



IJIRCCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 10, Issue 5, May 2022

ISSN INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA

Impact Factor: 8.165



9940 572 462



6381 907 438



ijircce@gmail.com



www.ijircce.com

A Review of Emotion Detection in Social Media during Covid-19

Archana R. Ugale, Pawan Devidas Chavan, Rohit Appasaheb Gunjal, Tejas Arun Kuldharan,
Mohammed Muzammil Razzaque Pathan

Department of Information Technology, SVIT, Nashik, Maharashtra India

Department of Information Technology, SVIT, Nashik, Maharashtra India

Department of Information Technology, SVIT, Nashik, Maharashtra India

Department of Information Technology, SVIT, Nashik, Maharashtra India

Department of Information Technology, SVIT, Nashik, Maharashtra India

ABSTRACT: Last two years since 2019 have proven to be very dreadful where a lot of lives were lost due to covid-19 pandemic. The Covid-19 pandemic proved introduced a lot of anxiety and uncertainty among the population. People shared a lot of things on the social media which proved to be helpful as well as panicky. Due to good thoughts on social media the general population found hope in the devastating pandemic, but with fake social media posts it also created panic among general population due to which a lot of lives were lost due to panic. So social media posts have to be monitored so that correct sentiment of the Covid-19 pandemic can be understood. So, a project has been designed with the help of latest technologies such as text mining, natural language processing and machine learning together. In our project we are going to use twitter as our social media platform as lakhs of tweets are posted daily and can give correct insight of the pandemic and other posts related to it. We are developing the application using a combination of java and python languages. First in our application we will fetch real time tweets using Twitter API. Then the incoming tweets will be cleaned. The cleaned tweets will then be more refined using stop words removal method. Then the stop words removed tweets will then be passed through lemmatization and only words with grammar will be kept. Thus, the final tweets after lemmatization can be used for machine learning. We will design a training dataset used for machine learning using various mediums available on the internet. Then finally we will apply machine learning algorithms SVM and Naïve Bayes and get correct classification of the tweets in two classes positive and negative. Thus, by using the analysis we will be able to understand the correct sentiment behind the tweets and understand emotions behind the Covid-19 pandemic.

KEYWORDS: Covid-19 Pandemic, Social Media, Twitter Tweets, Stop Words Removal, Data Lemmatization, Machine Learning, SVM, Naïve Bayes.

I. INTRODUCTION

Since 2019 a devastating pandemic has emerged due to which a lot of lives were lost. Lockdown was introduced where common communication medium of the general population was lost. The use of social media platforms increased rapidly where people shared positive as well as negative emotions about the pandemic. This sharing of information proved to be boon as well as hazardous. The misinformation caused panic and a lot of lives were lost. The main platform of it all was Twitter where a lot of information was shared, but as lakhs of tweets are generated daily it is very hard to track the correct sentiment of the pandemic and issues related to their emotions about the pandemic. There is need of a mechanism where a user or an authority who needs to track of the emotions of general population on various issues using social media.

In recent years many technologies and ideas have evolved in the text mining sector. The use data pre-processing is the main technology used in text mining of the data. In data pre-processing the data is cleaned and unwanted characters are removed from it. Then after cleaning the data stop words are removed from, it which tends to be a problem for text mining. Thus, another technique used in text mining is natural language processing-based lemmatization where correct meaning and grammar is understood and words with correct meaning are kept for text mining. Then latest machine learning algorithms are introduced which works and gives us classification intended using various algorithms. Thus, all these techniques can be combined to achieve our project target. So, the main motivation of our project is to.

- To develop emotion detection and analysis system.
- To use social media, natural language processing and machine learning technologies together.
- To make use of data pre-processing technology used in natural language processing.
- To make use of data lemmatization technology used in natural language processing.
- To get better insight of Covid-19 pandemic by understanding the emotions related to it and other issues.

Thus, the rest of the paper is structured as follows:

- Section II. discusses literature survey which help us understand and better design an emotion detection system using machine learning using studies published by other researchers.
- Section III. discusses problem definition with goals and objectives of emotion detection system using machine learning, methodology to be used with sub topics such as mathematical modelling and algorithmic use by the new framework and illustrates the proposed system with block diagram or system architecture and working of the new framework
- Section IV. illustrates the discussion of the new framework with the help of charts which shows how much time will it take to pre-process tweets and get classification results from the tweets.

II. LITERATURE REVIEW

This section discusses the idea and studies of various technicalities that can be used in designing the emotion detection system using machine learning. The literature review studies different ideas explained by other authors and help us understand the limitations of the system. Thus, by studying the limitations of other author techniques we can design one of the best and reliable emotion detection system using machine learning. Some of the researches are studied with the technique and limitations of it:

In 2020 Wenando et al. [1] presented the paper focusses mainly on “tweet-based sentiment analysis using machine learning algorithms”. This technique is quite good and covers all the things needed for a successful emotion detection but it lacks to explain how results can be improved using a different training dataset and not a common one for all.

