Interim Report of Twitter Data Analysis

Abstract

We examine the sentiment analysis of twitter data. In this project we are adressing the problem of sentiment analysis in twitter; where we are classifying the tweets that we have to a positive, negative or a neutral epxpression. This we are able to achieve based on the use of certain keywords.

Twitter is a growing online platform for social media where people express their thought on various topics. These expression are ussually in texts that are not more than 140 characters. This platform allows one to send and recieve short post that are called tweets.

The information that is desserminated in twitter are not always entitled to be read. Twitter services is rapidly expanding and it has a user of close to 200 million who are registered. out of this big numner, almost half of them are active users of twitter, who daily log into twitter and can generate upto 250 million tweets per day. Due to this big number, we are hoping to achieve a reflection of ehat the public says, based on the sentiments that the public has on different topics.

Introduction

Social media has a way in which the sentiments from the public can be useful in the society. It is the purpose of this data analysis, by the sentiment analysis that we are carrying out to identufy, by the use of the models that are going to be applied to make sense in the twitter data analysis. We are going to use the differrent methords that are used in the development of models in the carryong out of the model creation. Twitter data i allows oen to trach the brand and also to identify the customers that are able to cintribute to the development of a brand or a product. The focus of this analysis is not only to get the sentiment analysis of covid19 in Africa but as well to get the sentiment analysis of the different countries that are in Africa.

In the process of this analysis it is very important to set up the MLOps pipeline so as to identify the different steps up to the deployment stage of the data analysis.

Data

The data that we are going to use in this analysis is a data that is availed via a json file, but collected in a tweeter. The data was collected by the use of the APIs and thus, the process of the scrapping of the data and the preprocessing stage of the data was very useful as to ensure that the data we collected was the required data that was needed for the creation of the model. The data was as well provided in two forms, the first data was one that included the Africa alone, while the second data is the data that involves the individual countries.

How it is collected - it is from twitter

Key information about the data: format of data, number of rows, keywords used to collect the data, etc.

\* Method - the MLOps components you plan to implement

A flow chart demonstrating the different components in your MLOps design and their interaction

How you implemented the different components

Implementation details e.g. Travis CI, unit testing coverage, Streamlit for dashboard, etc.

Result and Discussion

The data that we have has been

The challenge you faced

During the analysis of this data, there are challanges that were faces in the course of the analysis.