Analysing Twitter Data to Observe Emotional Behavior of People in COVID-19 Pandemic

Aditya Raj
Indian Institute of Information Technology, Allahabad
Prayagraj, Uttar Pradesh, India
itm2017001@iiita.ac.in

Ayush Sinha
Indian Institute of Information Technology, Allahabad
Prayagraj, Uttar Pradesh, India
sinhamnnit@gmail.com

Ranjit Singh
Indian Institute of Information Technology, Allahabad
Prayagraj, Uttar Pradesh, India
ranjitsingh@iiita.ac.in

O P Vyas
Indian Institute of Information Technology, Allahabad
Prayagraj, Uttar Pradesh, India
opvyas@iiita.ac.in

***Abstract*—A study with an objective to analyse how people are reacting to coronavirus pandemic was done using sentiment analysis on about 273,000 tweets in English language related to COVID-19, Which were posted between April 1,2020 and June 30,2020. Tweets were analysed with lexicon based approach and Bayesian Classification based approach separately. Majority of tweets showed association to emotions like trust, fear ,sadness,anger and anticipation. Surprise was least observed emotion.Majority of tweets showed positive sentiment due to high association with trust. Results clearly demonstrate that the 20 percent of the collected tweets are positive and about 10 percent are Negative. Moreover 55.5 percent tweets showed trust, 46.2 percent tweets showed fear, 42.6 percent tweets showed sadness, 39.5 percent tweets showed anger and 38.3 percent tweets showed association towards anticipation.**

***Index Terms*—coronavirus, COVID-19, Lexicon, Sentiment, Bayesian**

I. Introduction

COVID-19 is a viral disease originated in Wuhan,China in late 2019.It is caused by a novel corona virus.On December 31,2019, China reported to World Health Organisation (WHO) about the virus outbreak. On January 11,2020, first death from COVID-19 was reported in China.First case of the disease outside china was reported in Thailand on January 13, 2020. On January 30,2019,WHO declared COVID-19 a global emergency as about 10000 people were affected till date and on March 11,2020, WHO declared COVID-19, a pandemic after reaching more than 1.26 lakh cases globally.

As strict social distancing measures were being implemented around the globe,People started limiting themselves to their homes.Business started to suffer,people either lost jobs or were awarded work from home and were forced to live inside their houses to save themselves. These things created an imbalance in daily life of people.There was an environment of sadness and fear about the disease but people were also reacting positively to the situation. In this paper we have performed sentiment analysis on the Twitter data related to COVID-19 to analyse how people were reacting to the incidents related to COVID-19. Twitter is a micro blogging site and is popular among the people in sharing thoughts and information because of maximum size of tweets which is currently 280 characters and method of interaction among the users in which any user can comment on other user's tweet and retweet it too.People use twitter to share their thoughts and propagate information. The rapid sharing of opinions and large user activity on the platform makes it first choice of the researchers for doing analysis of people's opinion on almost any topic.

II. Literature Review

Sentiment Analysis is a natural language processing task in which we determine underlying sentiments of a text.The sentiments can be either positive,negative or neutral.Emotions are also an important part of communication between individuals. The exchange of emotions trough text is done by using words that symbolise the emotion. These can be expressed in a word, a sentence or a full page document consisting of hundreds of words.

The analysis of feelings and emotions of people is a potential research problem and Businesses also use it to collect feedback for their products. Analysis of feelings and sentiments can be done in any area to gain insights of what people are thinking and how they are reacting to the situation. We can categorise sentiment analysis approaches into two groups namely machine learning approach and Lexicon based approach.

A. Machine Learning Approaches

Many machine learning approaches like Naive Bayes, Maximum Entropy (MaxEnt), Support Vector Machines(SVM), Long Short Term Memory(LSTM) have been used in past for sentiment analysis.

1) Naive Bayes: Naive Bayes [21] is a classification method which uses Bayes' Theorem [22] with strong independence assumptions between the features of the text. A Naive Bayes classifier assumes that the closeness of a specific feature in a class is disconnected to the closeness of some other element.Bayes Theorem is a method of computing for distinguishing likelihood P(a/b) from P(a),P(b) and P(b/a) as following:

$$P(a/b) = [P(b/a) * P(a)]/P(b) \tag{1}$$

where P(a/b) is the posterior probability of a given predictor b and P(b/a) is the likelihood of b given a.P(a) is prior probability of class a and P(b) is prior probability of class b.

2) Support Vector Machine: Support Vector machines are supervised learning models in which a set of hyperplanes are constructed which are used in tasks like classification,regression and outlier detection. [23]

3) Long Short Term Memory(LSTM): LSTM are special category of recurrent neural networks capable of learning long term dependencies. [24]

4) Maximum Entropy(MaxEnt): MaxEnt classifier [25] works by estimating the conditional distribution without assuming that features are conditionally independent as in Naive Bayes.It is based on the principle of maximum entropy [26].

