SENTIMENT ANALYSIS USING UNSUPERVISED LEARNING FOR LOCAL GOVERNMENT ELECTIONS IN SOUTH AFRICA

Penelope Matloga

u2282646

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Denkleiers • Leading Minds • Dikgopolo tša Dihlalefi

Faculty of Engineering, Built Environment & IT,
Department of Computer Science, University of Pretoria,
Pretoria

A mini-Dissertation submitted to the Faculty of Science in fulfilment of the requirements for the Master degree in Big Data Science.

27 November 2023

Supervised by: Prof. Vukosi Marivate

Co-Supervised by: Kayode Olaleye

Contents

Declaration

DECLARATION OF ORIGINALITY

UNIVERSITY OF PRETORIA

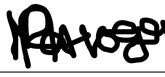
The University of Pretoria places great emphasis upon integrity and ethical conduct in the preparation of all written work submitted for academic evaluation. While academic staff teach you about referencing techniques and how to avoid plagiarism, you too have a responsibility in this regard. If you are at any stage uncertain as to what is required, you should speak to your lecturer before any written work is submitted. You are guilty of plagiarism if you copy something from another author's work (e.g. a book, an article or a website) without acknowledging the source and pass it off as your own. In effect you are stealing something that belongs to someone else. This is not only the case when you copy work word-for-word (verbatim), but also when you submit someone else's work in a slightly altered form (paraphrase) or use a line of argument without acknowledging it. You are not allowed to use work previously produced by another student. You are also not allowed to let anybody copy your work with the intention of passing if off as his/her work. Students who commit plagiarism will not be given any credit for plagiarised work. The matter may also be referred to the Disciplinary Committee (Students) for a ruling. Plagiarism is regarded as a serious contravention of the University's rules and can lead to expulsion from the University. The declaration which follows must accompany all written work submitted while you are a student of the University of Pretoria. No written work will be accepted unless the declaration has been completed and attached.

Full names of student: Mokgadi Penelope Matloga

Student number: u22826476

Declaration

1. I understand what plagiarism is and am aware of the University's policy in this regard.
2. I declare that this mini-dissertation report is my own original work. Where other people's work has been used (either from a printed source, Internet or any other source), this has been properly acknowledged and referenced in accordance with departmental requirements.
3. I have not used work previously produced by another student or any other person to hand in as my own.
4. I have not allowed, and will not allow, anyone to copy my work with the intention of passing it off as his or her own work.

SIGNATURE:  DATE: 27/11/2023

Abstract

Understanding public sentiment is vital for political parties in order for them to be able to structure their election campaigns around voter expectations. The study focuses on unsupervised learning to assess the variation of polarity sentiment in tweets during the 2021 South African local government election campaign. The study uses a pre-trained twitter-roberta-base-sentiment-latest model from Hugging Face and unsupervised lexicon based pre-trained approaches, namely: VADER and TextBlob to determine the polarity sentiment in order to gain insight that could be applied towards informing political campaigns and to see if there are any distinct sentiment patterns or shifts during different phases of the 2021 local government elections campaigns. Furthermore, the study applies the use of suspicious patterns and K-Means methods to classify the users as either bots and human using to be able to identify the user behind the keyboard. The study also make use of OpenAI GPT model to label the dataset for fine-tuning and addresses the issue of class imbalance. VADER and TextBlob results show a significant difference from that of the twitter-roberta-base-sentiment-latest models when comparing the statistical distribution based on the sentiment results and the user classification results. Based on the results, there is a significant variation across all sentiment classes and they vary over time. Furthermore, the results revealed TRBSL and TRBSL** outperforms VADER and TextBlob based on the scores for weighted accuracy and F1-scores. It was discovered that most of the tweets were generated by humans, with only few being identified as bot-generated and having a negative sentiments.

Keywords: Sentiment analysis, Unsupervised, OpenAI, Fine-tuning, User classification, Suspicious patterns.

Acknowledgement

I extend my heartfelt gratitude to the following individuals, whose invaluable assistance bolstered my successful completion of this mini-dissertation and played a pivotal role in the success of my master's degree.

First and foremost, I express my deepest appreciation to my supervisor, Prof. V Marivate, and co-supervisor, Kayode Olaleye, for their unwavering support, guidance, and insightful feedback throughout this project.

A sincere thank you goes out to my classmates and fellow members of Data Science For Social Impact for their fantastic collaborations, the sharing of brilliant ideas, and their encouraging spirits.

My eternal thanks are reserved for my family, particularly my mother and partner, for their unwavering belief in me, continuous support, and the patience they exhibited, allowing me the time needed to complete my studies. Their belief in me has been the driving force behind my perseverance throughout this journey. I also extend a special thanks to N Mabitsela for the emotional support and encouragement provided.

Above all, I acknowledge and express gratitude to God for His guidance, providing me with the knowledge, understanding, and wisdom required to overcome every obstacle encountered along the way.

In loving memory, I dedicate this mini-dissertation to my late father, confident that he would have been proud of his princess.

List of Figures

List of Tables

Chapter 1

Introduction

Over the years, there has been a significant change in South African politics as more people have developed an interest in it. Additionally, more people are participating in political discourse on social media platforms and freely expressing their opinions on the administration of government services and South African politics. Social media is an application or a web application that individuals use to share or create content and engage with each other. It has become an ideal tool for expressing the thoughts and feelings of users [1]. Over the past few years, there has been uncontrollable political unrest in South Africa [2]. To understand public opinion and develop effective strategies, political decision-makers are finding it useful to examine public sentiment [3]. Since most people share their opinions on politics as well as other economic and social challenges on social media, the data is available for analysis but most of it is not labelled. Since there is a lot of unstructured text data available on social media, the unsupervised learning technique is very useful. Labeling data is time-consuming and expensive, but unsupervised learning eliminates the need for labelling by taking into account the inherent structure of the data. In addition, the technique helps in discovering patterns and correlations in data that can be used to increase model accuracy [4].

Understanding public sentiment is essential for political parties so that they can structure their election campaigns around voter expectations. During local elections or any other type of government elections, political parties find creative ways to influence voters' thinking, making use of bots or influencers on social platforms. Nowadays, most political parties rely on social media for campaigning and persuading people to support them in elections. It is not surprising that some people develop a strong loyalty to one political party despite the difficulties they experience with the provision of basic services and other economic problems. According to Singhal, Agrawal and Mittal [5] and, Elbagir and Yang [6] using social media data for the analysis of politics is gaining more attention from many researchers. This is done to understand people's views and specific political trends during election time. To ensure that the analysis is not skewed, it is crucial to be able to understand user classification in terms of a bot or a human. Bots are social media profiles that interact with other users. Political bots have become significant actors during elections in a number of nations due to astroturfing and other techniques [7] as a means of influencing the results of the elections [8].

This study uses unsupervised sentiment analysis (SA) to detect the polarity of the tweets and classification of the users, following the work of Ledwaba and Marivate [2], which focused on semi-supervised learning techniques for predicting political behaviour in South African local elections. The authors classified a large volume of data from Twitter (now called "X") into positive and negative sentiment tweets using a semi-supervised method and a graph-based method. SA is a subset of Natural Language Processing (NLP) approaches that classify texts or sentences as negative, neutral, or positive, and is also divided into supervised and unsupervised learning. This study uses pre-trained Twitter-roberta-base-sentiment-latest and Lexicon-based classifications to determine sentiment. Twitter-roberta-base-sentiment-latest (TRBSL) model is a pre-trained Robustly Optimised BERT Pretraining Approach (RoBERTa) [9] model which has been built for sentiment analysis tasks with the TweetEval benchmark on English text and is also part of the Hugging Face Transformers library. Approximately 124 million tweets were extracted

between the beginning of January 2018 and end of December 2021 which made up the training dataset. Although RoBERTa is a supervised model, its pre-training was done using unsupervised methods. The model has achieved the highest level of accuracy on numerous NLP tasks, including sentiment analysis.

In the effort to leverage unsupervised techniques to perform SA using the Lexicon-based classification approaches, the study employs the Valence Aware Dictionary for sEntiment Reasoner (VADER) and TextBlob methodologies. Both VADER and TextBlob are unsupervised models for classifying text into positive, negative, or neutral attitudes. Both models can learn how to accomplish this without labelled training data. According to Pinto and Murari [3], VADER is the most effective NLP tool. The model is used to analyse sentiment and can tell how diverse the data is by how strong the current emotional power is based on the available Lexicon data dictionary [10]. VADER is very useful for sentiment analysis when used on data from social media platforms, and it produces good results. TextBlob is one of the straightforward Python library APIs for carrying out specific tasks involving natural language processing [11]. TextBlob offers the subjectivity and polarisation of the line of text, furthermore, it supports complex analyses based on textual information.

The study deals with stopwords using the Natural Language Toolkit (NLTK) default stopword list and the Ranks NL custom list of English stopwords. The stopwords are a list of popular words that do not provide anything useful to most of the text analysis techniques [12]. It is not necessary to manually define the stopword process while using NLTK. The class imbalance is also addressed, as are the contractions. The study also looks at labelling a sample dataset with the OpenAI model generative pre-trained transformers 3.5 (GPT-3.5), which uses few-short learning (FSL) where the model is trained to handle new sentiment classes based on a few labelled examples for each sentiment class in harnessing previous information knowledge gained from other tasks to easily adjust to new classes with fewer labelled data.

1.1 Problem Statement

Among the issues that the South African government is experiencing are service delivery and unemployment. Numerous protests have been held in response to issues with unemployment, corruption, and service delivery, and these protests have become increasingly popular on social media. Elections are being used by South Africa as a tool to make sure that the right candidates are chosen to meet the country's citizens' social and economic needs [2]. Political parties are now using social media platforms as tools to win public support and to make the most of the data they can gather there. With the growth of social media, there are concerns about bots being used to sway the electorate, particularly during local election campaigns. As a result, being able to identify the users behind the keyboard and their behaviours is critical. The study focuses on mainly four political parties (ANC, EFF, DA and ActionSA) in South Africa which participated in the 2021 local elections to determine the sentiment of the users, the variation of the different sentiment overtime and the classification of the users as bots or human.

The following research questions with their sub-questions are addressed in the study:

1. How do the polarity sentiments vary across the four political parties during the South African local elections campaign period between September and October 2021?
    - How can this insight be applied towards informing political campaigns?
2. How does the sentiment expressed by Twitter users during the 2021 local government elections campaign evolve over time?
    - Are there distinct sentiment patterns or shifts during different phases of the 2021 local government election campaigns?
3. What are the different classifications of Twitter users across the four political parties?
    - what insights can be gained from these user classifications?

1.2 Significance And Contribution Of The Study

The main aim of the study is to explore the unsupervised Twitter dataset using pre-trained models to answer the main questions and the sub-questions outlined in Problem Statement.

1.2.1 Significance Of The Study

The study addresses several challenges related to the use of unsupervised dataset. Firstly, the study overcomes the issue of having access to labelled data for SA. Data that is labelled takes time to collect and it can be expensive. By using unsupervised learning pre-trained models for sentiment analysis, the study offers a solution that does not require previously labelled dataset, enabling the analysis to be conducted on large dataset without requiring manual annotation. Secondly, the study shows the potential of the unsupervised models to discover sentiment patterns which are hidden in the dataset that are not clearly visible when using already labelled or manually labelled dataset. Thirdly, understanding how the public feels is essential for acquiring insights related to how politics works. The study uses SA to determine the polarity of users based on the local government elections in South Africa using Twitter dataset. Lastly, the results of this study will help political parties as well as potential candidates in decision-making. Understanding the feelings of people who vote will assist political parties to effectively prepare their campaigns, policies and procedures, and how to effectively communicate.

1.2.2 Contributions Of The Study

Using unsupervised learning techniques, the study first presents a unique unsupervised SA technique that identifies sentiment trends in the dataset. This technique adds to the growing studies in the field of unsupervised sentiment analysis. Secondly, the study presents the use of FSL to label unsupervised datasets using GPT-3.5. This contributes to the growing body of studies regarding the application of GPT models in the field of machine learning (ML). Lastly, the study contributes to the expanding field of studies on SA with an emphasis on geographic regions, i.e., South Africa and helps to make unsupervised sentiment analysis far simpler to understand for end users. The inclusion and focus of only South Africa provide a cultural and geographical aspect to the study. Understanding these nuances is important to insightful analysis, as cultural and geographical factors contributes to different sentiments.

The study is crucial in overcoming issues with data labelling, detecting sentiment trends, generating insights, and making decisions. Its methodological contributions to sentiment analysis, together with its cultural and regional focus and insights into the application of GPT models, are what make the study significant.

This study comprises of the following structure: Chapter 1 discusses the relevance of the study, the problem statement including the research questions and the contributions.The literature review, which is included in Chapter 2, includes summaries of various studies arranged according to different themes, as well as the the main findings in the studies and their limitations. Chapter 3 of the study describes the methodology in detail, including how the dataset was preprocessed, the models and procedures used, how the dataset was fine-tuned, the evaluation metrics, the ethical considerations surrounding the use of social media data, and finally the tools used in this study. The analysis and findings of this study based on the models selected are presented in Chapter 4. Chapter 5 presents the discussion of the study, which includes comparisons with other studies and drawbacks encountered in this study. Lastly, Chapter 6 presents the conclusions of the study together with the implications, future work and recommendations.

Chapter 2

Literature Review

An overview of earlier research on sentiment analysis (SA) and user classification (UC) are conducted in this chapter. Firstly, in Sections 2.1, 2.2, and 2.3, studies on SA are reviewed, along with the various analysis techniques and the key findings. Secondly, in Section 2.4, the study reviews the various methods applied in previous studies to distinguish actual humans from bots using the collected Twitter data and the key findings. Thirdly, Section 2.5 presents the key gaps. Lastly, the summary of the main takeaways and the gaps, together with the limitations addressed in this study are addressed in Section 2.6.

Twitter and other social media platforms have evolved into significant information sources for various application methods to analyse data which include sentiment analysis, opinion mining, and social media monitoring. However, given their briefness, informal language, use of emojis, and hashtags, Twitter texts pose obstacles for NLP. Several lexicon-based methods have been recommended to address these issues, including the use of pre-trained models like VADER and TextBlob as well as the Twitter-roberta-base-sentiment-latest (TRBSL) model from the Hugging Face Transformers library. The literature review examines the use of VADER, TextBlob and TRBSL models for sentiment analysis of the Twitter dataset regarding the 2021 South African local elections.

2.1 Sentiment Analysis Using Social Media Dataset

The area of sentiment analysis utilising social media dataset is dynamic and ever-changing, necessitating ongoing adaptation to shifts in language usage, online trends, and platform-specific capabilities.

In their study, Abiola et al. [13] employed sentiment analysis of Twitter data to assess the perception of people regarding the 2019 COVID (COVID-19) outbreak in Nigeria using COVID-19 hashtags. A pickle format was created out of the 1,048,575 tweets that were collected in comma-separated values (CSV) format for user convenience. To conduct the study, the authors used sentiment analysis techniques, TextBlob and VADER. Their results suggest that the perceptions of users with TextBlob are largely neutral, but VADER's results reveal a higher proportion of positive sentiments. The study conducted by Abiola et al. [13] provides a comprehensive examination of Nigerians' perceptions on the pandemic. In addition, the authors stated that the findings of their study could contribute to a better understanding of how to address pandemic-related obstacles.

Using information from Twitter, Bengesi et al. [14] analysed the public's perception of the monkeypox outbreak. The authors gathered more than 500,000 tweets in multiple languages about monkeypox and used VADER and TextBlob to analyse their sentiments. By using stemming and lemmatisation methods for vocabulary normalisation and vectorisation based on Term Frequency - Inverse Document Frequency (TF-IDF) and CountVectorizer (CV) methodologies, the authors further constructed and assessed 56 classifiers. Bengesi et al. [14] employed the learning methods Multilayer Perceptron, Naïve Bayes (NB), K-Nearest Neighbour (KNN), Random Forest (RF), Extreme Gradient Boosting (XGBoost), Support Vector Machine (SVM) and Logistic Regression (LR), and assessed the models' f1-score, accuracy, Precision, and Recall. According to their study, TextBlob annotation, lemmatisation, CV, and SVM were

the features of the model that produced higher accuracy. The authors argued that decision-makers could find the findings of their study helpful in developing health policies and disease-mitigation plans.

In their study, Darad and Krishnan [1] undertook an empirical analysis to discern the sentiment exhibited by individuals on prominent social media platforms, notably Twitter, during the zenith of the 2019 Corona virus pandemic in April 2021. Employing advanced methodologies, the authors conducted a comprehensive sentiment analysis by employing an advanced deep learning model known as "Bidirectional Encoder Representations from Transformers" (BERT), in conjunction with a suite of traditional machine learning models for textual analysis. Their study included the following machine learning models: Stochastic Gradient Descent, LR, XGBoost, FR, SVM, and NB. Part of their analysis involved a thorough comparison of these various models' performances. Strict evaluations of these models were carried out to determine their accuracy in several sentiment categories. Their findings demonstrated that the aforementioned models produced classification accuracies of 78.6%, 77.7%, 75.5%, 74.7%, 74.5%, and 66.4%, respectively. It is noteworthy that the BERT model excelled, exhibiting the highest accuracy rate at 84.2%. Sentiment classifier model obtained an exceptional degree of accuracy, much more than the 75% criterion, a notable success within the realm of text mining algorithms [1]. The results of their study provided statistical evidence that, over the defined period, the dominating sentiment on social media platforms was primarily oriented towards positive and neutral sentences. Their main findings suggested that, during the peak of the COVID-19 epidemic in April 2021, people tended to be more negative and depressed globally. However, because of the inherent complexity and nuance of social media discourse, it is vital to stress the inherent challenges in achieving perfect sentiment prediction accuracy [1].

Alabrah et al. [15] focused on overcoming vaccine resistance in Gulf nations by examining attitudes toward 2019 Coronavirus vaccination using approaches in machine learning. The authors collected Twitter data, filtered and tokenised it, and performed sentiment analysis using three distinct methods, i.e. Ratio, TextBlob, and VADER. Using the proposed Long short-term memory (LSTM) approach, the authors further categorised the sentiment scores attained as positive and negative. Alabrah et al. [15] extracted detailed features from the suggested LSTM and feed them to four different machine learning classifiers to maximise the classification accuracy. The findings of their study demonstrated that by employing Ensemble Boost and Fine-KNN, the sentiment scores of VADER offered the best classification performance of 94.01 percent. The authors concluded that the suggested method was convenient and reliable for categorising and identifying sentiments in the Twitter debate regarding COVID-19 immunisation in Gulf countries.

Using data from Twitter, Illia et al. [16] conducted a study to evaluate how Indonesians feel about the 'PeduliLindungi' application during the Covid-19 diseases outbreak. The authors used primary data about the data leaks regarding the application that occurred between 31 August and 7 September 2021 which caused the media storm. Illia et al. [16] used TextBlob and VADER libraries to analyse sentiment. Their results suggested that VADER's lexicon approach, which is focused on social media, was more efficient in conducting semantic analysis. In their study, Illia et al. [16] provided insight into the general perception of people regarding the 'PeduliLindungi' application. The authors argued that based on those opinions, the results could assist in improving the application. According to the authors, the model was not validated, normalising the data was handled manually and the data was translated into English manually also. They authors suggested that stopwords and normalisation, both of which need preprocessing, should be carried out automatically. Additionally, they recommended evaluating the sentiment model to ascertain its accuracy.

In summary, the evaluations of the literature on SA utilising social media datasets demonstrated the range of techniques used to measure public opinion during important events like the COVID-19 epidemic and applications. To evaluate the feelings stated on social platforms such as Twitter, the authors used sentiment analysis techniques such as TextBlob, VADER, SVM, LSTM models, advanced models like BERT, and other machine learning classifiers. Sentiment analysis was used by Abiola et al. [13] and

Alabrah et al. [15] to investigate public feelings on the 2019 corona virus pandemic in Nigeria and the vaccine for the virus in the Gulf countries respectively. Comparably, the study conducted by Bengesi et al. [14] used roughly half a million tweets to analyse public opinion on the outbreak of monkeypox. Furthermore, during April 2021 when the COVID-19 pandemic was at its worst, Darad and Krishnan [1] examined sentiment analysis on social media platforms, and Illia et al. [16] examined public opinions of the 'PeduliLindungi' application. Both studies highlighted the significance of SA in application development based on input from users. The results of their investigation provided complex insights into public attitudes, ranging from TextBlob neutral perceptions from users to the positive perceptions found by VADER. Their research also highlighted the usefulness of sentiment analysis in managing vaccination resistance, evaluating public sentiment regarding applications, and improving health policies. While some research, such as those by Darad and Krishnan [1], provided insight into the intensity of negative feelings around a certain periods, the examination by Mustaqim [10] of hidden patterns provided a more comprehensive view of sentiment polarity. The collective summary of these studies of the literature emphasises the significance of SA in understanding how people feel on social media, however, methods-related problems and the need for ongoing improvement are noted for future studies.

2.2 Sentiment Analysis In Political Context

Sentiment analysis is gaining significant attention in the contemporary world. However, it is encountering substantial challenges because of the rise of online platforms where interactions occur in unstructured forms [17]. In the context of politics, SA is a useful technique for measuring and addressing public opinion. Recently, there has been a substantial increase in the use of social media sentiment analysis in election campaigns around the world. Social media is being used by a lot of political parties and politicians to interact with existing supporters and win over new ones. It may provide insights that help influence political strategies, options about policies for the public, and activities requiring involvement by the public. Hence, establishing a foundation of loyal supporters and taking advantage of data to determine user perception and other potential voters serves the objectives. But it is imperative to confront the difficulties and moral issues that come with analysing feelings in the dynamic and complicated field of politics.

Oyewola et al. [18] employed sentiment analysis of tweets to conduct research on the feelings of users towards the leading three candidates in the Nigerian presidential election held in 2023. Employing three models: First, the authors preprocessed the dataset to remove noise and unnecessary information and then employed three models to classify tweets as negative, neutral, or positive, i.e., peephole LTSM (PLSTM), LSTM, and two-stage residual LTSM (TSRLSTM). The findings of their study demonstrated that the TSRLSTM model exhibited remarkable performance in tweet classification, accurately identifying the sentiments of each candidate. The results of their study are beneficial to researchers and decision-makers since they offered insightful information on public opinion about candidates and election strategies. Oyewola et al. [18] also noted that more model improvement was necessary, and that larger and more diverse datasets are crucial for a thorough understanding of public opinion, nevertheless they acknowledged the limitations of the model that resulted from using a relatively small set of data.

Shevtsov et al. [19] studied the discussion on social media during the November 2020 United States presidential election, specifically concentrating on Twitter and YouTube. The authors collected tweets pertaining to trending election hashtags and original YouTube videos, after which they undertook preprocessing actions, such eliminating punctuation marks, emoticons, and hyperlinks. The authors analysed user and tweet traffic, identified entities, and looked for relationships across various aspects of YouTube videos. Using daily sentiment scores for every user, the authors investigated how actual events affected conversations on social media. They discovered that positive sentiment was stronger for Donald Trump than for Joe Biden using the VADER algorithm. Their study also explored how conversations on social media are driven by real life events, utilizing daily sentiment scores for each user. Additionally, based on July to September 2020 dataset, Shevtsov et al. [19] looked at the graph for retweet at distinct time

periods, i.e., six time periods. They identified "Trump" and "Biden" as the primary entities and showed how the linked density component increased with time. The authors intended to do further study on the detection of sarcasm in the future to generate an appropriate ground truth dataset for training following the election season, by employing crowd-sourcing methodologies.

To better understand the political climate in South Africa, Ledwaba and Marivate [2] analysed the opinions expressed on Twitter during the local government elections. The authors used a semi-supervised approach with a graph-based method to classify the huge amount of publicly available Twitter data for classifying tweets into either positive or negative sentiment. To find hidden issues of concern related to each political party, Ledwaba and Marivate [2] further analysed the tweets revealing negative sentiment using the extraction method latent topic. The authors discovered that most Twitter users in South Africa are opposed to all four parties. Users expressed concerns about misconduct, incompetence, and load shedding, with the current governing party(ANC), receiving the most negative feedback. The limitation of the study, according to Ledwaba and Marivate [2], is that the model was not taught to recognise tweets that might not have either a negative or positive sentiment.

Endsuy [20] compared exploratory data analysis (EDA) and sentiment analysis using Twitter data regarding the election in the United State in their study. The author discovered that neutral sentiments made up most sentiments as compared to negative and positive sentiments. Endsuy [20] used EDA and VADER, which they asserted to be relatively accurate and produced good visualisations. The author further suggested that more work should be done on those methods to make them much more effective. Endsuy [20] concluded that their results for the sentiment analysis study should be enough to inspire readers to conduct related studies to assist political parties in trying to analyse election outcomes, and possibly even assist politicians to learn about the perceptions of their supporters.

Mustaqim [10] performed an analysis process which included insights into the data that were hidden for extraction, visualisations, and sentiment classification of people's perception of politics and religion using 5433 datasets collected on 12 November 2019 from Twitter. To ascertain the sentiment in the datasets, Mustaqim [10] first pre-processed the data, then used K-Means data clustering and the VADER model for SA. The results revealed hidden meanings in the form of fifty distinct words, each of which was further classified into five clusters of ten words, and each of which underwent sentiment analysis. According to the author, texts like "hate" and "radical" frequently appeared in the polarity of negative sentiment, whereas the words "like" and "god" frequently appeared in the polarity of positive sentiment in the results of the wordcloud visual analysis and the hashtag clustering. The findings suggested that there were different polarities of sentiment in public perception, including positive, negative, and neutral. Furthermore, Mustaqim [10] asserted that texts with the same occurrence frequency as the dominant word seemed to be more likely to generate a neutral sentiment score.

To conduct sentiment analysis of a multi-classification system for tweet analysis, Elbagir and Yang [6] used Twitter data related to the US election of 2016. They applied NLTK and classified the sentiment polarity using VADER. The authors asserted that VADER classified large volumes of data fast and easily. The findings of their study demonstrated that the sentiment analyser for VADER was the appropriate option for classifying sentiment using data from Twitter provided a good accuracy. The use of a small dataset to conduct the analysis and the use of a generic vocabulary to identify specific details on the data were two points that highlighted the limitations of their study. Elbagir and Yang [6] further asserted the lack of training in the data as the another drawback for their study. Future research, will require a substantial amount of data, along with a lexicon relevant to the dataset and a list of texts to train the data for improved outcomes [6].

In order to help political parties with campaigning, decision-making, and election outcomes prediction, Nandi and Agrawal [21] conducted a hybrid method study on politics. The authors combined the features of a SVM with the Vocabulary approach to overcome the limitations of each algorithm and take advantage of their strengths. Using this approach, they were able to achieve an accuracy of 93% for

Linear Support Vector Classifier (LinearSVC) with SVM and 91% for SVC with "Kernel equal to linear", which is considered acceptable for a sentiment analyzer. According to their findings, the SVC with Kernel = linear was outperformed by the LinearSVC. The authors claimed that there were difficult features to dealing with negation and intensifier phrases. Nandi and Agrawal [21] proposed a creation of a multilingual sentiment classification model, that would modify their system to accommodate additional languages and gather lexicon data that allows them to compare the features over several languages.

In summary, numerous studies that investigated sentiment analysis in political contexts provided insights into how people feel towards politicians and elections. While Ledwaba and Marivate [2] examined Twitter opinions during South Africa's local government elections and found a strong backlash to political parties, particularly the ANC, Oyewola et al. [18] concentrated on the 2023 Nigerian presidential election and used LSTM models for sentiment classification, revealing that the TSRLSTM model performed exceptionally well in accurately identifying sentiments towards candidates. In contrast, Shevtsov et al. [19], Endsuy [20], Elbagir and Yang [6] examined the presidential election in the United States and used the VADER model in identifying significant positive sentiment favouring specific presidential candidates, and also demonstrated a large proportion of neutral sentiments, and acknowledged limitations due the use of a small dataset, respectively. Furthermore, Mustaqim [10] found that phrases with the same frequency were more likely to provide a neutral sentiment score when analysing sentiment polarity on Twitter around politics and religion. A noteworthy insight is the identification of a hybrid strategy to political analysis, as demonstrated in the study of Nandi and Agrawal [21], which integrated SVM with a lexicon approach to reach high accuracy. By emphasising on the need for advanced methods to sentiment analysis, larger datasets, and enhancement of the models, these studies collectively provided insightful information on public views and opinions in various kinds of political contexts.

