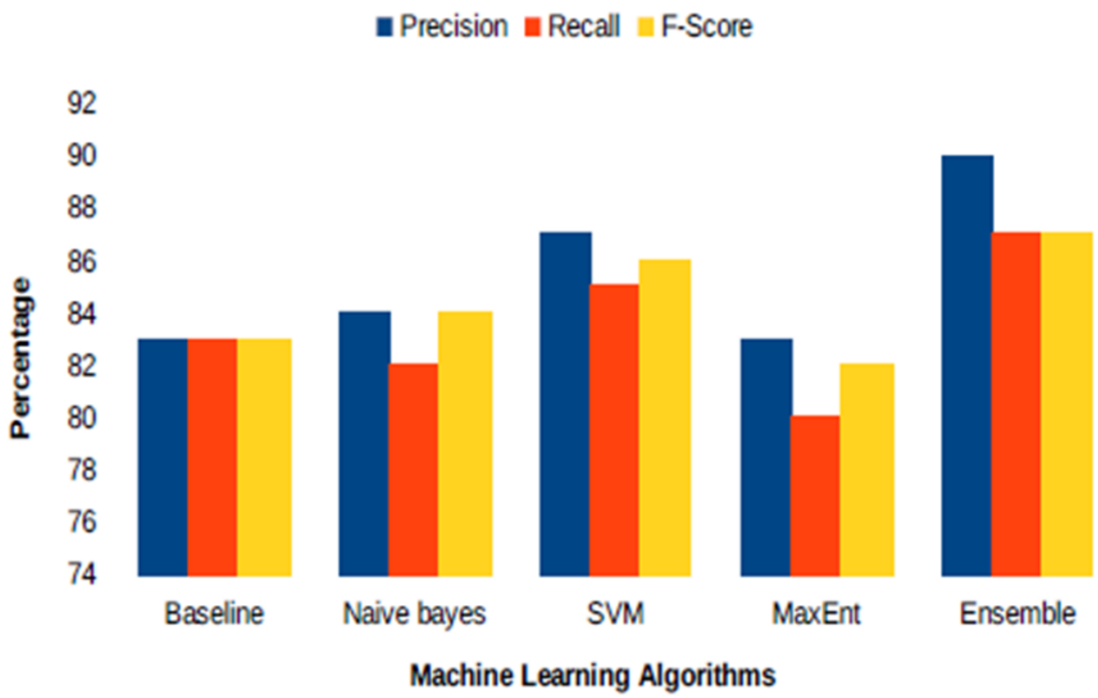


Figure 2.3: Performance of ML Algorithms (Sadhasivam and Kalivaradhan 2019)

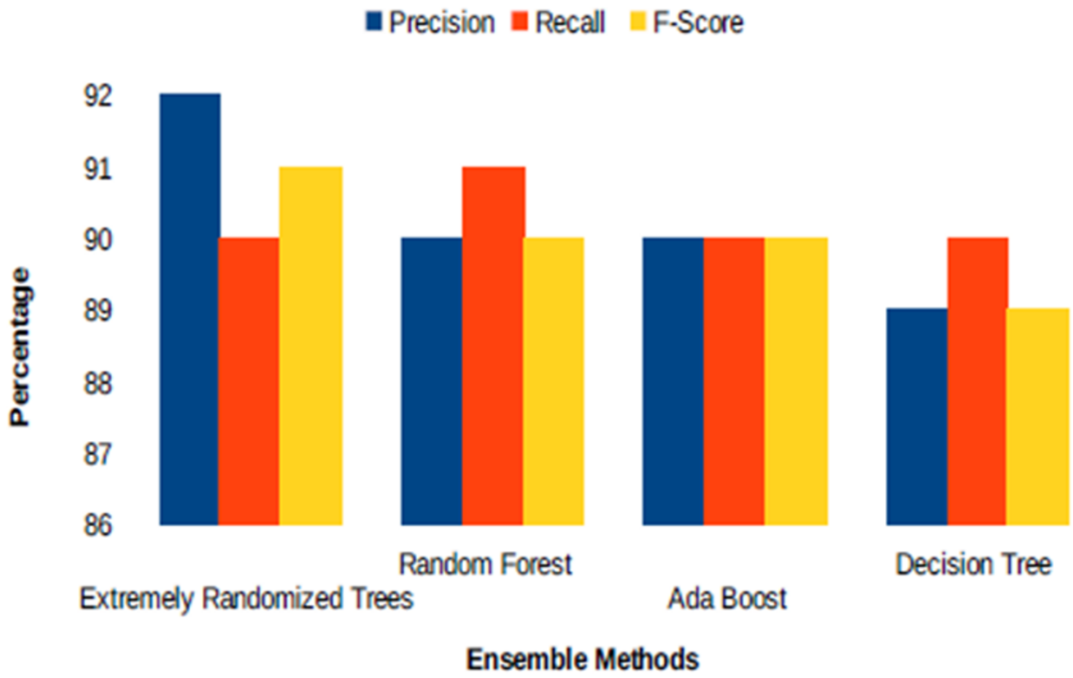


Figure 2.4: Performance of Ensemble Models (Sadhasivam and Kalivaradhan 2019)

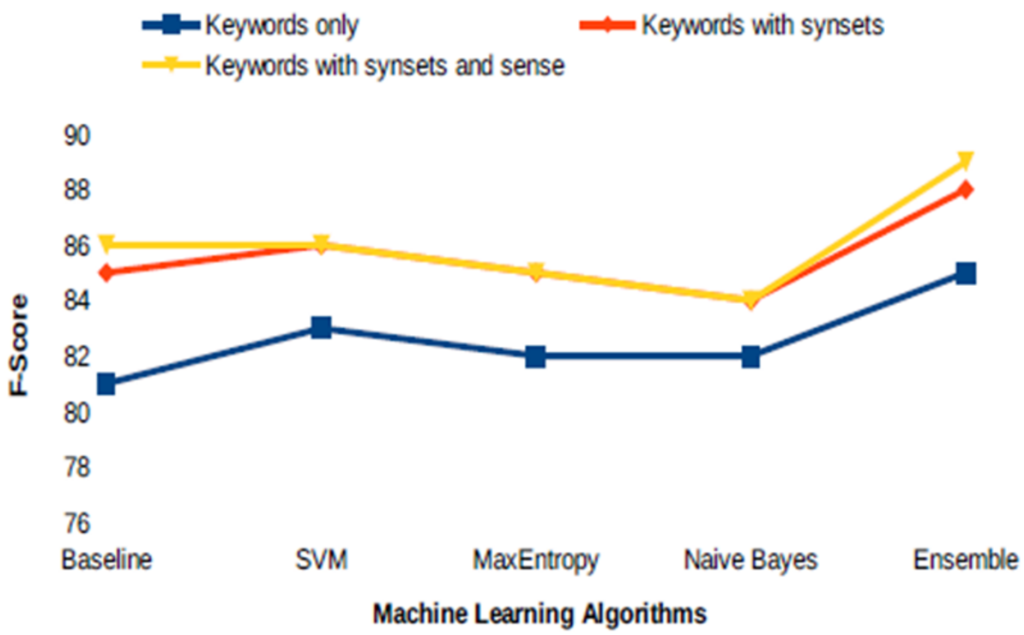


Figure 2.5: Comparison of ML Algorithms and Ensemble Models (Sadhasivam and Kalivaradhan 2019)

Table 2.2 describes the advantages and disadvantages of the recommended ML techniques that have been used for sentiment analysis in the literature review.

Table 2.2: Recent ML techniques used in sentiment analysis

References	ML Classifier	Description	Advantage	Disadvantage
(Mishev *et al.* 2020; Yadav and Vishwakarma 2020; Noori 2021)	KNN	Due to its simplicity and accuracy, the KNN algorithm is a well-known instance-based approach that has been widely applied to text categorisation. The KNN classifier ranks the document's neighbours among the training documents to label it. The k-nearest neighbour number is the most critical parameter impacting classification when using the KNN.	KNN is simple and is also used for multiclass categorisation of documents.	KNN requires more time to categorise when huge numbers of data are incline. Takes lot of memory to run a process
(Peikari *et al.* 2018; Soong *et al.* 2019)	Decision Tree	The decision tree is a well-known ML technique for automating classification tree induction from training data. There are usually two phases in a decision tree training method. The first phase is tree growth, which involves greedily dividing each tree node to create a tree. Because the tree can overfit the training	Decision tree is fast in learning data sets. It is easy to understand the purpose of the data set	It has a problem that is difficult handle data with noisy data

References	ML Classifier	Description	Advantage	Disadvantage
		data, the overfitted branches of the tree are deleted in the second phase.		
(Luo and Weng 2019; Han *et al.* 2020)	Naïve Bayesian	NB Tree is a Bayesian decision tree and Bayesian rule supervised classifier. The Bayes rule is used to calculate the probability of each class of cases in this procedure. A Naïve Bayes classifier assumes that the presence (or lack) of a specific feature of a class (i.e., attribute) has no bearing on the presence (or absence) of any other feature. The Naïve Bayes classifier has the advantage of requiring less training data to estimate the means and variances of the variables needed for classification. Because independent variables are assumed, only the variances of the variables for each label, rather than the whole covariance matrix, must be computed.	Simple and work well with textual as well as numerical data. Easy to implement Computationally cheap	Performs very poorly when feature set is highly correlated. Gives relatively low classification performance for large data set. Independent assumption of attribute may lead to inaccurate result.

References	ML Classifier	Description	Advantage	Disadvantage
(Ren *et al.* 2020; Sebai, Wang and Wang 2020)	Support Vector	SVM is a categorisation approach that uses supervised learning. It is an effective methodology for separating positive and negative samples by determining the optimum possible surface. Behind the training process, the main purpose of SVM is to identify a maximum margin hyperplane to accomplish the feature review classification problem. There are no limits on how far the two classes can be separated. It is critical to choose a decision boundary that has a maximum margin between any points from both classes when determining the best class.	High accuracy even with large data set. Works well with many numbers of dimensions	Problem in representing document into numerical vector.

2.25 Current challenges with sentiment analysis

Due to the volume, diversity, velocity, variability, and authenticity of data, which are the major characteristics of big data, analysing it is a difficult process. The filtering of relevant data from non-related data is limited by data volume, which may impair sentiment analysis results (Yue et al., 2019). The diversity and velocity of big data are restricting feature extractions, which is one of the most important tasks in the pre-processing of sentiment analysis data sets.

Extraction of opinion words and sentences, as well as POS tagging, are difficult tasks when the dataset is large and the data is diverse (Chen, Lee and Chen 2020).

Another challenge with sentiment analysis with big data is the large amount of memory necessary for the pre-processed dataset to be analysed. Storage of large amounts of data in various formats is a technical issue that is handled by sophisticated storage strategies (Iglesias and Moreno 2019). Another problem is big data velocity, because sentiment analysis on dynamic and real-time events in the big data world is a difficult operation that must be handled effectively while considering how people's opinions change over time (Yue et al., 2019).

There is a rapid change in the data set in social media data. There are problems that are content-related and incorporate hashtags, links, pictures, and spam. Multiple opinions in a statement, negation handling, sarcasm detection, implicit opinion, comparison sentences, and opinion spam are all issues that need to be addressed (Chen, Lee and Chen 2020).

The researcher identifies three commonly used performance results which, as found in the studies, are precision, recall and accuracy.

Precision is the ratio of correctly predicted positive observations to the total predicted positive observations (Reichert *et al.* 2017).

Precision = $\frac{TP}{TP+FP}$ *Equation 2.5*

Recall is the ratio of correctly predicted positive observations to all observations in the actual class (Reichert *et al.* 2017).

Recall = $\frac{TP}{TP+TN}$ *Equation 2.6*

Accuracy is the most intuitive performance measure and it is simply a ratio of correctly predicted observations to the total observations (Reichert *et al.* 2017, equation 2).

Table 2.3 presents a summary of studies that implemented supervised, unsupervised and deep learning algorithms. The aim of this table is to identify the commonly used algorithms in supervised, unsupervised and deep learning. It helps the researcher identify which algorithms can be explored to develop a data analytics framework for Twitter in order to enhance the quality of responses to post-sale complaints.

Table 2.3: Summary of studies that implemented supervised, unsupervised and deep learning algorithms

<table>
<tr><th>Author</th><th>The study</th><th>Dataset/Sample</th><th>ML Techniques Applied</th><th colspan="4">Algorithm Performance Results Found in the Literature</th></tr>
<tr><td rowspan="6">(Reichert et al. 2019)</td><td rowspan="6">A supervised ML study of online discussion forums about Type 2 diabetes</td><td rowspan="6">The dataset consists of 39 425 texts collected from 42 online forums.</td><td rowspan="6">Support Vector machine (SVM), Neural Network, linear and logistic Regression, Random Forest</td><td>Algorithm</td><td>Precision</td><td>Recall</td><td>Accuracy</td></tr>
<tr><td>Multinomial NB</td><td>0.67</td><td>0.66</td><td>0.68</td></tr>
<tr><td>Random Forest</td><td>0.69</td><td>0.68</td><td>0.68</td></tr>
<tr><td>Linear and Logistic Regression</td><td>0.63</td><td>0.62</td><td>0.61</td></tr>
<tr><td>Neural Network</td><td>0.69</td><td>0.69</td><td>0.69</td></tr>
<tr><td>SVM</td><td>0.65</td><td>0.64</td><td>0.63</td></tr>
<tr><td rowspan="6">(Vanhaeren et al. 2020)</td><td rowspan="6">A comparative study of supervised ML algorithms for the prediction of long-range chromatic interactions</td><td rowspan="6">Chia-PET datasets targeting the complex component</td><td rowspan="6">Support Vector machine (SVM), Neural Network, linear and logistic Regression, Neural Network, Convolutional Neural Network (CNN) and Multilayer Perceptron (MLP)</td><td>Algorithm</td><td>Precision</td><td>Recall</td><td>Accuracy</td></tr>
<tr><td>SVM</td><td>0.869</td><td>0.870</td><td>0.869</td></tr>
<tr><td>Neural Network</td><td>0.843</td><td>0.846</td><td>0.842</td></tr>
<tr><td>Linear and Logistic Regression</td><td>0.963</td><td>0.963</td><td>0.965</td></tr>
<tr><td>CNN</td><td>0.821</td><td>0.822</td><td>0.821</td></tr>
<tr><td>MLP</td><td>0.822</td><td>0.23</td><td>0.792</td></tr>
<tr><td>(Amruthnath and Gupta 2018)</td><td>A research study on unsupervised ML algorithms for fault detection in predictive Maintenance</td><td>Kaggle Dataset</td><td>PCA T2 statistic, Hierarchical clustering, k-means clustering, Association and model-based clustering to test its accuracy, performance, and robustness</td><td colspan="4">Based on these results, when data was fitted in hierarchical clustering, K-means, and C-means, the results were nearly identical.</td></tr>
<tr><td>(Mackey et al. 2020)</td><td>ML to detect self-reporting of symptoms, testing access, and recovery associated with COVID-19 on Twitter: Retrospective big data infoveillance study</td><td>Tweets were collected from the Twitter public streaming application programming interface from March 3-20, 2020, filtered for general COVID-19-related keywords</td><td>Multilayer Perceptron Neural Network, Convolutional Neural Network, Long Short-Term Memory, Generative, Adversarial Network, Restricted Boltzmann Machine, Deep Belief Network</td><td colspan="4">In light of the research results, Restricted Boltzmann is the best suited algorithm for the detection of Symptoms, Testing
Access, and Recovery Associated With COVID-19
Restricted Boltzmann can be used to generate genome-wide maps of chromatin interactions, and information on a few chromatin features at the
anchors may be enough to yield accurate predictions.</td></tr>
<tr><td>(Ferrag et al. 2020)</td><td>Deep learning</td><td>Datasets used by IDSs: it indicates whether the study focused</td><td>DNN,
Feed forward,</td><td>Algorithm</td><td>Accuracy</td><td>Detection</td><td>Overall</td></tr>
</table>

Author	The study	Dataset/Sample	ML Techniques Applied	Algorithm Performance Results Found in the Literature			
	for cyber security intrusion detection: approaches, datasets, and comparative study	on the datasets used for intrusion detection systems	Recurrent Neural network, Convolutional Neural network, Restricted Boltzmann Machine				**Accuracy**
				Deep Neural	0.57	0.78	99%
				Feed Forward	0.52	0.85	99.59 %
				Recurrent	0.49	0.75	92%
				Convolutional	0.53	0.77	96%
				Restricted Boltzmann	0.66	0.68	91%

2.26 Identifying barriers to data analytics adoption in retail

While there have been studies on the use of big data analytics, there appears to be a dearth of studies in South Africa that focus on industry-specific applications. The study's key finding was that South African retailers do not use big data analytics. However, some retailers are turning to big data analytic systems to speed up the processing of vast amounts of structured data and deliver insights at a lower cost. According to (Ridge, Johnston and O'Donovan 2021), South African merchants are having difficulty identifying a use case to justify their investment in big data analytics.

Every hour, retailers like Walmart acquire about 2.5 petabytes of data from customer transactions alone. With the massive amounts of data being generated on a terabyte and even petabyte scale, big data analytics is required to extract insights from big data (Lacinski *et al.* 2023; Tang *et al.* 2023). While organisations such as The Data Warehousing Institute have performed research on the use of big data analytics, there appears to be a lack of studies that analyse industry-specific use of big data analytics in South Africa (Ridge, Johnston and O'Donovan 2021).

Amazon, the world's most well-known online retailer, has access to all of its customers' personal information, including names, search histories, payment methods, and addresses. Amazon develops personalised suggestions and provides efficient customer service by using all of their information. One of Amazon's most essential and valuable strengths is its powerful recommendation engine, which is built on big data integration. It gathers information from consumers' purchase histories, as well as viewings, clicks, search searches, and in-cart goods. Based on this information, Amazon makes extremely accurate recommendations while also driving purchases to slow-moving products and listings (Rathor, Agarwal and Dimri 2018; Lohiniva *et al.* 2022). The information that Amazon is gathering is a constant flow of real-time information, making it agile and adaptive to market needs.

The video streaming service has access to all of its users' viewing choices and behaviours from across the world. Netflix uses sentiment analysis based on their user's viewing preferences to recommend content that the audience will enjoy, as well as select films or series that may grab the attention of specific individuals (Smith and Telang 2018). The shortage of data science and big data professionals in this area imposed a financial challenge on Netflix (Sakoda, Takayasu and Takayasu 2019; Lohiniva *et al.,* 2022).

Retailers in South Africa analysed big data because they could not find a use case for big data. Based on studies conducted by (Ridge, Johnston and O'Donovan 2021), big data vendors and professional services confirmed that retailers were not using big data but that some of them were in the process of investigating it. The response reflected a clear lack of consensus on the definition of big data. Some use the term big data interchangeably with business intelligence. Some retailers confirmed that they could potentially start using big data analytics once they perfected the mining of their structured transactional data.

The most likely barriers for retailers are the availability of analytical skills, the costs of big data analytics platforms, contextualising big data to provide meaning, return on investment of big data ventures, leadership buy-in, product selection and the need to structure current data. There was a general recognition by respondents of the importance of having skilled individuals discover insights from big data. Identifying and attracting employees with the appropriate abilities to take advantage of large datasets is a challenge for South African retail companies (Ridge, Johnston and O'Donovan 2021).

The literature review revealed that supervised, unsupervised and deep learning had yet been used in identifying retail complaints on Twitter. This led to the identification of a further gap for additional research. This field is new and emerging in South Africa and this study will contribute to it.

2.27 Conclusion

A specific review of the academic literature revealed retailers in South Africa were not generating or analysing big data because they could not find a use case for big data. The literature review explains why Twitter is a more popular platform for customer complaints compared to Facebook and LinkedIn. The business benefit of customer complaint handling is that it aids in the identification of flows, providing possibilities for improvement, increasing customer loyalty and improving their offering. Reasons for sentiment analysis, which can provide online advice and suggestions for both customers and businesses. The literature review revealed that supervised, unsupervised and deep learning have yet to be used in identifying retail complaints on Twitter. This led to the identification of a further gap for additional research. This field is new and emerging in South Africa, and this study wants to contribute towards it. The advantages and disadvantages of the selected algorithms have been identified to assist in the methodology.

Chapter 3 Research Methodology

3.1 Introduction

This chapter provides a comprehensive description of the study's research design. The study is first positioned in the post-positivism research paradigm (Section 3.2), followed by an overview of the research methodology and time horizon (Section 3.3). The chapter then proceeds to discuss the data source (Section 3.4); this is then followed by a detailed breakdown of the study's proposed conceptual framework (Section 3.5). The chapter ends with a detailed description of the experiment design (Section 3.6).

3.2 Research paradigm

The emphasis on meaning and the creation of new knowledge placed by the post-positivist research paradigm makes it possible to support dedicated social movements, or movements that seek to alter the course of history and advance social justice. Positivism is a thought paradigm in scientific philosophy that emphasises the value of observation for the expansion of knowledge and views the measurement of phenomena as essential to the process of understanding how things work. However, it acknowledges the necessity for a theoretical framework to organise data in its more complex characterisations (Vega 2022; Ariail Reed 2023). Data is the core of this study; the conceptual framework works on data collection, data preparation and data extraction.

In post-positivism, the research is broad rather than specific. It is difficult to keep theory and practice apart. The post-positivist perspective affirms the importance of politics, values, and passion in research. This type of research necessitates the ability to perceive the big picture, to step back, and to take an overview. The natural sciences, which rely on actual observation to produce generalisable theories and models, have made extensive use of positivism. This disregards sources of information that cannot be observed and are hence untestable as unscientific (Dahal 2023).

This research aims to determine the most suitable ML technique for sentiment analysis by the construction of a conceptual framework to improve the identification of customer complaints on Twitter using a hybrid approach that combines lexicon-based and machine learning (ML) techniques for sentiment analysis. Post-positivist researchers do not see themselves as inevitably solving the problems they set out to investigate (Vega 2022; Mustofa *et al.* 2023). The positivist paradigm gave rise to the post-positivist paradigm. It is focused on the subjectivity of reality and departs from the logical positivists' entirely neutral viewpoint. This research observes and analyses customer experience on the Twitter platform using machine learning techniques, as well as the performance of the selected machine learning techniques for sentiment analysis, which is the objective of the post-positivist paradigm.

3.3 Research methodology and time horizon

Quantitative research usually entails a systematic and empirical analysis of events using statistics and mathematics, as well as numerical data processing. In quantitative research, the process of estimating numbers serves as a vital link between empirical observation and the mathematical expression of quantitative relationships. Data is often selected and analysed numerically in quantitative research (Panhwar, Ansari and Shah 2017, Rahi 2017).

Qualitative methodology focuses on interpretation and the quality of things (Creswell and Creswell, 2017). When there is a need to analyse and process large volumes of quantitative data to verify hypotheses and test a theory, there is uncertainty related to the theories under consideration, the data obtained can be quantified and compared. Statistics, which are used in quantitative research, are an important area of mathematics. Data processing in quantitative research is usually done with specialised statistical tools (Pozzi et al., 2016; Chen, Lee and Chen 2020). The study's goal is to automatically analyse customer experience, which results in a quantitative mode of study.

Thus, a quantitative methodology was adopted for this study due to the researcher's intention of numerically analysing the opinions expressed in large volumes of data collected from a social media platform. The view is that a quantitative understanding of sentiment trends and patterns not only significantly reduces the amount of time and effort usually found in manual approaches but can be utilised on a broader scale due to its portability.

As data was collected over a period of 12 months in this research, the time horizon is longitudinal. Longitudinal studies are a type of correlational research in which researchers observe and collect data on a number of variables without trying to influence those variables (Bouadjenek *et al.* 2022). Researchers repeatedly examine the same data to detect any changes that might occur over a period of time.

3.4 Data Source

Data for this study was drawn from TrackMyHashTag, which is a Twitter analytic tool. TrackMyHashtag has been described earlier in Section 2.9.

3.4.1 Data Collection

TrackMyHashTag focuses on monitoring social media campaigns, gathering data, and making various deductions from the data acquired from Twitter. TrackMyHashTag enables the user to choose or specify the desired data. The researcher could select the region, year, and hashtags from which the data must be retrieved. This tool also provides necessary support and engagement with a consultant that can provide the required data directly from Twitter in real time.

Quantitative research is primarily concerned with numbers and statistics in order to test hypotheses and generalise findings to a large population. Twitter data has enabled the gathering and analysis of data for the retail experience. There are studies that look at Twitter for a specific health problem, collect data, and analyse it using a variety of quantitative research approaches. A total of 240 000 rows of Twitter dataset were received. Twenty-four percent (24%) of this data was retained because it was retail related data. The aim of this study is to analyse and observe retail complaints on Twitter using ML techniques for sentiment analysis.

A time frame from January 2017 to January 2019 was selected to collect historical data. A fee of R876,54 was paid per month for the duration of one year, which included the tracking of Twitter activities based on research data requirements, the gathering of the required Twitter dataset, data accuracy and authenticity, and email consultation. During the consultation, it was requested that retail-related data be included in the historical data collection posted on Twitter within South Africa only (TrackMyHashTag, 2023).

Each month, a dataset was received, which included 20 000 tweets from January 2017 until the end of January 2019. The dataset presented Tweets with the following columns (i.e., the tweet metadata): creation of tweet, gender, created, description, username, profile image, tweet, hashtag, tweet_count, tweet_created and tweet location, all in chronological order.

The Twitter data collection normally starts on the 1st of every month, starting from 5:00 until 23:00and then continues to the next date until the end of each month, which continues to the next month in chronological order.

During each month, the dataset was received from the TrackMyHashTag portal. The researcher analyses the dataset that best fits the study, which is then highlighted on the provided CSV file and uploaded back on the TrackMyHashTag portal. The consultant often contacts the researcher to understand and further refine the tweets, which leads to further improvement of the quality in each month.

The first receipt of the dataset was in August 2021, in which the researcher only retained 2 000 rows of the data out of 20 000 rows of dataset. The reason for this is because 18 000 rows of data were from outside South Africa. 1892 rows of data were within South Africa and 108 rows of data inside South Africa included tweets related to retail. The 108 rows of data were retained providing an objective of the required data for this research.

In the month of September 2021, only 1000 rows of data were retained, but there was an improvement in the quality of the data. 13 768 rows of data were found in South Africa and the balance was international. 978 of the retained data was retail-related. 2500 rows of data were retained for the month of October; the dataset now included Tweets within South Africa only, but only 2489 were related to retail. Thereafter, the researcher requested that the geolocation of the tweets be improved. Geolocation should now state where in South Africa the tweet was from and should not only state “South Africa”, but it should also be more specific by stating the province as well.

Noise data was also requested by the researcher in order to test the model to be constructed because the conceptual framework would have to search through a variety of Tweets within Twitter. Once the researcher identified the retail tweets within the dataset provided each month, the balance of the tweet was retained as noise data, thereby providing a complete dataset. For instance, in Table 3.1 for the month of July 2021, 7582 rows of retail data were retained, and a further 418 rows were retained as noise data.

