



REAL TIME TWITTER SENTIMENTAL ANALYSIS

Suraj Kumar*1, Sonu Kumar*2, Shivam Bharatiya*3, Shanvi Gautam*4, Prof. Mr. Shakun Garg *5

* 1 Student, Department of Information Technology, IIMT College Of Engineering, GreaterNoida, Uttar Pradesh India.

*2 Student, Department of Information Technology, IIMT College of Engineering Greater NoidaUttar Pradesh India

*3 Student, Department of Information Technology, IIMT College of Engineering, Greater Noida,Uttar Pradesh, India

*4 Student, Department of Information Technology, IIMT College of Engineering, GreaterNoida, Uttar Pradesh, India

*5HOD, Department Of Information Technology, IIMT College of Engineering, Greater Noida , UttarPradesh , India

1.ABSTRACT

The goal of tweet sentiment analysis is to find the positive, negative, or neutral sentiment part in the tweeter data. Sentiment analysis can help any organization to find people's opinions of their company and products. We have applied sentiment analysis on twitter dataset. Our model takes input tweet, sentiment, and output selected text starting and ending in input tweet. We are using the BERT (Bidirectional Encoder Representations from Transformers) model with one more layer to find sentiment positions in tweets. The extra layer is used to find the sentiment part of a tweet. Our model is able to achieve 70 percent accuracy onthe validation dataset.

Keywords: deep learning, natural language processing, sentimental analysis, supervised learning.

2.INTRODUCTION

With an exponential rise in social media usage to share emotions, thoughts and opinion, Twitter has become the goldmine to analyze brand performance. Opinions found on Twitter are casual, honest and informative than what can be picked from formal surveys etc. Millions of users express their sentiment about brands they interact with. These sentiments, if identified, can be useful for companies to not only monitor their brand's performance but also locate aspects and time periods that receive polar sentiments. These brands can be products, celebrities, events or political parties. Therefore, teams were dedicated to analyze candidate's performance during national polls¹, or to analyze public response during the launch of a gadget or a movie.

However, with more than 500M tweets sent per day², this data has already become huge enough to be analyzed by any team manually. Likewise, the diversity of tweets presumably cannot be captured by fixed set of rules designed by hand. It is worth noting that the task of understanding the sentiment in a tweet is more complex that of any well formatted document. Tweets do not follow any formal language structure, nor they contain words from formal language (i.e., out-of-vocabulary words). Often, punctuations and symbols are used to express emotions (smileys, emoticons etc).

In this work, we use machine learning approach and natural language processing techniques to understand the patterns and characteristics of tweets and predict the sentiment (if any) they carry. Specifically, we build a computational model that can classify a given tweet as either positive, negative or neutral based on the sentiments it reflects. A positive and negative class would contain polar tweets expressing a sentiment. However, a neutral class may contain an objective or subjective tweet - either a user reflect neutrality in an opinion or contain no opinion at all. Examples of each class can be found in Table 1. The decision to use three classes is made to accommodate the complexity of the problem and is consistent with ongoing research in the field. The experiments conducted on our sentiment predictor informally shows that our system is amongst the best performing systems for this task.

Using our sentiment predictor, we also build an interactive visualization tool to help businesses interpret and visualize public sentiments for their product and brands. This tool enables the user to not only visualize plain sentiment distribution over the entire dataset, but also equips user to conduct sentiment analysis over the dimension of time, location, and influencing power of a user.

Twitter data can help

- I. Companies to identify weak points and how to make better brands
 - II. Twitter is a direct link between the organization and the client.
 - III. find customer sentiment on products
- a. 0 – Negative
 - b. 1 -Neutral
 - c. 2 – Positive

3. LITERATURE SURVEY

Sentiment Analysis is the automated process of analyzing text data and sorting it into sentiments positive, negative, or neutral. Performing Sentiment Analysis on data from Twitter using machine learning can help companies understand how people are talking about their brand. Sentiment Analysis or Opinion Mining is the computational treatment of opinions, sentiments, and subjectivity of text. One of the original works on sentiment analysis is done on this movie review dataset. This was done by Bo Pang and Peter Turnkey using unsupervised learning to cluster the words, and they create word vector. After that, they applied different classifiers to find the polarity of the reviews. Then they used word vector to find the sentiment of the review.

