JeDEM Issue 17 (1): 144-169, 2025
ISSN 2075-9517
http://www.jedem.org
Date of submission: 07.17.2024
Date of acceptance: 10.02.2025
Article DOI: 10.29379/jedem.v17i1.945

Sentiment analysis using unsupervised learning for local government elections in South Africa

Penelope Matloga

ORCID Nr: 0009-0008-3123-0913
University of Pretoria, Pretoria, South Africa
u22826476@tuks.co.za

Vukosi Marivate

ORCID Nr: 0000-0002-6731-6267
University of Pretoria, Pretoria, South Africa
vukosi.marivate@cs.up.ac.za

Kayode Olaleye

ORCID Nr: 0000-0002-2622-7305
University of Pretoria, Pretoria, South Africa
kayode.olaleye@cs.up.ac.za

Abstract: This study examines public sentiment during the 2021 South African local government election campaign by analysing Twitter posts. The research uses advanced techniques such as fine-tuned RoBERTa model, VADER, and TextBlob to assess the sentiments of tweets about four political parties, addressing the difficulties of understanding political sentiment on social media. The research also distinguishes tweets from real human users and those from chatbots, employing the K-Means method to detect suspicious activity. To gain deeper insights into the analysis, OpenAI GPT is employed for dataset labelling and managing class imbalance. The results show that sentiment varied significantly over time, with the fine-tuned RoBERTa model providing the most accurate analysis. The results further indicated that most tweets came from real human users, with a small number from bots, which tended to be negative. The findings offer useful insights for shaping political campaigns based on public sentiment trends.

Keywords: GPT, Local elections, Sentiment analysis, Suspicious patterns, User classification, Unsupervised learning

1. Introduction

Social media has become a significant force in South African politics, amplifying public discourse against a backdrop of political unrest, economic challenges, and corruption (Ledwaba & Marivate, 2022; Ngcamu, 2019; Mamokhere, 2019). Social media platforms like X (formerly Twitter) provide a space for free expression (Darad & Krishnan, 2023; Endsuy, 2021), making them essential tools for political parties to connect with voters and examine public sentiment (Pinto & Murari, 2019). However, analysing social media data for political insights is challenging (Singhal et al., 2015; Elbagir & Yang, 2019) due to privacy issues and the vast amount of unstructured data, which complicates extracting meaningful information (Kumar, 2023).

This study addresses these challenges by using advanced methods to analyse sentiments expressed on Twitter during the 2021 South African local elections. Liu (2020) defines sentiment analysis (SA) as evaluating the views, emotions, reviews, behaviours of people, and sentiments expressed in written communication. It employs the twitter-roberta-base-sentiment-latest (TRBSL) model (Liu, et al., 2019) and lexicon-based models: Valence Aware Dictionary for sEntiment Reasoner (VADER) and TextBlob. The TRBSL model is notable for its unsupervised pre-training and task-specific fine-tuning, making it highly effective for sentiment analysis that is based on the Robustly Optimised BERT pre-training (RoBERTa) techniques, even without labelled data. VADER and TextBlob provide a complementary lexicon-based sentiment analysis without being explicitly trained with the labelled data. A novel aspect of this research is the use of OpenAI's Generative Pre-trained Transformer 3.5 (GPT-3.5) for labelling data and managing class imbalances, demonstrating the flexibility and efficiency of few-shot learning (FSL). This allows the study to bypass the labour-intensive process of data labelling, revealing hidden sentiment patterns and offering a more comprehensive analysis.

The study understands the influence of sentiments shared on social platforms and the influence of automated accounts on political discussions, particularly during critical electoral periods. This study addressed the following research questions:

1) What were the variations in sentiment polarity among four political parties during the South African local elections campaign (September—October 2021) and what are its implications for political campaigns?

2) How did the sentiment of Twitter users evolve throughout the 2021 local government elections campaign, and what were the sentiment shifts observed during different phases?

3) What are the classifications of Twitter users based on their political party affiliations, and what insights can be derived from these classifications?

The study focuses on sentiments towards four major political parties (African National Congress (ANC), Democratic Alliance (DA), ActionSA, and Economic Freedom Fighters (EFF)) and examines how these sentiments evolved during the local government election campaign. To offer objective insights, this study also differentiates between tweets generated by humans and those generated by chatbots (bots) or machines. By integrating unsupervised learning and advanced Natural Language Processing (NLP) models, this study offers methodological advancements in sentiment analysis and valuable insights for political campaign strategies. The findings highlight the potential of these techniques to better understand public sentiment and inform decision-making in political contexts.

2. Literature review

Twitter and other social media platforms have evolved into significant information sources for various application methods to analyse data which include SA, opinion mining, and social media monitoring (Al-Shabi, 2020; Elbagir & Yang, 2019). However, given their briefness, informal language, and use of emojis, and hashtags, Twitter texts pose obstacles for NLP (Hutto & Gilbert, 2014). 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 SA of the Twitter dataset regarding the 2021 South African local elections.

An overview of earlier research on sentiment analysis (SA) and user classification (UC) are conducted in this Section. The remainder of the study is structured as follows: in Sections 2.1, 2.2, and 2.3, studies on SA are reviewed, along with the various analysis techniques, and the key findings and the gaps. 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, the key findings and the gaps. Lastly, the summary of the main takeaways and the gaps, together with the limitations addressed in this study are given in Section 2.5.

2.1. Sentiment analysis using social media datasets

The area of SA using social media datasets has seen dynamic growth and evolution, driven by a need to adapt to changing language usage, online trends, and platform-specific capabilities. Several studies have significantly contributed to understanding public sentiment in diverse contexts.

Several studies have contributed to the understanding of SA in diverse contexts. Abiola et al. (2023) delved into public perceptions during the 2019 COVID-19 outbreak in Nigeria, utilising TextBlob and VADER. They argued for the necessity of diverse SA techniques, revealing nuanced perspectives during health crises. In contrast, Bengesi et al. (2023) extended SA to the monkeypox outbreak, emphasising the potential of specific features such as TextBlob annotation and support vector machine (SVM) for achieving higher accuracy. Their study presented practical implications for shaping effective public health strategies. Darad and Krishnan (2023) employed advanced deep learning models, highlighting the superior performance of BERT in analysing sentiment during the 2019 COVID pandemic peak on Twitter. Their argument underscored the challenges in achieving perfect sentiment prediction accuracy due to the intricate nature of social media discourse. Alabrah et al. (2022) shifted focus to vaccine resistance in Gulf nations, using SA methods and LSTM for classification. They contended that the findings offered valuable insights for categorising sentiments in the context of COVID-19 immunisation debates. Illia et al. (2021) explored Indonesian sentiments during the COVID-19 outbreak, advocating for the efficiency of VADER's lexicon approach for semantic analysis on social media. Their study suggested practical improvements for enhancing the reliability of sentiment analysis. In a different vein, Mustaqim (2020) conducted sentiment classification related to politics and religion, revealing distinct clusters of sentiments. Mustaqim (2020) argued for the complexity of sentiment in public perception, emphasising the importance of nuanced approaches in sentiment analysis.

