Journal of Computer Science & Computational Mathematics, *Volume 12, Issue 3, September 2022*
DOI: 10.20967/jcscm.2022.03.001

Twitter sentiment analysis of lockdown/pandemic in India by using deep learning : case study of COVID-19

Pranita P. Inamdar [*], N. V. Dharwadkar

Department of Computer Science and Engineering, Rajarambapu Institute of Technology, Rajaramnagar, Islampur, India
[*]inamdarpranita@gmail.com

***Abstract*:** Internet users today have access to a tremendous amount of data, as well as a massive volume of data has been generated, thanks to the innovation and spread of online technologies. Everyone in the world uses social networks to discuss and expresses ideas on different subjects, take an interest in debates, and exchange news. Sentiment analysis is a technology which mines the opinions from different social media sites and categorize them in the form of positive, negative and neutral polarities. Sentiment analysis is effective for evaluating data in tweets when opinions are unstructured, variable, and either positive or negative, or neutral in some cases. This paper is a literature survey on Twitter sentiment analysis of Indialockdown/lockdown/pandemic keywords. This paper provides a detailed comparative overview of existing opinion mining methodologies, such as Machine learning, dictionary-based approaches, etc. Also are discussed here in common topics and usage of sentiment analysis on Twitter.

***Keywords*:** Lockdown, Coronavirus, COVID-19, opinion mining, sentiment analysis, machine learning, deep learning Introduction.

1. Introduction

Sentiment Analysis is a method for examining people's feelings, attitudes, and opinions about an item, association, or administration. Opinion investigation is a kind of statistical surveying that utilizes text examination, biometrics, normal language handling (NLP), and computational semantics to decide the condition of information.

It's normal as far as we're concerned to consider about what others think while deciding. Large numbers of us depended on loved ones for item or administration ideas or data before the Internet's appearance. The Internet facilitates our efforts to obtain public opinion. The volume of data generated by users produced every minute is humongous. It is not humanely possible to gather entire data and decipher the meaning conveyed through the same. Effectively, it becomes very much essential to bring automation in classification of evaluations. Such evaluations are labeled as neutral, positive or negative. An example could be part of an email, social media content description, or a blog article. Sentiment analysis is useful to dissect public opinion and extract the understanding of consumer experiences.

There is a huge growth in volume of stream data continuously generated through social media platforms like Facebook, Twitter, etc. Behavioral science is one of the highly suitable domains providing fresh Twitter content. Twitter is also historically known as a very potent source of data for research of psychology. It's very interesting to analyze how catastrophic events affects sentiments of people e.g. How people react to Coronavirus outbreak and in the subsequent lockdown [1, 2].

It's also observed that researchers have taken keen interest to analyze tweets when different COVID-19 vaccines were made available to public. There are interesting observations at the global, national, and state levels in such incidents. Sentiments are extracted using latent Dirichlet allocation analysis. Temporal analysis is used to look at patterns across time. Geographic analysis, as the name suggests, localities. It deals with contents generated in the form of tweets from different residential areas [3].

Also, Twitter data has been used to investigate human psychology under catastrophic events like COVID-19 pandemic. With the COVID-19, there is huge impact on economy, jobs, social transformations which, to some extent, has resulted into depression and psychological disorders [4]. As psychology of people is reflected in the content generated, feelings of people about lockdown / social distancing enforced by states are a potential area of research [5].

Authors studied multiple data sources in Massachusetts [6] and investigated alterations in polarization of tweets, opined anxiety. It is also composed of discussion on health concerns. This was executed on subset of all tweets chosen stochastically. It can be done using symbolic as well as ML techniques. The prior is time consuming and tedious, scientists' resort to use ML [7].

To elicit feedback from fellow countrymen, Indonesian government has given a lot of emphasis too social network. Analysis of Twitter contents for coronavirus vaccines has proved very effective in their case. Naïve Bayes Algorithm was employed for the exercise [8]. SVM and CRF techniques have been successfully employed by researchers to classify feelings at the sentence level considering emoticons [9] and opinion mining using recommender system.