One of the early sentimental analyses on Twitter is done by Alex Goin 2009. They did Tokenization of all unique words in the dataset, used as classifier vocabulary. Alec Go used distant learning for sentiment classification and SVM outperforms (they use multiple classifiers). later, SVM and KNN were used by Mohammad Rezwana Huq, Ahmad Ali (2017) on Twitter Data to find Sentiment Analysis. Sara Rosenthal, Noura Farra, Preslav Nako work on Tweet Quantification using SVM. The A. Pak and P. Paraben used CRF, SVM, to find sentence-level sentiment then work on different techniques to find the sentiment of the document. They find differences in distributions among positive, negative, and neutral using Tree-Tagger for part of speech tagging. The beginning era of deep learning NLP networks started After Word vector representations were proposed by Tomas Mikolov. They used a dataset with more than 1.6 words. Neural networks train word representation from a dataset to get word vector representation. Tomas Mikolov used a billion words to train neural-network to find the word vector for words, which improves NLP task accuracy. The Continuous-Bag-of-Words model takes input as surrounding words and tries to predict current words as output. Skim-gram model takes input as a current word and tries to output the surrounding. The era of deep learning NLP model input parallelization started with “Attention Is All You Need” paper which helps to parallelize the processing of input, uses attention to boost the models. The transformer changes sequence to sequence processing to parallelize as a convolution neural network.

4. MOTIVATION FOR WORK

Tweets sometimes express opinions about different topics. These opinions are important in many business- related decisions and even political sentiments about a candidate.

Consumers can use sentiment analysis to research products or services before making a purchase.E.g., Kindle

Marketers can use this to research public opinion of their company and products, or to analyze customer satisfaction. E.g., Election Polls

Organizations can also use this to gather critical feedback about problems in newly released products. E.g., Brand Management (Nike, Adidas).

5. SIGNIFICANCE OF STUDY/ FUTURE SCOPE

Twitter sentiment analysis allows you to keep track of what's being said about your product or service on social media, and can help you detect angry customers or negative mentions before they escalate.

Some of the major significance of the study of twitter sentimental analysis are as follows;

Social Media Monitoring

Twitter sentiment analysis can provide valuable insights that drive decisions. What do customers love about your brand? What aspects get the most negative mentions? This tweet, for example, indicates that fast shipping is one of the most valued aspects for this Amazon customer

Customer Service

It has become increasingly important for customer service agents to be present on Twitter. They need to engage with customers and respond quickly to customer queries : six out of ten users expect a brand to respond to any customer service requests within one hour.

But how can you detect which support queries are most urgent? Twitter sentiment analysis allows you to track and analyse all the interactions between your brand and your customers, so you can make sure you respond to the most critical issues first.

Market Research

Twitter is a major source of consumer insight. In fact, people use it to express all sorts of feelings, observations, beliefs, and opinions about a variety of topics.

You can use Twitter sentiment analysis to track specific keywords and topics to detect customer trends and interests. Understanding what customers like, what their behaviours are, and how this changes over time is essential if you are planning to launch a new product.

Brand Monitoring

Whether you are launching a new feature on your platform, a site redesign, or a new marketing campaign, you may want to track customer reactions on Twitter.

