



Literature Review on Sentiment Analysis and Opinion Classifications on the impact of COVID19 Outbreak

¹Kirti Bala Bahekar, ²Dr.Pratima Gautam, ³Dr.Shailja Sharma

¹Research Scholar, ²Dean, ³Professor

¹Department of Computer Science, Rabindranath Tagore University, Bhopal, India

Abstract : The world is suffering from a very destructive pandemic caused by a novel coronavirus named COVID19. Countries imposed lockdowns, various guidelines, and measures to prevent the spread of COVID19 among the citizens. Due to lockdown and strict guidelines for the public, a large number of people took the support of Social Media to express their opinions, thoughts, suffering's and emotions. Twitter is a microblogging site, a platform for the public to write views in text format. Several researchers conducted studies on public sentiments during the pandemic period. This paper presents a review of some of the studies conducted, on community opinions during the pandemic, based on tweets. Further, we analysed various researches on the prediction of tweets using a variety of machine learning algorithms. Moreover, the analysis also focuses on datasets and performance analysis measures are recognized. The research gap is also identified where more upcoming researches can concentrate.

IndexTerms - COVID19, Machine Learning, Sentiment analysis, Performance analysis.

I. INTRODUCTION

The Covid19 outbreak cause a worldwide public health crisis, which disturbed mental, financial, and physical situations. On March 11, 2020, the World Health Organisation declared it as a global pandemic, till now approx. 53.5 lakh deaths are recorded, worldwide due to this pandemic. This harmful coronavirus travels in tiny water droplets, transmit from human to human. Therefore, the World health organizations (WHO) advised covid19 appropriate behaviour which includes wearing masks at public places, washing hands, proper sanitization, etc. to prevent the spreading of disease.

Many countries imposed lockdowns to prevent Covid19 and suggest their citizens stay home and be safe. So during the period public has undergone various emotional ups and downs, which they freely shared on social media as the only medium of communication during lockdowns. They shared their suggestions, feelings, reactions, and views on several government policies. Twitter is a wonderful microblogging site that helps in sharing opinions in form of text etc. In the course of the pandemic, many studies are performed to analyse the public sentiments on various topics like awareness on washing hands, wearing a mask, public views on leaderships of the countries, satisfactory health facilities, etc. In this proposed paper, we have done a review of several papers related to the public sentiment analysis in the times of the COVID19 pandemic.

II. REQUIREMENT OF SENTIMENT ANALYSIS

Sentiment analysis or NLP is the process of identifying the public views on any event, occasion, or policy. Such analysis guides authorities to take future acts and decision regards the wellbeing of the society. Natural language processing and text mining are the computerized methods to extract and analyse the reviews posted online or newspaper articles or headlines etc. The task of analysis starts from cleaning data known as Data pre-processing, which includes Tokenisation, removing symbols and stop words, stemming, and Normalization[1].

Sentiment classification techniques can be segregated into three categories (Fig. 1.).These are machine learning, lexicon-based and hybrid approaches [2]. The first of these techniques involve popular machine learning (ML) algorithms and involves using linguistic features. The second involves analyses through a collection of sentiment terms that are precompiled into the sentiment lexicon. This is further divided into the dictionary- and corpus-based approaches that use semantic or statistical methods to gauge the extent of the polarity of sentiment. The hybrid approach involves combining ML and lexicon-based approaches. Machine learning is further divided into three categories Supervised Learning, Unsupervised Learning, and Reinforcement Learning. Supervised learning methods include Decision Tree, Linear Classification, Rule-based classification, and probabilistic classification.