2.3 Methodological Approach To Sentiment Analysis

With a focus on the significance of carefully developed attributes, selecting a model, and interpreting every aspect of the data source, this literature review offers insights into numerous methods, techniques, and tools implemented in sentiment analysis.

In the study Ashir [17], a novel approach to sentiment analysis was proposed, combining rule-based and lexical procedures with unsupervised machine learning. The goal was to enhance sentiment analysis and make it more adaptable across various sources, even those lacking clear syntactic and grammatical structures. Their approach incorporated several techniques, such as the Rule-Based Method which included emoticon detection, word contraction expansion, and noise removal to handle sources with limited structure, and Lexicon-Based Preprocessing which included the use of lexical features. The experimental results from their study, employing different machine learning classifiers, demonstrated an improved performance and a high level of adaptability when using sources that are organised and those that are not organised. Notably, the findings indicated that carefully designed lexical features significantly enhanced the unsupervised learning process, surpassing the use of word embeddings alone as features. In summary, their study achieved its two primary objectives which was to improve the performance and increase the generalisation ability in sentiment analysis across diverse sources.

To analyse the polarity and subjectivity, Mujahid et al. [22] employed deep learning methods and machine algorithms on a dataset of 17,155 tweets from Twitter to assess how individuals felt about e-learning. The authors employed TextBlob, VADER, and SentiWordNet to analyse the polarity and subjectivity scores, and classified the sentiment using different types of machine learning models. Mujahid et al. [22] developed and assessed the models using the feature extraction methods TF-IDF and Bag of Words. F1-score, recall, accuracy, and precision were used by the authors as key metrics to measure the performance of the models. The authors also compared the effectiveness of TextBlob, VADER, and SentiWordNet. Their comparison of the performance of VADER with SentiWordNet showed that VADER performed significantly better than SentiWordNet, due to it being specifically created for anal-

yses of social media data. Their findings demonstrated that the LSTM performed barely slightly better than average, with accuracy scores of 0.94 for Textblob, 0.91 for VADER, and 0.85 for SentiWordNet. According to their findings, TextBlob was more effective in annotating data than VADER and SentiWordNet.

To fully understand consumers' perceptions on market research, Gujjar and Kumar [11] proposed a technique called TextBlob, a cost efficient method in computation. Python was used by the authors to access the TextBlob API. The proposed technique, according to the authors, should assist decision-making of establishing standards for goods and services. But, despite being found to be very useful for business intelligence and benchmarking of services and products in decision making, the suggested technique had drawbacks. The technique was found to be unreliable when analysing sentiment using emoticon-rich data, and performing poorly when dealing with biased and code-switched data, proving it to be incapable in analysing emotions. Gujjar and Kumar [11] proposed the use of either supervised or unsupervised learning for tackling the problems of sentimental analysis, especially for emojis in future works.

In their study, Al-Shab [23] assessed how well five widely used sentiment analysis lexicons (Sentiment Strength Detector, AFINN-111, VADER, Liu and Hu opinion lexicon, and SentiWordNet) performed on Twitter data. To classify polarities, Al-Shab [23] examined the F1-score and total accuracy of the five lexicons. According to their findings, VADER outperformed the other four mentioned lexicons in categorising positive and negative attitudes. Additionally, it worked well for categorising short phrases as positive, neutral, or negative. The author also found that, on two datasets, VADER performed best, while the other lexicons achieved outcomes that were the same. They further added that the Stanford dataset outperformed the Sandars dataset when using Liu and Hu's lexicon. The fact that the lexicons were used directly, without any changes or preprocessing, was noted by Al-Shab [23].

In their study Hutto and Gilbert [24] evaluated the performance of VADER on Twitter texts, against eleven well-known advanced benchmarks which included the SVM algorithm. A "gold standard" list of lexicons specifically tailored to sentiment in weblog contexts was developed and verified by the authors using a mix method technique. Hutto and Gilbert [24] expanded on these vocabulary characteristics by considering five basic rules that encapsulate grammar and syntax conventions for emphasising and expressing the intensity of the sentiment. The results of their study showed that VADER outperformed other tools on the Twitter dataset with an accuracy of about 96% and F1-score of 0.84%, including human raters. The authors suggested that the findings of their study were exceptional. Their study provided more evidence of the success that could be achieved in computer science when people are provided key roles in the development process.

Alqaryouti et al. [25] proposed a model for extracting crucial elements from reviews and categorising the correlating sentiments. To address several challenges in sentiment analysis, the authors used language processing methods, regulations, and vocabulary to generate results that are simplified. According to their findings, when underlying attributes were taken into account, the way to obtain accuracy significantly increased. Alqaryouti et al. [25] asserted that, when applied to the same dataset, the suggested method performed better than the SVM. The authors further stated that making use of the vocabulary and rules as the entry attributes to the SVM model improved the model to perform better than other SVM models in terms of accuracy. The authors proposed using an aspect-based hybrid sentiment analyser that integrates domain vocabulary items and rules to evaluate intelligent review applications.

Luong et al. [26] proposed methods to extract sentiment information from posts on social networking platforms. This can be viewed as either an extraction of features or a sequential classification problem. The authors argued that people are more willing to share their emotions, ideas, and intentions on social platforms nowadays. Luong et al. [26] further asserted that understanding consumers' online intentions is crucial in various business sectors. The authors utilized Bi-directional LSTM, Conditional Random Fields, and a complex statistical graphical model to develop machine learning models for sequence data. According to their findings, the proposed methods were able to effectively and accurately extract infor-

mation about intentions from texts online.

In summary, the examination of the literature examined several approaches taken in sentiment analysis, emphasising the models as well as techniques used in different studies. The solution proposed by Ashir [17] stood out for its combination of rule-based, lexical aspects and unsupervised ML methods, resulting in increased flexibility across multiple sources. Mujahid et al. [22], on the contrary, focused on approaches using deep learning for sentiment analysis on online learning tweets, demonstrating dominance of VADER over TextBlob and SentiWordNet. Furthermore, while recognising its shortcomings with data containing emoticon and biassed data, Gujjar and Kumar [11] suggested TextBlob for market research. For sentiment analysis on Twitter, VADER was demonstrated as being more effective by Al-Shab [23] evaluation, and Hutto and Gilbert [24] demonstrated this by using Twitter to highlight the importance of human contribution to the development of tools. Whilst the techniques by Luong et al. [26] concentrated on sentiment extraction from social media, the model by Alqaryouti et al. [25] outperformed approaches based on SVM-based approaches. Each method made a distinct contribution to the changing field of SA, demonstrating the significance of taking the particular context and analytic objectives into account.

2.4 Identification Of Bots And Spammy Tweets

Since people interact with each other on social media platforms like Twitter without really knowing who is behind the account, bot accounts can easily control the network by imitating real users' actions [27]. In order to better identify Twitter users as either bots or humans and extract insights from the 2021 local election data to assist political parties in making decisions during elections, these study firstly review previous studies on user classification. The most difficult component of understanding Twitter bots is figuring out how to interpret today's bots, which are more complex [28]. Due to the dynamic structure of the Twitter platform, traditional Twitter approaches cannot be implemented randomly or consecutively, therefore it typically takes months to filter out spam accounts [29]. Thus, it's critical to simplify and speed up the process of identifying the accounts used by bots. Bots can invalidate popular viewpoints or participate in a political discussion on social networking sites which can harm the evolution of public policy [28, 30]. These bots have been built to fulfil certain objectives and goals and are capable of carrying out complex interactions. More research is needed to determine the amount of content generated through bots that are being consumed by social media users, given the continued growth of social media bots and the spreading of disinformation [31].

In their study, Alarfaj et al. [28] aimed to identify automated Twitter users using specific content features and machine learning. Several sets of features were proposed for analysing tweet content, which is based on special characters, SA to classify Twitter users as bots or humans, part-of-speech, and messages. To normalize the data, the authors considered the minimum-maximum normalisation, and feature selection techniques were employed to determine the most important features in each feature set. The proposed methodology leverages deep learning algorithms and other classification-related techniques which included NB, rule-based classification, and RF. The application of these techniques enabled efficient and accurate classification of data, which was crucial for the success of various business and academic initiatives. In evaluating the accuracy of different techniques for detecting human and bot tweets based on separated features, NB, RF, and rule-based classification performed better than other methods. However, when all features were combined, deep learning algorithms proved to be the most accurate and precise choice. In fact, Alarfaj et al. [28] asserted that, when used with the suggested sets of features, deep learning performed better for accuracy and f-score than any advanced technique. Overall, it was observed that combined related content set of features are more accurately detected by deep learning than by any other method. The process of detecting Twitter bots could be significantly enhanced by the inclusion of image analysis. Images often contain valuable information that, when analysed, can augment the accuracy of bot-detection models. Thus, the authors suggested focusing on developing image analysis algorithms as a complementary method to improve the overall accuracy of bot detection for future works.

The study by Genfi [31] explored the availability of COVID-19-related bot-generated content. In the process of examining how social media contributes to the spread of false information and pandemic truth, the author brought attention to the difficulty in distinguishing the difference between actual users and bots on social media. The study employed a hybrid method to analyse 71,908 COVID-19-related tweets from January 22 to April 2020, during a global COVID-19 case count of less than 600. The study used the Weka machine learning Tool to identify bots and false information about the COVID-19 pandemic through sentiment features, user account attributes, and topic analysis. Genfi [31] in their study, investigated Twitter bot detection algorithms using 10-fold cross-validation on two datasets in test 1. The results showed that category 1 features, combined with the random forest algorithm, yielded the best prediction accuracy, outperforming category 2 features. The second test employed the Random Forest algorithm to classify two labelled datasets, attaining 94% prediction accuracy for Category 1 characteristics and 60% accuracy for Category 2 features. The author further used test 1 and 2 data to categorize 39,091 accounts linked to COVID-19, identifying 15% as bots and 85% as humans using a random forest algorithm and identified traits. The study revealed that human accounts spread 70% of false COVID-19 information, while bot accounts produced 30%, with retweets of bot-generated content accounting for nearly 30% of disinformation. In addition, compared to human accounts, bots posted content on a smaller range of topics and tend to elicit more negative sentiments towards COVID-19-related issue [31]. Consequently, it seems that topic distribution and sentiment analysis improved the capacity to discriminate between accounts that are bots and those that are people.

The ability of automated users to imitate actual users is made possible by the advancements in artificial intelligence and chatbot technology, which allows these bots to learn and adapt rapidly. A recent study by Alothali et al. [32] suggested that identifying the most effective features for detecting bots was an area that requires further research. In their study, the authors evaluated profile details using a novel method called a hybrid feature selection. Their study aimed to pinpoint the most valuable features for classification tasks. The proposed method leveraged four popular ML algorithms, namely RF, NB, SVM, and neural networks, to evaluate the selected features. Their approach was expected to yield improved results compared to traditional feature selection methods. In their study, Alothali et al. [32] discovered that cross-validation attribute evaluation exhibited superior performance compared to other feature selection techniques. Their findings highlighted the importance of utilizing cross-validation attribute evaluation when selecting features for a model. It provided a useful tool for improving the accuracy and performance of predictive models, making it an indispensable method in the field of data science and ML. Based on their findings, the highest scorefor the area under the curve (94.3%) was achieved by the RF classifier with six optimal features. The overall accuracy of the findings was 89%, with a precision of 83.8% and a recall of 83.3% for the bot classification. Alothali et al. [32] also discovered that utilising a total of four features, including status_count, favourites_count, average_tweets_per_day and verified, resulted in significant performance measurements for bot identification, with a precision of 84.1% and a recall of 81.2%. The authors looked at leveraging advanced deep learning techniques to identify social bots effectively and accurately in the near future.

In their study, Vasterbo [8] used tweet-level features to conduct a comparative analysis of various ML approaches for labelling tweets as either human or bot-generated. The primary focus of their study was to assess the ability of the models to generalize on data that was previously unseen. The author used AUC-ROC and Average Precision metrics to compare the performance of ML techniques, which include RF, AdaBoost, and Contextual LSTM model. In order to assess the models, Vasterbo [8] used two different tests with five datasets containing tweets from bots and one dataset with tweets from humans. In the initial test, the models were trained on the four bot datasets and then subsequently tested on the omitted dataset. For the last test, separate datasets were used independently to train and evaluate the models. A very small performance difference was found between the models according to the results of the initial test. While the Contextual LSTM model performed poorly in several dataset combinations, RF and AdaBoost demonstrated comparable patterns in the last test. According to the author, the models demonstrated minimal performance variation, making it difficult to identify the model which was performing better for the task as a result of the relatively small variation in the initial evaluation. In

addition, Vasterbo [8] took into account the amount of time needed for testing and training the models and discovered that RF seemed to be the most efficient option. For a further study into the world of advanced bots, the author emphasised the importance of gathering labelled datasets that showed recent tweets which are generated by bots. The author highlighted the need to determine whether the results of their study apply to present bots of today. Moreover, considering the possibility of individual account tweeting in a partially automated way, Vasterbo [8] suggested conducting additional studies on classification using the features of the tweet.

Abkenar et al. [33] conducted a complete systematic literature review, which featured 55 most relevant research papers issued during 2010 and 2020. Abkenar et al. [33] reviewed the tools, assessment parameters and methodologies, research methodology, and statistical analysis of the applicable approaches in each research study selected. The authors discovered that recall, precision, accuracy, and F1-score were the frequently adopted criteria in the evaluation process of identifying and filtering out spam on Twitter, with scores of 23%, 17%, 14%, and 18%, respectively. Based to the evaluation methods and tools available for Twitter spam detection, Abkenar et al. [33] also observed that Weka was the most frequently used evaluation tool across all the research papers they reviewed. Additionally, Abkenar et al. [33] provided a taxonomy that outlined Twitter spam detection methods in detail. According to the authors, Twitter spam detection methods were developed using analytical features, and those methods were classified into five categories: approaches for content, user, tweet, and network analyses, and mixed methods (15%, 9%, 9%, 11%, and 56%, respectively). Class imbalance in Twitter dataset studies resulted in lower classification performance due to biased ML techniques favouring the class which has no spam and the class of the minority with spam being ignored [33]. Despite the fact that several studies have indicated strategies to improve the identification of spam in unbalanced datasets, Abkenar et al. [33] emphasised that additional study was required and that it constituted a severe challenge.

According to Kudugunta and Ferrara [34], it can be difficult to identify bots on social media due to their ability to masquerade as real people. The author proposed a deep neural network built on contextual LSTM architecture to identify bots at the tweet level using content and textual information. They suggested a synthetic minority method to produce a large dataset for training deep networks on a small portion of labelled data. The system could distinguish between humans and bots with an increased accuracy rate from just a tweet and could also detect bots at the account level with almost perfect accuracy. The authors proved that their framework could achieve high accuracy significantly bigger than 96% in distinguishing humans from bots using a tweet. Using the same framework, the authors were able to detect bot activity at the account level with practically perfectly balanced accuracy rate of over 99%. Their approach used less training data and performed better than the prior advanced models, all while utilizing a minimal and understandable collection of traits. Kudugunta and Ferrara [34] planned to make available the system and develop an API to enable researchers to use it to detect bots at the tweet level. For the purpose of assessing bot participation in public discourse and comprehending its increasing complexity and capabilities, the authors advised applying their technique to the analysis of social media discussions in a different context.

Washha et al. [29] examined the issue of spamming from the standpoint of finding information and generated searchable data that functioned as a search query to identify fake accounts. The authors proposed a framework for an automated, unsupervised approach that predicted fake name behaviours by employing searchable details from a particular set of hashtag-rich tweets as basic metadata. The authors asserted that Twitter users and Twitter-based projects handling massive volumes of tweets could both take advantage of their method to search for spam accounts. Their experimental study demonstrated how successful it is to forecast spam activities to detect fraudulent accounts based on adjusted discounted cumulative, accuracy, and recall at various levels. Their results demonstrated that no variable outperformed the others. Asserting that their study was the first in such area, Washha et al. [29] aimed to extend the searchable information from tweets and enhance the assessment metrics.

Dickerson et al. [35] developed a list of system, textual, and application based on factors that could

potentially be used as key elements, assisting in the identification of a particular element that effectively differentiates actual humans from bot users. The authors presented a sizable collection of sentiment features, which include mixtures of network and sentiment variables. The SentiBot framework was used by the authors to address the issue of identifying users as human versus bot using relatively limited local data. With reference to the 2014 Indian election, the authors examined more than 7.7 million tweets with 550,000 users. They showed that making using sentiment features effectively increased the accuracy rate by 53% as well as the "area under the receiver operating curve" (ROC curve) from 65% to 73%. According to Dickerson et al. [35], this was the first time that bot detection used such sentiment-based features.

In summary, studies on spam detection and bot detection has provided a variety of insights into feasible approaches for identifying automated bots behaviour from human behaviour on social networking sites. Alarfaj et al. [28] presented a strong argument for the efficacy of deep learning techniques in correctly classifying accounts on Twitter, notably when integrated with different types of features that focus on specific content. Furthermore, the study conducted by Genfi [31] explored how bots contributed to the spreading of false information during the 2019 corona virus pandemic and highlighted the difficulties in differentiating between content created by bots and those created by human. The systematic literature review by Abkenar et al. [33], in contrast, offered a thorough overview, classifying spam detection techniques and emphasising standardised assessment criteria and techniques. Furthermore, it was proposed that the accuracy of bot-detection algorithms may be further improved by integrating image analysis. By focusing on cross-validation attribute assessment and using a hybrid feature selection strategy, Alothali et al. [32] achieved significant performance metrics for bot classification. By comparing machine learning techniques and highlighting the effectiveness of Random Forest, Vasterbo [8], concentrated on tweet-level attributes. Kudugunta and Ferrara [34], on the other hand, presented an effective deep neural network for identification bots on tweet-level, exhibiting highly accurate results and recommending the use of their method to evaluate discussions on social media across various settings. Furthermore, Washha et al. [29] demonstrated the effectiveness of their automated, unsupervised method to spam identification by predicting the behaviours of fraudulent accounts. The application of sentiment-based traits had originally been introduced by Dickerson et al. [35], who also showed how important these traits were for correctly detecting bots in the Indian election of 2014. The collective findings of these research highlighted the constantly evolving nature of detecting bots, the significance of appropriate features, and the continuous struggle of remaining relevant in the presence of constantly evolving bot behaviours. These numerous methods and outcomes add to the growing knowledge of the complicated landscape of spam and bot detection within the social network sphere.

2.5 Key Gaps In The Literature

Several gaps emerged from these literature review. First, several studies acknowledged that in order to increase model accuracy and get a deeper understanding of public opinion, larger and more varied datasets is necessary. This can be difficult since processing huge datasets takes time and additional processing resources. Several studies, such as Illia et al. [16], have drawn attention to the fact that their models lacked validation and that manual data normalisation and translation were performed. This raise concerns over how well the results generalise and whether they can be reliable, emphasising the importance of strong validation techniques and automated data preparation. It should be highlighted that dealing with unsupervised data may be complex, time consuming, and costly. This study also uses unsupervised dataset which is from Twitter platforms, hence addressing this particular gap is a challenge to this study. Second, limitations on neutral sentiment label identification and model training are emphasised, highlighting the difficulties in capturing subtle emotions. Ledwaba and Marivate [2] drew attention to the limitation of their study, which highlighted that they only used positive and negative sentiment labels, therefore, their model was not taught to recognise tweets that might not be conveying either sentiment. The limitations of utilising a small dataset and a generic vocabulary were also highlighted by Elbagir and Yang [6], who emphasised the need of having sufficient training data and lexicons appropriate for the scenario. Therefore, neutral sentiment labels and bigger, more accurate datasets should be included

to address this gap. This study focuses on the limitations associated with using just positive and negative sentiment labels, by using models that are able to identify neutral sentiments and further uses GPT model to label the data for training purpose into neutral, negative and positive sentiments.

Third, even though the research make use of a variety of sentiment analysis techniques, there is still opportunity for advancement and investigation into new approaches. The methods used are quite heterogeneous, ranging from data preparation approaches to sentiment analysis methods (TextBlob, VADER, LSTM models, SVM, BERT, and other machine learning classifiers). It is difficult to directly compare outcomes and make generalisations about the most successful approaches because of this variability. Moreover, most studies concentrated on Twitter data, which resulted in the exclusion of more comprehensive public opinion from other sources. Because not everyone participates in online debates, and this could generate biases. To effectively understand the results, it is important to comprehend the limits and potential biases included in social media data. This study aims to address the issues of noise, stopwords, emoticons, and class imbalance during the preprocessing of the data and conducting sentiment analysis by employing the pre-trained TRBSL model, and the unsupervised lexicons VADER and TextBlob models. In addition, K-Means clustering, and suspicious patterns method are used in this study to classify users as bots or humans. Finally, future studies are generally urged to address the drawbacks, explore multilingual sentiment models, and include more sophisticated features for complex analysis in political contexts. Additionally, not enough attention is given to the ethical aspects of sentiment analysis, such as user privacy and appropriate use of social media data. Therefore, in order to contribute to the standards of appropriate sentiment analysis practises, ethical issues should be taken into account. To contribute to requirements for ethical sentiment analysis procedures, this study highlights ethical procedures when using social media dataset.

2.6 Summary

This literature review highlights the key insights that Sections 2.1, 2.2, 2.3, and 2.4 play in sentiment analysis and user classification using social media datasets by emphasising these key points. In addition to contributing to the field of knowledge on sentiment analysis and user classification using social media datasets, these outcomes are also a beneficial resource for supervised, semi-supervised, and unsupervised sentiment analysis and user classification tasks. First, The collective review of these literature studies emphasises the value of sentiment analysis in assessing public sentiment on social media, however, issues with techniques and the necessity of continuous development are pointed out for further research. Second, the studies based on SA in political context gave significant insights on public opinions and feelings in a variety of political scenarios by highlighting the need for larger datasets, improved models, and advanced methodologies for sentiment analysis. Third, the methodological approach review provided a significant addition to the evolving area of sentiment analysis, highlighting the need of taking the specific context and analytic aims into consideration. Last, the cumulative results of the studies conducted to distinguish between spammy tweets and bots brought to light the constant evolution of bot detection, the importance for appropriate features, and the ongoing challenge of staying relevant in an era of continuously changing bot activity. These many approaches and results contribute to expanding understanding of the complex environment around spam and bot identification in social networks. Moreover, addressing the gaps and expanding on these findings will be essential for enhancing understanding of the findings of this study on sentiment analysis for local government elections in South Africa through unsupervised learning and to guide future studies and practices in the area. Therefore, since the unsupervised dataset is used to analyse the models, this study focuses on the limitations associated with using just positive and negative sentiment labels and add the use of GPT model to label the data for training purpose. Taking into consideration the challenges that were encountered in the studies reviewed by previous researchers, this study also aims to address the issues of noise, stopwords, emoticons, and class imbalance during the preprocessing of the data and conducting sentiment analysis by employing the pre-trained TRBSL model, and the unsupervised lexicons VADER and TextBlob models. In addition, K-Means clustering, and suspicious patterns method are used in this study to classify users as bots or humans. Finally, to contribute to requirements for ethical sentiment analysis procedures, this

study highlights ethical procedures when using social media dataset. The methods used to carry out this study are discussed in the next chapter.

Chapter 3

Methodology

This section presents a description of the dataset, exploratory data analysis, preprocessing, models, the fine-tuning procedure, evaluation measures, the ethical considerations and the tools.

3.1 Data Description

The study uses raw data from social media platform Twitter collected by Mashadi Ledwaba and Vukosi Marivate [2] prior to the 2021 local government elections in South Africa. The data collected is unlabelled and was collected using hashtags, as well as the names of the party leaders, hashtags for local municipal elections, and hashtags for South African political parties dating between September 2021 and October 2021. The zip compression is used to read the dataset. Unsupervised learning SA is performed on the unlabelled dataset using the three models mentioned in the Introduction chapter. Multiple field formats, including integer, string, and datetime, are present in the dataset. The tweets are written in a variety of languages, the majority of which being English and code-switched languages. Because the raw data is quite large and has many columns, the sample data is not shown.

3.2 Exploratory Data Analysis

This study examines data before doing sentiment analysis and evaluating the models. This gives further context and background information on the dataset in question [1]. The detailed information for the entire dataset, includes the memory allocation of 1.4 gigabytes, total of 41 columns and 727083 rows which include nine (9) empty columns, and column datatypes (datetime64[ns](1), float64(12), int64(7), object(21)). This information assists in better understanding of the raw dataset. Since certain users refrain from using hashtags, some decide to keep their profiles private, while others prefer to keep their geographical location totally hidden, it makes it difficult to collect all the information from social media networks [1]. In addition, the study looks for duplicates and null values. The majority of the columns contain null values, which indicates that some users feel uncomfortable disclosing sensitive personal information.

The dataset is filtered based on tweets containing phrases like Our_DA or Democratic Alliance, EFF or Economic Freedom Fighters or EFFSouthAfrica, ANC or African National Congress or MYANC, ActionSA or Action4SA in order to reveal the most often used phrases based on counts and help in providing a clear understanding of the data. According to the bigram frequency plot in Figure 3.1 which is based on the original dataset, the most common words are “south africa”, “vote anc”, “south africans”, and etc. In the trigram frequency plot of the top 15 most common words also in Figure 3.1, the most common words are “local government elections”, “corrupt lying looting”, “lying looting thieving”, and etc. It can also be observed that the top five phrases in the trigram mainly consist of negative phrases, suggesting that users are dissatisfied. Figure 3.2 provides a visual representation of a wordcloud comprising of top fifty most frequently used words, which is used to better understand various topics that the users are discussing. Words related to South African political parties, such as “ANC”, “EFF”, “people”, “vote”, and others, are among the most frequently used terms in the original dataset, as the

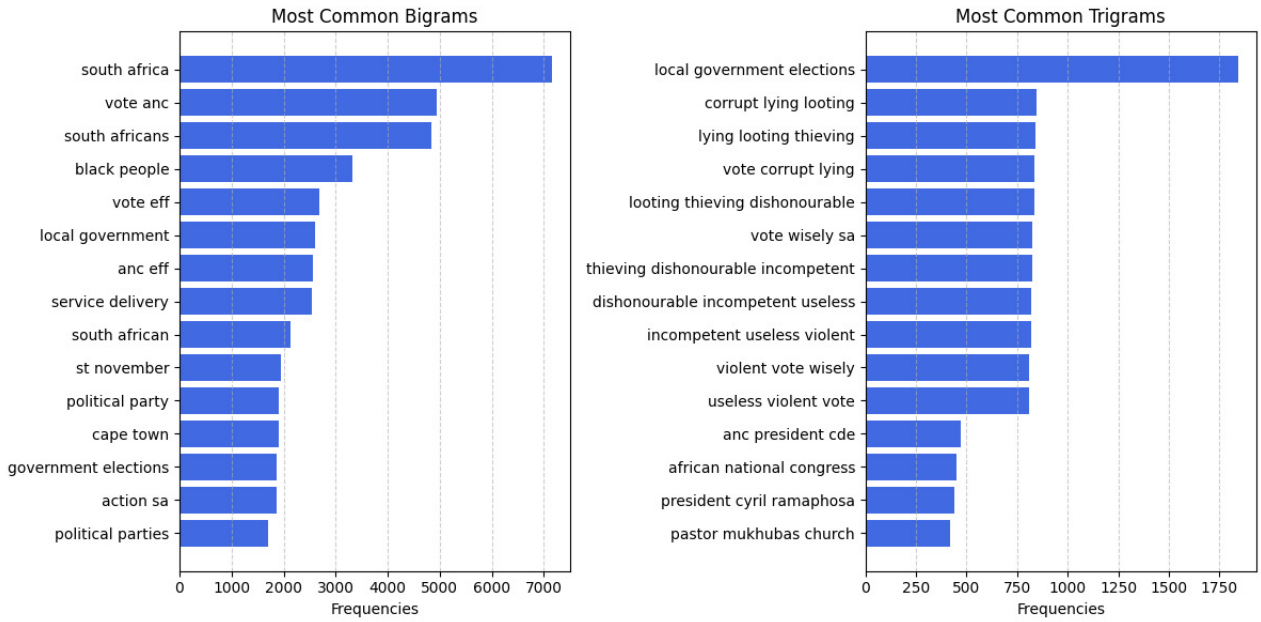


Figure 3.1: 15 Most Common Words In Bigrams And Trigrams

wordcloud shows. The data is then examined using the language column to determine the languages used in each tweet. Figure 3.3 shows that the top five languages of the majority of tweets are English (en), unknown (und), Lingala (in), Tagalog (tl), and Dutch or Flemish (nl). “African National Congress” (ANC), “Democratic Alliance” (DA), “ActionSA”, and “Economic Freedom Fighters” (EFF) are the political parties that are used in the study and are taken into account while analysing the tweets. The number of tweets for each political party prior to preprocessing is shown in Figure 3.4.

