SENTIMENT ANALYSIS

OF

NIGERIAN TWEETS

By

Henry Bankole Akinnawo

CS

COC251

B723411

Supervisor: André Schappo

Department of Computer Science

Loughborough University

April 2021

Link to website:

http://twitter-nigeriamusic-app.herokuapp.com/

Abstract

There are 28.15 million social network users in Nigeria, and this number is forecasted to grow to 44.63 million by the year 2025 (Clement, 2020). Twitter is an online microblogging and social-networking platform on which users interact with each other by sending and receiving short posts of up to 280 characters in length known as “tweets” (‘Twitter’, 2020). Due to the growing number of Twitter users in Nigeria, organizations need to monitor and analyze the reception of their products from their Nigerian consumers. This project is a web application that addresses the problem of sentiment analysis of text data by Nigerians on Twitter. Sentiment analysis is a field of Neuro-linguistic Programming that involves the classification of text into positive, negative, or neutral feelings. Due to the wide advantages sentimental analysis offers to private and public enterprises, many research works have been developed for sentimental analysis. This report focuses on the theoretical and practical aspects of the sentimental analysis. The practical aspect of the sentimental analysis was developed using the Python programming language and the structural query language (SQL) for sentimental analysis on tweet data relating to Nigerian music or musicians. This report describes the procedures for the implementation and results obtained from the experimental setup.

Keywords: sentimental analysis, Nigeria Tweet data

Acknowledgments

I would like to thank my project supervisor, André Schappo, for his advice and support throughout the completion of this project.

Contents

List of Figures

List of Tables

Chapter One: Introduction

1.1 Background

Sentiment analysis is the use of natural language processing, computational linguistics and text analysis to determine a speaker's or writer's attitude toward a specific topic. It basically aids in determining whether a text expresses positive, negative, or neutral sentiments. Sentiment analysis is a great way to find out how people, especially customers, feel about a certain product, topic or concept. The history of sentiment analysis can be traced back to the 1950s, when it was primarily used on written documents. Sentiment analysis, on the other hand, is now widely used to extract subjective information from online content such as text, blogs, tweets, news articles, social media, comments and reviews. This is accomplished through a variety of methods, including natural language processing (NLP), machine learning and statistics. The information gathered is then used by organizations to identify new opportunities and better target their messages to their target demographics.

Sentiment analysis is one of the most rapidly growing research areas in computer science, making it difficult to keep up with all of the activity. People use Social Media such as forums, blogs, review sites, wikis, tweets and social networks to share their knowledge, experiences, and thoughts with the rest of the world. In Web 2.0, this has changed the way people communicate and influence other people's social, economic and political behavior. Indeed, Web 2.0 gives everyone a voice, promising to improve human collaboration capabilities on a global scale by allowing people to share their opinions through user-generated content and read-write Web.

Sentiment analysis is a new branch of Natural Language Processing (NLP) research that aims to detect subjectivity in text and/or extract and classify sentiments and opinions. People's sentiments, opinions, evaluations, attitudes, emotions and appraisals toward services, individuals, , products , issues, topics, organizations , events, and their attributes are studied using sentiment analysis

Sims (2015) defines Sentiment Analysis as the use of natural language processing, text analysis, and computational linguistics to determine the attitude of a speaker or writer towards a specific topic. Sentiment analysis is a way to discover how people, consumers, in particular, feel about a particular topic, product, or idea. Sentiment analysis was originally done via written paper documents however, this becomes an expensive task when operating with large amounts of respondents. Today sentiment analysis is done digitally by extracting information from online content. Blogs, news articles, reviews, comments, and social media are all sources of data used for sentiment analysis. Sentiment analysis can be done by a range of different techniques such as neuro-linguistic programming (NLP), statistics, and machine learning methods.

Clement (2020) reports that millions of Nigerians are connected to the internet and are making use of social networking websites, such as Twitter. Twitter is a free micro-blogging platform that allows users to make public posts (referred to as "tweets") expressing their opinions openly on any topic ('Twitter', 2020). The growing number of Nigerians on the internet is contributing to a rapid rise of big data in Nigeria from social networking websites. According to Alabdullah et al (2018), the rapid rise in the availability of big data via various internet-based technologies such as social media platforms has left many market leaders unprepared for handling very large, random, and high-velocity data. Sentiment analysis, a field of Neuro-linguistic Programming, offers a solution to the problem of extracting useful information from large unstructured data and classifies them based on their polarity. This project is my first venture into the domain of Data Science. I have chosen to undertake this project because I intend on pursuing a career in Data Analytics in the future. Due to me not having any experience in this topic, I will be learning how to do sentiment analysis from scratch.

1.2 Aims and Objectives

This project aims to produce a real-time sentiment analysis monitoring system, which takes tweets from users in Nigeria and outputs the sentiment of the tweets as positive, negative, or neutral. The focus of the analysis is directed on popular Nigerian musicians. The insights gained from sentiment analysis is important for both music artists and the record labels that support the artists. Through real-time sentiment analysis of tweets, we can record how the music consumer's opinion of an artist and their music changes over time. Artists can use the insights to aid in the decision-making of the marketing and promotion of their future projects.

The objectives of this project are the following:

1. Develop a real-time sentiment analysis monitoring system.
2. Collect tweet data from Nigerian users.
3. The collected tweets must include a statement about a musician or their music.
4. Represent the analyzed data on interactive graphs.

5. Deploy the data visualization onto a web application

1.3 Project Scope

This project Focuses on the development of machine learning-based sentimental analysis using technologies and knowledge from the following topics: data mining, text mining, machine learning, natural language processing, python programming language, and structural programming language. The scope of the project also included the development of a web application for operating and visualizing sentimental analytic results. However, it is important to emphasize that the development and practical implementation carried out in this project is considered to be a prototype. The cost of developing a complex real working system led to the choice of the developing prototype system considerable and preferable

Chapter Two: Literature Review

The term ‘sentiment’ refers to attitudes, opinions, and emotions. This means that they are subjective to the individual they are expressed by as opposed to being objective facts. Sentiment analysis can be done in different ways depending on the type of analysis being conducted. The two main types of sentiment analysis are subjectivity/objectivity identification and feature/aspect-based sentiment analysis. Subjectivity/objectivity identification involves classifying a portion of text into one of two categories: subjective or objective. Sims (2015) points out that the key problem with subjectivity/objectivity identification is that a term or a phrase's meaning is often reliant on its context. Feature/aspect identification allows different opinions/sentiments (features) to be defined concerning distinct characteristics of an entity. Unlike subjectivity/objectivity identification, feature/aspect-based identification allows for a more precise analysis of sentiments.

In this project, the aim is to develop a sentiment analysis system concerning the sentiments expressed by Nigerians about music. The application of this system would enable music artists and record labels to gauge how people feel about the artists. Organizations can use the information they have extracted to identify new opportunities, optimize strategic decisions, and align their brand more towards their target demographics.

To understand how to develop a sentiment analysis system for Twitter, an online tutorial (Li, 2019) was followed. This tutorial aims to build an end-to-end real-time Twitter monitoring system. The system developed in this tutorial satisfies some of the central requirements of this project, (real-time analysis with data visualization via a web app) however modifications will need to be made to track the Nigerian consumer’s sentiments towards popular musicians.

To retrieve Twitter data in real-time, Twitter streaming API will need to be used. APIs allow computer programs to communicate with each other by providing software applications with the ability to call unique endpoints (Twitter, 2021). Endpoints are addresses that correspond with a specific type of information. Tweepy, a python library, will be used to access Twitter API. Tweepy makes it easy to access Twitter data by handling authentication, connection, creating and destroying the session, reading incoming messages, and partially routing messages. (Tweepy, 2021).

According to Guru99, (viewed 08/01/2021) Python codes are easier to maintain and more robust than R. The advantage that R has over python is that R is developed by academics and scientists and it is designed to answer statistical problems, machine learning, and data science. R also has a much denser variety of libraries than python, which makes it better for specialized statistical analysis. Guru99 argues that Python is the best choice if the analysis results are needed for use in an application or website, which is the intention of this project. The decision was made to build the sentiment analysis system in Python; its readable syntax makes it easier to work with than R, which is a more technical language.

Anber, Salah, and El-Aziz (2016) claim that NoSQL databases are a necessity when processing big data due to the huge amount of unstructured data from various sources such as Twitter. Furthermore, Jose and Abraham (2020) report that NoSQL databases perform better than relational databases in all cases of their analysis. They also observe that the performance of MongoDB is greater than that of MySQL. Despite this, the decision was made to use a MySQL local database for storage as I have experience working with standard query language which will aid in the speed of development. Implementation decisions were made with the time constraints that are imposed on this project in min. Thus, it was not possible to use a NoSQL database such as MongoDB due to the time required for learning how to use it.

In this chapter, the related topic to this project is studied using literature. The subsections below describe the details about the topics.

2.1 Data Mining

Data mining is the logical search for relevant information in the dataset, the goal of data mining is to find the pattern in the dataset that are previously unknown. The patterns or relationship compute from the data mining techniques are commonly used to determine the decision for the development of business (Ramageri, 2017)

Data mining techniques follow the three steps: exploratory, pattern identification, and deployment. Data mining techniques can be divided into the following categories: classification, regression, and clustering

Data mining is an interdisciplinary method that uses knowledge from mathematics, statistics, and machine learning, and computer science (Hand, 2000). It is considered the computational method for processing data. The algorithms for mining data are divided into two groups: supervised learning and unsupervised learning (Osmanbegovic & Suljic, 2012). Supervised learning is a process in which algorithms are used to learn labeled data for classification or predicting continuous values (regression). A supervised learning approach can be used to extract information in the dataset for classifying similar data. Examples of algorithm that falls into the supervised learning approaches are naïve Bayes algorithm, linear regression, logistical regression, support vector machine, and multilayer perceptron.

