Covid-19 Fear Sensitivity analysis of Twitter Users

Shubham Shirude[1], Krushnakumar Gudpalle[2], Purnima Shinde[3], Jeffrin Koshy[4], Namrata Gawande[5]

[1-4]Student, Dept. of Computer Engineering, PCCOE, Maharashtra, India
[5]Assistant Professor, Dept. of Computer Engineering, PCCOE, Maharashtra, India

---***---

Abstract - *A way to identify not only the panic level but also to map it on a scale of twitter users related to covid-19. Traditional sentiment analysis provides a part of the solution but main problem here is different people experience pain and pleasure differently. We propose a solution to this problem by also analyzing the tweets of a user related to previous incidents and compare then and relatively map it on a scale. This model measures the sentiment relatively and can map it on the scale for various topics as well. As people use twitter to express themselves in different calamities, we can use previous data of a user and see how that user reacts to a particular situation. First the tweets related to topic are analyzed using a sentiment analysis model and then the model to compare tweet sentiments is applied for each user.*

***Key Words*: Sentimental Analysis, Calamities, Tweets, Panic Level**

1. INTRODUCTION

Platform such as Twitter have redesigned the way people find, share messages, and broadcast sensible information. In this process, Short text messages such as tweets are being generated and shared at an unparalleled rate which have huge amount of noise and redundancy. In each situation like pandemic or natural calamities twitter users tweet regarding the situation and also express themselves. Traditional sentimental analysis can just classify tweets and cannot scale them so the use of traditional analysis is limited. If we can scale the sentiment of a user then we can it can be useful for various purposes like targeted marketing, social surveys, etc.

There is no such model available to scale the sentiment so in this project we will try to represent the relative sentiment on a scale. The social media platform generates sentimental data regarding the current events in huge quantity, our motivation is the unexplored facts in the areas of sentiment analysis and its use in social welfare.

2. LITERATURE REVIEW

There has been an exponential growth in the use of textual analytics, natural language processing (NLP) and other artificial intelligence techniques in research and in the development of applications. Despite rapid advances in NLP, issues surrounding the limitations of these methods in deciphering intrinsic meaning in text remain. The rise in emphasis on AI methods for textual analytics and NLP followed the tremendous increase in public reliance on social media (e.g., Twitter, Facebook, Instagram, blogging, and LinkedIn) for information, rather than on the traditional news agencies [1]. People express their opinions, moods, and activities on social media about diverse social phenomena (e.g., health, natural hazards, cultural dynamics, and social trends) due to personal connectivity, network effects, limited costs and easy access. Many companies are using social media to promote their product and service to the end-users .Twitter data has been used widely for textual and emotions analysis. a combination of psychological and linguistic analysis was used in past research to extract emotions from multilingual text posted on social media. The core of the timeline generation module is a topic evolution detection algorithm, which consumes online/historical summaries to produce real-time/range timelines. The algorithm monitors quantified variation during the course of stream processing. A large variation at a particular moment implies a sub-topic change, leading to the addition of a new node on the timeline. The demand for analyzing massive contents in social medias fuels the developments in visualization techniques[2]. Timeline is one of these techniques which can make analysis tasks easier and faster. proposed the evolutionary timeline summarization (ETS) to compute evolution timelines similar to ours, which consists of a series of time-stamped summaries. In contrast, our method discovers the changing dates and generates timelines dynamically during the process of continuous summarization. The tweet stream clustering module maintains the online statistical data. Given a topic-based tweet stream, it is able to efficiently cluster the tweets and maintain compact cluster information. Proposed to summarize documents from the perspective of data reconstruction & select sentences that can best reconstruct the original documents.

Document summarization is of great value to many real-world applications, such as snippets generation for search results and news headlines generation. Summarization can represent the document with a short piece of text covering the main topics, and help users sift through the Internet, catch the most relevant document, and filter out redundant information[3]. So, document summarization has become one of the most important research topics in the natural language processing and information retrieval communities. News sites usually describe hot news topics in concise headlines to facilitate browsing. Both the snippets and headlines are specific forms of document summary in practical applications. Most 10 of the existing summarization methods aim to obtain the summary which covers the core information of the document. We believe that a good summary should contain those sentences that can be used to reconstruct the document as well as possible, namely, minimizing the reconstruction error. In this study, the twitter data has been pulled out from Twitter social media, through python programming language, using Tweedy library, then by using TextBlob library in python the sentiment analysis operation has been done. After the measuring sentiment analysis.

