International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 08 Issue: 06 | June 2021 www.irjet.net p-ISSN: 2395-0072

Scaling Covid-19 Fear Sensitivity analysis of Twitter Users on Graph

Krushnakumar Gudpalle[1], Shubham Shirude[2], Jeffrin Koshy[3], Pornima Shinde[4], Shubham Biradar[5]

Krushnakumar Gudpalle
Computer Engineering Department
Pimpri Chinchwad College of Engineering, Pune, India

Shubham Shirude
Computer Engineering Department
Pimpri Chinchwad College of Engineering, Pune, India

Purnima Shinde
Computer Engineering Department
Pimpri Chinchwad College of Engineering, Pune, India

Jeffrin Koshy
Computer Engineering Department
Pimpri Chinchwad College of Engineering, Pune, India

Shubham Biradar
Computer Engineering Department
Pimpri Chinchwad College of Engineering, Pune, India

Prof. Namrata Gawande
Computer Engineering Department
Pimpri Chinchwad College of Engineering, Pune, India

Abstract :- *The popularity of social sites like Twitter, Facebook, and Instagram etc. is increasing rapidly. Huge number of tweets and short messages are being posted every single minute by the users from throughout the world, hence size of the data is very huge. On twitter platform tweets which are short in message are gets generated at enormous rate. For doing classification we mainly used naïve bayes algorithm. In each situation like pandemic or natural calamities twitter users tweet regarding the situation and also express themselves.The regular approaches of the summarization depends on the static or offline data. We removed this complexity of data and generated simple raw summarized text by using clustered data, And 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.*

Key Words :- **Sentiment Analysis, Calamities, Tweets, Panic Level, Naïve Bayes**

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 Survey :-

S. No	Studies	Algorithm / Method Used	Description
1.	COVID-19 Public Sentiment Insights and Machine Learning for Tweets Classification	Natural Language Processing (NLP)	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
2.	Twitter Sentiment Analysis on Worldwide COVID-19 Outbreaks	Natural Language Processing (NLP)	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
3.	On Summarization and Timeline Generation for Evolutionary Tweet Streams	Topic Evolution Detection Algorithm	Proposes an online tweet stream clustering algorithm to cluster tweets and maintain distilled statistics in a data structure called TCV The core of the timeline generation module is a topic evolution detection algorithm
4.	Document summarization based on data reconstruction	Clustering algorithm, Data Reconstruction	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
5.	Mining Twitter Data on COVID-19 for Sentiment analysis and frequent patterns Discovery	FP-Growth algorithm	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.

3. Challenges

Including data in the form of emoticons, images is difficult. Repetition of keywords in the summary. There is one big challenge in sentimental analysis is accuracy. Other system can do classification but fails to provide accuracy. The main challenge for our system is to provide greater accuracy than other previous systems

4. Motivation

As rapid growth in an internet, use of social media also increases. There are many social sites like Twitter, Facebook, Instagram etc. in which twitter has become one of the most popular social site for users to share information like text, audio, video etc. Short messages are being created and shared at massive rate. Twitter receives thousands of tweets per hour.

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 generate 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

5. Proposed 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 covid19 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.

