TWEET CLASSIFICATION AND TREND DETECTION USING NLP

Naman Veeramacheneni
VIT – AP, Amaravati

Abstract

Social media often plays a crucial role in disseminating information to warn the public about health concerns. Twitter is the most popular social media that allows its users to spread and share information. They publish these topics on the list called “Trending Topics”. It shows what is happening in the world and what people's opinions are about it. In this project, proposes a plan to develop a novel framework for topic sentiment trend detection and prediction in social media

Introduction

In social media, millions of active users express their opinions and interact with each other daily. Such users' content in the form of posts or tweets provides a vast amount of useful information if analyzed carefully. Therefore, the data streamed from social media such as Twitter, Facebook, or Instagram is so precious for researchers to perceive the users' social behavior through NLP.

A massive amount of user-generated online content is freely available to the real-time monitoring of public sentiment. It is difficult to find the contextual sentiment of a text. Sentiment analysis is one of the critical issues today. The primary job is to fast-pace the process of opinion extraction from the given subject. The subject here can be an excerpt from the written text, debate, or day to day conversation.

In sentiment analysis, we also evaluate the positive and negative intensities of symbols and words. Sentiment analysis helps to improve customer services, Political planning Policies, and manufacturing quality products.

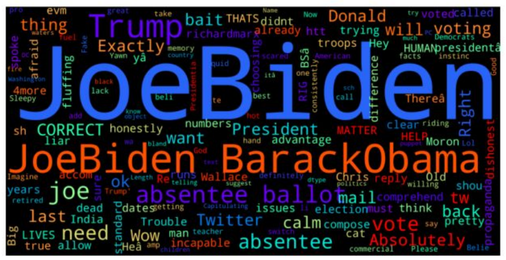
Methodology

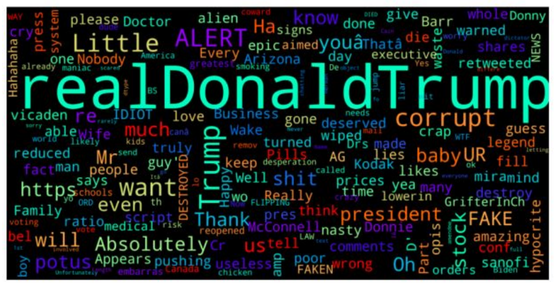
The basic methodology of the project: •Collecting high volumes of data to create an efficient prediction model. •Pre-processing data and cleansing of data. •Detecting topic trends and applying sentimental analysis •Displaying result interpretations and prediction.

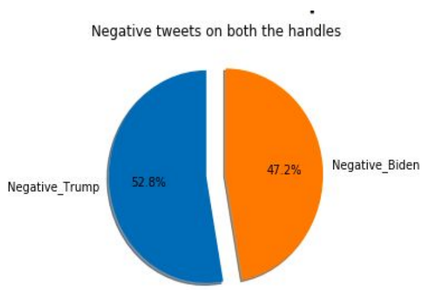
A huge amount of data is required to get accurate results. To create the dataset we would first need to scrape public tweets from home pages of both presidential candidates. Tweepy tool allows us to get tweets and re-tweets and also connects us to the twitter API. One would require a twitter developer account to get access to the API by using authentication keys. By giving account handle and account ID, we would be able to store all the tweets in a CSV file

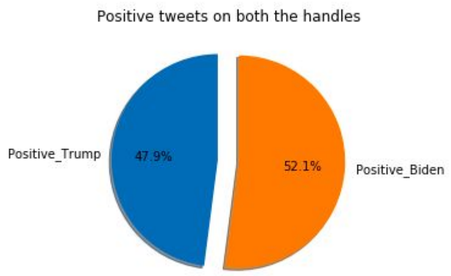
The next task is to analyze data and clean up the text which isn’t returning any meanings and apply our algorithm for classifying text into either positive sentiments or negative sentiments. The dataset contains two attributes in total, and only the replies column is for consideration, the other one wouldn’t add any value to the sentiment analysis. A correlation between different attributes is necessary to choose the most important ones which is also known as feature selection, a widely used technique for dimensionality reduction.

Results









From the pie-chart representation, it can be concluded that Joe Biden will win the 2020 US presidential elections with 52-53% votes.

Project Tools

Anaconda Navigator — Virtual Environment
- Python — a programming language
- Tweepy — a type of RESTful API specifically for Twitter
- Textblob — processed textual data library tool (already trained on numerous textual data.)
- Pandas — data manipulation and analysis library
- NumPy — scientific computing library
- Matplotlib — plotting library
- Plotly — plotting library
- Seaborn — Data visualization library based on Matplotlib
- Wordcloud — library for a visual representation of textual data

Conclusions

The data collected from twitter is very important to an organization using this model as it can determine the significant insights and provide results based on twitter data.

In addition to election results, this model could be used to get public feedback on consumer products which could give companies and organization a detailed feedback of their product performance in the global market.

Graphical and textual representation of data could be easily understood by the end users of this model.

Real-time applications

This analysis model can be used to:
- Categorize any product feedback (ex: medicine, electronics etc)
- Compare product competitors based on reviews.
- Election results based on public opinion on social media.
- Analysis of Restaurant/ Movie reviews.

Contact Information

Naman.V

VIT-AP
Email:
naman.veeramacheneni@vitap.ac.in
Phone: +917013524626

References

- https://medium.com/analytics-vidhya/tweet-analytics-using-nlp-f83b9f 7f7349
- https://ieeexplore.ieee.org/abstract/document/7219856/
- https://ieeexplore.ieee.org/abstract/document/6113240/
- https://ieeexplore.ieee.org/abstract/document/8397589/
- https://www.aclweb.org/anthology/S16-1040.pdf
- https://towardsdatascience.com/simple-twitter-analytics-withtwitter-nlp-toolkit-7d7d79bf2535
- Topic Sentiment Trend Detection and Prediction for Social Media by Aashish Thota
- https://www.youtube.com/watch?v=gUFDtuz73gI

Acknowledgements

I would like to thank Dr. Karthikeyan Saminathan for his valuable guidance and immense support throughout the project work as my project guide and advisor. Data analytics is growing very fast, Dr. Karthikeyan always keeps herself up to date with the latest research and encourages his students to work towards cut edge technologies. I am amazed by his positive energy and patience. His unparalleled knowledge, agile and prompt feedback coupled with smart ideas have helped me immensely in putting up the whole work.