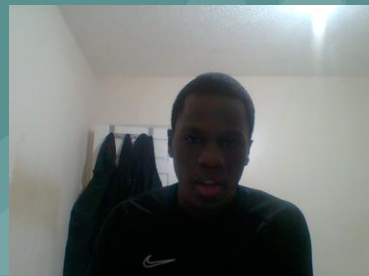
Sentiment Analysis of Nigerian Tweets

By Henry Bankole Akinnawo



Introduction

- Sentiment analysis was originally done via written paper documents
- This becomes an expensive task when operating with large amounts of respondents.
- Today sentiment analysis is done digitally by extracting information from online content.
- Sentiment analysis can be done by a range of different techniques such as neuro-linguistic programming (NLP), statistics, and machine learning methods.

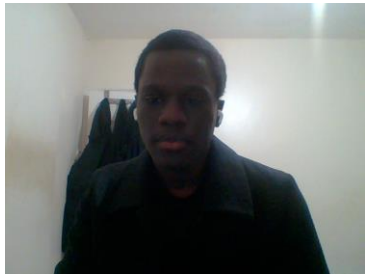
Introduction

Project Aim

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.

Project Objectives

- To demonstrate sentimental analysis using tweet data
- To develop real time sentimental analytic dashboard using Heroku App
- To store the analytic results in Postgre database



Literature Review:Sentimental Analysis

- 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 sentiment analysis more needed. (Prakash & Aloysius, 2019).

Literature Review : Related Topics

01 **Data Mining**

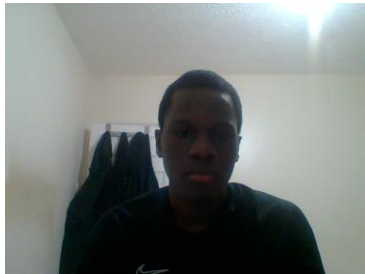
Data mining is the logical search for relevant information in the dataset, the goal of data mining is to find the pattern in the dataset that are previously unknown. (Ramageri, 2017)

02 **Text Mining**

Text mining is the branch of data mining that involves extracting relevant information from text data.

03 **Natural Language Processing**

Natural language Toolkit (NLKT) is considered one of the most powerful libraries that offer statistical processing of text and voice.



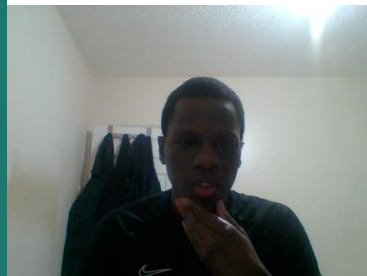
Literature Review: Sentimental analysis methods

Lexicon base method

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.

Machine Learning Base

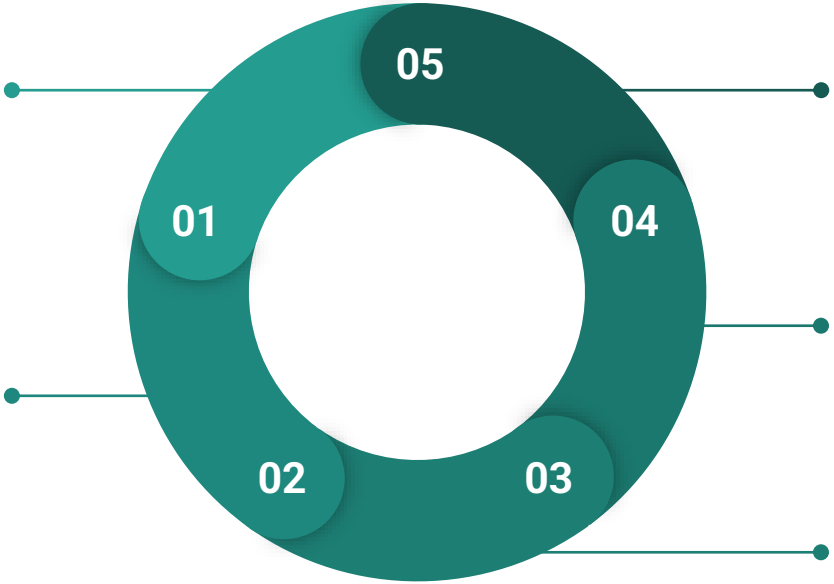
The machine-based sentimental analysis involves the use of a classification model to classify text into sentiment. It uses algorithms such as support vector machine (SVM), naïve Bayes, k nearest neighborhood, decision tree, and neural network (Prakash & Aloysius, 2019)