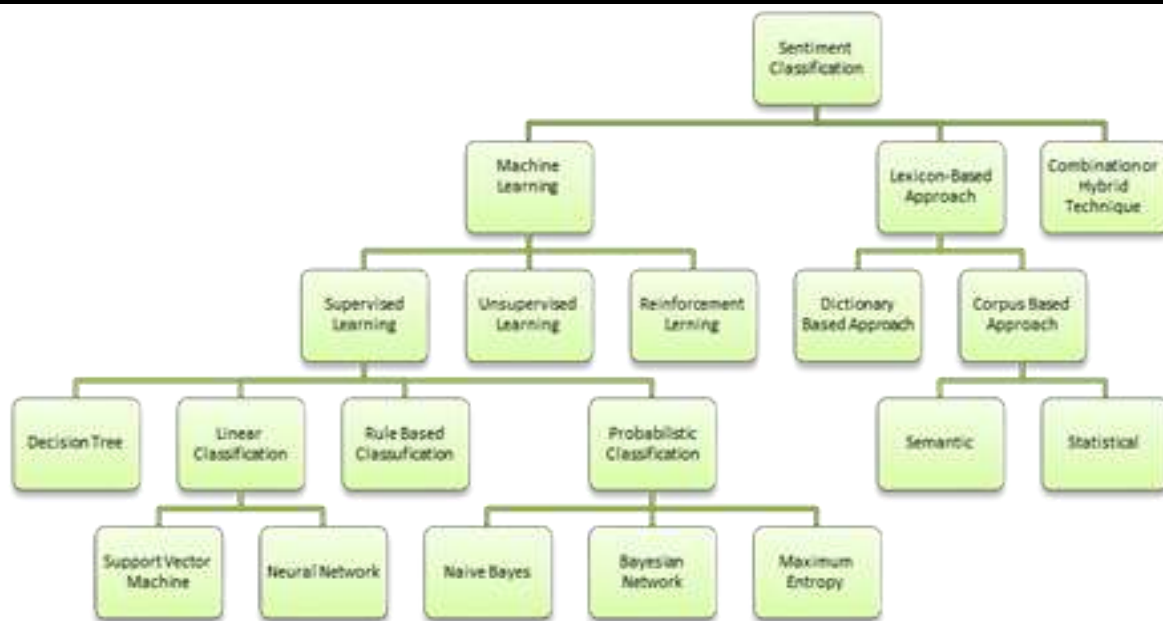


Fig. 1 Sentiment Classification Techniques

III. LITERATURE REVIEW

A. STATE OF ART: CLASSIFICATION ON COVID19 RELATED TWEETS

In 2021, M. Singh et al.[3] have aimed to do sentiment analysis to know the eagerness and curiosity of the people on novel disease due to newly identified coronavirus. They have tested accuracy on two data sets firstly tweets collected from the whole world and in the second dataset tweets from India. Bi-directional Encoding Representation for a Transformer (BERT) model is used for emotional classification. Sentiments are divided into Negative as weakly, strongly and mild, Neutral, Positive as weakly, strongly and mild. Accuracy of 93.89% is noted in the BERT classification.

In 2021, N. Chintalapudi et al.[4] Collected tweets including various COVID19 related hashtags and BERT classifier is used to analyze tweets from Indian citizens, these are labelled as fear, sadness, anger, and joy. A deep learning model BERT is used for text analysis and results are compared with other models as Logistic regression (LR), support vector machine (SVM), and Long-short term memory (LSTM). BERT has displayed 89% accuracy whereas other algorithms produced 75%, 74.75%, and 65% respectively.

In 2020, M. Mansoor et al.[5] have studied the behaviour of the public on the coronavirus and how the sentiment changed over time, they had collected dataset on three formats, firstly tweets related to coronavirus, secondly sentiments related to work from home and lastly collected data related to online education. For predicting further opinions they have applied machine learning algorithms like long short term memory (LSTM), and Artificial neural network (ANN), and accuracies are compared. On coronavirus tweets, the accuracy of 84.5% and 76% are noted in LSTM and ANN algorithms respectively. General positive tweets are in the case of work from home and tweets related to online learning.