In [3],Neethu M S and Rajasree R have used Naive Bayes Classifier, SVM Classifier, Ensemble Classifier and Maximum Entropy classifier to perform sentiment analysis on tweets.They concluded that Naive Bayes had better precision compared to other classifiers but slightly lower accuracy and recall. In [2],Alec Go, Richa Bhayani, and Lei Huang have introduced an approach for sentiment analysis in which they have used feature extractors like unigrams, bigrams with part of speech tags along with machine learning classifiers such as naive bayes,Maximum Entropy and Support Vector Machines to classify tweets in English language. They created a framework that treats classifiers and feature extractors as two different entities and have tried different combination of classifiers and feature extractors. They concluded that support vector machines had better performance with unigrams whereas Naive bayes and MaxEnt classifiers performed well with bigrams. They have also concluded that part of speech tags were not useful for sentiment analysis with all the three algorithms.

In [18] Bo Pang,Lillian Lee, and Shivakumar Vaithyanathan have used movie review data and performed classification using Naive Bayes, Maximum Entropy and support vector machine. They have randomly sampled 700 positive and 700 negative reviews and performed classification using all the three methods. Among the three, Support Vector Machine performed best and naive bayes had minimum accuracy.

In [19],Priyanka Thakur and DR. Rajiv Srivastava have used long short term memory deep learning algorithm for sentiment analysis. They have extracted lexicons from the review and tagged their part of speech. Then part of speech tags' occurrence are extracted and each part of speech tags' sentiment score was calculated by referring to SentiWordNet dictionary.Finally they fed the extracted features into LSTM deep learning classifier for sentiment classification.

In [28],Prachi Sanghvi, Disha Shah, Prof. Bharathi H. N. have implemented movie rating system based on sentiment analysis of reviews.They have implemented Naive Bayes and SVM to classify sentiments as positive, negative and neutral. In this classification Naive Bayes had better accuracy than support vector machines.

In [27],Dhiraj Gurkhe,Niraj Pal, Rishit Bhatia have used different feature extraction method like unigram, bigram and unigram with bigram and have applied naive bayes classification to perform sentiment analysis.They observed that highest accuracy was obtained from unigram and least from bigrams. Combiantion of unigram and bigram had average result.This may be due to the reason that naive bayes assumes strong independence between features of the text.

In [4] R. K. Bakshi, N. Kaur, R. Kaur and G. Kaur have introduced a new sentiment analysis method to analyse stock marked trends of a company based on tweets related to the company.

In [5] Z. Jianqiang, G. Xiaolin and Z. Xuejun have introduced a word embedding method obtained by unsupervised learning based on large twitter corpora which uses latent contextual semantic relationships and co-occurrence statistical characteristics between words in tweets. These word embeddings were combined with n-grams features and word sentiment polarity score features to form a sentiment feature set of tweets. The feature set was then integrated into a deep convolution neural network for training and predicting sentiment classification labels.

In [6] R. Wagh and P. Punde have surveyed about sentiment analysis types and techniques used to perform extraction of sentiment from tweets. They have performed comparative study of different techniques and approaches of sentiment analysis having tweets as data.

In [7] V. S. Pagolu, K. N. Reddy, G. Panda and B. Majhi have employed two different textual representations, Word2vec and N-gram, for analyzing the public sentiments in tweets.They have applied sentiment analysis and supervised machine learning principles to the tweets extracted from Twitter and analyze the correlation between stock market movements of a company and sentiments in tweets.

In [8],G. Gautam and D. Yadav have used Naive Bayes, Maximum entropy and SVM along with the Semantic Orientation based WordNet to perform sentiment analysis of customers' review for a particular company.

In [9],M. Hao et al. have introduced three novel time-based visual sentiment analysis techniques namely "topic-based sentiment analysis that extracts, maps, and measures customer opinions", "stream analysis that identifies interesting tweets based on their density, negativity, and influence characteristics" and "pixel cell-based sentiment calendars and high density geo maps that visualize large volumes of data in a single view". They have applied these techniques to a variety of twitter data to show their distribution and patterns, and to identify influential opinions.

In [10], B. Gokulakrishnan, P. Priyanthan, T. Ragavan, N. Prasath and A. Perera have discussed an approach where a publicised stream of tweets from the Twitter was preprocessed and classified based on their emotional content as positive, negative and irrelevant; and analysed the performance of various classifying algorithms based on their precision and recall in such cases.
In [11] S. Bahrainian and A. Dengel have introduced a novel solution to target-oriented sentiment summarization and sentiment analysis of short informal texts with a main focus on tweets. They have compared different algorithms and methods for sentiment polarity detection and sentiment summarization and showed that their hybrid polarity detection system not only outperforms the unigram state-of-the-art baseline, but also could be an advantage over other methods when used as a part of a sentiment summarization system.
In [12] P. Mishra, R. Rajnish and P. Kumar have performed sentiment analysis on tweets related to Digital India campaign using dictionary based methods.

