Sentiment and Political Analysis in Twitter

Shokan V, Vishaal K J and Naveen SVE

Department of Analytics, SCOPE
Vellore Institute of Technology
Vellore-14,TN,India

v.shokan2018@vitstudent.ac.in,kj.vishaal2018@vitstudent.ac.in

Department of Software Systems, SCOPE
Vellore Institute of Technology
Vellore-14,TN,India

naveen.sve2018@vitstudent.ac.in

Abstract – **Sentiment Analysis is the study of whether the sentiments of the text are either positive, negative or neutral. Specifically the sentiment analysis from a major social media platform says a lot about a person or a group. As this is the time for US Presidential Elections, we are going to gather data from the major social media platform, Twitter and compare the sentiment analysis for the appropriate query and also we are going to find which candidate has a better chance and can win the election by plotting the sentiment analysis graph for them.**

I. INTRODUCTION

Sentiment Analysis is the method of finding whether a statement or opinion is positive, negative or neutral. So, it is also known as opinion mining, getting the opinion or attitude of a speaker.

In the political field, it is used to keep check of political views, to detect consistency and inconsistency between statements and actions at the government level. It can be used to predict election results too.

As the audience on the media platforms grows continuously, data from these sites can be used to analyze the sentiments of the people.

Social Media is used by many people nowadays. During this pandemic time, people are more hooked to their phones and internet all the time. Thus, the news, hashtags and trends are more reached than ever. The netizens are trending many things and thus this is the perfect time to analyze these data and trends. Social network analysis is the study of people's interactions and communications on different topics and nowadays it has received more attention. Millions of people give their opinion on different topics on a daily basis on social media like Facebook, Twitter, Instagram, etc. It has many applications in different areas of research from politics, sciences to business.

NLTK is an open source natural language processing (NLP) platform made ready for Python. It can do tasks like textual tokenization, parsing, classification, stemming, tagging, semantic reasoning and other computational linguistics. NLTK is a community driven project and can be used on Linux, Mac OS X and Windows.

There are many reasons why we chose twitter data for sentiment analysis. Some of the reasons are that twitter is used by people ranging from regular people to actors, politicians, businessmen, and various religious and social leaders to post their opinion. So, twitter reflects views of all groups of people.

Also this is the time for the US Presidential Elections, we have decided to put into use the sentiment analysis and predict which presidential candidate has a better chance to win based on their reception they have in Twitter. Twitter is a social media platform unlike others that only allows tweets of 280 characters long which makes it even easier to do the sentiment analysis. Not stopping with that we also included the opinions netizens have of the vice presidential candidates and done sentiment analysis for them too.

II. SENTIMENT ANALYSIS

1.1 Authentication from Twitter
    For collecting the data set from Twitter API , we need to register an app and generate consumer key and access token. Then, we can get the data directly from Twitter API itself

1.2 Data Gathering
        Collecting datasets involves the fetching of tweets and parsing them .When the object is called , the tweets are extracted excluding the retweets and parsed in lists and dictionaries.

1.3 Pre Processing:
In preprocessing we use the Textblob module. It is actually a high level library built over top of NLTK library. Then ,we create a Textblob object which performs the following process:

1.3.1 Data cleaning :
This step involves the cleaning tweets by removing the links and special characters using regular expressions. This is helpful to store and process the dataset effectively.

1.3.2 Tokenizing and removing stop words
Tokenize the tweet ,that is to split words from the body of text. Next, remove stop words from the tokens.(stop words are the commonly used words which are irrelevant in text analysis like I, am, you, are, etc.)
Lastly, do POS(part of speech) tagging of the tokens and select only significant features/tokens like adjectives, adverbs, etc.

1.4 Sentiment Analysis:
Sentiment analysis involves the classification of the tweets for the related query into good, bad and neutral.