In 2020, L. Chen et al.[6] stated that various controversial terms emerged during this pandemic; such terms are used by the public and politicians to express their distress. Terms like "Chinese virus", "#Chinese virus is controversial terms, and terms like "coronavirus", "COVID19" are non-controversial terms. Tweets regarding these heads are collected and classified for prediction using Bi-LSTM, BERT, XLNet, performance of these algorithms is evaluated which gives an F score of 0.6723, 0.8732, and 0.9499 respectively.

In 2020, Sijia Li et al.[7] collected posts from the Weibo online platform in China to study public sentiments on coronavirus and calculated word frequency, the score of emotions indicating thoughts of anxiety, depression, anger, happiness, and cognitive indicators are also evaluated like social risk judgment and life satisfaction. The T-test is applied to study the changes before and after the declaration of COVID19 on 20 January 2020.

In 2020, Imran et al. [8] evaluated cross-cultural polarity and emotions on novel coronavirus various Deep learning models are used to evaluate the sentiments in different states like positive, negative, and neutral.

In 2020, S. Dutta et al.[9] used Machine Learning approach for COVID19 cases study algorithms like Long short-term memory (LSTM), and Gated Recurrent Unit (GRU), the results have obtained an accuracy of 87% for "Confirmed cases", 67.8% for "Negative cases", 62% for "Deceased Case", and 40.5% for "Release case". Root mean square error is also calculated 30.15% for confirmed, 49.4% for negative, 4.16% for deceased, and 13.72% for released cases.

In 2020, Iyer et al.[10] analysed the reaction of people on the pandemic outbreak, identified how people are aware of the disease, and its symptom, what precautionary measures are taken by them, and whether people are following guidelines are not. Naive Bayes classifier is used to evaluate accuracy after tweets collection and pre-processing tweet cleaning, build tweet dictionary, find word density, and feature extraction is done.

In 2020, Jim Samuel et al.[11] identified public sentiment on pandemic using two machine learning algorithms for analysing textual-based content, which presents 91% accuracy with Naïve Bayes and 74% accuracy with Logistic Regression.

TABLE 1. PERFORMANCE MEASURES ARE CONCERNED WITH TWITTER SENTIMENT ANALYSIS UNDER DIFFERENT RESEARCHES.

Ref no.	Models	Measures of Evaluation					
		Accuracy	Precision	Recall	Mcc	RMSE	F-measure
3	BERT(Bi-directional Encoding Representation for a Transformer)	√			√		
4	BERT,LR,SVM,LSTM	√					
5	LSTM,ANN	√					
6	BERT, XLNET(Generalised autoregressive pertaining method), Bi-LSTM (Bidirectional Long Short Term Memory)						√
7	t-test						
8	DNN,LSTM+fastText,LSTM+GloVe,LSTM+GloVe twitter,BERT	√					√
9	LSTM, GRU (Gated Recurrent Unit), Combined LSTM-GRU model	√				√	
10	NB(Naïve Bayes)	√					
11	Logistic Regression, Naïve Bayes	√					
12	MultinomialNB, BernoulliNB,LR,LinearSVC,AdaBoostclassifier,Ridgeclassifier,PassiveAggressiveClassifier,Perceptron(For unigram, Bigram, and Trigram)	√					
13	SVM(Support Vector Machine), RF(Random Forest), LR(Logistic Regression), DT(Decision Tree), and KNN(K-nearest neighbor)	√	√	√			√
14	DT(Decision Tree), SVM(Support Vector Machine), RF(Random Forest), and KNN(K-nearest neighbor)	√	√	√			√

In 2020 P. Gupta et al. [12] analysis tweets from India under the hashtag Lockdown classified the tweets with numerous different classifiers like MinomialNB, Logistic Regression (LR), Perception, etc. They have also generated accuracy curves for frequent unigrams, bigrams, and trigrams in the tweets and calculated accuracy for each n-gram.