B. Lexicon Based Approach

Lexicon based approaches are based on assumption that the contextual sentiment criteria is the sum of sentiment orientation of each word in the text. These approaches utilise the sentiment lexicon to describe the polarity of a text. It is more understandable and can be easily implemented in contrast to machine learning based algorithms. But it requires human involvement in the process of text analysis.
In [13], Y. Arslan, A. Birturk, B. Djumabaev and D. Küçük have employed several sentiment analysis techniques by using dynamic dictionaries and models, and performed some experiments on limited but relevant datasets to understand the popularity of some terms and the opinion of users about them.
In [14], M. E. Basiri and A. Kabiri have compared four lexicons in persian language.
In [15], K. Mao, J. Niu, X. Wang, L. Wang and M. Qiu have used domain based lexicons along with machine learning methods to perform sentiment analysis on product reviews.
In [16], N. A. Abdulla, N. A. Ahmed, M. A. Shehab and M. Al-Ayyoub performed corpus based and lexicon based sentiment analysis on manually annotated datasets.
In [17] E. W. Pamungkas and D. G. P. Putri have performed lexicon based sentiment analysis in Bahasa Indonesia.

In [29], Prabu Palanisamy, Vineet Yadav and Harsha Elchuri have proposed a sentiment analysis tool which uses lexicon based approach for discovering sentiments of text.
In [30], Florian Wogenstein,Johannes Drescher,Dirk Reinel,Sven Rill,Jörg Scheidt have presented a study of aspect based opinion mining using a lexicon based approach. They have used a phrase based opinion lexicon for German language to perform opinion mining in insurance domain.
In [31],Habiba Drias and Yassine Drias have done sentiment analyis and frequent pattern mining on tweets related to COVID-19. They have used fp-growth algorithm [33] for frequent pattern mining and NRC emotion lexicon [36] for sentiment analysis.
In [32], Kamaran H. Manguri,Rebaz N. Ramadhan,Pshko R. Mohammed Amin have used TextBlob [34] python library to analyse sentiments of the tweets related to COVID-19.

III. Methodology

A. Data Retreival

Twitter provides Twitter API which is an interface to connect any application with the social media website. Each API is assigned four unique keys which are used to authenticate and verify the users. We have used Tweepy [35] python module to download tweets related to corona virus.The tweet ID of tweets related to corona virus were obtained from github repository of Pancea Lab [1].This dataset contains tweets ID from several languages stored in separate files for each day. After collecting tweet IDs, tweets were downloaded from twitter using Tweepy module and Twitter API. API serves as an access provider to the twitter data. Tweets were downloaded using their tweet IDs and tweets in English language were filtered out.3000 English tweets were collected for each day from 1 April,2020 to 30 June,2020 for sentiment analysis.

B. Pre-processing of Tweets

Tweets which were downloaded initially contained links to web pages,emoticons , hashtags and ID of other twitter users.Each tweet was processed and all the redundant things were removed from the tweets.

1) Link Removal: URL links are often mentioned as a part of tweet but they don't contribute to sentiment expressed and hence are removed from the tweet.

2) Cleaning: Hashtag Sign from text of each tweet was removed and words were kept in the database. IDs of other twitter users were also removed from the tweets as they don't contribute to the sentiment of the tweet.

3) Text Extraction: English words were extracted from the preprocessed tweets for each tweet and kept in the final database of tweets.

4) Tokenisation: Each tweet is divided into phrases like noun, adjective, adverb and words like a,the,are and other pronouns,auxiliary verbs have been omitted to reduce redundant words. Each word is then assigned a part of speech tag.

C. Sentiment Analysis

1) Naive Bayes: Sentiment analysis has been done using TextBlob library in python.It gives two parameters as output namely polarity and subjectivity. Value of polarity lies between -1 and 1 where 1 means positive sentiment and -1 means negative sentiment. Value of subjectivity lies between 0 and 1 where 0 means a factual information and 1 means personal opinion, emotion or judgement. Algorithm for sentiment analysis using Naive Bayesian Classifier has been discussed below.

2) Emotion Lexicon based analysis: NRC Lexicons [36] are crowd sourced list of words which identify different emotions. These emotions are:

- Anger
- Anticipation

Algorithm 1: Sentiment analysis using Naive Bayesian Classifier

Result: Class with higher probability is assigned to the text

- Given a data set D consisting of text which belong to classes Positive and Negative, Prior probability of both classes is calculated as:
  P(Positive)= Number of texts classified as positive / Total Number of text in D
  P(Negative)= Number of text classified as negative / Total Number of text in D

- The total number of word frequencies of both classes is given by:
  n_P = the total number of word frequency of Positive Class
  n_N = the total number of word frequency of Negative Class

- Let $count_{i,j}$ be the count of $word_i$ in a given class j,then the conditional probability of keyword occurrence of any for a given class is:
  $P(word_i/Positive) = count_{i,Positive}/n_P$
  $P(word_i/Negative) = count_{i,Negative}/n_N$

- For classifying a new text T having words w_k Calculate
  $P(Positive/T) = P(Positive) * \prod_k P(w_k/Positive)$
  $P(Negative/T) =$
  $P(Negative) * \prod_k P(w_k/Negative)$

- Disgust
- Fear
- Joy
- Sadness
- Surprise
- Trust

There are 27000 lexicons for emotions in NRCLex python module. NRCLex measures emotional affect from a body of text. Affect dictionary contains approximately 27,000 words, and is based on the National Research Council Canada (NRC) affect lexicon and the NLTK library's WordNet synonym sets. Association of each emotion to the tweet text is calculated and accordingly each tweet is said to belong to the emotion with highest association to it.