A fee of $129 was paid for a final CSV file, which included 30 000 Tweets but 27 233 tweets with 11 columns were used. The CSV file included Tweets within South Africa only, and the majority of the dataset were tweets related to retail (TrackMyHashTag 2023).

Table 3.1 provides a brief description of the extracted metadata, which contains the Creation of Tweet, Gender, Created, Description, Username, Profile Image, Tweet, Tweet_count, Tweet_Created and Tweet Location

Table 3.1: Description of the dataset's columns

S/N	Column Name	Description of Column
1	Creation of Tweet	The time the tweet began to be created.
2	Gender	Gender of the profile.
3	Created	The time the tweet was posted on Twitter
4	Description	Profile description
5	Username	Username of the tweet
6	Profile Image	Profile Image
7	Tweet	The actual tweet
8	Hashtag	Hashtags found within the tweet
9	Tweet_count	The number of characters of the tweet.
10	Tweet_created	The time the tweet was posted on Twitter. This has been identified as an unused column within the dataset and is removed.
11	Tweet Location	The location of where the tweet was created and posted.

Table 3.2 presents the date the dataset that was received, the total rows of the dataset received, followed by the number of retained retail data rows. Retained retail data are tweets related to retail, they were either advertisements, complaints, compliments, or mentions within a tweet. A total of 240 000 rows of the Twitter dataset were received, 24% of which were retail-related and retained. A further 2% of the 240 000 dataset were retained as noise data.

Table 3.2: Amount purchased and the usage of data

Collection Date	Total Data Size (in rows)	Retained Retail Data (in rows)
31-Aug-20	20 000	108
30-Sep-20	20 000	978
28-Oct-20	20 000	2489
29-Nov-20	20 000	2789
30-Dec-20	20 000	2776
27-Jan-21	20 000	3891
27-Feb-21	20 000	5459
28-Mar-21	20 000	7423
28-Apr-21	20 000	7938
28-May-21	20 000	6723
29-Jun-21	20 000	7669
22-Jul-21	20 000	7582
Total amount of data	**240 000**	**55825**

3.4.2 Data Preparation

This research selected exploratory data analysis (EDA) and missing values for data preparation. The selected data preparation methods prepared the CSV file, which contained Twitter data for data extraction and pre-processing.

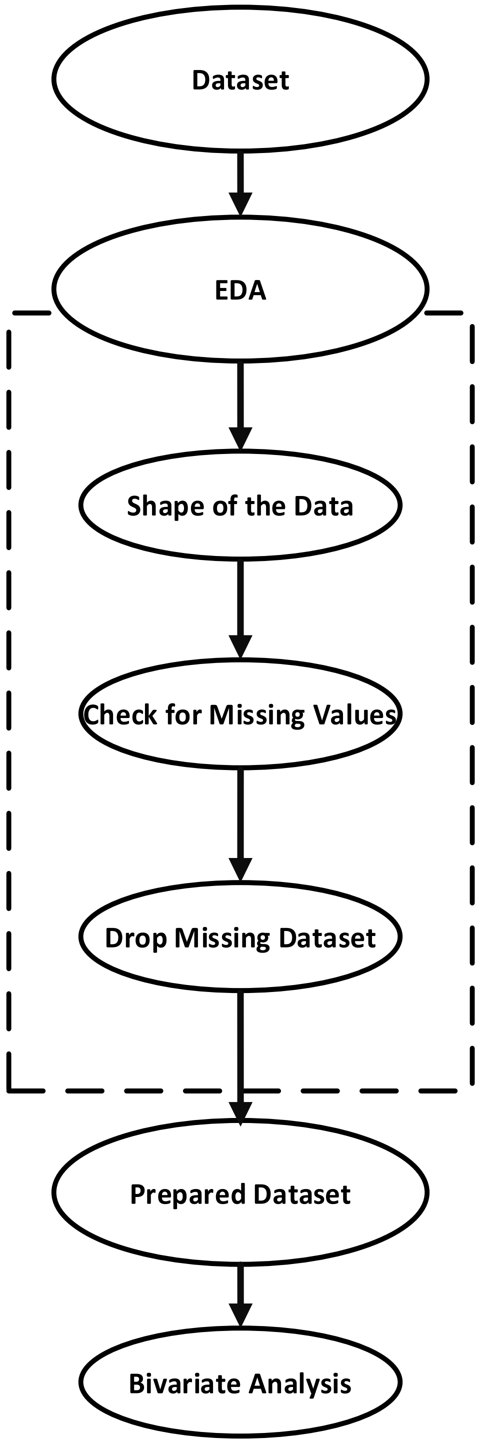


Figure 3.1: Selected EDA Methods

3.4.2.1 Exploratory Data Analysis and Missing Values

Figure 3.1 illustrates the methods and steps followed to perform EDA. Before the dataset can be used within the conceptual framework, it must first be prepared. In order to carry out data preparation, this research selected exploratory data analysis (EDA), which has been explained earlier in the literature review chapter (Section 2.11). The first step is to understand the shape of the data, which includes the columns and rows of the dataset. This is followed by identifying missing values within the dataset and the final step is to remove the missing values within the dataset by dropping the missing dataset. The EDA pseudocode, which was implemented using the pandas library methods, is described and presented in Algorithm 1.

Bivariate analysis will then be performed to determine the correlations between two variables which will be the time of the tweet and the day of the tweet for each province in South Africa, as explained in Section 2.12. This will include Limpopo, Mpumalanga, the Free State, the Eastern Cape, KwaZulu-Natal, Gauteng, the Western Cape, the Northern Cape and the North West.

Algorithm 1: EDA

Input: Unprepared Data set
Output: Prepaid Dataset

```
function EDA (dataset)
df = pd.read_csv(dataset)  //Initialise the dataset into a dataframe
 df.head() //Identify the beginning of the dataset
  df.tail() //identify the end of the dataset

      for values in dataset
        df.dtypes // the type or format of the variable

      if output(df.isnull().sum()) //identify null columns
    df = df.drop() // drop null values
  end if
end for
df.shape
output(df.shape() //dimension of the prepared dataset
end function
```

The first step involves understanding the meaning of the variables as well as the shape (i.e., the number of columns and rows). The second step is to identify and visualise missing values in the dataset. The *isnull()* method in pandas identifies missing values; adding *.any()* to the end of this method will return a boolean (true or false) column depending on whether the column is complete or not. Another approach is to use the opposite method, *notna*(), which returns a count of the filled values in the dataframe. Several visualisation techniques can be used to visualise missing data. The missingno Python library was adopted for visualising missing values in this study. It presents a convenient visualisation and understanding of the missing values' patterns within a dataframe, created from the research dataset.

In the third step, the identified missing values are dropped using the *dropna()* method in the pandas framework phase of this study. This further reduces the dimension of the dataset and prepares it for the conceptual framework.

3.5 Conceptual Framework

The first step in the conceptual framework as shown in Figure 3.2 involves data extraction (Section 3.5.1). This is followed by pre-processing techniques for NLP (Section 3.5.2) and, thereafter, the feature extraction stage (Section 3.5.3). A brief introduction to sentiment analysis using the hybrid approach is presented (Section 3.5.4). This is then followed by the sentiment lexicon (sub-section 3.5.4.1), then training and testing of machine learning models for sentiment analysis (sub-section 3.5.4.2). Thereafter, model evaluation and visualisation are presented (Section 3.5.5), and the conceptual framework ends with the selection of the most appropriate ML model (Section 3.5.6).

In this study, the proposed conceptual framework comprises of eight stages, as illustrated in Figure 3.2, which are described below.

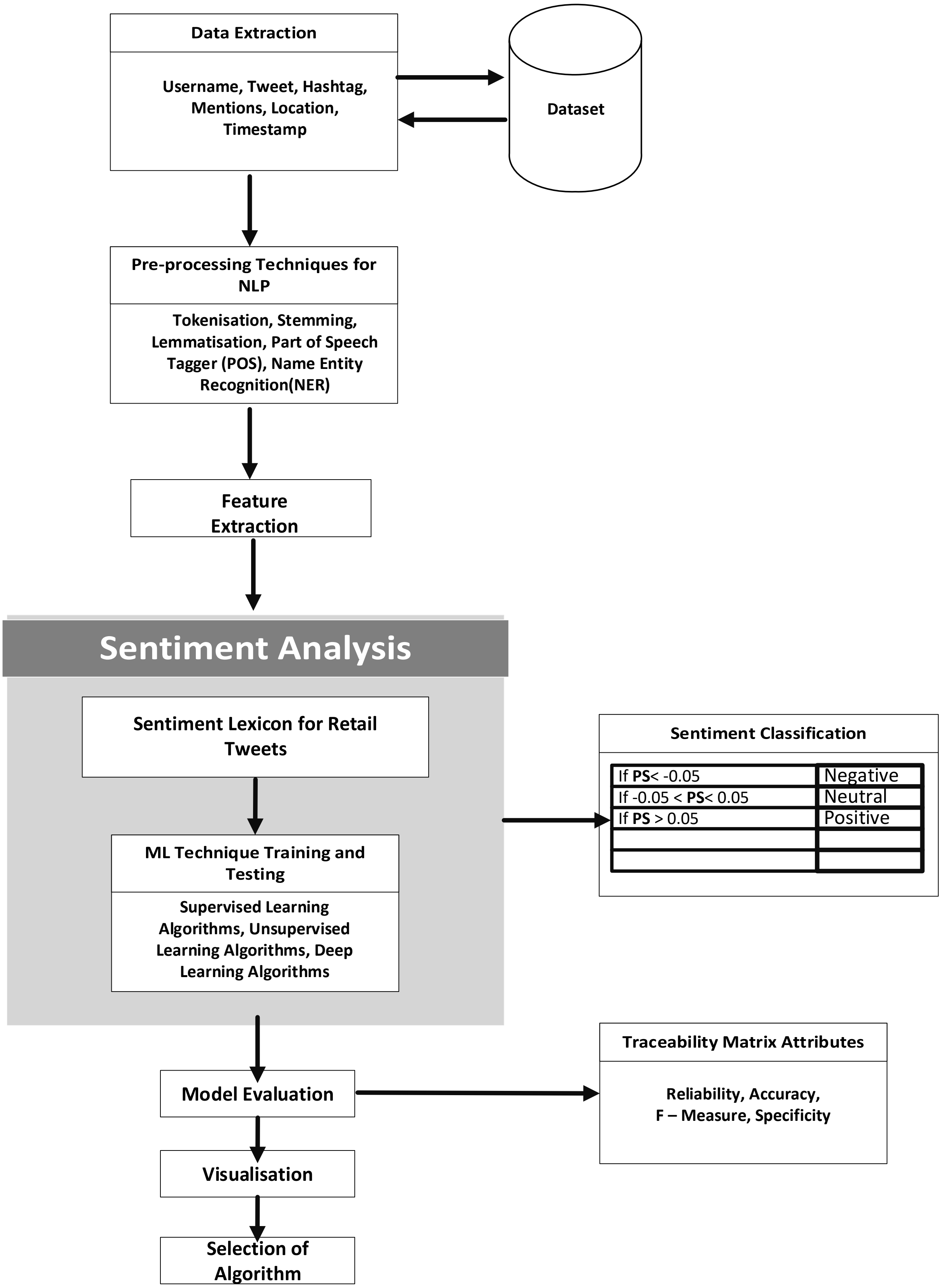


Figure 3.2: Proposed Conceptual Framework

3.5.1 Data Extraction

The dataset, which comprises the key attributes (i.e., variables) that are critical for building the sentiment analysis model, was extracted from the prepared dataset in the previous section. These attributes are *username, tweet, hashtag, mentions, location and timestamp.* To conduct the data extraction task, pandas, a Python library, was utilised in this study.

Pandas and its methods for EDA have been described earlier in Section 3.4.2. Some examples of methods in pandas that are applicable for data extraction tasks are *read_csv()* and *query().* The *read_csv()* is useful for extracting data from a CSV file into a dataframe. The *query()* method allows the querying of the dataframe with structured query language (SQL) syntax.

A total of 27233 tweets were extracted for the subsequent phases of the conceptual framework.

3.5.2 Pre-processing Techniques for NLP

The five pre-processing methods selected are specifically for NLP (Section 2.13). The pre-processing methods must be done in a certain order to produce optimal results. This further prepares the data for the next feature extraction phase of the conceptual framework. The order is tokenisation, stemming, lemmatisation, part-of-speech tagging (POS) and named entity recognition. Once the five pre-processing methods for NLP are completed, the data is moved onto the 3rd stage of the conceptual framework which is feature extraction (Section 3.5.3). The five pre-processing methods for NLP are described and presented in Algorithm 2 (Tokenisation), Algorithm 3 (Stemming), Algorithm 4 (Lemmatisation), Algorithm 5 (POS Tagging) and Algorithm 6 (NER).

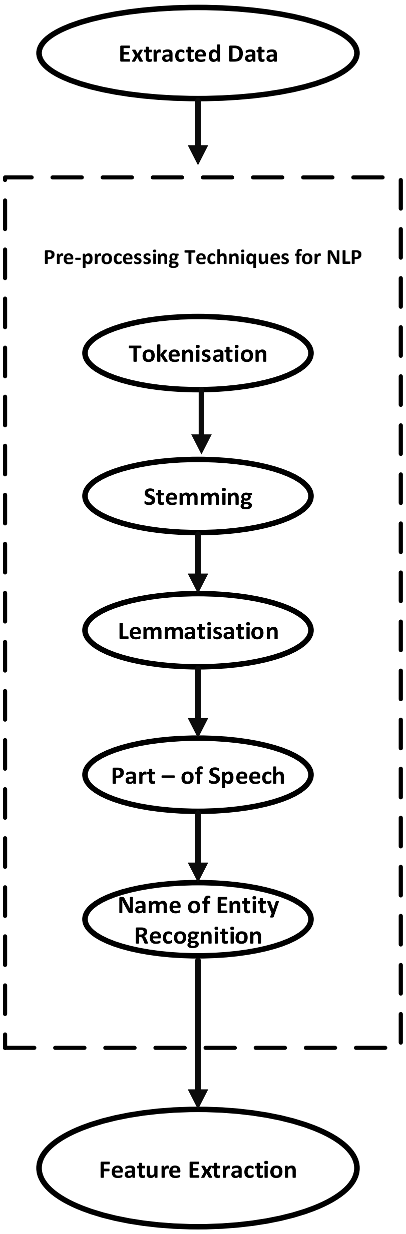


Figure 3.3: Pre-processing Techniques for NLP

Figure 3.3 displays the order of the NLP pre-processing methods that the data must go through. Each pre-processing method prepares the data for the following pre-processing method to receive the best result. The extracted data will first need to go through tokenisation, using the method *tokenise()* to break the text into tokens. To prepare the data to go through stemming utilise the *PorterStemmer()* method, which reduces the text into its stem form, followed by lemmatisation, which utilises the *WordNetLemmatiser()* function that aims to reduce the inflection forms to a basic form. These methods prepare the data ideally for part-of-speech tagging *(nltk.pos_tag(tokenise),* which are tokens implemented in the pos_tag method and named entity recognition using the *pos_ner_text()* function to identify noun, adjective, adverb and verb within a tweet, as well as the location, businesses, people and time expression within a tweet. These are important components for NLP and for this study.

3.5.2.1 Tokenisation

Tokenisation involves breaking down textual data into individual units, which could be words or subwords. The pseudocode for tokenisation of the dataset earlier extracted in this study is presented in Algorithm 2. The functions and iterative procedures in the algorithm were implemented using appropriate methods in the Natural Language Tool Kit (NLTK) Python library. The *word_tokenize()* method in NLTK generates word tokens from any input textual dataset. Tokenised data is the output of Algorithm 2.

Algorithm 2: Tokenisation

Input: Extracted Dataset
Output: Tokenised Dataset

```
function Tokenise (Dataset) //initialise the tokDatset as an empty string array
 for character in Text do

   data = input() //Get the string of text and split it into tokens
    tokens = data.split()
     end for

         output ("List of tokens:", tokens) // verify the list is populated by displaying it

          for token in tokens  //Loop through the tokens and output each token
         output ("Token:", token)
        tokDataset = tokDataset.append(token)
      end for
   return tokenised dataset
end function
```

3.5.2.2 Stemming

In NLP, stemming reduces words to their root form, known as “stem”. It removes prefixes or suffixes from words so that different variations of a word are mapped to the same root form. To perform stemming in this study, the Porter stemmer algorithm (Porter, 1980) was adopted among several other available options. It is a widely used stemming algorithm for NLP because of its high effectiveness in terms of mildness, intuitiveness and simplicity. This algorithm is available in the Python NLTK library and it contains the stem() method for stemming input tokens. The pseudocode that was implemented for the stemming task in this work is succinctly presented in Algorithm 3. Tokenised data, which is the output from Algorithm 2, will now become the input for Algorithm 3.

Algorithm 3: Stemming

Input: Tokenised Data
Output: Stemmed Data

```
function Stemming (tokenised Data) //pass the tokenised data as a parameter into the function

  load PorterStemmer // This is the algorithm from NLTK library for stemming

    stemDataset = "" //Initialise the stemDataset as empty string

     for word in tokenised data do

       ps = PorterStemmer() //convert tokens into their corresponding stems
         stemming_string = join([wnl.PorterStemmer(words) for words in Tokenised Data])
//Convert the text into stem

          end for

//loops through the stemming string and output each stem

             for each stems in stemming)_string  //Loop through the stemming and output each stem

        output(stem)


     return Stemming

 end function
```

3.5.2.3 Lemmatisation

Lemmatisation reduces tokenised words to their lemmas, which are their root forms, in order to further generate standardised and normalised texts. This pre-processing task for NLP goes beyond just the removal of prefixes and suffixes (as was done through stemming) but also inherently factors in the word's part-of-speech (POS) as well as its context to ensure the generation of accurate lemmas.

The Python NLTK library contains the WordNetLemmatiser, which was adopted for the lemmatisation task in this work. WordNetLemmatiser utilises WordNet, an English lexical database and incorporates part-of-speech (POS) tags for each word in order to lemmatise the stemmed tokens that it receives as inputs. The POS used within the WordNetLemmatiser is critical because the lemma of a given word may vary based on its grammatical role (such as adjective, adverb, verb or noun) within the sentence (Saranya and Usha 2023).Furthermore, in order to utilise WordNetLemmatiser, appropriate methods are used in the CorpusReader's. Algorithm 4 contains the pseudocode that was implemented using Python within the PyCharm

IDE for the lemmatisation task in the study. Stemmed data, which is the output from Algorithm 3, will now become the input for Algorithm 4.

Algorithm 4: Lemmatisation

Input: Stemmed Data
Output: Lemmatised Data

```
function Lemmatisation (Stemmed Data)
load WordNetLemmatizer // load the lemmatization library from NLTK
 lemmatised dataset = ‘’ //initalise lemDataset as an empty string
   for character in Stemmed dataset
     lemmatizer = WordNetLemmatizer()

//Lemmatised the Stemmed Data
        for word, tag in wordnet_tagged:
            if tag is None:   # if there is no available tag, append the token as is
               lemmatized_sentence.append(word)
             else:         # else use the tag to lemmatize the token
          lemmatized_sentence.append(lemmatizer.lemmatize(word, tag))
        lemmatized_sentence = " ".join(lemmatized_sentence)
    end for
  end for
output (lemmatized_sentence)

return Lemmatised Dataset
end function
```

3.5.2.4 Part – of – Speech

Once tokenisation, stemming and lemmatisation is completed POS – Tagging is introduced. To apply POS, the lemmatised data must be passed through. First, the text must be tokenized. POS tagging employs features such as the previous word, the next word, whether the first letter is capitalized. After the tokenization procedure, NLTK includes a method that retrieves pos tags*(nltk.pos_tag(tokenise)* which import the word tokenize object and NLTK library. To use this, a sentence must first be divided into tokens. After dividing into tokens, tagging operates. The research used the POS-Tag() procedure after tokenising the words in a sentence. For instance, the word "The" bears the tag "DT." The word "feet" is marked as "NNS.". POS – tagging results in the categorisation of words in a Tweet that are noun, adjective, adverb and verb. The pseudocode for POS – tagging is presented in Algorithm 5. Lemmatised data which is the output from Algorithm 4 will now become the input for Algorithm 5.

Algorithm 5: POS - Tagging

Input: Lemmatised Data
Output: POS – Tagged Dataset

```
function POSTagged (Lemmatised Data)
  for character in Text
    pos_tagged = nltk.pos_tag(nltk.word_tokenize(sentence))  // Find the POS tag for each
token

       output (POS – Tagged Data)
    return POS – Tagged Data
end function
```

3.5.2.5 Named entity recognition.

Once tokenisation, stemming, lemmatisation and POS – tagging have been implemented in the mentioned order, named entity recognition (NER) is introduced. The POS Tagged data is passed through the pos_ner_text() functionality to proceed with the NER pre-processing technique. The research imported Spacy to recognise named entities. Extracting the named entities from a text helped sort unstructured data and detect important information, which is crucial when dealing with large datasets. NER also requires the text to be tokenised first, and only then can the NER functionality be applied. Spacy was first applied in named entity recognition. The two steps that make up the fundamental process of NLP are as follows: Identification of the things in the text and classification into various categories in this implementation, NLTK and Spacy libraries were used to accomplish named entity recognition, which returned the NER_tree () method.

The result of NER is to determine if the word is a named entity or not, which could be places, organisations, people, or temporal expressions. To achieve the optimal output, pre-processing techniques for NLP have been executed in the specified sequence: tokenisation, stemming, lemmatisation, POS-tagging, and NER. Each technique will pass through its result to the next technique to improve the output. The pseudocode for NER is presented in Algorithm 6. POS Tagged data which is the output from Algorithm 5 will now become the input for Algorithm 6.

Algorithm 6: NER

Input: POS – Tagged Data
Output: NER_tree Dataset

function NER (POS Tagged Data)
  for character in Text
    NER_tree = ne_chunk(pos_tag(word_tokenize(pos_ner_text)))**output** (POS – Tagged Data) // Break the POS_Tagged data into tokens and implement NER

  return NER_tree
end function

3.5.3 Feature Extraction

Feature extraction in sentiment analysis transforms pre-processed text (i.e., lemmatised tokens in this study) into numerical features, which can be used to train ML models in order to predict the sentiment of input texts. Among other options in the literature, for example, bag of words (BoW) or term frequency-inverse document frequency (TF-IDF), word embedding is selected for the feature extraction task in this study. Word embeddings are dense and fixed-length vector representations of words that capture their semantic relationships. They are more powerful than BoW or TF-IDF, but more computationally expensive. The quality of the selected word embedding method determines the performance of the subsequent ML model. Based on comparative quality performance, the selected word embedding method in this work is Word2Vec (Marreddy and Mamidi 2023).

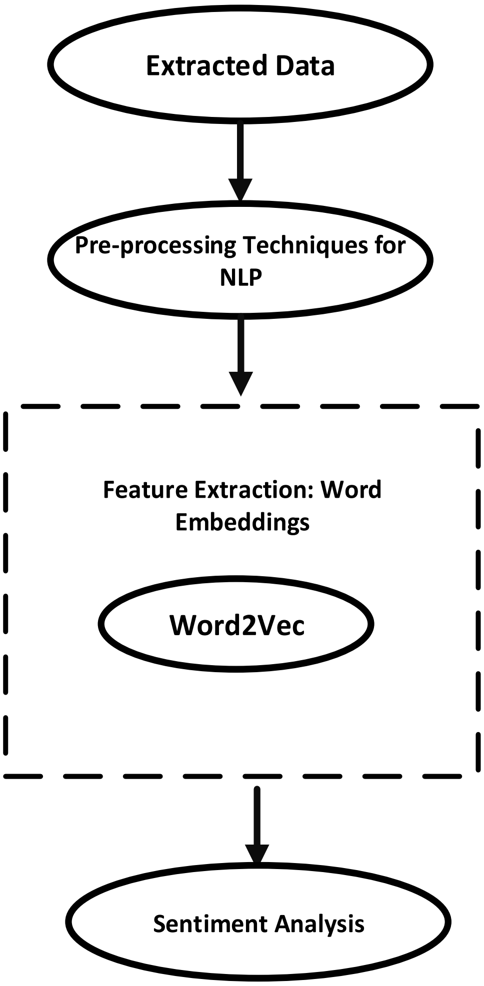


Figure 3.4: Feature Extraction

Figure 3.4 illustrates the process that would be followed to perform Feature Extraction. Word2Vec is a neural network-based method trained on a large corpus of text for generating word embeddings. Continuous bag-of-words (CBOW), which is one of the two Word2Vec architectures, predicts the fixed-length vector representation of the input words given the surrounding context words (Mikolov *et al.* 2013). Word2Vec includes continuous bag of words (CBOW), which determines the word that closely fits the context of the various input words utilising the *Sequential()* method. The CBOW training goal is to increase the likelihood that it will correctly predict context words given the target word. The goal can be expressed as the average log probability for the string of words w1, w2,... wT which is implemented utilising the training context's size. This probability is defined by the simple skip-gram formulation using the softmax function in Equation 3.1, which is the training context size. The model is trained on skip-grams, which are n-grams that allow tokens to be skipped. The context of a word can be represented through a set of skip-gram pairs of *target_word(), context_word()* where *context_word()* appears in the neighbouring context of *target_word().* The training objective of the skip-gram model is to maximise the probability of predicting context words given the target word.

$\frac{1}{T}\sum_{t=1}^{T} - \sum_{c<j<c,j=0} logp\ (W_{t+j}|\ W_t\ )$ *Equation 3.1*

The Gensim Python library contains a module named genism.models with classes and methods for different kinds of NLP tasks. The module contains the Word2Vec CBOW model, which was adopted for the feature extraction phase of the conceptual model in this work. Algorithm 3.1 presents the pseudocode that was implemented in the PyCharm IDE.

Algorithm 3.1: Word2Vec - CBOW

Input: PrepDataset
Output: Extracted Features

```
function Word2Vec-CBOW(prepDataset)
 load genism.models // load appropriate module from NLTK
  load Word2Vec // load Word2Vec from genism.models module of NLTK
    //Initialise the extracted features as a vector array
    extFeatures = []
//Re-train the Word2Vec model using the CBOW architecture
      model = Word2Vec(
          sentences= lemDataset // lemDataset is the list of pre-processed lemmatised
tokens
           vector_size=150,  // Indicate the dimensionality of the word vectors
        window=2,    //Indicate the maximum distance between the current and predicted
word
      sg=0,     // Set to 0 for CBOW (skip-gram: sg=1)
      min_count=5,   // Ignore words with a frequency less than the indicated value
    workers=3,  //indicate the number of CPU cores for training based on your system
   epochs=50   // Number of iterations over the corpus
)
# Generate the extFeatures using the trained model
for word in prepDataset
    extFeatures = extFeatures .append(model.wv['word'])
end for
return extFeatures
end function
```

3.5.4 Sentiment Analysis

Once feature extraction is conducted, this is followed by sentiment analysis. This study selected the hybrid approach for sentiment analysis. The hybrid approach accomplishes sentiment analysis in two stages: first, it searches for polarity of words in a pool of words that are previously predefined using the sentiment lexicon using the SentiWordNet() method, and second, it trains the machine learning algorithm, providing better accuracy of sentiment analysis using various ML algorithm methods.