Sentiment analysis nowadays can be considered as one of the most popular research topics in the field of natural language processing. The uses of sentiment analysis are covered by some interesting scientific and commercial areas, such as opinion mining, recommender systems[4]. Nowadays the social media platforms such as Twitter, Facebook and YouTube, are a great source of information known as social data [2]. The events occurring in normal daily life are discussed on social media and any individuals are free to discuss and express their opinion about these events. Twitter API used for collecting related tweets to the coronavirus, then positive, negative and neutral emotion analyzed by using machine learning approaches and tools. In addition, for pre-processing of fetched tweets NLTK library is used and Text blob dataset for analyzing tweets is used, after that show the interesting results in positive, negative, neutral sentiments through different visualizations. Language Processing methodology is use to determine the sentiment of users from extracted tweets. Opinions are treated as data for analyzation. A qualitative approach was also used in determining the effects of the extreme community quarantine. The researcher also analyzes the effect of extreme community quarantine and other effects of the Pandemic to personal lifestyle based on the tweets of the users. Naive Bayes 'machine learning approach has been produced better execution, and it's been thought to be the concept for basic learning. This also brings out another ensemble technique that uses sentiment score because the input function for the classifiers in machine learning, SVM, Max Entropy, Decision Tree, Boosting, and Random Forest. As a result, the Logit Boost, a blended approach, performed better with accuracy of 74%.

The FP-Growth algorithm was adapted to the tweets in order to discover the most frequent patterns and its derived association rules, in order to highlight the tweeters insights relatively to COVID-19.

The FP-Growth algorithm was adapted to the tweets in order to discover the most frequent patterns and its derived Several epidemic periods have been observed in the world. In the recent years, this epidemics phenomenon has grown because of the contagion favored by the globalization. Nowadays, epidemics provoke extreme economic crisis at the scale of countries as well at the individual level. People can reach psychosis symptoms because of the contagion and countries may suffer from economic crisis due to people traveling restriction and social isolation. In [8], the authors give several battle hints to fight the virus as it travels the world, the metaphor of war is used. association rules, in order to highlight the tweeters insights relatively to COVID-19[5].

3. METHOD

The main method in this project is comparing the users tweet related to covid-19 to the previous important events. To classify the sentiment, we will be using the Naïve bayes algorithm. Example: Let's consider the user is situated in Australia and she has tweeted regarding the Wild fires recently and now during the pandemic of covid-19 the user tweets related to the pandemic. Now the phase 1 consists of classifying he sentiments in the wild fire's tweets and then the covid-19 tweets. Phase 2 is comparing the sentiments and scaling it.

4. ALGORITHM

A User tweets related to a topic which is trending. In the tweet the user expresses his feelings positive or negative. Traditional sentiment analysis provide a part of the solution but main problem here is different people experience pain and pleasure differently. A user tweets related to various topics.

Step 1) Classify the various topics about which the user has tweeted many number of times. Various text clustering algorithms can be applied

Step 2) Calculate number of positive and negative tweets of a particular topic. Let this be vector "a"

Step3) Calculate the number of tweets of covid-19 topic. Let this be vector "b"

Step 4) Compare the positive and negative tweets with covid-19 topic. Compare "a" and "b"

Step 5) Calculate relative positivity and negativity of covid-19 tweets and map it on scale. To calculate the positive and negative sentiment we use Naive bayes Classifier.

Naive bayes is a well known probabilistic classifier

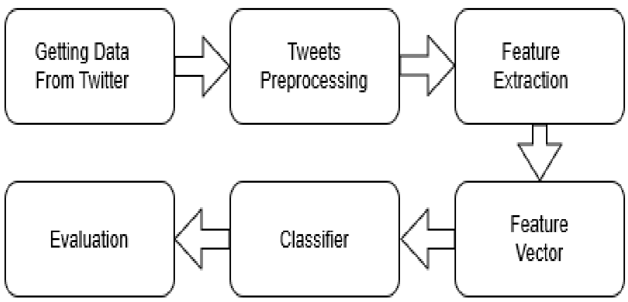

Fig -1: Sentiment Analysis Block Diagram and mixture of algorithms. The algorithm to classify positive and negative tweets will work as follows.

