|  |  |
| --- | --- |
| **Sentiment Analysis using OrangE**  Analysing Donald Trump’s Tweets | Abstract  This project uses Orange Data Mining to carry out sentiment analysis on the ‘Trump Tweets’ dataset from Kaggle.  Mrudav Mehta  BDS 21 |

INTRODUCTION

Sentiment analysis is a method for identifying the sentiment or emotional tone conveyed in textual data. It comprises reading the content to identify and categorise opinions, attitudes, and sentiments as neutral, unfavourable, or favourable. This technique is highly useful for understanding public opinion, social media sentiment, and consumer feedback, among other applications.

Orange is a software suite for data visualisation, machine learning, data mining, and data analysis, it is built around the elements of visual programming. Therefore, even though it is a very simple drag-and-drop tool which can be used by anyone, its uses are vast, and it is a very important tool in the realm of machine learning. Orange runs on the cross-platform Qt framework and uses the graphical user interface (GUI) of well-known Python open-source scientific computing packages, such as NumPy, SciPy, and Scikit-learn.

The primary objective of this project was to carry out a sentiment analysis on the ‘Trump Tweets’ dataset on Kaggle. To accomplish this, the text data needs to be examined in order to determine which main viewpoints were expressed in the tweets and categorise them as neutral, favourable, or negative. In addition, the programme will look into sentiment patterns over time or in response to certain events, providing a window into people's thoughts and perspectives on Donald Trump as they are expressed on Twitter. In addition, the effectiveness of using Orange Data Miner for sentiment analysis will be evaluated to determine whether or not it is suitable for these kinds of tasks.

We chose this topic because the Trump Tweets dataset is crucial in understanding public opinion on Donald Trump's words, actions, and policies due to his significant use of Twitter as a communication tool during his presidency. Analysing the mood expressed in these tweets can provide insights into the public's perceptions of Trump's policies, persona, and presidency, as examining the tone can yield insightful information.

Sentiment analysis of tweets can be beneficial for politicians for several strategic reasons. It allows them to modify their approach by monitoring public sentiment, identify important issues and sentiments, and provide customized messaging during campaigns. It also helps politicians avoid damage to their reputation by identifying and resolving negative sentiment. Sentiment analysis also helps lawmakers understand constituent issues, promoting effective engagement and responsiveness. Overall, politicians can use this information to optimize their strategies, engage constituents, and make informed decisions.

Data Description

As mentioned above, we went forward with the ‘Trump Tweets’ dataset that was obtained from Kaggle. The initial plan was to extract Tweets using an API pull directly from Twitter and then perform a sentiment analysis on that data, however, Orange does not have the facility of having a built-in data fetcher from Twitter.

After the recent shift in leadership at Twitter, there were some policy changes made, that specified a basic account will be charged USD100 for retrieval of Tweets. Therefore, we went forward with simply downloading the dataset off of Kaggle and working with it.

The dataset has 9 features or columns – id, link, content, date, retweets, favourites, mentions, hashtags, geo.

The dataset consists of 41,122 Tweets or rows, and the total size of the entire CSV file we are dealing with is 9.98 megabytes.

The description of each of the columns are as follows:

ID is the unique identifier of each Tweet, Link is the link to each Tweet, Content shows what each Tweet actually was, Date is the date on which the Tweet was posted, Retweets and Favourites are numerical fields that show the total number of retweets and favourites of that particular Tweet. Hashtags specifies which hashtags were used in that Tweet, and Geo shows the location from which the Tweet was made (if specified).

Sentiment Analysis

Method used for sentiment analysis (Liu Hu, Vader, etc)

Moving on to the main segment of the project, carrying out sentiment analysis on the Trump Tweets using Orange Data Mining. The sentiment modules from NLTK, Liu & Hu, multilingual sentiment lexicons from the Data Science Lab, LiLaH sentiment from Walter Daelemans et al., and SentiArt from Arthur Jacobs can all be usedin Orange. All of them rely on lexicons. Liu & Hu is available in both Slovenian and English. Vader can only speak in English. The list of compatible languages for Multilingual Sentiment can be found at the bottom of this page. SentiArt is compatible with both English and German. LiLaH sentiment supports Slovenian, Croatian, and Dutch. With the custom dictionary option, one can upload their own dictionaries of positive and negative sentiment.

Since all of Donald Trump’s Tweets are in English (in this dataset), we have chosen the Vader method as it can only comprehend English and would be best suited for our scenario.

A Vader sentiment analysis would do the following – tokenization, case normalization, punctuation handling, stop-word removal, sentiment scoring and compound scoring.

We then build a pipeline in Orange that looks as follows:

A diagram of a diagram

Description automatically generated

The data is imported, we select the columns that are relevant. We then have to choose the method we want to go forward with, which is Vader (Valence Aware Dictionary for Sentiment Reasoning) in our case. Vader understands the emotions behind sentences using the words built to make that sentence. It is also smart enough to comprehend that a phrase like ‘no enjoyment’ is a negative phrase, not positive due to the word ‘enjoyment’. Vader returns the score in 4 categories – positive, negative, neutral, or compound (normalised score of positive, negative, and neutral)

Once all Tweets have been analysed, we can select a sample of data to summarize our results. The results are available to us in the form of a heatmap, and we also have the option to see each Tweet’s compound score.

A colorful rectangular object with black lines

Description automatically generated A close-up of a website

Description automatically generated

The heatmap, and an example of a tweet along with its compound score.

We can see that most of the tweets are neutral/negative. But how did this help Donald Trump and what has been the learning out of this?

What we have performed is a very small-scale method of how sentiment analysis can prove to be useful to just take out the emotion behind the Tweet, however, Donald Trump has infamously taken help from Cambridge Analytica to win the 2016 US Presidential Elections. These are the following ways data mining and analysis were used by Cambridge Analytica which proved to be a successful venture for Donald Trump:

1. By data mining and psychographic profiling, targeted ads were deployed towards demographics and regions where Trump had his supporters. This ensured that the Trump campaign reached the right people in the right magnitude.
2. Cambridge Analytica also identified voters who were unbiased towards any party and were yet to take a decision. The Trump campaign then shifted resources towards these groups of people and used all their efforts in convincing them to vote for him.

Overall, Cambridge Analytica's data-driven approach allowed the Trump campaign to maximise its messaging, interaction, and advertising strategies, which also contributed to Trump's victory in the 2016 presidential election.

This shows the power that data holds and how it can be used to change the course of history

References

<https://orangedatamining.com/widget-catalog/text-mining/sentimentanalysis/>

<https://www.kaggle.com/datasets/austinreese/trump-tweets/data>

<https://towardsdatascience.com/sentimental-analysis-using-vader-a3415fef7664>

<https://www.kaggle.com/code/shailaja4247/sentiment-analysis-of-tweets-wordclouds-textblob>