3.5.4.1 Sentiment Lexicon for Retail Tweets

Sentiment lexicon is a list of lexical qualities that are typically classified as either positive, negative, or neutral depending on their semantic orientation. In this work, a Sentiment Lexicon for Retail Tweets was created through lexicon-based sentiment analysis by labelling and annotating the extracted numerical features (i.e., word2vec embeddings of the lemmatised tokens in Section 3.5.3) with the appropriate sentiment polarities. This first stage is the sentiment analysis component of the conceptual framework. In order to achieve this, the TextBlob Python library was utilised for automatic sentiment polarity labelling, as was done in an earlier study (Aljedaani *et al.* 2022). TextBlob is a popular Python library for lexicon-based sentiment analysis. It is built on top of the NLTK and provides sentiment polarity in text as either positive, negative, or neutral. It uses polarity values that range from -1 to 1 to indicate sentiment. A positive polarity sentiment is assigned a polarity value greater than 0, a neutral polarity sentiment is given a polarity value of 0 and a negative polarity sentiment is assigned a polarity value less than 0. Furthermore, it was reported in (Aljedaani *et al.* 2022) that TextBlob gives better performance than the Valence Aware Dictionary for Sentiment Reasoning (VADER) and Affective Norms for English Words (AFINN) for automatic annotation of sentiment lexicons in a similar problem domain to this study.

Algorithm 3.2 presents the pseudocode for the lexicon-based sentiment analysis, which generates the labelled Sentiment Lexicon for Retail Tweets in this study. This pseudocode was implemented in the PyCharm IDE. The labelled Sentiment Lexicon for Retail Tweets at this stage is a vital contribution of this study to the NLP body of knowledge. It provides a labelled dataset for the development of the supervised ML-based models in the second stage. Notably, the outcome of the second stage will validate or annul the hypothesis that ML-based models (either supervised or unsupervised) provide improved performance in sentiment analysis of retail tweets. However, the unsupervised ML models do not require a labelled Sentiment Lexicon but rather an unlabelled Sentiment Lexicon.

Algorithm 3.2: Sentiment Lexicon for Retail Tweet

Input: Extracted Features through Word Embedding (i.e word2vec) //exfeatures
Output: SentimentLexicon //sentLexicon

```
function SentLexicon(extFeatures) //load necessary library and module
 load textblob
  load TextBlob //load TextBlob module from the textblob library
  //Initialise an empty dictionary to store labelled sentiment scores
   sentLexicon = {}
    //Iterate through each feature in the lexicon
      for feature in extFeatures do
        //Calculate the sentiment score using TextBlob
            blob = TextBlob(value)
                sentScore = blob.sentiment.polarity
                //Assign a sentiment label based on the sentiment score
                               if sentScore > 0 do
                        sentLabel = "positive"
                  else if sentScore < 0 do
              sentLabel = "negative"
           else
        sentLabel = "neutral"
    end if
   // Store the feature, sentiment label, and sentiment score in the lexicon
    sentLexicon[feature ] = {
         "sentLabel": sentLabel,
         "sentScore": sentScore
     }
end for
return sentLexicon
end function
```

Table 3.3 Presents an example of output for Sentiment analysis using Sentiment Lexicon Approach

Table 3.3: Example of output for Sentiment Analysis

Text Data	Total Positive Score	Total Negative Score	Sentiment Polarity
['I', 'am', 'connect', 'with', 'shopping', 'walking', 'and', "it'", 'GOOD', 'Connect', 'each', 'other', 'with', 'friends', 'Jeans', 'Shirt', 'Music', 'connect', '2020', 'South Africa']	1.0	0.75	Positive
	0.0	0.0	Neutral
['I', 'am', 'child', 'woman', 'stock', 'good', 'variety', 'like', 'relaxing']	0.375	0.125	Positive
['I', "don't", 'enjoy', 'the', 'service', 'it', 'was', 'disgust', 'and', 'all', 'the', 'shoppers', 'was', 'upset']	0.375	0.125	Negative

3.5.4.2 Training and testing of machine learning models for sentiment analysis

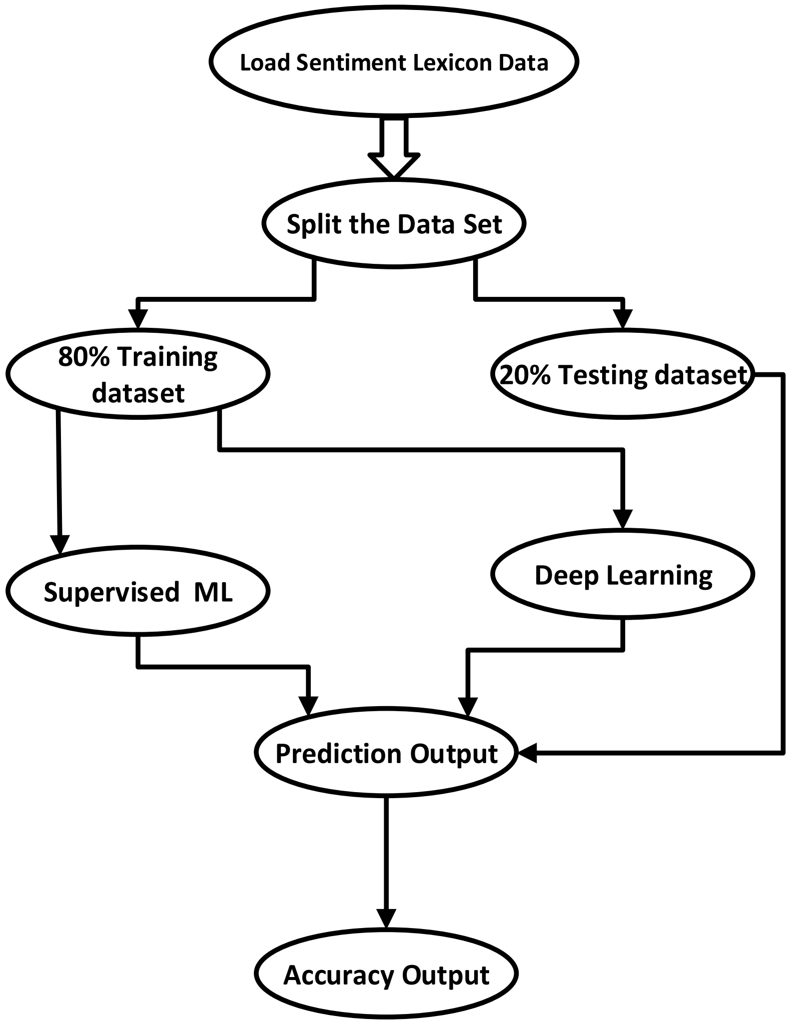


Figure 3.5: The ML techniques required to be trained for sentiment analysis

The block diagram in Figure 3.5 provides an overview of the training and testing of the machine learning model for sentiment analysis in this study based on the Sentiment Lexicon generated in Section 3.5.4.1. The dataset obtained from the Sentiment Lexicon is split into training and testing partitions of 80% and 20%, respectively similar to the splitting ratio utilised (Rodrigues *et al.* 2022). The training partition (80%) is then used to train a rich set of supervised (shallow and deep) and non-supervised ML models this approach allows for a wide exploration of model options with the aim of selecting the one that delivers optimal performance. After proper model hyperparameter configuration and training, the testing partition (20%) is utilised to evaluate the generalisation performance of the models. This will provide an insight on how the models will perform when deployed and tested with an out-of-sample dataset from customer tweets.

The scikit–learn Python ML library was used to carry out the configuration and experimental training of all the ML models. Other Python support libraries that were used are numpy and pandas. The splitting of the dataset into training and test sets can be achieved with the train_test_split() method, while the fit() and predict() methods are used to implement training and testing, respectively. The design and configurations of the various ML models in this study (as shown in Figure 3.6) are presented hereafter.

A) Supervised machine learning and deep machine learning models

Supervised deep machine learning is a subset of ML with multiple layers that has gained prominence due to its ability to automatically discover and represent intricate features within large, labelled datasets. Different variants of deep learning have achieved remarkable success in various fields, including NLP. For this study, the supervised deep ML architectures that were selected are deep neural networks (DNN), convolutional neural networks (CNN) and recurrent neural networks (RNN). The theoretical descriptions of these selected deep MLs have been presented earlier in Chapter 2. For this study, the architectural and hyperparameter configurations for training the DNN, CNN and RNN models with the labelled sentiment lexicons are presented in Table 3.4. During training using different variants of the optimisation algorithm (e.g., scale conjugate gradient (SCG)) and the labelled dataset, the parameters of the models (i.e., weights and biases) are optimally fine-tuned until the accuracies, losses, and all the other performance metrics achieve acceptable results. The training and testing results for these models are reported and discussed in Chapter 4.

Table 3.4: Architectural and Hyperparameter Configurations of the Supervised Deep Machine Learning Models

Model	Hyperparameter	Description
DNN	$\operatorname{def} sigmoid(x)$	Activation function which defines the output of a node.
	$Return\ 1/(1 + np.\exp(-x))$	Optimisation, the process of selecting the best hyperparameter.
CNN	N_c	Number of convolutional layers
	$kN_{i,i\in N_d}$	Number of kernels of each convolutional layer
	$kS_{i,i\in N_d}$	Kernel size in each convolutional layer
	$aF_{i,i\in N_d}$	Activation function in each convolutional layer
RNN	$R_{I,I\in N_d}$	Weight regularisation in each dense layer
	$R_{dropout}$	Dropout
	S_{batch}	Batch size
	L_{Rule}	Learning Rule

Algorithm 3.3: Supervised ML Model and Deep Learning Training and Testing

Input: Labelled Sentiment Lexicon for Retail Tweet//lblSentilexicon
Output: Trained and Tested Model //finalModel

function TrainTestSupML (lblSentLexicon)
//load necessary libraries
load dataProcessingLibrary // e.g., appropriate data processing modules from scikit-learn
load mlLibrary // e.g., appropriate ML modules from scikit-learn
// load the lblSentLexicon dataset
data = dataProcessingLibrary.load_dataset("lblSentLexicon ")

//Split the dataset into the features (X) and target labels (y)
X, y = dataProcessingLibrary.split_features_and_target(data)
// Split the data into training (80%) and testing (20%) sets
X_train, X_test, y_train, y_test = dataProcessingLibrary.train_test_split(X, y, train_size=0.8)

// Create the model
initModel = mlLibrary.create_model() // Select the ML algorithm from scikit-learn
// Hyperparameter tuning
finalModel = mlLibrary.tune_hyperparameters(initModel, X_train, y_train)
// Model training and testing
finalModel.fit(X_train, y_train)
yPredFinal = mlLibrary.predict(finalModel, X_test)
modelAccuracy = mlLibrary.calculate_accuracy(y_test, yPredFinal)

return finalModel
end function

B) Unsupervised machine learning models

In unsupervised ML, the algorithms automatically learn patterns from an unlabelled dataset in order to find the underlying structure and group similar data points together. As shown in Figure 3.3, the two unsupervised ML algorithms that were explored with the unlabelled version of the Sentiment Lexicon dataset are k-means clustering (KCM) and fuzzy C-means (FCM).

KCM is a simple unsupervised ML algorithm that groups similar data points together. This is done through iterative assignment of each data point to the cluster with the nearest mean (i.e., centroid), and the iteration terminates when no data point is reassigned to another cluster. On the other hand, fuzzy C-means clustering (FCM) is a soft clustering algorithm in which each data point could belong to multiple clusters with varying degrees of membership. This is different from KCM (a hard clustering algorithm) in which each data point is allocated to a single cluster (Chamarczuk *et al.* 2020).

Table 3.5 presents the architectural and hyperparameter configurations for training KCM and FCM with the unlabelled training dataset. In this work, the K-means class and skfuzzy.cmeans() function in Python Scikit-learn library were used for the implementation of KCM and FCM, respectively. The pseudocodes for that were implemented for these algorithms are presented in Algorithm 3.4.

Table 3.5: Architectural and Hyperparameter Configurations of the Unsupervised Machine Learning Models

Model	**Hyperparameter**	**Description**
KCM and FCM	$n\_clusters$	Determining how many clusters are present in the dataset
	init	Indicates random initialisation method
	n_init	Number of times K-Means would run with different sets of starting points.
	max_iter	Number of iterations taken
	error	Number of errors within an iteration
	maxiter	Maximum number of iterations

Algorithm 3.4: Unsupervised ML Model Training and Testing //KCM and FCM

Input: Unlabelled Sentiment Lexicon for Retail Tweet //ulblSentLexicon
Output: Trained and Tested KCM and FCM Models // kcmModel and fcmModel

```
function TrainTestKCM(ulblSentLexicon)
//Load necessary libraries
 load dataProcessingLibrary // e.g., from scikit-learn library
  load KCM library // e.g., sklearn.cluster  from scikit-learn library

    // Load the dataset
      data = dataProcessingLibrary.load_dataset("ulblSentLexicon")
       //Choose the number of clusters (K)
           K = determine_optimal_k(data)

         # Train the K-Means model
         kcmModel = KCM.KMeans(n_clusters=K)
        kcmModel.fit(data)

    // Obtain cluster assignments for each data point
  clusterAssign = kcmModel.labels()
// Analyze the cluster centroids
clusterCenters = kcmModel.cluster_centers()
return kcmModel
end function
```

```
function TrainTestFCM(ulblSentLexicon)
// load necessary libraries
load dataProcessingLibrary // e.g., from scikit-learn library
load skFuzzy //load skFuzzy from scikit-learn library

// Load the dataset
data = dataProcessingLibrary.load_dataset("ulblSentLexicon")

// Configure the FCM Hyperparameters
c = determine_optimal_c(data)
m = 2  // fuzziness parameter (m) is often set to 2 for FCM
error=0.005 // you can adjust this based on expected error value
maxiter=1000 // you can adjust this based on expected number of iterations
```

```
//Train the FCM model with the defined hyperparameters
fcmModel = skFuzzy.cmeans(data.T, c, m, error, maxiter)

# Obtain and display the cluster memberships and Centers
clusterMemberships = fcmModel [1]
display("Cluster Membership =", clusterMemberships)
clusterCenters = fcmModel [0]
display("Cluster Centers =", clusterCenters)

return fcmModel
end function
```

Table 3.6 displays the technologies selected for this study, along with the description and source.

Table 3.6: Summary of technological tools for the methodology

Name of technology	Purpose
Keyhole	Keyhole is used to compute or establish the effects of a brand or trend on Twitter, Facebook, and Instagram.
Principal Component Analysis	This method performs a direct mapping of the data to a lesser-dimensional space in a way that maximises the variance of the data in the low-dimensional representation (Hasan and Abdulazeez 2021).
Scikit-learn	Scikit-learn is an open-source data analysis library and it has become the standard for ML.
Sentiment Lexicon	Sentiment lexicon is a list of lexical features which are generally labelled according to their semantic orientation as either positive, negative, or neutral (Chiong, Budhi and Dhakal 2021).
Talkwalker	Talkwalker to identify or locate significant rumours and misleading facts.
TrackyMHashtag	TrackMyhashtag focuses on monitoring social media campaigns, gathering data, and making various deductions from the data acquired.

Name of technology	Purpose
VADER	VADER, also known as the Valence Aware Dictionary and Sentiment Reasoner, is an economical rule-based framework for analysing the sentiment of social media text (Pano and Kashef 2020).

3.5.4.3 ML approach sentiment analysis for retail Tweets

Once the selected ML techniques have been trained and tested, it is then used to further decipher the data. The selected ML techniques are supervised ML, unsupervised ML and deep learning. Algorithms within each technique were further selected based on the research done in the literature review (Section 2.22).

3.5.4.4 Supervised ML

A supervised ML approach was used when labelled data was available for the data analytic model. During training, a data set with labels is fed through a classification algorithm, which produces a model. The test data is then entered into the model, which provides a prediction.

3.5.4.5 Naïve Bayes

Naïve Bayes is a classification algorithm with a high probability of success. It does not consider the position of a term in the sentence; therefore, each word is treated independently. The probability of each phrase that corresponds to a label is calculated using the Bayes theorem by Naïve Bayes. The prior probability of the label in the dataset is p(label). The prior probability of a feature associated with a label is p(feature|label). The prior probability of a feature occurring is p(feature). SentiWordNet lexicon with Naïve Bayes was utilised in the study to improve the classification of the Twitter dataset by providing the score of positive and negative tweets.

The Naïve Bayes classifier cannot establish a semantic association between words since each word is treated independently, whereas the Bayesian network can. The words depend heavily on each other in a Bayesian network. The Bayesian network expresses dependency as an acyclic directed graph, with each node representing a variable and edges representing the relationship between variables.

3.5.4.6 Support Vector Machine

Support Vector Machine (SVM) is set up to solve binary classification issues. Its goal is to find the optimal hyperplanes, which serve as a separator to characterise the decision boundaries between data points from different classes. A hyperplane should be chosen that can maintain the maximum distance between two support vectors of different classes. The SVM is capable of handling both linear and non-linear classification tasks. For classification, SVM is employed with various weighting techniques such as TF-IDF, term occurrence, and binary occurrence. The study chose the chi-square as a feature for dimensional reduction (DR) and noise reduction. The study demonstrated that using chi-square feature selection with SVM improves accuracy with the help of the experiment.

3.5.4.7 Artificial Neural Network

Artificial Neural Networks (ANN) mimic the neural structure of the human brain. The basic unit of the neural network is the neuron. ANN comprises an input layer, a hidden layer, and an output layer (Vanhaeren *et al.* 2020). A vector "a (i)" is given as an input to neuron; vector denotes the frequency of a word in a document. There is a weight "A", corresponding to each neuron, which is used to calculate the function. Neural networks use a linear function:
x (i) = A (a (i)). The sign of x (i) is used to classify the class. ANN's training model consists of two steps: forward propagation and backward propagation. In forward propagation, the input is given at the input layer of neurons, which is multiplied by the weights, which are random numbers (Wong, Wu and Li 2008). Functions are used to normalise the output value between 0 and 1.

3.5.4.8 Decision Tree

The decision tree is a tree-like structure in which the non-terminal nodes represent features and the terminal nodes represent labels. The path is chosen based on a criterion. This is a recursive procedure that will lead to a terminal node that will assign a label to an input. A decision tree is an effective tool for sentiment analysis since it can handle vast amounts of data (Wong, Wu and Li 2008). The training data is divided into hierarchical groups via a decision tree. A condition based on the attribute value is used to divide the data. The presence or absence of a word determines the condition. The division process is repeated until the terminal nodes represent the limited number of attributes needed for categorisation.

3.5.4.9 Unsupervised ML

When the reliability of labelled data is a problem, this method is applied. Unlabelled data is easier to collect than labelled data. Each category's keyword lists are used to categorise the sentence. The unsupervised approach is more convenient for analysing domain-dependent data. During the testing, it was discovered that sentiment analysis using the unsupervised approach, resulted in tweets being clustered into positive and negative clusters using the spectral clustering approach.

3.5.4.10 Fuzzy C and K-Means clustering

With fuzzification, the positive and negative scores of each tweet obtained from the second phase are fuzzified using the triangular membership function. When the triangular fuzzy membership is used, each linguistic term T involves three key points, d, e, and f and is associated with the change in pattern of the fuzzy membership. A membership function (MF) for a fuzzy set S on the universe of discourse X is defined as $\mu S:X \rightarrow [0, 1]$, where each element of X is mapped to a value between 0 and 1. The equation for triangular function defined by a lower limit d, an upper limit f, and an intermediate value e, where d <e < f follows. The K-means technique in ML starts with a first group of randomly picked centroids, which serve as the starting points for each cluster, and then performs iterative (repetitive) calculations to optimise the centroids' placements.

3.5.4.11 Deep Learning

The most important concept involved in deep learning is the use of deep neural network to train the nodes for extracting complex features from the available information with limited contribution. These algorithms adapt easily and do not require manual input, giving them the ability to learn new complex features automatically. However, the drawback of using deep learning approaches is that deep learning requires a vast collection of data for high-efficiency output. The approach is utilised to produce a set of professionally nominated emotions from the text and group these emotions by using vocabularies or lexical resources like WordNet. Much of the work cited above focuses on identifying the prior polarity of the terms or phrases to use before assigning the sentiment polarity to the word using the WordNet lexical resource.

3.5.4.12 Convolutional Neural Network

Convolutional Neural Network (CNN) is a type of feedforward neural network that was developed for use in computer vision. Its design is based on the human visual cortex, which is a visual system in the brain of animals. The visual cortex comprises a large number of cells that sense light in small, overlapping sub-regions of the visual fields known as receptive fields. Over the input space, these cells act as local filters. CNN is made up of several convolutional layers, each of which performs the function that the cells in the visual cortex performs.

3.5.4.13 A Deep Neural Network

A Deep Neural Network (DNN) is a collection of layered neural networks, or networks with multiple layers. FF-DNN: FF-DNNs, also known as multilayer perceptron's (MLPs), are DNNs with more than one hidden layer and only one forward direction, as the name suggests (no loopback). Both categorisation and prediction are possible with these neural networks. This study employs a classification technique for spoken LID. The input and output nodes will match the input features and output classes when the FF-DNN is employed as a classifier. Weights, biases, nonlinear activation, and backpropagation are the most significant ideas in an FF-DNN. DNNs are an extension of traditional ANN. There are two major differences between DNNs and regular neural networks. One or two hidden layers are insufficient for ordinary neural networks. DNNs, on the other hand, have a lot of hidden layers. The Google Brain project, for example, uses a neural network with millions of neurons. DNNs come in a variety of models, including DNNs, CNNs, RNNs, and LSTMs. Attention-based networks, which focus on specific regions of a DNN, have even been developed as a result of recent research.

3.5.4.14 Recurrent Neural Networks

Recurrent Neural Networks (RNNs) work on the idea of preserving a layer's output and feeding it back into the input in order to predict the layer's output. The different activation functions, weights, and biases will be standardised by the RNN, ensuring that each hidden layer has the same characteristics. Rather than constructing numerous hidden layers, it will create only one and loop over it as many times as necessary. The information in RNNs cycles through a loop that passes through the middle-hidden layer.

Table 3.7 demonstrates the selected ML algorithm for each ML technique

Table 3.7: Select ML algorithm for each ML technique

Supervised	Unsupervised	Deep Learning
Support Vector Machine	K – Means Clustering	Multilayer Perceptron Neural Network
Naïve Bayes	Fuzzy C Means	Convolutional Neural Network
Bayesian Network		Long short-term memory
Artificial Neural Network		Adversarial Network
Decision Tree		Restricted Boltzmann Machine
		Deep Belief Network

3.5.5 Model Evaluation and Visualisation

To evaluate the performance of the supervised ML models, the performance metrics utilised are accuracy, precision, recall, F1-measure, and specificity. These metrics have been described with relevant mathematical equations (Chapter 2, Section 2.10). However, the evaluation metric that was used for the unsupervised ML models (KCM and FCM) is the silhouette score. It is an important metric for evaluating clustering algorithms by quantifying the cohesion within clusters and the separation between clusters. The score is a single value, which ranges from -1 to 1, with a score close to 1 indicating strong clustering. It was implemented in this work using the Python Sklearn's metrics.silhouette_score() method.

Visualisation of the training and testing results both for the supervised and unsupervised ML models was implemented using the Matplotlib Python library. This is done in order to graphically illustrate the performance of the models, which aids in the selection of the model with the best performance for the ML-based sentiment analysis task within the conceptual framework. Matplotlib supports line, bar, scatter, array and field plots. It also supports statistical plots such as histograms, boxplots, violin plots, and plots with unstructured coordinates. Furthermore, the Matplotlib visualisation tool was utilised to display the Twitter extracted data with attributes such as tweets, hashtags, location, date of tweet and time of tweet. The pseudocodes for the performance evaluation and visualisation of the supervised and unsupervised models are shown in Algorithms 3.4 and 3.5. All the evaluation metrics results and visualisation plots are presented and discussed in Chapter 4.

Algorithm 3.5: Supervised Model Evaluation and Visualisation

Input: Test Target and Predicted Target // yTestTarget, yPredTarget
Output: Metrics Values and Visualisations

function supMLEvalVisual(yTestTarget, yPredTarget)
// load necessary libraries
load perfMetric //e.g., load perfMetric from sklearn.metrics library
load visualPlot // e.g., load visualPlot from matplotlib.pyplot library

//Calculate the confusion matrix and other supervised model metrics
confMatrix = specificity.confusion_matrix(yTestTarget, yPredTarget)
accuracy = perfMetrics.accuracy_score(yTestTarget, yPredTarget)
precision = perfMetrics.precision_score(yTestTarget, yPredTarget)
recall = perfMetrics.recall_score(yTestTarget, yPredTarget) //Sensitivity = recall
f1_measure = perfMetrics.f1_score(yTestTarget, yPredTarget)
specificity = perfMetrics. specificity (yTestTarget, yPredTarget)

//Generate the printout of the metrics
display("Accuracy:", accuracy)
display("Precision:", precision)
display("Recall (Sensitivity):", recall)
display("F1-Measure:", f1_score)
display("Specificity:", specificity)

// Visualize and the metrics (e.g., using a bar chart)
metricsNames = ['Accuracy', 'Precision', 'Recall', 'F1-Measure', 'Specificity']
metricsValues = [accuracy, precision, recall, f1_measure, specificity]
visualPlot.bar(metricsNames, metricsValues)
visualPlot.xlabel('Metrics')
visualPlot.ylabel('Value')
visualPlot.title('Supervised ML Model Evaluation Metrics')
visualPlot.show()
//Visualise the confusion matrix
visualPlot.confMatrix()

end function

Algorithm 3.6: UnSupervised ML Model Evaluation and Visualisation

Input: Data, KCM and FCM //data, kcmModel, fcmModel
Output: Metrics Values and Visualisations

function uspMLEvalVisual(data, kcmModel, fcmModel)
// Load necessary libraries
load perfMetrics //e.g., load perfMetrics from sklearn.metrics library
load visualPlot // e.g., load visualPlot from matplotlib.pyplot library

//Calculate Silhouette Score for KCM and FCM
kcmSilhouetteScore = perfMetrics.silhouette_score(data, kcmClusters)
fcmSilhouetteScore = perfMetrics.silhouette_score(data, fcmClusters)

// Display Silhouette Scores
display("KCM Silhouette Score:", kcmSilhouetteScore)
display("FCM Silhouette Score:", fcmSilhouetteScore)

//Visualize K-Means Clustering
visualPlot.subplot(121)
visualPlot.scatter(data[:,0], data[:,1], data[:,2])
visualPlot.title("K-Means Clustering")

//Visualise FCM Clustering
visualPlot.subplot(122)
visualPlot.scatter(data[:,0], data[:,1], data[:,2])
visualPlot.title("Fuzzy C-Means Clustering")
visualPlot.show()

end function

3.5.6 Selection of ML technique

Based on the cumulative performance results for the various training and testing experiments, the most suitable ML model for the retail tweet sentiment analysis task was selected. The selected algorithm was further benchmarked against similar studies in the literature in order to situate the comparative performance of the proposed conceptual framework in the body of knowledge. In addition to the training and testing experimental results, the comparative results are also presented and discussed in Chapter 4.