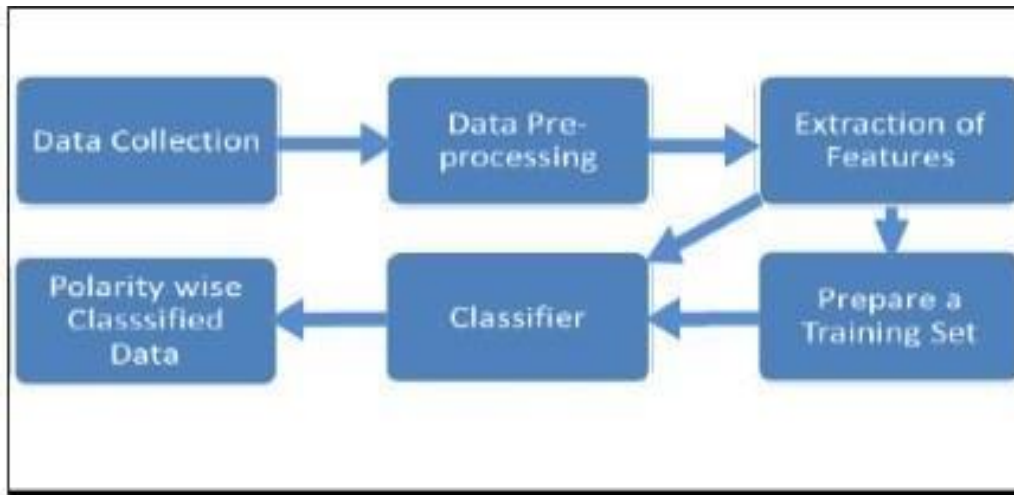
Taking action and making changes or improvements in real-time will help maintain customer loyalty, and can even help spread the word about your new feature.

Political Campaigns

A huge part of Twitter conversation revolves around news and politics. That makes it an excellent place to measure public opinion, especially during election campaigns. Twitter Sentiment Analysis can provide interesting insights on how people feel about a specific candidate (and you could even track sentiment over time to see how it evolves).

6. RESEARCH METHODOLOGY

The sentiment analysis of Twitter data is an emerging field that needs much more attention. Fig. 1 shows the steps to carry out the process of sentiment analysis on Twitter data. Firstly, the collected Twitter data is pre- processed to perform the data cleaning. Secondly, the important features are extracted from the clean text, applying any of the feature selection methods. Thirdly, the portion of the data is manually labelled as positive or negative Tweets to prepare a training set. Finally, the extracted features and the labelled training set are provided as an input to the built classifier to classify the remaining data i.e. test set. Each of the processing steps is discussed thoroughly in the following sub-sections.



A. Data Sources:

Selection of data source to conduct the sentiment analysis plays a significant role. Social media platforms as the data sources are broadly categorized into three general categories: blogs, micro-blogging sites, and review site . Among all categories, a micro-blogging site such as Twitter has gained higher popularity due to its limited strength of the content and publicly availability of data. From the following statistics of the Twitter growth rate, it's evident to use Twitter as the data source for sentiment analysis.

B. Twitter Data Collection Methods

The three possible ways to collect Tweets for research are as follows:

Data repositories such as UCI, Friendster, Kdnuggets, and SNAP

APIs: Twitter provides two types of APIs such as search API and stream API. Search API is used to collect Twitter data on the basis of hashtags and stream API is used to stream real time data from Twitter

Automated tools that are further classified into premium tools such as Radian6 , Sysmos, Simplify, Lithium and non-premium tools such as Keyhole, Topsy, Tagboard and Social Mention

C Data Preprocessing

Mining of Twitter data is a challenging task. The collected data is raw data. In order to apply classifier, it is essential to pre-process or clean the raw data. The pre-processing task involves uniform casing, removal of hashtags and other Twitter notations (@, RT), emoticons, URLs, stop words, decompression of slang words and compression of elongated words. The following steps show the pre-processing procedure.

Remove the Twitter notations such as hashtags (#), retweets (RT), and account Id (@).

Remove the URLs, hyperlinks and emoticon. It is necessary to remove non letter data and symbols as we are dealing with only text data.

Remove the stop words such as are, is, am etc. The stop words do not emphasize on any emotions, it is intended to remove them to compress the dataset.

Compress the elongated words such as happyyy into happy.

Decompress the slag words such as g8, f9. Generally slang words are adjectives or nouns and they contain the extreme level of sentiments. So it is necessary to decompress them.