Methodology: Process Flow Chart

Tweet Data Collection

Text Preprocessing



Building dashboard (web App)

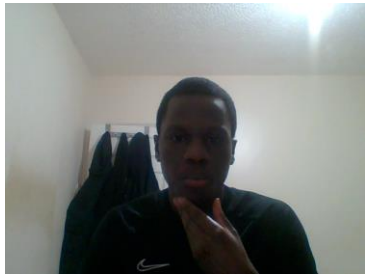
Storing result in database

L

Tweet Sentimental Analysis

Methodology : 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 developers.



Methodology : Text Preprocessing

- Removal of all special characters
- Removal of every single character
- Removal of multiple spacing
- Converting all case to lowercase
- Removal of all external link
- Removal of all username

Methodology: Machine learning based Sentimental Analysis

Machine Learning Pretrained Model

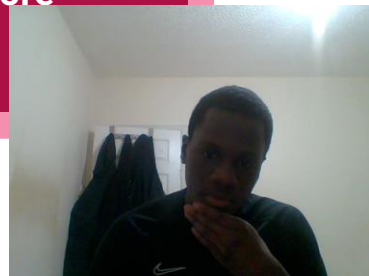
The Textblob library provided by python was used to compute sentimental score

Text classification

Classify the text based to positive, negative and neutral based on sentimental score

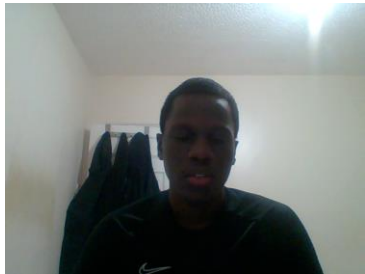
Processed Text input

The processed Text is fed to the pretrained model



Methodology : Tweet Sentimental Analysis

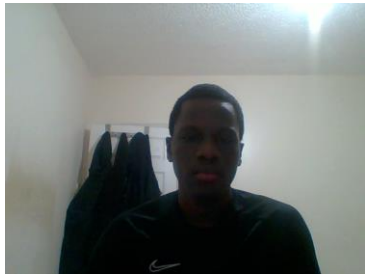
- Collecting the tweets extraction extracted from the API
- Applying text processing function on the extracted tweet
- Computing the sentimental score: The sentimental score is determined using the text blob framework provided by the python programming language
- Classifying the text based on the sentimental score



Methodology : Storing Sentimental analytic in Database

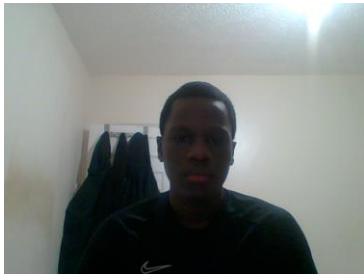
The sentimental analysis is stored in the form of tables in the Postgre database using the structural query language (SQL).

PostgreSQL is based on SQL and provides modern features. Similar to many database systems, PostgreSQL is an open-source descendant and provides the same data control elements such as data types, functions, operators, and procedure languages (PostgreSQL, 2021).



Methodology: Development of Dashboard

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.