3.6 Describing the Experiment Design

This section describes how the methodology of the experiment was determined to evaluate the best algorithm for sentiment analysts to identify retail complaints from the Twitter platform to improve retail services based on the conceptual framework. It begins a discussion on the sentiment analysis tools (Section 3.5.1) using the hybrid approach followed by mathematical description of algorithms and techniques (Section 3.5.2), and subsequently, a description on conducting the experiment (Section 3.5.3). A description of the experiment design concludes on a discussion on the applications of algorithms to Twitter customer experience (Section 3.5.4).

3.6.4 Sentiment analysis tools

Lexicon-based approach, ML approach, and hybrid approach are the three most often used approaches for sentiment analysis. The benefit of the lexicon-based method is that it does not need any training data and is regarded as an unsupervised method by certain experts. The fundamental drawback of lexicon-based approaches is that they are heavily domain-specific, and terms from one domain cannot be utilised in another (Mohamad Sham and Mohamed 2022). This study selected the hybrid approach.

The two main methods used to determine the sentiment of the text in the hybrid approach are lexicon-based and ML techniques. The lexicon-based approach focuses on finding the polarity of such lexicons after generating opinion-based lexicons from the text. Lexicons are collections of well-known and already-assembled emotive words. Dictionary-based methods and corpus-based approaches are two other categories for this approach. The dictionary-based approach first involves the process of identifying words that express opinions before consulting the dictionary to gather their synonyms and antonyms. In contrast, in the corpus-based approach, thereafter, construct a list of opinion terms and then search a sizable corpus for more related opinion words based on their context-specific orientations (Kauffmann *et al.* 2020; Zucco *et al.* 2020).

To carry out the lexicon approach, a small group of words that describe opinions are manually gathered with their known orientations as a pre-processing operation. The set is then gradually expanded by looking for their synonyms and antonyms in well-known and often-used lexicon dictionary tools like WordNet or Sentiful. Contrarily, the primary goal of ML approaches is to create an algorithm that enhances the system's performance by utilising training data like examples and/or prior knowledge and experiences (Zucco *et al.* 2020).

Supervised and unsupervised ML approaches are further divided into categories. Working with subjective data, supervised ML algorithms are often utilised to perform sentiment analysis. Unlike unsupervised ML approaches, supervised ML techniques rely heavily on training data that has already been tagged (Yue *et al.* 2019). The classifier will categorise the remaining data, or test data, using the provided training data. For sentiment analysis, a variety of supervised ML techniques are employed, including Logistic Regression, Naïve Bayes, Decision Trees, Support Vector Machines (SVM), Random Forests, Maximum Entropy, and Bayesian Networks (Ducange *et al.* 2019; Obaidi and Klünder 2021).

The SentiWordNet method obtains the scores of each word from this lexicon using syn.pos_score () and syn.neg_score (). The words that have a higher positive score than a negative score in a tweet are summed up to compute the positive score (TweetPos) of the tweet. Similarly, words that have a higher negative score than positive score in a tweet are summed up to compute the negative score (TweetNeg) of the tweet. These scores are computed for all tweets. Sentiment polarity was determined by filtering and analysing the data using natural language processing techniques based on the emotion terms found in the user tweets. Word tokenisation, stemming, lemmatisation, part-of-speech (POS) tagging, named entity recognition (NER), and parsing are among the natural language pre-processing techniques used to prepare the dataset. These techniques are used to extract emotions from the textual information contained in each tweet. The Natural Language Toolkit and the Python programming language are used to achieve this strategy (NLTK). When a word in a sentence has a meaning in the current context, the derived algorithm uses WordNet to extract emotional terms and assigns sentiment polarity using the SentiWordNet dictionary and a lexicon-based technique.

For sentiment analysis, the words that express the opinion are the most crucial. Positive opinion is the desirable label for an entity, whereas negative opinion is the unwanted term. A lexicon is a set of predefined words, each of which is assigned a polarity score. It is a straightforward method of sentiment classification. This classifier uses a lexicon to conduct word matching, which is then used to categorise a sentence. The effectiveness of this classification method is determined by the size of the lexicon. Sentiment lexicon is a list of lexical features that are generally labelled according to their semantic orientation as either positive or negative (Chiong, Budhi and Dhakal 2021). SentiWordNet is an extension of WordNet in which 147 306 synsets are annotated with three numerical scores relating to positivity, negativity, and objectivity (neutrality). It has high coverage of terms. Each score ranges from 0.0 to 1.0, and their sum is 1.0 for each synset. It is a useful and popular lexicon for a wide range of tasks in text mining. This study uses SentiWordNet via PyCharm's Natural Language Toolkit. The Neutral and negative sentiments are further deciphered using deep learning, supervised and unsupervised machine learning techniques.

3.6.5 Mathematical Description of Algorithms and Techniques

This section of the research aims to better understand the tools used for sentiment analysis. The research will compare a study that investigates customer support emails, finding and monitoring sites on the web, and distilling reviews and sentiment analysis of social media data.

Customer sentiment analysis was investigated in this study. VADER sentiment was used in conjunction with a Swedish sentiment lexicon to offer the first classification of the emails. Two support vector machine models were trained to extract and categorise the sentiment of emails using the email content and sentiment labels as input. Each email was initially sentiment-labelled using VADER sentiment, a lexical sentiment classifier (Borg and Boldt 2020).

Additionally, VADER Emotion can convert the sentiment scores of individual words into scores for complete sentences. Booster words are taken into consideration when determining sentence sentiment support. SVM has already been used to do text mining and sentiment analysis jobs effectively. SVC and LinearSVC, two alternative SVM implementations were examined in this paper. The two SVM models were trained using emails that contained sentiment polarity ratings from VADER sentiment and a Sorkbanken sentiment dictionary (Borg and Boldt 2020).

This study used two SVM models and VADER sentiment to investigate sentiment analysis in email data. Based on the email texts, the latter was used to grade sentiment, which might

range from good to negative. The mean F1-score and mean AUC of the LinearSVM model's ability to extract sentiment. It is also explored to what extent it is possible to anticipate, from an agent's sent email, the sentiment of the subsequent email in the thread. This is because SVM models have the ability to forecast the sentiment of future email answers. It was nevertheless able to predict the e-mail sentiment, albeit with less successful outcomes than for extracting e-mail sentiment (Borg and Boldt 2020).

Each website typically contains a large amount of opinion text, making it difficult for the average human reader to distinguish the polarity of each review and summarise the opinions in them. Finding and maintaining these websites on the internet and distilling the reviews from them remain challenging tasks (Bonta and Janardhan 2019).

The overall methodology follows four steps: data collection, pre-processing, sentiment extraction and classifying sentiment as positive or negative. This study demonstrates how NLTK divides a review into positive and negative categories. The stop words like a, an, the, for, is, etc. are removed first once the reviews have been tokenised into words. Stop words are eliminated, and then the words are stemmed to determine their base words. For instance, "disappointed" becomes "disappoint." This shortens the time needed to search for the term in SentiWordNet. All special symbols and digits are also removed from the reviews. The POS (parts of speech) tagging on the cleaned reviews are performed next. When doing the tagging, strict grammar requirements must be followed. As a result, the data is available for classification once the positive and negative words from the review are extracted and compared to SentiWordNet's respected sentiment score. The class is given the highest score after counting the number of positive and negative terms that may be discovered in the review and applying sentiment polarity (Bonta and Janardhan 2019).

ML algorithms build their vocabulary by training their modules on half of the data and using the other half for testing. The majority of algorithms only function in particular fields. VADER performs better than ML methods in a variety of domains. VADER has a number of benefits over ML methods. First off, it is computationally efficient and speedy. A corpus can be analysed in a fraction of a second using VADER, but it can take hours when using more complicated models like support vector machine. VADER runs directly from a typical modern laptop or PC (Bonta and Janardhan 2019).

When analysing comments or reviews from social media, Text Blob Sentiment Analysis performs better than VADER because emoticons can modify the meaning of a sentence. This makes VADER sentiment analysis more effective for texts from social media and other online sources. This is taken into account by VADER together with slang, capitalisation and word choice in the context of the sentence. As an illustration, "The movie is good" yields a composite score of 0.4404 whereas "The movie is GOOD" yields a composite score of 0.5622. The sentiment's strength is also increased by another factor (Bonta and Janardhan 2019).

The majority of earlier studies focused on binary classification; it was suggested that a multi-classification system for tweet analysis in this work. The VADER tool was employed to categorise tweets pertaining to the 2016 US election. The findings indicated that ternary and multiple class detection had good accuracy (Elbagir and Yang 2019).

Researchers in various domains have recently become interested in the sentiment analysis of Twitter data. However, the majority of cutting-edge studies have employed sentiment analysis to gather and organise data about the opinions expressed on Twitter about a variety of issues, including forecasts, reviews, elections, and marketing. Many technologies available today, such as linguistic inquiry and word count (LIWC) [5], provide the ability to extract sophisticated information from texts (Elbagir and Yang 2019).

The study discovered that the performance is incredibly poor when a classifier learned in one area is directly applied to other domains. The research demonstrated the efficacy of various algorithms for varying tweet counts, including Naïve Bayes, multi-nominal NB, linear SVC, Bernoulli NB classifier, logistic regression, and the SGD classifier. The outcomes demonstrated that the suggested system was more effective than the current systems (Elbagir and Yang 2019).

There are three stages to the current investigation. The first step is the gathering of Twitter data. The initial pre-processing work done to clean and eliminate extraneous information from the tweets is the subject of phase two. In phase three, the NLTK's VADER analyser is used, along with a scoring system that ranks the results on a scale of 1 to 5 to determine how well it can categorise tweets (Elbagir and Yang 2019).

To classify the dataset, a sentiment intensity analyser was first built. After that, the sentiment was ascertained using the polarity scores approach. The pre-processed tweets were categorised as positive, negative, Neutral, or compound using the VADER Sentiment Analyser. The study showed a good metric for assessing the sentiment in a particular tweet is the compound value. The threshold values in the suggested approach are used to classify tweets as either good, negative, or Neutral. This investigation referred to typical threshold values (Elbagir and Yang 2019).

The findings showed that the VADER Sentiment Analyser was a good option when using sentiment analysis to classify Twitter data. Massive volumes of data were swiftly and easily classified using VADER. The current study does, however, include the following drawbacks: a minimal amount of data was initially used; second, specific data were categorised using a generic lexicon; and third, no training was done on the data. The aims in future work should be to enhance the system by utilising massive amounts of data, a particular lexicon, and a corpus for training the data to get accurate results (Elbagir and Yang 2019).

In the context of exploratory data analysis, this research discussed methods and tactics for capturing the polarity of people's attitudes towards donating to any cause, focusing specifically on sentiment analysis techniques. Unstructured data was converted into structured and clean data by data parsing and cleaning (Shelar and Huang 2018).

This study focused on using Python 3 and its standard libraries, as well as Tweepy (to download tweets) and NLTK to do this (the Natural Language Toolkit). The VADER sentiment analysis was employed. A vocabulary and rule-based sentiment analysis tool called VADER (Valence Aware Dictionary and Sentiment Reasoner) is customised precisely to the sentiments expressed in social media. By examining the data, to evaluate and employ various natural language processing strategies. Then classify texts using NLTK 2.0.4 to perform sentiment analysis (Shelar and Huang 2018).

NLTK's VADER analyser computationally distinguishes and groups text into three sentiments—positive, negative, or Neutral. Utilise VADER because it manages the data by including a wide variety of letters, symbols, and other characters in its lexicon. This method rates each word in the lexicon according to its positivity or negativity and, in many cases, the degree of positivity or negativity. Using this method, whether a tweet is Neutral, or expressing a positive or negative sentiment can be determined. Twitter messages can only be 140 characters long at most; therefore, a tweet may only include one or two sentences. Therefore, our objective is to simply disassemble the tweet in order to retrieve the polarity from it (Shelar and Huang 2018).

An examination of data from Twitter was given. By using sentiment analysis techniques, able to identify people's polarised sentiments. As a future plan, to learn more about users and businesses in order to research potential donors for non-profit organisations. Additionally, sentiment analysis might be able to pinpoint areas where organisational procedures might be improved by classifying replies according to sentiment and topic. As a result, the strategy may be applied in various ways to raise customer satisfaction (Shelar and Huang 2018).

Data and linguistic issues have been identified in related studies as two issues. A large amount of research has not been done on sentiment analysis of email data. The issue with language is that most research has been done in English and, more recently, Chinese. To the authors' knowledge, no research has looked at the possibility of predicting the tone of return emails (Shelar and Huang 2018; Bonta and Janardhan 2019; Elbagir and Yang 2019; Borg and Boldt 2020).

3.6.6 Evaluation of Results

The performance metric measures how the ML technique is performing on a given dataset. Accuracy is defined as all true predicted cases against all predicted cases. If it produces 100% accuracy, it denotes that the predicted cases are precisely the same as the actual cases. Precision is defined as the true positive predicted cases against all positive predicted cases. Recall is defined as the true positive predicted cases against all actual positive cases. F1 is the harmonic average of the precision and the recall.

In recent studies, the performance of sentiment techniques was estimated using four indicators: accuracy, precision, recall and F1-score. These indicators are computed using the confusion matrix. Whereas this study focuses on evaluating the overall performance of the ML techniques, including four statistical measures: reliability, accuracy, specificity, and F-measure. The percentage of relevant items returned by the ML technique out of the total number of relevant items in the original data is referred to as reliability. The metric that measures a model's ability to forecast the real negatives of each accessible category is called specificity.

3.6.7 Application of Algorithms to Twitter Customer Experience Data

The cleaned data will be further deciphered using the following ML Techniques and their algorithms: supervised learning algorithms, unsupervised learning algorithms and deep learning algorithms. An algorithm is a group of calculations and heuristics used in machine learning to build a model from data. The algorithm initially examines the data you submit, searching for particular kinds of patterns or trends before building a model.

In numerous iterations, the algorithm uses the findings of the research to choose the best parameters for building the mining model. The full data collection is then subjected to these criteria in order to extract useful patterns and thorough statistics. It can be difficult to select the ideal algorithm to apply for a particular analytical assignment. While various algorithms can be used to complete the same business objective, each method yields a unique outcome, and some algorithms can yield more than one kind of result.

3.7 Conclusion

This chapter explained the study’s methodological and design choices. The chapter noted that this study was guided by the post-positivist philosophy and adopted a quantitative research methodology with a longitudinal approach. The key focus of this chapter was to explain the architecture of the conceptual framework. This chapter elaborated on the background of the conceptual framework and explained how the stages were applied in answering the research question. The associated decisions that were made to ensure the rigour and ethicality of the data analytics process were also justified. The discussion of the experiment is presented in the next chapter (Chapter 4).

Chapter 4 Results and Discussion

4.1 Introduction

Chapter 4 presents all of the study's findings and results, beginning with a discussion on the understanding of the experiment's hardware and software (Section 4.2). This is followed by the experiment's findings and results on data preparation (Section 4.3) and the conceptual framework results are presented in Section 4.4. In conclusion of this chapter, the selection of ML techniques for the conceptual framework is discussed (Section 4.6).

4.2 Experiment's Hardware and Software

This research aims to develop a conceptual framework to improve the identification of customer complaints on Twitter using a hybrid approach that combines lexicon-based and ML techniques for sentiment analysis. This research objective is to determine the most suitable ML technique for sentiment analysis.

The experiment was conducted on a Lenova Flex Core i7 1165G7 16GB RAM, 512GB SSD 2-in-1 laptop. This experiment required a device with a good amount of RAM for memory space. Initially, the experiment was conducted on an older model, specifically an i5, with an outdated Windows operating system, resulting in frequent device freezing. The experiment was conducted within PyCharm, which is an integrated development environment used in Python programming. PyCharm combines both Python, C# and HTML. PyCharm is a great Python IDE for ML that consists of a variety of features for ML development.

The collection of data in the .CSV file was prepared by exploratory data analysis (EDA) and the missing value method. The conceptual framework commenced with data extraction from the prepared .CSV file. This was followed by five NLP pre-processing methods to prepare the data for sentiment analysis. Dimensional reduction (DR) was introduced to reduce the dimensions of data. Supervised ML and deep learning needed to be trained and tested to improve the accuracy of the selected models.

This study used the hybrid approach for sentiment analysis, which is a combination of the Lexicon method and ML techniques. Sentiment lexicon was used for the Lexicon approach. Supervised ML, unsupervised ML and deep learning were selected for the ML approach. To evaluate the performance of the ML technique, the following performance metrics were introduced: reliability, accuracy, specificity and F-measures. This was then followed by visualisation to display the model evaluation line graph and the extracted data in tabular format. The conceptual models conclude with the selection of an ML technique for sentiment analysis.

4.3 Data Preparation

4.3.1 Exploratory Data Analysis and Missing values

The first step is to understand the shape of the dataset by knowing the number of rows and columns within the dataset. The shape of the dataset resulted in 27 233 rows of data and 11 columns within the dataset.

Figure 4.1: Shape of the dataset

Figure 4.1 displays the output of the shape of the dataset which includes 27 233 rows of data and 11 columns within the dataset. This is then followed by understanding the columns of the data. The following are the columns found within the dataset; Creation of Tweet, Gender, Created, Description, Username, Profile image, Tweet, Hashtag, Tweet_count, tweet_created, Tweet Location.

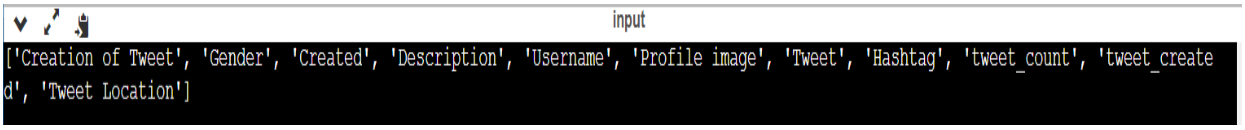


Figure 4.2: Columns of the dataset

Figure 4.2 displays the names of the columns found within the dataset.

```
Creation of Tweet        0
Gender                   154
Created                  0
Description              1899
Username                 0
Profile image            523
Tweet                    0
Hashtag                  7879
tweet_count              0
tweet_created            0
Tweet Location           4891
```

Figure 4.3: Null Values

Figure 4.3 displays the total rows with null values or missing values.

The second step is to check for null values, also known as missing values. Every tweet had a Creation of Tweet, which was the time of the tweet resulting in no missing values or null values. Out of the 27 233 tweets created, 154 had no gender assigned to it. 1899 descriptions were missing within the users' profiles. There were no missing usernames. 523 profile images were missing within the users' profiles. There was no null value for Tweets. 7879 Tweets were missing hashtags. Tweet count and tweet created were all accounted for. 4891 Tweets had no location assigned to them. Once these had been identified, the researcher proceeded to drop the columns with null values and columns that were not required for this study.

The third step in the EDA process is to drop missing values. As depicted in Figure 4.3, columns which revealed missing values were removed, because they contained no valid information for this study. The dataset was updated to consist only of the required columns: Creation of Tweet, Username, Tweet, Hashtag and Tweet Location which is presented in Figure 4.4. The following columns were no longer required in the research and were removed: Gender, Created, Description, Profile image, tweet_count, tweet_created. Bivariate analysis was performed after the data was prepared (Section 5.1.1.1).

```
input
['Creation of Tweet', 'Username', 'Tweet', 'Hashtag', 'Tweet Location']
```

Figure 4.4: Updated Table

Figure 4.4 displays only the required columns selected for this study. Columns that were not required and contained null values were dropped from the dataset. This resulted in the following columns to remain: Creation of Tweet, Username, Tweet, Hashtag and Tweet Location.

4.3.1.1 Bivariate Analysis

The objective in bivariate analysis is to compare two variables and discover how one property influences another. This is achieved using scatter plot graphs, which are data visualisations that show the relationships between different variables. This data is shown by placing various data points between the x and y axes.

Scatter plot graphs helped in determining the correlations between two variables, in which this research selected the time and day of the tweet. This was done for each of the nine provinces in South Africa: Limpopo, Mpumalanga, Free State, the Eastern Cape, KwaZulu-Natal, Gauteng, the Western Cape, the Northern Cape and the North West. This was done to determine the pattern between each province, a common time to create a complaint via Twitter, and the most prominent day to accomplish this. Each plot represents 100 tweets and the average time that they were posted on Twitter. By doing so, retailers can identify when they are prone to not meeting customers’ expectations or standards.

i. Bivariate analysis on Limpopo

Figure 4.5 shows that the majority of the complaints/comments about a retailer was conducted on a Sunday and Saturday as well as at night. Very few were produced on a weekday. No tweets were found on a Wednesday. Limpopo was one of the provinces to produce the least number of complaints/comments about a retailer. Limpopo created 978 retail tweets between January 2017 and January 2019.

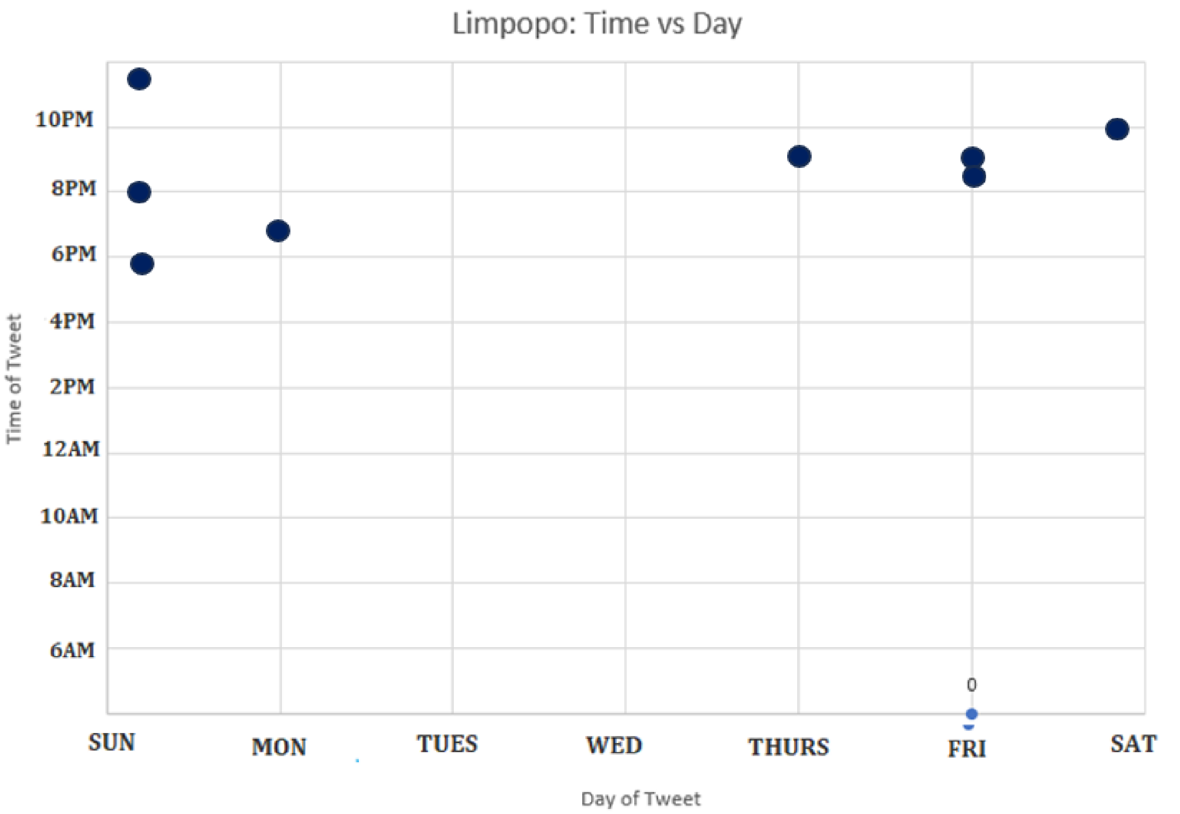


Figure 4.5: Bivariate analysis visualisation on Limpopo

Figure 4.5 displays the visualisation for bivariate analysis on Limpopo.

ii. Bivariate Analysis on Mpumalanga

Figure 4.6 shows that the majority of complaints/comments about a retailer was conducted on a Sunday and Saturday. No tweets were found on a Tuesday, Wednesday or Thursday. Mpumalanga was one of the provinces to produce the least number of complaints/comments about a retailer. Mpumalanga created 982 retail tweets between January 2017 and January 2019.