IV. Observation and Result

A. *Trends in emotional association of tweets*

Association of tweets to the emotions was balanced and didn't change much during the three months of observation namely April,May and June.Almost 55 percent tweets on average showed trust which remained highest in all the three months. Second most highest emotion observed was fear which remained at about 46 percent. Third significant emotion was fear to which nearly 43 percent tweets have shown association. Next most significant emotion observed in tweets was of anger which was at about 39 percent average. Joy,anticipation,disgust and surprise were observed in 36 percent,38 percent,36 percent and 34 percent respectively.

Algorithm 2: Assigning Emotions to text

- For each word in text,Search for it's presence in lexicon file and note which emotions are associated to it.
- Calculate the number times each emotion has appeared for whole text.
- Divide the frequency of each emotion by total number of words in the text.
- Emotions having highest frequency are assigned to the text
- A text can be assigned more than one emotions if frequency of two emotions are highest and same.

Colour	Emotion
Red	Anger
Black	Sadness
Green	Trust
Indigo	Anticipation
Maroon	Fear
Yellow	Joy
Purple	Disgust
Cyan	Surprise

TABLE I: Legend for Graph in Figure 1

Figure 1 shows the graph for trend of emotions in whole duration of observation.Table 1 shows the colour of the line in graph in Figure 1 to which emotions are associated. Legends couldn't be associated to graph directly due to size and complexity in visual representation.

B. *Average emotional association of tweets*

Most observed emotion in each month was trust and least observed emotion was surprise.The results have been tabulated in table II.

The Average percentage of tweets associated to different emotions are tabulated in Table III.

Most Observed Emotions	Least Observed emotions
Trust	Disgust
Fear	Joy
Sadness	Surprise
Anger	
Anticipation	

TABLE II

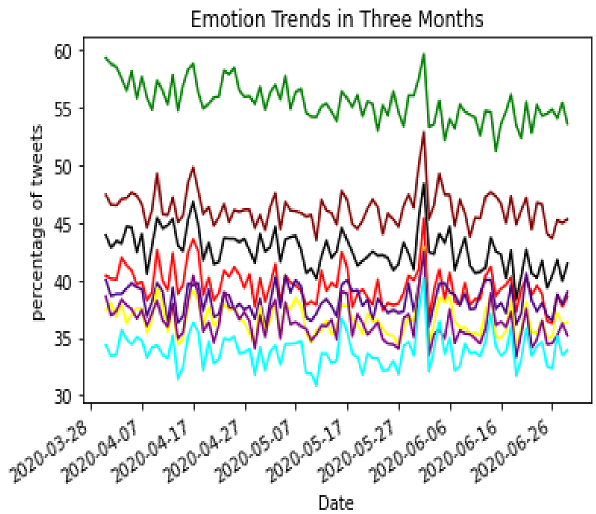


Fig. 1: Emotion trends in three months

Emotion	April	May	June	Average
Anger	40.3	39.2	39.1	39.5
Anticipation	38.4	38.2	38.4	38.3
Disgust	37	36.1	35.9	36.3
Fear	46.5	45.8	46.4	46.2
Joy	36.9	36.3	36.6	36.6
Sadness	43.5	42.4	42.0	42.6
Surprise	34	33.4	34.3	34
Trust	56.8	55.3	54.4	55.5

TABLE III: Average percentage of tweets associated to different emotions

C. Trends in Sentiment association of tweets

It is observed through experiment that 20 percent of the tweets have shown positive sentiment and less than 10 percent of the tweets have shown negative sentiment when sentiment analysis was done with Bayesian method.Also Positive sentiment has almost remained constant and negative sentiment tweets have shown some variations.Subjectivity of the tweets have remained above 50 percent which shows that people were just expressing their opinion and not stating facts which is clearly visible from objectivity values which are at about 45 percent.
Table IV shows the association of positive sentiment, negative sentiment, subjectivity and objectivity to colour of lines in the graph in figures 2,3,4 and 5. Legends couldn't be associated to graph directly due to size and complexity in visual representation.

Colour	Sentiment
Red	Negative
Green	Positive
Yellow	Subjectivity
Purple	Objectivity