Methodology: Tools for building Dashboard

Dash

Python Framework for building dashboard in web app

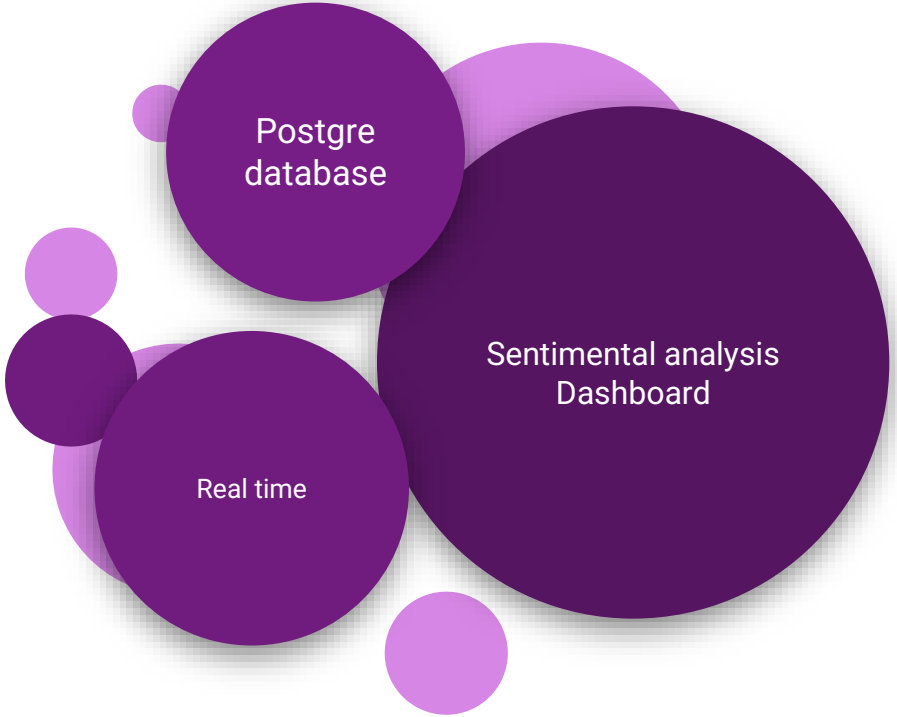
plotly

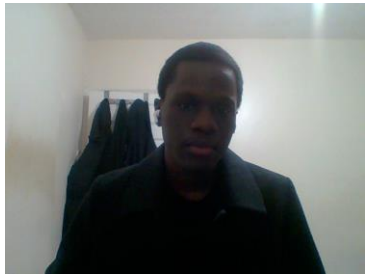
Plotly is used for making visualization

Heroku

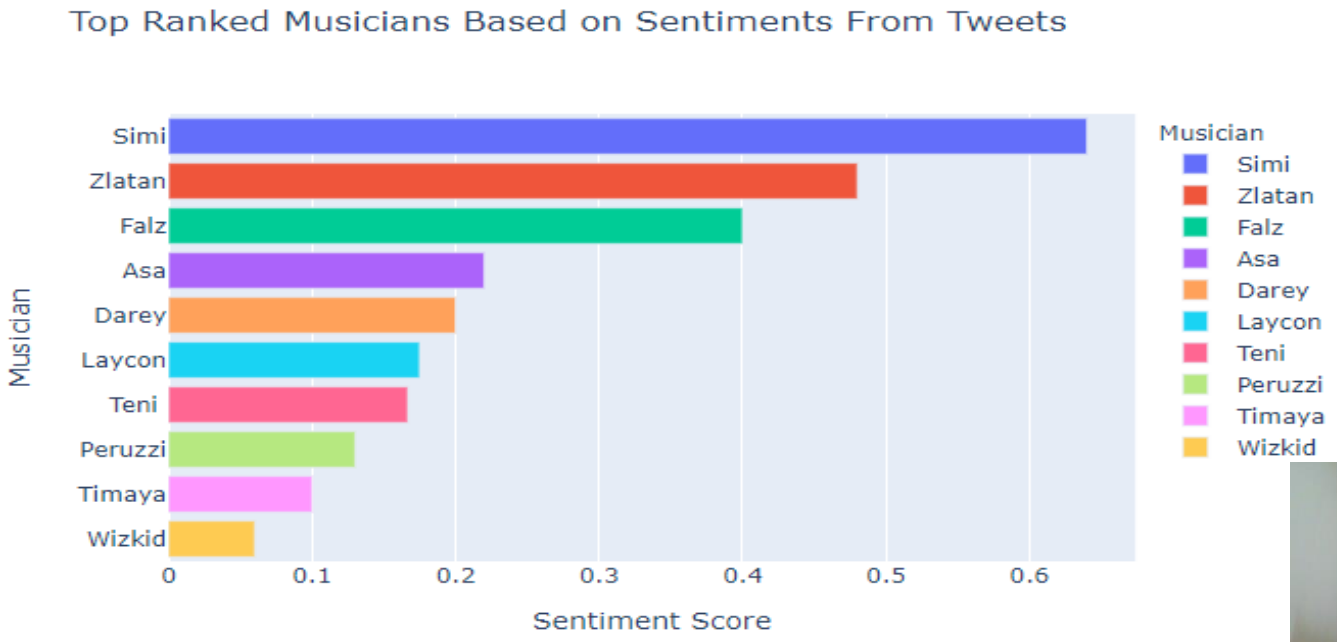
Heroku provides a platform as a service(PaaS) for our application