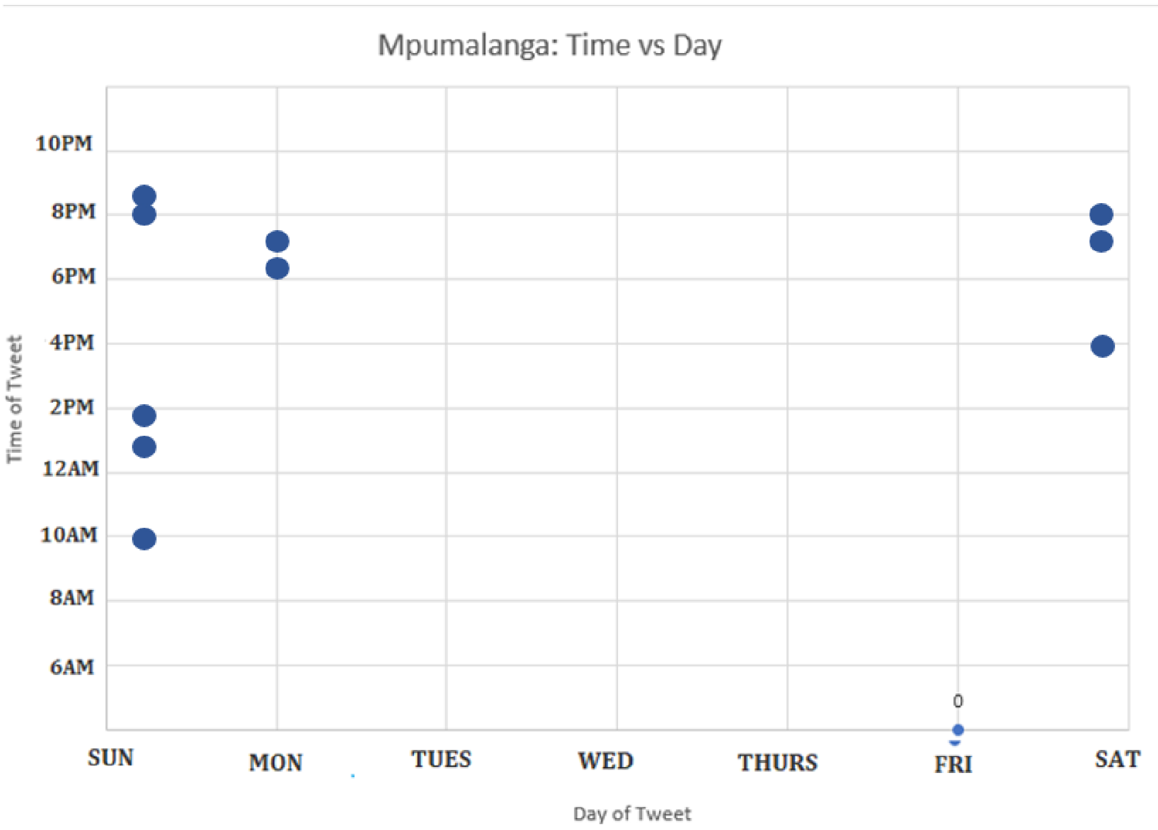


Figure 4.6: Bivariate analysis visualisation on Mpumalanga

Figure 4.6 displays the visualisation for bivariate analysis on Mpumalanga.

iii. Bivariate Analysis on Free State

Figure 4.7 shows that the majority of complaints/comments about a retailer was conducted on a Sunday and Saturday. Tweets were found throughout the week and at a variety of times, from 6:00 in the morning until past 22:00. The majority of the tweets in the Free state were found after 20:00. The Free State created 2561 retail tweets between January 2017 and January 2019.

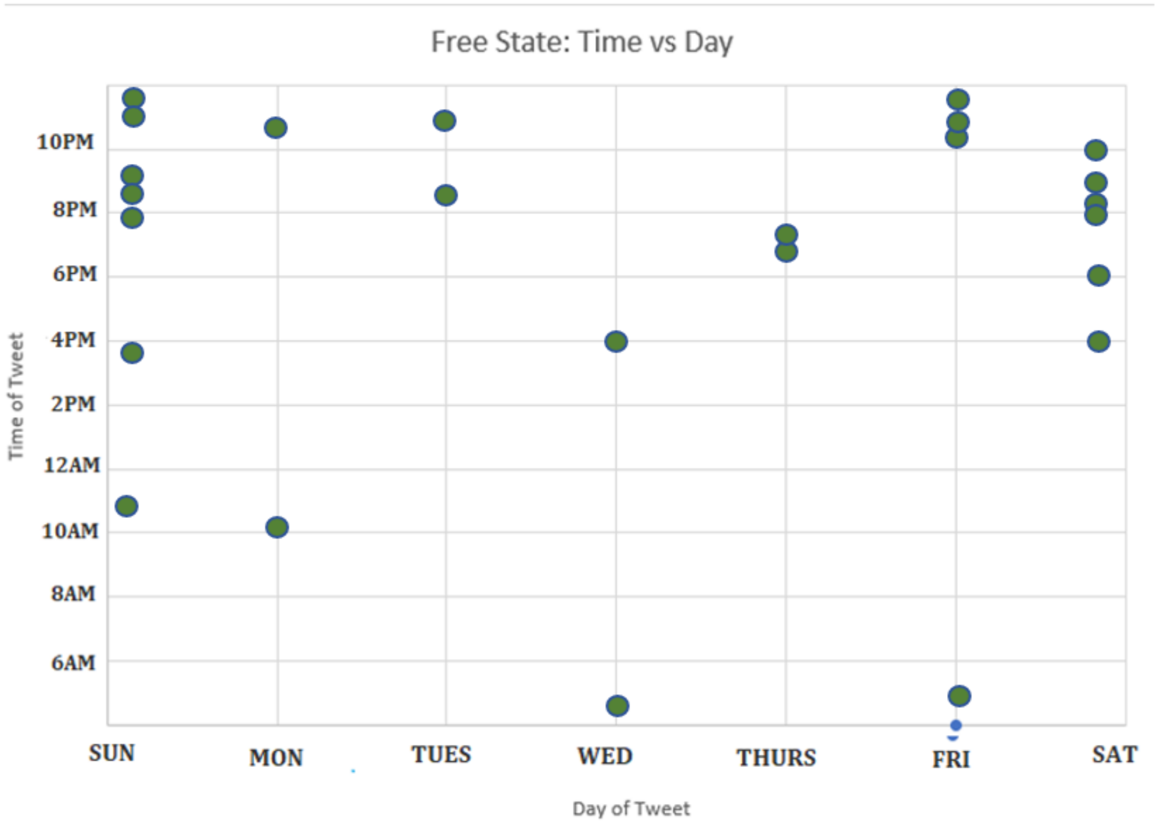


Figure 4.7: Bivariate analysis visualisation on Free State

Figure 4.7 displays the visualisation for bivariate analysis on Free State.

iv. Bivariate analysis on Eastern Cape

Figure 4.8 shows that the majority of complaints/comments about a retailer was conducted on a Sunday and Saturday. Tweets were found throughout the week and at a variety of times. Very few Tweets were found during the day. The Eastern Cape created 3489 retail tweets between January 2017 and January 2019.

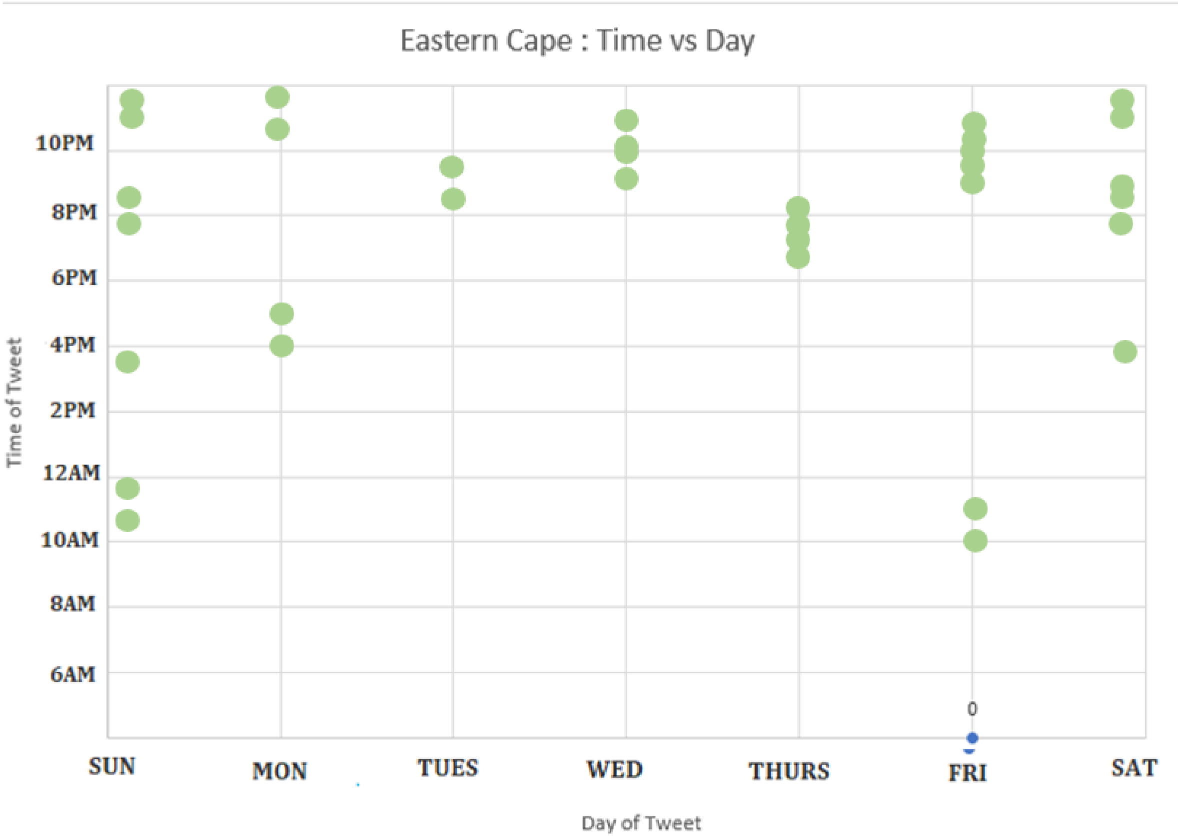


Figure 4.8: Bivariate analysis visualisation on Eastern Cape

Figure 4.8 displays the visualisation for bivariate analysis on Eastern Cape.

v. Bivariate Analysis on KwaZulu-Natal (KZN)

Figure 4.9 displays that the majority of complaints/comments about a retailer conducted on a Sunday and Saturday. Tweets were found throughout the week and throughout the day. Majority of the tweets were found after 6PM. KZN shown to have one of the highest complaints/comments on Twitter. Between January 2017 and January 2019 Eastern Cape created 3489 retail tweets.

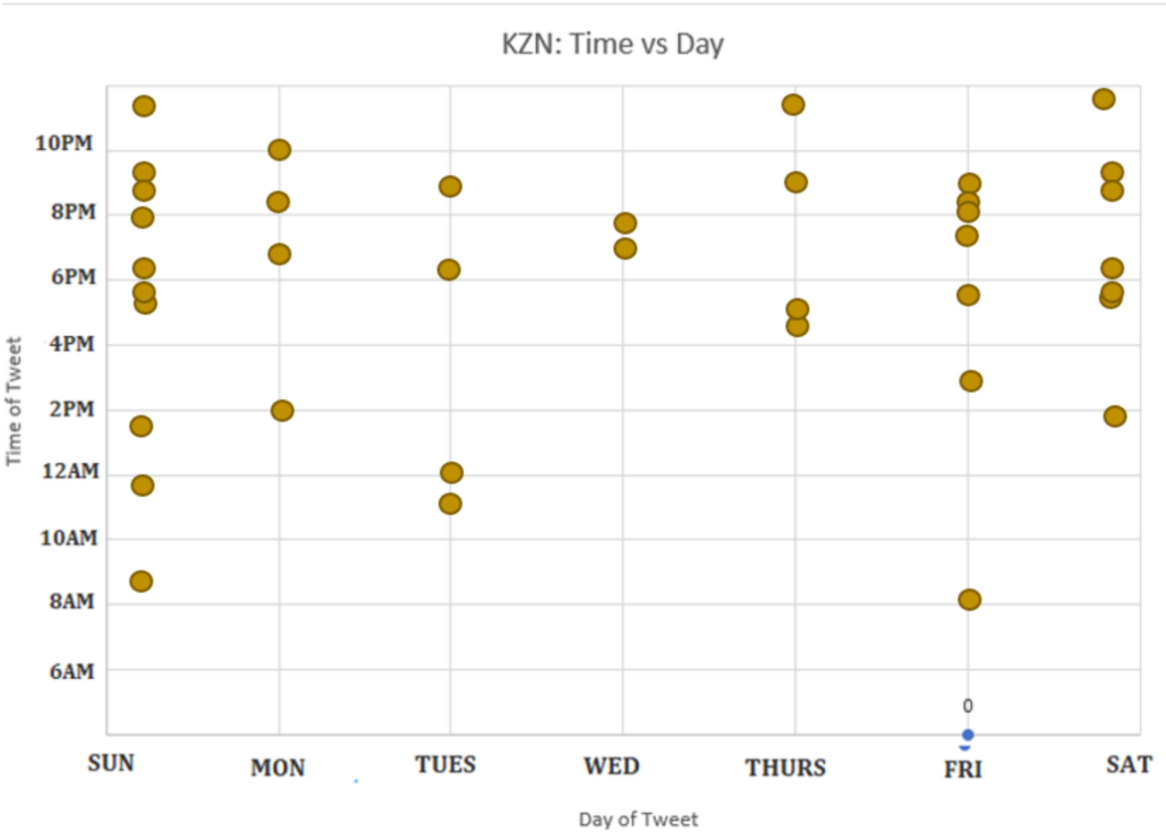


Figure 4.9: Bivariate analysis visualisation on KZN

Figure 4.9 displays the visualisation for bivariate analysis on KZN.

vi. Bivariate Analysis on Gauteng

Figure 4.10 displays Tweets that were found throughout the week and throughout the day. The majority of the tweets were found after 18:00. Gauteng was shown to have the highest complaints/comments on Twitter. Gauteng created 5477 retail Tweets between January 2017 and January 2019.

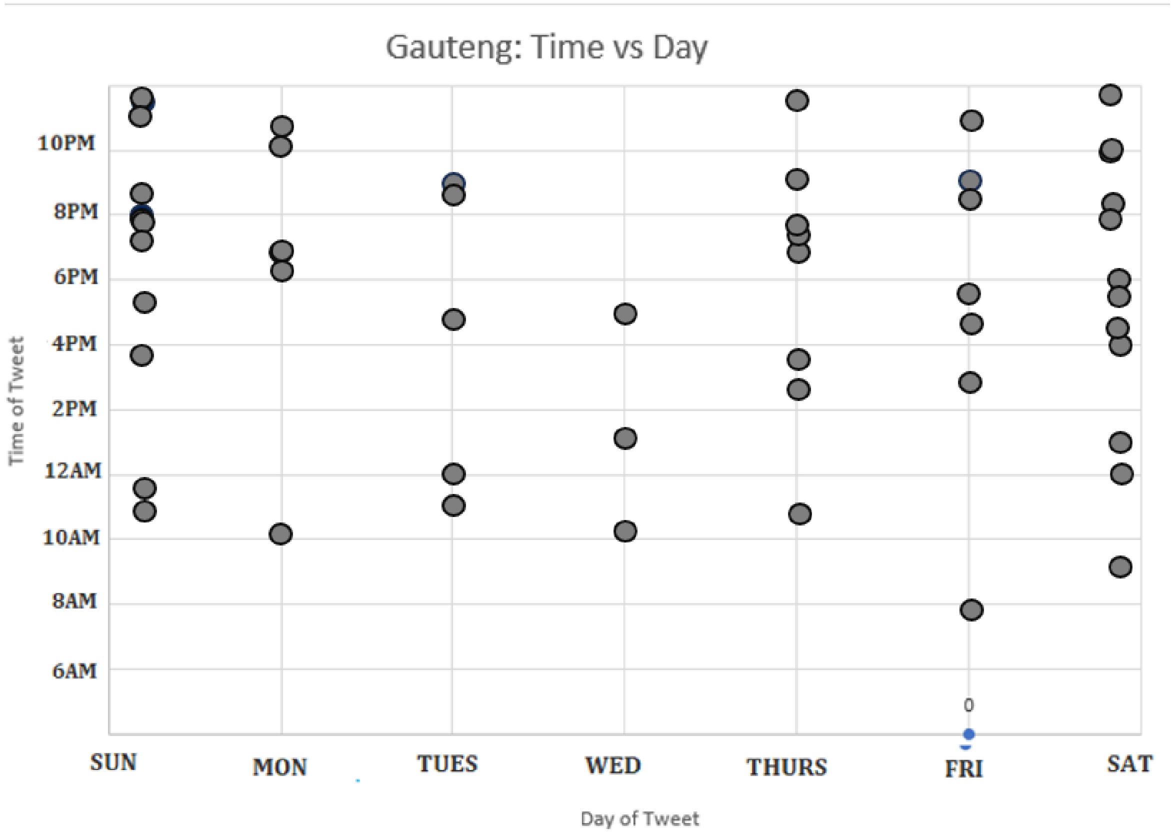


Figure 4.10: Bivariate analysis visualisation on Gauteng

Figure 4.10 displays the visualisation for bivariate analysis on Gauteng

vii. Bivariate Analysis on Western Cape

Figure 4.11 shows that the majority of complaints/comments about a retailer conducted on a Sunday and Saturday. Western Cape shown to have one of the highest complaints/comments on Twitter. Between January 2017 and January 2019 Western Cape created 5432 retail tweets between January 2017 and January 2019.

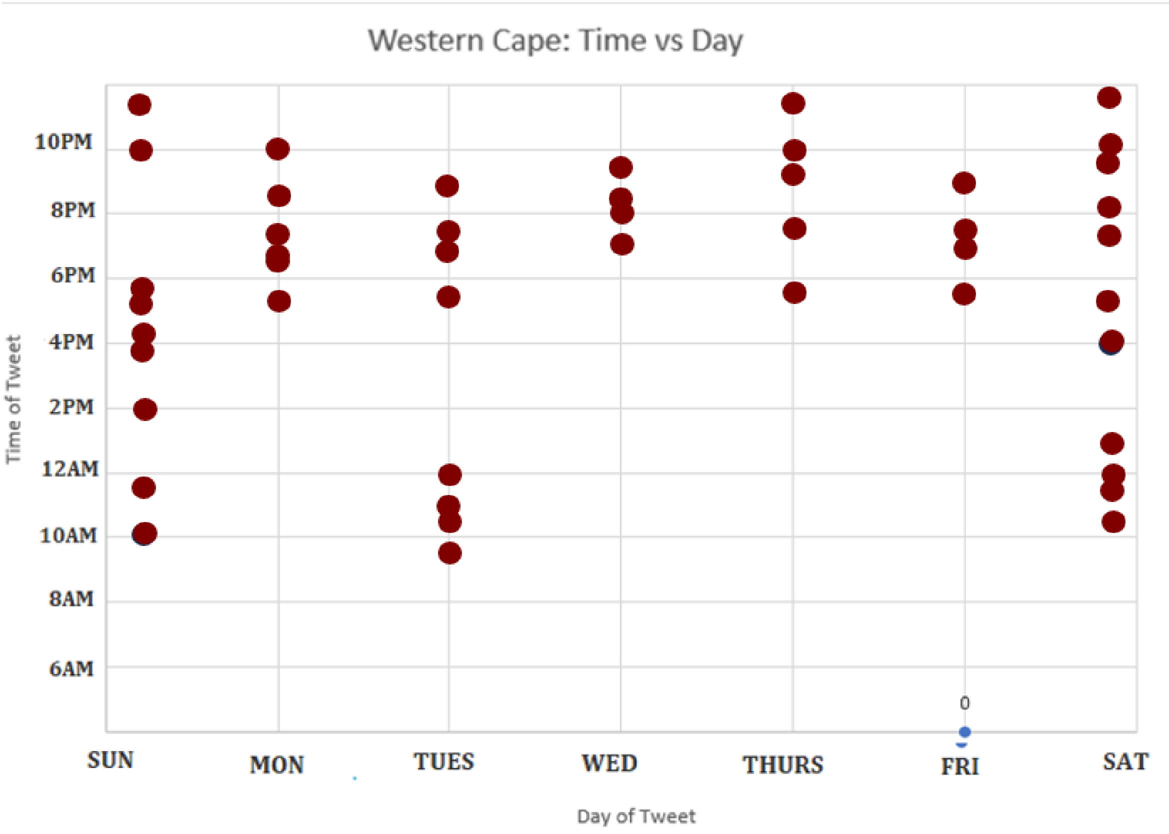


Figure 4.11: Bivariate analysis visualisation on Western Cape

Figure 4.11 displays the visualisation for bivariate analysis on Western Cape.

viii. Bivariate Analysis on Northern Cape

Figure 4.12 shows that very few Tweets were produced by Northern Cape, but it was found throughout the week. No tweets were produced on a Thursday at all. Northern Cape was one of the provinces to produce the least number of complaints/comments about a retailer. Northern Cape created 878 retail tweets between January 2017 and January 2019.

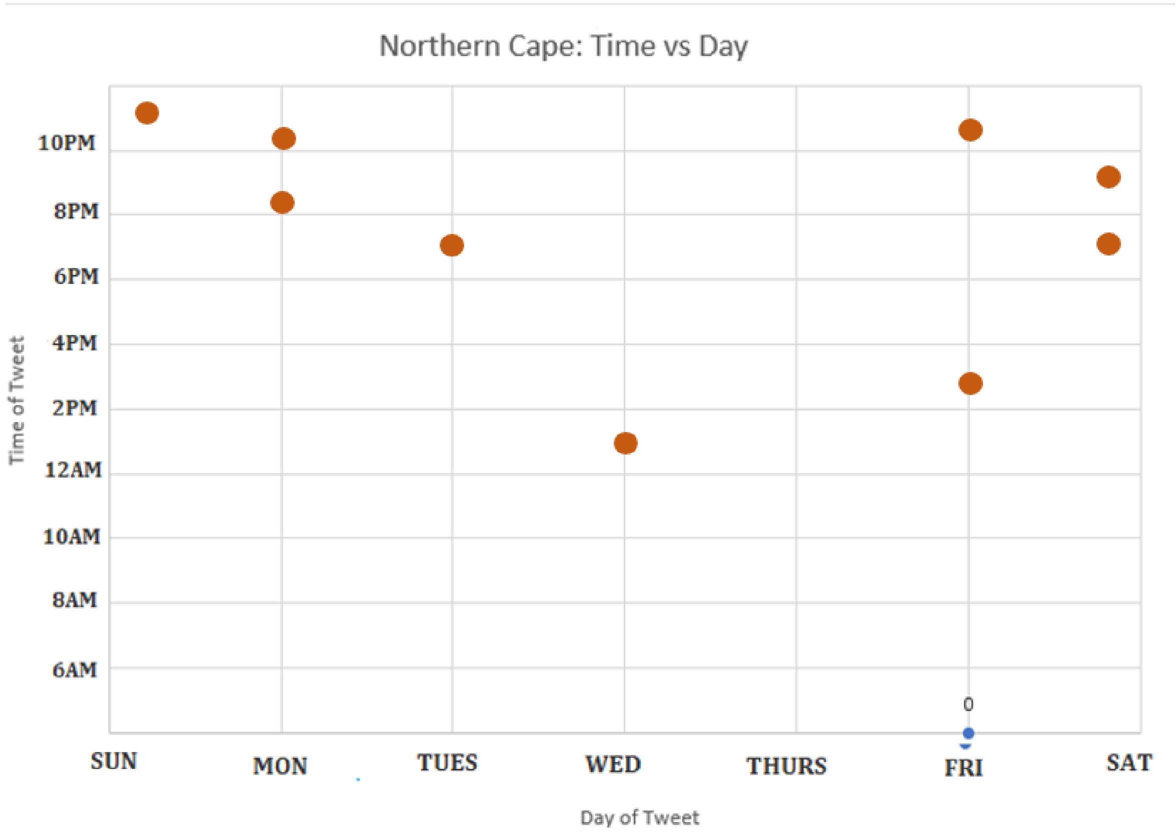


Figure 4.12: Bivariate analysis visualisation on Northern Cape

Figure 4.12 displays the visualisation for bivariate analysis on Northern Cape.

ix. Bivariate Analysis on North West

Figure 4.13 shows that very few Tweets were produced by North West, and it was found on a Sunday, Monday and Saturday. Northwest produced the least number of complaints/comments about a retailer. North West created 871 retail tweets between January 2017 and January 2019.

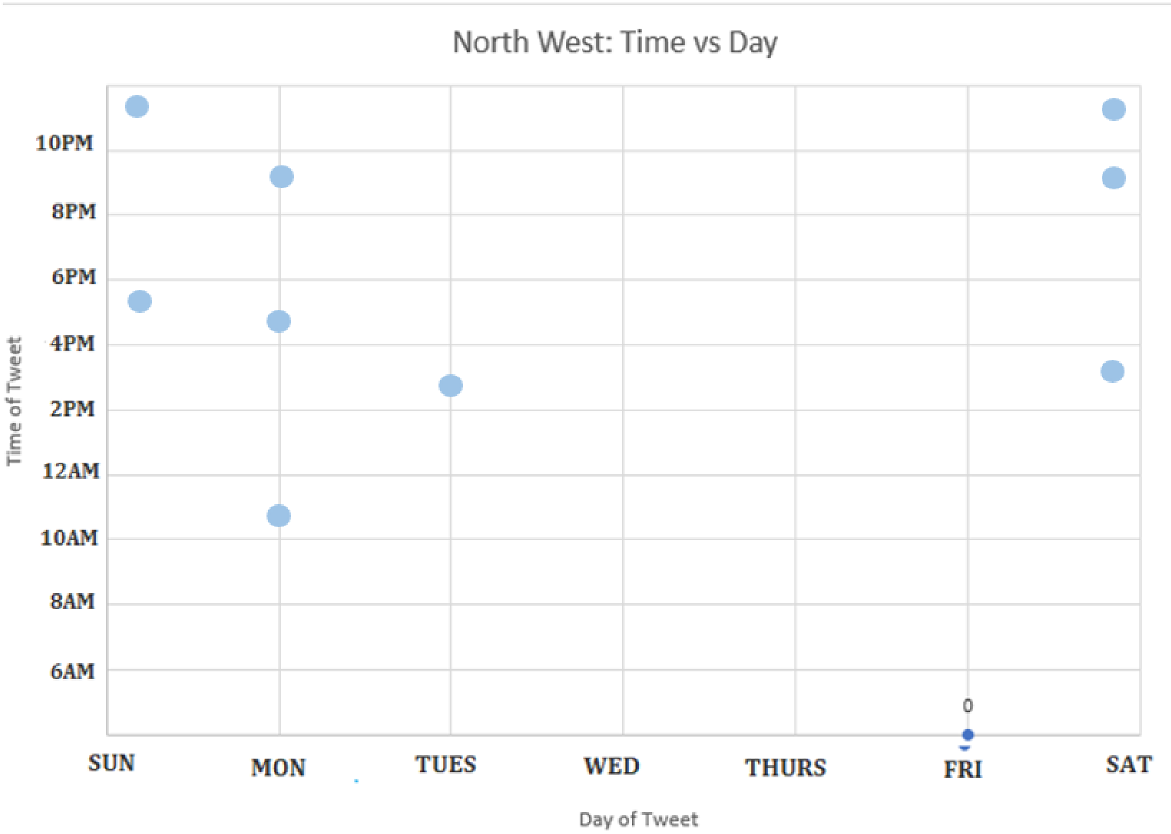


Figure 4.13: Bivariate analysis visualisation on North West

Figure 4.13 displays the visualisation for bivariate analysis on North West.

4.3.2 Retail complaints per province

An analysis was conducted on retail complaints for each province in South Africa. This data used the sentiment analysis results obtained through the sentiment lexicon approach, yielding a total of 22 298 tweets, of which 15 712 were Neutral sentiment tweets and 6 586 were negative sentiment tweets.

