2/20/2023

Twitter Sentiment Analysis

CIND820-XJH

Literature Review, Data Description, and Project Approach

Parth Dhananjay Desai
parth.desai@torontomu.ca
Student ID: 501145348
Prof: Zekiye Erdem (zekiye.erdem@ryerson.ca)

Theme & Introduction

Twitter Sentiment Analysis is a type of machine learning that aims to determine the overall sentiment of tweets or hashtags. Through natural language processing and machine learning techniques, we can extract subjective information from a dataset and classify it based on its polarity (positive, neutral, or negative). This analysis is valuable as it allows us to understand the public perception of a product or predict stock prices based on sentiment. However, sentiment analysis is a challenging problem due to the complexity of language and grammar.

For my project, I am using a probabilistic model to classify tweets on Twitter as having either a positive or negative sentiment. Twitter is a microblogging platform where users can quickly and spontaneously express their emotions through 140-character tweets. Users can also use the @ symbol and hashtags to join discussions or contact other users directly. Due to its popularity, Twitter is a valuable source of information for gauging public opinion on any topic.

Problem Definition

Sentiment analysis in the realm of micro-blogging is a relatively new field with immense potential for further research. Previous studies have focused on sentiment analysis in areas like user reviews, documents, web blogs, and general phrase-level sentiment analysis. However, sentiment analysis of tweets presents a unique challenge due to the 140-character limit that compels users to express their opinions succinctly.

Supervised learning techniques like Naive Bayes and Support Vector Machines have been the most effective sentiment classification methods to date, but they require expensive manual labelling. Some researchers have also explored unsupervised and semi-supervised approaches, but there is still room for improvement.

Researchers often compare their results to baseline performance, but there is a need for formal comparisons between different methods to determine the best features and most efficient classification techniques for specific applications. Further research is necessary to advance sentiment analysis of tweets and realize its full potential.

Introduction

What do you already know about the topic?

A Twitter sentiment analysis is also approached as Opinion mining is a process that helps to identify as well as divide the sentiment that is expressed on Twitter. It is a process of deciding the emotional tone of the words that have been used on Twitter. This is done through understanding the polarity of tweets as positive negative or neutral.

The analysis of this is generally accomplished using NLP (Natural Language Processing) and machine learning algorithm, which can be implied in various arena or different sectors for social media monitoring to get the best out of it can be useful in areas such as business intelligence, customer service, sport branding, as well as political science. If this data is correctly diagnosed, it can be used by many brands to figure out the aspects they lack and work on it.

What do you have to say critically about what is already known?

Based on my observations and understanding, I can confidently state that one of the most persistent challenges in sentiment analysis of tweets is accurately categorizing the different forms of emotions present.

This can often be confusing and misleading, making it difficult to discern the emotional tone and identify the true emotions and intended meaning behind a tweet, including the use of figurative language.

Mistakes in classification can lead to erroneous data, resulting in a larger failure. Additionally, emotions can vary based on context, making it challenging to create a comprehensive model for Twitter Sentiment Analysis.

These limitations need to be addressed to improve the accuracy and reliability of sentiment analysis in micro-blogging.

Has anyone else ever done anything the same?

It is highly improbable for an individual to obtain the same outcomes while scrutinizing the enormous quantity of tweets and hashtags, given the extensive amount of data accessible.

Nevertheless, numerous methods of Twitter Sentiment Analysis have investigated diverse categories of data with varying attributes and assessment techniques.

A multitude of research has been conducted regarding Twitter Sentiment Analysis, including the scrutiny of emotional support networks and the prediction of stock market trends through Sentiment Analysis.

Has anyone else done anything that is related?

It is highly unlikely to have completely identical findings since there are millions of tweets and hashtags being generated and analyzed, leading to a plethora of data.

However, multiple approaches to Twitter Sentiment Analysis may have explored various types of data, including different attributes and evaluation metrics.

It's possible that many studies have been conducted on Twitter Sentiment Analysis, which may include analyzing emotional support networks and predicting stock market trends using sentiment analysis.

Where does your work fit in with what has gone before?

Approaching the topic “Twitter sentiment analysis” our first task is to by integrate profile information, including location, age, gender, emotion, comments into the analysis model. This approach could help overcome some of the challenges faced in traditional Twitter sentiment analysis techniques, such as sarcasm and contextual variations in sentiment. To provide best example, I am

I am conducting Twitter sentiment analysis on the topic of Brexit, my work would fit into the larger body of research that has explored the public attitudes and sentiments towards Brexit and how country has taken this big step in positive, negative or in neutral way.

My research has examined the impact of Brexit on the UK economy, politics, and society, as well as the attitudes of different stakeholders such as politicians, business leaders, and citizens. Furter continuing the project I will on data related to Brexit and providing insights into the attitudes, opinions, and emotions expressed by Twitter users regarding the topic.

This could contribute to a better understanding of the public discourse surrounding Brexit and help to identify the most prevalent sentiments expressed on social media. Work will show use advanced machine learning techniques to improve the accuracy and efficiency of sentiment classification.

