IIETA International Information and Engineering Technology Association
Advancing the World of Information and Engineering

Analysis of Peoples' Opinion on Democratic Election in A Developing Nation Using Textblob Lexicon-Based Approach

Check for updates

Temitope Elizabeth Ogunbiyi[1] , Adedayo F. Adedotun[2*] , Johnson Adeleke Adeyiga[1] , Moses Joy Achas[1] , Abass I. Taiwo[1] , Abiodun A. Opanuga[2] , Onuche G. Odekina[2] , Osahon V. Adoghe[2]

[1] Department of Computer Science and Information Technology, Bells University of Technology, Ota 11001, Nigeria
[2] Department of Industrial Mathematics, Covenant University, Ota 11001, Nigeria

Corresponding Author Email: adedayo.adedotun@covenantuniversity.edu.ng

ABSTRACT

The field of Sentiment Analysis involves utilizing computational methods to identify and understand the emotional aspects present in a text and expressing emotions through written language. Hence, the objective of this research is to examine and analyze people's opinions regarding the Nigerian presidential election for a deeper understanding of their preferences, and concerns of the electorate on the person declared as the winner of the election. Using the Tweepy library and Twitter Application Programming Interface (API), 85,662 tweets were collected using some specific keywords and hashtags. The tweets were preprocessed using Natural Language Tool-kits and were analyzed using TextBlob lexicon-based processing techniques. The results showed that the Positive attitudes were 48%, neutral attitudes were 32%, and negative attitudes were 23%. Insights into popular sentiment and political inclinations that were gained from the sentiment analysis showed that most individuals had a positive reaction toward the presidential election for the betterment of their nation. In general, this research shows the importance and effect of sentiment analysis in determining people's opinions towards the conducted election.

1. INTRODUCTION

Nigeria, one of the developing nations in Africa with over 200 million population, has a lot of capacity for economic growth, due to the large part of the 1958 discovery of significant oil reserves in the Niger Delta area of the country. Nigeria is endowed with a lot of natural resources that can boost its economic development, and many people both inside and outside the country believe that Nigeria would soon rise to occupy a major position in Africa and international affairs [1]. On October 1, 1960, Nigeria became independent from the British colonial administration. At that time, the future of the nation seemed bright and expectations were high. Throughout its history, Nigeria's system of government has undergone numerous changes. The nation switched to a parliamentary form of government after gaining independence, with the prime minister serving as the head of state. This system, however, was not long-lasting because several military takeovers in the 1960s and 1970s resulted in prolonged periods of military control. Nigeria did not revert to civilian administration until 1999, at which point a presidential system of government was put in place. Since then, the nation has chosen its presidents through routine elections which take place every four years.

Nigeria, the most populous nation in Black Africa and the self-described "giant of Africa," has improved its election system through a number of initiatives over colonial and post-colonial times, leading to electoral reforms in the country. The election management body (EMB) is in charge of supervising electoral administration in Nigeria and has undergone numerous changes or adjustments under different regimes to improve its election procedures. Nigeria's electoral reforms date back to the colonial era, when the elective principle was first incorporated into the political system in 1922 by the Clifford Constitution. Elections were held for four elective seats in the Nigerian Legislative Council, which was maintained and expanded by later colonial governments [2]. The Richards Constitution of 1946 created a central legislature for the nation, with three representatives from Lagos and one from Calabar, and four of the 24 members elected to the central legislature [3]. The first national election, with the exception of Lagos, was made possible by the Macpherson Constitution of 1951. Only in the northern region, where the election was indirect, was the first widespread direct election made possible by the Oliver Lyttleton Constitution of 1954 [4]. To conduct the 1959 elections and the nation's first direct elections, the departing colonial government established the Electoral Commission of Nigeria (ECN) in 1959 as a special electoral body [2]. Nigeria's electoral provisions order-in-council of 1958 stipulated that the governor general could designate an electoral commission from among neutral-minded individuals [5]. The Federal Electoral Commission (FEC), which was led by a Nigerian and held the country's first federal and regional elections in 1964 and 1965, was renamed the electoral umpire after Nigeria gained independence [6].

In 1976, General Olusegun Obasanjo's military government

implemented a fresh election reform, replacing the FEC with the Federal election Commission (FEDECO). The 1979 and 1983 elections were held by FEDECO [7]. Also, the military government led by General Ibrahim Babangida instituted the National election Commission (NEC) in 1987 as part of yet another election reform. Elections for local government chairmen, lawmakers, governors, and presidents were held by NEC in 1990, 1991, and 1993, respectively. The 1995 establishment of the National Electoral Commission of Nigeria (NECON) by General Abacha's military government was another electoral reform initiative in Nigeria, spurred by the annulment of the presidential elections and other problems. The current Independent National Electoral Commission (INEC) was established in 1998 after General Abacha's death, when General Abdulsalami Abubakar dissolved the NECON. Seven elections have been held by INEC, the longest-serving EMB in Nigerian history: the 1999 transition election, the 2003 election, the 2007 election that resulted in the nation's first civilian regime change, the 2011 election [8], the 2015 election, the 2019 election, and the general elections that were just held in February 2023. Additionally, INEC has implemented a number of reforms over the years with the goal of strengthening democratic processes. The 1999 constitution (as modified), the Electoral Act of 2010 (as amended), and 2022 each govern elections in Nigeria.

Since 2011, there have been several modifications, such as the implementation of a permanent voter card, an electronic smart card reader, a simultaneous accreditation and voting system, and stakeholder participation prior to, during, and following all elections, including off-season ones [9]. Nigeria's democratic credentials were put to the test during the general election in 2023. While the governorship and state assembly elections were held on March 18, 2023, the presidential and national assembly elections were held on February 25, 2023. A year prior to the election, a new Electoral Act was passed, which authorised the electoral commission to use election technology, including electronic results transmission, and called for the early transfer of election funding to INEC. The Act enabled INEC to implement new technologies for the general election, including the INEC Election Results Viewing Portal (IReV) and the Bimodal Voter Accreditation System (BVAS). Interest in the election was raised by INEC's deployment of better technology for voter registration, particularly online pre-registration to vote. One positive aspect of the 2023 general election is the widespread implementation and high functionality rate of the BVAS for voter accreditation in the February 25 and March 18 elections.

The Nigerian presidential election is a highly anticipated moment with enormous political significance for the nation. Social media platforms like Facebook, Twitter, Instagram, and many others have significantly influenced how public opinions and political engagement have changed over the past years. A popular social media site called Twitter has emerged as an important tool for public dialogue, information exchange, and communication. It is impossible to ignore Twitter's impact in Nigeria. The website has developed into an essential forum for Nigerians, including activists, politicians, and citizens, to express their concerns, share news, inspire social movements, participate in political debates, and discuss pressing national issues by analyzing the language, tone, and context of tweets, voting patterns, and political preferences [10]. These platforms have developed into effective tools for political dialogue providing people with the opportunity to communicate their

thoughts and feelings in an uncensored manner in the digital era. Twitter in particular has become well-known as a forum where people actively discuss politics, making it a useful source for sentiment research. One can learn a great deal about the opinions and attitudes of the public about different candidates, political parties, and important issues by examining Twitter feeds on the Nigerian presidential election. The majority of the decisions made in our daily lives are influenced by data, particularly data from social media sites like Twitter [10], and have been used to keep track of and predict elections. Twitter and other social media channels provide a wealth of real-time information that may be utilized to assess public opinion [11] and opinions of political candidates and parties [12]. Twitter users in particular constantly hunt for material to consume or share. Being a unique social networking site that makes it simple to start trends around topics by using hashtags rather than individual posts, which are prevalent on other social media sites. As a result, it is easier to keep track of users' opinions on particular subjects and their moods. During a study aimed at exploring the potential of the Twitter-sphere, researchers stumbled upon an unexpected finding. They discovered that political hashtags and the associated discussions exhibited the highest level of persistence among all the categories analyzed on Twitter [13].

The use of sentiment analysis is just one of the many methods that scholars have used to analyse opinions [14], others are natural language processing, and machine learning [15] to evaluate Twitter data and forecast the results of elections. Several techniques have been developed to track and forecast elections using information from Twitter [16]. Therefore, sentiment is an opinion, attitude, or assessment brought on by feeling. Sentiment analysis, commonly referred to as opinion mining, is a computational approach used to discern the emotional underpinnings of a text. It involves analyzing written language to identify and extract people's ideas, feelings, attitudes, and emotions expressed within the text [17]. Sentiment analysis has been used to reveal peoples' opinion on elections in developing nations. Examples include the IndoBERT model and Twitter data that were used [18] to do a sentiment analysis of Indonesian public opinion on the 2024 election, and the study [19] computed sentiment scores by compiling tweets from major political parties that ran in the Gujarat Assembly Elections of 2022.

The primary purpose of sentiment analysis is to ascertain a speaker's or writer's viewpoint on certain topics or the overall contextual polarity of a work of literature. It establishes whether there are any positive, negative, or neutral feelings connected to a piece of writing. Sentiment analysis models are now frequently used in political methodology, starting with public political discourse [20]; this is consistent with the study of quantitative and qualitative methodologies used to understand politics and political systems [21]. Numerous studies have provided methods for comprehending politics that usually include formal theory, mathematics, and statistics. Election results have historically been anticipated analytically and statistically, with the technique depending on surveys and qualitative methodologies, researching political party manifestos, and keeping a watch on media trends [22].

