

## Data Programming in Python Module

The proposed project is based on twitter sentiment analysis for the five major banks in South Africa. This analysis is conducted for a period of one month in December 2022. South African people have been complaining a lot lately about the service they are receiving from all major five banks in South Africa, particularly in the festive season period (December 2022).

Data was scraped using both Tweeter API and snsrape to overcome some limitation with tweepy API. All data set generated are included in the submission of this project for the five banks. There are five major conventional banks in South Africa, and the analysis will be conducted in all of them, namely: First national bank (FNB), Capitec bank, Nedbank, Standard bank and ABSA Bank. Conventional banks refer to the banks that still use the old and modern way of banking, this means the banks that still have ATMs, physical branches etc. There are other banks in South Africa that use the modern ways of banking like fully digital, no ATM or branches, only done through online applications.

The banks that will be analysed in this proposed project will be ABSA bank with a twitter handle name of @ABSASouthAfrica, Capitec bank with a twitter screen name of @CapitecBankSA, Nedbank with a twitter handle name of @Nedbank, FNB with home name of @FNBSA and Standard Bank with a twitter screen name of @StandardBankZA.

The following are the five banks being investigated in the project



Python was used to write the code in Jupiter notebook. Jupiter notebook was also used as a reporting tool for the project. Exploratory data analysis was performed and machine learning algorithms such as Logistical regression, Naïve Bayes and Random Forest were used to train the models.

## **How to run the application/model**

The application can be run by opening the submitted Jupiter notebook. The data is attached for ease of reference. The component for twitter API won't run because the keys will not be there, however it does not matter because it is there to demonstrate its limitation. Just proceed and run the following steps. But if you have keys, you can use your or open a developer account on twitter developer dashboard for free. The collection of data needs internet connection to scrape the data with snsrape, depending on your network and processing power of your computer, it will take few minutes to collect the data. It also depends on the amount of data to be collected.

There will be few pip install requirements for this to run on anyone's environment. Installation like word cloud, download stop words, Tweepy API etc. but the error will indicate which is a missing module. Some clear details are on the Jupiter notebook report, if followed well, one can recreate the project.