TABLE IV: Legend for graphs in figure 2,3,4 and 5

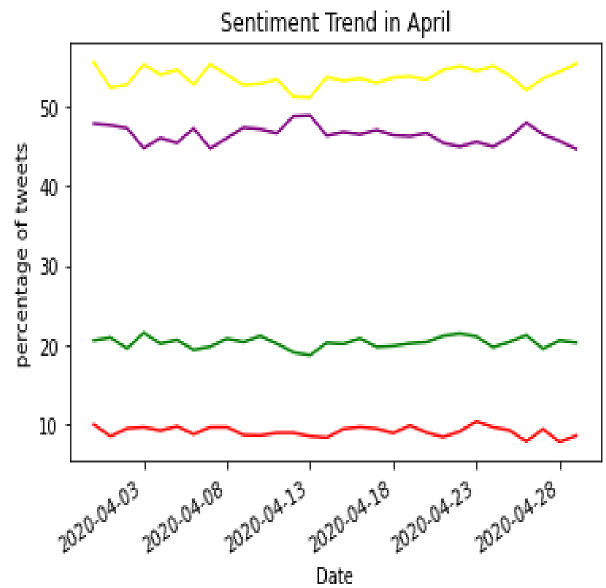


Fig. 2: Sentiment Trend in April

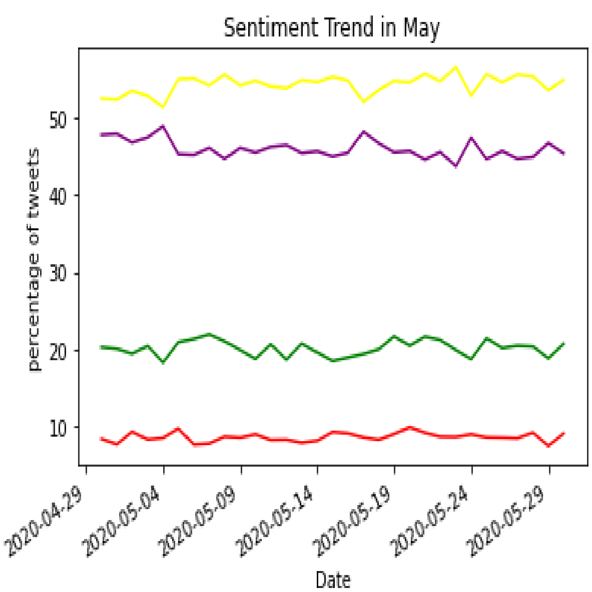


Fig. 3: Sentiment Trend in May

D. Average Sentiment association of tweets

With Bayesian classification, 20.3 percent tweets in April. 20.2 percent tweets in May and 20 percent tweets in June showed positive sentiments. While 10 percent tweets in April,8.7 percent tweets in May and 9.33 percent tweets in June showed negative sentiment which is tabulated in table V.

In April and May subjectivity value was at 54 percent and In June it was 55 percent. It shows that people were mostly sharing their opinions in the tweets. Continuous low values

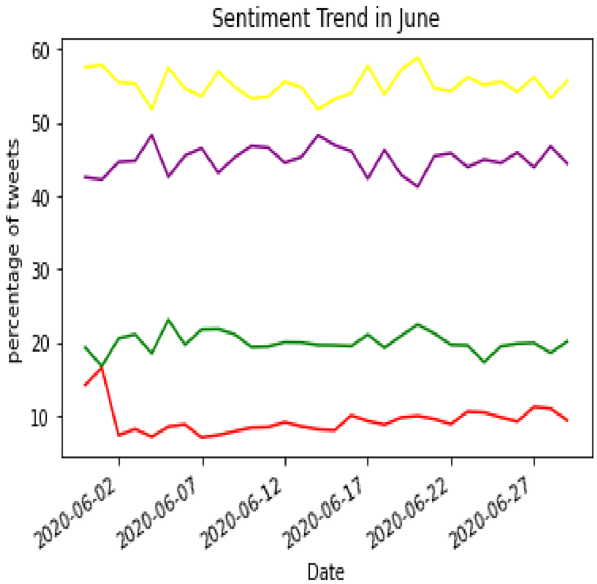


Fig. 4: Sentiment Trend in June

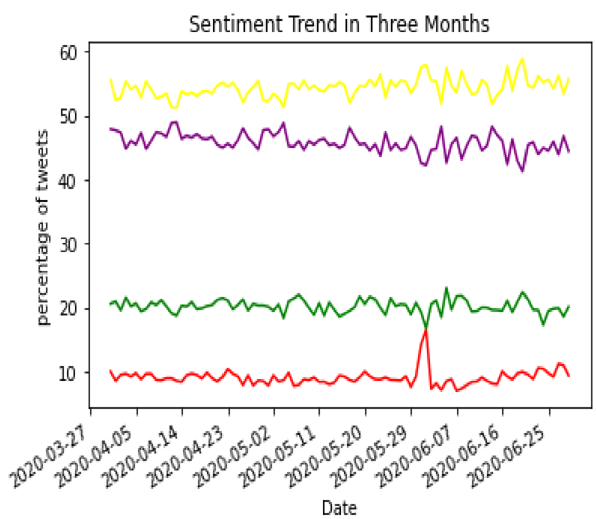


Fig. 5: Sentiment Trend in whole duration

of objectivity also justify this.Percentage of average subjectivity and objectivity of tweets in different months have been mentioned in Table VI.