The Nigerian presidential election of 2023 demonstrated the active involvement of individuals all around the nation, especially the lively youth population and knowledgeable elders including citizens who resided Overseas, this is due to their belief that they can change the bad economy of Nigeria through voting in the right candidate. Their participation, even

on social media sites like Twitter, demonstrated their common ideas to tackle and ensure that Nigeria elects a qualified and deserving leader. However, there have been diverse expressions towards the outcomes of the general elections in Nigeria. Most of these views were expressed by Nigerians on cyberspace via Twitter, Facebook, Instagram, and so on. Therefore, there is a need to know the opinions of people towards the general election, to determine the most frequently used hashtags, and know the locations of the most active users. The 2023 Nigerian presidential election presents challenges in determining public opinion due to the diverse ideas of voters and the dynamic nature of sentiment. However, social media platforms like Twitter provide a valuable means for individuals to express their thoughts on candidates and the election results. Invariably, this work is focused on knowing the opinions of different categories of the populace regarding the outcome of the election. The scope of this study is to analyze tweets for positive, negative, or neutral sentiment for the 2023 Nigerian presidential election. Twitter served as the main data source for this project, which helped in collecting and analyzing a sizable number of tweets about the election. Focusing on voter sentiment, public opinion, and the connections between sentiment and election-related variables, the work investigates the sentiment trends, patterns, and correlations. By contrasting the results of the sentiment analysis with other measures of public opinion, the study evaluates the findings and debates their implications for the Nigerian presidential election in 2023. To clearly and understandably convey the sentiment analysis results, the project made use of data visualization approaches.

1.1 Adoption of new technologies versus traditional methods of election

Over the past 10 years, political campaigning has moved from offline—which includes traditional media and physical locations—to online. A model that has been drastically expanded upon and changed subsequently, the 2008 Obama campaign is sometimes credited as being the first to use social media mobilisation for electoral purposes. Additionally, because social media platforms have broken down some of the obstacles to engaging in democratic debate, some groups of people now have a comparatively easier time participating in conversations about elections and candidate accountability. First of all, it is important to acknowledge that social media platforms have upended the dominance of traditional media organisations, which has helped a lot of people by allowing new groups to participate in the democratic debate who were previously shut out of those structures. From the perspective of an election campaign, this has helped independent candidates and minor parties who didn't have the funds to run campaigns using conventional methods. Despite these benefits, it must be recognised that some groups frequently map onto preexisting social, economic, and political hierarchies, and that not everyone is similarly situated in digital domains.

People assume that everyone has access to the internet or digital gadgets since election debate and campaigning have moved online. This is undoubtedly not the case, especially in the developing world, where access to the internet is still very limited and varies by class, race/ethnicity, gender, (dis)abilities, language, digital skill levels, and occasionally caste and religion. The International Telecommunication Union (ITU) estimated that 2.7 billion people worldwide were not connected in 2022, which is about half of the world's population, according to the UN [23]; the majority of these individuals are in developing nations [24]. Individuals without access to the internet continue to experience "a worsening cycle of disenfranchisement" as digitisation and digital changes expand annually. There is still a "Gender Digital Divide" in the world, with 69% of men and 63% of women using the internet [24]. The disparity is considerably more pronounced in "least developed countries," where it stands at 13%. The number of women who own a mobile phone is 131 million lower than that of men [23]. Furthermore, the "gender gap in meaningful digital use" is even more pronounced, as studies show that women use digital devices less frequently due to a lack of technical skills and, in certain countries, familial pressures, which affects their capacity to engage in online politics. Political engagement is affected by these disparities, which also affect who has access to crucial voting information. Even if social media "creates value by attracting the attention of mainstream media," it has the effect of only reaching an elite audience in nations with low digital density. This feature is likely to impede efforts to enhance grassroots organizing [25].

The direct use of technology to conduct elections is worth examining individually, even though social media platforms' influence on electoral politics is multifaceted. Election management technologies have been used by states, political parties, and civil society. First of all, governments and political parties use technology extensively to raise awareness and register voters. While there may be some overlap with social media campaigning and influence strategies, election management explicitly focusses on the use of technology to disseminate information about election procedures. Voter registration lists, voting station locations, polling timetables, and polling results are all examples of this. To increase voter turnout and enhance voter registration, governments worldwide have looked to information, communication, and technology (ICT). Social media and electronic platforms have been utilised by Election Management Board (EMBs), civil society, and governments in general to broaden voter education and outreach. Social media in particular has been utilised in several nations to raise voter education among young voters, many of whom are first-time voters [26]. To raise awareness and help voters register, EMBs have used SMS services in other nations with limited internet and social media access. It is good to say that; the world is evolving and the use of ICT is one of the major paradigms bringing this change on a daily basis. Therefore, the adoption of ICT in electoral processes need to be embraced for easier and fast processing of electoral procedures.

2. LITERATURE REVIEW

The review of literature in this study delves into the effect of ethnic diversity in Nigeria, techniques used in sentiment analysis, and related works.

2.1 The effect of ethnic diversity in Nigeria towards presidential selection

The convergence of multiple ethnic groups and the depth of their cultural legacies lay the groundwork for Nigeria's identity, which is a diversified tapestry. The major ethnic groups in each region and those competing for domination at the federal level were the Hausa/Fulani, Yoruba, Igbo, and Niger Delta

regions in the Northern, Western, Eastern, and Southern zones, respectively. Ethnic minorities frequently dispute the main ethnic groups' political power within each region [27]. They felt more and more cut off from the political system as a result, furthering the division of their identities and impeding the development of a strong, unifying Nigerian national identity. In the early 1960s, when it became clear that there was a lack of national cohesion, Nigerians addressed the problem. Academics, artists, and even some politicians have made attempts to establish a distinct Nigerian culture through their writings, articles, speeches, and laws. It was promoted that Nigeria would have a strong central government and a state-run economy that would give development projects priority. These programs all attempted to bring Nigerians closer together politically, economically, and culturally by highlighting commonalities and downplaying differences. Ultimately, though, these initiatives failed, partly due to the overwhelming political trend toward regional power consolidation at whatever cost [1]. Therefore, utilizing Twitter promotes national harmony by acting as a bridge between Nigeria's numerous ethnic groupings. By examining opinions posted on Twitter, one can learn more about the hopes of Nigerians as a whole, regardless of regional differences.

2.2 Techniques of categorizing sentiment

Sentiment analysis has been addressed at various granularity as a task in natural language processing. Before being handled at the phrase level [28, 29], it was first treated at the document level, then at the sentence level. To account for more precise results and classify emotions into clear groups like fear, grief, rage, enthusiasm, and pleasure, sentiment analysis with ambiguous handling might be used [30].

The three basic techniques for identifying and categorizing emotions in text. They are:
- Lexicon-based approach
- Methods based on machine learning
- Hybrid techniques [31]

Lexicon-Based Approach

The basic concept underlying the lexicon-based technique involves the following steps:
a. Breaking down sentences into individual words.
b. Comparing these words to the entries in the sentiment lexicon, considering their associated semantic relations.
c. Computing the overall polarity score of the entire text based on these comparisons.

These techniques are excellent in identifying whether a text has a positive, negative, or neutral tone [31]. The lexicon-based technique assigns semantic orientation to words using either a dictionary-based or corpus-based approach. The dictionary-based approach is simpler, wherein the polarity score of specific words or phrases can be determined in the text by referring to a sentiment dictionary that contains opinion keywords. The description of lexicon-based approaches is given as follows:

Approach Using a Built-In Lexicon. TextBlob and VADER (Valence Aware Dictionary and sEntiment Reasoner) are popular Python lexicon-based sentiment analysis models [32]. TextBlob uses averaging, providing polarity, subjectivity, and intensity scores. VADER, a rule-based tool, excels with informal language on social media due to relevant corpora. Sentiment orientation in VADER is determined by positive, negative, neutral, and compound values obtained through

heuristic rules and standardization. Each word in the sentiment lexicon is assigned scores for positive, negative, and neutral sentiments (-4 to +4). The compound score of a sentence is calculated with heuristic rules, considering punctuation, capitalization, modifiers, conjunctions, and negations. A formula is used to standardize word scores to a range of -1 to +1 as seen in Eq. (1) [33].

$$x = \frac{x}{\sqrt{x^2 + a}} \tag{1}$$

where, x is the total of the sentiment words' Valence scores and is a normalizing factor. By adding up the results from each standardized lexicon in the range of 1 (the most negative) to 1 (the most positive), the compound score is calculated. Table 1 displays the precise classification standards for TextBlob and VADER [34].

