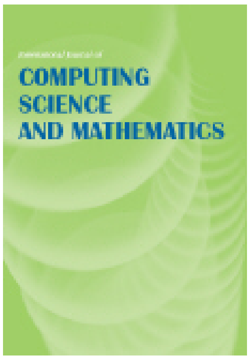



**International Journal of Computing Science and Mathematics**

ISSN online: 1752-5063 - ISSN print: 1752-5055
https://www.inderscience.com/ijcsm

---

# COVID-19: machine learning methods applied for twitter sentiment analysis of Indians before, during and after lockdown

H.S. Hota, Dinesh K. Sharma, Nilesh Verma

**DOI:** 10.1504/IJCSM.2023.10055652

**Article History:**

| | |
|---|---|
| Received: | 31 August 2020 |
| Last revised: | 12 December 2020 |
| Accepted: | 14 December 2020 |
| Published online: | 20 April 2023 |

# COVID-19: machine learning methods applied for twitter sentiment analysis of Indians before, during and after lockdown

## H.S. Hota*

Department of Computer Science and Application,
Atal Bihari Vajpayee University,
Bilaspur, 495001, India
Email: proffhota@gmail.com
*Corresponding author

## Dinesh K. Sharma

Department of Business, Management and Accounting,
University of Maryland Eastern Shore,
Princess Anne, MD, 21853, USA
Email: profdksharma@gmail.com

## Nilesh Verma

Department of Computer Science and Application,
Atal Bihari Vajpayee University,
Bilaspur, 495001, India
Email: me@nileshverma.com

**Abstract:** This paper emphasises the analysing sentiment of Indian citizens based on Twitter data using machine learning (ML) based approaches. The sentiment of about 1,51,798 tweets extracted from Twitter social networking and analysed based on tweets divided into six different segments, i.e., before lockdown, first lockdown, lockdown 2.0, lockdown 3.0, lockdown 4.0 and after lockdown (Unlock 1.0). Empirical results show that ML-based approach is efficient for sentiment analysis (SA) and producing better results, out of 10 ML-based models developed using N-Gram ($N$ = 1,2,3,1–2,1–3) features for SA, linear regression model with term frequency – inverse term frequency (Tf-Idf) and 1–3 Gram features is outperforming with 81.35% of accuracy. Comparative study of the sentiment of the above six periods indicates that negative sentiment of Indians due to COVID-19 is increasing (About 4%) during first lockdown by 4.0% and then decreasing during lockdown 2.0 (34.10%) and 3.0 (34.12%) by 2% and suddenly increased again by 4% (36%) during 4.0 and finally reached to its highest value of 38.57% during unlock 1.0.

**Keywords:** ML; machine learning; twitter; SA; sentiment analysis; logistic regression; COVID-19; lockdown.

**Reference** to this paper should be made as follows: Hota, H.S., Sharma, D.K. and Verma, N. (2023) 'COVID-19: machine learning methods applied for twitter sentiment analysis of Indians before, during and after lockdown', *Int. J. Computing Science and Mathematics*, Vol. 17, No. 1, pp.95–105.

**Biographical notes:** H.S. Hota is working as an Associate Professor in the Department of Computer Science of Atal Bihari Bajpayee University, India. He earned MCA and PhD in Computer Science and published more than 50 refereed papers in reputed journals along with one book. He has also delivered talk in international conferences and presented papers in many international and national conferences.

Dinesh K. Sharma is a Professor of Quantitative Methods and Computer Applications in the Department of Business, Management and Accounting at the University of Maryland Eastern Shore. He earned his MS in Mathematics, MS in Computer Science, PhD in Operations Research, and a second PhD in Management. He has published over 200 refereed journal papers and conference proceedings and has won 14 best paper awards. He is the Editor-in-Chief of the *Journal of Global Information Technology and Review of Business and Technology Research*.

Nilesh Verma is a student of MSc (Computer Science) in the Department of Computer Science of Atal Bihari Bajpayee University, India. His research interest area is machine learning.