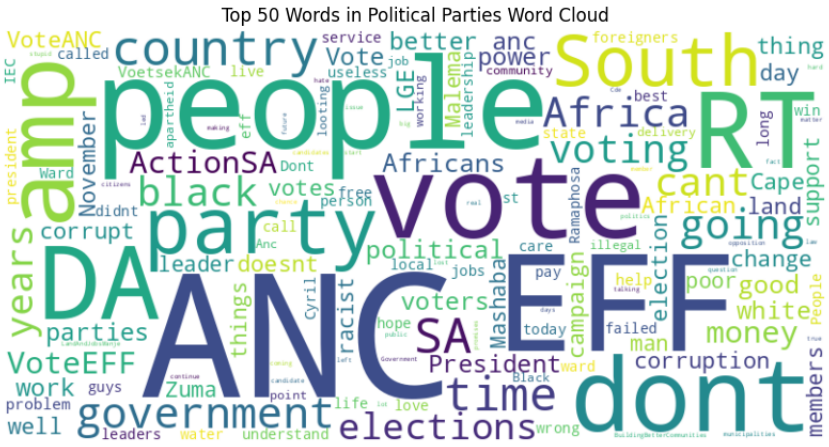


Figure 3.2: WordCloud of Top 50 Frequently Used Words

The time series plot in Figure 3.5 displays a time plot which demonstrate the variation over time for the tweet counts per political party. There seems to have been a lot of tweets about the ANC and EFF around 21st September 2021, and before 31st October 2021, which is roughly close to 2500 tweets. These seem to indicate that the ANC was the most frequently talked about political party, followed by the EFF, DA, and lastly ActionSA, which was the least talked about political party based on the original dataset time series plot (Figure 3.5).

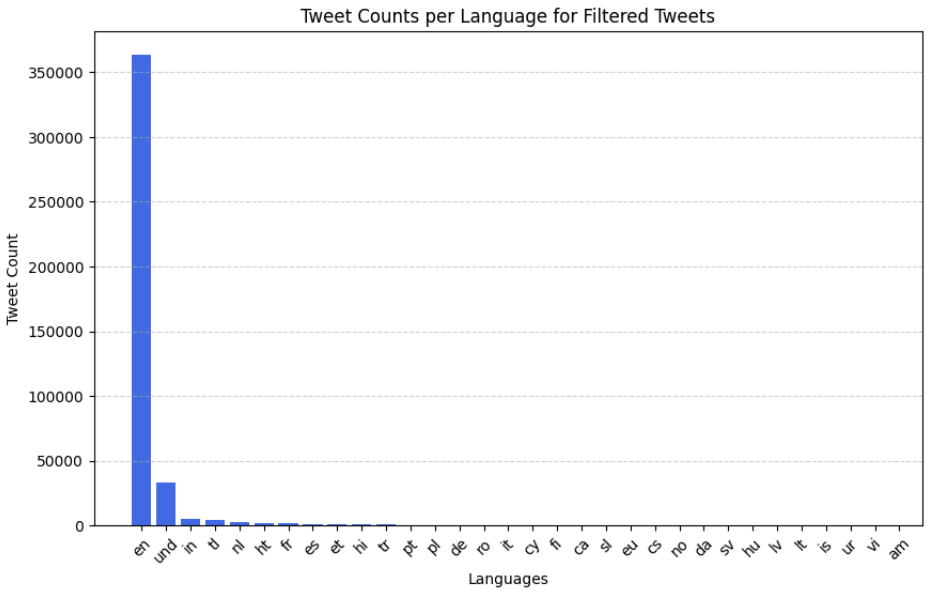


Figure 3.3: Language Distribution Of Tweets

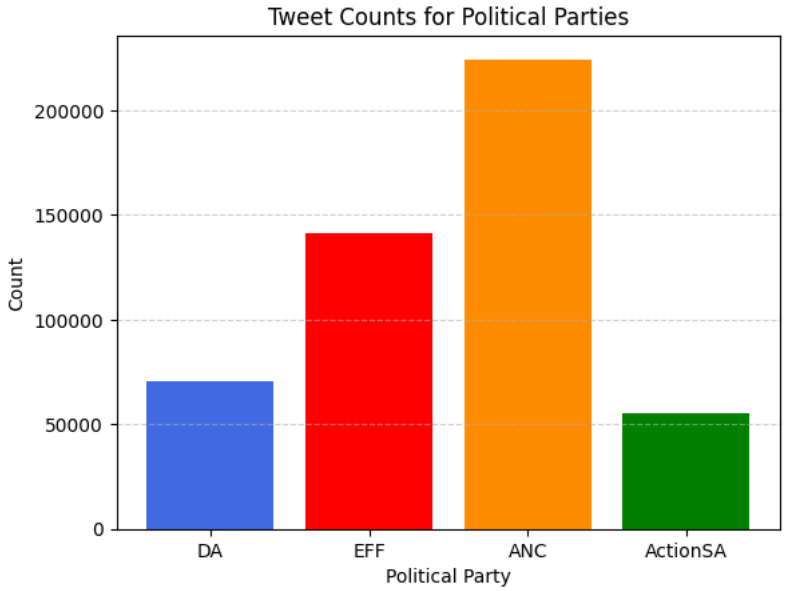


Figure 3.4: Political Party Tweet Count

3.3 Preprocessing

The most challenging task in natural language processing is dealing with unstructured texts. Tweets usually include out-of-character text with hyperlinks, emoticons, punctuation, and other unusual text formats. Since the chosen models don't have built-in features for removing hyperlinks, mentions, and hashtags, in this study preprocesses of the dataset is done. Although pre-processing is optional when utilising sentiment analyzers like VADER, TextBlob, and TRBSL, it is necessary to meet the requirements of this study. The dataset is filtered using the keywords displayed in Table 3.1 in order to greatly simplify and help speed up the preprocessing procedure. Additionally, the dataset is further refined by selecting tweets with more than ten words, and the "language" column is set to English (en). This process results in 264,267 rows from the original 727,083 rows. The following columns: user_id (data type: integer), username (data type: string), tweet (data type: string), hashtags (data type: string), language (data type: string), and datetime (data type: datetime) are kept as part of the reduced dataset, while the remaining columns from the original dataset are excluded because they contain null values with memory

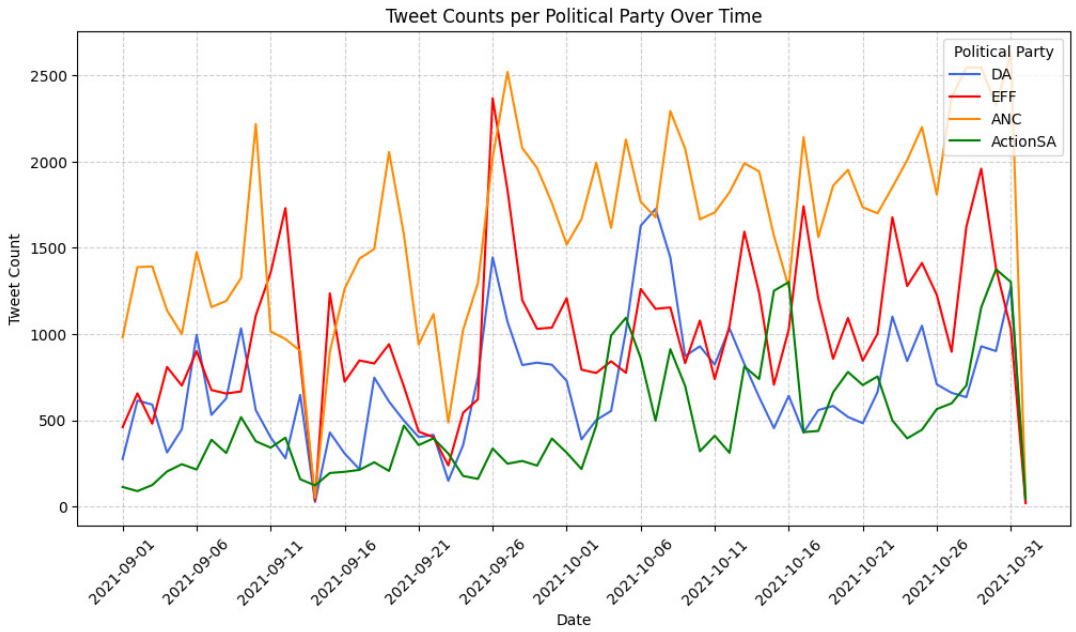


Figure 3.5: Time Series Plot Per Political Party Tweet Count

usage of plus 12.1 megabytes.

Keywords
Our_DA, VoteEFF, Economic Freedom Fighters, LGE2021, EFF, VoteActionSA, VoteDA, African National Congress, EFFSouthAfrica, VoetsekANC, Democratic Alliance, ANC, ActionSA, VoetsekDA, VoteANC, VoetsekActionSA Action4SA, VoetsekEFF, MYANC

Table 3.1: Keywords To Select The Dataset

The study uses pre-processing techniques like data cleaning, tokenizing, and NLTK Default-Stopword List for stop-word removal which is combined with custom list of stopwords to address these problems and get the data ready for analysis. Forty (40) stop words, which include words such as ”a, go, of, an, the” etc., are already included by default in NLTK. To add to the existing NLTK Default-Stopword Lis, a custom list of 667 stopwords is added from Ranks NL website [36]. For data cleansing, retweets, replies, usernames, unnecessary symbols, non-letters, punctuation, non-hashtags, hyperlinks, whitespaces, outliers, and missing values are removed. To carry out the analysis, all elements in the ”hashtag” column are transformed into tuples and the contractions in each tweet are substituted. A new column called ”date” with the datetime format was created by extracting the date from the datetime column. To make the texts of the tweets all lowercase and uniform, stop words such as “to”, “the”, “but”, and “their” are removed. The tweets are tokenized to generate a new column named ”filtered tweets. Table 3.2 displays a the details of the preprocessed dataset dataset, which has 56,993 rows and 9 columns free of duplicates and null values. Furthermore, Table 3.3 displays the sample tweets based on the original tweets and the filtered tweets which are preprocessed. The contents of the columns “username” and “user_id” are hidden and these columns are not shown as a result of ethical concerns, in contrast, tweets are paraphrased to preserve the main points of a tweet while respecting the privacy of users and sensitivity concerns. The final cleaned dataset contains 31,019 tweets about ANC, 15,628 tweets about EFF, 6,570 tweets about DA, and 5,359 tweets about ActionSA.

Column	Count	Data Type
user_id	56981	int64
username	56981	object
tweet	56981	object
datetime	56981	datetime64[ns]
hashtags	56981	object
language	56981	object
tweet_nousername	56981	object
filtered_tweets	56981	object
date	56981	datetime64[ns]

Table 3.2: Dataset Description

tweet	tweet_nousername	filtered_tweets
Is it not possible for us to send the whole @MYANC to Afghanistan in order to assist in the rebuilding of the cowntry? There is nothing else to screw up, therefore it can't get much worse there.	is it not possible for us to send the complete team to afghanistan to assist in rebuilding the cowntry there is nothing more to screw up there so it really cannot get worse	send entire afghanistan help rebuild cowntry things worse f**k
EFF RISE ***** 2020 EFF SOUTHAFRICA DOWNWITH RACIST downward ANC downwards DA downward EFF Rise EFF RISE https://##########	***** ANC DOWNWITH RACIST DOWN AND DA DOWN EFFSOUTHAFRICA RISE OF EFFLGESA RISE	***** effsouthafrica downwith racism anc da eff lge sa eff raise eff rise
We've been taken advantage of by the @MYANC for years. Perhaps it is time to repay the courtesy. https://##########	The has been deceiving us for a long time Perhaps it is now time to repay the courtesy	removing years time repay favours

Table 3.3: Sample Tweets From 2021 South African Election Dataset

3.4 Approach/ Models

The analysis of the study is conducted using unsupervised learning models. The use of artificial intelligence (AI) methods to find trends in datasets that are not labelled or even grouped is known as unsupervised learning. The TRBSL and Lexicon-based classifications (VADER and TextBlob) are pretrained models and do not require previously labelled data. Figure 3.6 depicts the flow diagram used to conduct sentiment analysis and user classification on local government election tweets and the classification of the users. Liu [37] defined sentiment analysis as the evaluation of people's views, emotions, reviews, behaviours, and sentiments as articulated in written communication.

3.4.1 Polarity Sentiment

To determine the polarity, the tweets are labelled into three categories: negative (numerical label 0), neutral (numerical label 1), and positive (numerical label 2). To begin, sentiment analysis for unsupervised learning is carried out using the TRBSL, VADER and TextBlob algorithms to determine the polarity of the tweet text. The TRBSL model uses the polarity score between 0 and 1 which represents the confidence level of the model's prediction, with higher scores indicating higher confidence. Vader decides how negative, neutral, or positive the texts are as well as how the vocabulary list is skewed. Its output is a Python dictionary with the following four key and value pairs: "neg", "neu", "pos", and "compound".

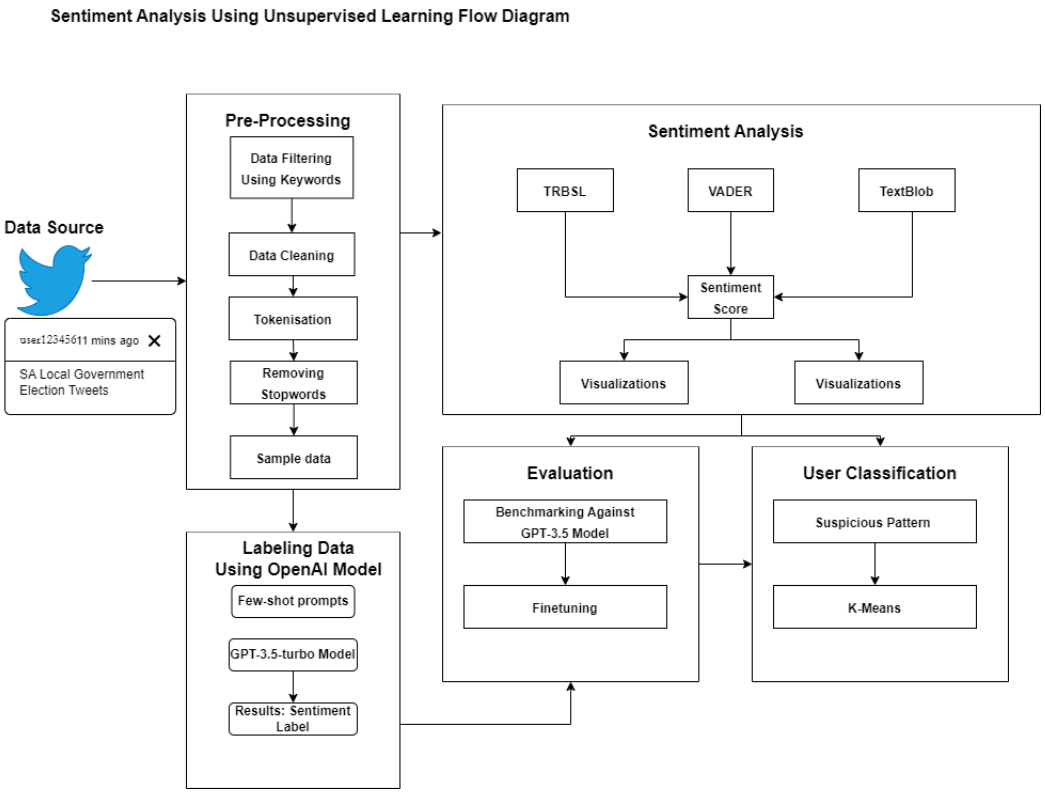


Figure 3.6: Flow Diagram For Unsupervised Sentiment Analysis And User Classification.

The compound score is taken as the overall polarity score for the texts and used to label the tweet The text and weight dictionary used by TextBlob contains scores that are used to determine the sentence's sentiment. The score for polarity ranges from -1 to 1, with 1 denoting the most positive texts and -1 denoting the most negative. Both models employ the same decision metric.

3.4.2 User Classification

Since it can be difficult to maintain a reasonable level of quality in data when there are malicious users on Twitter, a number of time-consuming detection techniques have been developed to check tweets or accounts of individuals for the presence of spam [29]. The study classifies users as either human or bot using suspicious patterns method and the K-Means clustering model. The study considers the following factors to help classify the tweets as being generated by humans or bots:

- Tweets that direct users elsewhere, such as another Twitter profile, product, or service.
- Tweets containing only links.
- Tweets that appear to be coordinated by a single person or several accounts from other users in an attempt to increase or control tweet engagement.
- Tweets with identical or similar content from a single or multiple accounts managed by a single user.

First, a list of frequently used spammy phrases in Table 3.4 is applied to find suspicious patterns in tweets. This allows us to classify users and identify whether a tweet was created by a human or a bot. Following that, a function is built that iterates through each row in the dataset using a for loop to extract the values of the "tweet" column, and it utilises an empty list that holds information about the suspicious tweets. The function then loops over the list of patterns passed in as an argument for each tweet. The current pattern that is in the suspicious patterns list is then searched for in the tweet text using the

re.search function. Additionally, the re.IGNORECASE flag is applied to make the pattern matching case-insensitive. If a pattern is found in the tweet text, the function checks if it contains the substring "CALL/WHATSAPP" or "link in bio" and label the tweet as "irrelevant", else it labels as "bot". The "irrelevant" tweets are tweets that are not related to the 2021 local elections, though they have used election hashtags to get attention. In order to reduce noise, the methods are used to eliminate the meaningless tweets from the analysis. At the end of the function, the suspicious_tweets list is returned, which consists of the twitter posts that resemble the defined patterns as well as their labels. The method is extremely useful for finding and categorising tweets in a dataset that display unusual behaviours based on predetermined trends, particularly for detecting bot-generated tweets on social media sites.

Suspicious Patterns
click here,
click the link,
link in bio,
visit https://t.co/,
CALL/WHATSAPP

Table 3.4: Suspicious Patterns Used For User Classification

Second, the study preprocesses the tweet texts using 'Term Frequency-Inverse Document Frequency" (TF-IDF) for the K-Means clustering model. The Elbow method is then applied to ascertain the ideal clusters. The TF-IDF algorithm converts a set of text documents into a numerical representation suitable for machine learning and other text analysis activities. It uses the scikit-learn library's TfidfVectorizer class id to transform a collection of text documents into a TF-IDF feature matrix and then creates an object. Words that appear in more than 85% of the documents are disregarded during the TF-IDF procedure in order to exclude relatively common words that might not offer much information. To restrict the dimensionality, the study limits the number of features in the TF-IDF matrix to a maximum of 2000. Stopwords is also employed in the removal of popular English words from tweets. The K-Means clustering process uses the TF-IDF matrix and a loop to fit the K-Means for a range of different cluster counts, from 1 to 10. It uses the initialization technique "k-means++" to build a K-Means model with the specified number of clusters for each iteration. A maximum of 300 iterations and 10 initializations are implemented to ensure the optimal clustering outcome. The "within-cluster sum of squares" (WCSS), which computes the total distances squared between data points and their designated cluster centroids, is determined after the model has been fitted to the TF-IDF data. The values of the WCSS measure, which is used to assess the quality of clustering, are kept in a list for each cluster. The "elbow method" plot is used to plot the WCSS values versus the number of clusters after iterating through the range of cluster numbers. The "elbow point," or the place on the plot where the rate of decline in WCSS abruptly changes, is provided by the "elbow method" plot, which assist in finding out the best number of clusters. This point is then used to choose how many clusters to take into consideration for the task.

The two methods mentioned above, are used one after the other to ascertain whether the tweets were created by a human or a bot. To authenticate the results, manual verification is conducted.

3.5 Fine-tuning Process

Model improvement entails fine-tuning models that performed poorly in the original sentiment analysis. Fine-tuning is the process of transferring learning for a specific issue using a pre-trained model on a limited subset of data. Since pre-trained models have already been trained on a bigger dataset, fine-tuning is a strong strategy for training the models used in this study. It enables the models to modify their features and transfer their learning to a smaller, unseen dataset. When optimizing natural language processing models, a number of approaches are used, including domain adaptation, transfer learning, and knowledge distillation. The advantages of fine-tuning a model include enhanced performance, the

ability to analyse and manage a set of data, and shorter training times.

The chat model endpoint gpt-3.5-turbo (GPT-3.5) from "Generative Pre-trained Transformer" (GPT) language models is used to first label the data with the help the of human verification because the dataset is unlabelled and fine-tuning requires labelled data. GPT language models are OpenAI models that have been pre-trained to understand natural language and coding while generating text outputs based on input [38]. A random sample of 6000 rows and 9 columns is used to determine the subset of the pre-processed dataset. The input for the OpenAI model is a column called "filtered_tweets", and the model provides a positive, neutral, or negative label for each tweet analysis. To reduce pressure on the running process, the "filtered_tweets" are divided into batches of 300. In order to help the OpenAI model adapt and finish each assignment effectively, the study uses a few shots to give the model a few instances in the prompts. A Few-shot learning technique is used in this study to give the OpenAI model a few examples in the prompts so that it can swiftly adjust and finish each task effectively (Table 3.5). GPT-3.5 has previously achieved near-state-of-the-art performance in open-domain NLP few-shot knowledge transfer [39,40]. Figure 3.7 displays the sentiment label distribution generated using GPT-3.5. For each tweet, the model gets called five times. The presence of humor, sarcasm, or irony in the tweets are taken into account (Table 3.5). Although there is no clear class imbalance, the classes are not evenly distributed. The labelled dataset by OpenAI contains 2195, 2026, and 1779 negative, neutral, and positive tweets, respectively. Class imbalance occurs when there is a meaningful difference in sample size between the minority class and the other classes [41]. This is handled during the model fine-tuning processing phase.

Prompts
"An example of a negative tweet or statement about South African political party EFF: Land expropriation without compensation? More like land grabs without a plan EFFLandGrab VoetsekEFF EFFMustFall."
"An example of a negative tweet or statement about South African political party ANC: The ANC's legacy of corruption and mismanagement is a stain on our country. Time for a change VoetsekANC ANCScandals ANCDivisions ANCMustFall."
"An example of a negative tweet or statement about South African political party DA: The DA's inability to address racial disparities effectively shows they're out of touch with South Africa's realities. DAFail DAMustFall."
"An example of a negative tweet or statement about South African political party ActionSA: ActionSA's lack of experience in governance makes me question their ability to lead effectively."
"Positive tweets or statements about these political parties usually include VoteForChange, HopeForSA, EFFForEquality, VoteANC, VoteMYANC, VoteEFF, VoteEFFSouthAfrica, VoteDA, VoteOur_DA, VoteAction4SA, VoteActionSA, ActionForChange, EFFForThePeople."
"The presence of humor, sarcasm, or irony in the tweets or statements must be considered."
"Classify this new tweets based on the examples provided above, the response should be either positive, or negative or neutral only, please do not provide justification to the results. Now classify this new tweets:"

Table 3.5: Few-shot Prompt For OpenAI GPT-3.5 Model

To improve the models that perform poorly, the following steps for fine-tuning are followed:

1. **Preparing Data**: The preparation of the data is necessary before fine-tuning a pre-trained model. The new dataset is divided into three sets: a training set for model training, a validation set for hyperparameter adjustment, and a testing set for model performance assessment. Prior to the split, the sentiment labels are translated into numerical values, i.e. negative: 0, neutral: 1 and positive: 2.

2. **Architecture of the Model**: The architecture of the model that was previously trained is employed, with certain layers trained while others are frozen. This phase entails employing the architecture of a pre-trained model that has been trained on an enormous textual corpus. The

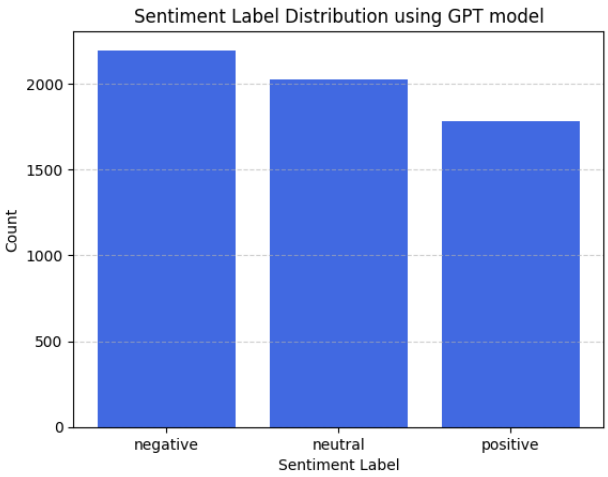


Figure 3.7: Overall GPT-3.5 Sentiment Distribution

measures described below have been implemented to address the class imbalance and enhance the performance of the fine-tuned model:

- The PyTorch train dataset and sample weights are generated together with the calculation of the class weights.
- The training dataset applied to the tokenization procedure, with the format set to torch.
- For padding, a data collator is produced, and for training data, a weighted sampler is created. This addresses the class imbalance throughout training and prioritises the underrepresented classes.
- The weighted sampler is applied to the PyTorch train dataloader.

3. **Hyperparameter Tuning**: In this stage, the parameters that are not transmitted during model training, including batch size and learning rate, are adjusted. The optimal values for the parameters are chosen to attain the best results on the validation set.

4. **Training**: The pre-trained model is refined using the new dataset labelled by utilising the GPT-3.5 model, and its performance is tested using the validation set. The pre-trained model is initialised using the weights that were learned from the prior training.

5. **Evaluation**: In order to assess if the model can be deployed for usage in the future, its performance is assessed on the test set in this last phase. This measures how well the model performs in general.

The assessment and testing phases are crucial to fine-tuning pre-trained models since they ascertain how effective the model is and whether it is appropriate for the issue that is investigated.

3.6 Evaluation Metrics

To evaluate performance, this study uses basic statistics (mean, median, and standard deviation), a confusion matrix, precision, recall, F1-score, accuracy, and average measures.

1. **Basic statistics**: Basic Python methods are used to determine the scores for mean, median, and standard deviation based on the sentiment labels.

    i Mean - is the mean sentiment intensity of the text samples that is analysed, and the formula is given by:

$$\text{mean} = \frac{\sum(\text{compound\_scores})}{\text{len}(\text{compound\_scores})}$$

The mean score closer to 1 indicates that, on average, the analysed texts tend to be very positive, while a mean score close to -1 suggests they tend to be very negative.

ii Median - is the middle value of the sentiment scores when they are sorted in ascending order. The formula is given by:

$$median = (\text{sorted}(\text{compound\_scores}))(\text{len}(\text{compound\_scores}))//2$$

It provides a central tendency metric that is less impacted by the highest or lowest values.

iii Standard deviation (std_dev) - measures the spread or dispersion of the sentiment scores. The formula is given by:

$$std\_dev = \frac{\sum[(x-mean)^2 \text{ for x in compound\_scores}]}{\text{len}(\text{compound\_scores})^{0.5}}$$

The scores are more widely distributed when the standard deviation is larger, which indicates that the scores are significantly more divided around the mean when it is less.

2. **Confusion Matrix (CM)**: It describes the classification performance of the model by showing the count of the predicted true positives (TP), true negatives (TN), false positives (FP) and false negatives (FN) which assist in understanding the performance of the model in classifying the data points correctly.

3. **Precision (P)**: It measures the number of the actual true predicted positives which assist in understanding the accuracy of the predicted positives. The formula to calculate the precision is given by:

$$\mathbf{P} = \frac{TP}{(TP+FP)}$$

4. **Recall (R)**: It measures the misclassification done by the model and shows the weighted average of the correctly predicted labels per class. Recall provides the full understanding of the ability of the model to correctly identify positive instances. It provides the ratio of the true positives to the sum of true positives and false negatives. The formula is given by:

$$\mathbf{R} = \frac{TP}{TP+FN}$$

5. **F1-score**: It combines the recall and precision to calculate their harmonic mean which provides the balance between the precision and recall. F1-scores are a very important measure as they rate the system by using both **P** and **R**. The formula is given by:

$$\textbf{F1-score} = 2 * \frac{(P*R)}{(P+R)}$$

6. **Accuracy (Acc)**: It measures how many times the correct sentiment happens whereby the correctly labelled texts count is divided by the total texts count. Accuracy is said to be the most significant performance evaluation measure for classification tasks but using it alone in some cases does not provide the results that are required [42]. The formula can be written as:

$$\mathbf{A} = \frac{(TP+TN)}{(TP+FP+FN+TN)}$$

7. **Weighted Average (WA)**: It is an average where each metric of the class is weighted by the count of instances in that class where it accounts for the class imbalances in the dataset.