In 2020, X. Zhang et al. [13] collected tweets on #COVID19 and #coronavirus and did feature extraction methods for textual data analysis like n-grams and TF-ID. Five regular machine learning algorithms are used for a comparative study like SVM, RF, LR, DT, and SVM is used. Presents the results that Random Forest algorithm produces good performance using Unigram feature.

In 2021, V. Kandasamy et al.[14] have used many supervised learning algorithms for prediction like DT, SVM, RF, and KNN and also extracted features like n-grams. Table no.1 shows details of performance measures used in various studies and criteria used for comparison.

Table 2 shows various data types used by different researchers for sentiment analysis related to COVID19 and its concerns in the public domain. Most of the researchers collected tweets including, #Covid19,#coronavirus, #lockdown, tweets from India as well as from the world is collected and analysed.

TABLE 2. DATASETS USED BY DIFFERENT CONTRIBUTIONS FOR SENTIMENT ANALYSIS ON COVID19

Author	Datasets
M. Singh et al.[3]	Impact of coronavirus on social life ,tweets on #Covid19,#Covid19+coronavirus 1)World2)India
N. Chintalapudi et al.[4]	COVID19 tweets from github(including tweets covid19,coronavirus and lockdown)from India
M. Mansoor et al.[5]	Three tweets datasets are used 1.Covid19 2. Work from Home 3. Online Learning (Worldwide, US, India, Brazil)
L. Chen et al.[6]	Terms used in Neutral & Controversial Terms for COVID19(Set of Controversial terms and Non-controversial terms)
Sijia Li et al.[7]	Impact of COVID19 on psychological consequences dataset Weibo post
Imran et al.[8]	Two data sets 1) Trending hashtags and 2) Kaggle dataset on COVID19 tweets
S. Dutta et al. [9]	Kaggle dataset on COVID19 tweets
Iyer et al.[10]	Twitter sentiment analysis on Coronavirus Outbreak in India
Jim Samuel et al.[11]	Tweets on keyword Corona are collected for public sentiment Insights and classification
P.Gupta et al.[12]	Tweets collected on "Indialockdown"
X.Zhang et al.[13]	Tweets are filtered by #COVID-19 and #coronavirus
V.Kandasamy et al.[14]	Tweets collected by filtering them by COVID19, COVID, sarascov2, coronavirus and covid_19 on IEEE data source

B. Researches on tweets analysis

Table 3 shows some research works used to evaluate sentiments and tweets are grouped in different categories. In these papers, sentiment classes are identified by different models.

In 2020, M.Hung et al.[16] applied machine learning method to analyze data collected from tweets from the United States for Covid19 related discussion, the study is leading on topics related to health care environment emotional support, business economy, social change, and psychological stress.

In 2021, J.Lang et al.[17]performed a study to examine types, themes, temporal, trends, and exchange patterns on mask related hashtags and also analyse temporal related to mask on Twitter. Tweets are further classified into pro-mask and anti-mask groups.

In 2021, M.Kausar et al.[18]tweets are collected from Covid19 infected 11 countries including gulf country Oman, having hashtags COVID19, coronavirus, etc. Research is used to analyze how people cope with the situation during pandemics.

TABLE 3. RESEARCH ON TWEETS TRENDS ON COVID19

Authors	Dataset Used	Categories of tweets
M.Hung et al.[16]	Tweets related to health care environment emotional support, business economy, social change, and psychological stress.	Positive, Negative, and Neutral
J.Lang et al.[17]	Pro-mask and Anti-Mask Hashtags	
M. Kausar et al.[18]	Emotion analysis of various countries (USA, UK, Brazil, India, Russia, South Africa, Peru, Mexico, Chile, Spain, Oman)	Anger, Anticipation, Disgust, Fear, Joy , Sadness, Surprise, Trust
S. Boon-Itt et al.[19]	Public perception on COVID19 pandemic on Twitter (coronavirus,covid_19,2019 n-COV,Covid19)	Evaluated public perception in form of three stages during the pandemic period

In 2020, Sakun Boon-Itt et al.[19]Public perception on COVID19 pandemic on Twitter in analysed ,tweets on coronavirus,covid_19,2019 n-COV,Covid19 are collected. Display the sentiments in several categories like anger, anticipation, disgust, fear, joy, sadness, surprise, and trust.