## 1 Introduction

As per the initial report published by World Health Organization (WHO, 2020) on 21st January, 2020, a new corona family virus emerged in December 2019 in Wuhan, Hubei province of China, and has spread globally at a fast rate. The situation in India is also not good, and it is panic. More than 2700000 cases found corona positive, out of which more than 59000 persons died. On the other hand, about 2460000 persons were cured (Ministry of Health and Family Welfare, GoI, 2020)**.**

As compared to developed countries, India is having fewer medical facilities and less health infrastructure. Also, the population of the country is one of the biggest challenges to fight against COVID-19. To protect the lives of Indians, Indian prime minister Shri Narendra Modi has announced a complete lockdown for 21 days on 24th March, 2020, and was effective from midnight of the same day. Further, this lockdown period has been extended for another 19 days till 3rd May, 2020. This period is known as lockdown 2.0. This process is continued as lockdown 3.0 and lockdown 4.0 till 31st May, 2020, and finally, from 1st June, 2020 government has given relaxations to open market, local transportation, railway, and air transportations. Despite all these precaution measures, situations in India are not right and continuously increasing day by day.

Sentiment analysis (SA) is a process of classification of text (Tweet in our case) into different categories like Positive and Negative sentiments. Twitter as a social network is popular and widely used online platform to raise opinion by the general public called user on a specific issue and also one of the best source of data for SA. Twitter is also playing a vital role to spread any social issues, crises or disaster in a rapid way through millions of its active users. The public often uses social media if any adverse situation arises in the country or outbreak like COVID-19 and shares their opinions, views, and emotions. It has often been criticised and analysed whether Twitter data is an authentic source of data to solve a real-world problem or not? In this direction, Lim and Tucker (2019) have presented an expert and intelligent system that identifies term groups having a causal relationship with real-world enterprise outcomes from Twitter data.

Sentiment analysis has always been an interest of researchers and is applied in various domains (Paramanik and Singhal, 2020; Chen and Alexender, 2020). Many others (Yang and Chen, 2017; Tang et al., 2018; Ruz et al., 2020) have applied machine learning techniques as well as deep learning techniques (Ramadhani and Goo, 2017) for sentiment analysis. Prabha and Srikanth (2019) presented a survey of SA that used deep learning techniques. Also, a detailed review of applications in fighting COVID-19 has been done by Alamoodi et al. (2020). In a recent paper, Barkur et al. (2020) have analysed the sentiment of Indians due to the COVID-19 pandemic based on Twitter data during the lockdown. They found that there is negative sentiment regarding many emotions like fear, disgust, and sadness about the lockdown. However, Indians still are more positive towards the situation due to COVID-19. Overall, it is safe to confirm from the several reviewed papers that Twitter is the most reliable data platform to analyse sentiment on any global issue like COVID-19. ML has also proven itself for SA to classify human sentiment as positive or Negative.

This research work is carried out to address the sentiment of Indians due to the novel coronavirus COVID-19 before, during and after lockdown. This piece of research work uses ML-based approaches for SA. A total of 1,51,798 tweets extracted from Twitter social networking and divided into six different segments i.e., Segment-1 for before lockdown, Segment-2 for first lockdown, segment-3 for lockdown 2.0, segment-3 for lockdown 3.0, segment-4 for lockdown 4.0 and segment-5 for unlock 1.0. Pre-processed applied on Twitter data to remove noises from the tweets. ML models were trained using the benchmark dataset obtained from Kaggle (2020). Results of this research work show that out of 10 different ML models, the linear regression model is outperforming with the highest accuracy, sensitivity, and specificity. Pre-processed Twitter data are then used to find out sentiment as Positive and Negative