D. Feature Extraction

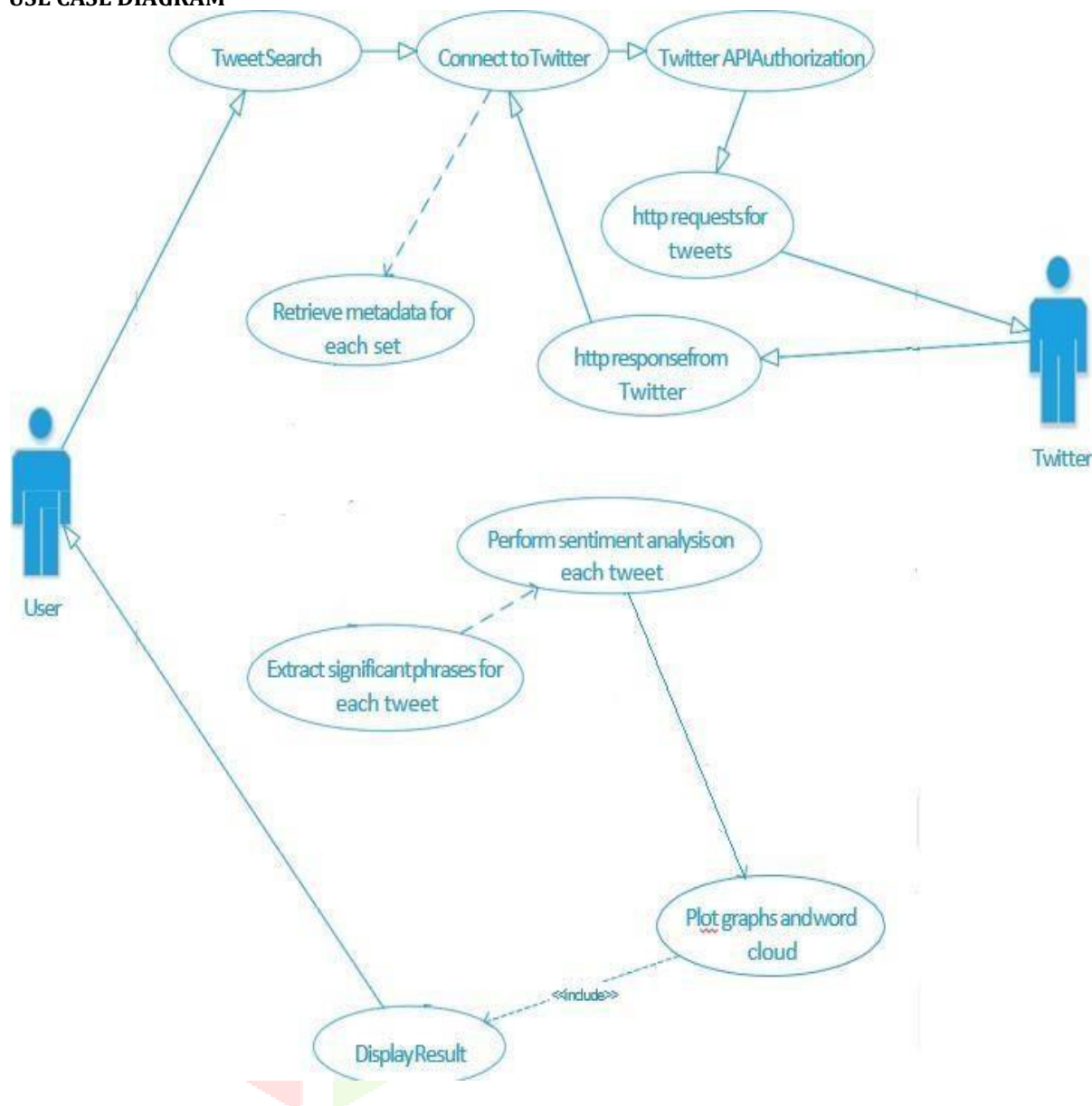
The pre-processed dataset has various discrete properties. In feature extraction methods, we extract different aspects such as adjectives, verbs and nouns and later these aspects are identified as positive or negative to detect the polarity of the whole sentence. Followings are the widely used Feature Extraction methods.