Sentiment	April	May	June	Average
Positive	20.3	20.2	20	20.1
Negative	10	8.7	9.33	9.11

TABLE V: Average percentage of tweets for different sentiment

	April	May	June	Average
Subjectivity	54	54	55	54.3
Objectivity	46	46	45	45.7

TABLE VI: Average subjectivity and objectivity of tweets in different months

Colour	Sentiment
Red	Negative(Lexical)
Green	Positive(Lexical)
Purple	Negative(Bayesian)
Yellow	Positive (Bayesian)

TABLE VII: Legend for graph in figure 6

E. Comparison of Trends in sentiment association of tweets (Bayesian Versus Lexical)

We can see that in Lexical Method, more than 69 percent tweets have shown association towards positive sentiment whereas In Bayesian classification, nearly 20 percent tweets were associated to positive sentiment. On the other side, Nearly 65 percent tweets have shown association to negative sentiment in lexical approach and 9 percent tweets have shown negative sentiments in Bayesian classification.So both the approaches confirm that more tweets have shown positive sentiments. Figure 6 shows the graph for comparison of trend of emotions in whole duration of observation in lexical and Bayesian approach. Table VII shows the association of positive and negative sentiment in both bayesian and lexical approach to colour of lines in the graph in figure 6. Legends couldn't be associated to graph directly due to size and complexity in visual representation.

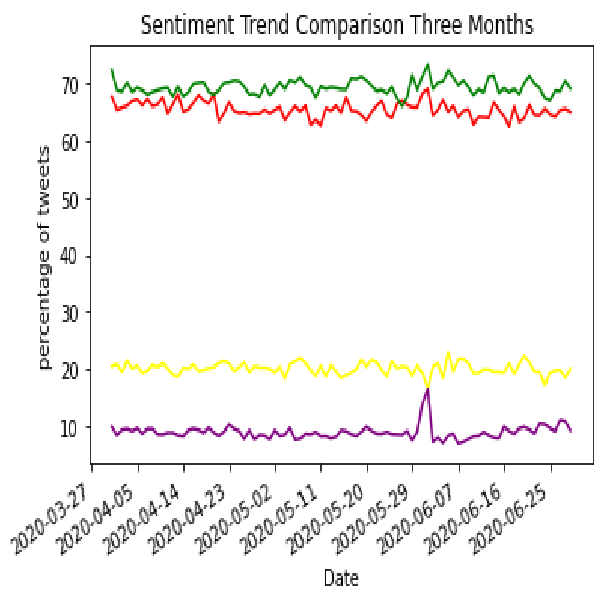


Fig. 6: Comparison of sentiment trends in whole duration

Sentiment	April	May	June	Average
Positive(Bayesian)	20.3	20	20.2	20.1
Negative(Bayesian)	10	8.7	9.33	9.11
Positive(Lexical)	69.1	69.2	69.6	69.3
Negative(Lexical)	66	65.2	65	65.4

TABLE VIII: Comparison of results in bayesian and lexical approach

F. Comparison of Average sentiment association of tweets (Bayesian Versus Lexical)

On an average ,more tweets have shown association to positive sentiments in both the approaches. In lexical approach positive as well as negative tweets have higher percentage because each tweet is assigned at least to an emotion. In Bayesian sentiment analysis, most of the tweets are classified as neutral because of the assumption of features for both the classes.This can be interpreted directly from table VIII.

V. Conclusion

People were in fear and sadness throughout the three months but they were also showing trust on authorities in their tweets which can be observed from the percentage of tweets associated to different emotions. People were also trying to predict upcoming situations which is evident from the percentage of tweets associated with anticipation.Despite having fear and sadness, people showed high amount of trust and Overall it can be concluded that people remained positive throughout the three months. Low performance of Bayesian learning method in classifying tweets may be due to it's assumption of independence between the features. In [32], Sentiment analysis of tweet data on COVID-19 for one week has been analysed using TextBlob .In addition to the previous work in [32],We have separately performed sentiment analysis of tweets with TextBlob and NRC emotion lexicons for twitter data of three months. Both the results were consistent to each other. In [31],unigram, Bigram and other frequent term set has been calculated using fp-growth algorithm. For our dataset, In [1], these things were provided with the dataset itself.Hence This paper only focuses on sentiment analysis of the tweets.

VI. Future Work

This paper's scope has been restricted to sentiment analysis of the data and hence we have used pre trained sentiment anlaysis models for the same. More sophisticated algorithms like SVM and LSTM can be trained to classify the tweets more precisely.Moreover granularity of time can be decreased or increased to observe the data more carefully. It would be interesting to observe how sentiment of tweets change within a day. More languages can be included in the study for better understanding of the sentiment of people. A better observation can be made if tweets are first classified based on geo-location of the person who tweeted the text. It will give a clear understanding of how people in different regions reacted to the pandemic.

References

[1] Juan M. Banda, Ramya Tekumalla, Guanyu Wang, Jingyuan Yu, Tuo Liu, Yuning Ding, Gerardo Chowell. 2020, June 21. ”A large-scale COVID-19 Twitter chatter dataset for open scientific research - an international collaboration.” Zenodo. http://doi.org/10.5281/zenodo.3902855.

[2] Alec Go, Richa Bhayani, and Lei Huang. 2009. ”Twitter sentiment classification using distant supervision”.Technical report, Stanford.