Table 1. The TextBlob and VADER classification thresholds

Sentiment Analysis Technique	TextBlob Score	Sentiment Orientation
TextBlob score	The polarity score > 0	Positive
	The polarity score < 0	Negative
	The polarity score $= 0$	Neutral
VADER compound score	The compound score $>= 0.05$	Positive
	The compound score $<= -0.05$	Negative
	The compound score > -0.05 and < 0.05	Neutral

Lexicon-Based Approach with SentiWordNet

The lexicon-based SentiWordNet approach uses WordNet Database to analyze sentiment. SentiWordNet contains lemmas organized in synsets, each with positive and negative polarity scores (ranging from 0 to 1). The SentiWordNet-based method is applied in multiple steps. It begins with data pre-processing, including cleansing, tokenization, stemming, and Part-of-Speech (POS) tagging for efficient searching in the SentiWordNet database [35]. When a tweet has a lemma with multiple interpretations, only the polarity score of the most common meaning is considered using specific Eqs. (2) and (3).

$$PosScore = PosScore1 \tag{2}$$

$$NegScore = NegScore1 \tag{3}$$

The emotion polarity scores of each tweet can be calculated by counting the number of positive and negative phrases within the tweet. The sentiment score of a word or a specific phrase in the SentiWordNet lexicon can be obtained by applying Eq. (4).

$$SynsetScore = PosScore - NegScore \tag{4}$$

The SynsetScore then calculates the absolute value of the top positive and negative scores ever recorded. In the case of a phrase with many synsets, the calculation is made as presented in Eq. (5).

$$Term\ Score = \frac{\sum_{n=1}^{k} SynsetScore\ (r)/r}{\sum_{n=1}^{k} 1/r} \tag{5}$$

For phrases not found in SentiWordNet, the overall score is recorded as 0. In the presence of negations, the sentiment value

is reserved based on the specified number of synsets (k). The final sentiment score for each tweet is determined by summing the sentiment scores of all phrases using Eqs. (6)-(8) [35].

$$PosScore\ (s) = \sum_{i=1}^{m} TermScore(T_i) \qquad (6)$$

$$NegScore\ (s) = \sum_{i=1}^{n} TermScore(T_i) \qquad (7)$$

$$SentiScore\ (s) = PosScore\ (s) + NegScore\ (s) \qquad (8)$$

where p is a simple tweet that has m positive and n negative words. $NegScore\ (s)$ represents the negative terms, $SentiScore\ (s)$ is the final sentiment score of tweets, and $PosScore\ (s)$ is the total score of all the positive terms.

Methods Based on Machine Learning

Machine learning techniques are utilized to build classifiers that perform sentiment categorization by extracting feature vectors. This process involves stages such as data collection and cleaning, feature extraction, training the classifier with the training dataset, and analyzing the outcomes [36]. The dataset is typically divided into a training dataset and a test dataset for machine learning techniques. The classifier learns text features from the training dataset and its performance is evaluated using the test dataset. Various classifiers, such as Naive Bayes, Support Vector Machines, Logistic Regression, and Random Forest, are employed to categorize text into predefined classes.

Machine learning, a widely adopted approach in text classification, is frequently used by researchers and practitioners. It is important to note that the performance of a single classifier may vary significantly across different types of text. Therefore, it is recommended to train feature vectors separately for each type of text. The process involves vectorizing the Twitter data and splitting the labeled tweet data into a training set (80%) and a test set (20%). Multiple classification models can then be trained to predict the sentiment labels.

Hybrid Technique

This is the combination of lexicon-based techniques and machine learning models. a hybrid approach combines both the lexicon-based and machine learning approaches to exploit the strengths of each method, thus enhancing the performance obtained. The combination is developed by first obtaining the semantic orientations of the tweets given by the lexicon-based approach, and these outputs are used as training data in the machine learning classifiers [37].

2.3 Related works

This section presents research works on sentiment analysis using data from microblog sites.

Pakistan's general election was predicted using social media data to carry out sentiment analyses about political parties [38]. Using supervised machine learning techniques, they were able to classify tweets as good, terrible, or neutral. The outcomes of their study show that social media content can serve as a reliable predictor of a party's political activity. Also, the study [39] aimed to understand the public sentiment towards the potential candidates in the 2020 United States presidential election. They employed a novel technique to initially collect tweets and user profiles, which allowed them to assess the perception of each presidential candidate across various user and tweet categories, including accessible, deleted, suspended, and inaccessible tweets and accounts. By comparing the sentiment scores obtained from different groups, the study provided valuable insights into the differences observed. Notably, the authors found that tweets removed after the election tended to exhibit more pro-Joe Biden sentiment, while tweets removed before the election were more pro-Donald Trump. Additionally, the study highlighted increased support for Joe Biden among older Twitter accounts.

Also, A dataset of 4.9 million tweets was gathered from 18,450 users and their contacts, totaling 4.9 million messages [40]. Researchers studied political homophily on Twitter during the 2016 US presidential election from August to November. Based on their opinions of Donald Trump and Hillary Clinton, they divided users into six categories. The results demonstrated that in each scenario examined, there was homophily among negative users, Trump supporters, and Hillary supporters. Additionally, they discovered that reciprocal linkages, multiplex ties, and comparable utterances all raise homophily.

The work [41] gathered more than 20,000 tweets from Twitter, examined them, and categorized them as good, negative, or neutral to provide a broad perspective on an Indian farmer protest. TF-IDF and Bag of Words were both utilized in the analysis, however Bag of Words fared better. Twitter data was used to investigate how different coronavirus-affected nations handled the situation [42]. The study used 50,000 tweets to evaluate tweets posted in English to determine how people felt and thought about the pandemic in their home countries. Unsupervised and ensemble machine learning models showed superior performance over traditional models in extracting opinions from text [43]. To enhance the sentiment analysis process and effectively detect tweet polarity, they combined various features and utilized ensemble models. This approach allowed for a more comprehensive analysis and improved the accuracy of sentiment classification. Twitter data has been used to demonstrate that social media is utilized to discuss ideas and opinions among users as well as express viewpoints [44]. To forecast the 2009 German federal election, authors utilized 100,000 tweets, which may be used as a political indicator. The study [45] suggests a method that gathers and examines tweets from blogs and uses the data to forecast selection outcomes. It surveyed the public and examined the factors that could influence the outcome of the 2018 presidential election candidates in Turkey. When compared to the actual election results, the final results showed a high prediction percentage of accuracy based on the tweets that were gathered.

Based on the review done, most of the works categorized the opinion of the populace into different classes [38, 41, 44] while some have some opinion about two personalities in US election [40]. In this study, emphasis was placed on the opinion of the electors on the president elect, commonly used hashtags, locations of the most active users, the mood polarity (happy, neural or sad).

Several other researchers have also carried out related works in the area of prediction using supervised machine learning techniques [44-48] for comprehensive applications of these techniques in enhancing cash crop production. Table 2 presents the summary of the related works.

Table 2. Summary of related works

s/n	Authors	Title	Approach	Strength
1	Razzaq et al. [49]	Prediction and analysis of Pakistan election 2013 based on sentiment analysis	Supervised machine learning techniques	They classified tweets into three categories: favorable, negative, and neutral using supervised machine learning techniques.
2	Ali et al. [50]	Large-scale sentiment analysis of tweets about the 2020 US presidential election	Sentiment analysis	Utilize Twitter's API to find tweets and user accounts in our database that Twitter later deleted or suspended.
3	Caetano et al. [51]	Using sentiment analysis to define Twitter political users' classes and their homophily during the 2016 American presidential election	Sentiment analysis	During the 2016 American Presidential Election, they identified six Twitter user categories, each of which comprises a range of potential opinions about Hillary Clinton and Donald Trump.
4	Neogi et al. [52]	Sentiment analysis and classification of Indian farmers' protest using Twitter data.	TF-IDF and Bag of Words	They gathered more than 20,000 tweets from Twitter, examined them, and categorized them as neutral, unfavorable, or favorably.
5	Wang et al. [53]	Public sentiment analysis on Twitter data during COVID-19 outbreak.	Sentiment analysis	They investigated how the various coronavirus-affected nations handled the circumstances using data from Twitter.
6	Bibi et al. [54]	A novel unsupervised ensemble framework using concept-based linguistic methods and machine learning for Twitter sentiment analysis.	Machine learning technique	They examined political homophily using information from Twitter.
7	Tumasjan et al. [55]	Predicting elections with Twitter: What 140 characters reveal about political sentiment.	Machine learning	They used over 100,000 tweets to predict the German federal election in 2009.

3. RESEARCH METHODOLOGY

This study used model research methods which hinged on quantitative and qualitative analysis of the data. To achieve the objectives of this research, four experimental procedures were followed and they were: data collection, data preprocessing, analysis of data, and data visualization. Figure 1 shows the research process flow, which is a step by step guide on how the research was done. Four distinct stages were identified for the realization of the research objectives.
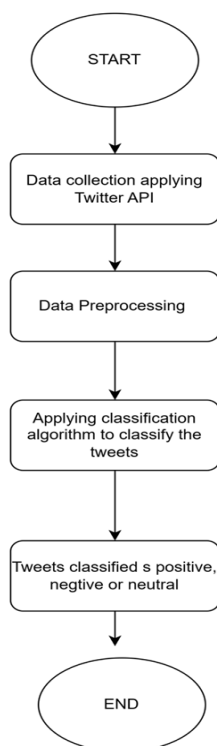
Figure 1. Research process flow

3.1 Method

This study utilised Python 3.8 and TextBlob version 0.17.1. Python 3.8 and TextBlob version 0.17.1 can be a powerful combo for political sentiment analysis, mainly because of TextBlob's compatibility and capabilities with this Python version. Python 3.8 environments may be relied upon thanks to version 0.17.1's stability and strong support. TextBlob uses NLTK to give users access to other capabilities that are essential for sentiment analysis, like tokenisation, stopword elimination, and POS tagging. It easily interfaces with other libraries, such as pandas, making it possible to preprocess and analyse massive text datasets more easily. Even people with no prior knowledge of natural language processing (NLP) can use TextBlob's simple API to process textual input. It has a built-in sentiment analysis function that applies a rule-based methodology based on models that have already been trained. TextBlob is a great tool for testing concepts or small-scale projects because it allows for quick prototyping of sentiment analysis pipelines. You can examine public opinion on political candidates, policies, or events using TextBlob's sentiment score feature, and you can monitor changes in sentiment during elections, debates, and emergencies. Since propaganda and false information are common in political discourse, it is essential to recognise if a piece is subjective or factual.