Extraction of salient topics, themes, and attitudes is done by authors [10] using LDA. They worked upon prominent unigrams and bigrams. Researchers of Typhoon Sandy found

out how client opinions vary reliant upon distance from the catastrophe and concluded that as passionate uniqueness rises, the chance of retweeting a tweet diminishes [11].

1.1 Types of sentiment analysis

Fine-grained:

This type of sentiment analysis version leads researcher to precision of polarity. Outcome of the polarity can be classified as extraordinarily high quality, superb, impartial, terrible, or very negative. When it comes to reviews and scores, fine-grained sentiment evaluation is beneficial.

Emotion Detection:

Emotion detection allows you locate feelings. This will encompass anger, sadistic behavior, exuberance, dissatisfaction, uneasiness, dismay, etc. Emotion detection exercises usually employ lexicons. It is nothing but a group of phrases that emote specific feelings.

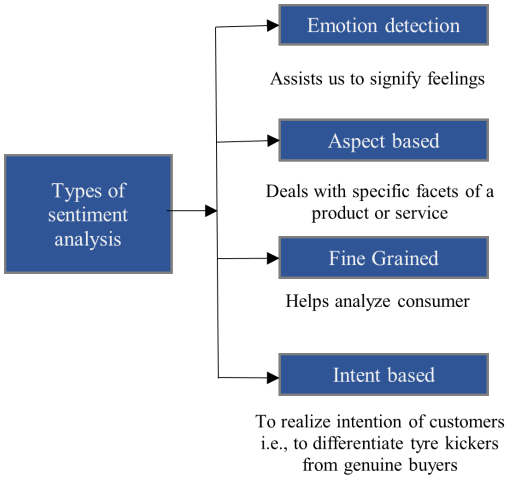

Figure 1. Types of sentiment analysis

Aspect-based:

It facilitates you to determine the factors of which one is speaking. Let's say; you're a mobile cellphone manufacturer, and you get a client evaluate declaring, "the digital camera struggles in synthetic lights situations." Tentatively, one can gauge that the customer is not happy about a specific camera functionality.

Intent analysis:

Many a times consumers are doing window shopping. They may be interested in a product but may not afford it immediately. Sometimes, they may be evaluating it or comparing with some other product. Such consumers are popularly known as tyre kickers. They do not contribute to the bottom line. Identifying such customers ahead of time precisely can help organizations to invest economically w.r.t. strategy, finance. It can also save physical efforts and time to market. Chasing such customers is often a waste of resources. Intent analysis can prove handy to resolve this challenge. It helps us identify whether customer is genuine or a time waster.

1.2 Applications of sentiment analysis

Applications of sentiment analysis has broad spectrum of domains. Politics, product reviews, movie reviews, social media posts, and other topics can all benefit from sentiment analysis. The below Table 1 shows the form of application of sentiment analysis.

Applications	Different rating modes
Review of a movie	★★★★★
Rating a product	
Exit polls for political elections	
Public opinion poll	
Social networking connections	

Figure 2. Applications of Sentiment analysis

1.3 Sentiment classification

Machine Learning

This method uses a ML algorithm and a miscellaneous feature to create a classifier. Such a classifier can recognize sentimental text. Deep-learning approaches are popular nowadays because they are compatible with data learning representations.

Lexicon-Based

To determine the general assessment score of a particular material, this method employs a variety of terms marked by polarity score. This technique's main feature is that it doesn't require any training data, but its biggest flaw is that lexicons do not consider many expressions. A lot of words are also often missed.

Hybrid

Hybrid is a Sentiment Analysis technique that combines machine learning and lexicon-based techniques. Even though it is not widely used, this strategy usually yields better outcomes than the approaches outlined above.