Unsupervised learning algorithms use unlabeled data to determine patterns and group data according to some characteristics.

The goal of data mining is to develop a model or system for classification, prediction, and clustering purposes. Data mining is carried out through the process of discovery and analysis of information from a large volume of data. The term big data analytics has emerged as the data analytics process on the large volume of data (Rao & M.Ramesh, 2019).

2.2 Text Mining

Text mining is the branch of data mining that involves extracting relevant information from text data. Text mining becomes popular due to the large volume of text data in every field of study and the increasing demand for text analytics. It is impossible to effectively extract information from text data manually (Salloum et al., 2018). The application of text mining includes web, internet, customer relationship, fraud detection, business, and finance (Preethi & P.Radha, 2019).

In-text mining, a substantive model is used to extract important information from a text database. The techniques used in text mining include information retrieval, summarization of text, clustering of text, and categorization of text (Babu & Srinivasu, 2019).

Text mining is a multidisciplinary study that involves statistics, data mining, information extraction, natural language processing, and information retrieval (Preethi & P.Radha, 2019).

It is a common term used in computational linguistics as the process of processing and structuring text input using different methods such as removal of unnecessary text, parsing, storing text in a database, and interpreting the meaning of the text (Preethi & P.Radha, 2019).

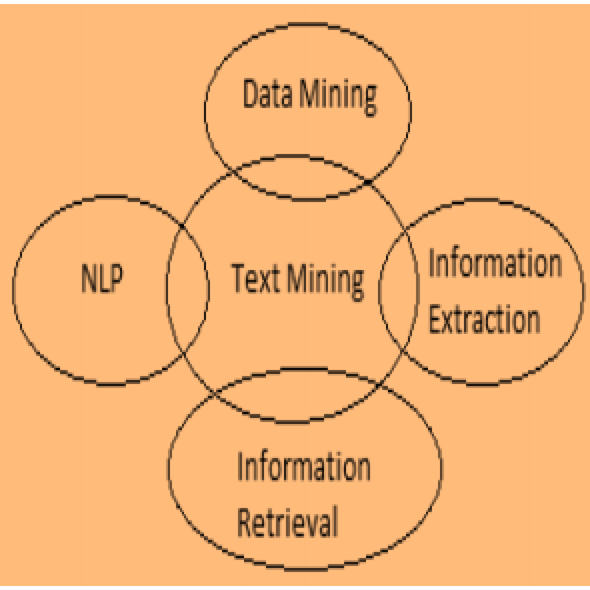


Figure 1 : The architecture of the Text mining

2.3 Natural Language Processing (NLP)

Natural language processing is the branch of artificial intelligence that is concerned with processing native language so that they are understandable by the computer system (Preethi & P.Radha, 2019). NLP involves gathering knowledge that will help computer systems understand the human language properly. NLP can be classified into the following areas: part of speech tagging, machine translation, understanding natural language, generating natural language, and speech recognition (Joseph et al., 2016).

NLP combines computations from knowledge of linguistics which is referred to as language modeling with artificial intelligence techniques and algorithms such as statistical and machine learning. NLP also extends to processing information from voice data to meaningful information understandable by computer (IBM_Cloud_Education, 2020).

In recent years, technology has helped to advance NLP, computer programs can be written to convert a native language to data structures understandable by the computer. Several needs and applications have motivated the use of NLP, one of the simpler applications of NLP is the bag of words features that are used for information retrieval. NLP can be used to extract syntactic information and semantic information from the native language (Collobert et al., 2011). Syntactic information involves extracting information relating to the importance of each word in the sentence such as part-of-speech and phrases. While semantic information involves determining the contextual meaning of the sentence.

Several software or frameworks have been developed for natural language processing. Examples include Natural language Toolkit (NLTK), Spacy, Gensim, Google Cloud translation, Amazon Comprehend, and so on (Bell & Olavsrud, 2020).

2.3.1 Natural Language Toolkit (NLTK)

NLTK is an open-source python framework that provides human language data and processing. NLTK was developed at the University of Pennsylvania, it offers different text processing techniques and libraries for lexical resources (Bell & Olavsrud, 2020).

Natural language Toolkit (NLKT) is considered one of the most powerful libraries that offer statistical processing of text and voice. NLTK can be designed for machine language to understand text or voice and also responds. It provides text processing techniques such as tokenization, word count, character count, stemming, word embedding, and so on (guru99, 2020).

2.3.2 Gensim

Gensim is an open-source python library that provides natural language processing for extracting syntactic and semantic information from the document. Gensim library is used for text processing such as topic modeling, similarity retrieval, and document indexing.

2.3.4 Spacy

Spacy is another library that is used for natural language processing, it offers statistical models and processing pipelines for text. Similar to other NLP libraries, spacy is multi-functional and flexible to use for language processing. It is also capable to perform functions such as part of speech Tagging, entity recognition, and rule-based matching (Joshi, 2020).

2.4 Sentimental Analysis

Huge amounts of user opinion texts have been posted on the Web in the last fifteen years, particularly on social media websites. All of these data, which is growing at an exponential rate, has created incentives for the development of automated methods and tools for analyzing user opinions about various products, brands, services, and other entities.

Sentiment Analysis, also known as Opinion Mining, is a set of techniques, methods, and practices for analyzing sentiments, opinions, or attitudes about specific entities automatically. Some of its most common and popular application realms include market predictions, business intelligence, online marketing and customer service.

Sentimental analysis is the analysis of the text that involves extracting sentiments from raw text. Sentiments are feelings, opinions or emotions express about a particular thing. A private and public enterprise uses sentimental analysis to determine public opinion. The sentimental analysis provides motivations for businesses during marketing and advertisement (Swati et al., 2015). The rapid growth in social websites with more people writing reviews on products has made sentimental analysis (Prakash & Aloysius, 2019).

The sentimental analysis is applied in business, stock market, and review analysis. In business, sentimental analysis can be used to determine the popularity of a product and also capture emotions and public options after displaying the product (Moorjani & Sadath, 2019). In stock market application, sentimental analysis helps the investor to know the best time to buy and sell stock to maximize profits or prevent loss.

The automatic analysis of text for sentiments has being found under different topics: sentimental analysis, opinion mining, evidentially, and subjectivity. Sentimental analysis is becoming more popular in computational linguistics (Taboada et al., 2011).

Sentimental analysis helps companies to determine the option and emotions of investors, buyers, selling and all other parties involved in the marketing of products. In review analysis, the product is reviewed according to the sentiments extracted by the comments made by customers after purchase. The review can be used to improve the customer regarding the customer's comments.

Sentimental analysis can also be regarded as an opinion investigation method for mining web-based information. In many cases, the opinion is classified as to whether they are positive, neutral, or negative. The results from sentimental analysis play a vital in making a business decision (Alsaeedi & Mohammad, 2019).

Public opinion is usually extracted from social media (such as Facebook, Twitter, Instagram, and so on), comments on blog posts, product review websites, and online forums. On many occasions, it is difficult to extract sentiments from all these texts because aside from sentiment. The text extracted from these sources also contains opinions, views, and beliefs.

Text is classified in sentiment analysis according to the following criteria:

- the polarity of the expressed sentiment (positive, negative, or neutral);

- the polarity of the outcome (e.g., in medical texts, improvement versus death)

Text polarity was mostly studied in the early 2000s, starting with phrases or words and working up to the entire document. It was common practice to create lexicons of affect terms and use them to infer word polarity. Later, as the web, social networks, and cloud services became more popular, users were enticed to provide more and more feedback on a daily basis.

User opinions on various items are expressed in a variety of ways. Apart from comments and textposts, social tags have grown in popularity, particularly as a tool for performing sentiment analysis on songs.

They're basically single-word descriptors like "rock," "sweet," or "awesome" that express users' feelings about a particular song or object. Various Music Emotion Recognition (MER) research papers frequently used tags from Last.fm (an online radio station and web platform for music listening with an open API).

At the same time, the growing popularity of microblogs with a lot of user feedback inspired a lot of researchers to create datasets of emotionally labeled texts. As a result, the focus for performing SA gradually shifted from unsupervised to supervised learning methods.

The development of highly effective machine or deep learning techniques, as well as the popularization of high-level libraries or frameworks for using them, fueled this trend. Neural networks have a broad range of applications and can automate feature extraction and selection in a variety of domains and tasks.

The introduction of powerful graphics processing units for scientific computing has improved the speed with which deep neural networks can be trained and used for difficult tasks like language modeling for predicting word combinations. Bengio et al. proposed one of the first neural language models in 2003.

In the years that followed, simpler and easier to train models were proposed. For training and generating word feature representations, also known as word vectors or embeddings, large and increasing amounts of text began to be used. These word features have a number of advantages over traditional bag-of-words text representations, including density,semantic relatability and reduced dimensionality

2.4.1 Approaches for sentimental analysis

Various ways have been developed for sentimental analysis by different researchers. In recent times, artificial intelligence has been adopted for artificial intelligence. Natural language processing (NLP) is used to extract meaning in human language. Hence, sentimental analysis is an interdisciplinary skill from NLP, opinion investigation, statistical analysis, and machine learning.

According to Kharde and Sonawane (2016), approaches in sentimental analysis can be classified into machine learning methods and lexicon-based approaches. The machine-based sentimental analysis involves the use of a classification model to classify text into sentiment. It uses algorithms such as support vector machine (SVM), naïve Bayes, k nearest neighborhood, decision tree, and neural network (Prakash & Aloysius, 2019). The stages in machine-based sentimental analysis can be classified into two: training and testing. Training involves the use of a sentimental database to build a model for future prediction while testing means evaluating the model built for sentimental analysis.