The predefined sentiment lexicons that TextBlob uses are essentially databases of words with corresponding sentiment scores. The dictionary contains words with subjectivity (a value ranging from 0 (objective) to 1 (subjective) and polarity (a value ranging from -1 (negative sentiment) to +1 (positive sentiment). TextBlob automatically does sentiment analysis using the PatternAnalyzer, which is based on the vocabulary of the Pattern library. After tokenising the input into discrete words or phrases, TextBlob retrieves the polarity and subjectivity values of the tokens by comparing them with entries in the sentiment lexicon. The following natural language processing (NLP) tools and techniques are used:

TextBlob, CountVectorizer, Truncated SVD, Latent Semantic Analysis (LSA), and Sentiment Intensity Analyser (e.g., VADER). When used together, they can enhance processes like political sentiment analysis.

3.2 Data collection

The dataset used in this analysis came from Twitter and included 85,662 tweets about the 2023 presidential election, since there is a need for a more accurate dataset, Twitter data from November 2022 to April 2023 had to be gathered, which indicated the time at which the presidential aspirants were made known to Nigerians and the commencement of discussions about the 2023 presidential elections on Twitter and other social media platforms. The Twitter API was used to stream real-time data alongside the Tweepy package in the Python programming language. Tweepy library is a package in Python that allow for real-time streaming of tweets via Twitter API. Since the analysis being carried out requires a real-time assessment, this account for the time frame which coincide with the election period in Nigeria. Each tweet was collected based on many factors like user_name, user_location, user_ description, user_description, user_created, etc., and the tweets made use of several trending hashtags like #NigeriaDecides2023, #NigeriaPresidentialElection2023, and many more, Nigerian Presidential candidates and tweets by citizens with top electoral preferences. The selection of the hashtag and keywords used for this study was based on the "trending" on Twitter App, individual hashtag common among the Nigeria citizens on their social media handles, and more importantly, selected keywords during the political talks aired on Arise TV, channel Television, TVC Political talks and the likes. A Twitter hashtag connects the conversations of several users into a single stream. This enables unrelated Twitter users to discuss the same subject. Their tweets show in the same stream because of the hashtag they employ. This resolves a coordination issue and helps to start a discourse. Since so many individuals are discussing the subject, some hashtag terms or phrases start to gain popularity. These are referred to as "trending topics."

3.3 Data preprocessing

This stage involves the cleaning of the dataset to reduce noise and increase the effectiveness of text analysis algorithms. The following techniques were used to clean up the tweets:

i. Text processing requires the removal of Unicode strings because some Unicode characters may not be supported by all systems or software, which might result in problems with text encoding and display.

ii. Substitution of blank lines for Re-tweets. In text processing, it's critical to remove the "RT" (re-tweet) prefix from text since it helps to avoid information duplication and lower noise levels in the text data. The effectiveness and precision of text processing algorithms can be increased by removing them and substituting blank spaces.

iii. Change any link to "URL" In text processing, it's crucial to convert URLs (Uniform Resource Locators) to the token "URL" because doing so helps to cut down on background noise in the text data. The main goal of this removal is to get rid of extraneous data and noise that could negatively impact sentiment categorisation systems' effectiveness. Frequently, URLs don't add any

significant sentiment-related information to the textual material. If not managed well, they might cause needless complexity and misclassification. The textual information is made more concise by eliminating URLs, which frees up the sentiment analysis algorithms to concentrate on the linguistic content that actually expresses sentiment. Additionally, special characters like symbols and punctuation can obstruct text analysis. Exclamation marks and other punctuation can convey emotion, but too many or unnecessary special characters can add noise. Removing superfluous special characters aids in text data normalisation, which improves feature extraction and tokenisation. This, in turn, reduces dimensionality and focusses on important features, improving the performance of machine learning models used in sentiment analysis. For text data to be cleaned, noise levels to be decreased, and sentiment analysis models to be more accurate, URLs and special characters must be eliminated.

iv. It is necessary to change any instances of "@username" to "user" because it helps to clean up text data by lowering noise. On social media sites like Twitter, "@username" is the symbol for mentioning or tagging other users. The findings may be skewed by these mentions, which are generally irrelevant to text-processing tasks.

v. Eliminate more white spaces since doing so helps text processing algorithms and machine learning models work more effectively and accurately. The performance of text processing algorithms can be hampered by excessive white spaces, which can also affect text encoding and display issues.

vi. Removal of hashtags placed in front of words. On social networking sites like Twitter, hashtags are used to organize and label tweets, however, they are not always applicable to the given job. By lowering the dimensionality of the data, removing hashtags can aid in enhancing the effectiveness and precision of text processing algorithms and machine learning models. Understanding the purpose of the hashtags is essential for topic model and topic classification. Replacing hashtags with their titles is a simple method of incorporating this information. This can be accomplished by processing the HTML response after performing an HTTP GET.

vii. Remove digits, multiple exclamation points, and numerous question marks because they can skew and bring noise to text data. Numbers may not be relevant to the work at hand; several exclamation or question marks may show feeling, but they may also signify a mistake or a joke rather than a genuine feeling. By removing them, the dimensionality of the data may decrease, which assisted the text processing algorithms and machine learning models to work more effectively and accurately. In a situation where punctations are useful in the analysis, feature engineering on the occurrences of various punctuation, particularly on consecutive appearances should be done.

viii. Emotions (such as emoticons and emojis) should be removed from text data while processing it since they can contribute to noise and bias. Though they might also be unrelated to the job at hand and add bias in text categorization, sentiment analysis, and other NLP tasks, emotions are frequently utilized to express sentiment. For emoticons that have impact in the analysis, it should be

converted to the English word.

ix. Lowercase is now used. Because it increases the effectiveness and precision of text processing algorithms and machine learning models, text to lowercase conversion in text processing is crucial. This is because a lot of machine learning and text processing algorithms are case-sensitive, and changing the text to lowercase can assist in making the data less dimensional and more consistent [54]. In case where the excitement is expressed using letter case, feature engineering needs to be done before conversion to lower case to preserve the meaning of the tweets.

3.4 Analysis of data

Text processing involves the crucial task of categorizing tweets as positive, negative, or neutral, accomplished through a text processing technique known as Count Vectorizer.

Count Vectorizer is a specific type of bag-of-words model that converts text into a numerical format. The CountVectorizer technique operates by initially tokenizing the text, breaking it down into individual words or phrases [44]. These tokens are subsequently tallied, and the counts are stored in a vocabulary. The vocabulary is a mapping from each token to its corresponding count. CountVectorizer in sentiment analysis is used to identify sentiment-related words. For example, words like "love" and "happy" are often associated with positive sentiment, while words like "hate" and "sad" are often associated with negative sentiment. CountVectorizer can be used to identify these words and then use them to create a sentiment lexicon. This lexicon can then be used to manually classify tweets or to train a machine learning model. Finally, CountVectorizer can also be used to analyze the sentiment of tweets over time. This can be done by tracking the frequency of sentiment-related words in tweets over time. This information can then be used to identify trends in public opinion.

Truncated SVD (Singular Value Decomposition) is a method for reducing the number of features in a dataset for dimensionality reduction. It works by performing a singular value decomposition (SVD) on the dataset. SVD decomposes the dataset into a set of orthogonal vectors, called singular vectors, and a set of singular values. The singular values are arranged in decreasing order, and the Truncated SVD algorithm keeps only the top k singular values. TruncatedSVD is made use of in sentiment analysis to identify sentiment-related features. For example, Truncated SVD was used in this study when trying to identify the top 15 tweets that are most correlated with negative sentiments.

Latent Semantic Analysis (LSA) is a method for drawing out semantic data from text. The number of features in a dataset can be decreased using this kind of dimensionality reduction method. To use LSA, you must first create a term-document matrix. This matrix is a table that shows how often each term appears in each document. LSA then performs a singular value decomposition (SVD) on the term-document matrix. SVD decomposes the matrix into a set of orthogonal vectors, called singular vectors, and a set of singular values. The singular values are arranged in decreasing order, and the LSA algorithm keeps only the top k singular values. The LSA algorithm then reconstructs the term-document matrix using the top k singular vectors. This results in a new term-document matrix with a lower dimensionality. Latent Semantic Analysis is made use of in sentiment analysis by implementing it with Truncated SVD in the dataset of the tweet of the 2023 presidential election to identify the top 10 topics that were prevalent during the elections.