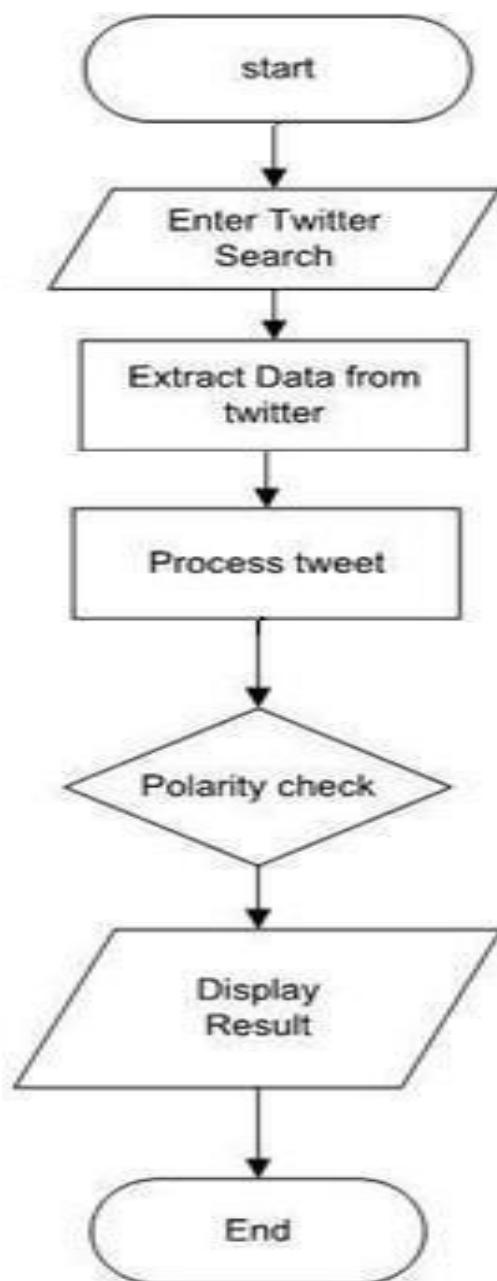
Terms Frequency and Term Presence: These features denote individual and distinct words and their occurrence counts.

Negative Phrases: The presence of negative words can change the meaning or orientation of the opinion. So it is evident to take negative word orientation in account.

Parts Of Speech (POS): Finding nouns, verbs, adjectives etc. as they are significant gauges of opinions.

USE CASE DIAGRAM



ACTIVITY DIAGRAM**7. CONCLUSION**

Twitter sentiment analysis comes under the category of text and opinion mining. It focuses on analyzing the sentiments of the tweets and feeding the data to a machine learning model to train it and then check its accuracy, so that we can use this model for future use according to the results. It comprises of steps like data collection, text preprocessing, sentiment detection, sentiment classification, training and testing the model. This research topic has evolved during the last decade with models reaching the efficiency of almost 85%90%. But it still lacks the dimension of diversity in the data. Along with this it has a lot of application

issues with the slang used and the short forms of words. Many analyzers don't perform well when the number of classes are increased. Also, it's still not tested that how accurate the model will be for topics other than the one in consideration. Hence sentiment analysis has a very bright scope of development in future.

8. REFERENCES

- <https://en.wikipedia.org/>
- <https://www.w3schools.com/>
- <https://www.youtube.com/>
- Dr. Khalid N. Alhayyan & Dr. Imran Ahmad "Discovering and Analyzing Important RealTime Trends in Noisy Twitter Stream" n.p
- R. Sharma, S. Nigam, and R. Jain, "Opinion mining of movie reviews at document level," arXiv preprint arXiv:1408.3829, 2014.
- R. Sharma, S. Nigam, and R. Jain, "Polarity detection at sentence level," International Journal of Computer Applications, vol. 86, no. 11, 2014.