8. **Macro Average (MA)**: It independently provides the metric for each class and takes the average which helps in giving the classes equal importance.

The metrics help in evaluating how well the model for sentiment analysis is performing. The confusion matrix provides an in-depth breakdown of both positive and negative predictions, while precision, recall, and F1-score focus on different aspects of model performance. Accuracy gives a general idea of overall correctness and weighted, and macro averages help address imbalances in multi-class classification.

3.7 Ethical Consideration For Using Twitter Data

To ensure the appropriate and ethical use of data while using Twitter datasets, acceptable standards of ethics must be considered and followed. This study adheres to the following key ethical considerations:

1. **Privacy** - This study ensures that the personal details of users, such as names, locations, and actual tweet texts, are not exposed. To safeguard user privacy and prevent the release of sensitive information without explicit permission, this study anonymises data.

2. **Openness And Research Integrity** - The dataset is exclusively used for the research study and not for any other reason. Additionally, this study adheres to ethical practices and standards for research, ensuring that the techniques used are providing accurate results and add to the field of knowledge in sentiment analysis and user classification.

3. **Data Security** - To prevent unauthorised people from accessing the dataset, it is kept in a password-protected location.

4. **Non-exploitative Use** – Although this study is conscious of the possible effects it may have on users, the dataset is not used in a way that would cause harm, promote destruction, or infringe upon the rights of the Twitter users.

5. **Fairness and Biases** - The goal of this study is to analyse and interpret the data in a way that is fair and does not reinforce or magnify pre-existing biases.

3.8 Tools

Python is used as a programming tool in this study for analysis and visualisation since it is user-friendly and contains all of the essential libraries for sentiment analysis.

3.9 Summary

This chapter provided an overview of the systematic approach used to explore sentiment analysis for South African local government elections using unsupervised learning. Under this methodology chapter, the study starts with an extensive overview of the dataset, providing more details about its features in the Data Description section. Subsequently, various methods are used in the Exploratory Data Analysis section to uncover insights and patterns, therefore building the foundation for informed preprocessing decisions. Moreover, the procedures used to clean, convert, and get the data ready for further analysis are described in Preprocessing section. The Approach/ Models section then delves into two crucial subsections: Polarity Sentiment and User Classification, where the specific methods and models used to evaluate sentiment and classify users are described in these two subsections. Moreover, the Fine-tuning Process section describes the details of the refinement made to enhance the performance of the models and the evaluation metrics that are used to determine if the applied techniques are effective are covered in the evaluation Metric section, providing a thorough discussion of each metric. The Ethical

Consideration For Using Twitter Data section delves deeply into ethical questions surrounding the usage of Twitter data, highlighting the dedication to ethical research methods. The methodology chapter is concluded with the Tools section, which describes the technology and software used during the study process. Overall, this methodology provides a clear framework for the way this study is conducted, the analysis, and ethical approach to Twitter data. The analysis and results from user classification and sentiment analysis using unsupervised learning for the South African local government elections are presented in the next chapter.

Chapter 4

Analysis And Results

This section focuses on the results and comparative analysis of the three models to assist in answering the research questions in Section 1.1. The analysis and results of the models are discussed separately, followed by a summary. First, sentiment analysis is done using the subset of the preprocessed dataset, secondly the fine-tuning and the implementation of weights to balance the classes on the model which did not perform well in the first analysis using the results of the GPT-3.5 model to train, evaluate, and test the model.

4.1 Data Sampling

For sentiment analysis, a random sample of 5000 rows is selected from the preprocessed dataset which is unlabelled to do the analysis for the three selected models. The sentiment labels are divided into three: negative, neutral and positive with a score between -1 and 1 for VADER and TextBlob, and 0 and 1 for TRBSL model. Similarly, for user classification, a random sample of 5000 tweets is selected to do identify the users based on the tweets using the suspicious patterns and K-Means methods.

4.2 Unsupervised Sentiment Analysis and Results

The primary objective of this study is to extract sentiments from Twitter datasets that had not been labelled, and since the dataset is not labelled, building a model capable of efficiently uncovering sentiments from tweets appears to be a feasible endeavour [43]. For the VADER and TextBlob models, the mapping function for polarity sentiment labels is defined in Table 4.1. The mapping function was not used on the TRBSL model as it is not required for it. For each model a batch size is defined, with TRBSL having a batch size of 150, Vader with a batch size of 300 and TextBlob with a batch size of 100. The results of this analysis answers the research questions 1 (How do the polarity sentiments vary across the four political parties during the South African local elections campaign period between September and October 2021?) and 2 (How does the sentiment expressed by Twitter users during the 2021 local government elections campaign evolve over time?) outlined in Chapter 1, Section 1.1.

label	**polarity score**
positive	$\geq$ 0.05
negative	$\leq$ -0.05
neutral	otherwise

Table 4.1: Sentiment Polarity Mappings

4.2.1 Twitter-roberta-base-sentiment-latest (TRBSL)

The Twitter-roberta-base-sentiment-latest model is used to perform the sentiment analysis in this section. Table 4.2 shows the overall distribution of the sentiment labels and their count where positive labels

have a count of 414, with 1650 for negative and 2936 for neutral. This clearly shows that the model is failing to label the positive tweets and there is clearly class imbalance.

Sentiment Label	Overall Count
neutral	2936
positive	414
negative	1650

Table 4.2: Sentiment Label Count - TRBSL

Statistic	Score
Mean	0.770454
Median	0.802965
Standard Deviation	0.132000

Table 4.3: Statistical Distribution - TRBSL

Basic Statistical Analysis

Table 4.3 shows the basic statistical analysis results mean, median and standard deviation with 0.7705, 0.8030 and 0.1320 scores respectively. The mean score of 0.7705 indicates that on average all the analysed tweets have a strong positive sentiment. The median score of 0.8030 indicates that the sentiment scores of the text samples tend to be positively skewed. In other words, a large proportion of the text samples likely have very positive sentiments, while there may be a smaller number of samples or no samples with neutral or negative sentiments. The standard deviation score of 0.1320 suggests that the sentiment scores are relatively tightly clustered around the mean of 0.7705, which indicates that most of the text samples have sentiment scores close to the mean, with relatively little variability in sentiment intensity. Overall, the statistics in Table 4.3 suggest that the tweets analysed using the TRBSL model have a strong positive sentiment on average, with most of the samples having sentiment scores close to the mean and the relatively small standard deviation indicates that the sentiment scores are consistent and not highly variable.

Sentiment Analysis and Results

The sentiment distribution in Appendix A Figure A1 shows the polarity sentiments variation across the four political parties during the 2021 South African local elections campaigns which suggests that tweets by the users have the highest positive sentiment across all political parties. The most positive tweets are regarding ANC, followed by EFF, then DA and lastly ActionSA, suggesting a slight variation among the four political parties in terms of the positive tweet count. The insights of the results of the TRBSL model also suggest that ANC is likely the most disliked political party with a lot of negative tweets count. It can also be seen on Figure A1 that EFF and DA have almost the same number of negative tweets, with ActionSA having the least number of tweets in all sentiment classes. Furthermore, based on the sentiment analysis conducted using TRBSL model, the positive tweets are way too less than the neutral and the negative tweets. The neutral tweets are significantly more as compared to both the negative and positive tweets. This suggest that there is a class imbalance in the dataset. Figure A3 in Appendix A shows the top 15 phrases frequency for ANC with positive sentiment (Figure A3b) suggesting phrases such as "building better communities", "anc president ancinjoburg" and words related to "campaign trail and president" as having higher frequency among other phrases. This suggest that the users are mostly tweeting about the ANC campaigns for the elections. In contrast, words such as "vote wisely sa", "lying looting thieving" and "corrupt lying looting" are seen amongst the phrases with the highest frequency in Figure a of the negative sentiment, which may suggest that the users are mostly tweeting about how "corrupt" the political party is seen. Moreover, the neutral sentiment trigram plot for the ANC in Figure c reveals phrases such as "local government elections", "voteanc anclge buildingbettercommunities" and "voteda voteda voteda" as highly dominant. Figure A4 in Appendix A shows the top 15 trigram frequency plot for DA based on the sentiment labels. The positive trigram plot in b has words "liberation transformation success", "mayoral candidate peter", "viva da viva" as the most dominant. In contrast, for negative sentiment (Figure a) the users are mostly tweeting about "racist da", the "massacre" and "phoenix massacre", which suggest that the political party is seen as "racist". Furthermore, phrases such as "local government elections", "federal council chairperson" and "leader john steenhuisen" are amongst the dominant phrases in the neutral sentiment (Figure c). The frequency

trigram plot in Figure A5 and A6 in Appendix A suggest neutral sentiment in the tweets with words mostly related to voting for EFF and also for ActionSA.

It can also be seen in the sentiment time series plot overtime in Appendix A Figure A7 that overtime the neutral sentiment expressed by Twitter users for the ANC political party (Figure A7 a) gradually increased as compared to the start of September 2021 where the negative sentiments was more dominant. It can also be observed that the positive sentiment seems to be having no significant difference overtime until around 22^{nd} October 2021, but also there is not much difference even during that period. The sentiment overtime for DA in Figure A7b shows a dominant neutral sentiment which suggest that most of the tweets about DA are likely neutral and there is evidence that the negative sentiment gradually increased between 1^{st} and 15^{th} October 2021. The sentiment overtime for EFF in Figure A7 c shows the most neutral tweets between 22^{nd} September and 1^{st} October 2021. The positive sentiment regarding EFF is not showing any significant increase but it can be observed that there is a gradual increase in the negative sentiment between 22^{nd} and 31^{st} October 2021. Furthermore, it can be observed on Figure A7d that there is gradually increase in the neutral and negative sentiment for Action SA, where it can be seen between 1^{st} and 10^{th} October 2021, and also a huge increase between 22^{nd} and 31^{st} October 2021. It should also be noted that ActionSA has almost the same tweet count for positive and negative sentiment around 22^{nd} October 2021. Hence, it can be seen that the neutral sentiment is very dominant over time in all the time series plots.

Therefore, although there seems to be class imbalance in the dataset, there is still significant distinct amongst the political parties and the different sentiment groups which clearly shows shifts during the different phases of the 2021 local government elections campaigns or preparations.

Benchmark Against the GPT-3.5 model

In this section, the evaluation metrics were used to benchmark the TRBSL model against the GPT-3.5 model. The sentiment analysis results between the TRBSL and GPT-3.5 models in Figure A8 in Appendix A suggest that both models disagree on sentiment counts. The sentiment counts for TRBSL are mostly distributed toward the neutral sentiment, whereas the negative sentiment for both model shows a slight variance. Furthermore, the models shows a significant variance when comparing the positive sentiment. Figure A8 shows the sentiment variation between the two models, where it can be clearly seen that the TRBSL model has less positive tweets as compared with the GPT-3.5 model, which indicate that the TRBSL model did not perform well on identifying the positive tweets from the dataset. The results also show that out of the 5000 tweets, 9.72% of them are positive tweets, 35.82% are actually negative tweets and 54.46% are neutral tweets. This suggest that most tweets in the dataset are actually neutral tweets, which is not the case when comparing to the results of the GPT-3.5 model. The confusion matrix in Figure 4.1 implies that the model was able to predict 1171 instances for negative, 1025 instances for neutral and 333 instances for positive. Some of the instances were incorrectly predicted as either positive, negative, and neutral.

	precision	**recall**	**f1-score**	**support**
0	0.65	0.70	0.68	1667
1	0.38	0.62	0.47	1666
2	0.69	0.20	0.31	1667
accuracy			0.51	5000
macro avg	0.57	0.51	0.48	5000
weighted avg	0.57	0.51	0.48	5000

Table 4.4: Classification Report For GPT vs TRBSL

The classification report shown in Table 4.4 includes some of the metrics mentioned in section 3.6: Eval-

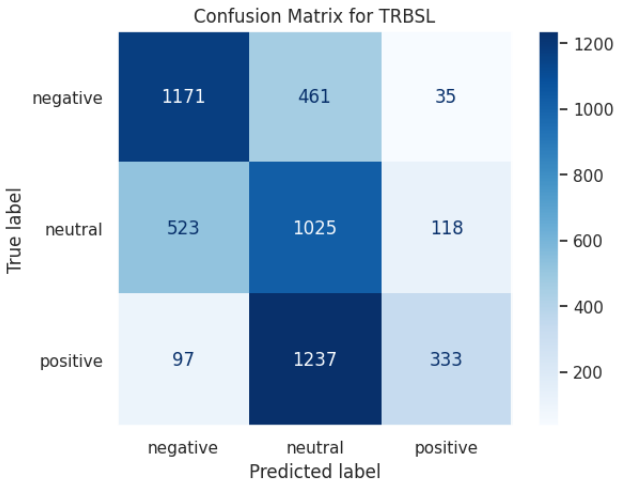


Figure 4.1: Confusion Matrix For GPT vs TRBSL

uation Metrics (in Chapter 3: Methodology) to evaluate the performance of a model, and it's based on the confusion matrix. The precision for the positive class i.e., class 2 is 0.69, which indicates that out of all the instances predicted as class 2, 69% are correctly classified as class 2. The recall for the positive class is 0.20 which indicates that only 20% of the actual positive class instances are correctly classified as positive. The F1-score for the positive class is 0.31, which indicates that there is no balance between precision and recall. The overall accuracy of the model across all classes is 0.51, which means that only 51% of all predictions are correct. The macro-averaged and the weighted average precision, recall, and F1-score are between 0.48 and 0.57 which indicate that there is a slight balance between the classes. Overall, the results in Table 4.4 indicates that the model was able to only predict the sentiments with approximately 51% accuracy and F1-scores of 68%, 47%, and 31% for negative, neutral, and positive sentiments, respectively.

In conclusion, the results indicate that the model performs well in identifying negative labels (class 0) and neutral (class 1) but poorly in classifying positive (class 2). This implies that the accuracy of the model in predicting class 2 may be inaccurate, suggesting further investigating and model enhancement. Therefore, to improve the model, fine-tuning will be done.

Fine-tuning Results

The first step in fine-tuning is making sure that the dataset to be used is preprocessed. In this section, the dataset labelled using the GPT-3.5 model is used. The dataset is split into three sets, the training set with 70% of the data, and testing and validation sets with 15% each of the data. The pre-trained TRBSL model is loaded using "AutoModelForSequenceClassification" from the Hugging Face Transformers library. The tokenize_function is also defined in order to tokenize the tweets in the dataset. The function takes a batch of examples as input and returns the tokenized version of the text, with padding and truncation to a maximum length of 152 tokens. The 124,647,939 million parameters of the TRBSL model are all trainable (Figure A9 in Appendix A).

The epoch is set to 3 with the evaluation metrics. Appendix A, Table A1 shows the training results of fine-tuning the TRBSL model. In the first epoch, the training loss of 0.6363 is relatively high, this indicates that the model is still learning, and the validation loss suggests that the model is not performing well at approximately 0.6677 at this stage. The accuracy and F1-score are both approximately 0.7245, indicating that the TRBSL model properly classifies approximately 72.45% of the training data. The training and validation losses are decreasing in the second epoch, indicating that the model is improving. The accuracy and F1-score of the training also show an improvement at approximately 0.7931, indicating that the model is predicting 79.31% of the training data correctly. The last epoch, which is the third epoch indicates that the training loss continues to decrease, however, the validation loss increases at

this epoch. This suggests that the training data may have been overfitted at epoch 3. The model appears to be successfully predicting the training data with 84.14%, as indicated by the accuracy and F1-score of the training, which both demonstrate an improvement with approximately 0.8414. It is important to note that although the training metrics continue to improve, overfitting may be the cause of the increasing validation loss in the later epochs. Overfitting happens when a model performs better on the data for training but stays away from generalising effectively onto new data. Therefore, since the TRBSL model is showing the signs of overfitting at epoch 3 with increased validation loss, the training is to be stopped at epoch 2 as the two epochs are sufficient for this dataset as we are getting accuracy and F1-score of 79%.

The results in Appendix A, Table A2 show the evaluation metrics for the TRBSL model. The performance of the model on the data used for validation is measured by the evaluation loss (eval_loss). Based on Appendix A, Table A2 the model is not performing well on the validation data with approximately 82.54% evaluation loss, this is a very high value for the eval_loss. The evaluation accuracy (eval_accuracy) and the evaluation F1-score (eval_f1_score) are both approximately 0.6822, suggesting that the model correctly classifies approximately 68.22% of the validation data and that there is a balance between the precision and the recall in the prediction of the model. Table 4.5 shows the classification report for the TRBSL model after the fine-tuning. The model accurately predicted approximately 57% of the negative sentiments, 68% of the neutral sentiments and correctly identifying approximately 87% of the positive sentiments. The F1-score for the negative, positive, and neutral class is approximately 0.66, 0.62, and 0.81, which indicates a reasonable balance between precision and recall. The overall accuracy of the model across all classes is 0.69, which means that approximately 69% of all predictions are correct. Overall, the model is performing best after fine-tuning.

	precision	**recall**	**f1-score**	**support**
0	0.79	0.57	0.66	345
1	0.57	0.68	0.62	303
2	0.76	0.87	0.81	252
accuracy			0.69	5000
macro avg	0.71	0.71	0.70	5000
weighted avg	0.71	0.69	0.69	5000

Table 4.5: Classification Report For TRBSL**

Using The Fine-tuned Model (TRBSL)**

The sentiment results of the fine-tuned model based on the new data predicted 3676, 880 and 444 count for neutral, negative and positive sentiment. Even with refinement, the model could not accurately classify all positive tweets, instead, it classified the majority of positive tweets as neutral, based on the sentiment analysis results. The results in Table 4.6 shows the summary statistics using the fine-tuned model (TRBSL**) to analyse the sentiment in the tweets which provides a basic understanding of the central tendency and spread of the values of the dataset. Firstly, the mean value of approximately 0.8062 represents the average value in the Twitter dataset being used for the analysis. Secondly, the median is approximately 0.8508, which indicates the middle value in the dataset when the values are sorted. The median is very useful as it provides the insights to the middle value of the overall dataset and is less affected by the outliers as compared to the mean. Lastly, the standard deviation is approximately 0.1545, which indicates that the data points are relatively close to the mean on average. In summary, the results suggest that there is no significant amount of variability in the dataset. Figure A2 in Appendix A demonstrate that there are more neutral sentiment tweets as compared to other sentiment classes, with ANC leading the neutral sentiment class with the highest count. The model does not perform well in predicting sentiments for positive tweets.

Statistic	Value
Mean	0.806183
Median	0.850807
Standard Deviation	0.154538

Table 4.6: Statistical Distribution - TRBSL**

Appendix A, Figure A10 shows the top 15 frequently used phrases for the ANC using the TRBSL** model. Figure b demonstrate that phrases such as "anc president cde", "local government elections" and "voteanc buildingbettercommunities anc" are dominant in the positive sentiment class. Furthermore, the negative sentiment in Figure a demonstrate phrases such as "leader julius malema", "dont vote anc" as the most frequently used phrases by the users. The top three most frequently used phrases in the negative sentiment trigram plot (Figure c) are the same as the top phrases in the positive sentiment trigram. Looking at the DA trigram for positive sentiment in Figure b in Figure A11 shows phrases such as "city ghlformayor voteda" and phrases related to "candidate dumezweni ngcamu" as the dominant phrases that are mostly used by the Twitter users. Furthermore, Figure a of the negative sentiment demonstrate phrases such as "idiots ur da", "wake idiots ur" and "voters wake idiots" as the most frequently used dominant words by the Twitter users. In contrast, the neutral sentiment trigram plot in Figure a demonstrate phrases such as "da mayoral candidate", "local government elections", and "leader john steenhuissen" as the mostly used words. The top 15 most frequently used words are shown in a the trigram plots in Figure A12 for each sentiment class. The phrases such as "landandjobsmanje voteeff eff", "councillor candidate fighter" and "ward councillor candidate" are amongst the top dominant phrases which are mostly used by the Twitter users for positive sentiment class in Figure b for the dataset used in this study. Figure b shows that dominant phrases in the negative sentiment class are mostly related to the party leader Julius Malema of the EFF. Furthermore, the neutral sentiment trigram (Figure b) for the frequently used phrases are words related to "vote eff". In the trigrams for the top 15 frequently used phrases plot in Figure A13, there seems to be predominately frequently dominant phrases in all the sentiment trigram frequency plots which suggest that the users are tweeting mostly about "voting ActionSA".

Moreover, the time series plot in Appendix A, Figure A14 shows the variation over time of the tweet counts for each political party. Figure a shows how sentiments of users vary over time and ultimately increase at end of the campaign trail in October 2021 for ANC in all sentiment classes. In contrast, the positive sentiment class seems to be constant between 1^{st} and 15^{th} September 2021. For political party DA, the neutral sentiment of users seems to fluctuate over time for most of the election campaign period, with the positive sentiment of users remaining constant throughout and a slight variation in the negative sentiment (Figure b). Figure b illustrates how user sentiment towards the political party EFF varies over time. Similarly, user sentiment towards positive political parties appears to be constant for a while, then slightly increases around September 15^{th}, and then returns to being constant and fluctuating at the end of September 2021. For the political party ACtionSA, the sentiment of users tends to be more neutral, with fluctuation over time (Figure d). Over time, there appear to be significant shifts in the neutral feeling towards 31^{st} October 2021, and a slight change in the negative sentiment. The overall positive attitude of users has not shifted all that much over time.

Benchmarking Against GPT-3.5 Model

The sentiment analysis results between the TRBSL** and GPT-3.5 models in Appendix A, Figure A15 suggest that both models disagree on sentiment counts. The sentiment counts for TRBSL** are distributed between negative, neutral and positive labels, with neutral sentiment for the TRBSL** model having more tweets count. This suggest that there is a significant variation between the models in all the sentiment classes. The results also indicate that the TRBSL** model did not perform well in analysing the tweets and labelling them as either positive or negative, as most of the tweets are seen as having

neutral sentiment. Furthermore, the results suggest that out of the 5000 tweets, the correctly classified tweets are 661 for positive, which is approximately 13.22% of 1667 positively classified tweets in GPT-3.5 model. In contrast, only 1174 tweets are correctly classified as negative and the other 3165 is classified as neutral. Figure 4.2 and Table 4.7 shows the confusion matrix and the classification report for the TRBSL** model. An estimated 50% of the negative predictions were correctly predicted by the model, 84% of the neutral predictions were accurately classified, and 35% of the positive predictions were also correctly classified. The F1-score for the negative, positive, and neutral class is 0.59, 0.58, and 0.51, which indicates a reasonable balance between precision and recall. The overall accuracy of the model across all classes is 0.56, which means that approximately 56% of all predictions are correct. Overall, the model is performing better on the test dataset. Further improvement on the model is recommended for better results.

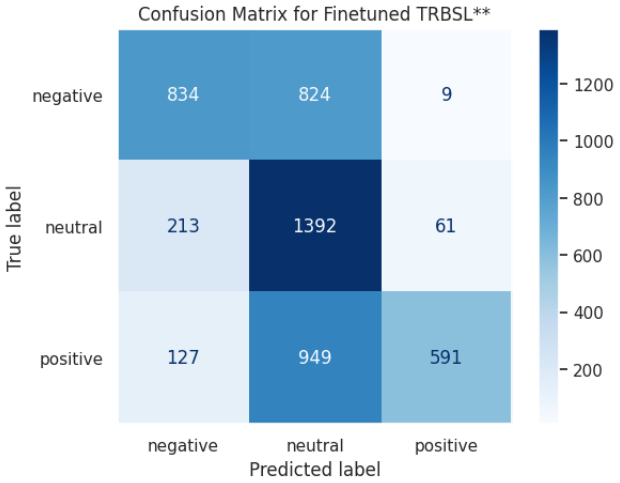


Figure 4.2: Confusion Matrix For GPT vs TRBSL**

	precision	recall	f1-score	support
0	0.71	0.50	0.59	1667
1	0.44	0.84	0.58	1666
2	0.89	0.35	0.51	1667
accuracy			0.56	5000
macro avg	0.68	0.56	0.56	5000
weighted avg	0.68	0.56	0.56	5000

Table 4.7: Classification Report For GPT vs TRBSL**

4.2.2 Valence Aware Dictionary for sEntiment Reasoner (VADER)

The VADER model is used for the sentiment analysis in this section. Table 4.8 shows the overall distribution of the sentiment labels and their count where positive labels have a count of 2103, negative with 1508 and neutral has a count of 1389. The positive class seems to have significantly more samples as compared with the minority classes negative and neutral.

Sentiment Label	Overall Count
neutral	1389
positive	2103
negative	1508

Table 4.8: Sentiment Label Count - VADER

Statistic	Score
Mean	0.053144
Median	0.000000
Standard Deviation	0..474259

Table 4.9: Statistical Distribution - VADER

Basic Statistical Analysis

Table 4.9 shows the scores of 0.0531, 0.0000 and 0..4743 for mean, median and standard deviation respectively. The mean score indicates that the sentiment is slightly positive on average, whereas the neutral median score suggests that there are an equal counts of negative, positive, and neutral sentiment scores, and the standard deviation suggests that the sentiment scores are relatively variable.

In summary, while the mean score is slightly positive, the sentiment scores in the data vary widely, with a neutral median score reflecting a mix of both positive, neutral, and negative sentiments suggesting that the performance of the VADER model is better. The standard deviation indicates that there is significant variability in sentiment intensities across the data.

Sentiment Distribution Analysis And Results

The polarity sentiment distribution amongst the political parties in Appendix B Figure B1 shows that most of the positive sentiments are tweets about ANC, followed by EFF, DA and lastly ActionSA, which suggests that there is a variation across the political parties. There is a very slight variation between the sentiment count for EFF and DA with negative sentiment roughly the same. The VADER model has managed to label the tweets into different sentiment labels, with most tweets labelled as positive over 2000 and the other two sentiments above 1000 counts. It can also be seen from the distribution that the most negative sentiments and neutral sentiments are tweets about ANC. Tweets about the ANC seem to be dominating more than tweets about the other three political parties. Most users are tweeting about ANC, this could be because ANC is the current governing party and the biggest political party in South Africa. It should be noted that although most users seems to be unhappy about the ANC, there are still some users who are either happy or uncertain about their stands. This suggests that the ANC may need to do better in campaigning for the local elections to keep the people happy and on their side. There is a slight variation between the EFF and the DA across all the sentiment distributions, while ActionSA has less tweets counts across all the sentiment distributions, with more being positive tweets and neutral, whereas negative tweets are slightly less.

Figure B2 in Appendix B shows the top 15 most frequently used phrases for political party ANC with different sentiments. For positive sentiment in Figure b, it can be seem that most dominant phrases are that which are related to those in the neutral sentiment (Figure c) such as “local government election”, and phrases that contains “building better communities”. In contrast, for the negative sentiment trigram frequency in Figure a, it can be observed that phrases such as “vote wisely sa”, “lying looting thieving”, and “corrupt lying looting” are at the top of the most frequently used phrases, which suggest that the political party ANC is seen as a “dishonest” party amongst other things. For the DA political party, the positive sentiment in Figure B3 in Appendix B shows “local government elections” and “leader john steenhuisen” as the top most used phrases in the positive trigram frequency plot (Figure b). The negative trigram plot (Figure a) for the DA suggest that the political party is seen by the Twitter users as “racist” and “killing black people” as the most top used phrases in the trigram plot are “daisracist daisracist daisracist” and “king black people”. In contrast, the neutral sentiment in Figure c shows the top most used phrases as “voetsekramaphosa nourda voetsekda” and “ramassacre voetsekramaphosa nourda”. The trigram frequency plot for the EFF in Figure B4 in Appendix B suggest phrases such as “local government elections”, “ward councillor candidate” and “eff cape metro” as the most dominant phrases that are being refered to in the positive class(Figure b). In contrast to the positive wordcloud, the negative sentiment trigram plot in Figure a shows phrases such as “eff vote eff”, “vote eff vote”, “leader julius malema” as the most top used phrases in the negative class. Similarly to the positive sentiment class, the neutral sentiment of EFF has dominant phrases which are related to “local government elections” and the “candidates”. The trigram frequency plot for ActionSA in Figure B4 suggest that the users are mostly tweeting about voting for ActionSA, whereas the most dominant phrases in the positive sentiment trigram plot (Figure b) are ralated to “voteaction”. The negative trigram frequency plot in Figure a suggest that the users are mostly tweeting about selecting ActionSA on the ballot papers with top phrases such as “ballot papers actionsa” and those related to ActionSA mayoral candidate. Overall, the trigram sug-

gest that the most negatively impacted parties amongst the four political parties are the ANC and the DA.