These studies collectively demonstrate the efficacy of SA techniques in understanding public perceptions during health crises, particularly through social media platforms like Twitter. Insights gleaned from these studies provide valuable information for policymakers and healthcare professionals in formulating effective crisis management strategies. However, several gaps remain which involve geographic variability, temporal dynamics, and methodological rigour. Most studies focus on specific regions, limiting the generalizability of findings to broader global contexts. Moreover, limited exploration of sentiment variations over different stages of a crisis hampers a comprehensive understanding of public perception dynamics. The manual preprocessing and translation of data as well as limited model validation, also underscore the need for more automated and validated methodologies in SA. Furthermore, limitations on neutral sentiment label identification (Ledwaba & Marivate, 2022) and model training are emphasised, highlighting the difficulties in capturing subtle emotions. Addressing these gaps will enhance the robustness and applicability of SA in understanding public perceptions during health crises, thereby facilitating more informed decision-making and crisis-management strategies. This study addressed some of the gaps identified in this section, focusing on SA for local government elections in South Africa through unsupervised learning. It acknowledges the limitations associated with using only positive and negative sentiment labels by introducing the neutral sentiment label.

2.2. Sentiment analysis in a political context

In the realm of SA, the contemporary surge in attention is undeniable, driven by the challenges posed by the unstructured nature of online interactions (Ashir, 2021). In the political context, SA emerges as a crucial tool, with increasing utilisation in election campaigns globally. Diverse political studies contribute nuanced perspectives on sentiments, preferences, and challenges within this dynamic field.

Oyewola et al. (2023) focused on the Nigerian presidential election in 2023, employing three models: peephole LTSM (PLSTM), LSTM, and two-stage residual LTSM (TSRLSTM). Their study argued that the TSRLSTM model exhibited exceptional performance in sentiment classification, providing valuable insights for researchers and decision-makers. However, they acknowledged the necessity for model improvement and emphasised the importance of larger and more diverse datasets to enhance understanding. In contrast, Shevtsov et al. (2023) delved into the November 2020 United States presidential election, concentrating on Twitter and YouTube. Their analysis revealed that positive sentiment was stronger for Donald Trump than for Joe Biden, utilising the VADER algorithm. The study argued that real-life events significantly influenced social media conversations. Additionally, the authors proposed future research on sarcasm detection through crowd-sourcing methodologies.

For a deeper understanding of the political climate in South Africa, Ledwaba and Marivate (2022) utilised a semi-supervised approach with a graph-based method to classify Twitter sentiments during local government elections using a Twitter (now called "X") dataset. Their study argued that most Twitter users in South Africa expressed concerns about political parties, with the governing party (ANC) receiving the most negative feedback. However, the study acknowledged limitations in recognizing tweets with neutral sentiments. In contrast, Endsuy (2021) conducted a comparative analysis between exploratory data analysis (EDA) and SA regarding the United States election. The study argued that neutral sentiments dominated, with EDA and VADER providing relatively accurate results. Endsuy (2021) suggested that further work should focus on enhancing the effectiveness of these methods for analysing election outcomes. The analysis of Mustaqim (2020) regarding politics and religion on Twitter underscored the existence of distinct clusters of positive, negative, and neutral sentiments. The study argued that words like "hate" and "radical" frequently appeared in negative sentiment, while "like" and "god" were associated with positive sentiment, revealing different polarities in public perception. Moreover, Elbagir and Yang (2019) conducted SA on Twitter data related to the 2016 US election, utilising NLTK and VADER. The study argued that VADER's sentiment analyser was an appropriate option for fast and accurate sentiment classification, despite limitations such as a small dataset and the use of a generic vocabulary. Furthermore, Nandi and Agrawal (2016) advocated for a hybrid method in politics, combining SVM with the Vocabulary approach to overcome limitations and leverage strengths. Their study argued that this approach achieved high accuracy, suggesting future research directions, including the creation of a multilingual sentiment classification model.

These studies highlight the importance of SA in understanding public opinion in political contexts. Insights from these studies offer valuable information for political campaigns, decision-making processes, and election outcome predictions. However, several gaps remain which include the data limitations, methodological refinement, and cross-language analysis. Many studies acknowledge the limitations imposed by relatively small or unbalanced datasets, emphasising the need for larger and more diverse data sources. Furthermore, while existing methods like VADER and deep learning models show promise, further refinement is necessary to improve accuracy, particularly in handling complex linguistic structures and nuances in political discourse. There are also limited studies addressing SA in multilingual contexts, indicating a gap in understanding public sentiment across diverse linguistic backgrounds. Addressing these gaps will contribute to the advancement of SA techniques in political contexts, enabling a more accurate and nuanced understanding of public opinion and enhancing the effectiveness of political strategies and decision-making processes. This study introduces the use of the GPT model for labelling data to improve the accuracy and further understanding of the models. Moreover, this study also addresses challenges such as noise, stopwords, emoticons, and class imbalance during data preprocessing, by using the TRBSL model, and the lexicon models VADER and TextBlob.

2.3. Methodological approach to sentiment analysis

This section of the literature review emphasises the importance of model selection, and comprehensive data source interpretation, providing insights into various methodological approaches, techniques, and tools used in sentiment analysis.

The innovative approach of Ashir (2021) deserves particular attention, as it ingeniously combines rule-based and lexical procedures with unsupervised machine learning. The incorporation of emoticon detection, word contraction expansion, and noise removal illustrates the meticulous attention to detail required for handling sources with limited syntactic structures. Notably, the study of Ashir (2021) not only improved SA performance but also demonstrated adaptability across various organised sources. In contrast, Mujahid et al. (2021) explored SA on e-learning tweets and highlighted the prowess of deep learning methods. Their comparison of TextBlob, VADER, and SentiWordNet reveals VADER's superiority in handling social media data, attributed to its specialised design for such analyses. The study also underscores the importance of feature extraction methods, with TF-IDF and Bag of Words contributing to the assessment metrics, including F1-score, recall, accuracy, and precision. Moreover, Gujjar and Kumar (2021) proposed using the TextBlob technique for market research, which is deemed cost-efficient and useful for business intelligence but acknowledges its limitations in handling emoticon-rich and biased data. The study not only serves as a practical recommendation but also provides insight into potential areas for improvement, suggesting the adoption of supervised or unsupervised learning for emojis in future applications. Furthermore, Al-Shab (2020) evaluated SA lexicons on Twitter data which offered a detailed comparison, emphasising VADER’s ability in categorising positive and negative attitudes, especially in handling short phrases and social media content. The direct use of lexicons without preprocessing adds an intriguing dimension to the findings, showcasing the robustness of VADER in its raw form. In contrast, Hutto and Gilbert (2014) evaluated VADER's performance on Twitter texts against advanced benchmarks, including human raters, and reinforced the tool's exceptional accuracy and F1-score. The study’s emphasis on a “gold standard” lexicon list tailored for sentiment in weblog contexts sheds light on the meticulous development process required for achieving high-performance results.

The above studies highlight the importance of methodological rigour and careful consideration of various techniques and tools in SA. Insights from these studies provide valuable guidance for researchers and practitioners in selecting appropriate approaches for analysing sentiment across different sources and domains. However, several gaps remain which include tool comparison, domain-specific analysis, interpretability and explainability. While some studies compare the effectiveness of different SA tools and techniques, further research is needed to comprehensively evaluate their performance across various datasets and domains. Furthermore, many studies emphasise the importance of domain-specific features for improved SA accuracy. Future research could explore the development of domain-adaptive SA models to enhance performance across diverse domains. Despite achieving high accuracy rates, some SA models lack interpretability and explainability. Future research could focus on developing transparent models that provide insights into the reasoning behind sentiment classifications. Addressing these gaps will contribute to the advancement of SA methodologies, enabling more accurate, interpretable, and domain-adaptive SA across various applications and domains. This study used basic statistical metrics such as mean, median, and standard deviation. These metrics were applied to sentiment labels to evaluate model performance. Additionally, a confusion matrix, precision, recall, F1-score, accuracy, and average measures were employed to assess and compare the models used in the study.