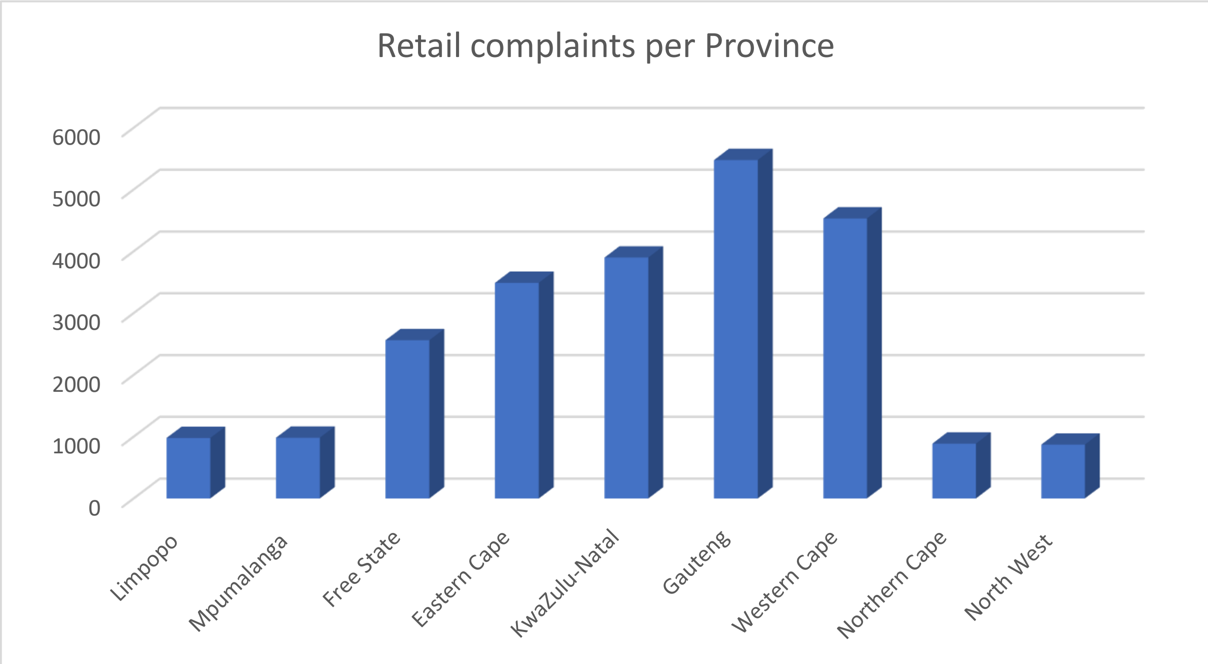


Figure 4.14: Retail complaints/comment per province bar graph

Figure 4.14 displays the average retail complaints/comment that were produced by each province in South Africa.

Table 4.1 presents the average retail complaints/comment per province from January 2017 to January 2019.

*Table 4.1: **Range retail complaints per province***

Provinces of South Africa	Retail complaints per Province
Limpopo	978
Mpumalanga	982
Free State	2 561
Eastern Cape	3 489
KwaZulu-Natal	3 899
Gauteng	5 477
Western Cape	5 432
Northern Cape	878
North West	871

4.4 Conceptual Framework Results

This section of Chapter 4 presents the results and discussions of the conceptual framework. The conceptual framework begins with data extraction (Section 4.4.1); the extracted data will then go through pre-processing techniques for NLP (Section 4.4.2). Thereafter, the extracted data will proceed to feature extraction (Section 4.4.3). Sentiment analysis will then be conducted (Section 4.4.4), which contains sentiment lexicon for retail tweets (sub-section 4.4.1.1) and then training and testing of ML models for sentiment analysis (sub-section 4.4.4.2). This is followed by model evaluation and visualisation (Section 4.4.5). The conceptual framework will end with the selection of ML techniques (Section 4.5).

4.4.1 Data Extraction

The conceptual framework will extract data from the .CSV file that contains retail Tweets directly from Twitter. The entire extracted data (.CSV file), which consists of 272 333 rows of Twitter data, will go through all eight stages of the conceptual framework.

4.4.2 Pre-processing Techniques for NLP

Table 4.2 explains the impact that pre-processing had on the data. stat function. Additionally, the file size was checked using st_size, and PyCharm profiler cProfile was used to assess CPU time as well as the length of time spent inside each function and the frequency with which each function was invoked. Each time a pre-processing method was applied, the file size and processing time were checked. Consistently, the file size and processing time recorded a decrease.

Pre-processing consisted of five methods and was done in the following order: tokenisation, stemming, lemmatisation, part-of-speech tagging and named entity recognition. Processing time is the amount of time it took for sentiword net to evaluate the file. Before the preprocessing methods could be applied, the file size was 4.308 KB and the processing time was 4.05 seconds. POS and NER required lemmatisation and tokenisation to be conducted first before implementing these two methods. The pre-processing methods had to be conducted in the following order: tokenisation, stemming, lemmatisation, part-of-speech tagging (POS) and named entity recognition (NER) to produce the best result. This is because each pre-processing method prepares the data precisely for the next method. Tokenisation is the process of breaking down the text in the NLP processing unit into a sentence called a token, and stemming is a process of finding the root of the words; therefore, tokenisation prepared the data for stemming, and so forth with the rest of the pre-processing methods. This caused the file to decrease in size to 3,420 KB and 1.15 seconds of processing time.

Table 4.2 presents the data's file size before and after pre-processing.

Table 4.2: Analysis on File Size and Data processing after Data has been pre-processing

Pre-processing tasks	File size in (KB)	File in size %	Processing time (sec)
Before pre-processing	4,308	100%	4.05
Tokenisation	3,695	85,77%	2.70
Stemming	3,518	81,66%	2.42
Lemmatisation	3,442	79,90%	2.06
Part-of-speech tagging (POS)	3,431	79,64%	1.32
Named entity recognition (NER)	3,420	79,13%	1.15

4.4.3 Feature Extraction

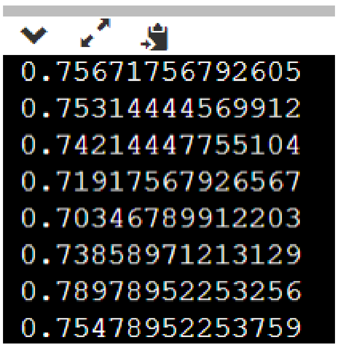


Figure 4.15: Word2Vec Results

Figure 4.15 demonstrates that the Word2Vec algorithm provided a representation of the 27 233 rows of data that has learnt the meaning of words within the dataset. Word2Vec generated embeddings that are context independent which is a one vector representation of a row of data.

4.4.4 Sentiment Analysis

As mentioned in Section 3.5.4, this study implemented the hybrid approach for sentiment analysis, which is the implementation of a lexicon-based approach (sub-section 4.4.4.1) and an ML-based approach (sub-section 4.4.4.2), which also provides results on training and testing of the required ML techniques.

4.4.4.1 Sentiment Lexicon for Retail Tweets

This approach used for analysing sentiment from linguistic data is known as a lexicon-based or dictionary-based approach. The derived architecture for sentiment analysis in NLP. For each word associated with each synset term. The sentiment scores obtained for all related terms or words with its POS tag using sentScore. SentiWordNet was used to evaluate positive and negative phrase scores to identify sentiment orientation for each term in the sentence or tweet. In this method, first list the emotion score for the first synset in the list of synsets. The score for each synset term is determined by the context or presence of the term in the supplied sentence. The positive opinion is defined by the word 'good,' which has a score of 0.25, and the negative opinion is defined by the word 'be,' which has a score of 0.25 positive and 0.125 negative sentiment. The term 'other', on the other hand, denotes a negative judgement with a score of 0.625. Finally, the data set for sentiment polarity for Twitter data was created, and only the negative and neutral results received were appended to ML approach for sentiment analyst to be further deciphered.

Table 4.3 presents the three classified sentiments and the overall total for the classified sentiments as follows:

Table 4.3: Sentiment Analysis Results

Sentiment Label	Tweets collected
Neutral	15 712
Positive	4935
Negative	6586

The sentiment analysis results were conducted using the sentiment lexicon approach. The dataset resulted in 57% being labelled Neutral, 18% being labelled positive, and 24% being labelled negative.

This was concluded by comparing the Neutral, positive, and negative sentimental tweets to the total number of extracted tweets, which was 27 233. 15 712 Neutral sentimental tweets against the total number of 27 233 tweets provided the result of 57% of the data that was calculated as Neutral. 4 935 positive sentimental tweets against the total number of 27 233 tweets provided the results of 24% of the data that was calculated as positive, and finally, 6

586 negative sentimental tweets against the total number of 27 233 tweets provided the results of 18% of the data that was calculated as negative. The Neutral dataset held positive and negative sentimental tweets. Only negative and Neutral sentiments were further selected for the experiment.

4.4.4.2 Training and Testing of Machine Learning Models for sentiment analysis

Only the negative and Neutral sentiments are to be further deciphered by the ML techniques. Methods were written in Python in PyCharm based on each ML algorithm. Each ML algorithm was linked to the main method, which was either supervised, unsupervised, or deep learning. The ML algorithms were selected based on the information that was found in the literature review in Chapter 2. These models could serve as the foundation for the tools that online merchants use to better analyse consumer feedback and, as a result, comprehend the wants of online shoppers. This study aims to identify the most acceptable ML model for sentiment analysis of customer complaints made online and posted to the social media site Twitter.

Before proceeding with the ML approach for sentiment analysis, the ML models first need to be trained and tested which is described and presented in Figure 4.16. To evaluate the impact the train and test method had on supervised ML techniques and deep learning, a comparison of the prediction output provided by the ML techniques against the test data was done first. Thereafter, an accuracy result is produced to tell how well the ML techniques have been trained to predict the values of the dataset.

4.4.4.3 Train and test Supervised Learning

A model is created during training by feeding a labelled data set through a classification algorithm. The model then uses the test data to make a category prediction. A classification algorithm with a high likelihood of success is Naïve Bayes. Each word was handled separately because the position of a keyword within a tweet was not taken into account. Naïve Bayes uses the Bayes theorem to determine the likelihood that each phrase matches a given label.

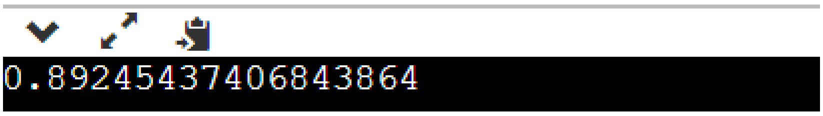


Figure 4.16: Accuracy Result for Supervised Learning

Figure 4.16 displays the accuracy result for supervised ML once it has been trained and tested, with an 89% accuracy result.

4.4.4.4 Train and test Unsupervised Learning

Unlabelled data is simpler to collect than labelled data. Each category's keyword lists are used to categorise the sentence. The unsupervised approach does not require the model to be trained and tested, as mentioned in Section 3.5.4.2.

4.4.4.5 Train and Test Deep Learning

The most important concept involved in deep learning is the use of deep neural networks to train the nodes for extracting complex features from the available information with limited contribution. These algorithms adapt easily and do not require manual input, giving them the ability to learn new complex features automatically. It took five hours and 22 minutes to train and test deep learning.

```
0.92366718808651217
```

Figure 4.17: Accuracy Result for Deep Learning

Figure 4.17 displays the accuracy result for deep learning once it has been trained and tested with a 92% accuracy result.

4.4.4.6 Supervised Learning

During training, a data set with labels is fed through a classification algorithm, which produces a model. The test data is then entered into the model, which predicts the category. Naïve Bayes is a classification algorithm with a high probability of success. It did not take into account the position of a term in a tweet; therefore, each word was treated independently. The probability of each phrase that corresponds to a label is calculated using the Bayes theorem by Naïve Bayes. SentiWordNet Lexicon with Naïve Bayes was utilised in the study to improve the classification of the Twitter dataset by providing the score of positive and negative tweets. The Naïve Bayes classifier cannot establish a semantic association between words since each word is treated independently, whereas the Bayesian network can. The words depend on each other in a Bayesian network.

SVM is set up to solve binary classification issues. Its goal is to find the optimal hyperplanes, which serve as a separator to characterise the decision boundaries between data points from different classes. A hyperplane that can maintain the maximum distance between two support vectors of different classes should be chosen. The study demonstrated that the selection of SVM improved accuracy with the help of the experiment.

```
input
SVC(C=2 , cache_size=200, class_weight=None, coef0=0.0,
   decision_function_shape ='ovr', degree=3, gamma='auto', kernel='linear',
   max_iter=-1, probality=False, random_state=None, shrinking=True, tol=0.001, verbose=False
   tol=0.001, verbose=False
```

Figure 4.18: SVM Model that has been built

Figure 4.18 demonstrates the SVM model that was built for this study. This is an output result of the model that was built but not the actual development model. The study had to begin by specifying the kernel and the radial basis was selected because it is the default kernel in SVM.

An input layer, a hidden layer, and an output layer are all parts of an artificial neural network (ANN). The input to the neuron is a vector called "a (i)" that represents the frequency of a word in a tweet. The function is calculated using a weight "A" that is assigned to each neuron. The neural network used the linear function x (i) = A. (a (i)). The class is categorised using the sign of x (i). The two steps in training an ANN model are forward propagation and backward propagation. The input is supplied to the input layer of neurons in forward propagation, where it is multiplied by weights that are generated at random.

```
ANN(activation= 'relu' , alpha=0.001, batch_size='auto' , beta_1=0.9,
    beta_2=0.99, early_stopping=False, epilon=1e-08,
    hidden_layer_sizes=(10, 10, 10), learning_rate= 'constant',
    learning_rate_init=0.001, max_fun=1500, max_iter=1000,
    momentum=0.9, n_iter_no_change=10, nesterovs_momentum=True,
    power_t=0.5, random_state=None, shuffle=True,
    tol=0.001, validation_fraction=0.1, verbose=False,
    warm_start = False)
```

Figure 4.19: Artificial Neural Network model that has been built

Figure 4.19 demonstrates the ANN model that was built for this study. This is the output result of the model that was built, not the actual development model. Weights and bias were required for this model to learn the input feature.

The decision tree is a tree-like structure in which the non-terminal nodes represent features and the terminal nodes represent labels. This is a recursive procedure that will lead to a terminal node that will assign a label to an input. A decision tree is an effective tool for sentiment analysis since it can handle vast amounts of data. The training data is divided into hierarchical groups via a decision tree. A condition based on the attribute value is used to divide the data. The presence or absence of a word determines the condition. The division process is repeated until the terminal nodes represent the limited number of attributes needed for categorisation.

```
DecisionTree(class_weight=None , criterion='entropy', max_dept='3',
           max_features=None, max_leaf_nodes=None,
           min_impurity_decrease=0.0, min_impurity_split=None,
           min_samples_leaf=5, min_samples_split=2,
           min_weight_fraction_leaf=0.0, presort=False, random_state=100,
           splitter='best'
```

Figure 4.20: Decision Tree model that has been built

Figure 4.20 demonstrates the decision tree model that was built for this study. This is an output result of the model that was built, not the actual development model.

All of the above models were built using the supervised ML method.

4.4.4.7 Unsupervised ML

Labelled data is more difficult to gather than unlabelled data. The sentence is categorised using the keyword lists for each category. The unsupervised method is more practical for domain-specific data analysis. During testing, it was found that sentiment analysis employed an unsupervised method in which tweets were divided into positive and negative groups using spectral clustering. Fuzzification uses a triangle membership function to fuzzy each tweet's positive and negative score from the second phase. Each linguistic word T involves three essential points, d, e and f, related to the change in pattern of the fuzzy membership when the triangu-lar fuzzy membership is applied. The K-means technique in ML starts with a first group of randomly picked centroids, which serve as the starting points for each cluster, and then performs iterative (repetitive) calculations to optimise the centroids' placements.

```
KMeans( algorithm='auto' , copy_x=true, init= K-means++ , max_iter=300,
        n_clusters=3, n_init=10, n_jobs=1, precompute_distances='auto',
        random_state=None, tol=0.0001, verbose=0)
```

Figure 4.21: K-Means model that was built

Figure 4.21 demonstrates the K-means model that was built for this study. This is the output result of the model that was built, not the actual development model. This model begins with defining the number of clusters; this study selected free clusters, followed by the selection of cluster points within the data. Observations were assigned to the closest centroid and these steps had to be repeated until the centroids did not change position.

```
def_calculate_fuzzyCov(self, X, membership, new_class_centers,
              fuzzyMem = Memberships ++ self.fuziness,
              n_clusters = self.n_clusters,
              FcovInv_Class = [],
              dim = x.shape [1],
              for i in range (n_ clusters)
              diff = x.new_class_centres [i])
              Fcov = n_clusters=3, n_init=10,
              FcovInv_Class.append(FcovInv)
```

Figure 4.22: Fuzzy C Model that was built

Figure 4.22 demonstrates the fuzzy C model that was built for this study. This is the output result of the model that was built, not the actual development model. This model begins with an understanding of the parameters and hyperparameters. This is followed by clustering, and 'I' is the fuzzifier, which controls how fuzzy the cluster boundary should be.

4.4.4.8 Deep Learning

The use of deep neural networks to train the nodes for extracting complicated characteristics from the available input with minimal contribution is the most crucial idea in deep learning. These algorithms can automatically learn new complex features because they are flexible and do not require operator input. However, the disadvantage of adopting deep learning techniques was that they required a large amount of data in order to provide high-efficiency results. WordNet was used to produce a set of emotions from text and group the emotions. Recurrent Neural Networks (RNN) worked on preserving a layer's output and fed it back into the input in order to refine the sentiment output.

```
MLPCClassifier(activation = 'relu', alpha=0.001, batch_size='auto', beta_1=0.9,
              beta_2=0.999, early_stopping = False, epsilon = 1e-08,
              hidden_layer_sizes=(100,), learning_rate='constant',
              learning_rate_init = 0.001, max_fun=15000, max_iter = 500
              momentum=0.9, n_iter_no_change = 10, nesterovs_momentum=True,
              power_t = 0.5, random_state=None, shuffle = True, solver = 'adam',
              tol=0.0001, validation_fraction=0.1, verbose=False
              warm_start = False,
```

Figure 4.23: Multilayer Perceptron Neural Network Model that was built

Figure 4.23 demonstrates the multilayer perceptron neural network model that was built for this study. This is the output result of the model that was built, not the actual development model. The six models that were selected for deep learning all required the same procedures for the model to be built in PyCharm.

4.4.5 Model evaluation and Visualisation

4.4.5.1 Model Evaluation

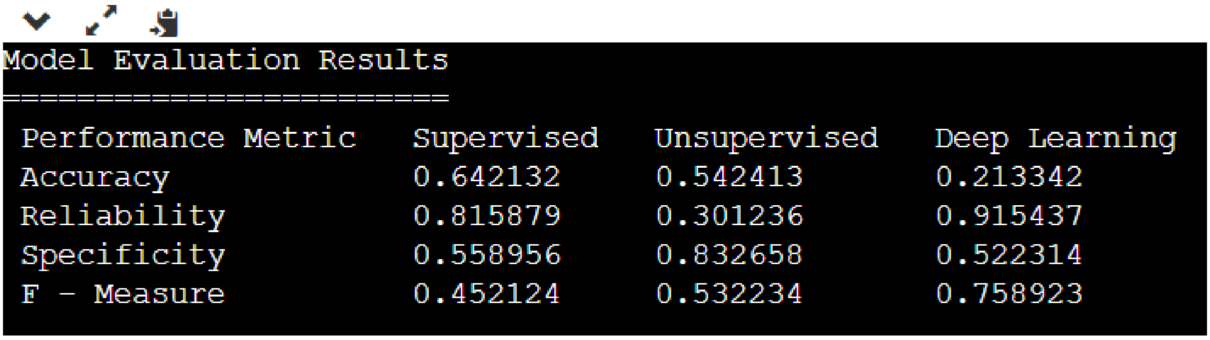
Using performance metrics methods, the methodology was able to identify the performance of the selected ML techniques.

Table 4.4 presents the model evaluation results.

Table 4.4: Model evaluation results

Performance Metric	Supervised	Unsupervised	Deep Learning
Accuracy	0.64	0.54	0.21
Reliability	0.82	0.31	0.92
Specificity	0.56	0.84	0.52
F – Measure	0.45	0.53	0.76
Final results	61.75	55.5	60.25

The performance metrics identified showed that the supervised ML technique produced the best results in the conceptual framework to identify retail complaints posted on Twitter. Deep learning followed, and previous research showed that unsupervised ML will produce better results than deep learning (Amruthnath and Gupta 2018; Pouyanfar *et al.* 2018; Kumar, Glisson and Cho 2020; Sebai, Wang and Wang 2020).



```
Model Evaluation Results
========================
 Performance Metric    Supervised   Unsupervised   Deep Learning
 Accuracy              0.642132     0.542413       0.213342
 Reliability           0.815879     0.301236       0.915437
 Specificity           0.558956     0.832658       0.522314
 F - Measure           0.452124     0.532234       0.758923
```

Figure 4.24: Model Evaluation Results

Figure 4.24 displays the model evaluation result output in PyCharm

4.4.5.2 Visualisation

The visualisation is a line graph of the model evaluation conducted. Once the model evaluation was conducted and the results were provided, the researcher developed a line graph in PyCharm using Matplotlib.

The line graph represents how the ML technique performed in the statistical measures, which are accuracy, reliability, specificity and F-measures. As each ML technique went through each statistical measure, you could compare the performance between the three ML techniques.

The supervised ML technique is represented in red, unsupervised ML technique is represented in blue and deep learning is represented in green. Supervised ML performed the best in accuracy, whereas deep learning produced the best result in reliability. Unsupervised ML performed the best in specificity, and deep learning again performed well in F-measures. No ML techniques produced consistent results. Yet, the supervised ML technique performed good results overall and did not perform badly in any of the statistical measures. Whereas deep learning gave good results and reliability but performed very badly in accuracy, and the unsupervised ML technique produced good results in specificity but very bad results in reliability, thereby bringing their overall score down.

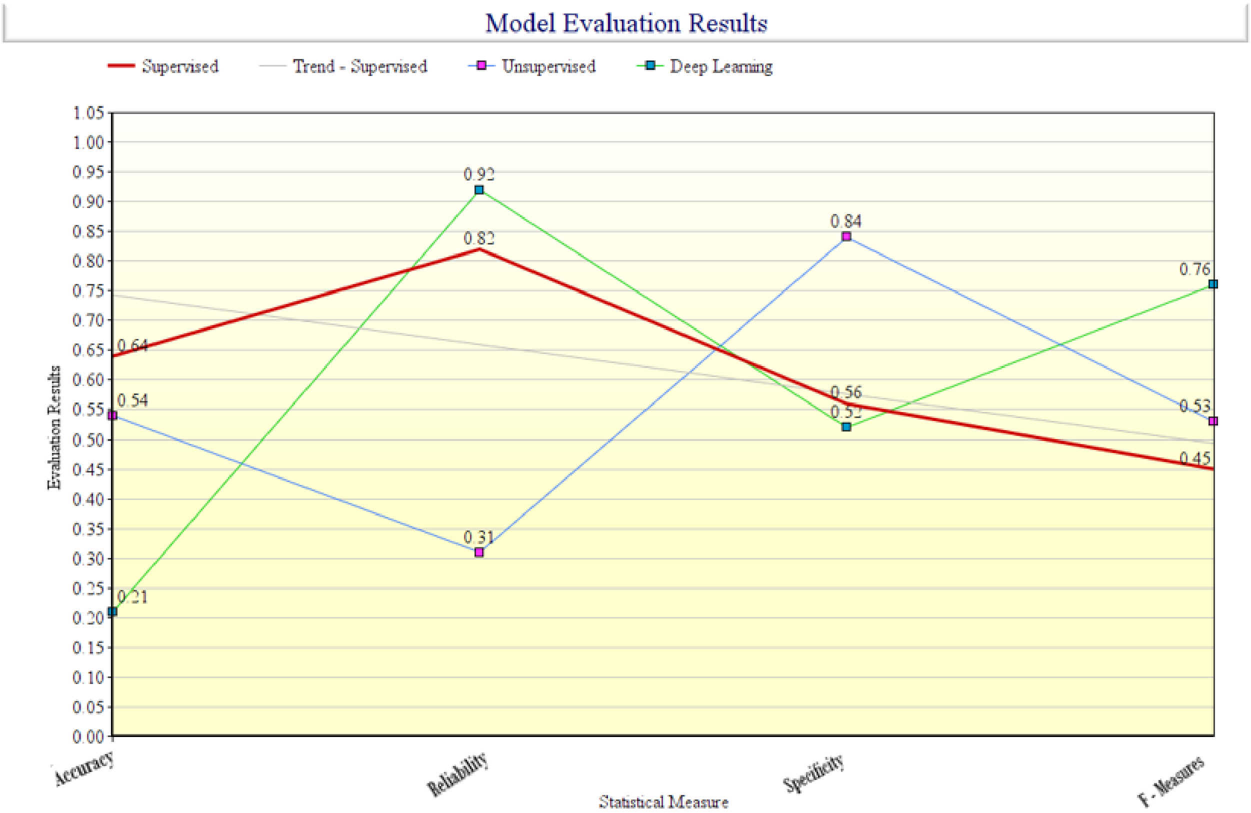


Figure 4.25: Visualisation Model Evaluation Diagram Displayed in PyCharm

Figure 4.25 displays the visualisation in the research based on the model evaluation results of the selected machine learning techniques.

4.5 Selection of ML technique

The result from each ML technique was added together, thereafter multiplied by 100, then divided by the number of statistical measures found in the traceability, which gives the value 4, which then gives the total performance result. The supervised ML technique performed the best with a result of 61.75, followed by deep learning, which yielded a result of 60.25, and finally the unsupervised ML technique, which gave a result of 55.5.

4.6 Conclusion

In summary, the results of this study showed that to get the best results on pre-processing, the techniques needed to be done in a certain order. The pre-processing techniques were applied in the following order: tokenisation, stemming, lemmatisation, part-of-speech tagging and then named entity recognition. Pre-processing decreased the file size by 888KB, which resulted in reducing the processing time to 1.15 seconds. The sentiment lexicon method was used for sentiment analysis, which resulted in 57% of the results being Neural, 18% being positive sentiment, and 24% being negative sentiment. Supervised ML was the best-performing technique, followed by deep learning and unsupervised. Previous studies have shown that unsupervised ML is better at performing sentiment analysis compared to deep learning (Amruthnath and Gupta 2018; Pouyanfar *et al.* 2018; Kumar, Glisson and Cho 2020; Sebai, Wang and Wang 2020). This study found that deep learning performed better than unsupervised ML.

Chapter 5 Conclusion and Recommendation

5.1 Introduction

Chapter 4 presented all of the study’s findings. This chapter presents a discussion of these findings and the research objectives. First, the summary of the study is discussed (Section 5.2); thereafter, the findings of research objective 1 (Section 5.3.1), research objective 2 (Section 5.3.2) and research objective 3 (Section 5.3.3) are discussed.