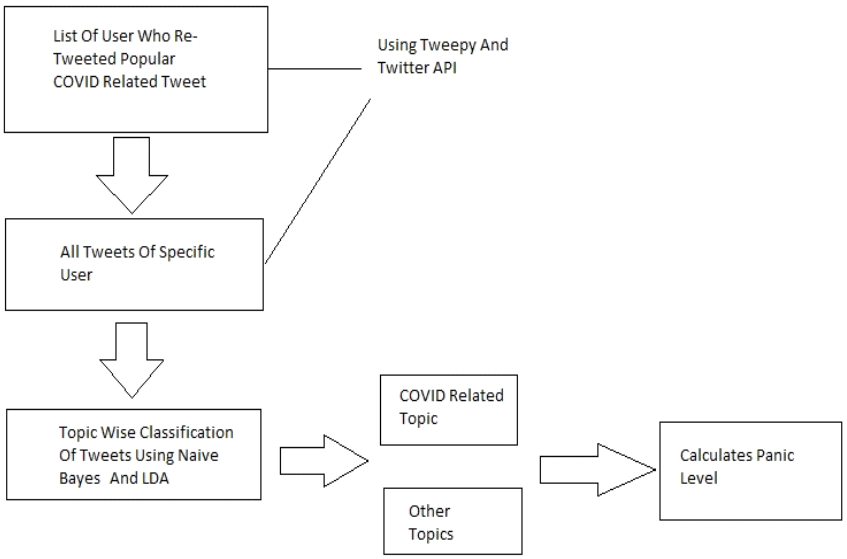

Fig -1: System Architecture

6. Algorithm Implementation:-

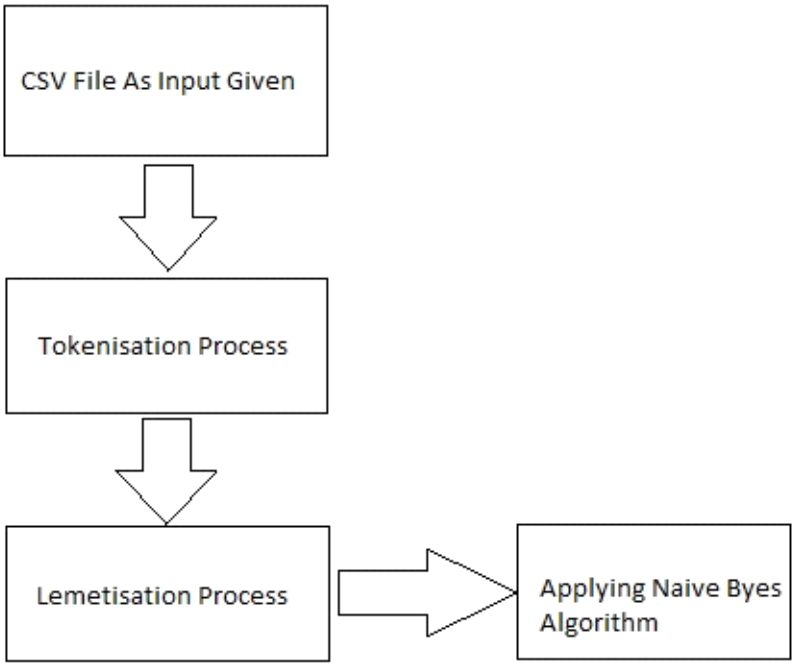

Fig -2: Module 1

Input:-

Our System Mainly Depends On Twitter For Data Input, For Taking Input From Twitter We Demanded Twitter API From Twitter. By Using Tweepy Library We Will Extract Tweets From Twitter

Library Used :- Tweepy

Tokenization:-

Tokenization is the process of breaking a sentence or splitting sentence in words, its task is chopping given sentence into words these words then called as tokens. While at the same time tokenization also removes uncertain characters from given documents these process is called as punctuation by default we are punctuating white space characters only.

In order to feed data to our classification model which is naïve bayes we first need to tokenize that data. For that we are using ***nltk (Natural Language Toolkit)*** library.

Library Used :- nltk (Natural Language Toolkit)

Example:- We are applying tokenization on data which is extracted by tweepy library from twitter and saved in CSV file. Suppose that one of the user twitted like

"The situation of covid-19 is getting worst day by day" then tokenization method will split this sentence into words like

["The", "situation", "of", "covid-19", "is", "getting", "worst", "day", "by", "day"]

Lemmatization:-

In lemmatization technique we actually remove the recursive words which are appearing in document more than once, for lemmatization we get input from tokenized dataset for accomplish this task we are using ***nltk's WordnetLemmatizer library.*** After lemmatization we get less noisy and redundant data from our dataset get removed so that our classification model will provide us best result

Library Used :- nltk's WordnetLemmatizer

Example:- If you Look closely then you will find that in our tokenized data which will be input for our lemmatization method there is one recurring word appeared which is day these will be removed from our data so lemmatization method will give you output like

["The", "situation", "of", "covid-19", "is", "getting", "worst", "day", "by"]

Naïve-Bayes Algorithm:-

After Tokenization and Lemmatization we will get suitable data which is less noisy to feed into our classification model

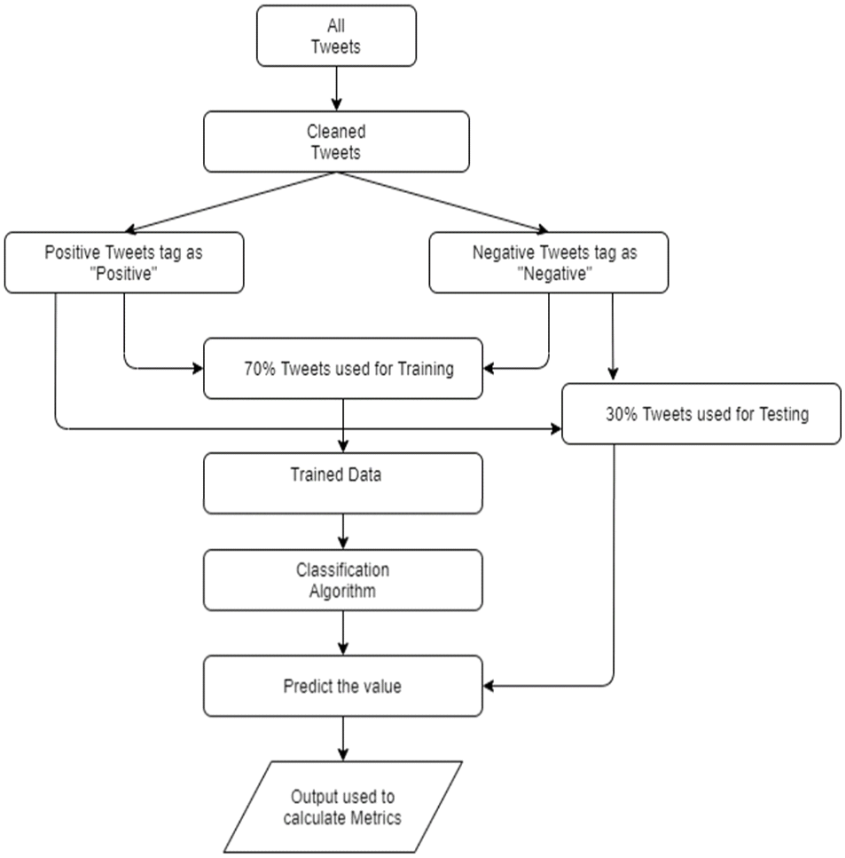

Fig -3: Naive Bayes Flowchart

9. 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. And after that we have mapped and scaled this data.

10. Conflict of Interest :-

All the group members contributed to the project equally. And there is no conflict of interest for this publication

11. Acknowledgment :-

All the group member are very thankful for twitter for providing an API. So we could run our process easily

All the group members thankful for our guide Ms. Namrata mam we guided throughout the process

12. 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.