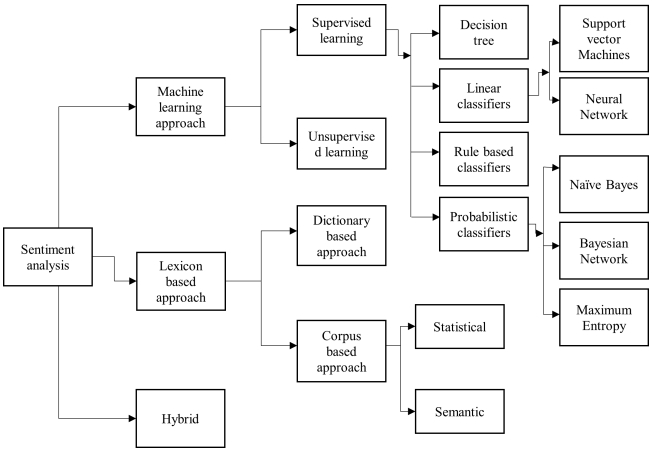

Figure 3. Classification of Sentiment analysis

2. Literature survey

In [1] the author of this study sought to understand Indian citizens' reactions to the Indian government's nationwide lockdown. The lockdown was applied to control outbreak of

corona virus. In this study, tweets written by Indian citizens were analyzed by NLP and ML classifiers. The dataset was gathered during 5th April, 2020 to 17th April, 2020. It provided 12, 741 tweets with the hashtag 'Indialockdown'. Tweepy library, TextBlob & Vader lexicons and NLTK were used.

This article examines [2] public opinions toward social separation as expressed in Twitter textual data. The SentiStrength algorithm was used to calculate the polarity of emotions in Twitter data from Canada, as well as tweets containing social distancing keywords. Sentiment was classified using SVM approach. In [3], Refers to sentiment analysis of tweets related to the covid-19 vaccine and to examine changes in sentiment at the global, national, and state levels in the USA. The compound score is produced using the Valence Aware Dictionary and the sentiment Reasoner.

This paper [4] is about human brain science under horrendous circumstances. The Coronavirus pandemic is a debacle that has brought about a huge number of mental hardships, including wretchedness, because of unexpected social changes and an absence of occupations. With the Coronavirus pandemic, numerous nations had various tops, with the ascent and reduction of new cases influencing lockdowns, which directly troubled occupations. During the surge of COVID-19 cases with tighter lockdowns, people have been venting their frustrations on social media. This could lead to a better understanding of human psychology in emergency situations.

In [5], author aim is to study sentiments of people about lockdown/social distancing. Besides using mask and sanitizers, Indian Government has also decided to maintain proper social distancing or lockdown. In this paper, they have collected twitter data of people across India, during March to June and then using NLP, the polarity is measured, i.e., positive, negative or neutral. After that, they have used SVM classifier and Logistic Regression.

In [6], temporal trends of Twitter posts in Massachusetts (MA) was area of focus. They researched changes in tweet feeling extremity, communicated nervousness, and conversation on hazard and well being themes on partial subset of tweets. They find tremendous contrasts between tweets. It justifies the use of Twitter data, particularly when paired with other observational data.

In [7], symbolic and ML are the two techniques for sentiment analysis. ML techniques are more sophisticated and time- consuming than symbolic alternatives. Different classifiers based on entropy and ensemble, such as Naïve Bayes, SVM, and Maximum Likelihood Classifiers, are used to assess the classification accuracy of the feature vector.

In [8], Indonesian government formed a team to foster development of COVID vaccines. With over 9.5 crores of confirmed cases, over 20 lacs deaths reported worldwide, the task force elicited feedback from countrymen using Twitter for the COVID-19 vaccine. The exercise used the Naïve Bayes technique to analyze Twitter data gathered for the phrase 'Vaccine COVID-19.'

In [9], the authors made the following contributions:

- Show usage of opinion mining with recommender systems.
- Visualize sentiment analysis to aid recommendations.
- Evaluate product reviews and urban policies.

They used SVM and CRF (conditional random field) to classify feelings. It employed hybrid learning strategy for sentence level models given partially labelled training data.

In [10], research envisages to look at COVID-19-related remarks, concerns, and sentiments using Twitter. 13 topics classified into five categories:

1. precautions for public health,
2. impact on society,
3. fresh cases and deaths,
4. US, and
5. rest of the world.

LDA (Latest Dirichlet Allocation) was used with unigram and bigrams. The study covered 4 million tweets related to COVID-19.