IV. CONCLUSION

Sentiment analysis is considered to be critical in terms of socio-economic standpoints. During the time of the pandemic, a not large number of researchers is present to study public opinion. This paper showcases a few types of research in the field, which might be useful in understanding the perception of citizens to overcome the situation. The present review explores machine learning implementation employed in sentiment analysis. Initially, the assessment is done on clarifying the contribution of every task and the type of machine learning methods operated. The assessment also focused on the data type and outcomes of the study. This review may helpful for considering public emotions in the time of the pandemic, where the third wave in India is approaching, due to the new variant Omicron of coronavirus. The results may help authorities to take decisions in the public interest and to save lives.

APPENDIX

NLP: Natural Language Processing helps developers to organize knowledge for performing tasks such as translation, automatic summarization, Named Entity Recognition (NER), speech recognition, etc.

RF: Random Forest: The "forest" it builds, is an ensemble of decision trees, usually trained with the "bagging" method.

DT: Decision Tree is a supervised learning technique that can be used for both classification and Regression problems, but mostly it is preferred for solving Classification problems.

ANN: Artificial Neural Networks are computational models and inspire by the human brain. Many of the recent advancements have been made in the field of Artificial Intelligence, including Voice Recognition, Image Recognition, and Robotics using Artificial Neural Networks.

SVM: Support Vector Machine is one of the most popular Supervised Learning algorithms, which is used for Classification as well as Regression problems.

NB: Naive Bayes Naïve Bayes algorithm is a supervised learning algorithm, which is based on Bayes theorem and used for solving classification problems.

ML: Machine Learning is a part of artificial intelligence and growing technology that enables machines to learn from past data and perform a given task automatically.

KNN: k-Nearest Neighbor. It is the simplest machine learning algorithm.

BERT: A Deep learning algorithm related to NLP.

LSTM: Long Short Term memory is a deep learning algorithm having recurrent neural network architecture.