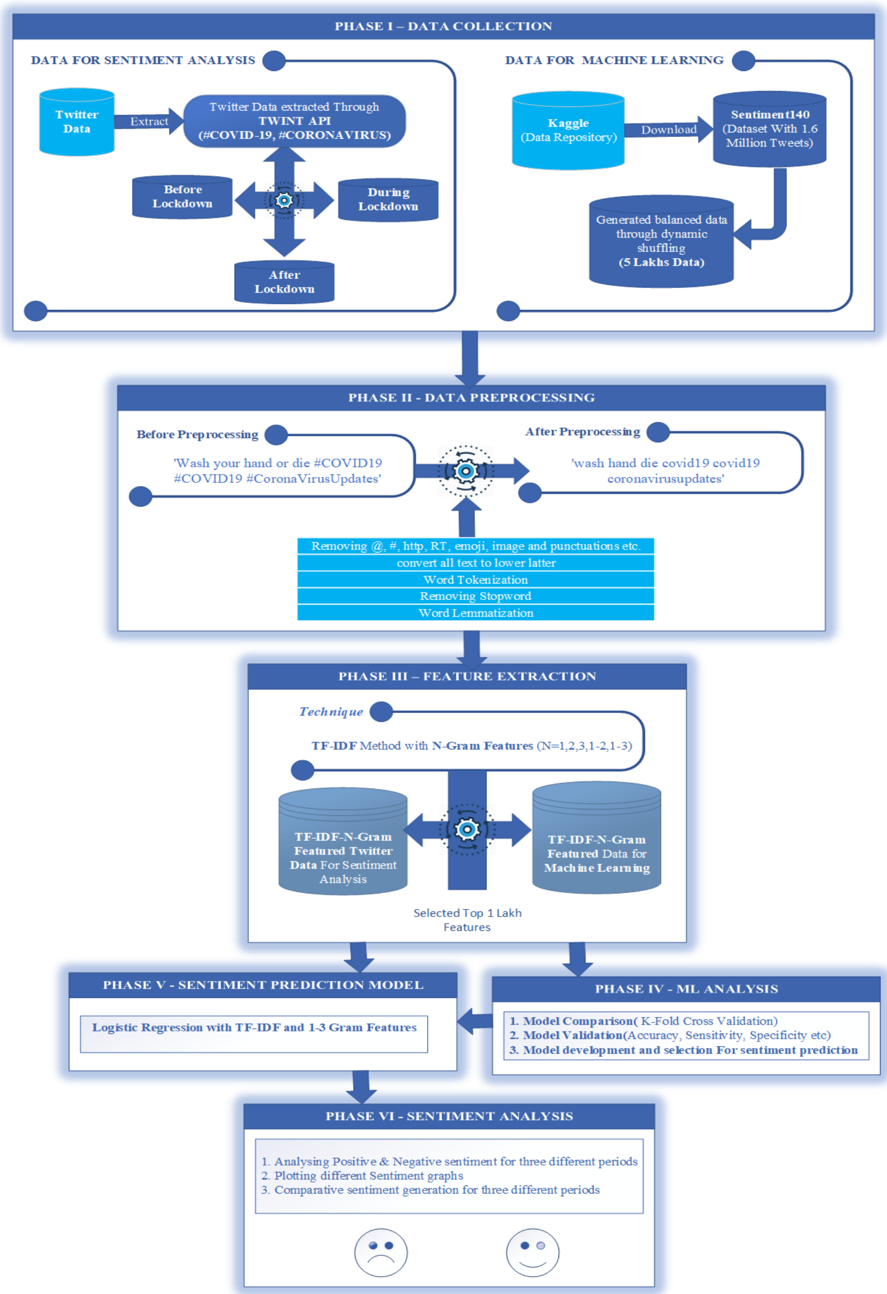
## 2 Proposed methodology

A systemic flow diagram of the proposed research work is depicted in Figure 1, which consists of 5 different phases of SA of Twitter data using the ML approach. The detail of phase I to III is explained below other phases (Phase IV to VI) are explained in the experimental section.

### *2.1 Data collection*

*Twitter data for sentiment analysis*: Data from Twitter has been extracted using TWINT API, known as a twitter intelligence tool, and divided into three different segments as follow: Segment-1: Before lockdown – 15th March to 24th March, 2020.Segment-2: First lockdown – 25th March to 4th April, 2020.Segment-3: Lockdown 2.0 – 15th April to 24rd April, 2020. Segment-4: Lockdown 3.0 – 4th May to 17th May, 2020. Segment-5: Lockdown 4.0 – 18th May to 31$^{st}$ May, 2020. Segment-6: Unlock 1.0 – 1st June to 9th June, 2020.