In [12], coronavirus pandemic tweets spanning from 15th January to 15th April, 2020 was used to research the elements that portray falsehood on Twitter. The data is classified as dependable and non-dependable. Tweet validation depended upon an assortment of variables, counting tweet. The strategy portrayed in this paper has a high exactness in perceiving dependable and false Coronavirus tweets.

In [11], researchers of Typhoon Sandy found out how client opinions vary based upon distance from the catastrophe and concluded that as passionate uniqueness rises, the chance of retweeting a tweet diminishes. Also, they investigated the impact of dissimilarity in enthusiasm on a tweet's retweet capacity and found that, as passionate uniqueness rises, the chance of retweeting a tweet diminishes. It helps in geo-planned opinion investigation calculations. The study highlights that removing fake opinions could assist authorities with laying out a superior situational consciousness of the calamity zone.

3. Sentiment analysis techniques

Sentiment analysis comprises of steps as given below:

Data collection: The gathering of data is the first and foremost important step. Without availability of reliable data, no ML experiment can be successful.

Pre-processing: After collecting the data, often the data is not in good shape. Preprocessing enables researcher to clean the data. It assists to lessen, if not eliminate, noise in the data. The noise usually causes trouble while extracting insights. Usual actions taken relate to removal of stop words, repeated words. Stemming and lemmatization are bare necessities in case NLP work. Other peripheral activities such as removal of emoticons and URLs are essential, too.

Feature extraction: If incorrect features are selected for analysis, it leads to poor result. Hence, precise selection of

features is a crucial activity in working out the accuracy of the model. Extraction of the features is obvious successor step to selection.

Sentiment classification: There is multiple ML and DL techniques used for classification. Variants of Bayesian techniques such as Naïve Bayes (NB) and SVM are the widely used ones.

Polarity detection: The polarity calculation comes next to sentiment classification. The eventual outcome of polarity detection is classes viz. positive, negative or neutral. It indicates what the text indicates overall psychologically.

Validation and evaluation: The model developed is of no use unless it's results are confirmed by any validation technique. Accuracy of the sentiment analysis model is determined by the evaluation methods. Only after this exercise, that the model can be practically applied.

4. Evaluation Metrics

Metrics like Precision (equation 1), Recall (equation 2), F1-score (equation 3) and Accuracy (equation 4) are used to evaluate performance of a model used for sentiment analysis

$$P = TP/(TP + FP') \tag{1}$$
$$R = TP/(TP + FN') \tag{2}$$
$$F = 2 * (P * R)/(P + R') \tag{3}$$
$$A = (TP + TN)/(TP + TN + FP + FN') \tag{4}$$

where TP, TN, FP, and FN are abbreviations of true positive, true negative, false positive, and false negative, respectively. Similarly, P, R, F, and A are Precision, Recall, F1-score, and Accuracy, respectively.

5. Challenges and limitations

Credibility/Behaviour

Only the traces that people leave in social media can be used to track their actions. It's impossible to validate the veracity of these behavioral patterns, even if they're analyzed on social media and similar patterns are obtained. In industries where critical judgments are made based on observations of human behaviour, evaluation becomes considerably more difficult.

Taunting

It's possible to utilize sarcasm in a variety of ways. Can be used to injure or anger someone, or it can be used to make people laugh. It's a satirical or ironic statement that appears to be complimenting someone or something but is insulting or cutting. Detecting sarcasm in expressions and determining the appropriate context-related attitudes is a difficult process.

Author Segmentation Analysis

Many people, referred to as review authors, can express their feelings about a certain target. These authors should be classified based on their commenting style so that credibility can be easily assessed. This credibility assessment is useful in making decisions.

Words with Grammatical Errors

While there are multiple ways to extract attitude, grammatical faults lead to poor evaluation of attitude. If such errors can be corrected, results of sentiment analysis tend to be more precise

Managing Dynamism and Noise

Because social media data is large, noisy, unstructured, and dynamic, it poses new challenges and research problems in social media mining. It's difficult to spot and eliminate noisy data.

Data uncertainty

The findings are based on tweets and messages shared on social media, and a tiny proportion of the population uses social media to express their emotions and opinions. On social media, not everyone has the freedom to express themselves.