2.4. Identification of bots and spam tweets

The literature review explores the challenges posed by bot accounts on social media platforms, particularly X, where the anonymity of users allows bots to imitate real users and influence the network dynamics. Bots are social media accounts that are automated, they have been demonstrated to propagate false information and control online conversations (Rossetti & Zaman, 2023). Studies emphasise the need for effective identification of bots and extraction of insights from social media data, especially during events like local elections. Bots can distort public opinion, participate in political discussions, and hinder the evolution of public policy.

One study by Alarfaj et al. (2023) focuses on identifying automated Twitter users using content features and machine learning. Deep learning algorithms, combined with specific feature sets, are found to be more accurate in classifying bots. The inclusion of image analysis is suggested to enhance bot detection. Moreover, the study of Genfi (2021) delves into COVID-19 related bot-generated content, employing machine learning tools to identify bots and false information. Their study emphasises the significance of topic distribution and SA in discriminating between bot and human accounts. In contrast, Alothali et al. (2021) propose a hybrid feature selection method to identify effective features for detecting bots. Their study highlights the importance of cross-validation attribute evaluation and explores the use of advanced deep-learning techniques for bot identification. The study of Vasterbo (2020) compares ML approaches for labelling tweets as human or bot-generated, emphasising the need for labelled datasets reflecting recent bot behaviour. In contrast, the systematic literature review of Abkenar et al. (2020) discusses evaluation criteria, methodologies, and tools for Twitter spam detection. Class imbalance in datasets is identified as a challenge, impacting classification performance. Furthermore, Kudugunta and Ferrara (2018) propose a deep neural network for bot detection at the tweet and account levels. Their approach achieves high accuracy rates, even with limited labelled data. In contrast, Washha et al. (2017) present an automated, unsupervised approach for predicting fake name behaviours and spam activities on Twitter, benefiting users who manage large tweet volumes. Moreover, Dickerson et al. (2014) focused on sentiment features to differentiate between human and bot users during the 2014 Indian election, achieving improved accuracy in bot detection.

These studies provide insights into various methods and techniques for identifying bots and spam tweets on Twitter. Machine learning and deep learning algorithms, combined with feature selection methods, SA, and metadata analysis, have shown promise in accurately classifying bot-generated content. However, challenges such as class imbalance, generalizability to unseen data, and the evolving nature of bot behaviour remain. Addressing these challenges requires further research into developing robust classification models, collecting labelled datasets with recent bot-generated tweets, and improving evaluation methodologies to ensure accurate performance assessment. Overall, the identification of bots and spam tweets on Twitter is an ongoing research area with significant implications for mitigating the spread of disinformation and preserving the integrity of online discourse. Continued efforts in this field are essential to developing effective detection methods and safeguarding the trustworthiness of social media platforms. This study addressed the challenges regarding the class imbalance and additionally, used the K-Means clustering method and suspicious patterns to classify users as either bots or humans.

2.5. Summary

The review highlights the importance of SA in assessing public sentiment on social media, provides insights into using social media datatsets in political contexts, emphasises methodological considerations, and sheds light on distinguishing spam tweets and bots. The study further addresses the gaps identified in the sections, focusing on SA for local government elections in South Africa through unsupervised learning. It acknowledges limitations associated with using only positive and negative sentiment labels and introduces the use of the GPT model for labelling data for further understanding of the performance of the models. Moreover, the study also addresses challenges such as noise, stopwords, emoticons, and class imbalance during data preprocessing, employing the pre-trained TRBSL model, and unsupervised lexicons VADER and TextBlob models. Additionally, the K-Means clustering method and suspicious patterns are used to classify users as either bots or humans. The research concludes by highlighting ethical procedures for using social media datasets, with detailed methods discussed in the next section.

3. Methodology

The description of the dataset, exploratory data analysis, preprocessing, approach, evaluation measures, ethical considerations and tools are outlined in this section.

3.1. Data

3.1.1. Data description and exploratory data analysis (EDA)

This study uses raw Twitter data collected before the 2021 South African local government elections, spanning from September to October 2021. The dataset includes tweets mentioning political party leaders, election-related hashtags, and political party hashtags. Initial data exploration addressed challenges such as privacy settings, duplicates, and null values. The dataset has been reduced from 727 083 tweets to 56 993 after being filtered for relevant political and election content.

3.1.2. Preprocessing

Managing unstructured texts, especially tweets with elements like hyperlinks and emoticons, poses a significant challenge in NLP. Since the selected models do not have built-in features for removing hyperlinks, mentions, and hashtags, the following preprocessing steps were performed to ensure that the data is ready for analysis:

Cleaning: Removed retweets, replies, usernames, symbols, non-letters, punctuation, non-hashtags, hyperlinks, outliers, and missing values.

Tokenization and Stopword Removal: Tokenised tweets and removed stopwords using a comprehensive list of 707 stopwords, which is a combination of forty (40) default Natural Language Toolkit (NLTK) stopwords and 667 from Ranks NL website[1].

Filtering: Limited tweets to those with more than ten words and set the language to English, further reducing the dataset size. The time series plot in Figure 3.1 illustrates the variation in tweet counts over time for different political parties. Notably, around 21 September 2021 and before 31 October 2021, there is a surge in tweets, reaching approximately 2 500 tweets, with the ANC being the most frequently discussed political party, followed by the EFF, DA, and ActionSA as the least talked-about party.

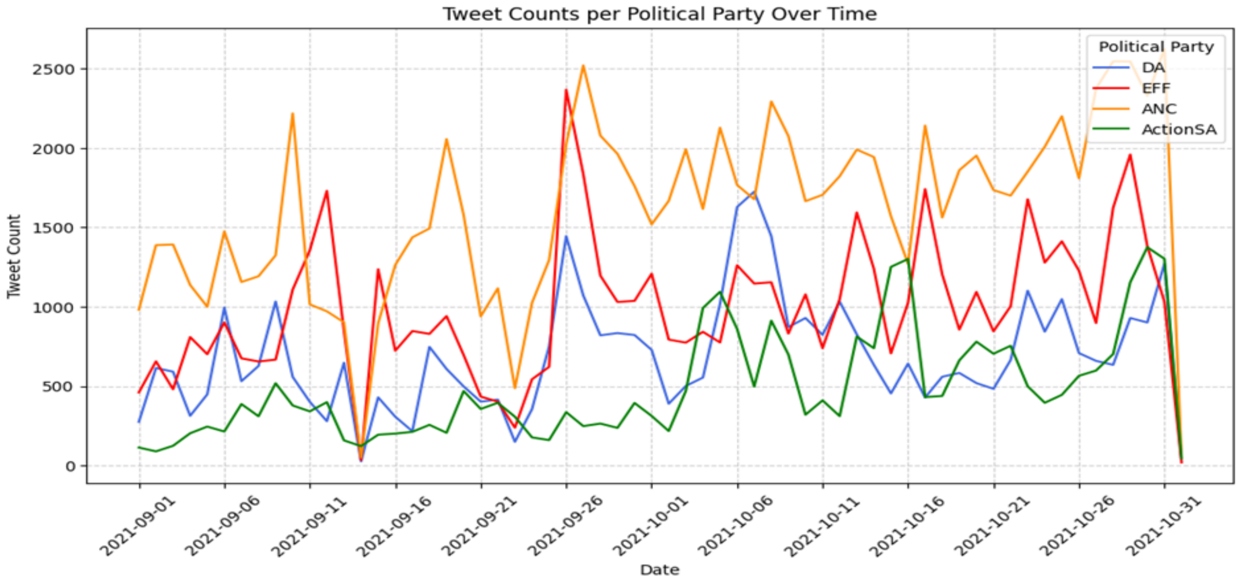
Ethical Considerations: Hidden usernames and user IDs, and paraphrased tweets to protect user privacy. The study maintained high ethical standards in handling Twitter datasets by prioritising user privacy through data anonymisation, which ensured that all personally identifiable information was removed, ensuring data security by storing the data in a password-protected location and striving for fairness and unbiased analysis by adjusting weights to balance the classes to reduce potential biases in the results of the study.