5.2 Summary of the Study

Sentiment analysis is a useful tool for businesses to understand consumers opinions about their services, but it contains several open challenges and research gaps. These challenges include building multilingual classifiers and the requirement to build domain-independent lexicons or classifiers to handle implicant word meaning.

Due to the ambiguities of natural language as well as the qualities of the content that was provided, the analysis of tweets is sometimes complicated with hashtags, emoticons and links, making detecting the expressed emotion challenging. Furthermore, automatic procedures that which require big datasets of annotated posts or lexical databases with emotional terms correlated with sentiment scores are required. Another key feature is that analyses are appropriate for the English language, whereas other languages have limitations.

There are various challenges in terms of design and application domains with ambiguous or scarce datasets in a variety of circumstances in the field of sentiment analysis. In addition, there is a shortage of labelled data, which can hinder progress in this field.

Getting the whole history of how customers arrived at their current sentiment is a critical research gap in sentiment analysis. Sentiment analysis of brief texts, such as single sentences and Twitter posts, is difficult due to the lack of contextual information in these texts. To complete this challenge successfully, it is essential to employ strategies that integrate the limited text content with prior knowledge and go beyond the bag-of-words approach. Elements of irony, humour, and other subtleties of human speech, such as the emoticon's ability to modify the tone of an otherwise negative message, are all examples of human speech. Social media sites are also populated with spam-posts which are posts that are perceived as unauthentic and untrustworthy.

False negatives occur when software recognises a negative term like "crap" but fails to recognise that it is positive in context—"Holy crap! This was fantastic!" Cultural differences exist, with some people from different countries being more or less effusive in their language use. A sentiment analytics framework has not been done in South Africa before, especially a sentiment analytics framework that is specifically for retail.

5.3 Research Objective

5.3.1 Research Objective 1: To investigate the characteristics of customer complaints on Twitter

According to the literature review, there are discrepancies between what a corporation claims in terms of products or services and what customers really get. There is a disconnect between how customers view the brand and the times when they do not get the right kind of customer service.

There could be various subcategories of consumer complaints. It could be related to a) inadequate communication, b) internal processes, or c) subpar service. When you recognise that a complaint is an opportunity, it is easier to convert conflict into positive change. If you pay close attention to consumer complaints, you can figure out how to fix them.

NLP pre-processing techniques concentrate on spotting intricate and interconnected patterns in consumer behaviour based on social media interactions. To interpret language more like humans do, NLP combines the disciplines of data science, computer science, and linguistics. NLP simplifies human language for autonomous machine analysis. Process enormous amounts of data in a matter of seconds or minutes, whereas human analysis would take days or weeks.

NLP pre-processing techniques were applied to the methodology to prepare the data for sentiment analysis. By calculating the frequency of positive, negative, and Neutral words and symbols in each tweet to establish its polarity, the researcher performed the analysis using the lexicon (i.e., positive, negative, or Neutral). The study awarded polarity scores to each word in the lexicon. Tweets with a negative polarity were regarded as complaints in the study.

5.3.2 Research Objective 2: To conduct a comparative performance analysis of ML algorithms used for customer sentiment analysis

The conceptual framework began with the extraction of data using TrackMyHashTag. Before The study entered in the retailer for the Name of retailer, Johannesburg for Area, followed by the Start date: 01 February 2017 and End: date: 30 November 2018. The conceptual framework extracted tweets, hashtag, location and timestamp from TrackMyHashTag based on the key words.

The extracted dataset was prepared using the following natural language pre-processing techniques (NLP): tokenisation, stemming and lemmatisation, part-of-speech (POS) tagging, named entity recognition (NER), and parser to extract emotions from the textual data from each tweet. The NLP pre-processing methods prepared the data for better sentiment analysis.

This resulted in the dataset decreasing in size from 100% to 79% and decreasing the processing time from 2.70 seconds to 1.15 seconds. The pre-processing stage prepared the data to go through the algorithms and sentiment analysis. A sentiment lexicon approach was used to analyse the sentiment of the dataset. The majority of the sentiments labelled were Neutral, the study found that sentiment lexicon method was not accurate in labelling the data.

The dataset resulted in 57% being labelled Neutral, 18% being labelled positive and 24% being labelled negative. Only the negative-labelled data was further used in the study. Within the realm of Neutral sentiment analysis, a substantial portion of the data could have been categorised as either negative or positive sentiment.

5.3.3 Research Objective 3: To develop and evaluate an ML-based customer sentiment analysis model incorporating the characteristics of customer complaints on Twitter to determine the best-performing ML algorithms

To prepare the data for evaluation for sentiment analysis, it was essential to first apply NLP pre-processing methods. The dataset was prepared using the following natural language pre-processing techniques: tokenisation, stemming and lemmatisation, part-of-speech (POS) tagging, named entity recognition (NER), and parsing to extract emotions from the textual data from each tweet.

The ML techniques were tested on their reliability, accuracy, specificity and F-measures. The study's results identified the supervised ML technique as the best-performing ML technique to identify retail complaints posted on social media. The supervised ML technique performed the best with a result of 61.75, followed by deep learning with a result of 60.25, and finally Unsupervised which gave the result of 55.5.

Within the performance metrics, the supervised ML technique performed the best in reliability, while deep learning produced the best results for performance and f-measure and specificity in the unsupervised ML.

5.4 Research Question: What ML technique can be used for accurate identification of customer complaints in Twitter data?

Training the dataset for supervised ML and deep learning was the most important phase for sentiment analysis. The research noted that supervised ML techniques were successful in classification problems where they classified a tweet as negative or positive. The performance of the classifier is dependent on the training procedure and the amount of labelled data fed into it as input for its learning purpose. Once the training was completed, testing was used to obtain the model's classification of accuracy and F-measure.

Deep learning methods are not better than traditional ML methods. Their performance depends on the appropriateness of the data structure and the size of the training data set. Deep learning effectively constructs text semantics compared to supervised and unsupervised methods.

Supervised ML did not require a trained dataset; this method can read unlabelled data. The K-means algorithm was used to find the hidden patterns in the unlabelled Twitter.

The accuracy and F-measures obtained by the unsupervised ML technique were 0.54 and 0.53. Whereas deep learning produced results of 0.21 and 0.76. Finally, the supervised ML obtained the overall best results of 0.64 and 0.45. The supervised ML produced the best results for accuracy, followed by unsupervised and then deep learning. Deep learning obtained the best results for F-measure followed by the unsupervised ML and then the supervised ML. The unsupervised method remained consistent between accuracy and F-measure.

5.5 Limitation

The results of this study should be interpreted in conjunction along with its four stated limitations. The first limitation is the scope of the study, which is the study is only conducted on Twitter. The second limitation of the research is that the Internet is changing and unpredictable suggesting the number of likes and comments on both Twitter and Facebook is not stagnant. The third limitation is that this study only addresses public communication, even though Twitter allows private Direct Messaging communication.

The fourth limitation is that companies and users delete posts, which are no longer available anymore and, therefore outside the scope of this research. Furthermore, all background information and social demographic characteristics of the users are missing. It is therefore recommended that researchers who use online research methods bear these limitations in mind. Future research should include other social media networks, especially Facebook. This research only targeted Twitter.

5.6 Recommendation for Future Research

Future work should consider comparing the state-of-the-art techniques presented in this research using the same data set across all the different techniques to be able to evaluate the best techniques used. This study identified the most efficient ML technique for sentiment analysis; future research should identify the best ML technique. Future works should identify methods to accurately identify the sentiment in Twitter data. More accurate pre-processing methods need to be investigated to only identify complaints posted on social media. This study had to sift through various data that was extracted from Twitter to only identify tweets with complaints. Moreover, as the task of collecting sentiment labels on social media is extremely time-consuming and tedious, unsupervised or deep learning approaches are required. Furthermore, this study only focused on Twitter data; future work should consider Facebook and Instagram data.

5.7 Conclusion

The study concludes that the extraction of customer complaints from Twitter necessitates the utilisation of NLP pre-processing methods. Subsequently, successful visualisation was achieved following data extraction. The NLP pre-processing methods selected for the study successfully prepared the data for sentiment analysis. The pre-processing stage reduced the size of the file and decreased the noise in the data. Even though sentiment analysis is an important tool to understand social media comments, it is not an accurate solution. Sentiment analysis at this stage cannot completely analyse social media ambiguities in natural language. This study noted that sometimes negative sentiments were classified as neutral sentiment tweets. In summary, the main research question was answered in this study. The discussions provided insightful explanations for the results and contributed to successfully addressing the aims and objectives of this study.

References

Adak, A., Pradhan, B. and Shukla, N. 2022. Sentiment analysis of customer reviews of food delivery services using deep learning and explainable artificial intelligence: Systematic review. *Foods*, 11 (10): 1500.

Adarsh, R., Patil, A., Rayar, S. and Veena, K. 2019. Comparison of VADER and LSTM for sentiment analysis. *International Journal of Recent Technology and Engineering*, 7 (6): 540-543.

Aggarwal, C. C. 2022. Opinion mining and sentiment analysis. In: *Machine learning for text*. Springer, 491-514.

Ain, Q. T., Ali, M., Riaz, A., Noureen, A., Kamran, M., Hayat, B. and Rehman, A. 2019. Sentiment analysis using deep learning techniques: a review. *Int J Adv Comput Sci Appl*, 8 (6): 424.

Akçay, M. B. and Oğuz, K. 2020. Speech emotion recognition: Emotional models, databases, features, preprocessing methods, supporting modalities, and classifiers. *Speech Communication*, 116: 56-76.

Alharbi, M. S. and El-kenawy, E.-S. M. 2021. Optimize Machine Learning Programming Algorithms for Sentiment Analysis in Social Media. *International Journal of Computer Applications*, 174 (25): 38-43.

Aljedaani, W., Rustam, F., Mkaouer, M. W., Ghallab, A., Rupapara, V., Washington, P. B., Lee, E. and Ashraf, I. 2022. Sentiment analysis on Twitter data integrating TextBlob and deep learning models: The case of US airline industry. *Knowledge-Based Systems*, 255: 109780.

Allem, J.-P., Dharmapuri, L., Unger, J. B. and Cruz, T. B. 2018. Characterizing JUUL-related posts on Twitter. *Drug and alcohol dependence*, 190: 1-5.

Alsaeedi, A. and Khan, M. Z. 2019. A study on sentiment analysis techniques of Twitter data. *International Journal of Advanced Computer Science and Applications*, 10 (2): 361-374.

Amruthnath, N. and Gupta, T. 2018. A research study on unsupervised machine learning algorithms for early fault detection in predictive maintenance. In: Proceedings of *2018 5th International Conference on Industrial Engineering and Applications (ICIEA)*. IEEE, 355-361.

Ariail Reed, I. 2023. Epistemology Contextualized: Social Scientific Knowledge in a Post-Positivist Era. In: *Sociology as a Human Science: Essays on Interpretation and Causal Pluralism*. Springer, 19-49.

Arisandy, T., Dasril, Y. B., Salahudin, S. N. B., Muslim, M. A., Adnan, A. and Wen, G. K. 2023. Buy Now Pay Later Services on Generation Z: Exploratory Data Analysis Using Machine Learning. *Journal of Theoretical and Applied Information Technology*, 101 (11): 4194-4204.

Arora, P., Gautam, S., Kalra, A., Negi, A. and Tyagi, N. 2022. Credit Risk Analysis Using EDA. In: *Emerging Technologies in Data Mining and Information Security: Proceedings of IEMIS 2022, Volume 2*. Springer, 563-574.

Arora, S., Malhotra, A., Kumar, A. and Singh, N. 2023. Sentiment Analysis Unleashed: Leveraging VADER for Evaluation of Public Opinion. In: Proceedings of *2023 11th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions)(ICRITO)*. IEEE, 1-6.

Ashfaq, M. 2019. After Sales Service, Customer Satisfaction and Loyalty in Telecom Sector. *Journal of Applied Structural Equation Modeling*, 3 (1): 31-41.

Balaji, M., Jha, S. and Royne, M. B. 2019. Customer e-complaining behaviours using social media. *The Service Industries Journal*, 35 (11-12): 633-654.

Balaji, T., Annavarapu, C. S. R. and Bablani, A. 2021. Machine learning algorithms for social media analysis: A survey. *Computer Science Review*, 40: 100395.

Basarslan, M. S. and Kayaalp, F. 2020. Sentiment Analysis with Machine Learning Methods on Social Media.

Beigi, O. M. and Moattar, M. H. 2018. Automatic construction of domain-specific sentiment lexicon for unsupervised domain adaptation and sentiment classification. *Knowledge-Based Systems*, 213: 106423.

Benyahia, S., Meftah, B. and Lézoray, O. 2022. Multi-features extraction based on deep learning for skin lesion classification. *Tissue and Cell*, 74: 101701.

Beysolow II, T. 2018. *Applied Natural Language Processing with Python: Implementing Machine Learning and Deep Learning Algorithms for Natural Language Processing*. Apress.

Bibi, M., Abbasi, W. A., Aziz, W., Khalil, S., Uddin, M., Iwendi, C. and Gadekallu, T. R. 2022. A novel unsupervised ensemble framework using concept-based linguistic methods and machine learning for twitter sentiment analysis. *Pattern Recognition Letters*, 158: 80-86.

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

Blum, A., Hopcroft, J. and Kannan, R. 2020. *Foundations of data science*. Cambridge University Press.

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

Bouadjenek, M. R., Sanner, S., Iman, Z., Xie, L. and Shi, D. X. 2022. A longitudinal study of topic classification on Twitter. *PeerJ Computer Science*, 8: e991.

Brewer, S. W. 2018. Come for a job, stay for the socializing: gratifications received from LinkedIn usage. *Online Journal of Communication and Media Technologies*, 8 (4): 345-361.

Buratti, N., Parola, F. and Satta, G. 2018. Insights on the adoption of social media marketing in B2B services. *The TQM Journal*, Carrera-Rivera, A., Larrinaga, F. and Lasa, G. 2022. Context-awareness for the design of Smart-product service systems: Literature review. *Computers in Industry*, 142: 103730.

Carvalho, N. M. V. 2018. A Data Analysis approach to study events' influence in Social Networks. Universidade do Minho (Portugal).

Chamarczuk, M., Nishitsuji, Y., Malinowski, M. and Draganov, D. 2020. Unsupervised Learning Used in Automatic Detection and Classification of Ambient-Noise Recordings from a Large-N Array. *Seismological Research Letters*, 91 (1): 370-389.

Chen, L.-C., Lee, C.-M. and Chen, M.-Y. 2020. Exploration of social media for sentiment analysis using deep learning. *Soft Computing*, 24 (11): 8187-8197.

COVID, I., Team, F. and Hay, S. I. 2020. COVID-19 scenarios for the United States. *medRxiv*,

Da Veiga, A. and Ophoff, J. 2021. Concern for information privacy: a cross-nation study of the United Kingdom and South Africa. In: Proceedings of *International Symposium on Human Aspects of Information Security and Assurance*. Springer, 16-29.

Dahal, N. 2023. Is doing research for people or on people? A book review of Emancipatory and Participatory Research for Emerging Educational Researchers: Theory and Case Studies of Research in Disabled Communities. *The Qualitative Report*, 28 (1): 69-74.

Dashtipour, K., Gogate, M., Adeel, A., Larijani, H. and Hussain, A. 2021. Sentiment analysis of persian movie reviews using deep learning. *Entropy*, 23 (5): 596.

Davis, J., Wolff, H.-G., Forret, M. L. and Sullivan, S. E. 2020. Networking via LinkedIn: An examination of usage and career benefits. *Journal of Vocational Behavior*, 118: 103396.

Donders, F., Lonnée-Hoffmann, R., Tsiakalos, A., Mendling, W., Martinez de Oliveira, J., Judlin, P., Xue, F., Donders, G. G., COVID, I. and Workgroup, G. 2020. ISIDOG recommendations concerning COVID-19 and pregnancy. *Diagnostics*, 10 (4): 243.

Dong, S., Wang, P. and Abbas, K. 2021. A survey on deep learning and its applications. *Computer Science Review*, 40: 100379.

Dong, X., Yu, Z., Cao, W., Shi, Y. and Ma, Q. 2020. A survey on ensemble learning. *Frontiers of Computer Science*, 14 (2): 241-258.

Dreisbach, C., Koleck, T. A., Bourne, P. E. and Bakken, S. 2019. A systematic review of natural language processing and text mining of symptoms from electronic patient-authored text data. *International journal of medical informatics*, 125: 37-46.

Ducange, P., Fazzolari, M., Petrocchi, M. and Vecchio, M. 2019. An effective Decision Support System for social media listening based on cross-source sentiment analysis models. *Engineering Applications of Artificial Intelligence*, 78: 71-85.

Duffy, B. E. and Chan, N. K. 2019. “You never really know who’s looking”: Imagined surveillance across social media platforms. *New Media & Society*, 21 (1): 119-138.

Duica, M. C., Florea, N. V. and Dobrescu, T. A. 2019. Obtaining Value and Performance by Managing Effectively the Customers' Complaints. *Valahian Journal of Economic Studies*, 10 (1): 29-38.

El Rahman, S. A., AlOtaibi, F. A. and AlShehri, W. A. 2019. Sentiment analysis of twitter data. In: Proceedings of *2019 international conference on computer and information sciences (ICCIS)*. IEEE, 1-4.

Elbagir, S. and Yang, J. 2019. Twitter sentiment analysis using natural language toolkit and VADER sentiment. In: Proceedings of *Proceedings of the international multiconference of engineers and computer scientists*. 16.

Elinda, E., Yuliansyah, H. and Latiffi, M. I. A. 2023. Sentiment Analysis of the Sheikh Zayed Grand Mosque’s Visitor Reviews on Google Maps Using the VADER Method. *International Journal of Advances in Data and Information Systems*, 5 (1): 71-84.

Erşahin, B., Aktaş, Ö., Kilinç, D. and Erşahin, M. 2019. A hybrid sentiment analysis method for Turkish. *Turkish Journal of Electrical Engineering and Computer Sciences*, 27 (3): 1780-1793.

Ferrag, M. A., Maglaras, L., Moschoyiannis, S. and Janicke, H. 2020. Deep learning for cyber security intrusion detection: Approaches, datasets, and comparative study. *Journal of Information Security and Applications*, 50: 102419.

Freihat, A. A., Khalilia, H., Bella, G. and Giunchiglia, F. 2023. Advancing the Arabic WordNet: Elevating Content Quality. *arXiv preprint arXiv:2403.20215*,

Froot, K., Kang, N., Ozik, G. and Sadka, R. 2017. What do measures of real-time corporate sales say about earnings surprises and post-announcement returns? *Journal of Financial Economics*, 125 (1): 143-162.

Froot, K., Kang, N., Ozik, G. and Sadka, R. 2019. What do measures of real-time corporate sales say about earnings surprises and post-announcement returns? *Journal of Financial Economics*, 125 (1): 143-162.

Geyik, S. C., Ambler, S. and Kenthapadi, K. 2019. Fairness-aware ranking in search & recommendation systems with application to linkedin talent search. In: Proceedings of *Proceedings of the 25th acm sigkdd international conference on knowledge discovery & data mining*. 2221-2231.

Goga, S. and Paelo, A. 2019. *Issues in the regulation and policy surrounding e-commerce in South Africa*. Working Paper 2019/6. Johannesburg: CCRED.

Goularas, D. and Kamis, S. 2019. Evaluation of deep learning techniques in sentiment analysis from twitter data. In: Proceedings of *2019 International Conference on Deep Learning and Machine Learning in Emerging Applications (Deep-ML)*. IEEE, 12-17.

Gowri, S., Jabez, J., Vimali, J., Sivasangari, A. and Srinivasulu, S. 2021. Sentiment Analysis of Twitter Data Using Techniques in Deep Learning. In: *Data Intelligence and Cognitive Informatics*. Springer, 613-623.

Greenhalgh, S. P., Rosenberg, J. M., Willet, K. B. S., Koehler, M. J. and Akcaoglu, M. 2020. Identifying multiple learning spaces within a single teacher-focused Twitter hashtag. *Computers & education*, 148: 103809.

Grégoire, Y., Salle, A. and Tripp, T. M. 2019. Managing social media crises with your customers: The good, the bad, and the ugly. *Business Horizons*, 58 (2): 173-182.

Gunarathne, P., Rui, H. and Seidmann, A. 2017. Whose and what social media complaints have happier resolutions? Evidence from Twitter. *Journal of Management Information Systems*, 34 (2): 314-340.

Gunarathne, P., Rui, H. and Seidmann, A. 2019. Whose and what social media complaints have happier resolutions? Evidence from Twitter. *Journal of Management Information Systems*, 34 (2): 314-340.

Han, H., Ma, W., Zhou, M., Guo, Q. and Abusorrah, A. 2020. A novel semi-supervised learning approach to pedestrian reidentification. *IEEE Internet of Things Journal*, 8 (4): 3042-3052.

Haupt, M. R., Jinich-Diamant, A., Li, J., Nali, M. and Mackey, T. K. 2021. Characterizing twitter user topics and communication network dynamics of the "Liberate" movement during COVID-19 using unsupervised machine learning and social network analysis. *Online Social Networks and Media*, 21: 100114.

Hayon, S., Tripathi, H., Stormont, I. M., Dunne, M. M., Naslund, M. J. and Siddiqui, M. M. 2019. Twitter mentions and academic citations in the urologic literature. *Urology*, 123: 28-33.

Hswen, Y., Xu, X., Hing, A., Hawkins, J. B., Brownstein, J. S. and Gee, G. C. 2021. Association of "# covid19" versus "# chinesevirus" with anti-Asian sentiments on Twitter: March 9–23, 2020. *American Journal of Public Health*, 111 (5): 956-964.

Huang, D., Chen, Q., Huang, J., Kong, S. and Li, Z. 2021. Customer-robot interactions: Understanding customer experience with service robots. *International Journal of Hospitality Management*, 99: 103078.

Huang, F., Li, R., Wang, G., Zheng, J., Tang, Y., Liu, J., Yang, Y., Yao, Y., Shi, J. and Hong, W. 2020. Automatic classification of single-molecule charge transport data with an unsupervised machine-learning algorithm. *Physical Chemistry Chemical Physics*, 22 (3): 1674-1681.

Ibrahim, N. F. and Wang, X. 2019. A text analytics approach for online retailing service improvement: Evidence from Twitter. *Decision Support Systems*, 121: 37-50.

Iglesias, C. Á. and Sánchez-Rada, J. F. 2021. Sentiment Analysis meets Linguistic Linked Data: An overview of the State of the Art.

Jagdale, R. S., Shirsat, V. S. and Deshmukh, S. N. 2019. Sentiment analysis on product reviews using machine learning techniques. In: *Cognitive Informatics and Soft Computing*. Springer, 639-647.

Jaller, M. and Pahwa, A. 2020. Evaluating the environmental impacts of online shopping: A behavioral and transportation approach. *Transportation Research Part D: Transport and Environment*, 80: 102223.

Jiang, T., Gradus, J. L. and Rosellini, A. J. 2020. Supervised machine learning: a brief primer. *Behavior Therapy*, 51 (5): 675-687.

Jindal, K. and Aron, R. 2021. A systematic study of sentiment analysis for social media data. *Materials today: proceedings*,

Jumaryadi, Y. 2019. Customer Complaint Information Systems at National Standardization Agency of Indonesia. *Int. J. Inf. Syst. Comput. Sci*, 3 (2): 43-49.

Kanakaraj, M. and Guddeti, R. M. R. 2019. NLP based sentiment analysis on Twitter data using ensemble classifiers. In: Proceedings of *2015 3Rd international conference on signal processing, communication and networking (ICSCN)*. IEEE, 1-5.

Kao, S.-C., Wu, C.-H. and Syu, S.-W. 2018. A creative idea exploration model: based on customer complaints. In: Proceedings of *Proceedings of the 5th Multidisciplinary International Social Networks Conference*. 1-8.

Karami, A., Lundy, M., Webb, F. and Dwivedi, Y. K. 2020. Twitter and research: a systematic literature review through text mining. *IEEE Access*, 8: 67698-67717.

Kauffmann, E., Peral, J., Gil, D., Ferrández, A., Sellers, R. and Mora, H. 2020. A framework for big data analytics in commercial social networks: A case study on sentiment analysis and fake review detection for marketing decision-making. *Industrial Marketing Management*, 90: 523-537.

Kaur, A. and Singh, R. Service recovery paradox: A study of after sale services in Tractor market.

Kausar, M. A., Soosaimanickam, A. and Nasar, M. 2021. Public sentiment analysis on Twitter data during COVID-19 outbreak. *Int. J. Adv. Comput. Sci. Appl*, 12 (2): 415-422.

Kazmaier, J. and van Vuuren, J. H. 2022. The power of ensemble learning in sentiment analysis. *Expert Systems with Applications*, 187: 115819.

Khai Tran, T. and Thi Phan, T. 2019. Deep learning application to ensemble learning—the simple, but effective, approach to sentiment classifying. *Applied Sciences*, 9 (13): 2760.

Khattak, A., Asghar, M. Z., Ishaq, Z., Bangyal, W. H. and Hameed, I. A. 2021. Enhanced concept-level sentiment analysis system with expanded ontological relations for efficient classification of user reviews. *Egyptian Informatics Journal*, 22 (4): 455-471.

Klassen, K. M., Borleis, E. S., Brennan, L., Reid, M., McCaffrey, T. A. and Lim, M. S. 2018. What people "like": Analysis of social media strategies used by food industry brands, lifestyle brands, and health promotion organizations on Facebook and Instagram. *Journal of medical Internet research*, 20 (6): e10227.

Korzer, T., Hübscher, M., Schade, K. and Ringel, J. 2020. Rethinking retail digitisation in urban settings. The case of Leipzig, Germany. *Proceedings of the Institution of Civil Engineers-Urban Design and Planning*, 173 (5): 197-209.

Kühl, N., Goutier, M., Ensslen, A. and Jochem, P. 2019. Literature vs. Twitter: Empirical insights on customer needs in e-mobility. *Journal of cleaner production*, 213: 508-520.

Kumar, A., Glisson, W. and Cho, H. 2020. Network attack detection using an unsupervised machine learning algorithm.

Kumar, V. V., Raghunath, K., Muthukumaran, V., Joseph, R. B., Beschi, I. and Uday, A. K. 2022. Aspect based sentiment analysis and smart classification in uncertain feedback pool. *International Journal of System Assurance Engineering and Management*, 13 (1): 252-262.

Kundu, S. G., Ghosh, A., Kundu, A. and GP, G. 2022. A ML-AI ENABLED ENSEMBLE MODEL FOR PREDICTING AGRICULTURAL YIELD. *Cogent Food & Agriculture*, 8 (1): 2085717.