Using Facebook messages to perform sentiment analysis

Facebook Graph API and surrounding strict rules prohibit researcher from gathering data from Facebook effectively. This has led to less work being done on Facebook textual contents.

Dependence on the domain

Context forms an essential sphere of influence while evaluating meaning of a phrase or sentence. It's quite possible that same word means something else when context changes. E.g. the word "unpredictable" indicates a positive rating for movie/drama, but means exactly opposite for vehicle's steering.

Identifying text's subjective elements

Sentimental content is represented by subjective components. In certain circumstances, the same term can be interpreted subjectively, while in others, it can be interpreted objectively. As a result, determining whether parts of the text are subjective is impossible.

Internationalization

The current study focuses on English-language content, yet Twitter has a global user base.

6. Conclusions

Sentiment Analysis is a technique for examining people's emotions, attitudes, and views about a product, service, or organization. Businesses and governments utilize sentiment analysis to make decisions based on the data.

This article highlights the numerous obstacles that exist in sentiment analysis, as well as what has been done and what still needs to be done in terms of various research opportunities. We also looked at the many forms of sentiment analysis, how to classify sentiment analysis, and how to employ sentiment analysis approaches.

We have largely focused on perceptions of people associated with Covid-19 pandemic, lockdown. We zero-in on influence of these on health and scientific discipline. We have tendency to gather publications that use varied machine learning approaches to perform sentiment analysis. The results and strategies utilized in used in these studies have been mentioned.

DL techniques are extremely well poised to be employed for sentiment analysis, as latter has widespread application

Table 1. Literature survey

Title	Author(s)	Method (s)	Results
Sentiment Analysis of Lockdown in India During COVID-19: A Case Study on Twitter	Gupta, P., Kumar, S., Suman, R. R., and Kumar, V. (2021).	Logistic Regression Multinomial Naïve Bayes,Linear SVC, Ridge classifier, AdaBoost classifier	The outcome shows that the LinearSVC classifier and unigrams perform the best. In general, Indians are actively supporting government's decision for nationwide lockdown to control the outbreak of Corona virus.
Sentiment Analysis on COVID-19-Related Social Distancing in Canada Using Twitter Data	Shofiya, C., and Abidi, S. (2021).	Support vector machine	The outcomes showed that an expansion in preparing information with positive and negative feeling extremity expanded the exhibition of the calculation
Public attitudes toward COVID-19 vaccines on English-language Twitter: A sentiment analysis	Liu, S., and Liu, J. (2021)	latent Dirichlet allocation analysis	Investigation of public feeling towards creating immunity against Coronavirus will help research conducted to develop relevant antibodies. It is also guiding indicator for general-well being-policymakers in conducting specialized custom immunization schooling activities.
COVID-19 sentiment analysis via deep learning during the rise of novel cases	Chandra R, Krishna A (2021)	latent Dirichlet allocation analysis	Authors examined that large part of the tweets contains words viz. "hopeful", "irritated" and "kidding" that communicates positive thinking, dread and vulnerability during the ascent of the Coronavirus cases in India. These expectations by and large demonstrate, albeit the greater part have been hopeful, a huge gathering of populace has been irritated towards how the pandemic was taken care of by the specialists.
Sentiment Analysis of People During Lockdown Period of COVID-19 Using SVM and Logistic Regression Analysis	Majumder, Sayan and Aich, Anuran and Das, Satrajit (2021)	SVM, Logistic Regression	Result shows that SVM performed better than Logistic Regression. It also shows that people across India have positive sentiments regarding the lockdown ordered by Indian Prime Minister as a protective measure to stop the spread of COVID19.
Exploring discussions of health and risk and public sentiment in Massachusetts during COVID-19 pandemic mandate implementation: A Twitter analysis	Thorpe Huerta, D., Hawkins, J. B., Brownstein, J. S., and Hswen, Y. (2021)	LIWC analytic score linear model	In this way, tweets have the potential to be a more informal and low-cost way of monitoring public sentiment, streamlining the data collection process and allowing for faster analysis. Compared to old-school survey methods, it enables quicker policy implementation tuned to public needs.
Sentiment Analysis in Twitter	Sayali P. Nazare1, Prasad S. Nar2, Akshay S. Phate3, Prof. Dr. D. R. Ingle	Naïve Bayes, SVM, Maximum Entropy and Ensemble	classifiers Nearly identical accuracy for classification of the feature vector is observed when comparing Naïve Bayes, SVM, Max Entropy, and Ensemble classifiers.
Sentiment analysis of COVID-19 vaccine in Indonesia using Naïve Bayes Algorithm	Pristiyono, Ritonga, Mulkan and Ihsan, Muhammad and Anjar, Agus and Rambe, Fauziah. (2021)	Naïve Bayes	Most of the tweets are negative. It indicates a cold rejection offered by countrymen towards COVID-19 vaccination.
Twitter discussions and emotions about COVID-19 pandemic: A machine learning approach.	Xue, J., Chen, J., Hu, R., Chen, C., Zheng, C., Liu, X., and Zhu, T (2020)	Latent Dirichlet Allocation	It showcases how researchers can investigate evolving public debates and feelings during the COVID-19 pandemic using ML techniques.