Furthermore, my analysis could help to identify potential gaps in previous research and suggest new directions for future investigations.

Why is your research worth doing in the light of what has already been done? ‘

Conducting Twitter sentiment analysis on Brexit my approach will be the following.

Firstly, my study will include vast research on topic which will consist of articles, blogs, trending hashtags, media resources, Newspaper etc. To get the better understanding of public review, attitudes, emotions, and sentiments.

Secondly, my approach will be to learn advanced machine learning techniques which will help me improve sentiment analysis accuracy and efficiency which I result yield more reliable results and nuanced insights into Twitter users' attitudes and opinions about Brexit.

Thirdly, my research will identify any potential gaps in previous research to bring up unseen aspects and facts about the topic which is being missed in previous research.

Finally, I will be being my best imply decision making and public discourse on Brexit. My work might help in future to improvise in decision making, raising local voice and improve communication in public.

Data Description & Methodology

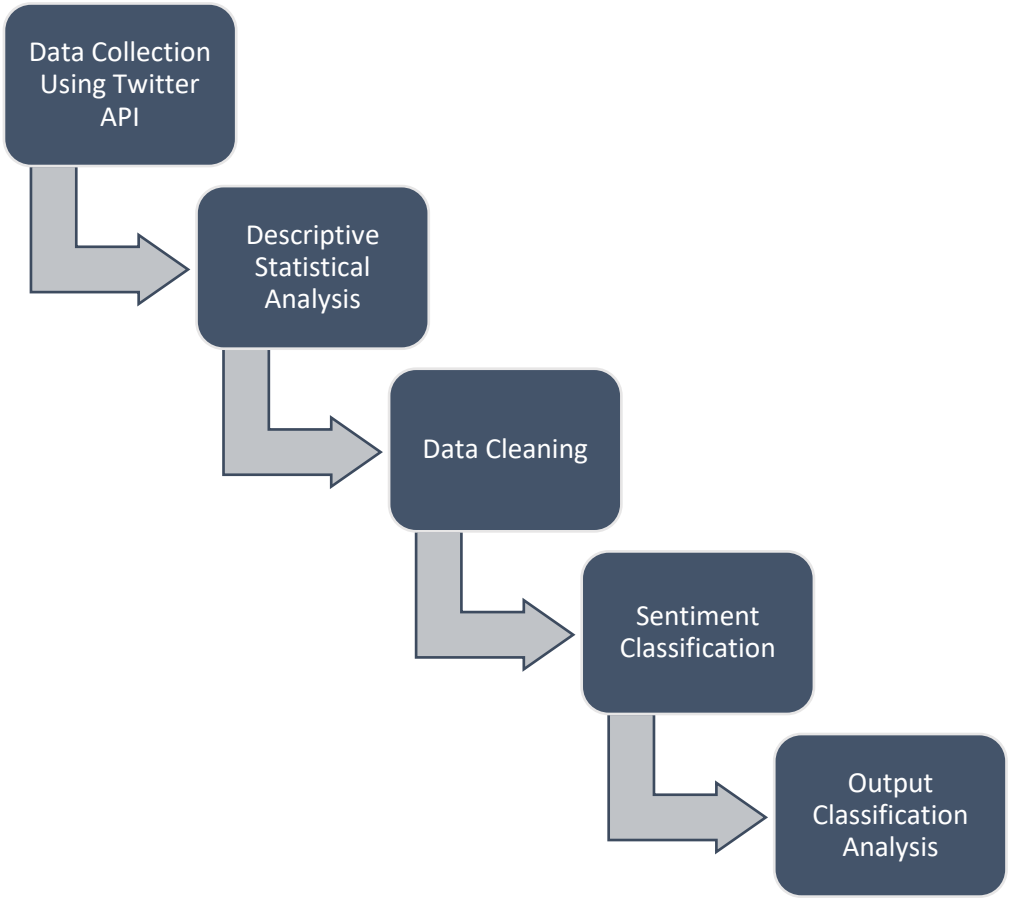
Brexit is a highly complex topic that has significant political, economic, and social implications. Sentiment analysis of Twitter data related to Brexit aims to determine the attitudes, opinions, and emotions expressed by users regarding the topic. To carry out such an analysis, a dataset is compiled that includes many tweets related to the Brexit. These tweets are collected using hashtag #Brexit.

The dataset would include relevant metadata such as user IDs, timestamps, and other information. The text data in the dataset would consist of tweets that express positive, negative, or neutral sentiments towards Brexit. To make the data suitable for analysis, pre-processing techniques such as cleaning, filtering, and tokenization would be applied to remove irrelevant or noisy data.

To classify the sentiments expressed in the tweets, various machine learning algorithms are available. In this project, I will be using SentimentIntensityAnalyzer from NLTK library.