In 2020 Arifiyanti et al. [2] presented the paper focusses mainly on “tweet sentiment classification using Naive Bayes and SVM”. This technique is quite good and covers all the things needed for a successful emotion detection but it lacks to explain how results can be improved using a different training dataset and not a common one for all.

In 2020 Jagdale et al. [3] presented the paper focusses mainly on “sentiment analysis using supervised learning algorithms”. This technique is quite good and covers all the things needed for a successful emotion detection but it lacks to explain how results can be improved using a different training dataset and not a common one for all.

In 2019 Hartanto et al. [4] presented the paper focusses mainly on “Classification of Tweet data using Naive Bayes algorithm”. This technique is quite good and covers all the things needed for a successful emotion detection but it lacks to explain how results can be improved using a different training dataset and not a common one for all.

In 2019 AlZu’bi et al. [5] presented the paper focusses mainly on “multilabel tweet classification using KNN, J48 and RF”. This technique is quite good and covers all the things needed for a successful emotion detection but it lacks to explain how results can be improved using a different training dataset and not a common one for all.

III. PROJECT DESIGN

This section discusses various issues arising while designing the emotion detection system using machine learning and how to resolve the problems by studying and proposing a methodology to solve and design a new and better system.

A. Problem Statement

With the rapid increase of Covid-19 pandemic there was also an exponential increase in information sharing about the pandemic and related issues. The information sharing was done on platforms like twitter etc. The social media proved to be both good and bad in the pandemic. The good thing was that people got confidence in facing the pandemic by the experiences shared by general population. But the other side of this information sharing proved to be bad as rumours were spread and a panic was created which lost lives. So as the pandemic progressed there arise a need of understanding the correct emotion behind the information shared using tweets on Twitter platform. So, sentiment analysis can be done on the tweets for emotion detection using a combination of technologies such as natural language

processing and machine learning together. By developing a system which can classify large number of tweets generated daily on the pandemic and related topics the general population can understand and handle the Covid-19 pandemic well. So, our major goals and objectives to solve this problem is stated as follows:

- To create emotion detection system using tweets as medium.
- To use Tweepy python library to access real time tweets about a specific keyword and language.
- To make effective use of data pre-processing technique stop words removal to get proper results.
- To make effective use of data lemmatization to clean the data for analysis and keep grammatically meaningful words.
- To create a different training dataset for keyword specific tweets.
- To apply machine learning algorithms Naïve Bayes and SVM to classify tweets in mainly two emotions positive and negative.

B. Methodology

This section will discuss the mathematical model that is to be used with relevant mathematics and algorithmic use for designing emotion detection system. It is discussed as follows:

1) Mathematical Model

Our emotion detection system using machine learning can be explained in two sets with probability, success and failure conditions.

- Pre-Processing Data Module:
Set (PD) = {PD0, PD1, PD2, PD3, PD4, PD5}

PD0 ∈ PD = Enter keywords to fetch tweets.

PD1 ∈ PD = Fetch and clean tweets.

PD2 ∈ PD = Remove stop words from tweets using NLTK.

PD3 ∈ PD = Apply feature extraction using NLTK.

PD4 ∈ PD = Keep tweets with grammatical meaning.

PD5 ∈ PD = View tweets.

- Machine Learning Module:
Set (ML) = {ML0, ML1, ML2, ML3, ML4, ML5}

ML0 ∈ ML = Create training dataset.

ML1 ∈ ML = Fetch feature extracted tweets.

ML2 ∈ ML = View tweets.

ML3 ∈ ML = Create testing dataset.

ML4 ∈ ML = Train and apply Naïve Bayes algorithm.

ML5 ∈ ML = Train and apply SVM algorithm.

ML5 ∈ ML = View results in 2 emotions positive & negative.

So, by studying the relevant sets as illustrated above we come to understand that some elements are common in both the illustrated modules and they can be forward by using intersection of the sets

$$x \in PD \cap ML \text{ if } x \in PD \text{ and } x \in ML. \quad (1)$$

Thus, the probability of intersection of elements in both modules can be given as

$$P(PD \cap ML) = P(PD) + P(ML). \quad (2)$$

Thus, the most probable use of the intersecting elements can be put forward by using probability of how the element will be used at given time

$$PD \cap ML = \{PD4\}. \quad (3)$$

So, the common elements intersection can be illustrated as

$$P(PD | ML) = \frac{P(PD \cap ML)}{P(ML)}. \quad (4)$$

Thus, we conclude that our project “Emotion Detection in Social Media During Covid-19 using Machine Learning” success and failure will depend upon the tweets fetched from Twitter, i.e., if the tweets are not fetched properly the project will not work thus the system will fail, so our new system supports NP-Hard and not NP-Complete.