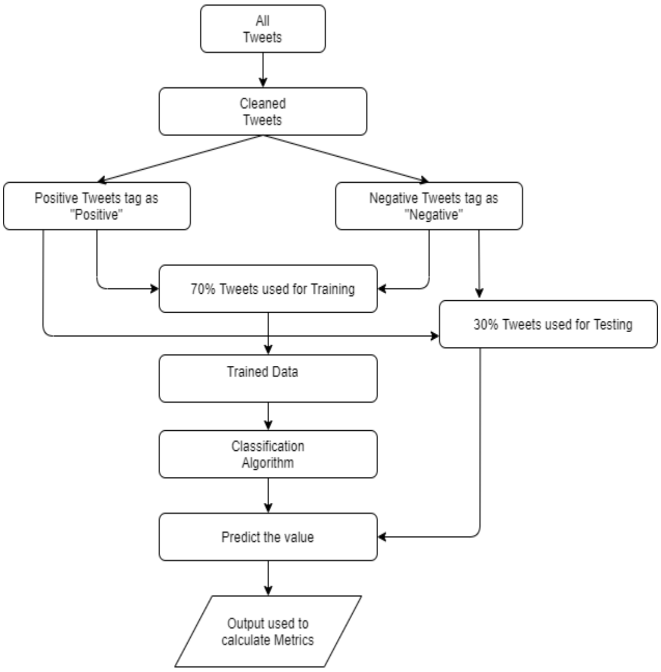

Fig -2: Naïve Bayes Flow Chart

5. CONCLUSIONS

Thus, we have provided a solution for mapping panic levels of a twitter user on a scale. We started with user tweets, classified them topic wise and then compared the various sentiments with Covid-19 sentiment.

ACKNOWLEDGEMENT

We are really grateful to Professor Namrata Gawande for guiding in completing our paper.

REFERENCES

[1] Jim Samuel 1, G. G. Md. Nawaz Ali 2, Md. Mokhlesur Rahman 3,4 , Ek Esawi 5 and Yana Samuel, "COVID-19 Public Sentiment Insights and Machine Learning for Tweets Classification," MDPI, Received: 28 April 2020; Accepted: 9 June 2020; Published: 11 June 2020.

[2] Zhenhua Wang; Lidan Shou; Ke Chen; Gang Chen; Sharad Mehrotra,"On Summarization and Timeline Generation for Evolutionary Tweet Streams", Published in: IEEE Transactions on Knowledge and Data Engineering (Volume: 27, Issue: 5, May 1 2015.

[3] Zhenhua Wang, Lidan Shou, Ke Chen, Gang Chen, and Sharad Mehrotra,"On Summarization and Timeline Generation for Evolutionary Tweet Streams",IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 27, NO. 5, MAY 2015.

[4] Kamaran Hussein Manguri, Rebaz Najeeb Ramadhan, Pshko Rasul Mohammed Amin, "Twitter Sentiment Analysis on Worldwide COVID-19 Outbreaks", May 2020 Kurdistan Journal of Applied Research.

[5] Hull David A. and Grefenstette Gregory. "A detailed analysis of English stemming algorithms". Rank Xerox Research Center Technical Report.1996.

[6] A. McCallum, and K. Nigam, "A comparison of event models for nave Bayes text classification", Journal of Machine Learning Research, Vol. 3, 2003, pp. 12651287

[7] T. Zhang, R. Ramakrishnan, and M.Livny, "BIRCH: An efficient data clustering method for very large databases", in Proc. ACM SIGMOD Int. Conf. Manage. Data, 1996, pp. 103-114.

[8] L Gong, J. Zeng, and S. Zhang, "Textstream clustering algorithm based on adaptivefeature selection", Expert Syst. Appl., vol. 38,no. 3, pp. 1393-1399, 2011.

[9] J. Zhang, Z. Ghahramani, and Y. Yang, "A probabilistic model for online document clustering with application to novelty detection", in Proc. Adv. Neural Inf. Process. Syst., 2004, pp. 1617- 1624.

[10] G. Erkan and D. R. Radev, "LexRank: Graph- based lexical centrality as salience intext summarization", J. Artif. Int. Res., vol. 22, no. 1, pp. 457-479, 2004.

[11] Z. He, C. Chen, J. Bu, C. Wang, L. Zhang, D. Cai, and X. He, "Document summarization based on data reconstruction", in Proc. 26th AAAI Conf. Artificial Intell., 2012, pp.620-626.