across industry verticals for variety of business domains. The Smart Deep Convolution Neural Network (SDCNN) framework will be used in future studies to attain better accuracy than present deep learning methods.

References

[1] 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 Transactions on Computational Social Systems*, vol. 8, no. 4, pp. 992–1002, 2020.

[2] C. Shofiya and S. Abidi, "Sentiment analysis on covid-19-related social distancing in canada using twitter data," *International Journal of Environmental Research and Public Health*, vol. 18, no. 11, p. 5993, 2021.

[3] S. Liu and J. Liu, "Public attitudes toward covid-19 vaccines on english-language twitter: A sentiment analysis," *Vaccine*, vol. 39, no. 39, pp. 5499–5505, 2021.

[4] R. Chandra and A. Krishna, "Covid-19 sentiment analysis via deep learning during the rise of novel cases," *PLoS One*, vol. 16, no. 8, p. e0255615, 2021.

[5] S. Majumder, A. Aich, and S. Das, "Sentiment analysis of people during lockdown period of covid-19 using svm and logistic regression analysis," *Available at SSRN 3801039*, 2021.

[6] D. T. Huerta, J. B. Hawkins, J. S. Brownstein, and Y. Hswen, "Exploring discussions of health and risk and public sentiment in massachusetts during covid-19 pandemic mandate implementation: A twitter analysis," *SSM-Population Health*, vol. 15, p. 100851, 2021.

[7] S. P. Nazare, P. S. Nar, A. S. Phate, and D. Ingle, "Sentiment analysis in twitter," *Int Res J Eng Technol (IRJET)*, vol. 5, 2018.

[8] M. Ritonga, M. A. Al Ihsan, A. Anjar, F. H. Rambe *et al.*, "Sentiment analysis of covid-19 vaccine in indonesia using naïve bayes algorithm," in *IOP Conference Series: Materials Science and Engineering*, vol. 1088, no. 1. IOP Publishing, 2021, p. 012045.

[9] K. Gandhe, A. S. Varde, and X. Du, "Sentiment analysis of twitter data with hybrid learning for recommender applications," in *2018 9th IEEE Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON)*. IEEE, 2018, pp. 57–63.

[10] J. Xue, J. Chen, R. Hu, C. Chen, C. Zheng, Y. Su, T. Zhu *et al.*, "Twitter discussions and emotions about the covid-19 pandemic: Machine learning approach," *Journal of medical Internet research*, vol. 22, no. 11, p. e20550, 2020.

[11] V. K. Neppalli, C. Caragea, A. Squicciarini, A. Tapia, and S. Stehle, "Sentiment analysis during hurricane sandy in emergency response," *International journal of disaster risk reduction*, vol. 21, pp. 213–222, 2017.

[12] M. S. Al-Rakhami and A. M. Al-Amri, "Lies kill, facts save: detecting covid-19 misinformation in twitter," *Ieee Access*, vol. 8, pp. 155 961–155 970, 2020.