The sentiment expressed by Twitter users during the 2021 local government elections campaign period evolved over time as shown in Appendix B, Figure B6. It can be seen that with the VADER model, the tweets regarding ANC are evolving over time (Figure B6a), in contrast to the tweets regarding DA which shows slight variations over time across all the polarity sentiments (Figure B6b). Furthermore, it can be observed in Figure B6c that there is a significant variations across the sentiment distribution overtime for EFF, which shows an increase in positive sentiment around 22^{nd} September and 1^{st} October 2021, then through to the last day of election campaigns. The sentiment time series plot for Action SA shows a minimal variation across the sentiment labels between 22^{nd} October and 31^{st} November 2021 (Figure B6d), with most of the variations due to positive tweets.

In summary, the results of the VADER model suggest that there is a variation in polarity sentiment across all the political parties. Therefore, there are significant distinct sentiment shifts during different phases of the 2021 local government elections campaigns overtime across all the political parties which suggests that overtime, people's sentiment changes.

Benchmark Against the GPT-3.5 model

In this section, the evaluation metrics were used to benchmark the VADER model against the GPT-3.5 model. The sentiment analysis results between the VADER and GPT-3.5 models in Figure B7 in Appendix B suggest that both the models slightly disagree on sentiment counts, i.e. there is a variation between the two models. The sentiment counts for VADER is distributed amongst negative, neutral and positive sentiment. The confusion matrix in Figure 4.3 shows the 3x3 matrix that represents the classification results for VADER multi classification. Each cell in the matrix demonstrates various aspect of the performance of the VADER model. The model correctly predicted 906 instances as negative (class 0), and incorrectly predicted 204 instances as a neutral class (class 1) and 557 instances as positive (class 2). The model also correctly predicted 460 instances as neutral, and incorrectly predicted 451 instances as negative and 755 instances as positive instead of neutral. This indicates that the model is not accurately able to correctly classify the neutral class, but instead predict most of the neutral instances as positive. Furthermore, the model correctly predicted 990 instances as positive and incorrectly predicted 153 as negative and 524 as neutral. The results suggest that they model can accurately classify the negative and positive sentiments.

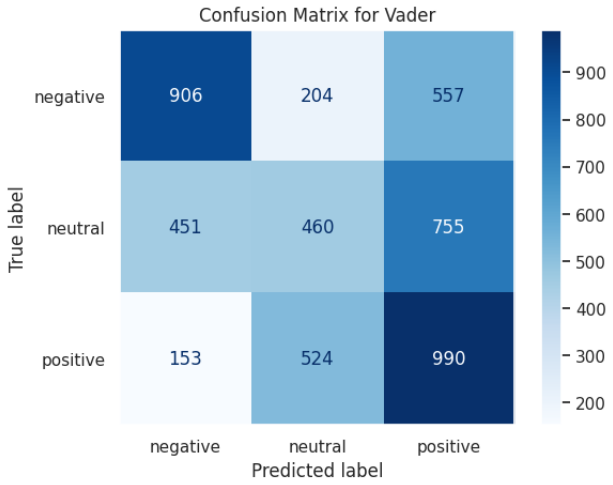


Figure 4.3: Confusion Matrix For GPT vs VADER

The classification report in Table 4.10 shows the various metrics to evaluate the performance of a multi-class classification for the VADER model. The precision of 0.60 for the negative label (class 0) indicates that approximately 60% of instances are correctly predicted as negative, whereas approximately 39% are

	precision	recall	f1-score	support
0	0.60	0.54	0.57	1667
1	0.39	0.28	0.32	1666
2	0.43	0.59	0.50	1667
accuracy			0.47	5000
macro avg	0.47	0.47	0.46	5000
weighted avg	0.47	0.47	0.46	5000

Table 4.10: Classification Report For VADER

correctly predicted as neutral (class 1) and 43% are correctly predicted as positive (class 2). The recall of 0.54, 0.28 and 0.59 indicates that the VADER model correctly predicted approximately 54%, 28%, and 59% of the instances for negative, neutral, and positive respectively. This suggest that the model is failing to predict the neutral instances correctly. The F1-score of 0.57, 0.32 and 0.50 for negative, neutral, and positive labels respectively, indicates a reasonable balance between precision and recall. The overall accuracy of the model across all classes is 0.47, which means that approximately 47% of all predictions are correct. The macro-averaged and the weighted average precision, recall, and F1-score are between 0.46 and 0.47, indicating that the model performs best for negative and positive labels and not well for the neutral labels.

In summary, the report in Table 4.10 indicates that the performance of the model varies across all the classes, with Class 0 and Class 2 having relatively better performance compared to Class 1. This indicates that in order to get better results for the model, more improvements is required on the model. Due to the computational powers, the model is not further improved for this study.

4.2.3 TextBlob

The TextBlob model is used in this section to perform sentiment analysis. Table 4.11 shows the overall distribution of the sentiment labels and their count where positive sentiment have a count of 1536, negative with 1028 and neutral have a count of 2436. It can be clearly seen that there is a class imbalance in the distribution.

Sentiment Label	Overall Count
neutral	2436
positive	1536
negative	1028

Table 4.11: Sentiment Label Count - TEXTBLOB

Statistic	Score
Mean	0.035006
Median	0.000000
Standard Deviation	0.268955

Table 4.12: Statistical Distribution - TEXTBLOB

Basic Statistical Analysis

The mean score of 0.0350 in Table 4.12 suggests a slightly positive sentiment on average, whereas there is a neutral median score of 0.0000 which suggests a balanced distribution of positive, neutral, and negative sentiment on average and the standard deviation of 0.2690, suggesting that the sentiment scores are not highly variable and tend to be clustered around the mean. Overall, the data exhibits a slightly neutral sentiment on average, with a neutral median sentiment and the sentiment scores are relatively consistent and not highly variable.

Sentiment Distribution Analysis And Results

The sentiment distribution in Figure C1 in Appendix C suggests that most of the tweet samples are neutral, followed by negative with a slight variance between negative and positive. There is a significant variation across all the political parties during the 2021 South African local elections campaign period. It can be seen that the ANC is the most dominant party in all the sentiment distributions, having the most neutral, positive and negative sentiments. There is not much difference between the DA and EFF across the positive sentiment and the negative sentiment, whereas the neutral sentiment seems to be significantly different. ActionSA seems not to have much significant variation across the different polarity sentiments. It must be noted that ActionSA was a new political party during the 2021 local election, which was formed in August 2020. The insights of the sentiment distribution suggest that, overall, ANC is still dominating in the local government elections.

The trigram plot in Appendix C, Figure C2 show the top 15 dominant phrases across the different sentiment labels using TextBlob model sentiment results for ANC. Figure C2b suggest that phrases such as "local government elections", "campaign trail anc", "building better communities" are amongst the most dominant phrases in the positive sentiment frequency. Similarly, the top phrase in the neutral trigram is the same (Figure C2 c). The negative trigram for the ANC in Figure C2a suggest that the governing party is seen as "lying looting thieving", "corrupt lying looting", and users are mostly encouraged to "vote wisely sa". The results of the trigram frequency plot for DA in Appendix C, Figure C3 have phrases such as "local government elections", "leader john steenhuisen", "da learder john" amongst the dominant phrases in the positive sentiment trigram (Figure b). In contrast, the trigram for neutral sentiment (Figure c) has phrases "daisracist daisracist daisracist", "federal council chairperson", "da things voteda" as dominant phrases. Furthermore, the negative sentiment suggest phrases such as "killing black people", "da different voteda", "voetsekramaphosa call", as dominant phrases(Figure a). Figure C4b suggest phrases such as "economic freedom fighters", "local government elections", "social media platforms" as the most frequent phrases used by users in the positive sentiment trigram plot, similary to the neutral frequency (Figure c) where phrases such as "eff leader julius" appears to be the most dominant. In the negative trigram frequency (Figure a) for EFF, it can be seen that phrases such as "eff vote eff" tops the chart as the most frequent phrases which suggest that most users are being encouraged to vote for EFF even in the negative sentiment trigram. On the trigram frequency plot for ActionSA in Figure C5, the most frequent phrases for positive sentiment (Figure b) is "voteactionsa actionsa voteactionsa" which means users are mostly encouraging each other to vote for ActionSA. On contrary, the neutral sentiment trigram in Figure c for the top frequently used phrases is more related to the mayoral candidates for the Tshwane and Joburg municipal elections. Furthermore, in Figure a, the most negative sentiment phrases are found to be having phrases such as "ballot papers actionsa" and "mashaba stands views" as the top phrases. Overall, the trigram suggest that the most negatively impacted parties amongst the four political parties are the ANC and the DA.

Appendix C, Figure C6 suggest that the sentiment expressed by the Twitter users during the 2021 local government elections campaign period evolved over time. For ANC, Figure C6a suggest that there is a significant variation overtime for the sentiment expressed, with the increase of negative sentiments during the last push for campaigning in October 2021. There is a significant variation between the positive and neutral sentiments for DA overtime, whereas neutral tweets seem to be standard for some period of time (Figure C6b). The EFF sentiment overtime Figure C6c suggests a significant variation over time across all the sentiment labels. The sentiment for EFF overtime suggests an increase from 22^{nd} September to 25^{th} September 2021 in both positive and neutral sentiments. ActionSA time series plot in Figure C6d suggest that there is no there is no obvious shift in the trend as the pattern appears to slowly increase and decrease until an increase in positive and neutral sentiment between 22^{nd} and 31^{st} October 2021.

In summary, the results of the TextBlob model suggest that there is a variation in polarity sentiment across all the political parties and also distinct sentiment patterns or shifts during different phases of the 2021 local government elections campaigns.

Benchmark Against the GPT-3.5 model

In this section, the evaluation metrics were used to benchmark the TextBlob model against the GPT-3.5 model. Figure C7 in Appendix C shows the sentiment analysis results between the TextBlob and GPT-3.5 model which suggest that the models slightly disagree on sentiment counts, i.e. there is a slight variation between the two models. The sentiment counts suggest that the TextBlob model has high count on neutral class, meaning that the model predicted more tweets as neutral with approximately 42.56%. Furthermore, there is a significant variation between the count for negative sentiment for both models, with TextBlob predicting slightly below (approximately 21.58%) the GPT-3.5 model. There is not much variation on positive sentiment, though it must be noted that the TextBlob model predicted more tweets than GPT-3.5 model as having positive sentiment (approximately 35.86%).

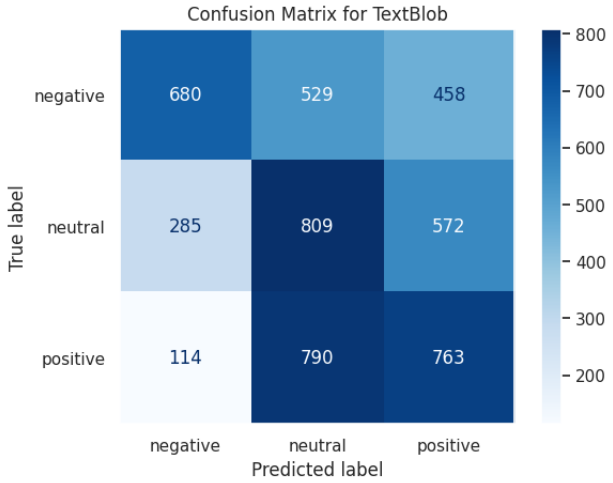


Figure 4.4: Confusion Matrix For GPT vs TextBlob

	precision	recall	f1-score	support
0	0.63	0.41	0.50	1667
1	0.38	0.49	0.43	1666
2	0.43	0.46	0.44	1667
accuracy			0.45	5000
macro avg	0.48	0.45	0.45	5000
weighted avg	0.48	0.45	0.45	5000

Table 4.13: Classification Report For TextBlob

Figure 4.4 shows the confusion matrix of the TextBlob model sentiment results against the GPT-3.5 model. The results for negative sentiment (Class 0), show that the model correctly predicted 680 instances as class 0, incorrectly predicted 529 instances of class 0 as class 1 and 458 instances of class 0 as class 2. For neutral sentiment (class 1), the model correctly predicted 809 instances as neutral, incorrectly predicted 285 instances of class 1 as negative and 572 instances of class 1 as positive. Lastly, for positive sentiment (class 2), the model correctly predicted 763 instances as positive, incorrectly predicted 114 instances of class 2 as negative and 790 instances of class 2 as neutral. This indicates that the model is failing to predict some of the tweets, therefore labelling them as neutral for the positive sentiment. Table 4.13 shows the classification report with evaluation metrics used to evaluate the performance of a multi-class classification for TextBlob model. The results of the precision show that approximately 0.63%, 0.38% and 0.43% of the instances are predicted correctly for negative, neutral and positive labels respectively. For the recall, the model correctly identifies approximately 41%, 49% and 46% for

classes 0, 1 and 2 respectively. The F1-score for negative labels is approximately 50%, which indicates a reasonable balance between precision and recall, whereas for neutral and positive labels, the F1-score is approximately 43% and 44% respectively, indicating a moderate balance between precision and recall. The overall accuracy of the model is 0.45, which means that approximately 45% of all predictions are correct across all classes. The macro-averaged and the weighted-average precision, recall, and F1-score are all between 0.45 and 0.48.

In summary, the model has low precision, recall, and F1-score for all three classes, with a 45% average accuracy. This indicates that in order to get better results for the model, there is a need for model improvement. Due to the computational powers and the time constraints, for this study, the model will not be improved.

4.3 User Classification

This section provides the analysis and findings for the user classification using suspicious patterns and K-Means methods. The results of this section address the last research question 3 (What are the different classifications of Twitter users across the four political parties?) described in Chapter 1, Section 1.1.

4.3.1 User Classification Analysis

The tweets were analysed to determine if they were generated by humans or bots. In order to identify the user based on the tweet, two user classes were first created, i.e., human and bot user classes. The following points elaborate on the two approaches that were used to assist in producing results.

- **Suspicious Patterns**: Firstly, suspicious patterns were used to identify the user class and also the irrelevant tweets in the dataset. The results based on the user classification analysis distribution using suspicious patterns suggest that out of the 5000 tweets which were analysed, 4975 tweets are human generated, 21 are generated by a bot and 4 were found to be irrelevant (Table‘4.14). All 21 tweets generated by a bot were verified manually. A sample dataset is illustrated in Table 4.15. The user classification results are stored and in a new column named “user_class” in the dataset.

user_class	**count**
human	4975
bot	21
irrelevant	4

Table 4.14: User Classification Using Suspicious Patterns

- **K-Means**: The study further applies K-Means clustering model based on the results found in using suspicious patterns. The appropriate optimal number of clusters for this K-Means clustering task is three (3) based on the elbow point, as illustrated by Figure 4.5. The K-Means clustering model is created using the three (3) clusters, and the same configuration used during the Elbow method analysis are also applied here. Cluster 0 has 4451 tweets, with cluster 1 having 68 tweets, and cluster 2 having 480 tweets. The cluster labels assigned by K-Means are extracted from the model and added to the dataset in a column named “user_class”, i.e., in this case only 68 tweets that are in cluster 1 are manually relabelled as bots. These are some of the tweets that were not discovered using the suspicious patterns approach. Appendix D, Figure D1 illustrate the information of the three (3) clusters, showing the cluster centroids visualised in the reduced 2D space. This provides a sense of the content and characteristics of the tweets in each cluster. The Principal Component Analysis (PCA) technique is used for the dimensionality reduction to transform the tweets into a lower-dimensional space before visualizing it.

index	tweet	user_class	manual_verification
0	Before the next local government elections, South Africans have one more chance to register to vote which is today. #SignUpToVoteDA To verify your status, simply go to https://########## right now.	bot	bot
1	Every voice must be heard! This weekend register to vote in the local government elections. Visit https://########## for more information on #LGE2021 #VoteSafe #EveryVoiceTogether. https://##########	bot	bot
2	Do you need help signing up to vote? We have a staff ready to help you with any questions you may have about registering. Visit https://########## to have a live online conversation with us. Call *** *** ****. #RegisterToVoteDA https://##########	bot	bot

Table 4.15: Manually Verified Bot-Generated Tweets Sample Dataset

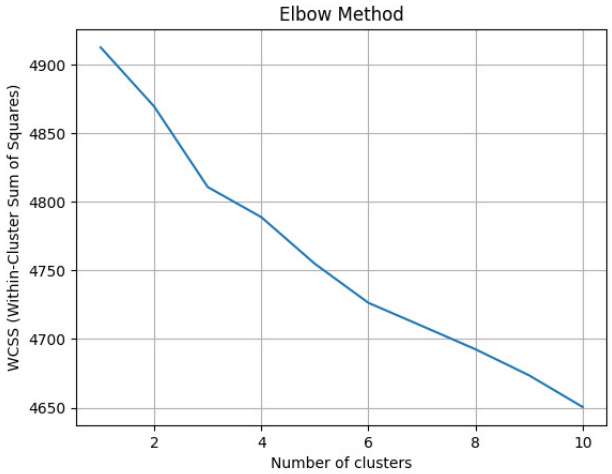


Figure 4.5: Elbow Point For Choosing Optimal Number Of Clusters

4.3.2 User Classification Results

Figure 4.6 illustrate the distribution of the final K-Means user classification results for all the tweets in the dataset (dataset of 5000 sample). There are 4,906 occurrences of tweets that are associated with real human users, 90 which are associated with bots or automated accounts and 4 users with tweets which are neither human nor bot generated, therefore irrelevant to the analysis. It appears that most of the tweets are categorized as human with a smaller number categorized as bot, and a very small number as irrelevant. The irrelevant classes are excluded from the discussion carried out in the study. To further understand the classification of the users, Figure 4.7 displays the classification of users into different categories, such as ”bot”, ”human”, and “norelevant” along with the counts of tweets falling into political party categories such as “DA”, “EFF”, “ANC” and “ActionSA”. There are about 71 users with tweets which are classified as bot for ANC, one (1) for DA and ActionSA respectively. Out of 4906 tweets generated by human users, 2125 are about ANC, 1013 are about EFF, 572 are about DA and lastly 388 are about ActionSA. There is only 1 tweet classified as irrelevant which is related to EFF. Table 4.16 illustrate the first rows sample of the human labelled tweets.

The sentiment distribution by user class for TRBSL in Appendix D, Figure D2 illustrate that 69, 811 and 0 occurrences of tweets are labelled as negative within the “bot”, “human” and “irrelevant” user class

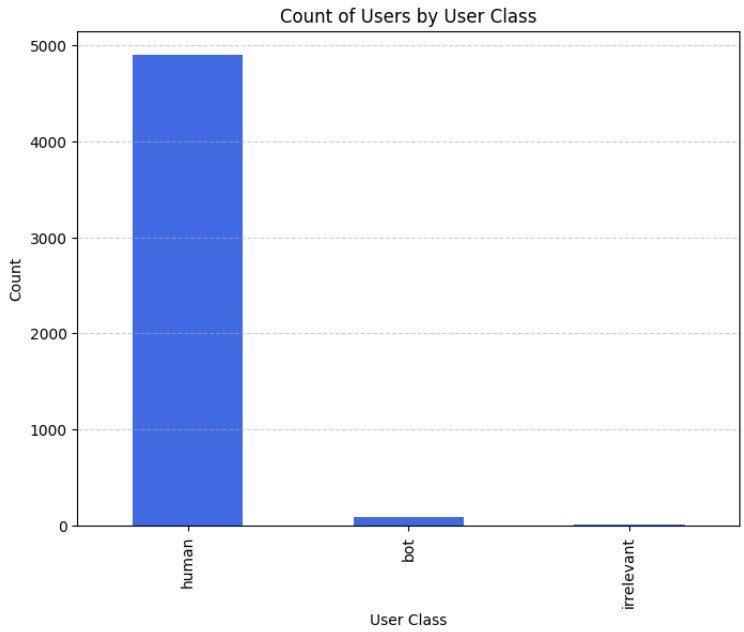


Figure 4.6: User Classification Count

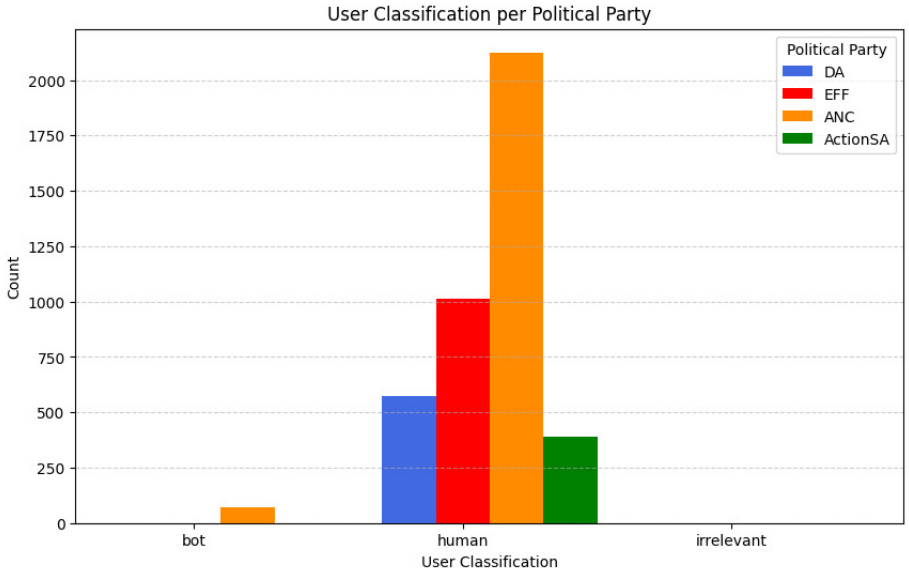


Figure 4.7: User Classification Distribution Per Political Party

categories respectively. For neutral sentiment, 20, 3652 and 4 occurrences of tweets within “bot”, “human” and “irrelevant” user class categories respectively. Lastly, in the positive sentiment class, only one (1), 443, and 0 are labelled as positive within the “bot”, “human” and “irrelevant” user class categories respectively. This suggests that out of all the users with tweets classified, they are only three (4) tweets which were found to be irrelevant according to the TRBSL** model, with 90 labelled as generated by a human and the rest of the tweets by human.The polarity sentiment (positive, neutral, and negative) breakdown for each political party (“DA”, “EEF”, “ANC”, “ActionSA”) suggest that there are 8, 55, 172, and 33 generated tweets by actual human and no tweet by a bot for the positive sentiment as predicted by TRBSL** model. For neutral, there are 449, 741, 1468 and 303 by human respectively, and 1, 0, 2, 1 by a bot respectively. For negative, there are 115, 217, 485 and 52 tweets generated by human respectively, and no tweet is generated by a bot in the negative sentiment class, as all are neutral.

The results for VADER in Appendix D, Figure D3 illustrate the distribution of sentiments across different user classes which is used to assess how well the model is performing for different user classes.

index	tweet	user_class	manual_verification
0	Together with Open Cities Lab, OpenUp created this helpful website for #LGE2021.	human	human
1	“Following the elections on November 1st, user, the independent candidate for Ward 15 councillor in #LGE2021 said they are prepared to address the corruption that has resulted in the shortage of water delivery in Lulekani IMS.”	human	human
2	”#LGE2021 the posters were placed on top of each other and they read as follows: “The #DA labels you as notable individuals and the ANC labels you as racists.”	human	human

Table 4.16: User Classification Human Tweets Sample Dataset

Firstly, the bot user class, has 72, 7 and 11 occurrences of the tweets are labelled as negative, neutral and positive respectively for bot users. Secondly, the user class for human has 1436, 1379 and 2091 occurrences of tweets labelled as negative, neutral, and positive respectively. Lastly, the tweets occurrences labelled as neutral and positive for the irrelevant class are 3 and 1 each. The irrelevant class is excluded for the analysis as it contains spammy tweets which are tweets not relevant to local elections in South Africa. Furthermore, this results indicate that the human class has many predictions in all three label, while the bot class has a smaller number of predictions. The polarity sentiment (positive, neutral, and negative) breakdown for each political party (“DA”, “EEF”, “ANC”, “ActionSA”) suggest that the distribution of the sentiment among user classes varies. For political parties “DA”, “EEF”, “ANC”, “ActionSA” there are 254, 454, 771, and 176 generated tweets by actual human and 1, 2, 0, and 1 by a bot as predicted by VADER model. In contrast, neutral sentiment has 98, 267, 532 and 119 tweets generated by human respectively, and no tweet is found to be generated by a bot. Furthermore, negative sentiment results using VADER model for the different user classes has least amount of tweets, i.e., 220, 292, 822 and 93 by human respectively, and 69 tweets for ANC which are bot generated and also negative.

The results for TextBlob in Appendix D, Figure D4 illustrate the distribution of sentiments across different user classes which is used to assess how well the model is performing for different user classes. Firstly, the bot user class, has 72, 9 and 9 occurrences of the tweets labelled as negative, neutral, and positive respectively. Secondly, the user class for human has 956, 2423 and 1527 occurrences of tweets labelled as negative, neutral, and positive respectively. Lastly, there are only 4 tweet occurrences labelled as neutral for the irrelevant class. The “irrelevant class” is excluded for the analysis as it contains spammy data. The user class results suggest that most of the predicted tweets are generated by human users, while the bot class has a smaller number of tweets predictions. Based on the polarity sentiment (positive, neutral, and negative) breakdown for each political party (“DA”, “EEF”, “ANC”, “ActionSA”) there are 175, 330, 605, and 124 generated tweets by actual human and 1, 0, 1, and 0 by a bot respectively, as predicted by the TextBlob model. Moreover, the neutral sentiment class has 250, 506, 976 and 196 tweets generated by human respectively, and only a single tweet generated by a bot for ActionSA. For negative sentiment class, there are 151, 177, 544 and 68 tweets generated by actual human respectively, and 70 tweets for ANC generated by a bot.

4.4 Summary

Although the Twitter-roberta-base-sentiment-latest (TRBSL) and TRBSL** models have a high average which suggest that most of the samples are having sentiment scores close to the mean and the relatively small standard deviation indicates that the sentiment scores are consistent and not highly variable, the

models have higher weighted accuracy. There seems to be an improvement in the TRBSL after fine-tuning using the OpenAI labelled data but the mean and the standard deviation shows a slight variation from the initial one. The VADER and TextBlob models seem to be performing much better, though there is evidence of class imbalance in the labelling of the tweets. For all the models, class weight is applied to address the issue of class imbalance in the dataset. All the models are benchmarked against the GPT-3.5 model in order to address the class imbalances in the dataset using the class weight. According to the results of SA, there is a significant variation in polarity sentiment between the four political parties throughout the 2021 South African municipal election campaign period, based on all of the models examined. Moreover, the findings indicate that there is a significant difference in the feelings expressed by Twitter users during the 2021 local government elections campaign. Overall, most tweets were found to be relevant and generated by human. The party with the most negative tweets generated by a bot was ANC, followed by DA. No bot generated tweets were found for EFF across all the applied methods. Most of the tweets generated by human users were neutral for TRSBL** and TextBlob, whereas for VADER, most tweets were positive. This indicates that the various user classes vary across the four political parties and further demonstrates that the majority of bots are employed for malicious purposes. The next chapter discusses the results, how they compare with other studies and the limitations faced in this study.

Chapter 5

Discussion

This section discusses the findings, how they compare to earlier studies, and the limitations.

5.1 Results

The sentiment analysis results of the three models show a significant difference as per the analysis done in the Analysis & Results section. The results in Table 5.1 show that TRBSL and TRBSL** appears to have the most neutral labels, with relatively high mean and median sentiment scores. VADER results suggest that sentiment is slightly positive on average and has the highest variability in sentiment scores among the models. TextBlob leans towards neutral and slightly positive sentiments, which suggest the model has moderate variability in its sentiment scores. The initial sentiment analysis results based on the statistical distributions suggest that the VADER and TextBlob models as compared to the TRBSL and TRBSL** models are able to achieve good performance when classifying tweets into negative, neutral, and positive. Furthermore, the results in Appendix E, Figures E1(a-d) demonstrate the trend of sentiment (polarity) for tweets over time, with a focus on changes in sentiment. The dashed line is significant in sentiment analysis as it represents the neutral sentiment point, where values above the line are positive sentiment, values below are negative sentiment, and values parallel to the line are considered neutral. The time series plot shows significant variation on sentiment over time for both VADER and TextBlob.