2) Algorithms Used

The new system will use Naïve Bayes and SVM algorithms to create a better emotion detection system using machine learning.

- Naïve Bayes:

This algorithm will be used to classify testing data in two classes positive and negative. It is a machine learning algorithm called as Naïve Bayes. It is a supervised learning model. It is one of the fastest classification algorithms. To predict the class of unknown datasets it uses Bayes theorem of probability. The working is quite simple i.e., a Naive Bayes classifier assumes that a presence of a particular feature in a class is unrelated to any other feature in the class. The naïve Bayes classifiers very easy to build and works on large datasets. It converts a dataset in to a frequency table than uses Naïve Bayesian equation to calculate posterior probability of each class and the highest outcome of posterior probability is the predicted class.

- SVM:

This This algorithm will be used to classify data in two classes positive and negative. It is a machine learning algorithm called as support vector machine. It is a supervised learning model. An SVM model can be representation of points in space mapped in such a way that the examples of each separate categories are placed in a separate plane by a clear gap that is wide as possible. It needs a training and testing dataset for classification. It needs to be trained and tested properly for better results with good datasets. It can give good results with small training dataset.

3) Proposed System

This section discusses 3 main modules and the sub modules associated with them. The modules and the sub modules associated with them can be explained using block diagram or system architecture as illustrated in Fig. 1. The working of the new system can be explained as:

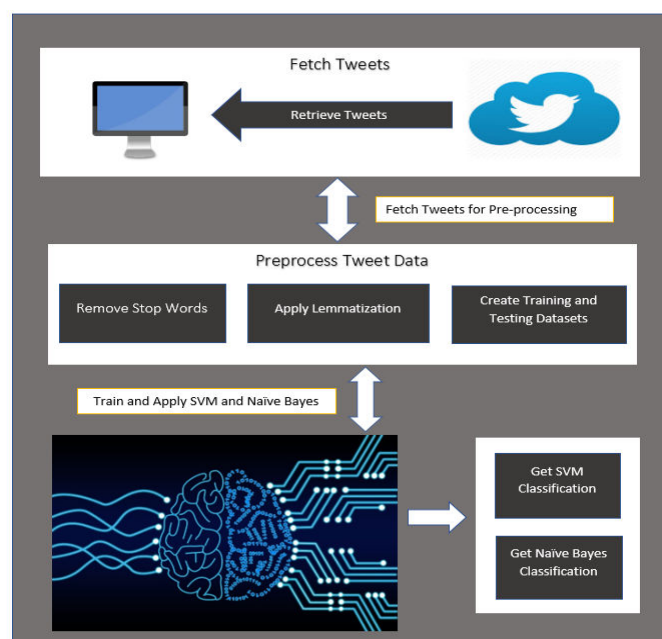


Fig. 1: Block Diagram/ System Architecture

- Fetch Tweets:

In this module explains how tweets will be fetched. First the user will enter a keyword. Then the tweets will be fetched at real time from Twitter server using Tweepy library. Then the fetched tweets will be cleaned for unwanted characters.

- Pre-process Tweet Data:

In this module the data after cleaning will be fetched and then the stop words will be removed from the tweets using NLTK stop words array. Then lemmatization will be applied on the tweets and only grammatically meaningful words in the tweets will be kept and other will be removed. Then a training dataset will be created using the positive and negativesentiment words taught to it. Then testing datasets will be created for testing.

- Train and Apply SVM and Naïve Bayes:

In this module the SVM and Naïve Bayes algorithms will be trained using the previously created dataset. Then testing dataset will be passed to it for analysis. The SVM and Naïve Bayes algorithms will classify the tweets in two labels positive emotion and negative emotion so the correct emotion behind the tweets can be understood by the user.