The lexicon-based method involves the use of a dictionary along with opinion words to determine the sentiments in text. This method mainly relies on precompiled words and sentences such as phrases that have been defined and assigned with sentimental scores or classes. The lexicon-based method can be build using a bag of words. It is considered as the process of using the semantic orientation of words to measure the subjectivity of opinion in the document (Taboada et al., 2011)

Lexicon base method is further classified into dictionary-based classification and corpus-based classification. In dictionary-based classification, the text is classified by manual searching for synonyms or antonyms of the word content. Examples of dictionary-based classification software are WordNet and sentiwordNet. Corpus-based classification involves the use of statistical and semantic methods to determine the sentiment of the text data through objectives of the dictionaries (Prakash & Aloysius, 2019).

2.4.2 Sentimental Analysis Software

Several tools have been developed for sentimental analysis with the aid of computer software. Sentimental analytic tools are software that uses technology such as social analytics, machine learning, NLP to analyze text coming from social media, online reviews, online discussion groups e.t.c (Awasthi, 2021).. Though some of the tools require some programming language to handle, however, attempts have been developing to make the sentimental analytic tool very flexible. The sentimental analysis tools rely on the sentimental analysis approaches described in the previous section. While some are dedicated to a particular purpose or platform (such as Twitter and Facebook). Examples of sentimental analytic software are Awario, Talkwalker, Social Searcher, Brandwatch, NCSU, MokeyLearn e.t.c (Bredava, 2020).

The sentimental analytic tool starts with tools that can search collect content on social media and online forums (web scrapper) then proceeded to the addition of external features for analytics. Sentimental analytic software is becoming popular for a wide range of marketing and other business tasks. While there is much sentimental analytics software that is not free, some are available for free. Examples of free sentimental analytic software are Intencheck, social searcher, and Text2data. The aforementioned sentimental analytic software offers different features; hence the choice of a particular software might depend on the factors like cost, flexibility, and purposes (e.g opinion mining, social listening) (Awasthi, 2021).

Sentimental analytic tools are developed using computer programming languages such as python, R, Java, C, C++ e.t.c. Due to the framework and libraries provided by the python programming language, it has become popular for sentimental analysis purposes. Python programming language has been considered as the fastest-growing programming language for data science due to its flexibility

Chapter Three Methodology

This chapter describes the methodologies required for the development of the sentimental analysis using the Twitter extractions. The proposed solution was solved by following the architecture described in the flow chart below. The architecture shows the data collection (tweets), sentimental analysis, database storage, and the displaying in the dashboard.

3.1 Data Collection

The tweet data are required in this project, the proposed method developed for sentimental analysis starts from the data collection. The twitter developer tool is always used for twitter scrapping. Tweets are collected by searching for a search tag or keyword and trend using the twitter API for developer.

3.2 Data Processing

Data processing operation aimed at enhancing the quality of the data obtained from the previous step to improve the effectiveness of the sentimental analysis. Data processing operation applied are specifically required by text. It includes operations required to remove unnecessary part of the text and obtain the highly relevant part that give better meaning. Text operation applied in this project includes removal of double spacing, removal of special characters, removal of website link and removal of username from tweet data

3.3 Tweet Sentimental Analysis

The goal of the sentimental analysis is to determine the sentiments of the text using text analytics. The sentimental analysis is the fundamental technique used in this project, several text mining techniques have been developed for sentimental analysis. The sentimental analysis is presented in the form of the sentimental score which could range from -1.0 to 1.0. The sentimental score can be classified into the classes: positive, negative, and neutral sentiment.

We proposed the machine learning-based sentimental analysis using the supervised learning approach to compute the sentimental score of the text. This method has been discussed in detail in the Literature review chapter. The pretrained machine learning model would be used for classification of text, hence they will be no need to train a new machine learning model for senti ental analysis.

3.4 Python Programming Language

The python programming language is an interpreter, flexible language that is used for web development, software development, data analytics. The python programming language is proposed for the implementation of the project. Python provides libraries that are suitable for the sentimental analysis. The python libraries required for this project are explained in the table 1:

Table 1: Description of the python library required for this project

	Python Library	Description
1	Tweepy	This library is used to connect to the twitter API for tweet extractions
2	NLTK	This library provides the natural language processing (NLP)
3	Pandas	This library provides dataframe extraction, processing and manipulations
4	Textblob	This library has the pretrained model for computing the sentimental score of text
5	Dash	The dash app helps to build dashboard
6	Plotly	The plotly app is used to make visualization on dashboard
7	wordcloud	The worldcloud appp is used to make word cloud figure on dashboard
8	sqlalchemy	This library is used to connect to the database for storing sentimental analytic table
9	apscheduler	This library is used for continuous running of tweet extraction at every interval

3.5 Jupyter Notebook

The Jupyter Notebook is considered an interactive environment for writing python codes. Jupyter notebook makes it easy for writing code through annotations.

3.6 Structural Query Language

The structural query language (SQL) is a relational data language used to organize, update and query data in the database. The SQL also provides relational operation that helps to effectively control and manage data in the database. SQL is specifically used to develop, organize and manage the relational database management system (Knight, 2017). There are several versions of the structural query language such as the SQLite , POSTGRESQL.

Query language allows users to organize and formulate their queries in a simple way without the need for technical requirements.

The structural query language is from the class of the data retrieval language that is specifically designed for data management. There is also a provision for users to specify the properties of the results obtained from the structural database without the need for specifying how the language can process the queried data (Schweikardt et al., 2017).

The structural query language can be classified into the following aspect. The other types of data retrieval languages are XQuery (query language for XML data format), SPAROL, RDOL, and XPath (query language for the RDF graph (Kaur, 2015).

The language elements in structural query language (SQL) are described below

1. Statements
2. Clauses
3. Expressions
4. Queries
5. Predicates

3.6.1 MySQL

MySQL is a database service or software that is used to create and manage a relational database system. A database is regarded as the collection of structured and organized data. The database provides a storage system for structured data. MYSQL was developed by a sweetish company in 1994.

The MYSQL becomes popular because it provides an easy-to-understand and flexible and high-performance querying language (Dwika, 2020).

3.6.2 PostgreSQL

The PostgreSQL is a database system developed at the University of California at Berkeley. It is considered an object-oriented database system that is available commercially. The PostgreSQL is based on SQL and provides modern features. Similar to many database systems, the PostgreSQL is an open-source descendant and provides the same data control elements such as data types, functions, operators, and procedure languages (PostgreSQL, 2021).

3. 7 Text Postprocessing Operation

The post-processing operation is the final step in our proposed framework for the development of the machine-learning-based sentimental analysis. At this stage, the sentimental analytic result is being processed to enhance better meaning and understanding. The postprocessing operation is used to gather the sentimental analytics result in a presentable way. The section below describes the preprocessing operations:

3.7.1 Visualization

Line graph, area graph, pie chart, and word cloud are recommended for visualizing the sentimental analytics result. The sentimental analytics result can be represented in the form. The sentimental analytics result can be arranged according to the time they are collected and determining the distribution of the class of sentiment classes at all times.

Word cloud is used to describe word frequency for all the text extracted for a particular time. Thie method can determine the most frequently used word.

3.7.2 Application Interface for sentimental Analysis

The application interface was is proposed for visualizing the sentimental analysis. With the aid of web app, the sentimental analytic result can be accessed on real time at anywhere. Hence, the Heroku programming interface is considered for app deployment.

3.7.2.1 Heroku

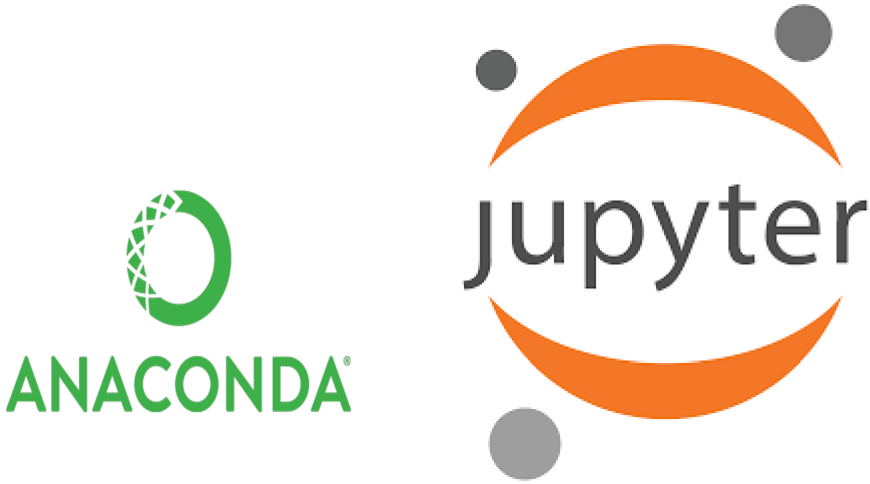
Heroku provides a platform as a service(PaaS) for web application, this service enable developer to deploy and operate programming application on the internet. It can also be considered as a container-based cloud service for enhancing developer productivity (Heroku, 2021). In this report, we use this platform for building the dashboard for our result.

Chapter Four: Implementation

This chapter describes the implementation procedure carried out in this project for the development of sentimental analysis for musicians. The implementation procedure is developed with the formulations in the methodology. The python programming language and the structural query language (SQL) are used in the development. The implementation steps can be classified into the following: Tweet data extraction, text cleaning, sentimental analysis, application programming interface, and model deployment.