3.2. Approach

This study employs three sentiment analysis models which are TRBSL, VADER and TextBlob to analyse and get sentiments based on tweets about the local government elections. The models categorised the tweets as negative, neutral, and positive based on polarity scores. To enhance model performance and address the presence of bots, the study uses detection techniques as outlined by Washha et al. (2017) and the K-Means clustering model to classify users as human or bot. Additionally, GPT-3.5-turbo chat is used for labelling the dataset by employing FSL for labelling, to improve class balance compared to the initial analysis with the unlabelled dataset. The FSL is used to give the model a few instances that are then labelled as neutral, negative, or positive. The labelled dataset consists of 2 195 negative, 2 026 neutral, and 1 779 positive tweets. The ability of GPT-3.5 to understand natural language, code and generate text outputs (Cribben & Zeinali, 2023) is leveraged for the labelling task. Despite the cost associated with labelling training datasets (Bach, et al., 2019), models trained on GPT-3 often match or exceed human-annotated data performance (Ding, et al., 2022). This study explicitly discloses efforts to address the potential impacts and limitations of utilising the GPT-3.5 model for data labelling. The methodology used in this study is illustrated in Figure 3.2.

[1] More information regarding Ranks NL can be found using this link: - https://www.ranks.nl/stopwords

Figure 3.1: Tweet count per political party over time

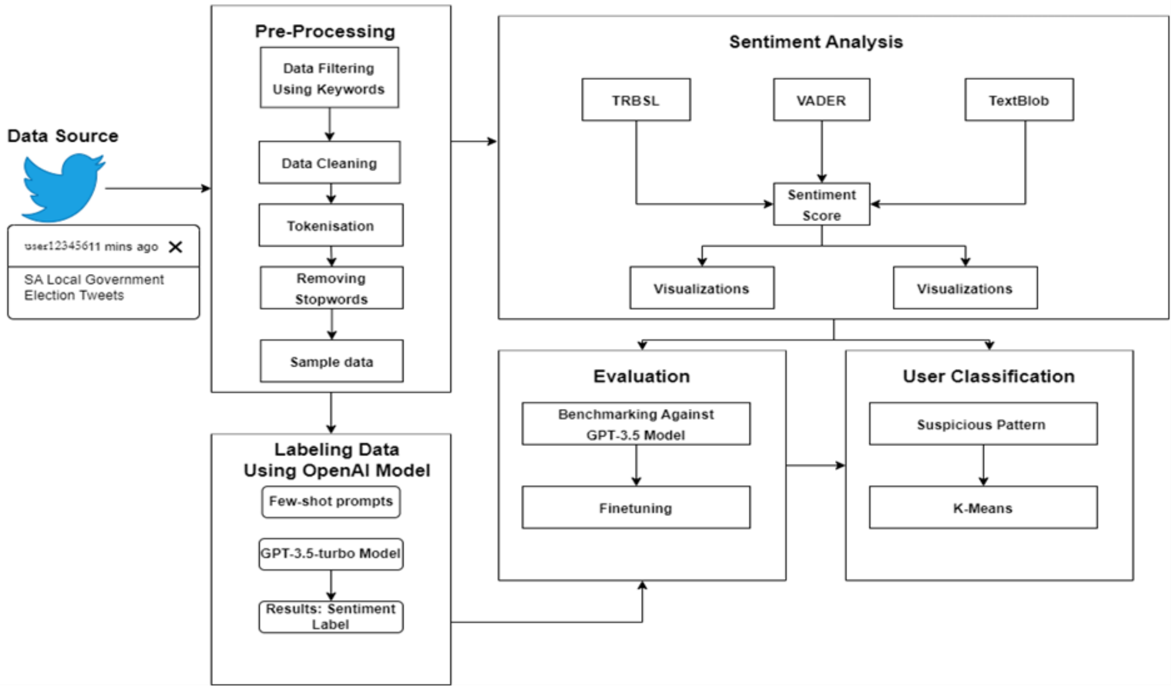


3.3. Evaluation metrics and tools

The performance of the models in this study is evaluated using mean, median, standard deviation, confusion matrix, precision, recall, F1-score, and accuracy. These metrics ensure a comprehensive assessment of sentiment classification accuracy and reliability. Importantly, the study acknowledges potential enhancements in data quality through GPT-3.5 annotation compared to human-annotated data. Python serves as the programming tool for analysis and visualisation in this study due to its user-friendly nature and comprehensive libraries for SA.

Overall, this methodology section provides a clear understanding of the dataset, the exploratory data analysis, the preprocessing, the approach, the fine-tuning process, the ethical approach, and the tools. The analysis and results from user classification and SA using unsupervised learning for the South African local government elections are presented in the next Section.

Figure 3.2: Methodological flow diagram



4. Results analysis

First, sentiment analysis is conducted using a subset of the preprocessed dataset. Secondly, fine-tuning is implemented along with the adjustment of weights to balance the classes in the model that initially underperformed. This fine-tuning process utilises results from the GPT-3.5 model to train, evaluate, and test the improved model.

4.1. Unsupervised sentiment analysis and results

4.1.1. TRBSL model

The TRBSL model reveals a potential class imbalance in the sentiment analysis. The distribution of the tweets for the model is summarised in Table 4.1 which indicates a strong positive sentiment, yet signs of model limitations in labelling positive tweets are evident.

Table 4.1: TRBSL sentiment analysis count

Sentiment Label	Count
Neutral	2 936
Positive	414
Negative	1 650

Table 4.2: TRBSL statistical analysis metrics

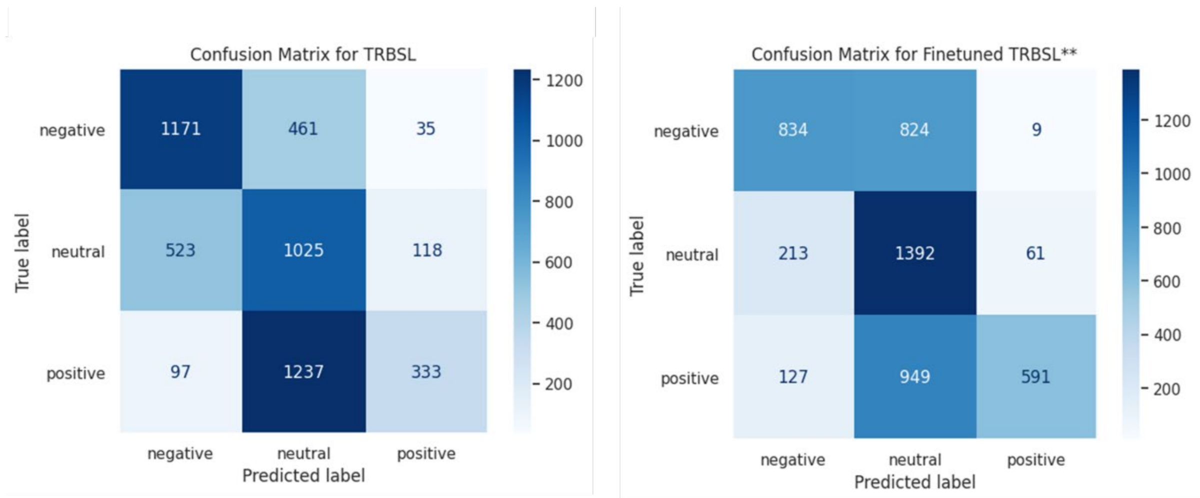
Metric	Score
Mean	0.7705
Median	0.8030
Standard Deviation	0.1320