Sentiment Intensity Analyser is a sentiment analysis tool that is made specifically to operate with social media text and is lexicon-based and rule-based. It works by first identifying sentiment-related words in a text. These words are then scored according to their intensity by getting a sentiment score between -1 and 1. A score of -1 represents a very negative sentiment, while a score of 1 represents a very positive sentiment. For example, the tweet "I'm so excited to vote for my candidate in the 2023 election!" would be scored as very positive, while the tweet "I'm so disappointed in the candidates in the 2023 election!" would be scored as very negative. Another example is the word "love" which is scored as very positive. This means that if the word "love" appears in a tweet about the 2023 presidential election, the Sentiment intensity analyser identifies it as a sentiment-related word.

3.5 Data visualization

Data visualization is the process of representing data in a visual format to facilitate easier understanding. To provide a comprehensive view of the factors considered when determining the type of tweets to collect, a graph that offers deeper insights into the data was generated. In this project, word clouds for visualization purposes were specifically utilized. A word cloud is an illustration of the most frequently appearing words in a text or dataset. It is widely employed in text analysis to quickly identify keywords and topics within a given text. Creating a word cloud typically involves measuring the frequency of each word in the text or dataset and sizing the words proportionally to visually represent their importance. Less frequently used words are typically displayed in smaller font sizes, whereas frequently used terms are typically printed in bigger font sizes. It is also possible to easily spot patterns and trends in the data thanks to the word arrangement [55].

4. RESULTS AND DISCUSSION

This section presents the outcomes of the Sentiment Analysis conducted on the Nigeria presidential elections conducted in 2023. The Jupyter Notebook, a free, open-source, interactive web-based environment that allows code generation and running as well as data visualization, served as the IDE for the sentiment analysis portion of this research work. The Natural Language Toolkit (NLTK), a widely used open-source Python package, provides an extensive collection of tools and resources for natural language processing (NLP). In this work, various NLTK functionalities such as Sentiment Intensity Analyzer, Count Vectorizer, TruncatedSVD (Singular Value Decomposition), Latent Semantic Analysis (LSA), and Word Clouds were utilized.

Figure 2 shows how different python libraries needed for the implementation of the study were imported. These NLTK tools presented in Figure 2 proved instrumental in the pre-processing and data visualization stages of our work. Through the application of these NLTK functionalities, the categorization of the tweets into positive, negative, and neutral attitudes were made possible.

Before the collection of data, some factors were considered. The Figure 3 shows the different factors which were taken into consideration during the data collection from Twitter.

```
#import all necessary packages
import numpy as np
import pandas as pd
import matplotlib
import seaborn as sns
import matplotlib.pyplot as plt
%matplotlib inline
from wordcloud import WordCloud, STOPWORDS
from sklearn.decomposition import LatentDirichletAllocation
from collections import Counter
from nltk.sentiment import SentimentIntensityAnalyzer
from textblob import TextBlob
plt.style.use("fivethirtyeight")

from sklearn.feature_extraction.text import CountVectorizer
import scipy.stats as stats

from sklearn.decomposition import TruncatedSVD
from sklearn.decomposition import LatentDirichletAllocation
from sklearn.manifold import TSNE
import warnings
warnings.simplefilter("ignore")
```

Figure 2. Python libraries and natural language toolkit used in this study

```
#checking for the shape of the data
print(f"data shape: {tweets_df.shape}")
```

data shape: (85662, 16)

```
## displaying some information about dataframe
tweets_df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 85662 entries, 0 to 85661
Data columns (total 16 columns):
 #   Column            Non-Null Count   Dtype
---  ------            --------------   -----
 0   id                85661 non-null   object
 1   user_name         85657 non-null   object
 2   user_location     64280 non-null   object
 3   user_description  78176 non-null   object
 4   user_created      85651 non-null   object
 5   user_followers    85651 non-null   object
 6   user_friends      85651 non-null   object
 7   user_favourites   85651 non-null   object
 8   user_verified     85649 non-null   object
 9   date              85650 non-null   object
 10  text              85651 non-null   object
 11  hashtags          45586 non-null   object
 12  source            85651 non-null   object
 13  retweets          85646 non-null   object
 14  favorites         85645 non-null   float64
 15  is_retweet        85645 non-null   object
dtypes: float64(1), object(15)
memory usage: 10.5+ MB
```

Figure 3. Factors considered during data collection

4.1 Preprocessing results

When the data was first gathered, the team decided to carry out an analysis to ensure that the right data had been collected, and the results indicated the accuracy of the data selection. The sample of data collected is shown in Figure 4.

Figure 4. Sample of dataset collected

To be sure of the accuracy of the data collection process, a check was conducted to verify for missing data fields. From Figure 4, it was observed that some individuals had missing data. Users such as Promise and Mo'6icky had no location specified, while Promise and KemKem did not use any hashtags or engage in any political conversations online. This prompted us to perform an additional analysis to calculate the percentage and total amount of missing data in the previously collected dataset. This analysis revealed that, based on the factors considered during data gathering, 5 users had not provided usernames, 21,382 users had not specified their location, 7,486 users without any provided description, and 40,076 users who had not participated in hashtags and likes. The result of missing data is presented in Figure 5, the missing data is when a user failed to provide information about his location, usernames etc., missing data can lead to a number of problems, such as poorer performance, trouble processing data, and biased findings because of the differences between complete and missing information.

To enhance the accuracy of the data, it was decided to replace the missing data with unique values. In all data-related professions, missing data can lead to a number of problems, such as poorer performance, trouble processing data, and biased findings because of the differences between complete and missing information. This approach ensures that the dataset remains reliable and provides a complete and comprehensive representation of the gathered information. Figure 6 shows the values which were used to replace the missing data.

Figure 5. Missing data profile

Figure 6. Presentation of unique values for the replacement of missing data for enhancement of the accuracy of the data

4.2 Latent semantic analysis results

After implementing the NLTK Stopwords to remove irrelevant words that could potentially corrupt the data and yield inaccurate results, the analysis proceeded based on the factors considered during data collection. As a result, word clouds and bar graphs were generated, depicting the Top 10 user_names, Top 10 devices used, Top 10 user_locations, and Top 10 most used hashtags. Figure 7 shows 10 user_names and the Top 10 devices used, and Figure 8 presents the Top 10 user_locations, and the Top 10 most used hashtags during the Nigeria presidential election. These visual representations provided valuable insights into the most prominent users, locations, user descriptions, devices, and hashtags associated with the #NigeriaDecides2023 election period.
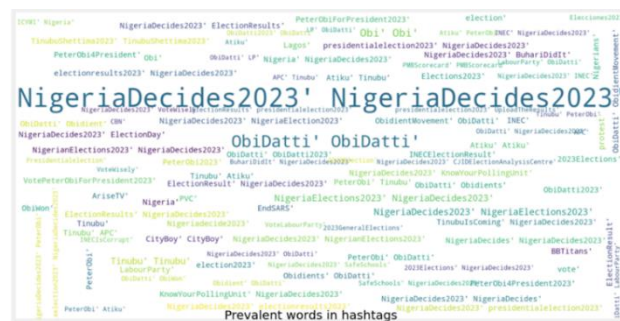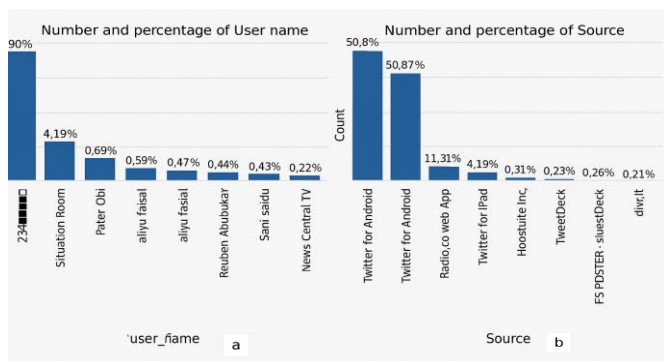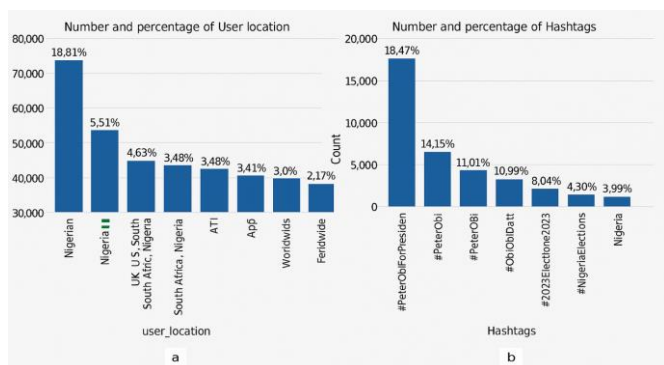
Figure 7. (a) Top 10 most active users based on their names, (b) Top 10 most frequently used devices by the users

Figure 8. (a) Top 10 locations of the users, (b) Top 10 most used hashtags

4.3 Visualization result

A word cloud was generated to know the prevalent words in tweets and the most used hashtags and these were shown in Figures 9 and 10 respectively. According to Figure 9, the most discussed presidential aspirants were Peter Obi, Tinubu, and Atiku. The majority of users who were active on Twitter during the election period were located in Nigeria. Additionally, the popular words used during the election period included "Election," "Support," and "Polling Unit." These findings provide valuable insights into the trending topics and discussions on Twitter during the 2023 election period in Nigeria.

Figure 9. Most prevalent words in tweets by users

Based on Figure 10, it is evident that #NigeriaDecides2023 was the most frequently used hashtag during the election period. This hashtag had the highest occurrence, indicating that it was widely used and became a central topic of discussion on Twitter throughout the election period.
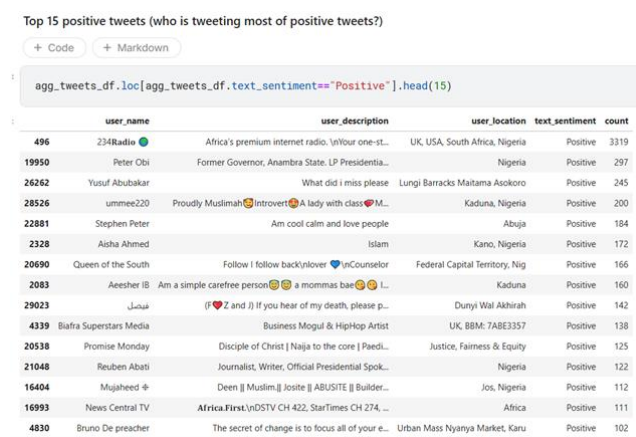

Figure 10. Most used hashtags

4.4 Sentiment intensity analysis result

In this study, the opinions of individuals were classified into three expressions. This allowed for the classification of tweets as either positive, negative, or neutral based on the sentiment expressed in the text as presented in Figures 11-13. The sentiment analysis provided valuable insights into the overall public sentiment and opinions during the presidential election period.

Based on Figure 11, it was evident that the user "234Radio" station had the most positive tweets. Following closely behind with more positive tweets were tweets about Peter Obi, a presidential aspirant and former governor of a Nigerian state, and Yusuf Abubakar, a well-known politician. These findings indicate that these users were actively engaging in positive discussions and interactions on Twitter during the presidential election period.

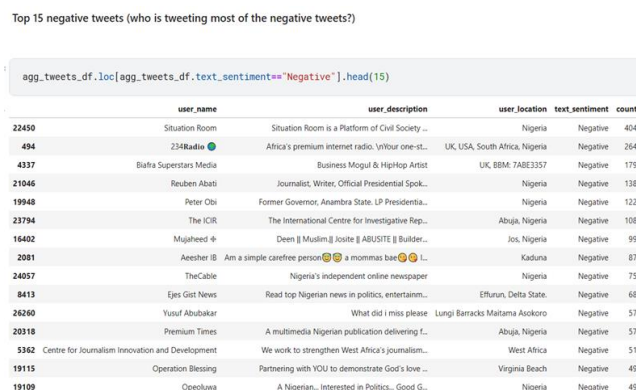**Figure 11.** Positive tweets of the active users

Figure 12. Negative tweets of the active users

Based on Figure 12, it is apparent that "Situation Room," a space on Twitter comprising Nigerian civil society groups focused on elections, good administration, and other concerns, had the most negative tweets towards the election period. This finding indicates that many users utilized this platform to express their negative feelings and concerns related to the presidential election. The prominence of negative tweets from this user suggests that it served as a significant avenue for critical discussions and commentary during the election period.

Some users, according to Figure 13, were indifferent about their emotions during the presidential election due to the reasons best known to them. The authors suggested reasons on their neutrality to be due to the past disappointed received from the past elections, poor governance and mistrust on the candidates running for the presidential position.

From Figure 14, the result shows that there were 48% positive reactions from the users, 31% neutral reactions from users, and 23% negative reactions from users.

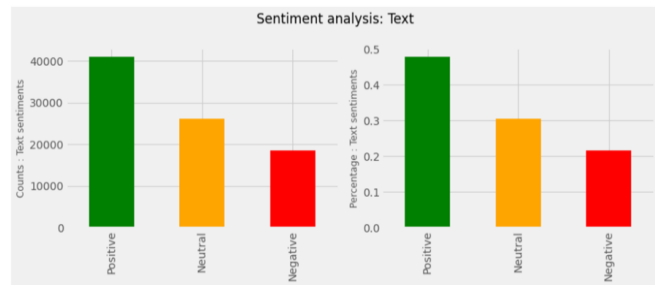**Figure 13.** Neutral tweets of the active users

Figure 14. Sentiment intensity analysis result of Nigeria presidential election

4.5 Topic modelling result

The trending topics presented in Figures 15 and 16 during the election period among the general public were analyzed. This analysis provided valuable insights into the most discussed and prominent topics on Twitter related to the presidential election. By identifying the trending topics, a deeper understanding of the public's interests, concerns, and opinions during the election period were known, and thereby allowing us to capture the pulse of the nation on social media.

Figure 15. Trending topics during the Nigeria presidential election period

The analysis of the most discussed topics as presented in Figures 15 and 16 shows topic 0: Nigeriadecides2023 with several headlines 35,000counts followed to this is topic:8 Obidatti with 10,000 headlines counts and the next topic 2: Vote Obi Atiku with 8,000 headlines counts. On the fourth position is topic 7: state APC Lagos with 7,000 headline counts. This result gave a clear description of the actual happenings during the Nigeria presidential election conducted in 2023.
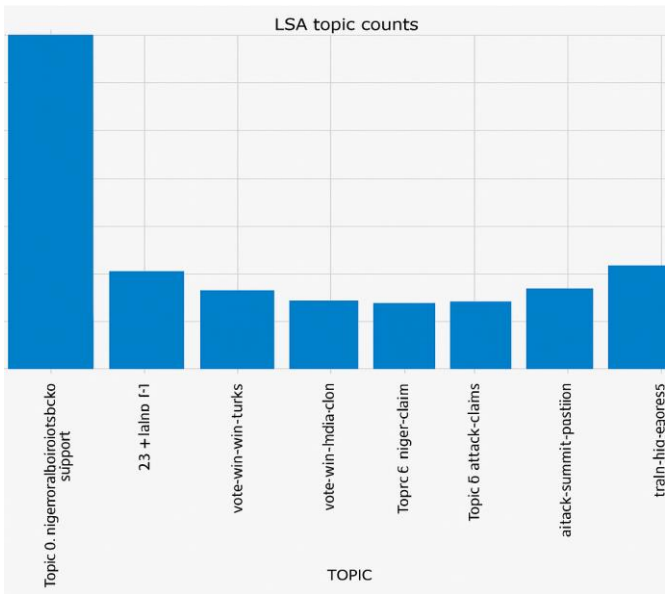

Figure 16. Latent semantic analysis topic counts for most discussed hashtags

4.6 General result discussion

The results of this study indicated that the tweets generated by Nigerian users on Twitter during the 2023 presidential elections predominantly focused on the personalities of the major candidates vying for the Nigerian presidency without emphasis on the political party. The Sentiment Analysis conducted on Twitter data related to the 2023 presidential election in Nigeria, specifically using the hashtag "#NigeriaDecides2023," revealed that the prominent candidates mentioned in these tweets were Peter Obi, Bola Tinubu, and Atiku Abubakar. Also, most Twitter users predominantly youth, had positive intentions towards Peter Obi; and neutral and negative intentions about Tinubu and Atiku. Meaning that most Nigerian youths were in support of Peter Obi to win the election, the reason being that, the Nigerian youths wanted a new and young leader different from the previous ones ruling them. The youth in Nigeria believe that a lasting positive effect on the economy can only start through the election of a young dynamic person and not the old persons. The result study shows that most Twitter users had neutral and negative things to say about Atiku Abubakar and Bola Ahmed Tinubu, meaning that some users were not sure if his tenure would be a good one if he was to win the election, while others have few positive and negative things to say about them.

TextBlob version 0.17.1 and Python 3.8 were used in this investigation. TextBlob has an integrated sentiment analyser that operates without the need for additional datasets or model training. When examining smaller datasets, such as political speeches or specific social media posts, its lexicon-based

approach's processing efficiency is beneficial. TextBlob does not require labelled training data, in contrast to sentiment analysis techniques based on machine learning. Furthermore, although TextBlob is easy to use and helpful for basic NLP applications, it struggles to identify context in text. Its reliance on pre-existing lexicons and rule-based techniques, however, is one of its primary disadvantages as it may overlook the nuanced meanings that context provides. The sentiment analysis provided by TextBlob, for example, sometimes examines individual words or phrases without considering context, which may alter sentiment. Idioms, cultural allusions, and sarcasm are a few examples. It can lead to erroneous interpretations of the true meaning of the text. There are also no advanced deep learning models that utilise contextual embeddings in TextBlob. Unlike transformer-based methods like BERT, which consider the context of a word by examining the words that comprise its surrounding phrase.

4.7 Comparison of result with the existing works

This section delves into comparative analysis of our result with the existing works as depicted in Table 3. Four factors were used to evaluate the performance of this study in relative to similar existing works. The factors are data source, analysis techniques, performance metrics, and sentiment granularity.

Table 3. Comparative analysis of our result with the existing works

Author	Data Source	Analysis Technique	Visualization Technique	Sentiment Granulariy
Al Barghuthi and Said [38]	Twitter	Naive Bayes	Word cloud, pie chart	Positive, Negative, and Neural
Ali et al. [50]	Twitter	VADER	Word count, timelines	Positive, Negative, and Neutral
Caetano et al. [51]	Twitter	Dictionary lexicon-based approach	Timelines, graph	Positive, Negative, and Neural
Proposed study	Twitter	Textblob lexcion-based approach	Word count, word cloud, topic modelling, graph	Positive, Negative, and Neutral

From the Table 3, the existing works sourced their datasets from Twitter. The reason for chosen Twitter is because it is widely subscribed social medium platform by the populace. More importantly, Twitter is a medium where emotion and opinion are expressed. From this comparative study, it can be seen that analysis technique differs for different authors and this could be based on the opinion being analysed. From these works, our study used more visualization techniques to showcased the result for better comprehension by all. In a nutshell, the present study has demonstrated the used of current sentiment analysis technique with more visualization tool to showcase their findings.

4.8 The practical value and societal impact based on this research outcome

With the advent of internet technology, social media platforms have been a place where majority of people express their feelings. Nigerians are not left out in this regard as it shows during the 2023 presidential election. Individual, groups, ethnic groups, political parties and others took this advantage to campaign for their candidate. It was recorded that, there was massive influx of Nigerians in diaspora into the country prior to the election day. The impact of social medium on the Nigerian to exercise their right in the 2023 election and also, to talk about different presidential candidates cannot be overemphasized. According to the definition of civic education, it is the process of teaching young people to be involved and active citizens through civics, government, history, and other related disciplines taught in the classroom, as well as through service learning, democratic procedures, and other activities. One of the lessons learnt based on this research outcome is that, human capacity development is key in gaining the heart of the people. Political candidate at any level should cultivate the habit of good stewardship towards the citizen and also, in infrastructural development. Often times, winning an election is not about rigging or using power, it has to do with the personality of the aspirant.

5. CONCLUSION

The sentiment analysis of the #NigeriaDecides2023 using Twitter data provides valuable insights into the sentiments expressed by the Nigerian public during this crucial electoral event. By analyzing the sentiments shared on Twitter, a deeper understanding of the opinions, preferences, and concerns of the electorate were gained. The results and findings can inform decision-making processes, guide political strategies, and contribute to ongoing discussions on the role of social media in shaping political landscapes.

By illuminating voter sentiment and public opinion, this study demonstrated how sentiment analysis research could have a substantial impact on elections. Sentiment analysis utilizes Natural Language Processing techniques to assess various sources of data, such as social media posts, news stories, and other content, to determine the underlying sentiment or emotion. This analytical approach proves valuable for political parties and candidates as it allows them to gauge public sentiment towards their policies and themselves before an election. By examining social media posts and news stories within the context of sentiment analysis, patterns can be identified in how people respond to political events and speeches. This information becomes a useful tool for political campaigns to adapt their strategies and tactics accordingly. For example, if sentiment analysis reveals that voters disapprove of a particular policy proposal, a campaign may choose to modify or abandon that idea to garner more support. Additionally, sentiment analysis enables campaigns to identify key themes that resonate with voters and adjust their messaging to highlight those concerns.

In general, sentiment research offers valuable insights into public perceptions and the overall mood of the electorate, aiding political campaigns in making informed decisions and refining their approaches to better connect with voters. It is important to note, however, that not all eligible voters express their thoughts on Twitter, and there is variation in the usage of hashtags. Therefore, the findings of this study may not fully

align with the actual election outcome. Furthermore, the tweets analyzed in this study were not geographically specific, making it challenging to determine if they solely originated from Nigerian accounts.

Nevertheless, by providing insights into the networks, impressions, and public sentiments surrounding the three prominent presidential candidates during the #NigeriaDecides2023 election, this study has contributed to our overall understanding of the election process.

5.1 Limitation to the Study

Generalization of Results: Not all Nigerian voters may agree with the results of sentiment analysis on Twitter data. Twitter users are a particular segment of the population, so they might not accurately reflect the opinions of the general public. Interpreting the sentiment analysis results in light of this restriction while avoiding over-generalization is difficult.

Linguistic and Cultural Nuances: The sentiment analysis may encounter challenges in accurately capturing sentiment due to the diverse linguistic and cultural expressions specific to Nigeria. Variations in dialects, local languages, and the use of slang or colloquialisms may introduce difficulties in sentiment classification and interpretation.

5.2 Future research

Several suggestions for further study and advancements can be made based on the results and findings of the sentiment analysis of the #NigeriaDecides2023 elections utilizing Twitter data:

Incorporate Data from Multiple Social Media Platforms: It would be advised that in future studies, the data collection be expanded beyond Twitter to incorporate information from Facebook, Instagram, and online discussion forums. This would give an improved understanding of the attitudes and viewpoints of the people across different online channels.

Investigate Emotion Analysis Techniques: Future studies should consider combining sentiment analysis with emotion analysis techniques. Insights into the intensity and nuance of public opinion would be gained as a result, allowing for a fuller understanding of the emotional reactions connected to the feelings voiced during the elections.

ACKNOWLEDGMENT

REFERENCES

[1] Lamak, K. (2022). A history of Nigeria, by Toyin Falola and Mathew M. Heaton. African Historical Review, 53(1-2): 95-97.

[2] Ogbeidi, M.M. (2010). A culture of failed elections: Revisiting democratic elections in Nigeria, 1959-2003. Historia Actual Online, 21: 43-56.

[3] Sklar, R.L. (2004). Nigerian political parties: Power in an emergent African nation. New Jersey, USA: Africa World Press.

[4] Ikelegbe, A. (2016). Politics and Governance in Nigeria:

[5] Kurfi, A. (2005). Nigerian General Elections: My Roles and Reminiscences. Ibadan, Nigeria: Spectrum Books.

[6] Adibe, J. (2019). INEC and the challenges of free and fair elections in Nigeria. https://www.inecnigeria.org/wp-content/uploads/2019/02/Conference-Paper-by-Jideofor-Adibe.pdf.

[7] Nwaodu, N. (2011). Theory and practice of development administration in Nigeria: A survey of Institutional Development in Selected Sectors of the Nation's Economy. Owerri, Nigeria: Mega Atlas Projects Limited.

[8] Adele, J. (2011). Nigeria. In M.H.A.L.J. Ismaila M. Fall & P.K. (Eds.), Election Management Bodies in West Africa (OSIWA).

[9] Oliji, V. (2021). Improvements in the Electoral System in Nigeria Since Independence. https://von.gov.ng/improvements-in-the-electoral-system-in-nigeria-since-independence/.

[10] Bansal, B., Srivastava, S. (2018). On predicting elections with hybrid topic-based sentiment analysis of tweets. Procedia Computer Science, 135: 346-353. https://doi.org/10.1016/j.procs.2018.08.181

[11] Kumar Singh, S., Verma, P., Kumar, P. (2020). Sentiment analysis using machine learning techniques on Twitter: A critical review. Advances in Mathematics: Scientific Journal, 9(9): 7085-7092.

[12] Yanjiv, T. (2023). Unlegalized activities and practice of political parties to raise funds to support women candidates. Studia Philosophiae Et Juris, pp. 80-93.

[13] Kwak, H., Lee, C., Park, H., Chun, H., Moon, S. (2011). Novel aspects coming from the directionality of online relationships: A case study of Twitter. ACM SIGWEB Newsletter.

[14] Mk, S., Sudha, M.K. (2020). Social media sentiment analysis for opinion mining. International Journal of Psychosocial Rehabilitation, 24(5): 3672-3679.

[15] Wang, M., Hu, G. (2020). A novel method for Twitter sentiment analysis based on attentional-graph neural network. Information, 11(2): 92. https://doi.org/10.3390/info11020092

[16] Katalinić, J., Dunđer, I., Seljan, S. (2023). Polarizing topics on Twitter in the 2022 United States elections. Information, 14(11): 609. https://doi.org/10.3390/info14110609

[17] Yadav, J. (2023). Sentiment Analysis on Social Media. https://doi.org/10.31219/osf.io/d6wv2

[18] Geni, L., Yulianti, E., Sensuse, D.I. (2023). Sentiment analysis of tweets before the 2024 elections in Indonesia using Indobert language models. Jurnal Ilmiah Teknik Elektro Komputer Dan Informatika (JITEKI), 9(3): 746-757. https://doi.org/10.26555/jiteki.v9i3.26490

[19] Patel, S.B., Dharwa, J., Patel, C.D. (2024). Harnessing Twitter: Sentiment analysis for predicting election outcomes in India. ITM Web of Conferences, 65: 03008. https://doi.org/10.1051/itmconf/20246503008

[20] Roberts, M.E. (2018). What is political methodology? PS: Political Science & Politics, 51(3): 597-601.