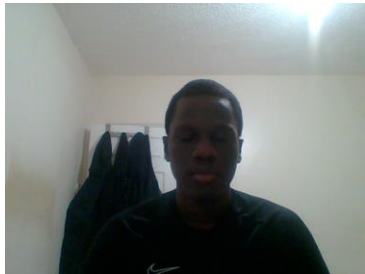
Methodology : Web App Features



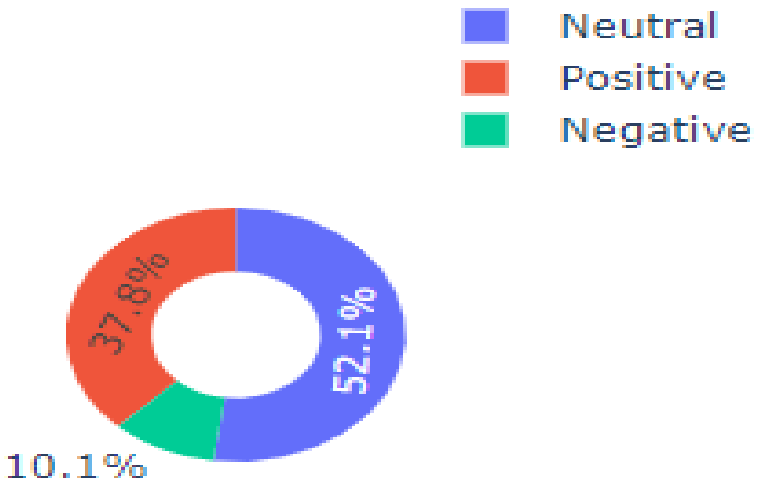


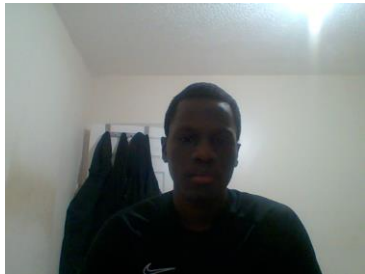
Project Achievement : Sentimental analytic results





Project Achievement: sentimental analytic results

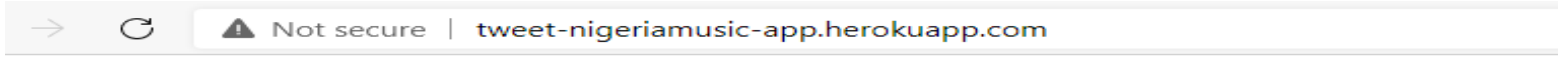




Project Achievement: sentimental analytic results

Most Frequently Used Word (in WordCloud)

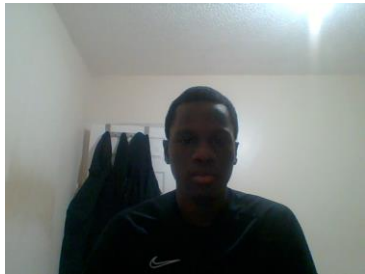
Project Achievement: web App



Conclusion: Project Summary

The machine learning-based sentimental analysis was adopted and the implementation was carried out using the python programming language, structural query language, and the Heroku web platform

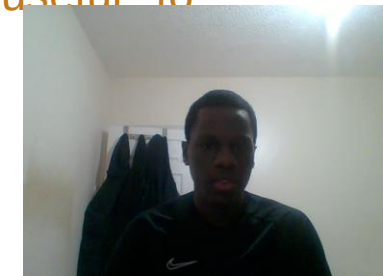
Real-time sentimental analysis of the tweet extracted was achieved using a web application. The application generated sentimental results for the last 24 hours in the form of plots and images every 2 hours.



Conclusion: Potential Future Development

Sentiment of Tweets with a Mixture of Languages

Pidgin English, Yoruba, Hausa, and Igbo are two of the most widely spoken indigenous languages within Nigeria. Many Nigerian Twitter users may choose to tweet in their ethnic language and important insights may be lost if non-English tweets are ignored. Thus, in the potential future development of this project, it will be useful to incorporate a mixture of languages.



References

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