[3] M. S. Neethu and R. Rajasree, ”Sentiment analysis in twitter using machine learning techniques,” 2013 Fourth International Conference on Computing, Communications and Networking Technologies (ICCCNT), Tiruchengode, 2013, pp. 1-5, doi: 10.1109/ICCCNT.2013.6726818.

[4] R. K. Bakshi, N. Kaur, R. Kaur and G. Kaur, ”Opinion mining and sentiment analysis,” 2016 3rd International Conference on Computing for Sustainable Global Development (INDIACom), New Delhi, 2016, pp. 452-455.

[5] Z. Jianqiang, G. Xiaolin and Z. Xuejun, ”Deep Convolution Neural Networks for Twitter Sentiment Analysis,” in IEEE Access, vol. 6, pp. 23253-23260, 2018, doi: 10.1109/ACCESS.2017.2776930.

[6] R. Wagh and P. Punde, ”Survey on Sentiment Analysis using Twitter Dataset,” 2018 Second International Conference on Electronics, Communication and Aerospace Technology (ICECA), Coimbatore, 2018, pp. 208-211, doi: 10.1109/ICECA.2018.8474783.

[7] V. S. Pagolu, K. N. Reddy, G. Panda and B. Majhi, ”Sentiment analysis of Twitter data for predicting stock market movements,” 2016 International Conference on Signal Processing, Communication, Power and Embedded System (SCOPES), Paralakhemundi, 2016, pp. 1345-1350, doi: 10.1109/SCOPES.2016.7955659.

[8] G. Gautam and D. Yadav, ”Sentiment analysis of twitter data using machine learning approaches and semantic analysis,” 2014 Seventh International Conference on Contemporary Computing (IC3), Noida, 2014, pp. 437-442, doi: 10.1109/IC3.2014.6897213.

[9] M. Hao et al., ”Visual sentiment analysis on twitter data streams,” 2011 IEEE Conference on Visual Analytics Science and Technology (VAST), Providence, RI, 2011, pp. 277-278, doi: 10.1109/VAST.2011.6102472.

[10] B. Gokulakrishnan, P. Priyanthan, T. Ragavan, N. Prasath and A. Perera, ”Opinion mining and sentiment analysis on a Twitter data stream,” International Conference on Advances in ICT for Emerging Regions (ICTer2012), Colombo, 2012, pp. 182-188, doi: 10.1109/ICTer.2012.6423033.

[11] S. Bahrainian and A. Dengel, ”Sentiment Analysis and Summarization of Twitter Data,” 2013 IEEE 16th International Conference on Computational Science and Engineering, Sydney, NSW, 2013, pp. 227-234, doi: 10.1109/CSE.2013.44.

[12] P. Mishra, R. Rajnish and P. Kumar, ”Sentiment analysis of Twitter data: Case study on digital India,” 2016 International Conference on Information Technology (InCITe) - The Next Generation IT Summit on the Theme - Internet of Things: Connect your Worlds, Noida, 2016, pp. 148-153, doi: 10.1109/INCITE.2016.7857607.

[13] Y. Arslan, A. Birturk, B. Djumabaev and D. Küçük, ”Real-time Lexicon-based sentiment analysis experiments on Twitter with a mild (more information, less data) approach,” 2017 IEEE International Conference on Big Data (Big Data), Boston, MA, 2017, pp. 1892-1897, doi: 10.1109/BigData.2017.8258134.

[14] M. E. Basiri and A. Kabiri, ”Translation is not enough: Comparing Lexicon-based methods for sentiment analysis in Persian,” 2017 International Symposium on Computer Science and Software Engineering Conference (CSSE), Shiraz, 2017, pp. 36-41, doi: 10.1109/CSICSSE.2017.8320114.

[15] K. Mao, J. Niu, X. Wang, L. Wang and M. Qiu, ”Cross-Domain Sentiment Analysis of Product Reviews by Combining Lexicon-Based and Learn-Based Techniques,” 2015 IEEE 17th International Conference on High Performance Computing and Communications, 2015 IEEE 7th International Symposium on Cyberspace Safety and Security, and 2015 IEEE 12th International Conference on Embedded Software and Systems, New York, NY, 2015, pp. 351-356, doi: 10.1109/HPCC-CSS-ICESS.2015.75.

[16] N. A. Abdulla, N. A. Ahmed, M. A. Shehab and M. Al-Ayyoub, ”Arabic sentiment analysis: Lexicon-based and corpus-based,” 2013 IEEE Jordan Conference on Applied Electrical Engineering and Computing Technologies (AEECT), Amman, 2013, pp. 1-6, doi: 10.1109/AEECT.2013.6716448.

[17] E. W. Pamungkas and D. G. P. Putri, ”An experimental study of lexicon-based sentiment analysis on Bahasa Indonesia,” 2016 6th International Annual Engineering Seminar (InAES), Yogyakarta, 2016, pp. 28-31, doi: 10.1109/INAES.2016.7821901.

[18] Bo Pang,Lillian Lee, and Shivakumar Vaithyanathan. 2002.”Thumbs up? Sentiment Classification using Machine Learning Techniques.”

[19] Priyanka Thakur, Dr. Rajiv Srivastava.2019. ”Sentiment Analysis of tourist review using supervised long short term memory deep learning approach.”