**Figure 1** Process flow diagram for sentiment analysis (see online version for colours)




### 2.1.1 *Benchmark data for machine learning (ML)*

Most authors have either used data generated through a lexicon-based approach or used Twitter data after pre-processing and labelling it through a vocabulary dictionary for ML. Since similar benchmark data are available in the repository, it is good to utilise benchmark data for model development. So this research work uses a separate dataset from the Kaggle data repository named Stanford sentiment140 corpus (Kaggle, 2020) for ML. This dataset is quite popular and used by many authors (Vyas and Uma, 2018; Ankit and Saleena, 2018) to develop a sentiment-based classification model. Sentiment140 data mainly used to train the ML models for SA containing 1.6 lakh tweets with 6 features, i.e., target, ID, date, flag, user, and text with 0.80 lakh positive and negative class labels. Due to computational limitation, a subset of the original dataset with 5 lakhs data selected based on reshuffling method with balanced samples of Positive and Negative (2.5 lakhs each) was considered and used to train the ML models. However, ML models' training with high sample size is always recommended to develop a robust ML model.

### *2.1.2 Data pre-processing*

Data pre-processing is an important and essential step, especially in the case of Twitter data. Twitter data (Singh et al., 2016) are highly unstructured; small message data containing emoji, hashtag, stop word, unidentified word, symbols, abbreviations, etc. as noise and user creates these data with their own shortcut words and spelling, which makes tweets so complicated. Hence, it is challenging to remove these and pre-process them before using them for ML-based classification tasks. Singh and Kumari (2016) have worked on the role of text pre-processing in Twitter SA. This research work has also applied data pre-processing to remove many irrelevant contents from Twitter data. Many other studies revealed that pre-processing of Twitter data certainly enhances the accuracy of classifiers. In general, the following activities (Pandey et al., 2019) performed in the pre-processing of textual data like Twitter: First of all noises which does not play important roles are need to be removed and then tweets are converted into lower case after that tokenisation process is performed followed by removing stop word like papers, prepositions and conjunctions and finally lemmatisation (Agarwal, 2018) process is performed.

## *2.2 Feature extraction*

Feature extraction is a process of generating new feature space from existing feature space. Text is exceptionally sparse and high-dimensional, which causes off-the-shelf multidimensional models to behave in unexpected ways. The frequency of a single term often contains little predictive power, and it is only by using combinations of many features that robust classification can be achieved (Agarwal, 2018).

i *N-Gram*: N-Gram is a process of extracting features from textual data like a tweet in which tweets can be broken down into words and appended to the feature vector (Chakraborty et al., 2020).

ii *Term frequency-inverse term frequency* (*Tf-Idf*): The more a word appears in a document, the more likely it is crucial to that document. We call this term frequency (TF). In contrast, if a word appears in many documents, it is less important to any individual document. We call this document frequency (DF). By combining these two statistics, we can assign a score to every word representing how important that word is in a document. Specifically, we multiply Tf to the inverse of document frequency (IDF).

### *2.3 Machine learning (ML) methods*

There are three popular approaches widely used for sentiment analysis: Lexicon based, machine learning based, and Hybrid. All these approaches involve various computational steps. Lexicon based approach determines polarity or sentiment to classify tweets in three categories: Negative, Neutral, and Positive. On the other hand, the ML-based approach is widely used and more robust than the lexicon-based approach. ML methods like support vector machine (SVM) and Naïve Bayes were proven to obtain the best results in accuracy using different benchmark datasets (Hassonath et al., 2020). The 10 ML algorithms are found to be best for SA related to COVID-19. Naive Bayes (NB) Variants can be used together with the Bayes theorem to obtain the posterior probability of the class variable *C* given an input data point (Ruz et al., 2020; Pearl, 1998). