**Chapter One**

**Introduction**

Social media platforms have experienced exponential growth over the years. Recently alone, the number of social media users globally grew from 4.62 billion in January 2022 to 4.72 billion in January 2023 (Chaffey, 2023). This trend has drawn the attention of many organizations, companies and researchers to social media over of the outburst of social data they made available, and partially accessible online, also for the reliability of emotions expressed spontaneously by people towards a topic without any social boundaries or psychological complexity, and moreover psychological researchers found that more the people are relaxed at their homes, the more their sentiment and expression are true (Esteban, 2019).

The concept of sentiment analysis, also referred to as opinion mining, is an approach to Natural Language Processing ([NLP](https://www.techtarget.com/searchbusinessanalytics/definition/natural-language-processing-NLP)) that identifies the emotional tone behind a body of text. This is a popular way for organizations to determine and categorize opinions about a product, service or idea. Sentiment analysis involves the use of [data mining](https://www.techtarget.com/searchbusinessanalytics/definition/data-mining), machine learning ([ML](https://www.techtarget.com/searchenterpriseai/definition/machine-learning-ML)), artificial intelligence and [computational linguistics](https://www.techtarget.com/searchenterpriseai/definition/computational-linguistics-CL) to mine text for sentiment and subjective information such as whether it is expressing positive, negative or neutral feelings.

In this thesis, we seek to benefit from the tremendous development that artificial intelligence and semantic web brings, and also the tools they offer in order to extract the sentiments of the contents posted on twitter and get insights about their opinions towards Nigerian 2023 Presidential Elections.

Elections are fundamentally about the will of the people. Sentiment analysis allows political parties, candidates, and analysts to gauge public sentiment, providing insights into what voters are thinking and feeling. It helps in understanding the electorate's preferences, concerns, and priorities. Sentiment analysis aids political campaigns in tailoring their messaging and strategies to resonate with the electorate. By analysing sentiment, campaigns can better address issues that matter the most to voters, and they can adapt their messages based on real-time feedback. Sentiment analysis is used for early predictions and tracking trends in voter sentiment throughout the election cycle. As well, it can uncover new or emerging issues that may not be adequately addressed by candidates. By monitoring social media and news platforms, political actors can be more responsive to the evolving concerns of the electorate.

In this digital age, the spread of misinformation and disinformation is of great concern. Sentiment analysis can help in identifying false narratives and rumours, enabling fact-checking efforts and corrections. Candidates can also use sentiment analysis to assess how their campaign messages are resonating with the public. They can make data-driven adjustments to their communication strategies based on sentiment feedback.

Sentiment analysis is valuable not only during but also after an election. It helps in post-election analysis, understanding why voters made certain choices, and assessing the effectiveness of campaign strategies.

By monitoring public sentiment, elected officials can better understand their constituents' expectations and hold themselves accountable for campaign promises. This encourages transparency and responsive governance. Sentiment analysis can encourage civic engagement. When voters see that their opinions are being listened to and considered, they may be more likely to participate in the electoral process and engage in discussions about important issues. Sentiment analysis during elections generates valuable data for academic research and policy analysis. It provides a rich source of information for studying voter behaviour and the impact of political messaging.

Sentiment analysis is a powerful tool for enhancing democratic processes, improving campaign strategies, and promoting informed decision-making. It helps candidates, political parties, and governments better understand the electorate, respond to their needs, and ensure that the democratic system functions as intended.

**Problem Statement**

Existing literatures clearly shows that social media plays key role in electioneering campaign, sharing manifesto and reaching out to electorates in developed nations like France, USA and Germany etc. But we do not know to what extent can social media help in reaching out to Nigerians and to what extent will the social media users in Nigeria influence the outcome of an elections. While literature also indicates the exact number of social media users in developed nations, same cannot be said about Nigeria.