The key statistical analysis presented in Table 4.2 indicates a positively skewed sentiment as many tweets are extremely positive and there are minimal variations around the mean. The comparison between TRBSL and GPT-3.5 models suggests disagreement in sentiment counts, notably in positive sentiments where TRBSL performs less effectively. The confusion matrix (Figure 4.1) indicates the model's ability to predict instances, but the classification report summary in Table 4.3 highlights imbalances in the TRBSL model. To enhance the results, fine-tuning with the GPT-3.5 labelled dataset improved accuracy to approximately 69%, demonstrating an improved balance across sentiments. The dataset is split into three sets illustrated in Table 4.4 for fine-tuning.

*Table 4.3: Classification report summary for TRBSL and TRBSL** models*

Metric	TRBSL Model	TRBSL** Model (Fine-tuned Model)
Accuracy	0.51	0.69
F1-score	0.48	0.69
Precision	0.57	0.71
Recall	0.51	0.61

Figure 4.1: Confusion Matrix for TRBSL and fine-tuned TRBSL Models



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 demonstrated an improvement. The evaluation accuracy and the evaluation 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. The sentiment class performance for the model (Table 4.5) demonstrated an improved overall accuracy of 69% across all the classes, indicating that approximately 69% of all the predictions are correct.

The polarity sentiment analysis indicates that most positive tweets by users are directed at the dominant party, the ANC, followed by the EFF, the DA, and ActionSA However, most negative tweets are directed against the ANC, implying that it is both the most liked and possibly the most disliked party. Overall, the model is performing better on the test dataset. Further improvement on the model is recommended for better results.

Table 4.4: Fine-tuning dataset

Dataset	Split Percentage	Sample Size
Training Set	70%	4200
Testing Set	15%	900
Validation Set	15%	900

Table 4.5: The class performance for the model

Sentiment Label	Accurate Predictions	F1-Score
Neutral	1392	0.81
Positive	591	0.62
Negative	834	0.66

4.1.2. VADER model

The sentiment analysis in this section employs the VADER model, revealing the illustrated sentiment classification summarised in Table 4.6. The sentiment analysis indicates that the dataset has more positive tweets compared to negative and neutral tweets.

The statistical analysis results are summarised in Table 4.7. The mean score indicates that the sentiment is slightly positive on average, whereas the median score suggests that there is an equal count of negative, positive, and neutral sentiment scores, and the standard deviation suggests that the sentiment scores are relatively variable. 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.

Table 4.6: VADER sentiment analysis count

Sentiment Label	Count
Neutral	1 389
Positive	2 103
Negative	1 508

Table 4.7: VADER statistical analysis

Metric	Score
Mean	0.0531
Median	0.0000
Standard Deviation	0.4743

The statistical analysis results are summarised in Table 4.7. The mean score indicates that the sentiment is slightly positive on average, whereas the median score suggests that there is an equal count of negative, positive, and neutral sentiment scores, and the standard deviation suggests that the sentiment scores are relatively variable. 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.

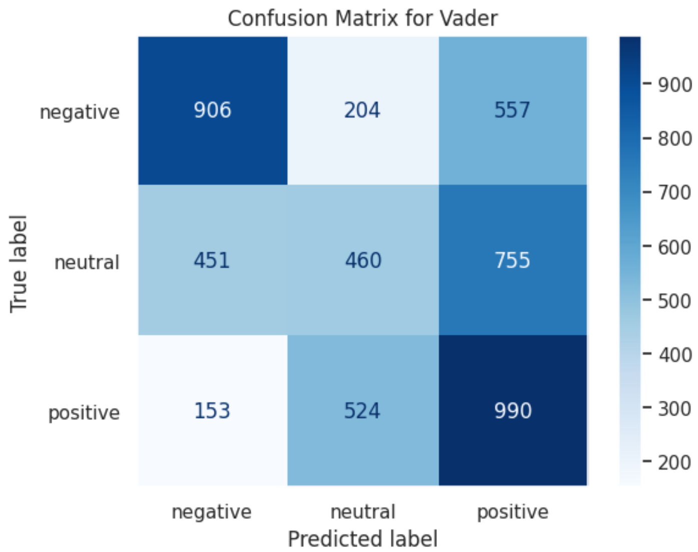
Table 4.8: Classification report summary for the VADER model

Sentiment Class	Precision	Recall	F1-Score
Neutral	0.39	0.28	0.32
Positive	0.43	0.59	0.50
Negative	0.60	0.54	0.57
Accuracy			0.47

The confusion matrix (Figure 4.2) demonstrates VADER's classification results, with challenges in accurately predicting the neutral class. The classification report is summarised in Table 4.8 revealing the precision, recall, and F1-score metrics, highlighting a reasonable balance for negative and positive labels but a challenge in predicting neutral instances. The overall accuracy is 47%, with better performance for negative and positive labels than for neutral. In summary, VADER's performance varies across classes, suggesting potential improvements for better results, although computational limitations prevent further enhancements in this study.

The polarity sentiment distribution among political parties indicates varying sentiments, with the ANC leading in positive tweets. The SA results between the VADER and GPT-3.5 models also suggest that both the models slightly disagree on sentiment counts, i.e., there is a variation between the two models. In summary, the performance of the model varies across all the classes, with negative labels and positive labels having relatively better performance compared to neutral labels. This indicates that to get better results for the model, more improvements are required on the model. Due to the computational powers, the model is not further improved for this study.

Figure 4.2: Confusion Matrix for VADER Model



4.1.3. TextBlob model

The sentiment analysis in this section employs the TextBlob model, revealing a class imbalance in the illustrated sentiment classification summarised in Table 4.9. The sentiment analysis shows that the high number of neutral tweets points to possible challenges with the classification of the model.

Table 4.9: TextBlob sentiment analysis count

Sentiment Label	Count
Neutral	1 028
Positive	1 536
Negative	2 436

Table 4.10: TextBlob statistical analysis

Metric	Score
Mean	0.0350
Median	0.0000
Standard Deviation	0.2690

The statistical analysis results are summarised in Table 4.10 which indicates a slightly positive average sentiment, a balanced distribution of all the sentiments, and a moderate variability in the sentiment scores. The confusion matrix in (Figure 4.3) highlights misclassifications, especially in predicting positive sentiment as neutral.

The classification report is summarised in Table 4.11 revealing the precision, recall, and F1-score metrics, highlighting 42.56% neutral predictions and a slight difference in predicting negative sentiments. The model further demonstrates a moderate performance, however it has challenges in accurately classifying positive tweets. The TextBlob model reveals clear polarity variations and distinct sentiment patterns across the parties. The ANC dominates all sentiment distributions, recording the highest counts of neutral, positive, and negative sentiments. While the TextBlob model shows potential for improvement, computational constraints and time limitations prevent further enhancements in this study.