The output of the sentiment analysis would include the percentage of positive, negative, and neutral tweets related to Brexit, as well as the most common emotions expressed in the tweets. The results of the analysis would be presented visually using charts, graphs, or other forms of data visualization, making it easier to interpret and understand the sentiment expressed on Twitter about Brexit.

In summary, sentiment analysis of Twitter data related to Brexit involves collecting many tweets, pre-processing the data, and analysing the sentiments expressed in the tweets using natural language processing techniques. The output provides valuable insights into the attitudes, opinions, and emotions of Twitter users towards Brexit, which can be used to inform decision-making and public discourse.



Brief Descriptive Statistic

Total number of Tweets	3000
Average Word Count per Tweet	20.47
Average # count	1.17
Average @ count	1.16
Average emoji count	0.10

%Tweets having @	89.97%
%Tweets having emojis	7.10%

References:

Vishal A. Kharde (2016, April 11). *Sentiment analysis of twitter data: A survey of techniques.* International Journal of computer Applications(0975-8887). https://arxiv.org/ftp/arxiv/papers/1601/1601.06971.pdf

Faizan (2019, February 2). *Twitter Sentiment Analysis.* International journal of innovative science and research technology. https://ijisrt.com/wp-content/uploads/2019/02/IJISRT19FB242.pdf

Adwan, Omar., et all (2020, August). *Twitter Sentiment analysis Approaches.* International journal of emerging technologies in learning. https://www.researchgate.net/publication/343666530_Twitter_Sentiment_Analysis_Approaches_A_Survey

Agarwal, Apoorv, et all. *Sentiment analysis of twitter data.* Department of computer science, Columbia University. http://www.cs.columbia.edu/~julia/papers/Agarwaletal11.pdf

Sharma, Ankita. Ghose, Udayan (2020). *Sentimental Analysis of Twitter data with respect to general elections in India.* Procedia Computer Science. https://www.sciencedirect.com/science/article/pii/S1877050920315428

Alsaeedi, Abdullah, et all. (2019). A study on sentiment analysis techniques of twitter data. International Journal of advance computer science and applications. https://thesai.org/Downloads/Volume10No2/Paper_48-A_Study_on_Sentiment_Analysis_Techniques.pdf

Yaswanth, Vanama, et all (2022, May 31). Sentiments analysis using tweets. https://assets.researchsquare.com/files/rs-1706745/v1_covered.pdf?c=1654018429

Schwartz, Cassilde, et att (2020, May 04). A Populist paradox? How Brexit softened Anti-Immigrant Attitudes. Published online by Cambridge university Press. https://www.cambridge.org/core/journals/british-journal-of-political-science/article/abs/populist-paradox-how-brexit-softened-antiimmigrant-attitudes/BAEFB80FD773AF203C36B71D7518A5B7#

Hobolt, Sara, et all (2020, July 07). Divided by the vote: Affective Polarization in the wake of the Brexit referendum. Published online by Cambridge university Press. https://www.cambridge.org/core/journals/british-journal-of-political-science/article/abs/divided-by-the-vote-affective-polarization-in-the-wake-of-the-brexit-referendum/2393143858C3FA161AF795269A65B900

Delis, Agelos, et all (2018, July 04) Electoral Spillovers in an Intertwined world: Brexit effects on the 2016 Spanish vote. Published online by Cambridge university Press. https://www.cambridge.org/core/journals/british-journal-of-political-science/article/abs/electoral-spillovers-in-an-intertwined-world-brexit-effects-on-the-2016-spanish-vote/126DDE685EFFF565543CABA908C7B6CE

Goodwin, Matthew, et all (2018, February 05). For and against Brexit: A survey experiment of the impact of campaign effects on public attitudes toward EU membership. Published online by Cambridge university Press. https://www.cambridge.org/core/journals/british-journal-of-political-science/article/for-and-against-brexit-a-survey-experiment-of-the-impact-of-

campaign-effects-on-public-attitudes-toward-eu-membership/83A412AC102A3E28389F9FD66DF84AFE

Green, Jane, et all (2021, February 16). Who gets what: The Economy, relative Gains and Brexit. Published online by Cambridge university Press. https://www.cambridge.org/core/journals/british-journal-of-political-science/article/who-gets-what-the-economy-relative-gains-and-brexit/092683A0C4CE85FD53FAFAEA6B7207E5

Murphy, Justin. Devine, Daniel (2018, July 31). Does media Coverage Drive public support for UKIP or does public support for UKIO drive media coverage? Published online by Cambridge university Press. https://www.cambridge.org/core/journals/british-journal-of-political-science/article/does-media-coverage-drive-public-support-for-ukip-or-does-public-support-for-ukip-drive-media-coverage/81B77DDCA9B0DE26A8DF18B15158EF16

Bove, Vincenzo, et all (2021, April 21). Did Terrorism affect voting in the Brexit referendum? Published online by Cambridge university Press. https://www.cambridge.org/core/journals/british-journal-of-political-science/article/did-terrorism-affect-voting-in-the-brexit-referendum/132666AB0B1163DBD9E055E95A0C0938