Model	**Mean**	**Median**	**Standard Deviation**
TRBSL	0.7705	0.8030	0.1320
TRBSL**	0.8061	0.8508	0.1550
VADER	0.05314	0.0000	0.4743
TextBlob	0.0350	0.0000	0.2690

Table 5.1: Overall Statistical Distribution For Sentiment Analysis

Model	**Weighted Accuracy**	**Weighted F1-score**
TRBSL	0.51	0.48
TRBSL**	0.56	0.56
VADER	0.47	0.46
TextBlob	0.45	0.45

Table 5.2: Weighted Accuracy And F1-score For All Models

The results for benchmarking the models with the GPT model (Table 5.2), suggest that the standard TRBSL model has relatively high weighted accuracy and F1-score. This indicates that the model is not performing bad in classifying the tweets, both in terms of overall accuracy and precision-recall balance (as indicated by the weighted F1-score). The model performed well in classifying the negative tweets. Furthermore, the TRBSL**, which is the improved or modified version of the standard TRBSL model

(original TRBSL model) exhibits significantly better performance compared to the standard TRBSL. The model has high weighted accuracy and F1-score, indicating that it is proficient in correctly classifying the tweet data and achieving a good balance between precision and recall. The VADER model achieves moderate accuracy and F1-score. The model shows better results than the standard TRBSL and TRBSL** models when comparing the statistical distribution. In contrast, the model is not performing better than the standard TRBSL and TRBSL** models when comparing the accuracy and precision-recall balance scores. The VADER model performed well in classifying tweets as negative and positive but failed to classify neutral sentiment tweets. The performance of the TextBlob model is almost similar to that of VADER. It has moderate accuracy and a reasonable F1-score, indicating that it is decent at text classification and performed well in predicting the sentiments. Overall, the VADER model is good in classifying the negative and positive tweets as compared to the other three models, although it fails to correctly classify tweets which are neutral. The findings from the four models also show that the polarity sentiment of each of the four political parties varies, and that user sentiment varies with time. Additionally, the identification of two user classes revealed that user classifications varied across all parties and that the majority of the negative tweets directed at the ANC were generated by a bot.

5.2 Comparison With Previous Work

Sentiment analysis have become popular in political campaigns over the years where political parties uses social media data to analyse the sentiment of the users to use the results to put some corrective measures. This study addresses some of the limitations identified from other studies which were highlighted in Chapter 2 Literature Review section 2.5 "Key Gaps In The Literature". Although many studies highlighted that a large dataset is required to get deeper understanding of the model and better results, this study only uses limited data due to the limitation of computational power, time constraints and also the fact that the study uses unsupervised dataset. Furthermore, the study addresses the limitation of not including neutral sentiment label in identification and model training. To address the limitation of only using positive and negative sentiment labels, this study introduces neutral sentiment label to the sentiment analysis of the tweets. This enables tweets that are neither positive nor negative to be identified, in contrast to previous studies, particularly the work by Ledwaba and Vukosi [2]. According to the findings of Ledwaba and Vukosi [2], the ANC party was the target of the majority of the negative sentiments expressed in the tweets. "The ANC, which is the ruling party, received the worst results in the election which was below 50%" [2]. In agreement with the study of Ledwaba and Vukosi [2], the findings of this study demonstrate that the majority of tweets are negative, followed by neutral, and finally positive.Additionally, this study discovered that the majority of negative tweets were about ANC after examining the outcomes of the three models.

Moreover, in this study an unsupervised sentiment analysis is explored using pretrained techniques such as VADER and TextBlob, together with the hugging face model TRBSL, contrary to the study of Ledwaba and Vukosi [2] which employed a single model to analyse the same dataset. The TRBSL and TRBSL** models demonstrates that the majority of tweets have neutral sentiment, followed by negative sentiment, and finally positive sentiment with relatively few tweets. According to the results of the model, there are a lot of tweets with neutral sentiment, and the majority of them are regarding the ANC, subsequently followed by the EFF, DA, and ActionSA. It is worth noting that negative sentiment label has the second highest count of tweets, with the highest count of tweets regarding ANC, followed by EFF, DA and lastly ActionSA with fewer negative tweets. On the contrary hand, the general sentiment count of the VADER model indicates that the majority of tweets are positive, followed by negative and neutral. The sentiment results of this model indicate that the ANC is the political party with the highest percentage of positive, negative, and neutral sentiment. The EFF, DA, and ActionSA follow respectively. Furthermore, using the TextBlob model, the results suggest that the majority of tweets are neutral, followed by positive and negative attitudes. The political party with the most neutral, positive, and negative tweets is the ANC, same as TRBSL and TRBSL** models. In agreement with the findings of the study of Ledwaba and Vukosi [2], the overall results of this study indicate that the ANC

is the political party that received the most tweets, with the greatest count across all sentiment distributions. Additionally, this study employed the GPT-3.5 OpenAI model to classify the unsupervised dataset, which sets it apart from previous studies. Labelling datasets may be difficult, time consuming, and costly. Despite its own set of challenges, the GPT-3.5 model made the labeling of the dataset easier using few-shot learning.

Finally, this study uses the suspicious pattern technique and the K-Means algorithm to categorise Twitter users based on the dataset as either real humans or bots, which is different methods from previous studies. Bots are social media accounts that are automated, they have been demonstrated to propagate false information and control online conversations [44].The results shows that the selected methods perform well in identifying spammy tweets as the results were manually verified.

5.3 Limitations

There are several limitations that were faced while conducting the analysis of this time which limited the progress and consumed a lot of time.

- One of the main drawbacks in this study is the use of the unsupervised dataset that has lots of class imbalanced. Using class weights failed to provide the desired enhanced outcomes.
- Another issue faced is that the dataset used for this study was collected in the 2021 South African local election campaign period, which is now outdated as the country is currently undergoing another election period which will be held in 2024.
- In this study, the TRBSL model is the only model that is fine-tuned due to the computational power issues and the size of the labelled data which was labelled using OpenAI. Fine-tuning requires a lot of time and computational powers. It must be noted that the TRBSL model requires high computational power, GPU, and a large time to fine-tune the model. Labeling unsupervised data is time consuming and costly, hence OpenAI model GPT-3 is used for this. Using OpenAI GPT-3 model to label the dataset has its own challenges. The model requires high computational power, it is time consuming and requires a set of examples for it to understand the requirements and if this is not set correctly, the model may not provide the correct results. Some of the issues faced with using GPT-3 model included the following errors below due to the model taking time to run:
  - Errors associated with Request timed out.
  - Errors associated with bad gateway.
  - Errors resulting from the server being overloaded or not yet ready.
- Due to cost and time constraints, the sentiment labels for the tweets were personally verified. Only 1000 tweets were confirmed, and there was no assistance from external persons or political experts to verify the tweet sentiments.

5.4 Summary

In summary, the TRBSL model performs better in this classification task, with high accuracy and F1-score, whereas the TRBSL** model is the best-performing model, with higher accuracy and F1-score compared to the latter, making it the most effective classifier among the models provided. The VADER and TextBlob models have similar, moderate-level performance, with accuracy and F1-scores falling below the performance of TRBSL and TRBSL**. Since high accuracy and precision-recall balance are crucial evaluation metrics for the analysis of this study, TRBSL** seems to be the favourable option, but the best option will be using VADER followed by TextBlob as the models were able to classify tweets for negative and positive accurately compared to both TRBSL and TRBSL**. The choice of which model to use depends on the specific requirements and objectives of the classification task [8]. However, the choice also depends on other factors like computational cost and the specific nature of the text data

being classified. TRBSL** requires a lot of time and high computational power to fine-tune the model. In contrast, without any adjustments, the TextBlob and VADER models work sufficiently well to get better sentiment predictions. The next chapter present the conclusion of this study, together with the implications, future work and recommendations.

Chapter 6

Conclusion, Implications, Future Work and Recommendations

This section summarises the general result of this study, the implications based on this study, future work, and recommendations.

6.1 Conclusion

This study set out to explore the concerns regarding the influence of bots on election campaigns through social media, emphasizing the need to identify genuine users and understand their behaviours. Political parties must understand public feelings to be able to strategically execute their election campaigns according to the expectations of the community. The primary goal of this study was on sentiment analysis, classifying tweets as human or bot-generated, and providing insights for political parties. The analysis explored the changing sentiment over time and the polarity variation for different political parties. The results revealed that, in terms of both average and standard deviations, VADER (mean = 0.0964, std_dev = 0.4848) and TextBlob (mean = 0.0460, std_dev = 0.2578) models outperformed TRBSL (mean = 0.7629, std_dev = 0.1333) and TRBSL** (mean = 0.7976, std_dev = 0.1551) models in the comparison of the four sentiment analysis models. Furthermore, the study compared four sentiment analysis models on weighted accuracy and F1-scores, revealing that TRBSL (accuracy = 0.51, F1-score = 0.48) and TRBSL** (accuracy = 0.56, F1-score = 0.56) outperformed VADER (accuracy = 0.46, F1-score = 0.45) and TextBlob (accuracy = 0.44, F1-score = 0.45), indicating room for improvement for all models. Based on these results, this study was able to answer the research questions relating to the variation in polarity sentiment and change over time of sentiments, which indicated that there was a variation in polarity sentiment for all the political parties and also the variation in the sentiments expressed by users over time. According to the results of the models, the political party with the most negative tweets overall was the ANC, followed by the EFF, then DA, and finally ActionSA. Despite dissatisfaction by users regarding the ANC, there were also positive and neutral sentiments expressed. Furthermore, it emerged that there were more tweets generated by humans than by bots. According to the findings, 2106 tweets regarding the ANC political party were generated by humans. However, 79 tweets with explicit negative sentiments were generated by bot users, with the EFF, DA, and ActionSA following in that order. It should be noted that the analysis of this study was based on a sample set of data from Twitter and does not reflect the sentiments of the entire South African population.

6.2 Implications

This study has numerous implications that may affect several parties. This study provides useful information for political parties, particularly the ANC, by demonstrating the number of negative sentiments in tweets. Parties could use this information as guidelines to improve their methods of interaction in order to tackle societal problems and enhance their overall reputation. Moreover, the sentiment analysis findings from this study could potentially be used by political campaign strategists to modify their ap-

proaches in light of public perceptions. It is easier to develop communications that resonate with voters and successfully address problems when one is aware of the various kinds of sentiments. Understanding public discourse is improved by knowledge of how sentiment changes over time and how various political parties differ in their polarities. With this information, individuals could engage in more thoughtful and nuanced conversations. Additionally, the study brings attention to the prominence of bots on social media and how they could affect political debate. This knowledge could inspire people to assess internet content attentively and participate in political conversations more carefully. The implications highlighted affect the political, social, and technological spheres. They have an impact on how political parties interact with the general population and how individuals participate in and navigate online political discourse.

6.3 Future Work

Overall, there is still significant opportunity for improvement in all of the models. Further study will focus on further model improvement and the incorporation of different techniques to address the challenge of class imbalance. In order to prevent biases in the dataset, the research will also take into account several data sources. Furthermore, using the same dataset, the study will take into consideration the analysis of the code-switched tweets.

6.4 Recommendations

In politics, sentiment analysis of social media data can produce the most beneficial outcomes, particularly during election campaigns. Political parties frequently seek to gain the trust of the community they are running for office by learning about it. Understanding the sentiments of the voters they hope to win over is crucial for political parties running for local government. By using social media and other communication channels, sentiment analysis may also assist political parties in structuring their campaigns in a way which helps them and attracts more devoted supporters.

Bibliography

[1] S. Darad and S. Krishnan, "Sentimental analysis of covid-19 twitter data using deep learning and machine learning models," *Ingenius. Revista de Ciencia y Tecnología*, no. 29, pp. 108–117, 2023.

[2] M. Ledwaba and V. Marivate, "Semi-supervised learning approaches for predicting south african political sentiment for local government elections," in *DG. O 2022: The 23rd Annual International Conference on Digital Government Research*, pp. 129–137, 2022.

[3] J. P. Pinto and V. Murari, "Real time sentiment analysis of political twitter data using machine learning approach," *International Research Journal of Engineering and Technology (IRJET)*, no. 4, pp. 4124–4129, 2019.

[4] M. Nawaz, "Unsupervised sentiment analysis using vader and flair." https://soshace.com/unsupervised-sentiment-analysis-using-vader-and-flair/, 2023.

[5] K. Singhal, B. Agrawal, and N. Mittal, "Modeling indian general elections: sentiment analysis of political twitter data," in *Information Systems Design and Intelligent Applications: Proceedings of Second International Conference INDIA 2015*, pp. 469–477, 2015.

[6] S. Elbagir and J. Yang, "Twitter sentiment analysis using natural language toolkit and vader sentiment," in *Proceedings of the international multiconference of engineers and computer scientists*, p. 16, 2019.

[7] B. García-Orosa, P. Gamallo, P. Martín-Rodilla, and R. Martínez-Castaño, "Hybrid intelligence strategies for identifying, classifying and analyzing political bots," *Social sciences*, no. 10, p. 357, 2021.

[8] S. Västerbo, "Classifying twitter botsa comparasion of methods for classifying whethert weets are written by humans or bots," 2020.

[9] Y. Liu, M. Ott, N. Goyal, J. Du, M. Joshi, D. Chen, O. Levy, M. Lewis, L. Zettlemoyer, and V. Stoyanov, "Roberta: A robustly optimized bert pretraining approach," *arXiv preprint arXiv:1907.11692*, 2019.

[10] T. Mustaqim, "Analysis of public opinion on religion and politics in indonesia using k-means clustering and vader sentiment polarity detection," in *Proceeding International Conference on Science and Engineering*, pp. 749–754, 2020.

[11] J. P. Gujjar and H. Kumar, "Sentiment analysis: Textblob for decision making," *Int. J. Sci. Res. Eng. Trends*, no. 2, pp. 1097–1099, 2021.

[12] O. D. Tijani, S. Onashoga, and A. Akinwale, "An auto-generated approach of stop words using aggregated analysis," in *13th International Conference on Information Technology Innovation for Sustainable Development*, 2017.

[13] O. Abiola, A. Abayomi-Alli, O. A. Tale, S. Misra, and O. Abayomi-Alli, "Sentiment analysis of covid-19 tweets from selected hashtags in nigeria using vader and text blob analyser," *Journal of Electrical Systems and Information Technology*, no. 1, pp. 1–20, 2023.

[14] S. Bengesi, T. Oladunni, R. Olusegun, and H. Audu, "A machine learning-sentiment analysis on monkeypox outbreak: An extensive dataset to show the polarity of public opinion from twitter tweets," *IEEE Access*, no. 11, pp. 11811–11826, 2023.

[15] A. Alabrah, H. M. Alawadh, O. D. Okon, T. Meraj, and H. T. Rauf, "Gulf countries' citizens' acceptance of covid-19 vaccines—a machine learning approach," *Mathematics*, no. 3, p. 467, 2022.

[16] F. Illia, M. P. Eugenia, and S. A. Rutba, "Sentiment analysis on pedulilindungi application using textblob and vader library," in *Proceedings of The International Conference on Data Science and Official Statistics*, pp. 278–288, 2021.

[17] A. M. Ashir, "A generalized method for sentiment analysis across different sources," *Applied Computational Intelligence and Soft Computing*, no. 1, pp. 1–8, 2021.

[18] D. O. Oyewola, L. A. Oladimeji, S. O. Julius, L. B. Kachalla, and E. G. Dada, "Optimizing sentiment analysis of nigerian 2023 presidential election using two-stage residual long short term memory," *Heliyon*, no. 4, 2023.

[19] A. Shevtsov, M. Oikonomidou, D. Antonakaki, P. Pratikakis, and S. Ioannidis, "What tweets and youtube comments have in common? sentiment and graph analysis on data related to us elections 2020," *Plos one*, no. 1, p. e0270542, 2023.

[20] R. D. Endsuy, "Sentiment analysis between vader and eda for the us presidential election 2020 on twitter datasets," *Journal of Applied Data Sciences*, no. 1, pp. 08–18, 2021.

[21] V. Nandi and S. Agrawal, "Political sentiment analysis using hybrid approach," *International Research Journal of Engineering and Technology*, no. 5, pp. 1621–1627, 2016.

[22] M. Mujahid, E. Lee, F. Rustam, P. B. Washington, S. Ullah, A. A. Reshi, and I. Ashraf, "Sentiment analysis and topic modeling on tweets about online education during covid-19," *Applied Sciences*, no. 18, p. 8438, 2021.

[23] M. Al-Shabi, "Evaluating the performance of the most important lexicons used to sentiment analysis and opinions mining," *IJCSNS*, no. 1, p. 1, 2020.

[24] C. Hutto and E. Gilbert, "Vader: A parsimonious rule-based model for sentiment analysis of social media text," in *Proceedings of the international AAAI conference on web and social media*, pp. 216–225, 2014.

[25] O. Alqaryouti, N. Siyam, A. A. Monem, and K. Shaalan, "Aspect-based sentiment analysis using smart government review data," *Applied Computing and Informatics*, 2020.

[26] T.-L. Luong, M.-S. Cao, D.-T. Le, and X.-H. Phan, "Intent extraction from social media texts using sequential segmentation and deep learning models," in *2017 9th International Conference on Knowledge and Systems Engineering (KSE)*, pp. 215–220, 2017.

[27] A. Dehghan, K. Siuta, A. Skorupka, A. Dubey, A. Betlen, D. Miller, W. Xu, B. Kamiński, and P. Prałat, "Detecting bots in social-networks using node and structural embeddings," *Journal of Big Data*, no. 1, p. 119, 2023.

[28] F. K. Alarfaj, H. Ahmad, H. U. Khan, A. M. Alomair, N. Almusallam, and M. Ahmed, "Twitter bot detection using diverse content features and applying machine learning algorithms," *Sustainability*, no. 8, p. 6662, 2023.

[29] M. Washha, A. Qaroush, M. Mezghani, and F. Sèdes, "Information quality in social networks: Predicting spammy naming patterns for retrieving twitter spam accounts," in *19th International Conference on Enterprise Information Systems (ICEIS 2017)*, pp. 610–622, 2017.

[30] D. M. Beskow and K. M. Carley, "Its all in a name: detecting and labeling bots by their name," *Computational and mathematical organization theory*, pp. 24–35, 2019.

[31] E. K. Genfi, "Detecting bots using a hybrid approach," *Theses, Dissertations and Culminating Projects*, no. 736, 2021.

[32] E. Alothali, K. Hayawi, and H. Alashwal, "Hybrid feature selection approach to identify optimal features of profile metadata to detect social bots in twitter," *Social Network Analysis and Mining*, pp. 1–15, 2021.

[33] S. B. Abkenar, M. H. Kashani, M. Akbari, and E. Mahdipour, "Twitter spam detection: A systematic review," *arXiv preprint arXiv:2011.14754*, pp. arXiv–2011, 2020.

[34] S. Kudugunta and E. Ferrara, "Deep neural networks for bot detection," *Information Sciences*, pp. 312–322, 2018.

[35] J. P. Dickerson, V. Kagan, and V. Subrahmanian, "Using sentiment to detect bots on twitter: Are humans more opinionated than bots?," in *2014 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM 2014)*, pp. 620–627, 2014.

[36] "Stopwords." https://www.ranks.nl/stopwords. [Online; accessed 2023-11-17].

[37] B. Liu, *Sentiment analysis: Mining opinions, sentiments, and emotions*, pp. 1–17. Cambridge university press, 2020.

[38] I. Cribben and Y. Zeinali, "The benefits and limitations of chatgpt in business education and research: A focus on management science, operations management and data analytics," *Operations Management and Data Analytics (March 29, 2023)*, 2023.

[39] M. Moradi, K. Blagec, F. Haberl, and M. Samwald, "Gpt-3 models are poor few-shot learners in the biomedical domain," *arXiv preprint arXiv:2109.02555*, 2021.

[40] T. Brown, B. Mann, N. Ryder, M. Subbiah, J. D. Kaplan, P. Dhariwal, A. Neelakantan, P. Shyam, G. Sastry, A. Askell, *et al.*, "Language models are few-shot learners," *Advances in neural information processing systems*, pp. 1877–1901, 2020.

[41] J. M. Johnson and T. M. Khoshgoftaar, "Survey on deep learning with class imbalanc," *Journal of Big Data*, no. 1, pp. 1–54, 2019.

[42] S. M. John and K. Kartheeban, "Sentiment scoring and performance metrics examination of various supervised classifiers," *International Journal of Innovative Technology and Exploring Engineering*, no. 2, p. 9, 2019.

[43] K. Chakraborty, S. Bhattacharyya, R. Bag, and L. Mršić, "Sentiment analysis on labeled and unlabeled datasets using bert architecture," *Soft Computing*, pp. 1–18, 2023.

[44] M. Rossetti and T. Zaman, "Bots, disinformation, and the first impeachment of us president donald trump," *Plos one*, no. 5, p. e0283971, 2023.

Appendix A

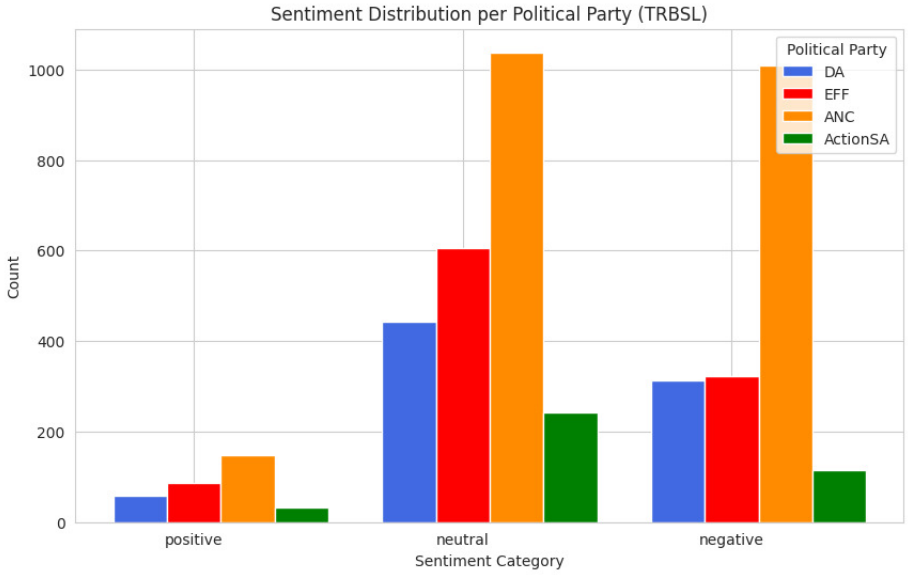


Figure A1: Sentiment Distribution Per Political Party- TRBSL

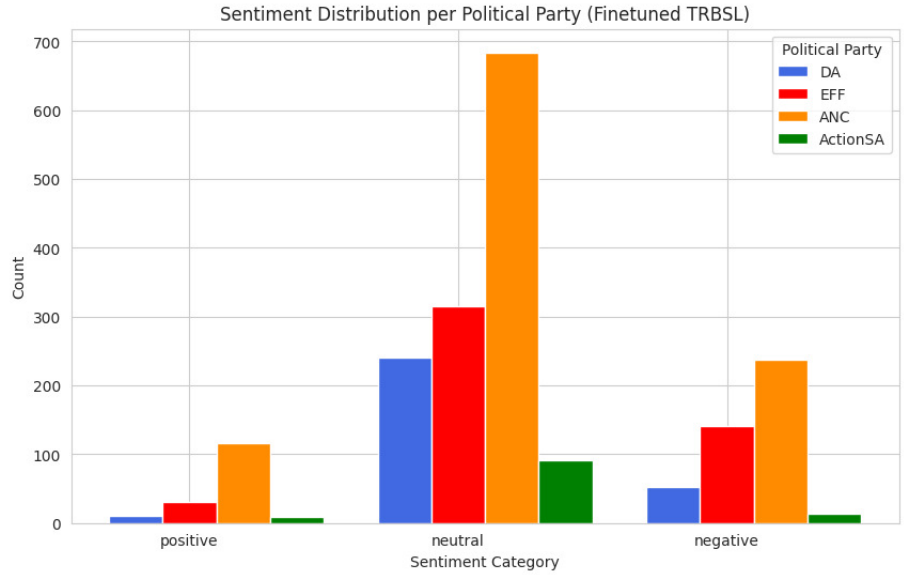
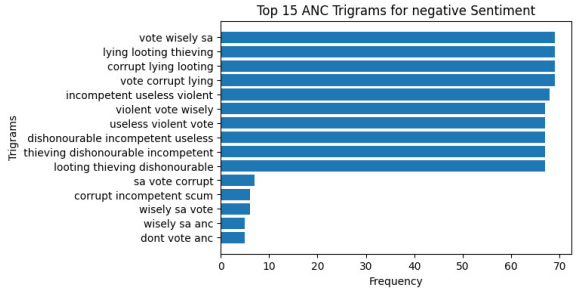
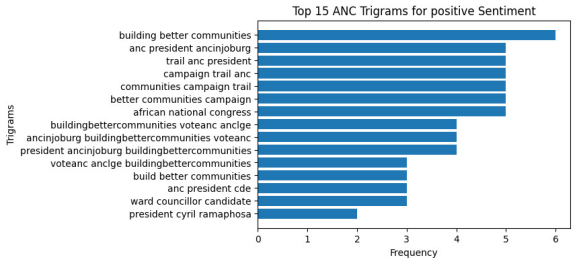


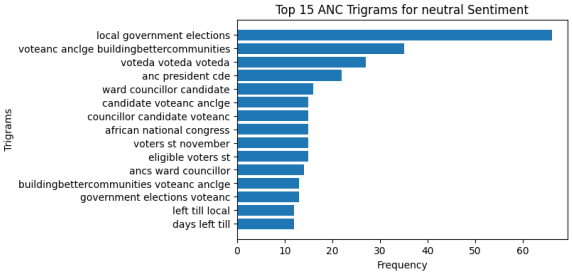
Figure A2: Sentiment Distribution Per Political Party- TRBSL**



(a) Negative Trigram Frequency

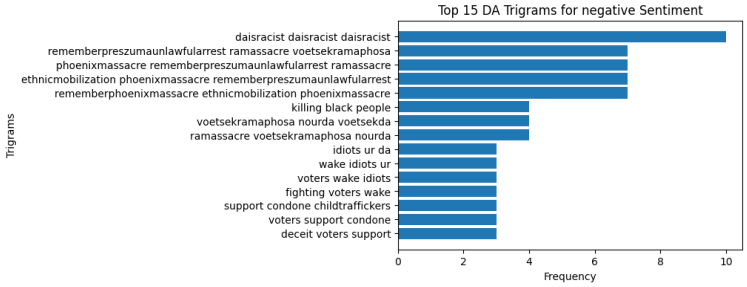


(b) Positive Trigram Frequency

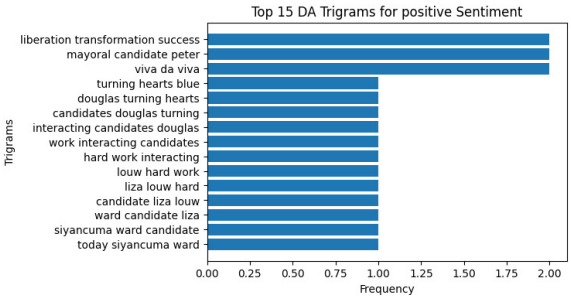


(c) Neutral Trigram Frequency

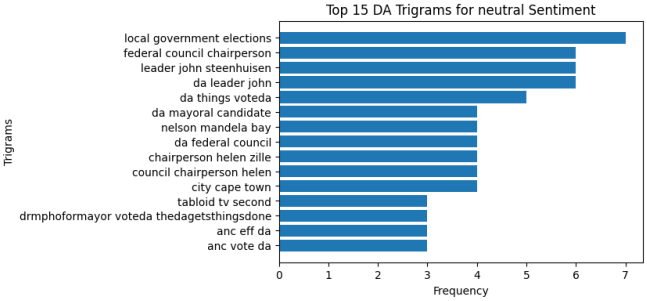
Figure A3: Trigram Frequency For ANC Using TRBSL Model