Figure 4.3: Confusion matrix for TextBlob models

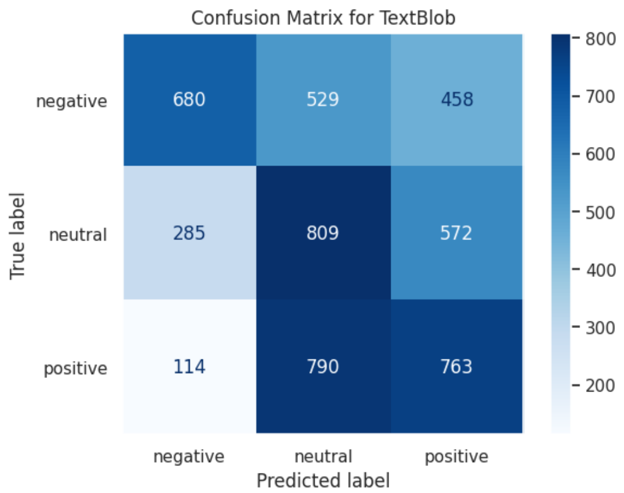


Table 4.11: Classification report summary for TextBlob model

Sentiment Class	Precision	Recall	F1-Score
Neutral	0.38	0.49	0.43
Positive	0.43	0.46	0.44
Negative	0.63	0.41	0.50
Accuracy			0.45

4.2. User classification analysis and results

This study employs two techniques for user classification. Firstly, suspicious patterns are utilised, revealing the summarised details in Table 4.12. All bot-generated tweets undergo manual verification.

Table 4.12: User classification report summary for suspicious patterns

Tweets	No of Tweets
Human-generated	4 975
Bot-generated	21
Irrelevant	4

Subsequently, the study applies a K-Means clustering model based on suspicious patterns results, determining an optimal cluster number of three as summarised in Table 4.13. This approach uncovers additional bot tweets not identified by suspicious patterns (Table 4.14). The tweets which are neither human nor bot-generated, are therefore excluded from the analysis as they are deemed as spammy and non-political tweets.

Table 4.13: K-Mean clustering using 3 clusters

Cluster	No of Tweets
Cluster 0	4 451
Cluster 1	68
Cluster 2	480

Table 4.14: Final user classification report summary

User Class	No of Tweets
Human	4 906
Bots	90
Irrelevant	4

To further understand the classification of the users, Figure 4.4 displays the classification of users into various categories, such as "bot", "human", and "not relevant" along with the counts of tweets falling into political party categories such as "DA", "EFF", "ANC" and "ActionSA". Table 4.15 highlights the sentiment distribution by user class summary in the three models.

Figure 4.4: User classification per political party

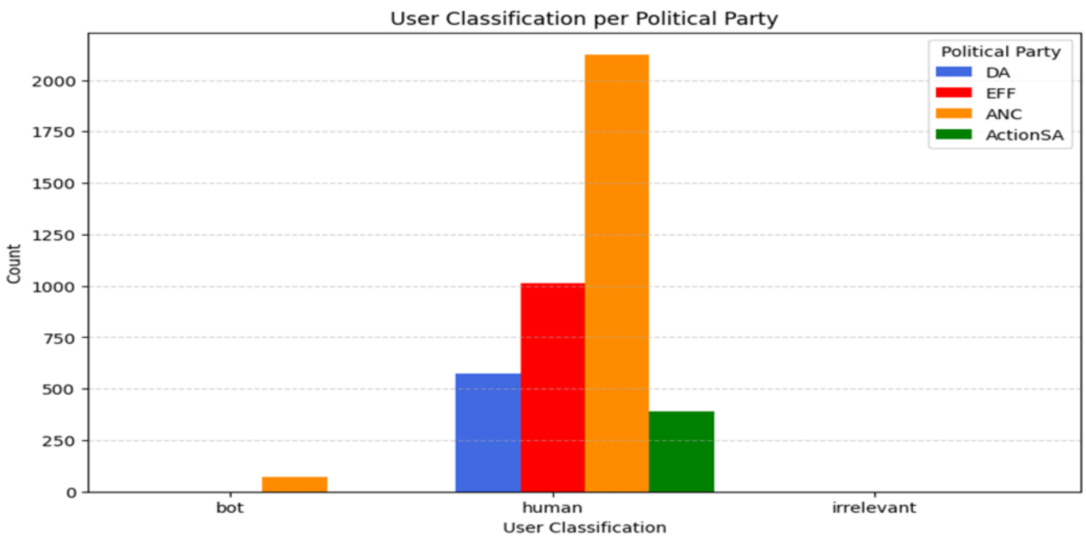


Table 4.15: Sentiment distribution by user class

Model	User Class	Negative	Neutral	Positive
TRBSL**	Human	811	3 652	443
	Bots	69	20	1
VADER	Human	1 436	1 379	2 091
	Bots	72	7	11
TEXTBLOB	Human	956	2 423	1 527
	Bots	72	9	9

The polarity sentiment breakdown for each political party (DA, EFF, ANC, ActionSA) in Table 4.16 suggests varying distributions among user classes. Human-generated tweets dominate all the political parties. The ANC has received the most negative bot-generated tweets, followed by DA, while EFF shows no bot-generated tweets in both TRBSL** and TextBlob models. Despite the TRBSL and TRBSL** models showing high averages and relatively low standard deviations, indicating consistent sentiment scores, their weighted accuracy is higher. VADER and TextBlob models exhibit better performance, though tweet labelling reveals class imbalance, addressed through class weight application across all models. Therefore, the results reveal significant polarity sentiment variations among the four political parties during the 2021 South African municipal election campaign, with diverse user sentiment expressions detected. The next Section discusses the results of this analysis, how they compare with other studies and the limitations faced in this study.

Table 4.16: Polarity sentiment by political party per model

Model	Political Party	User Class	Negative	Neutral	Positive
TRBSL**	ANC	Human	485	1 468	172
		Bots	0	2	0
	DA	Human	115	449	8
		Bots	0	1	0
	EFF	Human	217	741	55
		Bots	0	0	0

	ActionSA	Human	52	303	33
		Bots	0	1	0
VADER	ANC	Human	822	532	771
		Bots	69	0	0
	DA	Human	220	98	254
		Bots	0	0	1
	EFF	Human	292	267	454
		Bots	0	2	2
	ActionSA	Human	93	119	176
		Bots	0	1	1
TEXTBLOB	ANC	Human	544	605	605
		Bots	70	1	1
	DA	Human	151	175	175
		Bots	0	0	1
	EFF	Human	177	330	330
		Bots	0	0	0
	ActionSA	Human	68	124	124
		Bots	0	0	0

5. Results discussion

5.1. Sentiment analysis

The analysis reveals significant differences between the sentiment analysis models used in the study:

TRBSL and TRBSL:** These models indicate that neutral sentiments had higher mean and median sentiment scores (Table 5.1) which shows that the model predicts more neutral sentiments and fewer positive sentiments. Table 5.2 indicates that the standard TRBSL model exhibits high weighted accuracy and F1-score, performing well in classifying tweets with a good balance between overall accuracy and precision recall. The improved TRBSL** model surpasses the standard TRBSL, demonstrating significantly better performance with high accuracy and F1-score.

VADER: The result of this model indicates slightly positive sentiments on average, but further showed the greatest variability in scores (Table 5.1). Furthermore, the model achieves moderate accuracy and F1-scores, excelling in classifying negative and positive tweets but struggling with neutral sentiments (Table 5.2).

TextBlob: This model leans towards neutral and slightly positive sentiments, displaying moderate variability (Table 5.1). Furthermore, the model has moderate accuracy and an acceptable F1-score, similar to that of the VADER model (Table 5.2).