4.1 Setting up coding Environment

The setup procedure carried out on the coding environment includes installing the python software, preparation of the development environment, and installing the necessary libraries and framework needed for the sentimental analysis. The code was developed using the anaconda software, which is downloadable from the link: https://www.anaconda.com/products/individual. The jupyter notebook and spyder (software from the anaconda) were used for writing the code.



4.2 Tweet Data Extraction

The tweet data form the basis of this project, the sentimental analysis was carried out on the tweet data. Twitter API was created to extract tweet data from the Twitter developer page using the link: https://developer.twitter.com/en. The procedure used to collect Twitter data is described as follows:

4.2.1 Creating Twitter API

The Twitter API provides access to tweets data programmatically for learning and analytic. To create Twitter API account, the Twitter account is required. The developer account is created by writing a request to Twitter customer service and specifying the purpose (which is sentimental analysis for tweet extraction).

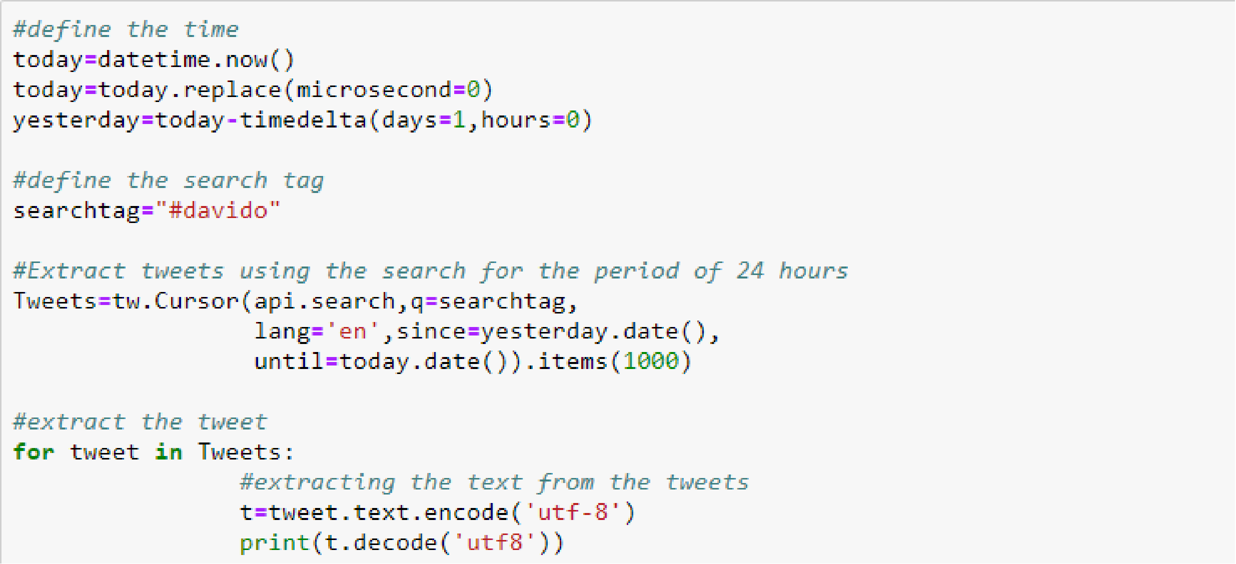
Figure 2 : Twitter API

4.2.2 Obtaining the Access key and Access secrete Key

The authentication details such as the API key, API secret key, access token, and the access secret token, are required to connect to the API for the extraction of tweets are obtained from the API tab on Twitter developer page.

4.2.3 Tweet Extraction

Tweet Extraction code was written to extract tweets for a particular time given a search tag is provided. The search tag is obtained from the list of Nigerian musicians obtained from Wikipedia. The duration for the extraction of the tweet was defined as 24 hours and the total number of tweets to be extracted from each search tag was set to 1000.



```python
#define the time
today=datetime.now()
today=today.replace(microsecond=0)
yesterday=today-timedelta(days=1,hours=0)

#define the search tag
searchtag="#davido"

#Extract tweets using the search for the period of 24 hours
Tweets=tw.Cursor(api.search,q=searchtag,
                 lang='en',since=yesterday.date(),
                 until=today.date()).items(1000)

#extract the tweet
for tweet in Tweets:
                #extracting the text from the tweets
                t=tweet.text.encode('utf-8')
                print(t.decode('utf8'))
```

Figure 3 : Python code for Tweet extraction

4.3 Text Cleaning and Preprocessing

The text extracted from tweets contains characters that portray little or no meaning to the sentimental analysis such as website link, Twitter username, and alphanumeric characters. Hence, it is necessary to subject tweets to text cleaning techniques. The following text preprocessing techniques are applied:

1. **Removal of all special characters**

   The special characters were removed from the tweets such as the alphanumeric symbols such as “/”,”#”, “%”, “@”, “^”

2. **Removal of every single character**

   The single and double character was removed from the text because they seem to portray little or no meaning to the tweet.

3. **Removal of multiple spacing**

   It involves removing multiple spacing from the text to make it more manageable. Sometimes working with tweets, we may have multiple spacing in tweets which is mostly unwanted. So we remove them to make it easy to analyse.

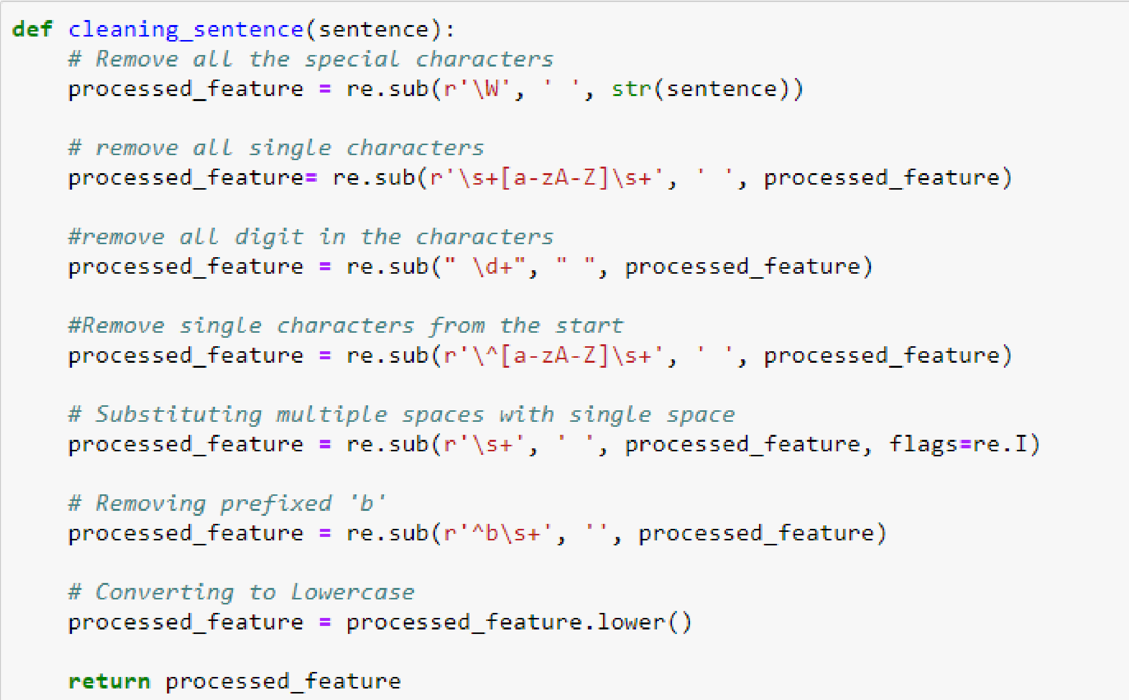
4. **Removal of numerical characters**

   The number characters are removed from the text, though the numerical character might possess a numeric value, however due to the simplicity of the analytic. It was removed from the text.

5. Converting all case to lowercase

All text was converted into the lower case to ease analytics.

The text cleaning was implemented in the python programming language using the python function. The figure below shows the function written in python to clean the text.



```
def cleaning_sentence(sentence):
    # Remove all the special characters
    processed_feature = re.sub(r'\W', ' ', str(sentence))

    # remove all single characters
    processed_feature= re.sub(r'\s+[a-zA-Z]\s+', ' ', processed_feature)

    #remove all digit in the characters
    processed_feature = re.sub(" \d+", " ", processed_feature)

    #Remove single characters from the start
    processed_feature = re.sub(r'\^[a-zA-Z]\s+', ' ', processed_feature)

    # Substituting multiple spaces with single space
    processed_feature = re.sub(r'\s+', ' ', processed_feature, flags=re.I)

    # Removing prefixed 'b'
    processed_feature = re.sub(r'^b\s+', '', processed_feature)

    # Converting to Lowercase
    processed_feature = processed_feature.lower()

    return processed_feature
```

Figure 4 : Python for the text cleaning

4.4 Sentimental Analysis on Text

The sentimental analysis was applied to the processed tweet (that was subjected to the previous technique). The text sentiment was classified into positive, negative and neutral. The following procedure describes the detailed step carried for the application of the sentimental analysis on the processed tweets:

4.4.1 Collecting the tweets extraction extracted from the API

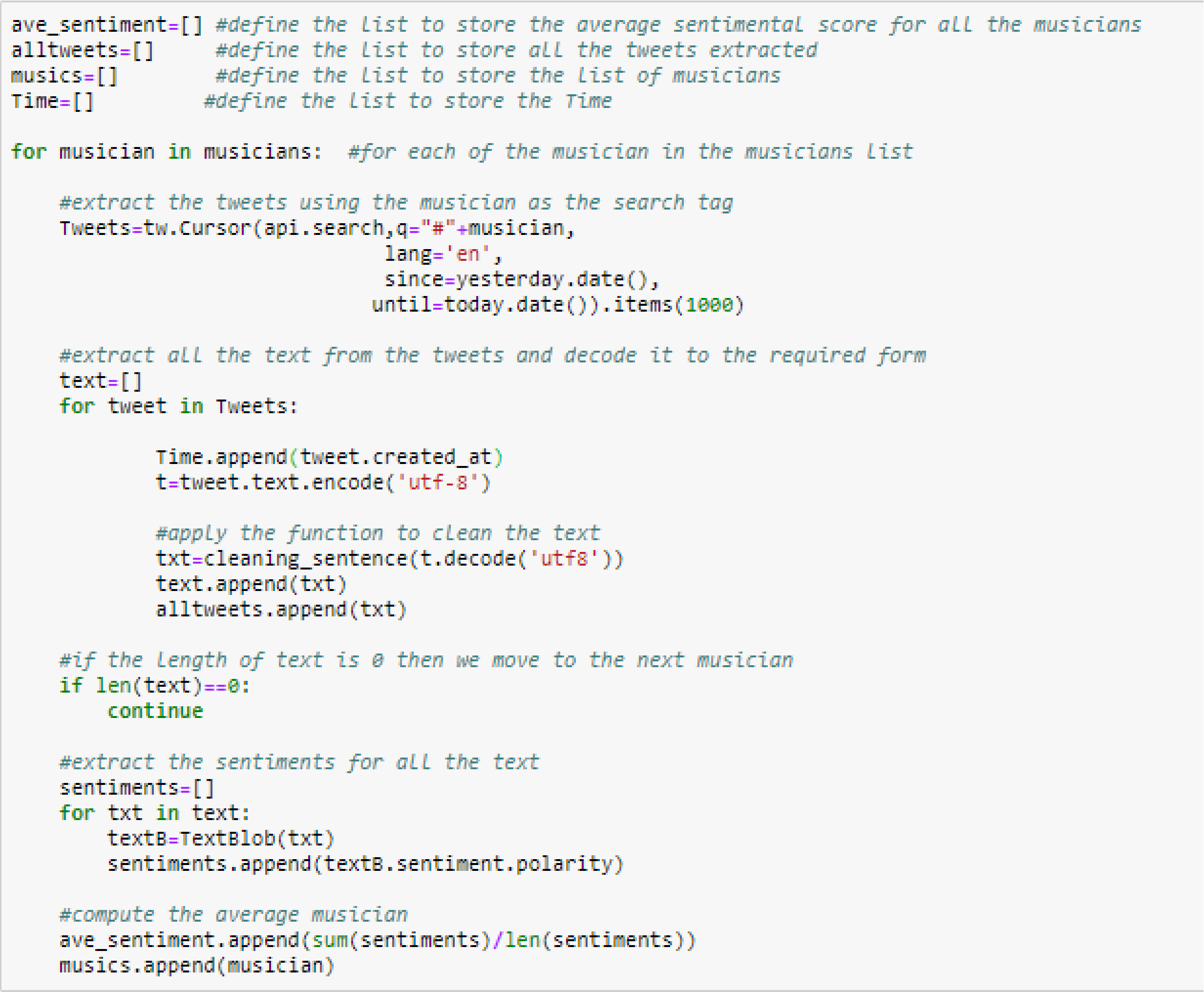
The tweet extraction code uses the procedure explained in section 4.2 to extract the tweet on a real time on an interval of 24 hours. The tweet extraction is designed with a scheduler for this purpose. The analyses from the extracted tweet are saved in a POSTSQLRE database.

4.4.2 Computing the sentimental score

The sentimental score is determined using the text blob framework provided by the python programming language.

4.4.3 Classifying the text based on the sentimental score

The sentimental score is in the range from -1 to 1. Hence, the text is classified as positive sentiment when the sentiment score is greater than or equal to 0.5, the text is classified as neutral if the text is between 0.0 and 0.49 while the text is classified as negative sentiment. when the sentimental score is less than 0.0.



```
ave_sentiment=[] #define the list to store the average sentimental score for all the musicians
alltweets=[]     #define the list to store all the tweets extracted
musics=[]        #define the list to store the list of musicians
Time=[]         #define the list to store the Time

for musician in musicians:  #for each of the musician in the musicians list

    #extract the tweets using the musician as the search tag
    Tweets=tw.Cursor(api.search,q="#"+musician,
                           lang='en',
                           since=yesterday.date(),
                          until=today.date()).items(1000)

    #extract all the text from the tweets and decode it to the required form
    text=[]
    for tweet in Tweets:

            Time.append(tweet.created_at)
            t=tweet.text.encode('utf-8')

            #apply the function to clean the text
            txt=cleaning_sentence(t.decode('utf8'))
            text.append(txt)
            alltweets.append(txt)

    #if the length of text is 0 then we move to the next musician
    if len(text)==0:
        continue

    #extract the sentiments for all the text
    sentiments=[]
    for txt in text:
        textB=TextBlob(txt)
        sentiments.append(textB.sentiment.polarity)

    #compute the average musician
    ave_sentiment.append(sum(sentiments)/len(sentiments))
    musics.append(musician)
```

Figure 5 : Python code for the tweet extraction and sentimental analysis

4.5 Storing sentimental analysis

The sentimental analysis is stored in the form of tables in the database using the structural query language (SQL). The POSTSQLRE database is required for this purpose.

The sentimental analysis tables are described below:

1. The table showing the number of positive, negative, and neutral sentiments extracted for 24 hours
2. The table showing the average sentimental score for each musician based on the sentimental analysis on tweet extracted.

The structural query language (SQL) is used to select and export the required table to the database. The following SQL tags such as “SELECT”,” FROM” and UPDATE

4.6 Tweet Analytics

The tweet analytics was developed using visualizations showing the number distributions of tweet sentiments for all classes and the top-rank politicians based on the tweets. The visualization was developed using the area graph, pie chart, and image. A word cloud was created to shows the commonly used word.

4.7 Application programming interface (API)

The API was built using the python dash-plotly framework, the API code was designed to describe visualizations showing the percentage for each class of the sentiments extracted for the last 24 hours, the high ranked musicians based on sentiments.

4.7.1 Dash-plotly Dashboard

Dash is a framework provided by the python programming language for building the friendly user interface for data visualization and analytics. The application developed by the dash framework is rendered in a web browser. The dash app provides application features such as Html tags, graph layout, button, and sectioning.

The Plotly Framework is used for making visualization such as pie chart, bar chart, pie chart, line chart, and the area chart.

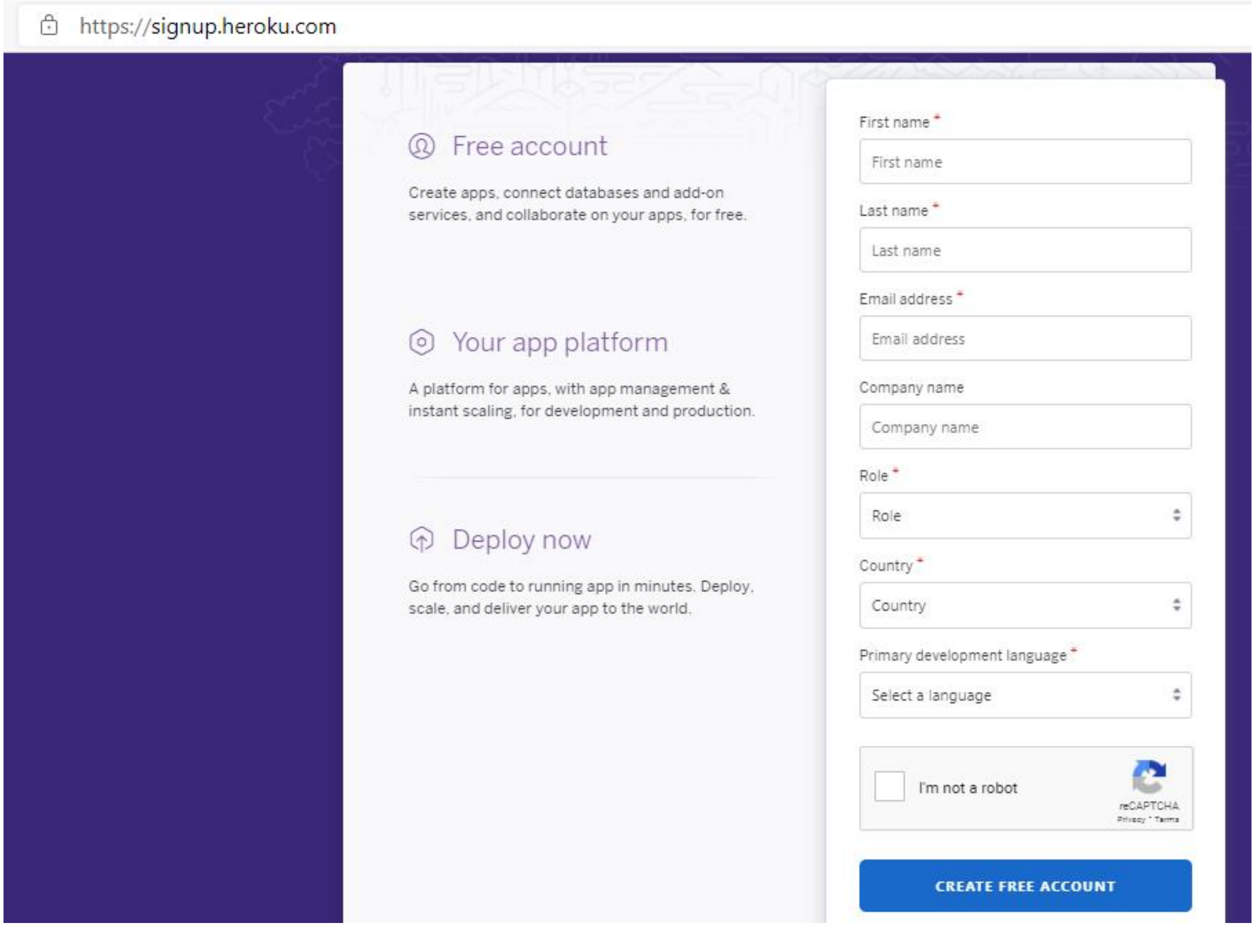
4.8 Model Deployment

The application programming interface (API) was deployed in Heroku platform and can be accessed using the web link. The detailed procedure carried out for the development of the API runs the code.