[20] Shashank Gupta.2018,January 7.”Sentiment Analysis: Concept, Analysis and Applications.” TowardsDataScience. https://towardsdatascience.com/sentiment-analysis-concept-analysis-and-applications-6c94d6f58c17

[21] Rohith Gandhi.2018,May 5.”Naive Bayes Classifier.”TowardsDataScience.https://towardsdatascience.com/naive-bayes-classifier-81d512f50a7c

[22] Adam Hayes.2020,June 5.Bayes Theorem. Investopedia. https://www.investopedia.com/terms/b/bayes-theorem.asp

[23] Rohith Gandhi.2018,June 7.Support Vector Machine — Introduction to Machine Learning Algorithms.TowardsDataScience. https://towardsdatascience.com/support-vector-machine-introduction-to-machine-learning-algorithms-934a444fca47

[24] Christopher Olah.2015,August 27. Understanding LSTM Networks. Github.io. https://colah.github.io/posts/2015-08-Understanding-LSTMs/

[25] Vasilis Vryniotis.2013,November 20.Machine learning Tutorial: The Max Entropy Classifer. DatumBox. https://blog.datumbox.com/machine-learning-tutorial-the-max-entropy-text-classifier/

[26] Maximum Entropy Principle. StatisticsHowTo. https://www.statisticshowto.com/maximum-entropy-principle/

[27] Dhiraj Gurkhe,Niraj Pal, Rishit Bhatia.2014. ”Effective sentiment analysis of social media datasets using naive bayesian classification.”

[28] Prachi Sanghvi, Disha Shah, Prof. Bharathi H. N. 2019. ”A study on sentiment analysis techniques of twitter data.” IJRASET.

[29] Prabu Palanisamy, Vineet Yadav and Harsha Elchuri.2013.”Serendio: Simple and Practical lexicon based approach to Sentiment Analysis.”

[30] Florian Wogenstein,Johannes Drescher,Dirk Reinel,Sven Rill,Jörg Scheidt.”Evaluation of an Algorithm for Aspect-Based Opinion Mining Using a Lexicon-Based Approach.”

[31] Habiba Drias, Yassine Drias.May 18,2020. ”Mining Twitter Data on COVID-19 for Sentiment analysis and frequent patterns Discovery.” medRxiv.

[32] Kamaran H. Manguri,Rebaz N. Ramadhan,Pshko R. Mohammed Amin. May 15,2020.”Twitter Sentiment Analysis on Worldwide COVID-19 Outbreaks.”KJAR.

[33] Frequent Pattern (FP) Growth Algorithm In Data Mining.June 30,2020.Software Testing Help. https://www.softwaretestinghelp.com/fp-growth-algorithm-data-mining/

[34] Susan Li.2019,January 10.Introducing TextBlob.TowardsDataScience. https://towardsdatascience.com/having-fun-with-textblob-7e9eed783d3f

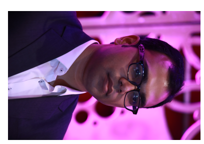
[35] Richard Chadwick.2019,July 1.Tweepy for beginners.TowardsDataScience. https://towardsdatascience.com/tweepy-for-beginners-24baf21f2c25

[36] Saif M. Mohammad and Peter D. Turney.August 28,2013. ”Crowdsourcing a Word–Emotion Association Lexicon.” arXiv.

AUTHORS’ PROFILE

Aditya Raj is currently pursuing Bachelor of Technology and MBA (dual degree) in Information Technology from Indian Institute of Information Technology Allahabad,Prayagraj(Declared as an Institute of National importance by Parliament of India).

Dr Ranjit Singh is Associate Professor in the Department of Management Studies, Indian Institute of Information Technology Allahabad, Prayagraj, India. He has more than 17 years of Post-Graduate Teaching experience. His specialization is in the area of Accounting, Finance and Entrepreneurship. He has authored more than 110 numbers of research papers and 16 books till date. He has already completed three funded projects. Seven research scholars have completed their PhD under his guidance.



Ayush Sinha is a research scholar in department of Information Technology, Indian Institute of Information Technology Allahabad, Prayagraj, India. He has close to 9 years of Industry Research and Development Teaching experience in software development. His research interest is in deep learning and cyber physical systems specially energy infrastructures.

Pro. O.P. Vyas is an Alumnus with Technical University of Kaiserslautern (Germany).His PhD in ‘Technomathemtik’ from TU Kaiserslautern as DAAD Fellow (1995-97) jointly with TU Kaiserslautern IIT-Kharagpur and Master of Technology (Computer Sc.) from Indian Institute of Technology-Kharagpur.He Worked at ‘Fraunhofer Institute for Software Engineering.’ (Kaiserslautern - Germany) and Currently working as Professor (IT) at IIIT Allahabad (India).He has completed many Collaborative projects with Germany, France, Italy and Japan in Data Science, Machine Learning IoT. He has Guided 16 Ph.D. scholars, 25 Master Theses, B.Tech. Students’ theses and more than 160 publications and 04 Books.