1.4.1 Sentiment Classifier

We pass the tokens to a sentiment classifier which classifies the tweet sentiment as positive, negative or neutral by assigning it a polarity between -1.0 to 1.0

Sentiment classifier is created using textblob module TextBlob uses a dataset in which reviews have already been labelled as positive or negative.
Positive and negative features are taken from each positive and negative review respectively.
Training data then consists of labelled positive and negative features. This data is trained on a Naive Bayes Classifier .

1.4.2 Fixing polarity for the sentiment
We use sentiment. Polarity method of TextBlob class to get the polarity of tweet between -1 to 1 and classify the sentiment as positive or negative which is used for the analysis.

1.5 Plotting graph :
From the output ,simultaneously we use matplotlib to plot the sentiment graph with the graph denoting the number tweets which are either pos/neg/neutral which can be seen from the shape of the graph.

In recent years a lot of progress has been made in the field of Sentiment Analysis by number of researchers. In its early stage, it was initially made for binary classification which assigns opinions or reviews to bipolar classes such as positive or negative only. But later on, the study was further enhanced and all researchers and they had included neutral in the analysis. Earlier only sentences and phrases were used in the study but in the recent years the researchers are able to do sentiment analysis for each lexicon and most of them are able to include special characters, symbols and emoticons which require more access and better software in comparison to beginners and amateurs. With the usage of machine learning the field of sentiment analysis is no longer research but there are a lot of people using it and running as a business.
But there is a long way to go, even though sentiment analysis is able to predict recent fair enough, there are many drawbacks too like it is yet to make the analysis in other languages than English and the code by many people are not able to identify humor speech and criticisms for example. So let's hope in the coming years, people are able to overcome the obstacles and present a successful model of Sentiment analysis.

III Design and Implementation so far:

This team has begun the implementation of the Twitter sentiment analysis, by using the Twitter API. Tweepy is a library of Twitter API for getting the tweets directly from Twitter that are posted by different people. The streaming tweets are retrieved and then saved into CSV files to sentiment analysis. In this project, many libraries have been used.

The approach to extract sentiment from tweets is as follows: Firstly, Import Tweepy for creating the connection with Twitter API .Then, Fetch tweets as a dataset and then save into a CSV file. After that ,pre-processing of tweets by removing the stop words , punctuations, tags, etc. Later, Tokenize each word in the dataset and save it into the dataset. For each word, compare it with positive, negative and neutral sentiments words in the dictionary. Then increment the positive, negative and neutral counts. At last, based on the positive, negative and neutral count, we get the result percentage about sentiment to decide the polarity.

IV. ACKNOWLEDGMENT

V. PROPOSED WORK

The team has completed the collection of datasets after authentication from twitter API and the pre-processing and the work up to the count has been done as of now. The team has to plot the graph and calculate the result percentage to determine the polarity. The US elections are scheduled on November 5 and the election scenario is getting tougher day by day in the United States. So the coming weeks are the best time to make the prediction. We are planning to update our datasets and make it fail-proof too.

VI .REFERENCES

[1] David Osimo and Francesco Mureddu,Research Challenge on Opinion Mining and Sentiment Analysis, 2012
[2] Shubham Goyal,Sentiment Analysis of Twitter Data Using Text Mining and Hybrid Classification Approach, 2017
[3] Aisopos and Fotis ,Sentiment analysis of social media content using graphs, 2012
[4] L. Barbosa, J. Feng. "Robust Sentiment Detection on Twitter from Biased and Noisy Data". COLING 2010:
[5] Priyanka Thakur, Sentiment Analysis of Tourist Review using Supervised Long Short Term Memory Deep Learning Approach, 2019
[6] Vishal Kharde, SS Sonwane. Sentiment Analysis of Twitter Data: A Survey of Techniques
[7] Shravan IV. Analysing Sentiments with NLTK
[8] Chhinder Kaur ,Anand. Twitter Sentiment Analysis on Coronavirus using Textblob
[9] Lokesh Mandoli, Ruchi Patel. Twitter Sentiments Analysis Using Machine Learninig Methods