The procedures carried out for the deployment of the model are described below:

4.8.1 Creating Heroku account

The Heroku account was created by visiting the official website provided by the link: https://signup.heroku.com and supplying the necessary information such as the user First name, last name, email, password, and other information as described below:



https://signup.heroku.com

Free account

Create apps, connect databases and add-on services, and collaborate on your apps, for free.

Your app platform

A platform for apps, with app management & instant scaling, for development and production.

Deploy now

Go from code to running app in minutes. Deploy, scale, and deliver your app to the world.

Figure 6 : Creating Heroku account

4.8.2 Setting Environment for Heroku with Python

The Heroku server provides an app development platform for a different versions of the Heroku can be set up on different types of operating software such as Windows, Linus, and Ubuntu. Heroku provides the command line code for working around the application on the Heroku server. The installer for the Heroku command line was downloaded from the link: https://devcenter.heroku.com/article.

Once the installer is downloaded, the command line interface was tested by accessing the Heroku server using the code below:

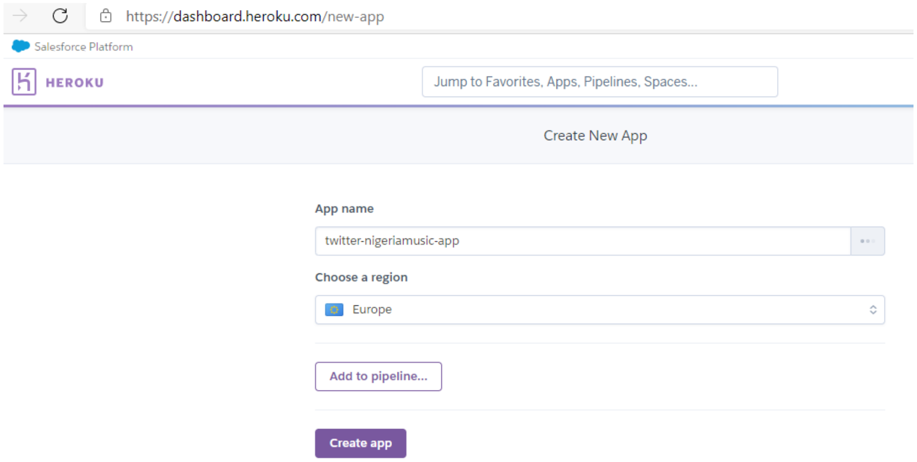
4.8.3 Installing the GIT software

The GIT is the command-line codes that are used for committing and browsing through the Heroku web server through other third-party tools. The GIT was downloaded from the https://git-scm.com/download and installed on the Windows computer.

4.8.4 Creating App on Heroku

Heroku provides free cloud services for the deployment of application programming interfaces on websites. The Heroku app is deployed to dynos, a Linux container that forms the basis of the Heroku platform. The Heroku application is accessible all the time (24/7) and provides continuous commitment if changes to the webserver

The Twitter app was created was by visiting the link: https://dashboard.heroku/new-app, the app name (Twitter-Nigeria-app was defined and deployment region (Europe) was defined as shown in the figure below.



https://dashboard.heroku.com/new-app

Create New App

App name: twitter-nigeriamusic-app

Choose a region: Europe

Figure 7 : Creating App on heroku for the dashboard

4.8.5 Setting up Heroku Postgres

The Heroku Postgres is the infrastructure on the Heroku that manages the SQL database which is accessible through any programming language using the SQL scripting language. The Heroku Postgres is installed using the command line code in figure 1 and the free plan is specified. Consequently, Postgres is linked with the application using the figure.

4.8.6 Setting up the deployment environment

The set up for the model is described using the architecture described below

a) **Musician list**: The musician file is the text file that contains the list of musicians in Nigeria

b) **Requirements:** The requirements that is described by the requirements.txt file contain the python libraries that are required for the working process of the application

c) **Procfile :** The procfile describes the protocol understandable by the Heroku server that defines the rules of the application

d) **gitignore:** This file is part of the app development files that is used to commit the transient file.

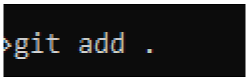
e) **Clock File:** This is the python file responsible for the tweet extraction. It is being controlled by the Heroku scheduler application. The python file is designed to run at every interval of one hour for tweet extraction.

f) **App file:** The application file is the python file for building the dashboard. The python file function is to extract the sentimental analytic results from the database as visualizations display on a dashboard.

4.8.7 Deploying App on Heroku

The app is deployed on Heroku anytime changes are made, the procedure for deployment of the app is described below:

4.8.7.1 Adding the changes

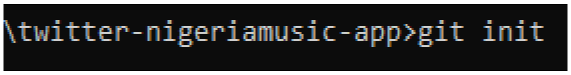


```
>git add .
```

Figure 8 : Adding the changes to the webserver

4.8.7.2 Initialization

It is required to initialize the website every time changes are required on the web site

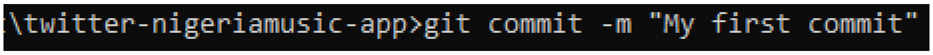


```
\twitter-nigeriamusic-app>git init
```

Figure 9 : Initializing the changes

4.8.7.3 Committing the changes

The changes made on the application can be committed to the website using the Heroku command line code shown in the figure below

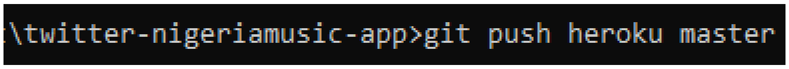


```
:\twitter-nigeriamusic-app>git commit -m "My first commit"
```

Figure 10 : Committing the changes to Heroku server

4.8.7.4 Pushing the application on Heroku server

The application is pushed to the web server so that changes can be effected using the link for the webserver.



```
:\twitter-nigeriamusic-app>git push heroku master
```

Figure 11 : Pushing the application to the Heroku server

Chapter Five: Result and Discussion

This chapter describes the sentimental analytic results obtained from the dashboard app created by following the implementation procedures described in the previous chapter. The result for the sentimental analysis is described below:

5.1 Tweet Extraction

The table 2 shows the samples of the processed text extracted from the tweet data at a different time. The processing operation wa carried out as describes by the methodology and implementation.

Table 2 : Samples of the processed text extracted from tweet data

	Search Tag (Musician)	Time	Tweet Extraction
1	Davido	2021-04-30 00:03:23	notice_com_ng growing demand democratic system local government nigeria collins okeke opinion
2	Simi	2021-04-30 00:04:51	timayo_ priceless moment rooting years season voice nigeria timothyayodele _nuelayo teamayobr
3	Olamide	2021-04-30 00:09:15	after inter milan italian title sunday romelu lukaku couldn resist taking swipe zlatan ibrahimov h2q9u6lvg0
4	Falz	2021-04-30 23:57:01	RT @chizzyofficia: I'm Badboychizzy... I'm the sound of the New Generation...I motivate cos hustle comes before love #dahbadboyyear taking…
5	Phyno	2021-04-30 23:59:59	kaywise phyno lyrics follow ofofonobs ofofonobs trending moneyheist moneyheist5 ad7knzhlxr
6	Burnaboy	2021-04-30 00:03:23	refinedng burna certified united states burnaboy congratulations odogwu interesting articles
7	Rema	2021-04-30 00:04:51	Show some love guys❤❤🔥 #BurnaBoy #burnathrowback #explorepage #Explorer #Rema #Accounting #access #kizzdaniel…
8	Laycon	2021-04-30 00:09:15	Work “n” Grind has been on a stoopid repeat in here.....The guy make a stoopid hit on that song #COTW #conti
9	Falz	2021-04-30 23:57:01	I'm Badboychizzy... I'm the sound of the New Generation...I motivate cos hustle comes before love #dahbadboyyear taking…
10	Timaya	2021-04-30 23:59:59	timaya eedris abdulkareem dirty looking smoker blames

5.2 Sentimental Analytic Result

The sentimental analytic result was presented in the form of tables and visualizations. Table 3 shows the sample of 10 records from a sentimental analytic table. The table describes the distribution of positive, negative, and neutral sentiment at different times. The table shows that the Neutral class has the highest number of records for samples observed.

The table 4 shows the top musicians that are rank based on the sentimental score for 24 hours. Since the sentimental analysis was carried out in real-time, the result keeps changing showing the top musicians ranked based on sentiments from tweets for the last 24 hours.