Table 5.1: Summary of overall statistical distribution for SA

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 5.2: Weighted accuracy and F1-score for all models

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

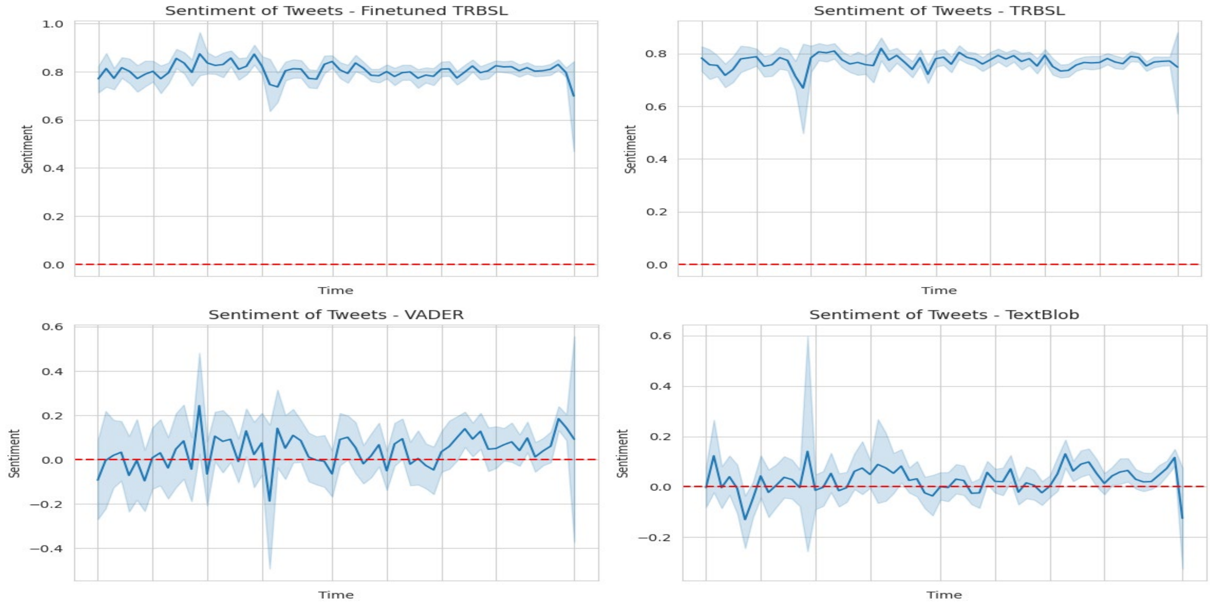
5.2. Sentiment trends and user classification

The time series plot demonstrates significant sentiment variations over time for both VADER and TextBlob, indicating fluctuating user sentiments throughout the election campaign period. Figure 5.1 illustrates the sentiment trend over time, using a dashed red line to signify the neutral point. Values above the dashed line signify positive sentiments, whereas below signify negative sentiments, and those parallel are neutral. The solid blue line shows the average sentiment score over time, whereas the shaded blue area indicates the variations in sentiment scores. The TRBSL model shows consistently high positive sentiment with minimal variability but with slightly more fluctuations. Similarly, the fine-tuned TRBSL model indicates that the sentiment scores are consistently high and positive, with minor variations, indicating a strong positive sentiment over time. Contrary to the TRBSL models, the sentiment scores for VADER show a fluctuation around the neutral line, indicating a mix of positive and negative sentiments, with higher variability. Furthermore, the sentiment scores for TextBlob suggest a more balanced sentiment with occasional strong positive or negative sentiments. The study further identifies two user classes, revealing variations in user classifications across all parties. A significant number of negative tweets directed at the ANC are found to be originating from bots.

5.3. Comparison with related work

SA has become popular in political campaigns over the years where political parties use social media data to analyse the sentiment of the users to use the results to put some corrective measures in place. Unlike previous studies that use single models, this study explores multiple models (TRBSL, VADER, TextBlob) and introduces a neutral sentiment label, in contrast to previous studies, particularly the work by Ledwaba and Vukosi (2022). Consistent with previous findings by Ledwaba and Vukosi (2022), the ANC is the target of most negative tweets, and positive and neutral sentiments are also predominantly about the ANC. Moreover, in agreement with the study of

Figure 5.1: Trend analysis for all models over time



Ledwaba and Vukosi (2022), the findings of this study demonstrate that most tweets are negative, followed by neutral, and finally positive. The use of the GPT-3.5 model for dataset labelling and the employment of unsupervised learning techniques set this study apart from others. Labelling datasets may be difficult, time-consuming, and costly. Despite its own set of challenges, the GPT-3.5 model made the labelling of the dataset easier using FSL. Bots play a significant role in spreading negative sentiment, highlighting the importance of accurately classifying user types. Furthermore, the use of identifying suspicious patterns and using the K-Means algorithm to categorise Twitter users based on the dataset as either real humans or bots, also adds to the uniqueness of this study. The results show that the selected methods perform well in identifying spam tweets which were manually verified.

5.4. Implications

The findings offer valuable insights for political campaigns, suggesting that understanding sentiment patterns can inform strategic decisions. The methodology of this study highlights the effectiveness of combining multiple sentiment analysis models and unsupervised learning techniques for robust political sentiment analysis. The identification of bot-generated content emphasizes the need for advanced methods to ensure unbiased sentiment analysis and accurate user classification in political contexts.

6. Conclusion and limitations

6.1. Conclusion

This study explored the concerns regarding the influence of bots on election campaigns through social media, emphasising the need to identify genuine users and understand their behaviours. Political parties must understand public perceptions to be able to strategically execute their election campaigns according to the expectations of the community. The primary goal of this study was on SA, 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 sentiment and standard deviations, VADER (mean = 0.0964, std_dev = 0.4848) and TextBlob (mean = 0.0460, std_dev = 0.2578) out-performed TRBSL (mean = 0.7629, std_dev = 0.1333) and TRBSL** (mean = 0.7976, std_dev = 0.1551) models in the comparison of the four SA models. Furthermore, the study compared four SA 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 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.

6.2. Limitations

Several limitations impeded the progress of this analysis:

Dataset Issues: The use of an unsupervised dataset labelled with GPT-3.5 presented challenges such as class imbalance and instability in the labelling process. Future work could involve using larger, more balanced datasets and exploring alternative labelling techniques to improve stability and accuracy. In addition, adding human verification at various points might improve the accuracy of the labelled data.

Outdated Data: The dataset, collected during the 2021 election, may not reflect current political sentiments as the country approaches the 2024 elections. Future studies could use more recent data to ensure the findings are relevant to the current political climate. Continuous data collection and real-time analysis could also provide more dynamic insights.

Computational Constraints: Only the TRBSL model was fine-tuned due to high computational power and time limitations. Addressing these constraints in future work could involve leveraging cloud-based solutions and more efficient algorithms to handle larger datasets and multiple models simultaneously. Collaborating with institutions that have greater computational resources might also mitigate these limitations.

Verification: The author of this study personally confirmed sentiment labelling for only 1000 tweets, with no external support or political experience. Future research could include external validation from multiple experts to enhance the credibility and accuracy of the sentiment labels. Establishing partnerships with political analysts or domain experts can provide more nuanced insights into the data.

Future research can expand on the existing findings by dealing with these limitations and provide deeper and more accurate user classification and sentiment analysis within the framework of political contexts.

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. Implications suggest parties could use SA to improve interaction methods and address societal problems. Future work involves further model improvement, addressing class imbalance, and considering code-switched tweets in the analysis. Recommendations highlight the benefits of SA for political parties in gaining community trust during election campaigns. It's important to note that this study's analysis is based on a Twitter sample dataset and may not represent the sentiments of the entire South African population.