In this research, we will try to find out if there is a correlation between social media posts and the outcome of the 2023 presidential election in Nigeria. We will be analyzing users tweets online and perform sentiment analysis on to find the sentiment of users on one of the three (3) most popular candidates of the 2023 Nigerian Presidential Election i.e. Atiku Abubakar of the PDP, Peter Obi of the Labour party and Bola Ahmed Tinubu of the APC political parties respectively. We used twitter api to gather about 78,946 tweets and will focus our research on this figure and test the effectiveness of two models i.e. VADER Model in comparison with RoBERTa Model to see which classifier performs better in classifying tweets sentiment to each candidate.

We will be cleaning the data by removing all empty or missing values and all features must be properly defined to enable us perform the analysis. In addition, we will be preprocessing our dataset by replacing usernames with user, http addresses with http and hashtags within the text column of our tweet dataset.

**Purpose of the Study**

The aim of this study is to find out whether there is a correlation between tweets of Nigerians on twitter and the outcome of the 2023 Presidential Elections using VADER and RoBERTa sentiment analysis models.

The data will comprise of the 36 states of the federation, FCT and beyond Nigeria as reflected on the dataset. Nigeria is a country with the estimate of about 220million people. The Independent National Electoral Commission said about 93.46 million Nigerians are eligible to vote during the 2023 General Elections. They reported that about 87 million PVCs have been collected prior to the elections.

**Objectives of the Study**

The following forms the objectives of this study:

1. Develop Effective Sentiment Analysis Models: Develop and evaluate machine learning models for sentiment analysis that can accurately classify voters' opinions into positive, negative, or neutral categories.
2. Analyze Social Media Data: Design models that specialize in analyzing sentiment in social media data, where opinions are often expressed in short and informal text.
3. Predict Election Outcomes: Investigate the potential of machine learning models in predicting election outcomes based on sentiment analysis data.
4. Contribute to Academic Research: Generate valuable data and insights that can contribute to academic research in political science, machine learning, and natural language processing.
5. Assess Public Opinion: To analyze and assess public sentiment, opinions, and attitudes expressed by voters, political stakeholders, and the general public during the Nigerian Presidential Election.
6. Monitor Candidate Perception: To track and evaluate how the candidates and their campaigns are perceived by the public. This involves assessing sentiment towards individual candidates and their party platforms.
7. Evaluate Campaign Strategies: To analyze the effectiveness of campaign strategies and messaging. This includes assessing which campaign messages resonate with voters and which do not.
8. Predict Election Outcomes: To use sentiment analysis as a predictive tool to forecast or estimate potential election outcomes based on public sentiment and trends.
9. Measuring Candidate Performance: Candidates can use sentiment analysis to assess how their campaign messages are resonating with the public. They can make data-driven adjustments to their communication strategies based on sentiment feedback.
10. Assessing Voter Turnout: Sentiment analysis can provide insights into voter enthusiasm and the likelihood of voter turnout. Understanding the emotional connection voters have with candidates and issues can help in mobilization efforts.

**Research Question(s) and Hypotheses**

In this study, we intend to answer the following research questions and hypotheses

1. Can we rely on tweets in predicting outcome of elections?
2. Do we know the exact number of Nigerians on social media and twitter in particular?
3. Is the sentiment accurately classified by our models?

**Hypotheses**

H0: There is no significant relationship between tweets sentiment and the outcome of Nigerian 2023 Presidential Elections.

H1: There is significant relationship between tweets sentiment and the outcome of Nigerian 2023 Presidential Elections

**Rationale, Relevance, and Significance of the Study**

**Rationale of the study**

In this study, we will be using VADER model and RoBERTa model to perform sentiment analysis. VADER One of the well-known rule-based algorithms is [VADER from the NLTK](https://www.nltk.org/_modules/nltk/sentiment/vader.html) package. It isspecifically attuned to sentiments expressed in social media. It is quite easy to implement yet a very powerful model. Here we deal with bag of words approach, we analyze each word individually and score it to be able to say if the tweet is positive, negative or neutral. While, RoBERTA is a transformer based deep learning model that can pick up on the context at which a text is written. [Twitter-roBERTa from the transformers package](https://huggingface.co/cardiffnlp/twitter-roberta-base-sentiment-latest). This model was pretrained with 58 million tweets and finetuned for sentiment analysis.