[21] Achen, C.H. (2002). Toward a new political methodology: Microfoundations and ART. Annual review of political science, 5(1): 423-450. https://doi.org/10.1146/annurev.polisci.5.112801.080943

[22] Borghetto, E., Belchior, A.M. (2020). Party manifestos,

opposition and media as determinants of the cabinet agenda. Political Studies, 68(1): 37-53. https://doi.org/10.1177/0032321718820738

[23] Secretary-General, U.D. (2021). With Almost Half of World's Population Still Offline, Digital Divide Risks Becoming 'New Face of Inequality', Deputy Secretary-General Warns General Assembly.

[24] Word Bank. (2023). Digital Development Partnership Annual Review 2023: Transitioning Towards Scale (English). https://documents.worldbank.org/en/publication/documents-reports/documentdetail/099810207152431146.

[25] Pal, J., Gonawela, A. (2016). Political social media in the global South. In Social Media: The Good, the Bad, and the Ugly: 15th IFIP WG 6.11 Conference on e-Business, e-Services, and e-Society, I3E 2016, Swansea, UK, pp. 587-593. https://doi.org/10.1007/978-3-319-45234-0_52

[26] Dad, N., Khan, S. (2023). Reconstructing elections in a digital world. South African Journal of International Affairs, 30(3), 473-496. https://doi.org/10.1080/10220461.2023.2265886

[27] Ayang, L.O., Abonyi, C.L. (2022). Variation of prenatal ultrasound-estimated cephalic index; A comparative study of Igbo, Hausa and Yoruba ethnic groups in Nigeria. Journal of Radiography and Radiation Sciences, 32(1).

[28] Wilson, T., Wiebe, J., Hoffmann, P. (2009). Recognizing contextual polarity: An exploration of features for phrase-level sentiment analysis. Computational Linguistics, 35(3): 399-433. https://doi.org/10.1162/coli.08-012-R1-06-90

[29] Agarwal, P., Shrivastava, P. (2021). Twitter sentiment analysis using machine learning. Computology: Journal of Applied Computer Science and Intelligent Technologies, 1(12): 37-46.

[30] Wang, Z., Ho, S.B., Cambria, E. (2020). Multi-level fine-scaled sentiment sensing with ambivalence handling. International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, 28(04): 683-697. https://doi.org/10.1142/S0218488520500294

[31] Alharbi, J.R., Alhalabi, W.S. (2020). Hybrid approach for sentiment analysis of Twitter posts using a dictionary-based approach and fuzzy logic methods. International Journal on Semantic Web and Information Systems, 16(1): 116-145. https://doi.org/10.4018/IJSWIS.2020010106

[32] Oyebode, O., Orji, R. (2019). Social media and sentiment analysis: The Nigeria presidential election 2019. In 10th Annual Information Technology, Electronics and Mobile Communication Conference, IEMCON 2019, Vancouver, BC, Canada, pp. 140-146. https://doi.org/10.1109/IEMCON.2019.8936139

[33] Büschken, J., Allenby, G.M. (2020). Improving text analysis using sentence conjunctions and punctuation. Marketing Science, 39(4): 727-742. https://doi.org/10.1287/mksc.2019.1214

[34] Illia, F., Eugenia, M.P., Rutba, S.A. (2021). Sentiment analysis on pedulilindungi application using textblob and vader library. Proceedings of the International Conference on Data Science and Official Statistics, 2021(1): 278-288. https://doi.org/10.34123/icdsos.v2021i1.236

[35] Muhammad, A.B., Dahiru, A.A. (2019). Lexicon-based sentiment analysis of web discussion posts using SentiWordNet. Journal of Computer Science and Its Application, 26(2): 1-11. https://doi.org/10.4314/jcsia.v26i2.1

[36] Patil, S.A., Adhiya, K.P., Patnaik, G.K. (2023). Various feature extraction and selection techniques for lexicon based and machine learning sentiment classification. The Ciência & Engenharia - Science & Engineering Journal, 11(1): 1939-1951. https://doi.org/10.52783/cienceng.v11i1.361

[37] Sham, N.M., Mohamed, A. (2022). Climate change sentiment analysis using lexicon, machine learning and hybrid approaches. Sustainability, 14(8): 4723. https://doi.org/10.3390/su14084723

[38] Al Barghuthi, N.B., E. Said, H. (2020). Sentiment analysis on predicting presidential election: Twitter used case. In International Symposium on Intelligent Computing Systems, Cham: Springer International Publishing, pp. 105-117. https://doi.org/10.1007/978-3-030-43364-2_10

[39] El-Badaoui, M., Gherabi, N., Quanouni, F. (2024). TED talks comments sentiment classification using machine learning algorithms. Revue d'Intelligence Artificielle, 38(3): 885-892. https://doi.org/10.18280/ria.380315

[40] Saha, S., Yadav, J., Ranjan, P. (2017). Proposed approach for sarcasm detection in twitter. Indian Journal of Science and Technology, 10(25): 1-8. https://doi.org/10.17485/ijst/2017/v10i25/114443

[41] Adegoke, M.A., Adedotun, A.F., Odeyinde, T., Grace, A.O., Sotonwa, K.A., Aribisala, B.S., Oluwole, O.A., Onuche, O.G. (2024). Development of stroke prediction system using neural network with sigmoid activation function. Journal Européen des Systèmes Automatisés, 57(6): 1753-1760. https://doi.org/10.18280/jesa.570623

[42] Gupta, K., Jiwani, N., Afreen, N. (2023). A combined approach of sentimental analysis using machine learning techniques. Revue d'Intelligence Artificielle, 37(1): 1-6. https://doi.org/10.18280/ria.370101

[43] Durga, P., Deepthi Godavarthi, D. (2024). ERS-GARNET: An ensemble recommendation system for sentiment analysis using gated attention-based recurrent networks. Ingénierie des Systèmes d'Information, 29(3): 839-852. https://doi.org/10.18280/isi.290305

[44] Eweoya, I.O., Odetunmibi, O.A., Odun-Ayo, I.A., Agbele, K.K., Adedotun, A.F., Akingbade, T.J. (2023). Machine learning approach for the prediction of COVID-19 spread in Nigeria using SIR model. International Journal of Sustainable Development and Planning, 18(12): 3783-3792. https://doi.org/10.18280/ijsdp.181210

[45] Nakka, R., Lakshmi, T.S., Priyanka, D., Sai, N.R., Praveen, S.P., Sirisha, U. (2024). LAMBDA: Lexicon and aspect-based multimodal data analysis of tweet. Ingénierie des Systèmes d'Information, 29(3): 1097-1106. https://doi.org/10.18280/isi.290327

[46] Adesina, O.S., Adedotuun, A.F., Adekeye, K.S., Imaga, O.F., Adeyiga, A.J., Akingbade, T.J. (2024). On logistic regression versus support vectors machine using vaccination dataset. Journal of the Nigerian Society of Physical Sciences, 6(1): 1092. http://doi.org/10.46481/jnsps.2024.1092

[47] 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, 9(4): e14836.

https://doi.org/10.1016/j.heliyon.2023.e14836

[48] Wang, Z., Ho, S.B., Cambria, E. (2020). Multi-level fine-scaled sentiment sensing with ambivalence handling. International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, 28(4): 683-697. https://doi.org/10.1142/S0218488520500294

[49] Razzaq, M.A., Qamar, A.M., Bilal, H.S.M. (2014). Prediction and analysis of Pakistan election 2013 based on sentiment analysis. In 2014 IEEE/ACM International conference on advances in social networks analysis and mining (ASONAM 2014), Beijing, China, pp. 700-703. https://doi.org/10.1109/ASONAM.2014.6921662

[50] Ali, R.H., Pinto, G., Lawrie, E., Linstead, E.J. (2022). A large-scale sentiment analysis of tweets pertaining to the 2020 US presidential election. Journal of Big Data, 9(1): 79. https://doi.org/10.1186/s40537-022-00633-z

[51] Caetano, J.A., Lima, H.S., Santos, M.F., Marques-Neto, H.T. (2018). Using sentiment analysis to define twitter political users' classes and their homophily during the 2016 American presidential election. Journal of Internet Services and Applications, 9: 1-15. https://doi.org/10.1186/s13174-018-0089-0

[52] Neogi, A.S., Garg, K.A., Mishra, R.K., Dwivedi, Y.K. (2021). Sentiment analysis and classification of Indian farmers' protest using twitter data. International Journal of Information Management Data Insights, 1(2): 100019. https://doi.org/10.1016/j.jjimei.2021.100019

[53] Wang, H., Li, Y., Hutch, M., Naidech, A., Luo, Y. (2021). Using tweets to understand how COVID-19-related health beliefs are affected in the age of social media: Twitter data analysis study. Journal of Medical Internet Research, 23(2): e26302. https://doi.org/10.2196/26302

[54] Bibi, M., Abbasi, W.A., Aziz, W., Khalil, S., Uddin, M., Iwendi, C., Gadekallu, T.R. (2022). A novel unsupervised ensemble framework using concept-based linguistic methods and machine learning for twitter sentiment analysis. Pattern Recognition Letters, 158: 80-86. https://doi.org/10.1016/j.patrec.2022.04.004

[55] Tumasjan, A., Sprenger, T., Sandner, P., Welpe, I. (2010). Predicting elections with twitter: What 140 characters reveal about political sentiment. In Proceedings of the International AAAI Conference on Web and Social Media, 4(1): 178-185. https://doi.org/10.1609/icwsm.v4i1.14009