References

Abiola, O., Abayomi-Alli, A., Tale, O. A., Misra, S., & Abayomi-Alli, O. (2023). Sentiment analysis of COVID-19 tweets from selected hashtags in Nigeria using VADER and Text Blob analyser. *Journal of Electrical Systems and Information Technology*, 1--20.

Abkenar, S. B., Kashani, M. H., Akbari, M., & Mahdipour, E. (2020). Twitter spam detection: A systematic review. *arXiv preprint arXiv:2011.14754*, arXiv--2011.

Alabrah, A., Alawadh, H. M., Okon, O. D., Meraj, T., & Rauf, H. T. (2022). Gulf countries' citizens' acceptance of COVID-19 vaccines—A machine learning approach. *Mathematics*, 467.

Alarfaj, F. K., Ahmad, H., Khan, H. U., Alomair, A. M., Almusallam, N., & Ahmed, M. (2023). Twitter Bot Detection Using Diverse Content Features and Applying Machine Learning Algorithms. *Sustainability*, 6662.

Alothali, E., Hayawi, K., & Alashwal, H. (2021). Hybrid feature selection approach to identify optimal features of profile metadata to detect social bots on Twitter. *Social Network Analysis and Mining*, 1--15.

Al-Shabi, M. (2020). Evaluating the performance of the most important Lexicons used for Sentiment analysis and opinions Mining. *IJCSNS*, 1.

Ashir, A. M. (2021). A Generalized Method for Sentiment Analysis across Different Sources. *Applied Computational Intelligence and Soft Computing*, 1--8.

Bach, S. H., Rodriguez, D., Liu, Y., Luo, C., Shao, H., Xia, C., . . . others. (2019). Snorkel drybell: A case study in deploying weak supervision at industrial scale. In *Proceedings of the 2019 International Conference on Management of Data* (pp. 362--375).

Bengesi, S., Oladunni, T., Olusegun, R., & Audu, H. (2023). A Machine Learning-Sentiment Analysis on Monkeypox Outbreak: An Extensive Dataset to Show the Polarity of Public Opinion From Twitter Tweets. *IEEE Access*, 11811--11826.

Cribben, I., & Zeinali, Y. (2023). 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)*.

Darad, S., & Krishnan, S. (2023). Sentimental analysis of COVID-19 twitter data using deep learning and machine learning models. *Ingenius. Revista de Ciencia y Tecnolog'ia*, 108--117.

Ding, B., Qin, C., Liu, L., Bing, L., Joty, S., & Li, B. (2022). Is gpt-3 a good data annotator? *arXiv preprint arXiv:2212.10450*.

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

Endsuy, R. D. (2021). Sentiment analysis between VADER and EDA for the US presidential election 2020 on twitter datasets. *Journal of Applied Data Sciences*, 08--18.

Genfi, E. K. (2021). Detecting Bots Using a Hybrid Approach. *Theses, Dissertations and Culminating Projects*.

Gujjar, J. P., & Kumar, H. (2021). Sentiment analysis: Textblob for decision making. *Int. J. Sci. Res. Eng. Trends*, 1097--1099.

Hutto, C., & Gilbert, E. (2014). 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).

Illia, F., Eugenia, M. P., & Rutba, S. A. (2021). 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).

Kudugunta, S., & Ferrara, E. (2018). Deep Neural Networks for Bot Detection. *Information Sciences*, 312--322.

Kumar, N. (2023). Harnessing the Power of Big Data: Challenges and Opportunities in Analytics. *Tuijin Jishu/Journal of Propulsion Technology, 44*, 2.

Ledwaba, M., & Marivate, V. (2022). 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).

Liu, B. (2020). *Sentiment analysis: Mining opinions, sentiments, and emotions.* Cambridge university press.

Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., . . . Stoyanov, V. (2019). Roberta: A robustly optimized bert pretraining approach. *arXiv preprint arXiv:1907.11692*.

Mamokhere, J. (2019). An exploration of reasons behind service delivery protests in South Africa: A case of Bolobedu South at the Greater Tzaneen Municipality. *International Conference on Public Administration and Development Alternatives (IPADA)*, 9.

Mujahid, M., Lee, E., Rustam, F., Washington, P. B., Ullah, S., Reshi, A. A., & Ashraf, I. (2021). Sentiment analysis and topic modeling on tweets about online education during COVID-19. *Applied Sciences*, 8438.

Mustaqim, T. (2020). 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).

Nandi, V., & Agrawal, S. (2016). Political sentiment analysis using hybrid approach. *International Research Journal of Engineering and Technology*, 1621--1627.

Ngcamu, B. S. (2019). Exploring service delivery protests in post-apartheid South African municipalities: A literature review. *The Journal for Transdisciplinary Research in Southern Africa*, 9.

Oyewola, D. O., Oladimeji, L. A., Julius, S. O., Kachalla, L. B., & Dada, E. G. (2023). Optimizing sentiment analysis of Nigerian 2023 presidential election using two-stage residual long short term memory. *Heliyon*.

Pinto, J. P., & Murari, V. (2019). Real time sentiment analysis of political twitter data using machine learning approach. *International Research Journal of Engineering and Technology (IRJET)*, 4124--4129.

Shevtsov, A., Oikonomidou, M., Antonakaki, D., Pratikakis, P., & Ioannidis, S. (2023). What Tweets and YouTube comments have in common? Sentiment and graph analysis on data related to US elections 2020. *Plos one*, e0270542.

Singhal, K., Agrawal, B., & Mittal, N. (2015). 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).

Stopwords. (2023). Retrieved from https://www.ranks.nl/stopwords

Vasterbo, S. (2020). CLASSIFYING TWITTER BOTSA comparasion of methods for classifying whether tweets are written by humans or bots.

Washha, M. Q., Mezghani, M., & Sèdes, F. (2017). 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).

About the Author

Penelope Matloga

Penelope Matloga works as a senior specialist in revenue assurance for a private sector based in Johannesburg, South Africa. She holds a Masters degree in Data Science from the University of Pretoria, a BSc honours in Statistics from the University of South Africa and an undergraduate BSc degree in Mathematical Sciences from the University of Limpopo.

Vukosi Marivate

Prof Vukosi Marivate is a Professor of Computer Science and holds the ABSA UP Chair of Data Science at the University of Pretoria. He specialises in developing Machine Learning (ML) and Artificial Intelligence (AI) methods to extract insights from data, with a particular focus on the intersection of ML/AI and Natural Language Processing (NLP). His research is dedicated to improving the methods, tools and availability of data for local or low-resource languages. As the leader of the Data Science for Social Impact research group in the Computer Science department, Vukosi is interested in using data science to solve social challenges. He has worked on projects related to science, energy, public safety, and utilities, among others. Prof Marivate is a co-founder of Lelapa AI, an African startup focused on AI for Africans by Africans. Vukosi is a chief investigator at the Masakhane Research Foundation, which aims to develop NLP technologies for African languages. Vukosi is also a co-founder of the Deep Learning Indaba, the leading grassroots Machine Learning and Artificial Intelligence conference on the African continent that aims to empower and support African researchers and practitioners in the field.

Kayode Olaleye

Kayode Olaleye is a Postdoctoral Fellow in the Computer Science department at the University of Pretoria, South Africa. He holds a PhD in Electronic Engineering from the University of Stellenbosch, where his research focused on using images to train speech models for low-resource African languages. Currently, his work explores speech and language processing for under-resourced African languages, aiming to make language technologies more accessible and inclusive. Kayode is driven by a commitment to supporting linguistic diversity and bridging gaps in technology for underserved communities.