Kursuncu, U., Gaur, M., Lokala, U., Thirunarayan, K., Sheth, A. and Arpinar, I. B. 2019. Predictive analysis on Twitter: Techniques and applications. In: *Emerging research challenges and opportunities in computational social network analysis and mining*. Springer, 67-104.

Lacinski, L., Liming, L., Turoscy, S., Harr, C., Chard, K., Dart, E., Durack, P., Ames, S., Hoffman, F. M. and Foster, I. T. 2023. Automated, Reliable, and Efficient Continental-Scale Replication of 7.3 Petabytes of Climate Simulation Data: A Case Study. *arXiv preprint arXiv:2404.19717*,

Lahti, T. 2018. Improving retailer satisfaction: Commissioner Green Room Ltd.

Li, X., Zhao, H., Yu, L., Chen, H., Deng, W. and Deng, W. 2022. Feature extraction using parameterized multisynchrosqueezing transform. *IEEE Sensors Journal*, 22 (14): 14263-14272.

Li, Y., Tang, B. and Jiao, S. 2023. SO-slope entropy coupled with SVMD: A novel adaptive feature extraction method for ship-radiated noise. *Ocean Engineering*, 280: 114677.

Lohiniva, A.-L., Nurzhynska, A., Hudi, A.-h. and Anim, B. 2022. Infodemic Management Using Digital Information and Knowledge Cocreation to Address COVID-19 Vaccine Hesitancy: Case Study From Ghana. *Jmir Infodemiology*, 2 (2): e37134.

Lu, H., Uddin, S., Hajati, F., Moni, M. A. and Khushi, M. 2022. A patient network-based machine learning model for disease prediction: The case of type 2 diabetes mellitus. *Applied Intelligence*, 52 (3): 2411-2422.

Luo, S. and Weng, Y. 2019. A two-stage supervised learning approach for electricity price forecasting by leveraging different data sources. *Applied energy*, 242: 1497-1512.

Mackey, T., Purushothaman, V., Li, J., Shah, N., Nali, M., Bardier, C., Liang, B., Cai, M. and Cuomo, R. 2020. Machine learning to detect self-reporting of symptoms, testing access, and recovery associated with COVID-19 on Twitter: retrospective big data infoveillance study. *JMIR public health and surveillance*, 6 (2): e19509.

Mahesh, B. 2020. Machine Learning Algorithms-A Review. *International Journal of Science and Research (IJSR).[Internet]*, 9: 381-386.

Majumder, M. G., Gupta, S. D. and Paul, J. 2022. Perceived usefulness of online customer reviews: A review mining approach using machine learning & exploratory data analysis. *Journal of Business Research*, 150: 147-164.

Mao, R., Li, X., Ge, M. and Cambria, E. 2022. MetaPro: A computational metaphor processing model for text pre-processing. *Information Fusion*, 86: 30-43.

Marreddy, M. and Mamidi, R. 2023. Learning sentiment analysis with word embeddings. In: *Computational Intelligence Applications for Text and Sentiment Data Analysis*. Elsevier, 141-161.

Maudslay, R. H., Teufel, S., Bond, F. and Pustejovsky, J. 2023. ChainNet: Structured Metaphor and Metonymy in WordNet. *arXiv preprint arXiv:2403.20308*,

McKibbin, W. and Fernando, R. 2020. The economic impact of COVID-19. *Economics in the Time of COVID-19*, 45 (10.1162)

Mei, X. Y., Bagaas, I. K. and Relling, E. K. 2019. Customer complaint behaviour (CCB) in the retail sector: why do customers voice their complaints on Facebook? *The International Review of Retail, Distribution and Consumer Research*, 29 (1): 63-78.

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

Milo, T. and Somech, A. 2020. Automating exploratory data analysis via machine learning: An overview. In: Proceedings of *Proceedings of the 2020 ACM SIGMOD International Conference on Management of Data*. 2617-2622.

Mishev, K., Gjorgjevikj, A., Vodenska, I., Chitkushev, L. T. and Trajanov, D. 2020. Evaluation of sentiment analysis in finance: from lexicons to transformers. *IEEE Access*, 8: 131662-131682.

Mishra, S. and Singh, S. P. 2020. Designing dynamic reverse logistics network for post-sale service. *Annals of Operations Research*: 1-30.

Mohamad Sham, N. and Mohamed, A. 2022. Climate Change Sentiment Analysis Using Lexicon, Machine Learning and Hybrid Approaches. *Sustainability*, 14 (8): 4723.

Molinillo, S., Aguilar-Illescas, R., Anaya-Sánchez, R. and Carvajal-Trujillo, E. 2022. The customer retail app experience: Implications for customer loyalty. *Journal of Retailing and Consumer Services*, 65: 102842.

Moon, N. N., Salehin, I., Parvin, M., Hasan, M., Talha, I. M., Debnath, S. C., Nur, F. N. and Saifuzzaman, M. 2021. Natural language processing based advanced method of unnecessary video detection. *International Journal of Electrical & Computer Engineering (2088-8708)*, 11 (6)

Moore, S., Bulmer, S. and Elms, J. 2022. The social significance of AI in retail on customer experience and shopping practices. *Journal of Retailing and Consumer Services*, 64: 102755.

Munthali, N., Van Paassen, A., Leeuwis, C., Lie, R., van Lammeren, R., Aguilar-Gallegos, N. and Oppong-Mensah, B. 2021. Social media platforms, open communication and problem solving in the back-office of Ghanaian extension: A substantive, structural and relational analysis. *Agricultural Systems*, 190: 103123.

Murthy, D. 2018. *Twitter*. Polity Press Cambridge.

Mustofa, M. Y., Nadhifah, N., Djamil, A. and Irsyad, M. F. 2023. A Critical Analysis of Auguste Comte's Positivism for Islamic Digital Research. *Islamic Review: Jurnal Riset Dan Kajian Keislaman*, 12 (1): 1-18.

Naldi, M. 2019. A review of sentiment computation methods with R packages. *arXiv preprint arXiv:1901.08319*, Nanda, C., Dua, M. and Nanda, G. 2018. Sentiment analysis of movie reviews in hindi language using machine learning. In: Proceedings of *2018 International Conference on Communication and Signal Processing (ICCSP)*. IEEE, 1069-1072.

Noori, B. 2021. Classification of customer reviews using machine learning algorithms. *Applied Artificial Intelligence*, 35 (8): 567-588.

Obaidi, M. and Klünder, J. 2021. Development and application of sentiment analysis tools in software engineering: A systematic literature review. *Evaluation and Assessment in Software Engineering*: 80-89.

Okazaki, S., Díaz-Martín, A. M., Rozano, M. and Menéndez-Benito, H. D. 2020a. Using Twitter to engage with customers: a data mining approach. *Internet Research*,

Okazaki, S., Plangger, K., West, D. and Menéndez, H. D. 2020b. Exploring digital corporate social responsibility communications on Twitter. *Journal of Business Research*, 117: 675-682.

Ozili, P. K. and Arun, T. 2020. Spillover of COVID-19: impact on the Global Economy. *Available at SSRN 3562570*,

Pandian, A. P. 2021. Performance Evaluation and Comparison using Deep Learning Techniques in Sentiment Analysis. *Journal of Soft Computing Paradigm (JSCP)*, 3 (02): 123-134.

Pano, T. and Kashef, R. 2020. A complete VADER-based sentiment analysis of bitcoin (BTC) tweets during the era of COVID-19. *Big Data and Cognitive Computing*, 4 (4): 33.

Pantano, E., Giglio, S. and Dennis, C. 2019. Making sense of consumers' tweets: Sentiment outcomes for fast fashion retailers through Big Data analytics. *International Journal of Retail & Distribution Management*,

Peikari, M., Salama, S., Nofech-Mozes, S. and Martel, A. L. 2018. A cluster-then-label semi-supervised learning approach for pathology image classification. *Scientific reports*, 8 (1): 1-13.

Penu, S. L. and Nkanbia-Davies, L. O. 2019. Implications of Exchange Rate Fluctuation and Economic Performance: The Nigerian Experience. *International Journal of Economics and Financial Research*, 3 (10): 206-217.

Pilkington, M. 2021. *Retail Recovery: How Creative Retailers Are Winning in their Post-Apocalyptic World*. Bloomsbury Publishing.

Pouyanfar, S., Sadiq, S., Yan, Y., Tian, H., Tao, Y., Reyes, M. P., Shyu, M.-L., Chen, S.-C. and Iyengar, S. 2018. A survey on deep learning: Algorithms, techniques, and applications. *ACM Computing Surveys (CSUR)*, 51 (5): 1-36.

Pradha, S., Halgamuge, M. N. and Vinh, N. T. Q. 2019. Effective text data preprocessing technique for sentiment analysis in social media data. In: Proceedings of *2019 11th international conference on knowledge and systems engineering (KSE)*. IEEE, 1-8.

Preotiuc-Pietro, D., Gaman, M. and Aletras, N. 2019. Automatically identifying complaints in social media. *arXiv preprint arXiv:1906.03890*,

Priyavrat, A. 2017. Sentiment Analysis: A Comparative Study of Supervised Machine Learning Algorithms Using Rapid miner. *International Journal for Research in Applied Science & Engineering Technology (IJRASET) ISSN*: 2321-9653.

Rahimi, N., Eassa, F. and Elrefaei, L. 2020. An ensemble machine learning technique for functional requirement classification. *symmetry*, 12 (10): 1601.

Rathor, A. S., Agarwal, A. and Dimri, P. 2018. Comparative study of machine learning approaches for Amazon reviews. *Procedia Computer Science*, 132: 1552-1561.

Reichert, J.-R., Kristensen, K. L., Mukkamala, R. R. and Vatrapu, R. 2017. A supervised machine learning study of online discussion forums about type-2 diabetes. In: Proceedings of *2017 IEEE 19Th International Conference on E-health Networking, Applications and Services (Healthcom)*. IEEE, 1-7.

Reichert, J.-R., Kristensen, K. L., Mukkamala, R. R. and Vatrapu, R. 2019. A supervised machine learning study of online discussion forums about type-2 diabetes. In: Proceedings of *2017 IEEE 19Th International Conference on E-health Networking, Applications and Services (Healthcom)*. IEEE, 1-7.

Ren, H., El-Khamy, M. and Lee, J. 2020. Stereo disparity estimation via joint supervised, unsupervised, and weakly supervised learning. In: Proceedings of *2020 IEEE International Conference on Image Processing (ICIP)*. IEEE, 2760-2764.

Ren, H., Raj, A., El-Khamy, M. and Lee, J. 2020. Suw-learn: Joint supervised, unsupervised, weakly supervised deep learning for monocular depth estimation. In: Proceedings of *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*. 750-751.

Ridge, M., Johnston, K. A. and O'Donovan, B. 2021. The use of big data analytics in the retail industries in South Africa. *African Journal of Business Management*, 9 (19): 688-703.

Rodrigues, A. P., Fernandes, R., Shetty, A., Lakshmanna, K. and Shafi, R. M. 2022. Real-time twitter spam detection and sentiment analysis using machine learning and deep learning techniques. *Computational Intelligence and Neuroscience*, 2022

Sadhasivam, J. and Kalivaradhan, R. B. 2019. Sentiment analysis of Amazon products using ensemble machine learning algorithm. *International Journal of Mathematical, Engineering and Management Sciences*, 4 (2): 508.

Saranya, S. and Usha, G. 2023. A Machine Learning-Based Technique with IntelligentWordNet Lemmatize for Twitter Sentiment Analysis. *Intelligent Automation & Soft Computing*, 36 (1)

Savitha, K. 2021. A Study on Using the Internet of Things (Iot) to Analyse the Consumer behavior in Household Appliances in Tamilnadu. *Annals of the Romanian Society for Cell Biology*, 25 (6): 18367-18373.

Sebai, M., Wang, X. and Wang, T. 2020. MaskMitosis: a deep learning framework for fully supervised, weakly supervised, and unsupervised mitosis detection in histopathology images. *Medical & Biological Engineering & Computing*, 58 (7): 1603-1623.

Serravalle, F. 2020. Augmented reality in retail: an analysis of this immersive technology through consumers' and retailer's perception.

Shearer, E. and Mitchell, A. 2021. News use across social media platforms in 2020.

Shin, H. and Larson, L. R. 2020. The bright and dark sides of humorous response to online customer complaint. *European Journal of Marketing*,

Shinde, G. K., Lokhande, V. N., Kalyane, R. T., Gore, V. B. and Raut, U. M. 2021. Sentiment analysis using hybrid approach. *International journal for research in applied science and engineering technology (IJRASET)*, 9: 282-285.

Silva e Sousa, R., Pinho, T. and Simões, D. 2022. How can customer experience improve retail operations sustainability?

Singh, A., Saha, S., Hasanuzzaman, M. and Jangra, A. 2021. Identifying complaints based on semi-supervised mincuts. *Expert Systems with Applications*: 115668.

Smirnova, A. and Cudré-Mauroux, P. 2018. Relation extraction using distant supervision: A survey. *ACM Computing Surveys (CSUR)*, 51 (5): 1-35.

Smith, M. D. and Telang, R. 2018. Data can enhance creative projects—Just look at Netflix. *Harvard Business Review*, 23

Soong, H.-C., Ayyasamy, R. K. and Akbar, R. 2021. A review towards deep learning for sentiment analysis. In: Proceedings of *2021 International Conference on Computer & Information Sciences (ICCOINS)*. IEEE, 238-243.

Soong, H.-C., Jalil, N. B. A., Ayyasamy, R. K. and Akbar, R. 2019. The essential of sentiment analysis and opinion mining in social media: Introduction and survey of the recent approaches and techniques. In: Proceedings of *2019 IEEE 9th symposium on computer applications & industrial electronics (ISCAIE)*. IEEE, 272-277.

Sridevi, P. C. and Velmurugan, T. 2023. Twitter Sentiment Analysis of COVID-19 Vaccination Integrating SenticNet-7 and SentiWordNet-Adjusted VADER Models. *International Journal of Computer Information Systems and Industrial Management Applications*, 16 (2): 22-22.

Srivastava, R., Bharti, P. and Verma, P. 2022. Comparative Analysis of Lexicon and Machine Learning Approach for Sentiment Analysis. *International Journal of Advanced Computer Science and Applications*, 13 (3)

Stracqualursi, L. and Agati, P. 2022. Tweet topics and sentiments relating to distance learning among Italian Twitter users. *Scientific reports*, 12 (1): 1-11.

Struwig, M. 2018. Consumers' perception of eco-labels in South Africa.

Sudirman, I. D. and Nugraha, D. Y. 2023. SentiWordNet and LDA Analysis of Foreign Tourist. In: Proceedings of *2024 3rd International Conference on Digital Transformation and Applications (ICDXA)*. IEEE, 1-6.

Sultana, N. and Islam, M. M. 2020. Meta classifier-based ensemble learning for sentiment classification. In: Proceedings of *Proceedings of International Joint Conference on Computational Intelligence*. Springer, 73-84.

Takashima, M., Suh, S.-O., Bai, F.-Y. and Sugita, T. 2019. Takashi Nakase's last tweet: what is the current direction of microbial taxonomy research? *FEMS yeast research*, 19 (8): foz066.

Tang, C., Fan, B., Zhao, J., Liang, C., Wang, Y., Wang, B., Qiu, Z., Qiu, L., Ding, B. and Sun, S. 2024. Data Caching for Enterprise-Grade Petabyte-Scale OLAP. *arXiv preprint arXiv:2406.05962*,

Taylor, S., Wilson, E., Murfitt, K. and Balandin, S. 2022. Self-advocacy with retailers by people with complex communication needs. *International Journal of Speech-Language Pathology*: 1-10.

Teh, P. L., Low, J. Q. and Si, Y.-W. 2020. Geographical distance and news diffusion associating with the sentiment of tweets: Case Study on London Bridge Attack 2017. In: Proceedings of *2020 IEEE Intl Conf on Parallel & Distributed Processing with Applications, Big Data & Cloud Computing, Sustainable Computing & Communications, Social Computing & Networking (ISPA/BDCloud/SocialCom/SustainCom)*. IEEE, 1481-1486.

TEHCİ, A. and Ersoy, Y. 2020. Investigation of digital retail consumer complaints in the food industry during COVID-19: Market chain example of Turkey. *The Journal of International Scientific Researches*, 5 (Ek): 22-27.

Vanhaeren, T., Divina, F., García-Torres, M., Gómez-Vela, F., Vanhoof, W. and Martínez-García, P. M. 2020. A comparative study of supervised machine learning algorithms for the prediction of long-range chromatin interactions. *Genes*, 11 (9): 985.

Vashishtha, S. and Susan, S. 2019. Fuzzy rule based unsupervised sentiment analysis from social media posts. *Expert Systems with Applications*, 138: 112834.

Vega Bernal, M. 2017. Retail payments innovations in Peru: Modelo Peru and financial inclusion. *Journal of Payments Strategy & Systems*, 10 (4): 343-351.

Vega Bernal, M. 2019. Retail payments innovations in Peru: Modelo Peru and financial inclusion. *Journal of Payments Strategy & Systems*, 10 (4): 343-351.

Vega, J. 2022. Constructivist Metaphors and Law's Autonomy in Legal Post-Positivism. In: *Human Dignity and the Autonomy of Law*. Springer, 89-111.

Wang, G., Zhao, B., Wu, B., Zhang, C. and Liu, W. 2023. Intelligent prediction of slope stability based on visual exploratory data analysis of 77 in situ cases. *International Journal of Mining Science and Technology*, 33 (1): 47-59.

WANG, Q. 2018. Using Natural Language Processing and Data Mining Techniques for Amazon Reviews Data Analytics: A study.

Wankhade, M., Rao, A. C. S. and Kulkarni, C. 2022. A survey on sentiment analysis methods, applications, and challenges. *Artificial Intelligence Review*: 1-50.

Wedel, I., Palk, M. and Voß, S. 2021. A Bilingual Comparison of Sentiment and Topics for a Product Event on Twitter. *Information Systems Frontiers*: 1-12.

Wei, Q., Shi, X., Li, Q. and Chen, G. 2020. Enhancing Customer Satisfaction Analysis with a Machine Learning Approach: From a Perspective of Matching Customer Comment and Agent Note.

Widiastuti, N. 2019. Convolution neural network for text mining and natural language processing. In: Proceedings of *IOP Conference Series: Materials Science and Engineering*. IOP Publishing, 052010.

Wilson, R. E., Gosling, S. D. and Graham, L. T. 2021. A review of Facebook research in the social sciences. *Perspectives on psychological science*, 7 (3): 203-220.

Wojcik, S. and Hughes, A. 2019. Sizing up Twitter users. *PEW research center*, 24

Wong, K.-F., Wu, M. and Li, W. 2008. Extractive summarization using supervised and semi-supervised learning. In: Proceedings of *Proceedings of the 22nd international conference on computational linguistics (Coling 2008)*. 985-992.

Xian, Y., Lampert, C. H., Schiele, B. and Akata, Z. 2018. Zero-shot learning—a comprehensive evaluation of the good, the bad and the ugly. *IEEE transactions on pattern analysis and machine intelligence*, 41 (9): 2251-2265.

Yadav, A. and Vishwakarma, D. K. 2020. Sentiment analysis using deep learning architectures: a review. *Artificial Intelligence Review*, 53 (6): 4335-4385.

Yilmaz, C., Varnali, K. and Kasnakoglu, B. T. 2016. How do firms benefit from customer complaints? *Journal of Business Research*, 69 (2): 944-955.

Yilmaz, C., Varnali, K. and Kasnakoglu, B. T. 2019. How do firms benefit from customer complaints? *Journal of Business Research*, 69 (2): 944-955.

Yue, L., Chen, W., Li, X., Zuo, W. and Yin, M. 2019. A survey of sentiment analysis in social media. *Knowledge and Information Systems*, 60 (2): 617-663.

Zaid, S. and PATWAYATI, P. 2021. Impact of customer experience and customer engagement on satisfaction and loyalty: A case study in Indonesia. *The Journal of Asian Finance, Economics and Business*, 8 (4): 983-992.

Zaver, F., Battaglioli, N., Rose, C. C., Montrief, T., Alvarez, A. a. and Lin, M. 2021. Professional development during a pandemic: a live virtual conference for emergency medicine chief residents. *Canadian Journal of Emergency Medicine*, 23 (5): 696-699.

Zayani, M. 2021. Digital journalism, social media platforms, and audience engagement: The case of AJ+. *Digital Journalism*, 9 (1): 24-41.

Zhang, C., Mousavi, A. A., Masri, S. F., Gholipour, G., Yan, K. and Li, X. 2022a. Vibration feature extraction using signal processing techniques for structural health monitoring: A review. *Mechanical Systems and Signal Processing*, 177: 109175.

Zhang, J., Hsieh, C.-Y., Yu, Y., Zhang, C. and Ratner, A. 2022b. A survey on programmatic weak supervision. *arXiv preprint arXiv:2202.05433*,

Zucco, C., Calabrese, B., Agapito, G., Guzzi, P. H. and Cannataro, M. 2020. Sentiment analysis for mining texts and social networks data: Methods and tools. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 10 (1): e1333.

Appendix

Data Extraction

```
import re
import pandas as pd
import tweepy
Retailer =input (f'Name of Retailer:')
Area=input(f'Write your Area of complaint for tweet:')
print('Enter Start Date.')
day=int(input(f'Enter the day:'))
month=int(input(f'Enter the month:'))
year=input(f'Enter the year:')
proper_date=f'(Jindal and Aron)/{month}/{year}'
# date.append(proper_date)
print('Enter End Date.')
day1=input(f'Enter the day:')
month1=input(f'Enter the month:')
year1=input(f'Enter the year:')
proper_date1=f'{day}/{month}/{year}'
# date.append(proper_date)
date_List=['Retailer',
'Area'f'{day}/{month}/{year}',f'{day+1}/{month+1}/{year}',f'{day+2}/{month}/{year}',f'{day+3}/{mo
nth}/{year}' ,f'{day1}/{month1}/{year1}']
result = re.search(date_List)
def printtweetdata(n, date_List):
    print()
    print(f"Tweet {n}:")
    print(f"Username:{ith_tweet[0]}")
    print(f"Location:{ith_tweet[2]}")
    print(f"Timestamp Count:{ith_tweet[3]}")
    print(f"Hashtags Count:{ith_tweet[4]}")
    print(f"Mentions
Tweets:{ith_tweet[5]}")
```

Equation 4 Data Extraction Source Code

Pre-processing

```
Import nltk
Tokenisation = TwitterData(map(str.split, test_list))
from nltk.stem import PorterStemmer
from nltk.tokenize import word_tokenize
 ps = PorterStemmer()
TwitterData = open(result)
print(f.read())
From nltk.stem import WordNetLemmatizer
Lemmatizer = WordNetLemmatizer()
Print (lemmatizer.lemmatize ( pos = 'n'))
nltk.pos_tag(Lemmatizer)
NER_tree = ne_chunk(pos_tag(word_tokenize(pos_ner_text)))
print(NER_tree)
```

Equation 5 Pre-processing source code

Open-source sentiment analysis tool

```
Input: Sentences, SentiWordNet,WordNet, NegationWords
Output: Labelled Dataset
for each sentence S:
taggedSentence = POS(S)
for eachWordCandidate (verb, adverb, and adjective) in taggedSentence
LookupSentiWordNet (WordCandidate)
ifWordCandidate not in SentiWordNet
LookupWordNet (WordCandidate)
else ifWordCandidate > 0
polarity (WordCandidate)   positive
else ifWordCandidate < 0
polarity (WordCandidate)   negative
else if
polarity (WordCandidate)   neutral
else (there is NegationWords near WordCandidate)
polarity (WordCandidate)   opposite (polarity (WordCandidate))
PolarityScore += LookupSentiWordNet (WordCandidate)
TotalWordCandidateCount++
AveragePolarity = PolarityScore/ TotalWordCandidateCount
```

Equation 6 Sentiment analysis source code

Performance Metrics for Model Evaluation

```
Import Accuracy
Import Reliability
Import Specificity
Import F1
Accuracy=cross_val_score(rfc,x_best,y,cv=10,scoring='accuracy')
A_float = np.mean(accuracy)
Formatted_float = "{: .2f}".format(a_float)
Print('Accuracy', formatted_float, accuracy))

Reliability=cross_val_score(rfc,x_best,y,cv=10,scoring= 'reliability')
A_float = np.mean(reliability)
Formatted_float = "{: .2f}".format(a_float)
Print('Reliability', formatted_float, reliability))

Specificity=cross_val_score(rfc,x_best,y,cv=10,scoring= ' specificity')
A_float = np.mean(specificity)
Formatted_float = "{: .2f}".format(a_float)
Print('Specificity', formatted_float, specificity))

F1=cross_val_score(rfc,x_best,y,cv=10,scoring= 'f1')
A_float = np.mean(f1)
Formatted_float = "{: .2f}".format(a_float)
Print('F1', formatted_float, f1))
```

Equation 7 Performance Metrics source code

Visualisation

```
def Graph_Plot():
    # x axis values
    x = [1, 2, 3, 4, 5]
    # corresponding y axis values
    y = [0.00,  0.25, 0.30, 0.35, 0.40, 0.45, 0.50, 0.55, 0.60, 0.65, 0.70, 0.75, 0.80, 0.85, 0.90, 0.95, 1.00, 1.05]

    x_axis_labels = ['Accuracy', 'Reliability', 'Specificity', 'F -Measure']
    # Deep Learning
    x1 = [0, 1, 2, 3, 4, 5]
    y1 = [0, 0.21, 0.92, 0.52, 0.76]
    # Unsupervised
    x2 = [0, 1, 2, 3, 4, 5]
    y2 = [0, 0.54, 0.31, 0.84, 0.53]
    # Supervised
    x3 = [0, 1, 2, 3, 4, 5]
    y3 = {0, 0.64, 0.82, 0.56, 0.45]
    # plotting the points
    plt.plot(x1, y1, label="Deep Learning", color='green', linestyle='dashed', linewidth=3, marker='o',
             markerfacecolor='blue', markersize=7)
    plt.plot(x2, y2, label="Unsupervised", color='blue', linestyle='dashed', linewidth=3, marker='o',
             markerfacecolor='blue', markersize=7)
    plt.plot(x3, y3, label="Supervised", color='red', linestyle='dashed', linewidth=3, marker='o',
             markerfacecolor='blue', markersize=7)

    # setting x and y axis range
    new_ticks = np.linspace(0, 110, 12)
    plt.yticks(new_ticks)
    # plt.ylim(0,100)
    # plt.xlim(1,6)
    plt.xlim(0, 6)
    plt.xticks([0, 1, 2, 3, 4, 5, 6],
               ['$0$', '$Accuracy$', 'Reliability\ n Specifity','F -Measure'])

    plt.xlabel('Statistical Measure')
    plt.graph_plot('Model Evaluation Results')

    plt.ylabel('Evaluation Result')
```

Equation 8 Visualisation source code