(a) Negative Trigram Frequency



(b) Positive Trigram Frequency



(c) Neutral Trigram Frequency

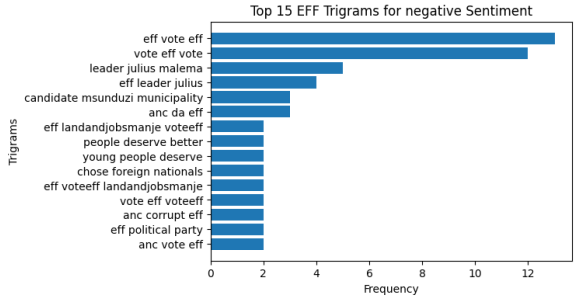
Figure A4: Trigram Frequency For DA Using TRBSL Model

Epoch	Training Loss	Validation Loss	Training Accuracy	Training F1 Score
1	0.6353	0.6678	0.7245	0.7245
2	0.5138	0.6488	0.7931	0.7931
3	0.4171	0.8186	0.8414	0.8414

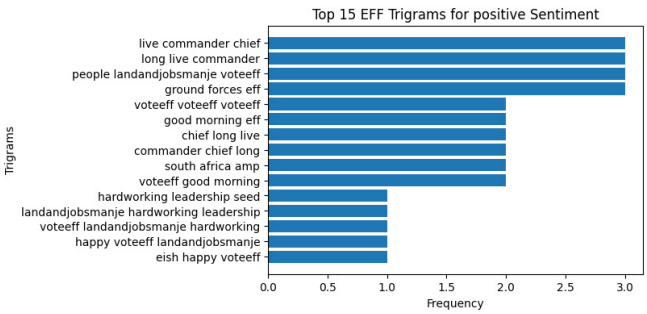
Table A1: Training Results Of Finetuning TRBSL Model

eval_loss	eval_accuracy	eval_f1_score	eval_runtime	eval_samples_per_second	eval_steps_per_second
0.8251	0.6822	0.6822	27.1277	33.1760	0.5530

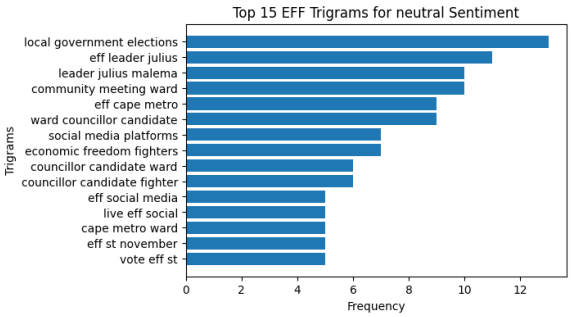
Table A2: Evaluation Results Of Finetuning TRBSL Model



(a) Negative Trigram Frequency

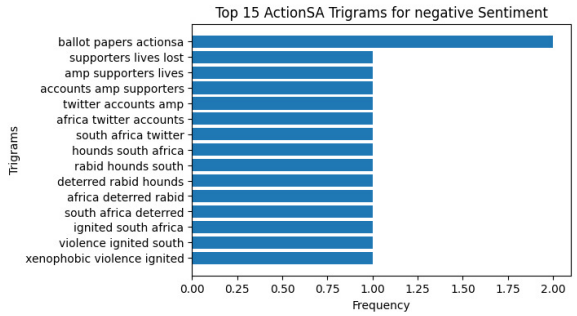


(b) Positive Trigram Frequency

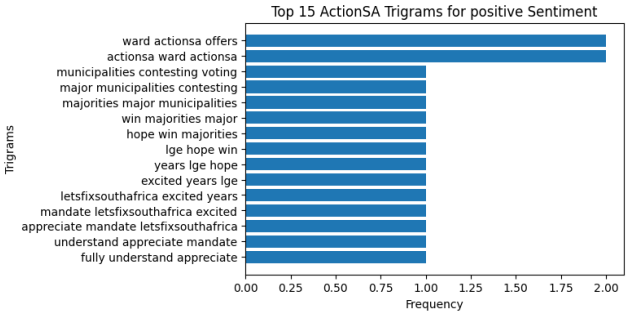


(c) Neutral Trigram Frequency

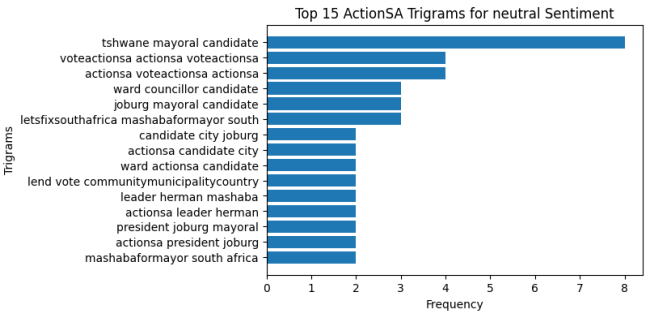
Figure A5: Trigram Frequency For EFF Using TRBSL Model



(a) Negative Trigram Frequency

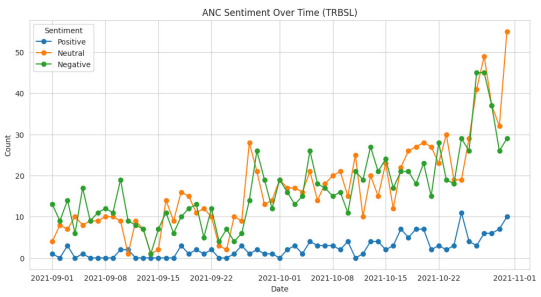


(b) Positive Trigram Frequency

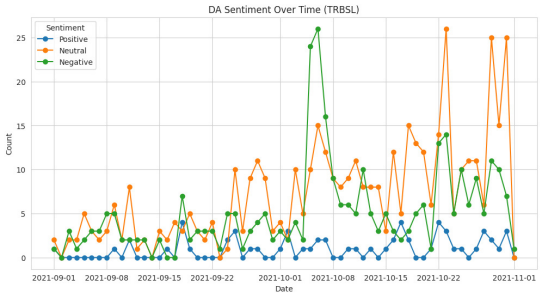


(c) Neutral Trigram Frequency

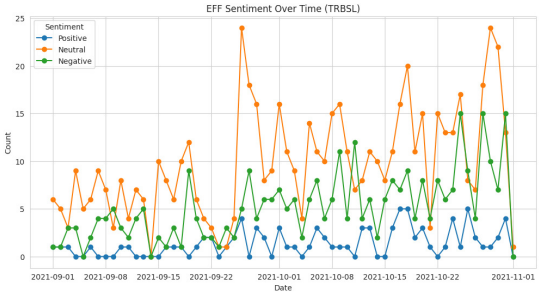
Figure A6: Trigram Frequency For ActionSA Using TRBSL Model



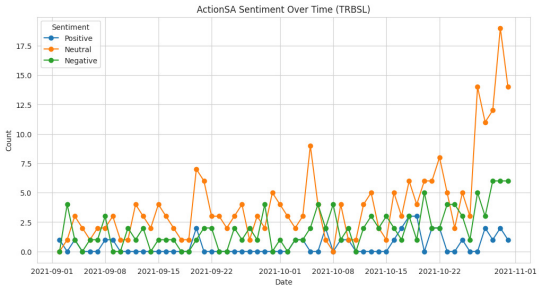
(a) ANC Time Plot



(b) DA Time Plot



(c) EFF Time Plot



(d) ActionSA Time Plot

Figure A7: Time Series For ActionSA Using TRBSL Model

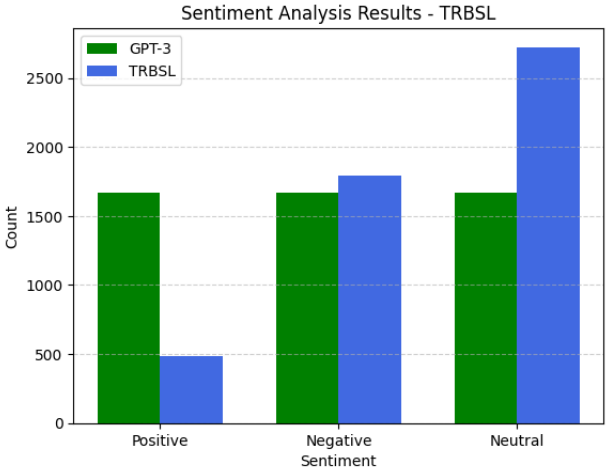
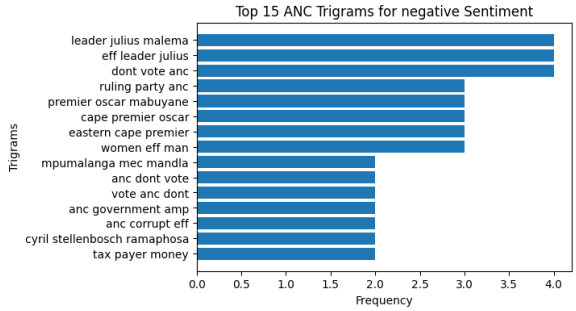


Figure A8: Sentiment Distribution For GPT-3.5 vs TRBSL

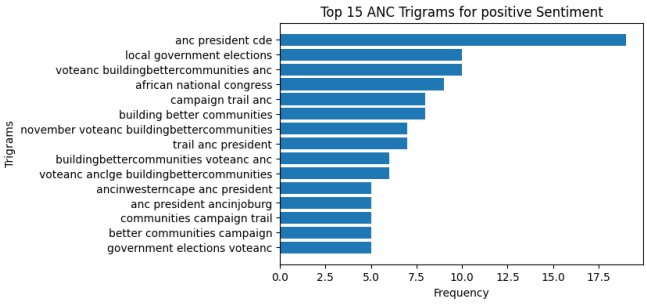
```
1 summary(model)
```

```
==========================================================================
Layer (type:depth-idx)                                   Param #
==========================================================================
RobertaForSequenceClassification                         --
├─RobertaModel: 1-1                                      --
│    └─RobertaEmbeddings: 2-1                            --
│    │    └─Embedding: 3-1                               38,603,520
│    │    └─Embedding: 3-2                               394,752
│    │    └─Embedding: 3-3                               768
│    │    └─LayerNorm: 3-4                               1,536
│    │    └─Dropout: 3-5                                 --
│    └─RobertaEncoder: 2-2                               --
│    │    └─ModuleList: 3-6                              85,054,464
├─RobertaClassificationHead: 1-2                         --
│    └─Linear: 2-3                                       590,592
│    └─Dropout: 2-4                                      --
│    └─Linear: 2-5                                       2,307
==========================================================================
Total params: 124,647,939
Trainable params: 124,647,939
Non-trainable params: 0
==========================================================================
```

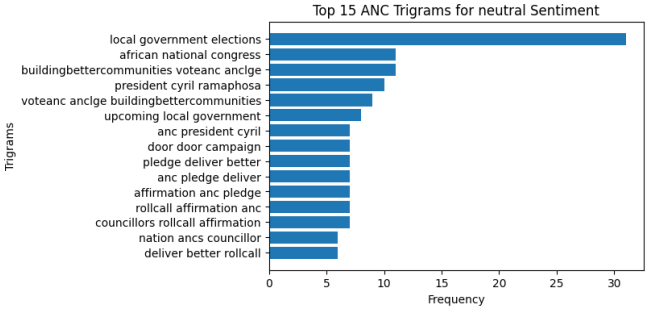
Figure A9: Model Summary For TRBSL



(a) Negative Trigram Frequency

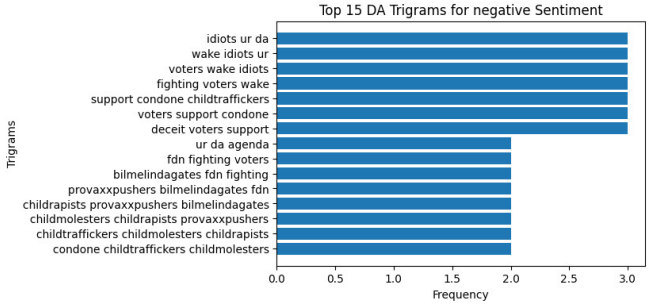


(b) Positive Trigram Frequency

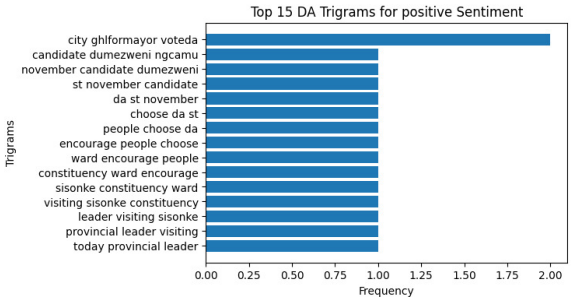


(c) Neutral Trigram Frequency

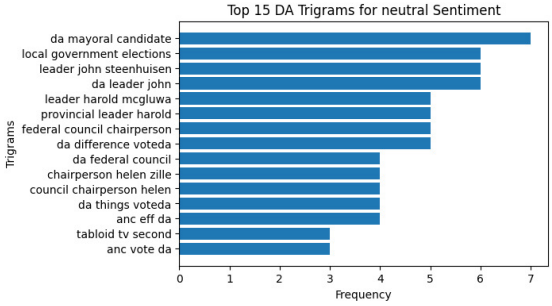
Figure A10: Trigram Frequency For ANC Using TRBSL** Model



(a) Negative Trigram Frequency

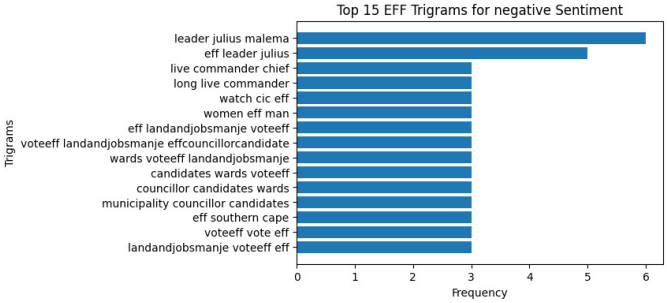


(b) Positive Trigram Frequency

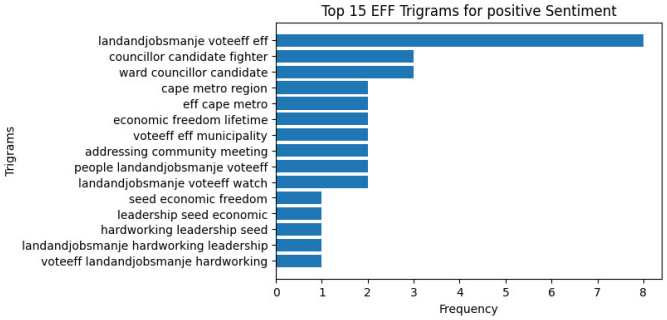


(c) Neutral Trigram Frequency

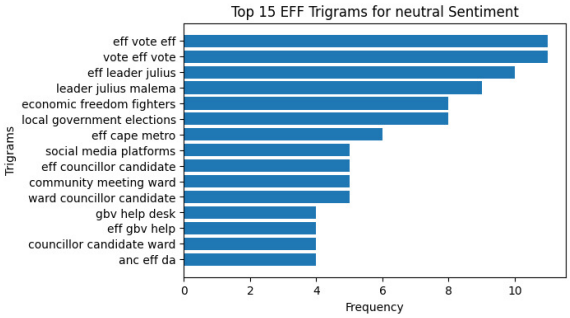
Figure A11: Trigram Frequency For DA Using TRBSL** Model



(a) Negative Trigram Frequency

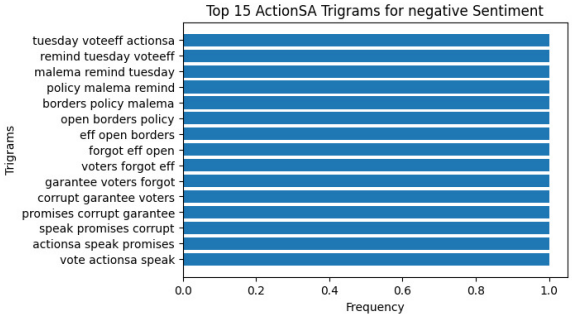


(b) Positive Trigram Frequency

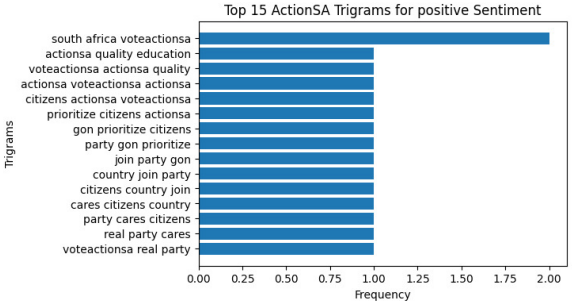


(c) Neutral Trigram Frequency

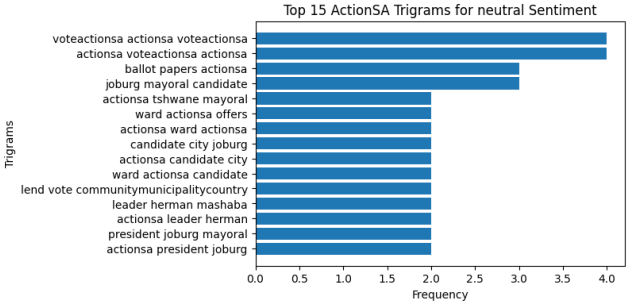
Figure A12: Trigram Frequency For EFF Using TRBSL** Model



(a) Negative Trigram Frequency

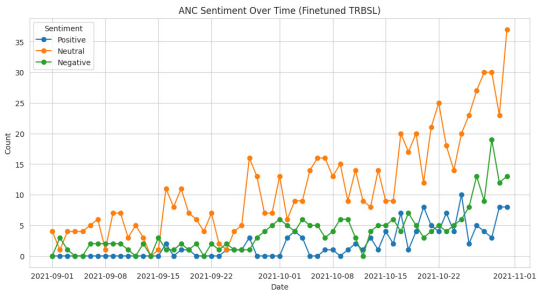


(b) Positive Trigram Frequency

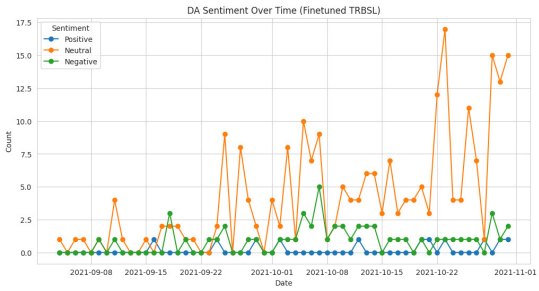


(c) Neutral Trigram Frequency

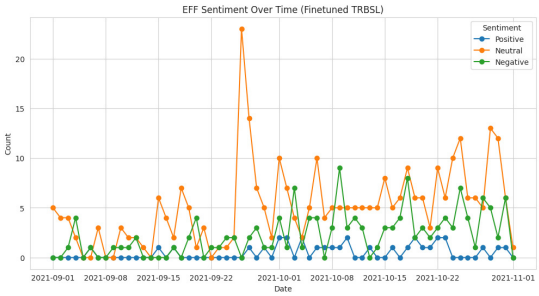
Figure A13: Trigram Frequency For ActionSA Using TRBSL** Model



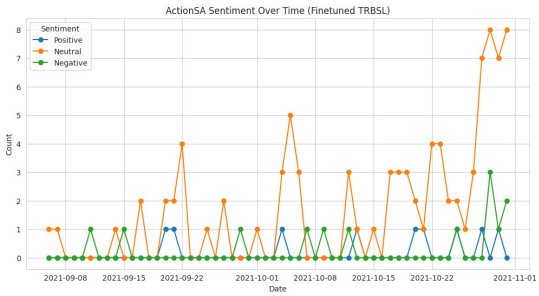
(a) ANC Time Plot



(b) DA Time Plot



(c) EFF Time Plot



(d) ActionSA Time Plot

Figure A14: Time Series For ActionSA Using TRBSL** Model

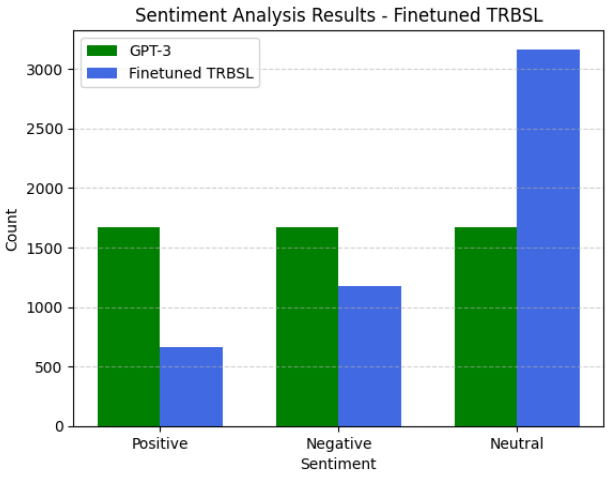


Figure A15: Sentiment Distribution For GPT-3.5 vs TRBSL**

Appendix B

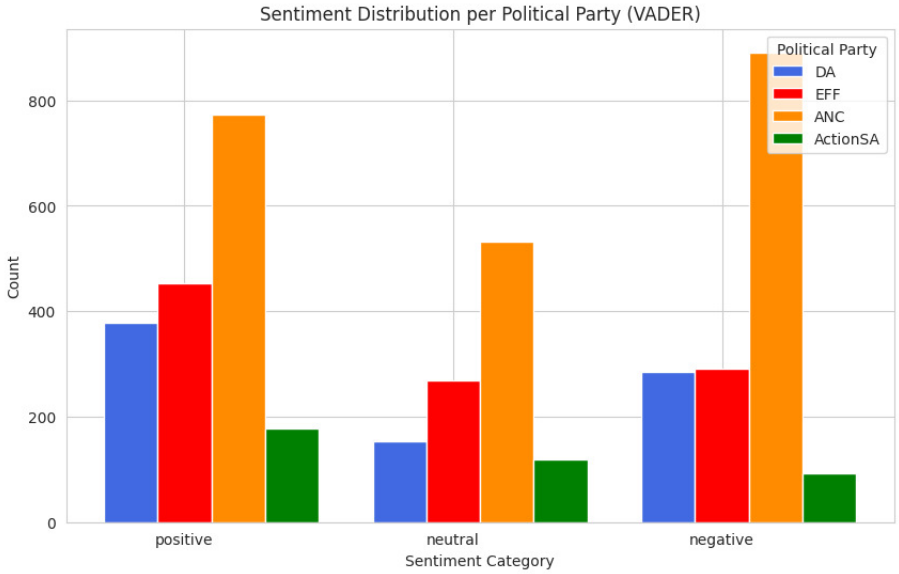
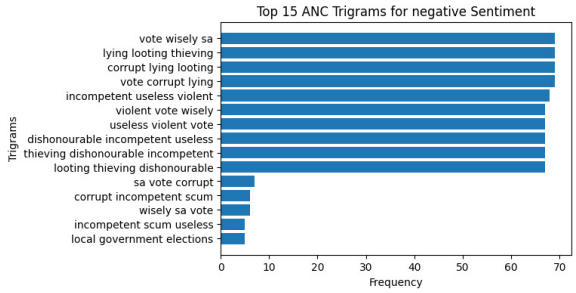
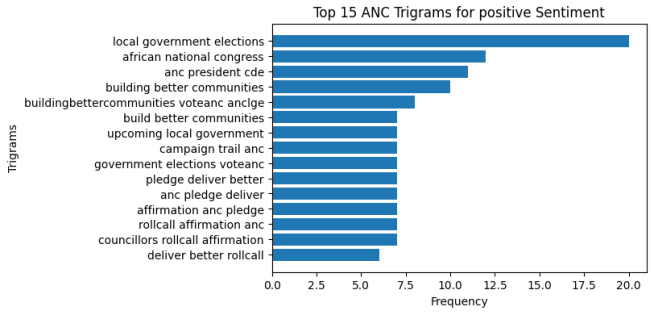


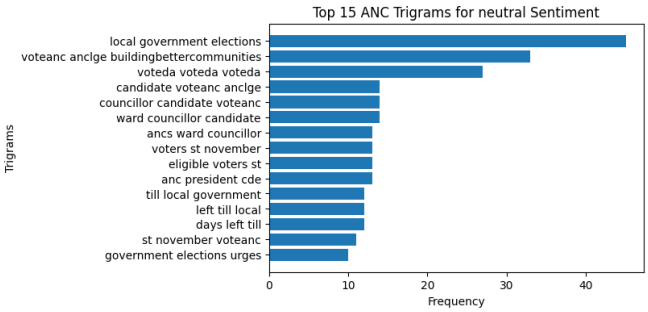
Figure B1: Sentiment Distribution Per Political Party- VADER



(a) Negative Trigram Frequency

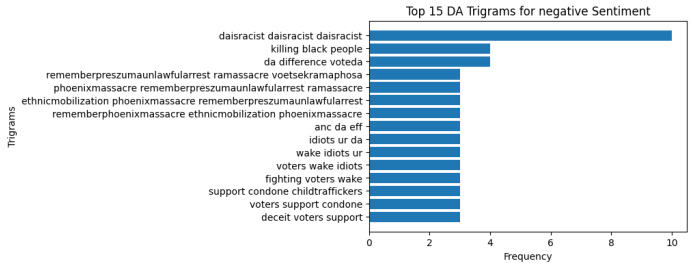


(b) Positive Trigram Frequency

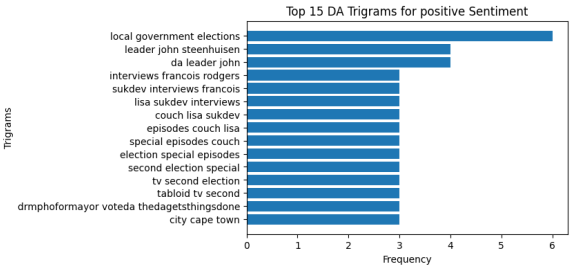


(c) Neutral Trigram Frequency

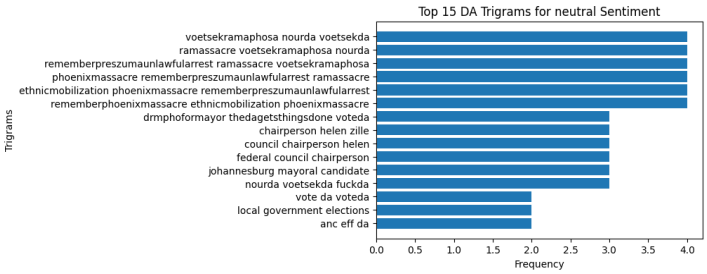
Figure B2: Trigram Frequency For ANC Using VADER Model



(a) Negative Trigram Frequency

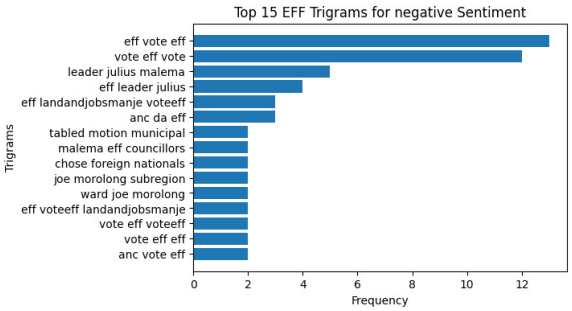


(b) Positive Trigram Frequency

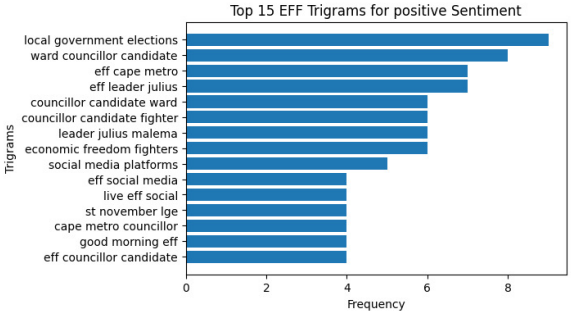


(c) Neutral Trigram Frequency

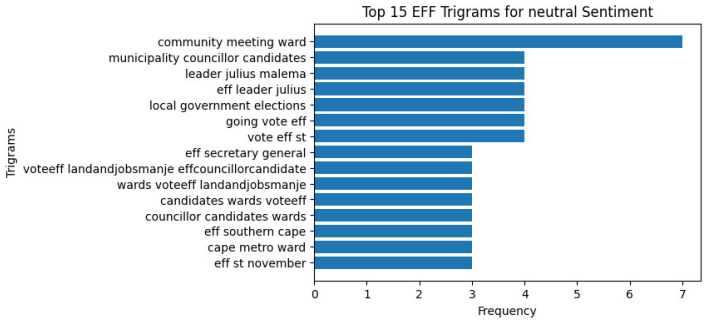
Figure B3: Trigram Frequency For DA Using VADER Model