IV. RESULTS AND DISCUSSIONS

This section discusses the time taken to fetch tweets, pre-process them and apply SVM and Naïve Bayes algorithms with the help of charts which are explained as

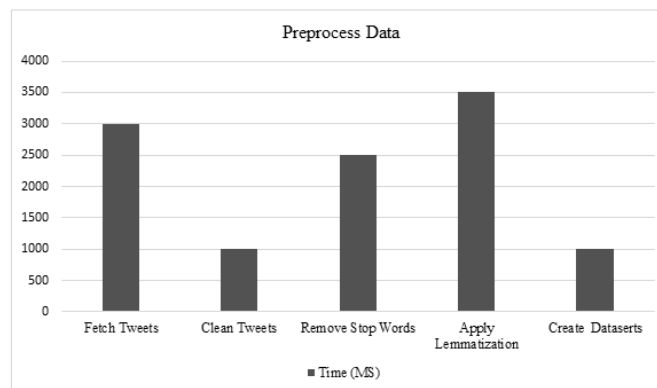


Fig.2: Pre-Process Data Time Line Chart

The Fig. 3.illustrates chart which explains how much time is taken to fetch tweets, clean tweets, remove stop words, apply lemmatization and create datasets. The chart suggests that applying lemmatization on tweets takes more time than other processes.

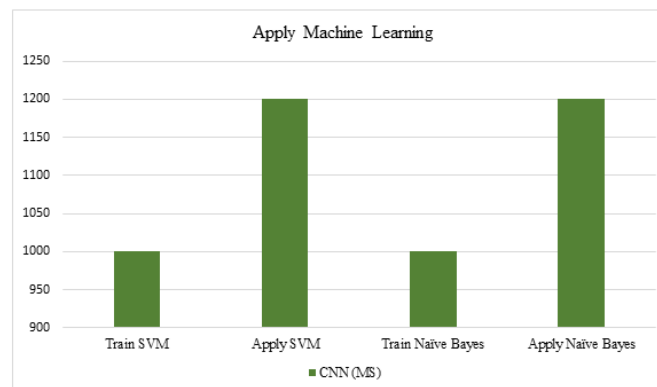


Fig.3: Apply Machine Learning Time Line Chart

The chart in Fig.3.illustrates chart which explains how much time is taken to train SVM, apply SVM, train Naïve Bayes, apply Naïve Bayes.The chart suggests that applying SVM and Naive Bayestakes same time.

V. CONCLUSION

In this paper we developed a sentiment analysis approach to understand the sentiments behind the emotions in complicated pandemic like Covid-19. The paper presents a sentiment analysis technique with a combination of technologies such as natural language processing and machine learning together. The main idea of application is to enter a keyword and fetch tweets and classify them in two emotions positive and negative. In our paper we have followed papers [1][2][3][4][5], and studied the techniques used in them and how sentiment analysis is done and how we can improve the results of the same by our new technique. We have used Tweepy python library to fetch tweets. We have used NLTK python library to pre-process tweets. We are creating different training datasets and not using a common one for all. We are using machine learning algorithms SVM and Naïve Bayes to classify the tweets in two emotions mainly positive emotion and negative emotion. We have developed the application using a combination of both Java and Python languages together to achieve the desired results. Thus, we conclude that our application will help the general population understand the correct sentiment in a trying time like Covid-19 pandemic.

REFERENCES

- [1] Febby Apri Wenando, Regiolina Hayami, Bakaruddin and Ali Yunda Novermahakim., "Tweet Sentiment Analysis for 2019 Indonesia Presidential Election Results using Various Classification Algorithms.," in IEEE-2020.
- [2] Amalia Anjani Arifiyanti and Eka Dyar Wahyun., "Emoji and Emoticon in Tweet Sentiment Classification.," in IEEE-2020.
- [3] Rajkumar S. Jagdale, Sonal S. Deshmukh. "Sentiment Classification on Twitter and Zomato Dataset Using Supervised Learning Algorithms.," in IEEE-2020.
- [4] Anggit Dwi Hartanto, Ema Utami, Sumarni AdiAmikom Yogyakarta Yogyakarta and Harish Setyo Hudnanto, "Job Seeker Profile Classification of Twitter Data Using the Naïve Bayes Classifier Algorithm Based on the DISC Method.," in IEEE-2019.
- [5] Shadi AlZu'bi, Omar Badarneh, Bilal Hawashin, Mahmoud Al-Ayyoub, Noh Alhindawiz and Yaser Jararwehy. "Multi-Label Emotion Classification for Arabic Tweets.," in IEEE-2019.
- [6] P. Nakov, A. Ritter, S. Rosenthal, F. Sebastiani, and V. Stoyanov, "SemEval-2016 Task 4: Sentiment Analysis in Twitter," arXiv:1912.01973 [cs], Dec. 2019.
- [7] F. Nausheen and S. H. Begum, "Sentiment analysis to predict election results using Python," in 2018 2nd International Conference on Inventive Systems and Control (ICISC), Jan. 2018.



INNO  SPACE
SJIF Scientific Journal Impact Factor

Impact Factor: 8.165

 **doi**[®]
cross **ref**

ISSN INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

 9940 572 462  6381 907 438  ijircce@gmail.com



www.ijircce.com

Scan to save the contact details