Table 3: The table showing the distribution of sentiments extracted at every time

	Time	Positives (%)	Negatives (%)	Neutral (%0
1	2021-04-30 00:03:23	35	38	10
2	2021-04-30 00:04:51	35	38	10
3	2021-04-30 00:09:15	35	38	10
4	2021-04-30 23:57:01	35	38	10
5	2021-04-30 23:59:59	35	38	10
6	2021-04-30 23:57:26	35	38	10
7	2021-04-30 23:59:59	35	38	10
8	2021-04-30 23:59:59	35	38	10
9	2021-04-30 23:59:59	35	38	10
10	2021-04-30 23:59:59	40	38	10

Table 4 : The table showing the average sentiments extracted for each of the musicians

	Musicians	Sentimental Score
1	Brymo	0.75
2	Simi	0.39
3	Don Jazzy	0.37
4	Falz	0.293
5	Phyno	0.194
6	Timaya	0.1604
7	Davido	0.1533
8	Olamide	0.140167
9	Adekunle Gold	0.140167
10	Laycon	0.128004

5.3 Result obtained from the Dashboard

The dashboard shows the sentimental analytic results in the form of visualizations such as area graphs, ring pie charts, bar charts and images. Figure 12 shows the overall view of the dashboard as users see it.

Figure 13 shows a similar result as table 3 which the area graph is describing the distribution of the positive, negative, and neutral tweets extracted every 1min for the last 24 hours. Figure 14 show a similar result as table 4 that is the horizontal bar chart of the sentimental analytics result obtained for ranking the musicians based on the sentimental score.

The ring pie chart in figure 15 shows the summary for all sentimental scores in the form of percentage distribution, the figure indicates more than half of the tweets extracted for 24 hours fall into the neutral sentimental class. Furthermore, only 10.5% of tweets fall into the negative sentimental class, while the positive sentimental class taking the second largest after neutral.

Figure 16 shows the most commonly used for all the tweets extracted for the period in the form of an image known as the word cloud. The frequency of words is described by the size of the word on the word cloud. The word with largest size describes the most frequently used word for all the tweets extracted.

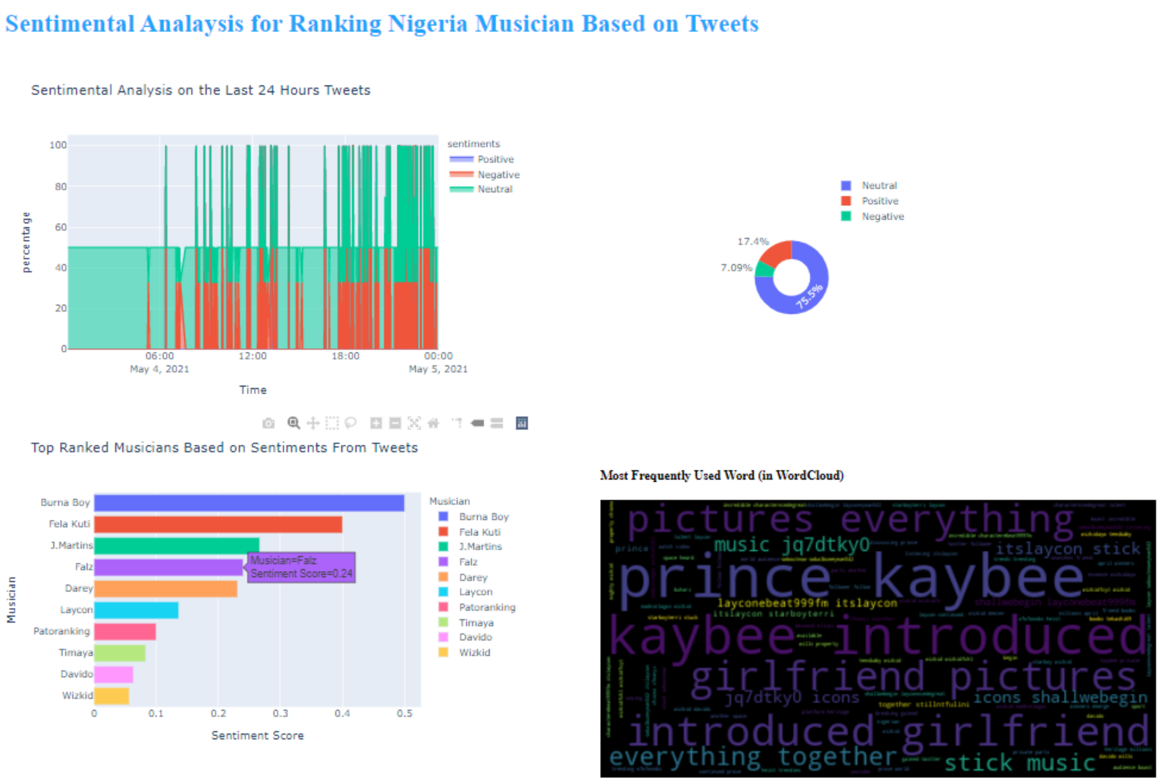


Figure 12: Overall view of dashboard

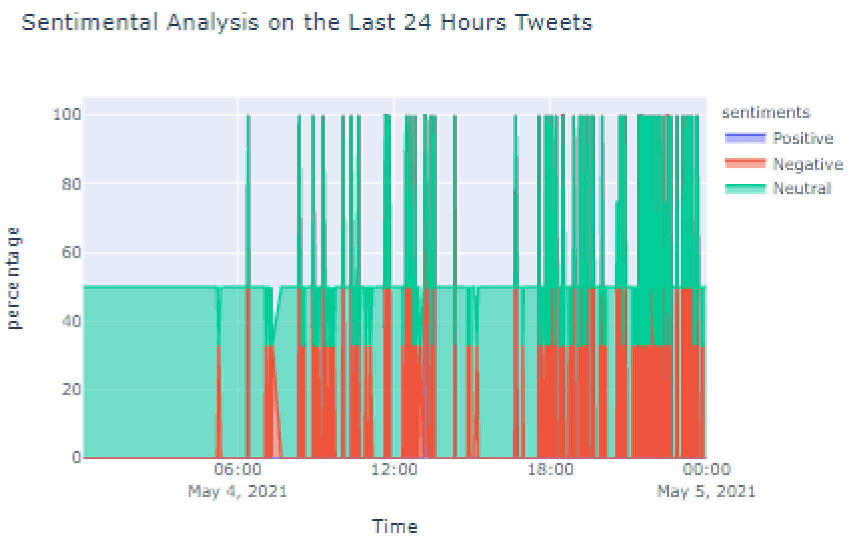


Figure 13: Area graph describing distribution of positive, negative and neutral tweets

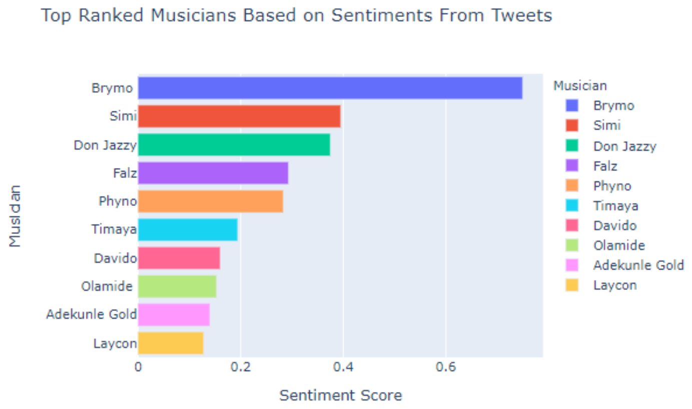


Figure 14: Visualization showing the Top Ranked Musicians based on the sentiments from tweet

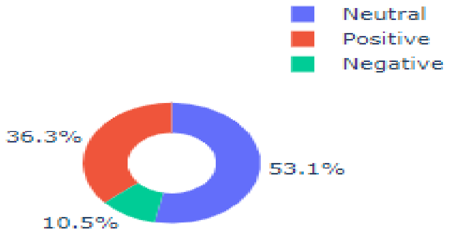


Figure 15: Visualization showing the summary of the sentimental classes for all the tweets extracted

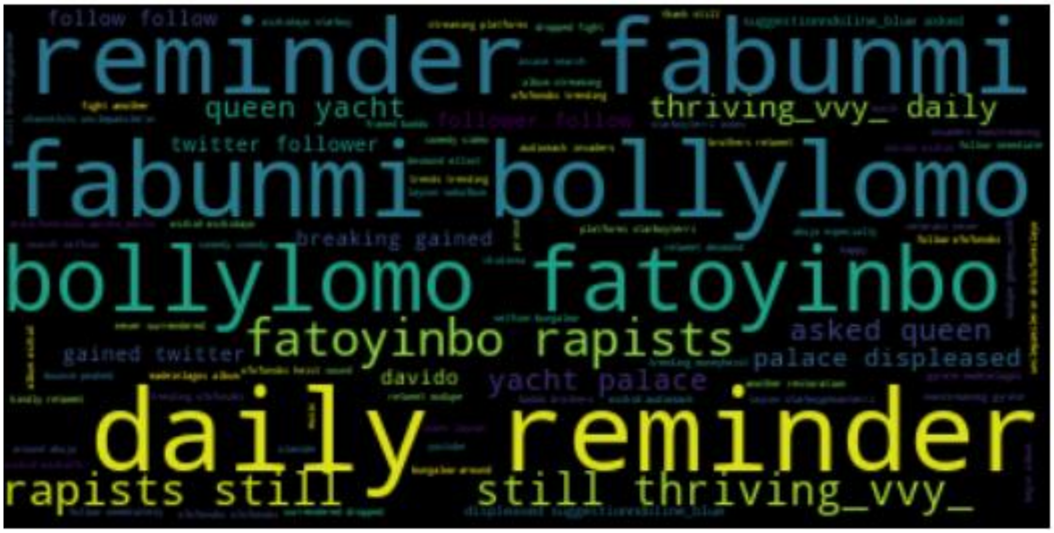


Figure 16: Visualization showing the most commonly used words

Chapter Six: Evaluation

This chapter describes the procedure carried out for evaluating the implementation procedures and overall project. The evaluation chapter aims to determine if the objectives of the project were met by reviewing all the chapters. The evaluation was done based on the results obtained and the studies from the theories in this project. The details of this chapter are included in the following subsections:

6.1 Sentimental Analysis

The result obtained in the previous sections shows that the implementation procedures carried out were able to develop proper analytic from the tweet data extracted. The analytics which was presented in the form of plots of different types is considered to be a sentimental analytic results. This is because it describes the summary of opinion, expression, and views, which is known as sentiment expressed by large numbers of people on Twitter. From literature studies, it was learned that sentimental analysis can be presented in the form of scores that ranges from -1.0 to 1.0 or 0.0 to 1.0. The results presented in the previous chapters utilized this knowledge for tweet classification into three classes which are positive, negative, and neutral. Therefore, this project uses theoretical knowledge from literature to achieve sentimental analysis

Though the sentimental analysis was applied to Nigerian musicians, considering our implementation, it can also be applied to another area with little changes. Consequently, the sentimental analysis developed by this project is not only restricted to an application

6.2 Sentimental analytic Software

The software developed for sentimental analytics was evaluated by testing the performance of the application. The application which is accessible from a web page through a web link was visited every hour of the day. The application was accessible all the time it was visited and shows different results. The sentimental analytic results throughout the time it was visited differs, which indicates the sentimental analytics kept updating thereby operating in real-time.

Furthermore, the backend of the application is evaluated to determine the logs if there is an error in our scheduler. The scheduler is responsible for extracting tweets and saving it in the Postgre database. The log error was observed to evaluate if the application and the database are working perfectly. The application was developed on Heroku platform by running the codes written in python language on the platform. The implementation of the API was in line with the objective of the project.

Chapter Seven: Conclusion

This report describes the theories, formulation, and implementation of the sentimental analysis. In this project, the related topics to sentimental analysis such as Natural language processing, machine learning, text mining, and so on were described in detail using literature studies. The studies describe the approaches used for sentimental analysis which was classified as the lexicon-based and machine learning-based sentimental analysis. In this project, the machine learning-based sentimental analysis was adopted. The implementation was carried out using the python programming language.

A real-time sentimental analysis of the tweet extracted was achieved using a web application. The application generated sentimental results for the last 24 hours in the form of plots and images every 2 hours. The application was developed on Heroku web platform with a provision to save information extracted on a Postgre database.

Though the implementation was able to achieve substantial results for data analysis, our limitation is the inability to collect a larger number of tweet due to the free version

The result obtains shows that the sentimental analytics result in different plot formats fitted in a dashboard. Musicians were able to develop. The evaluation procedure carried out in the previous chapter was aimed at evaluating the effectiveness of the project towards achieving its objectives as described in chapter one.

`

References

‘Twitter’ (2020). Wikipedia. <https://en.wikipedia.org/wiki/Twitter> (Accessed: 24 December 2020).

About Twitter’s APIs. (n.d.). https://help.twitter.com/en/rules-and-policies/twitter-api. (viewed 7th January 2021).

Alabdullah, B., Beloff, N., White, M. (2018). Rise of Big Data – Issues and Challenges. 1-6. 10.1109/NCG.2018.8593166.

Anber, H. Salah, A. and El-Aziz, A. (2016). A Literature Review on Twitter Data Analysis. International Journal of Computer and Electrical Engineering. 8(3). pp.241-249.

Clement, J (2020), Number of social network users in Nigeria from 2017 to 2025, <https://www.statista.com/statistics/972907/number-of-social-network-users-in-nigeria>, (Accessed: 24 December 2020).

Jose, B. and Abraham, S., (2020). Performance analysis of NoSQL and relational databases with MongoDB and MySQL. Materials Today: Proceedings, 24, pp.2036-2043.

Li, Chulong. (2019). Real-time Twitter Sentiment Analysis for Brand Improvement and Topic Tracking (Chapter 1/3). https://towardsdatascience.com/real-time-twitter-sentiment-analysis-for-brand-improvement-and-topic-tracking-chapter-1-3-e02f7652d8ff. (viewed 7th January 2021).

R Vs Python: What’s the Difference? (n.d.). <https://www.guru99.com/r-vs-python.html>. (viewed 8th January 2021).

Sims, Scott. (2015). Sentiment Analysis 101. https://www.kdnuggets.com/2015/12/sentiment-analysis-101.html. (viewed 6th January 2021).

Tweepy Documentation. (n.d.). http://docs.tweepy.org/en/latest/. (viewed 8th January 2021).

Alsaeedi, A. & Mohammad, Z.K., 2019. A Study on Sentiment Analysis Techniques of Twitter Data. *International Journal of Advanced Computer Science and Applications,* 10(2).

Awasthi, B., 2021. *Top 3 Free Twitter Sentiment Analysis Tools*. [Online] Available at: https://www.softwareadvice.com/resources/free-twitter-sentiment-analysis-tools/ [Accessed 30 April 2021].

Babu, G.L.A. & Srinivasu, B., 2019. A Conceptual Based Approach in Text Mining: Techniques and Applications. *International Journal of Innovative Techniology and Exploring Engineering*, 8(7), pp.1779-81.

Bell, T. & Olavsrud, T., 2020. *What is natural language processing? The business benefits of NLP explained*. [Online] Available at: https://www.cio.com/article/3258837/what-is-natural-language-processing-the-business-benefits-of-nlp-explained.html [Accessed 30 April 2021].

Bredava, A., 2020. *Top 10 best free and paid sentiment analysis tools*. [Online] Available at: https://awario.com/blog/sentiment-analysis-tools/ [Accessed 3 May 2021].

Clement, 2020. *Number of social network users in Nigeria from 2017 to 2025*. [Online] Available at: https://www.statista.com/statistics/979207/number-of-social-users-in-nigeria [Accessed 20 April 2021].

Collobert, R. et al., 2011. Natural Language Processing (Almost) from Scratch. *Journal of Machine Learning Research*, 12.

Dwika, H., 2020. *What is MySQL: MySQL Explained For Beginners*. [Online] Available at: https://www.hostinger.com/tutorials/what-is-sql [Accessed 2021].

guru99, 2020. *NLTK Tutorial in Python: What is Natural Language Toolkit?* [Online] Available at: https://www.guru99.com/nltk-tutorial.html [Accessed 30 April 2021].

Hand, D.J., 2000. Data Mining: New Challenges for Statisticians. *SAGE*.

IBM_Cloud_Education, 2020. *Natural Language Processing*. [Online] Available at: https://www.ibm.com/cloud/learn/natural-language-processing [Accessed 30 April 2021].

Joseph, S.R. et al., 2016. Natural Language Processing: A Review. *International Journal of Research in Enginnering and Applied Sciences*, 6(3), pp.207-10.

Joshi, P., 2020. *spaCy Tutorial to Learn and Master Natural Language Processing (NLP)*. [Online] Available at: https://www.analyticsvidhya.com/blog/2020/03/spacy-tutorial-learn-natural-language-processing/ [Accessed 30 April 2021].

Kaur, B., 2015. Data Retrieval Languages. *International Journal of Emerging Trends in Engineering and Development*, 4(5), pp.28-32.

Kharde, V.A. & Sonawane, S.S., 2016. Sentiment Analysis of Twitter Data: A Survey of Techniques. *International Journal of Computer Applications*, 139(11), pp.5-15.

Knight, M., 2017. *What Is Structured Query Language (SQL)?* [Online] Available at: https://www.dataversity.net/structured-query-language-sql/# [Accessed 25 April 2021].

Moorjani, G. & Sadath, L., 2019. Sentiment Analysis- A tool for Data Mining in Big Data Analytics. *International Journal of Innovative Technology and Exploring Engineering*, 8(9).

Osmanbegovic, E. & Suljic, M., 2012. Data Mining Approach for Predicting Student Performance. *Econstor*, 10(1), pp.3-12.

PostgreSQL, 2021. *PostgreSQL 10.16 Documentation*. University of California.

Prakash, T.N. & Aloysius, A., 2019. A Comparative study of Lexicon based and Machine learning based classifications in Sentiment analysis. *Integrated Intelligent Research* , 8(1), pp.43-47.

Preethi, B.M. & P.Radha, 2019. A Survey Paper on Text Mining - Techniques, Applications And Issues. *Journal of Computer Engineering*, pp.46-51.

Ramageri, B., 2017. Data Mining Techniwues and Applications. *Indian Journal of Computer Science and Engineering* , 1(4), pp.301-05.

Rao, N. & M.Ramesh, 2019. A Review on Data Mining & Big Data, Machine Learning Techniques. *International Journal of Recent Technology and Engineering* , 6S2(914-916), p.7.

Salloum, S.A., Al-Emran, M., Monem, A.A. & Shaalan, K., 2018. Using Text Mining Techniques for Extracting Information from Research. *Studies in Computational Intelligence*, pp.373-97.

Schweikardt, N., Schwentick, T. & Segoufin, L., 2017. Database theory: Query languages. In *Query languages*. CRC Book.

Swati, U., Pranali, C. & Pragati, S., 2015. Sentiment Analysis of news articles using machine learning approach. *International Journal of Advances in Electronics and Computer Science*, 2(4), pp.114-16.

Taboada, M. et al., 2011. Lexicon-Based Methods for Sentiment Analysis. *Computational Linguistics*, 37(2).