REFERENCES

- [1] G. D. Devi and S. Kamalakkannan, "Literature Review on Sentiment Analysis in Social Media : Open Challenges toward Applications Literature Review on Sentiment Analysis in Social Media : Open Challenges toward Applications," no. January, 2020.
- [2] A. Baldominos Gómez, N. Luis Minguenza, and M. C. García del Pozo, "OpinAIS: An Artificial Immune System-based Framework for Opinion Mining," *Int. J. Interact. Multimed. Artif. Intell.*, vol. 3, no. 3, p. 25, 2015, doi: 10.9781/ijimai.2015.333.
- [3] M. Singh, A. K. Jakhar, and S. Pandey, "Sentiment analysis on the impact of coronavirus in social life using the BERT model," *Soc. Netw. Anal. Min.*, vol. 11, no. 1, pp. 1–11, 2021, doi: 10.1007/s13278-021-00737-z.
- [4] N. Chintalapudi, G. Battineni, and F. Amenta, "Sentimental analysis of COVID-19 tweets using deep learning models," *Infect. Dis. Rep.*, vol. 13, no. 2, pp. 329–339, 2021, doi: 10.3390/IDR13020032.
- [5] M. Mansoor, K. Gurumurthy, A. R. U, and V. R. B. Prasad, "Global Sentiment Analysis Of COVID-19 Tweets Over Time," no. October 2020, 2020, [Online]. Available: <http://arxiv.org/abs/2010.14234>.
- [6] L. Chen, H. Lyu, T. Yang, Y. Wang, and J. Luo, "In the Eyes of the Beholder: Analyzing Social Media Use of Neutral and Controversial Terms for COVID-19," 2020, [Online]. Available: <http://arxiv.org/abs/2004.10225>.
- [7] S. Li, Y. Wang, J. Xue, N. Zhao, and T. Zhu, "The impact of covid-19 epidemic declaration on psychological consequences: A study on active weibo users. *Int J Environ Res Public Health* [revista en Internet] 2020 [acceso 13 de abril de 2021]; 30(3): 201-205," *Int. J. Environ. Res. Public Health*, vol. 17, no. 6, 2020, [Online]. Available: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7143846/pdf/ijerph-17-02032.pdf>.
- [8] A. L. I. S. Imran, S. M. Daudpota, Z. Kastrati, and R. Batra, "Cross-Cultural Polarity and Emotion Detection Using Sentiment Analysis and Deep Learning on COVID-19 Related Tweets," vol. 8, 2020, doi: 10.1109/ACCESS.2020.3027350.
- [9] S. Dutta and S. Kumar Bandyopadhyay, "Machine Learning Approach for Confirmation of COVID-19 Cases: Positive, Negative, Death and Release," *medRxiv*, vol. 03, pp. 172–177, 2020, doi: 10.1101/2020.03.25.20043505.
- [10] K. B. Priya Iyer and S. Kumaresh, "Twitter sentiment analysis on coronavirus outbreak using machine learning algorithms," *Eur. J. Mol. Clin. Med.*, vol. 7, no. 3, pp. 2663–2676, 2020.
- [11] J. Samuel, G. G. M. N. Ali, M. M. Rahman, E. Esawi, and Y. Samuel, "COVID-19 public sentiment insights and machine learning for tweets classification," *Inf.*, vol. 11, no. 6, pp. 1–22, 2020, doi: 10.3390/info11060314.
- [12] P. Gupta, S. Kumar, R. R. Suman, and V. Kumar, "Sentiment Analysis of Lockdown in India during COVID-19: A Case Study on Twitter," *IEEE Trans. Comput. Soc. Syst.*, vol. 8, no. 4, pp. 939–949, 2021, doi: 10.1109/TCSS.2020.3042446.
- [13] X. Zhang, H. Saleh, E. M. G. Younis, R. Sahal, and A. A. Ali, "Predicting Coronavirus Pandemic in Real-Time Using Machine Learning and Big Data Streaming System," vol. 2020, 2020.

- [14] V. Kandasamy et al., “Sentimental Analysis of COVID-19 Related Messages in Social Networks by Involving an N-Gram Stacked Autoencoder Integrated in an Ensemble Learning Scheme,” 2021.
- [15] A. D. Dubey, “Decoding the Twitter Sentiments towards the Leadership in the times of COVID-19: A Case of USA and India,” SSRN Electron. J., no. April, pp. 19–22, 2020, doi: 10.2139/ssrn.3588623.
- [16] M. Hung et al., “Social network analysis of COVID-19 sentiments: Application of artificial intelligence,” J. Med. Internet Res., vol. 22, no. 8, pp. 1–13, 2020, doi: 10.2196/22590.
- [17] J. Lang, W. W. Erickson, and Z. Jing-Schmidt, “#MaskOn! #MaskOff! Digital polarization of mask-wearing in the United States during COVID-19,” PLoS One, vol. 16, no. 4 April, pp. 1–25, 2021, doi: 10.1371/journal.pone.0250817.
- [18] M. A. Kausar, A. Soosaimanickam, and M. Nasar, “Public Sentiment Analysis on Twitter Data during COVID-19 Outbreak,” Int. J. Adv. Comput. Sci. Appl., vol. 12, no. 2, pp. 415–422, 2021, doi: 10.14569/IJACSA.2021.0120252.
- [19] S. Boon-Itt and Y. Skunkan, “Public Perception of the COVID-19 Pandemic on Twitter: Sentiment Analysis and Topic Modeling Study,” doi: 10.2196/21978.