(a) Negative Trigram Frequency

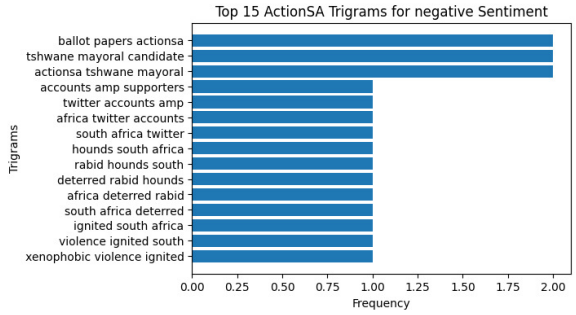


(b) Positive Trigram Frequency

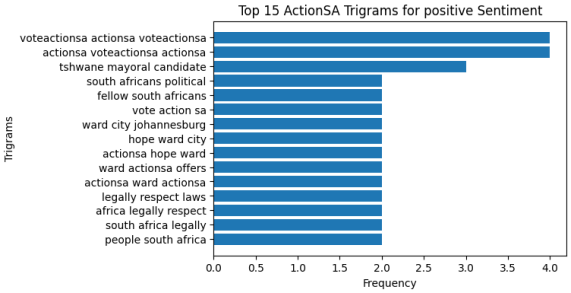


(c) Neutral Trigram Frequency

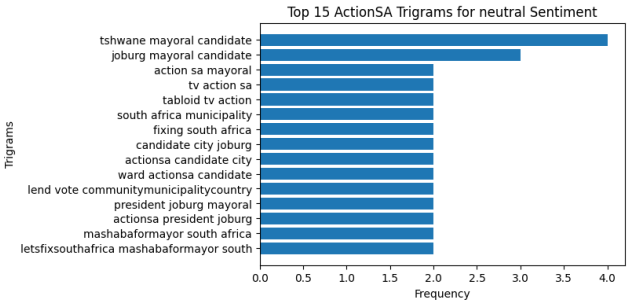
Figure B4: Trigram Frequency For EFF Using VADER Model



(a) Negative Trigram Frequency

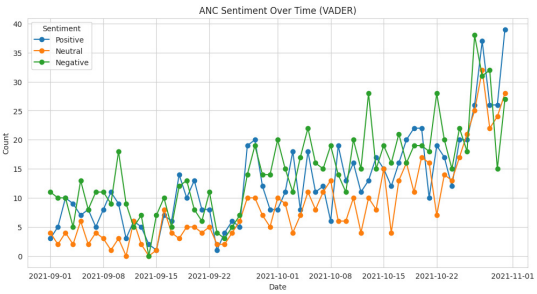


(b) Positive Trigram Frequency

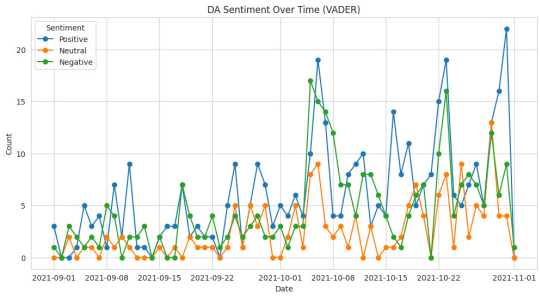


(c) Neutral Trigram Frequency

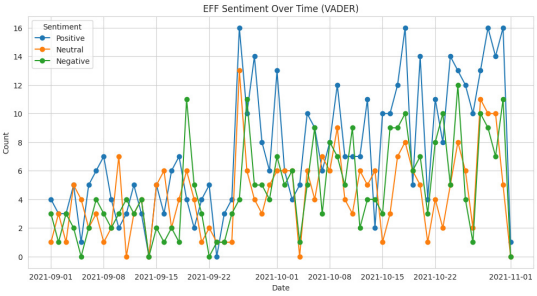
Figure B5: Trigram Frequency For ActionSA Using VADER Model



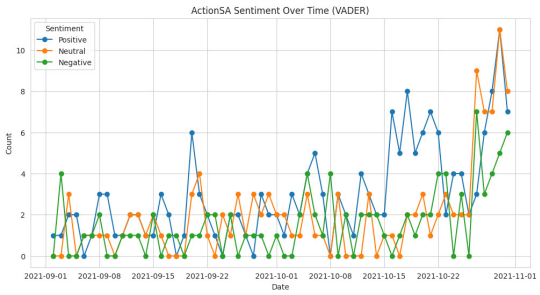
(a) ANC Time Plot



(b) DA Time Plot



(c) EFF Time Plot



(d) ActionSA Time Plot

Figure B6: Time Series For ActionSA Using VADER Model

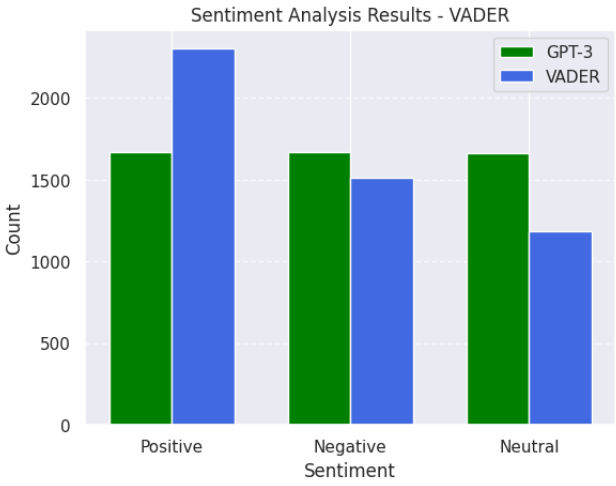


Figure B7: Sentiment Distribution For GPT-3.5 vs VADER

Appendix C

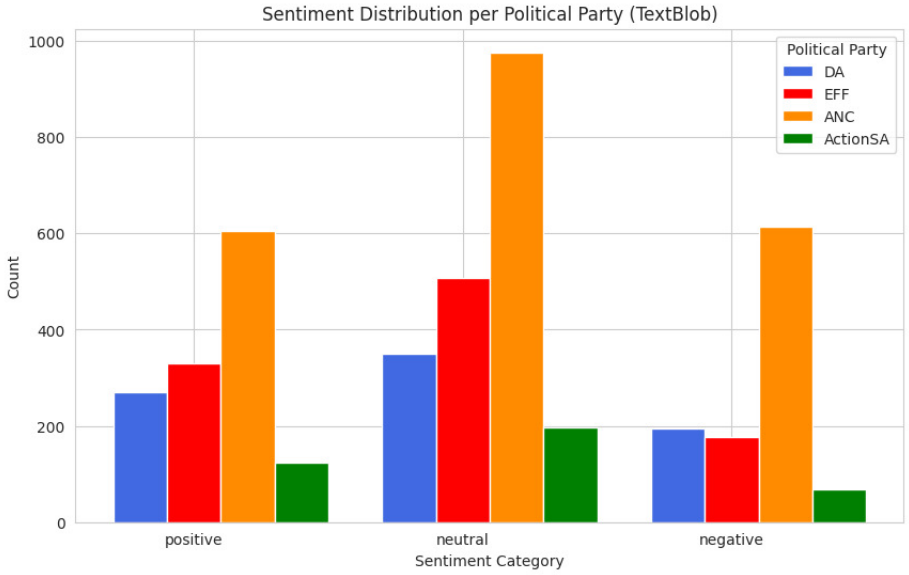
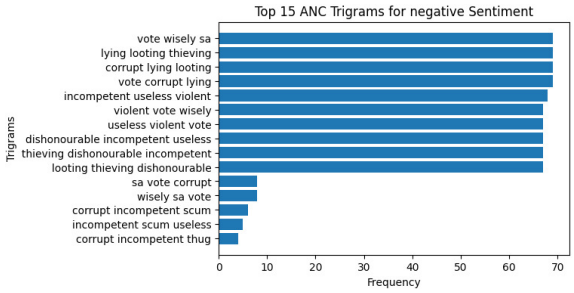
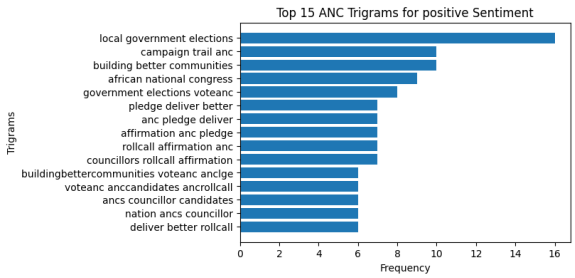


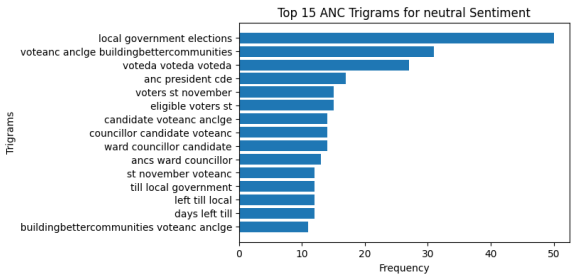
Figure C1: Sentiment Distribution Per Political Party- TextBlob



(a) Negative Trigram Frequency

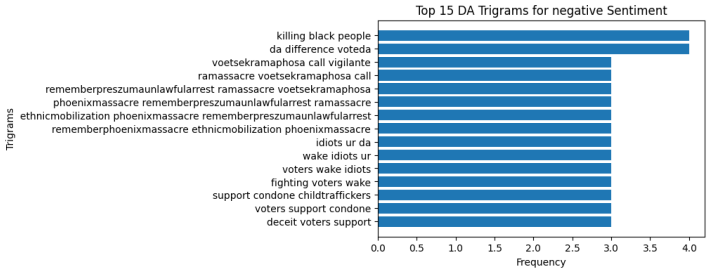


(b) Positive Trigram Frequency

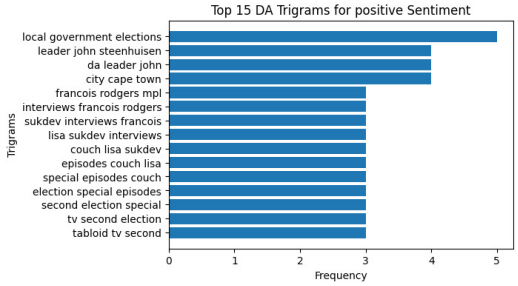


(c) Neutral Trigram Frequency

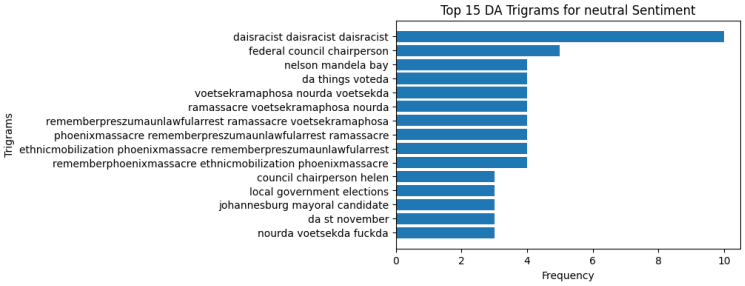
Figure C2: Trigram Frequency For ANC Using TextBlob Model



(a) Negative Trigram Frequency

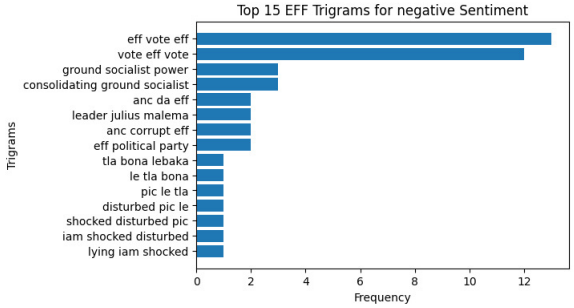


(b) Positive Trigram Frequency

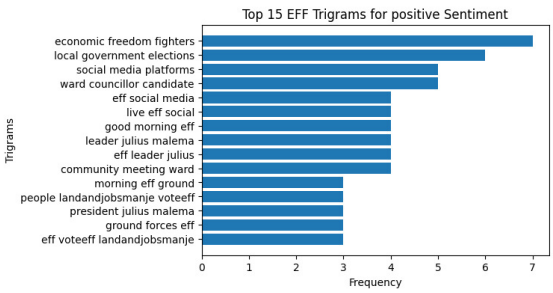


(c) Neutral Trigram Frequency

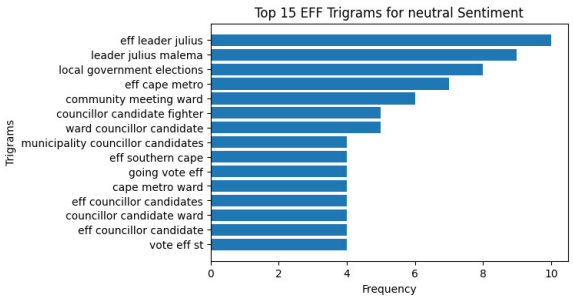
Figure C3: Trigram Frequency For DA Using TextBlob Model



(a) Negative Trigram Frequency

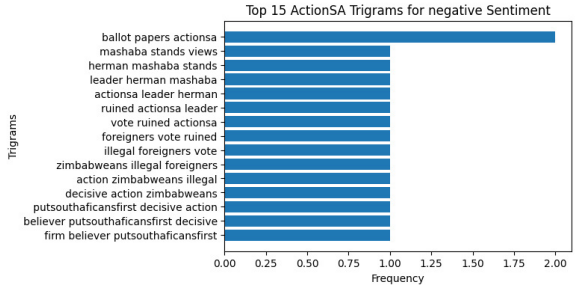


(b) Positive Trigram Frequency

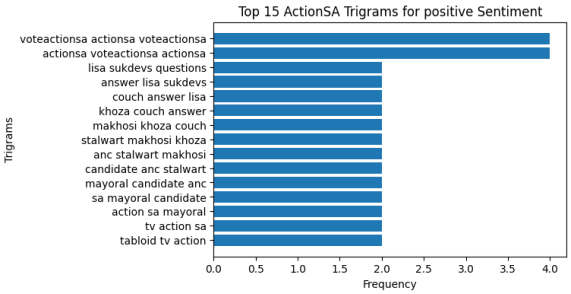


(c) Neutral Trigram Frequency

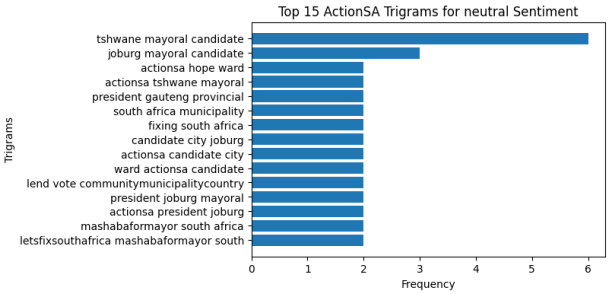
Figure C4: Trigram Frequency For EFF Using TextBlob Model



(a) Negative Trigram Frequency

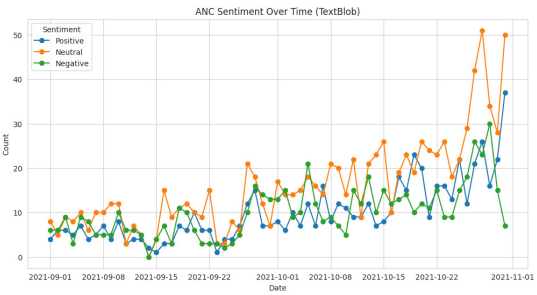


(b) Positive Trigram Frequency

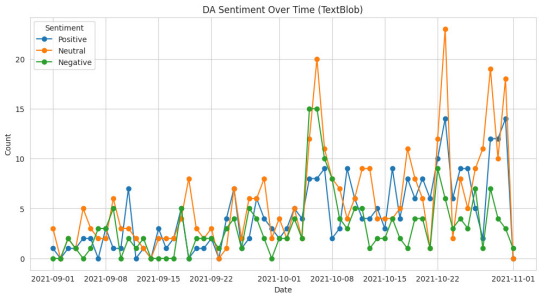


(c) Neutral Trigram Frequency

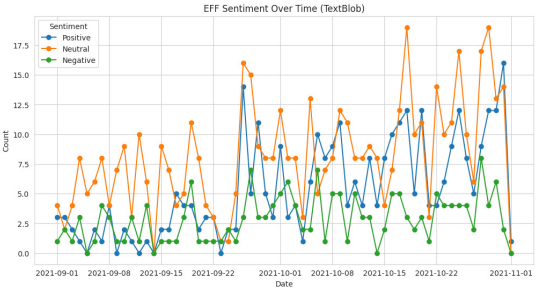
Figure C5: Trigram Frequency For ActionSA Using TextBlob Model



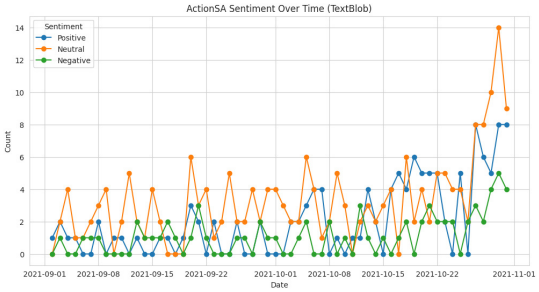
(a) ANC Time Plot



(b) DA Time Plot



(c) EFF Time Plot



(d) ActionSA Time Plot

Figure C6: Time Series For ActionSA Using TextBlob Model

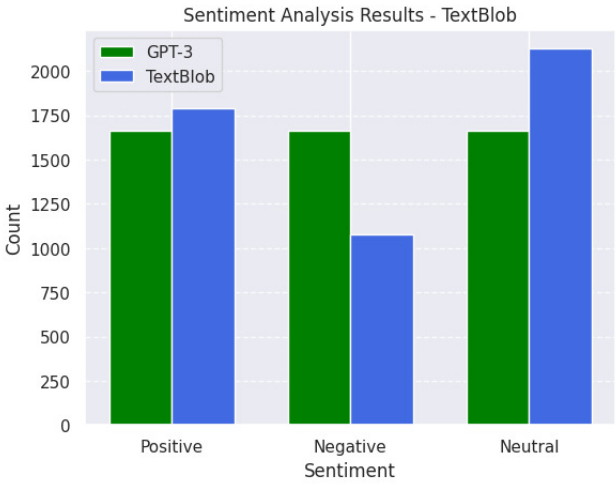


Figure C7: Sentiment Distribution For GPT-3.5 vs TextBlob

Appendix D

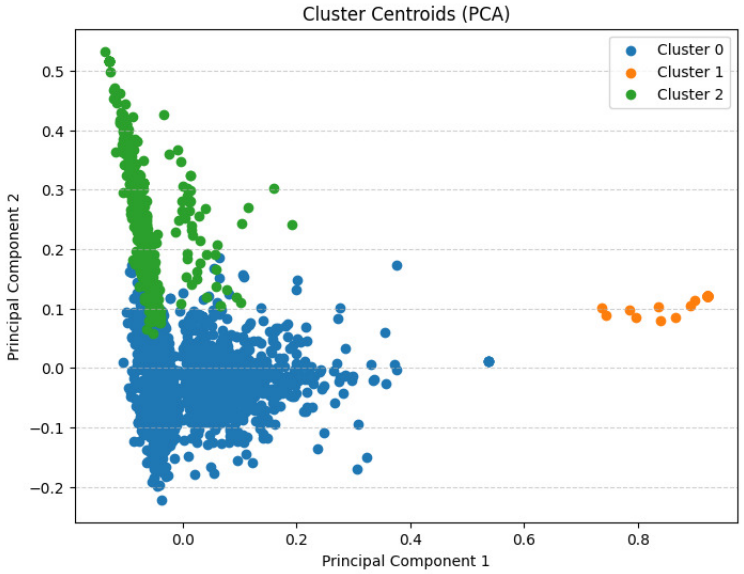


Figure D1: Cluster Centroids Visualised In The Reduced 2D Space

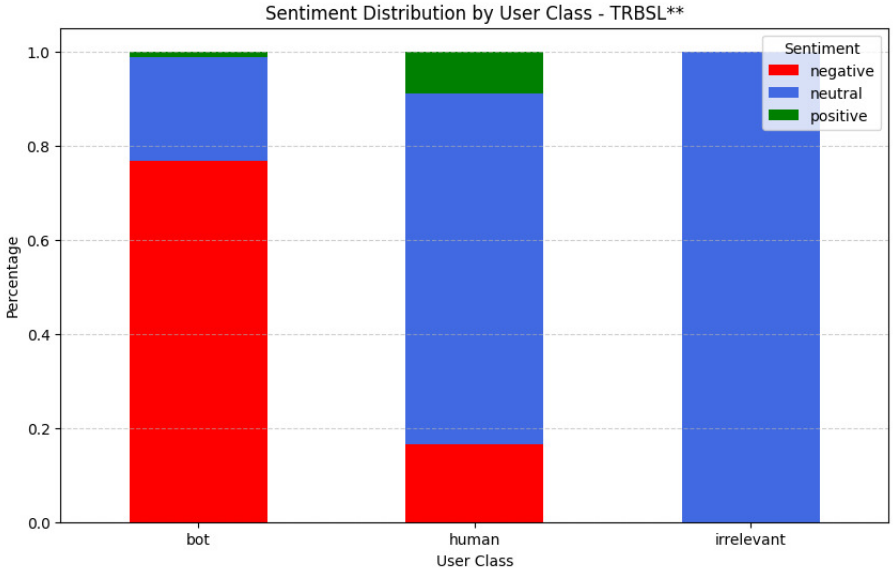


Figure D2: Classification Distribution - TRBSL**

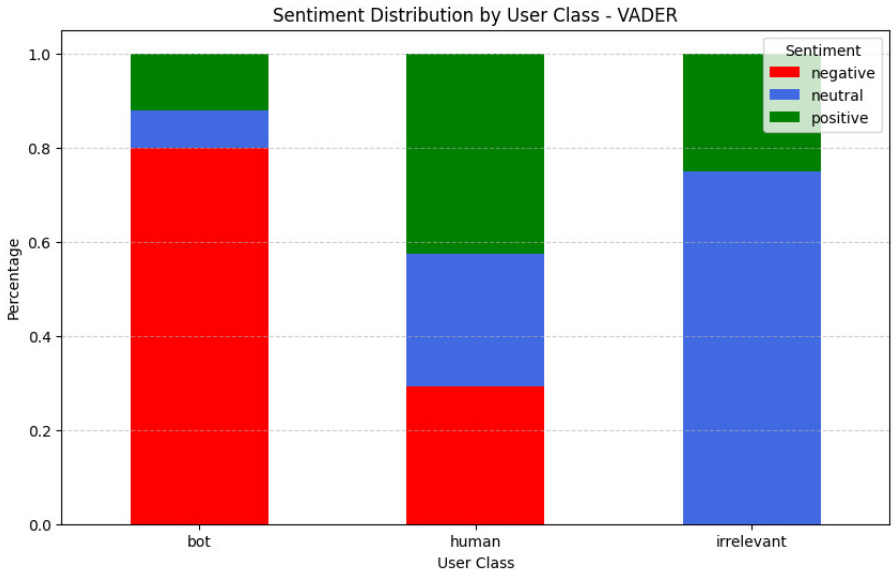


Figure D3: Classification Distribution - VADER

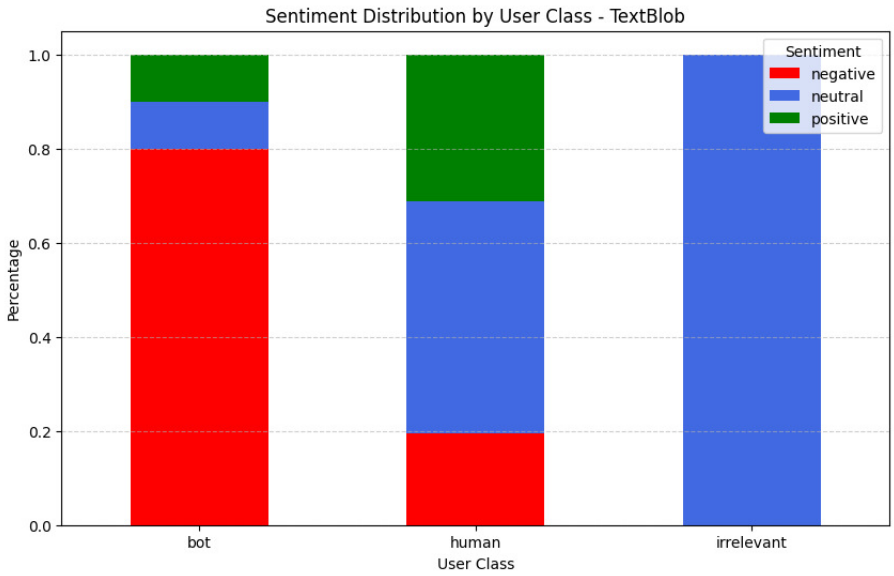
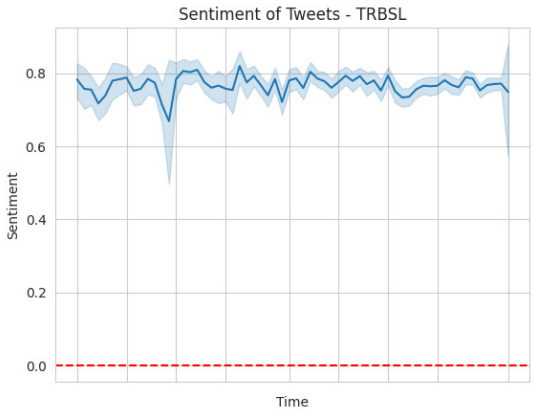
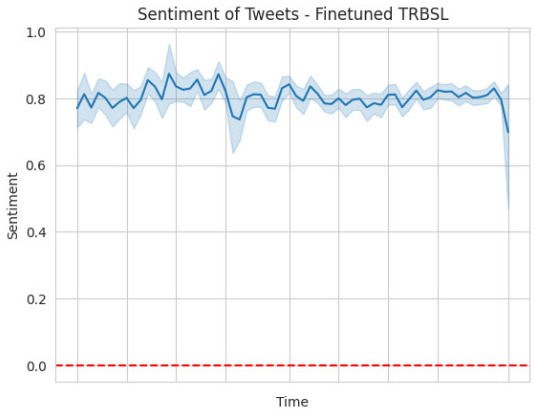


Figure D4: Classification Distribution - TextBlob

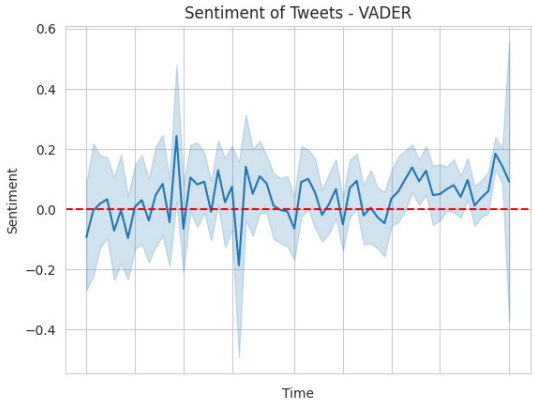
Appendix E



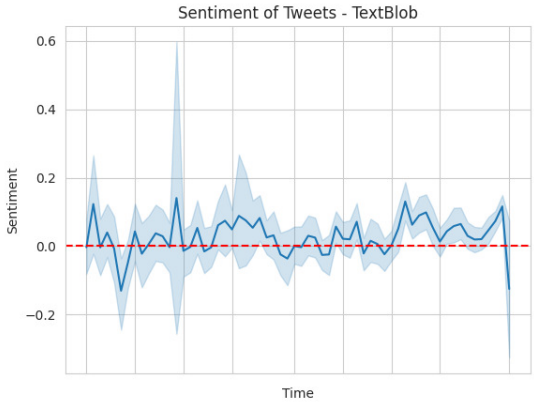
(a) Trend of Sentiment Over Time - TRBSL



(b) Trend of Sentiment Over Time - TRBSL**



(c) Trend of Sentiment Over Time - VADER



(d) Trend of Sentiment Over Time - TextBlob

Figure E1: Wordcloud For ActionSA Using TextBlob Model