Both of the models do not require any training, which means they can be directly applied to textual data out of the box.

**Significance of the Study**

The research aims to provide insights into the public sentiment surrounding the 2023 Nigerian Presidential Election, which can be valuable for understanding voter sentiment, campaign strategies, and post-election analysis.

This study will have a direct effect on Nigerian elections in the future. Whether the political parties are to take social media sentiment seriously or not. It is also going to show to what extent does social media affect the outcome of the elections and whether candidates are to focus on reaching out to the electorates on social media or offline. Most importantly, this study will enable the researcher implement two (2) different models and to see which will perform better in predicting the sentiment of social media users and how much is it correlated to the outcome of the elections.

As Nigerians continue to embrace social media, issues of governance of the nation will continue to be discussed on social media and the attention of young Nigerians will continue to be drawn to the media. Therefore, it is importantly to study the level of acceptance of social media at the moment and to what extent it can influence future elections and issues of governance of the nation.

So far, we know that a significant number of Nigerians are on social media but we do not have the exact figure as the number keeps changing every day. This study will try to find out the exact fraction of the population of Nigerians on social media and specifically on Twitter(X).

This study will try to find out whether the available technology is able to accurately predict the sentiment of Nigerians on social media and it will as well try to find out among the available models, which performs better. Considering the fact that Nigerians use different ways of expressing their thoughts. Sometimes they use pidgin or native languages to express their concerns on social media.

**Definition of Terms**

**NLTK**: NLTK stands for Natural Language Tool Kit, is a leading platform for building Python programs to work with human language data. It provides easy-to-use interfaces to [over 50 corpora and lexical resources](https://www.nltk.org/nltk_data/) such as WordNet, along with a suite of text processing libraries for classification, tokenization, stemming, tagging, parsing, and semantic reasoning, wrappers for industrial-strength NLP(Natural Language Processing) libraries, and an active [discussion forum](https://groups.google.com/group/nltk-users).

**ML**: The use and development of computer systems that are able to learn and adapt without following explicit instructions, by using algorithms and statistical models to analyse and draw [inferences](https://www.google.com/search?sca_esv=583796676&sxsrf=AM9HkKnlEOHU0Q2TjyJ9IW3s3COaxfwGhg:1700406465138&q=inferences&si=ALGXSlbK6dNKc3P-z0hratVoTzWIhY-4vt2UeNXm6S47WAhi0ib7-NOBKtRseXvZqyFh841SoRFlEvsNwhh11vJsXW6D6W9olK6jsXGiwQTb2tKQX8PNHQs%3D&expnd=1) from patterns in data.

**AI:** Artificial intelligence, the ability of a computer or computer-controlled robot to perform tasks commonly associated with intelligent beings.

**VADER**: Valence Aware Dictionary for Sentiment Reasoning is an NLTK module that provides sentiment scores based on the words used. It is a rule-based sentiment analyzer in which the terms are generally labeled as per their semantic orientation as either positive or negative.

**roBERTa**: short for “Robustly Optimized BERT Approach” is a variant of the BERT (Bidirectional Encoder Representations from Transformers) model, which was developed by researchers at Facebook AI.

**Assumptions, Limitations, Delimitations**

It is assumed that tweets gathered from twitter users from 23rd January, 2023 to 15th March, 2023 can be used to provide inference on level of acceptance or rejection of a particular candidate running for office the President of the Federal Republic of Nigeria.

The dataset gathered for the study is limited to Twitter(X) in order to save time and resources and to be able to conduct this research within an appropriate time. The requirement for scrapping data from social media platforms are expensive and time consuming thereby we could not go across various platforms to collect data but limit ourselves to Twitter.

Since every candidate is on Twitter and almost all media outlets have twitter handles as well as social media influencers, we discover that twitter will be the best platform to focus on when trying to carry out research of this magnitude.

While there are so many models available to be chosen from to carry out this research. The NLTK VADER and Huggin Face roBERTa models were chosen to test the best model that will well suit this research.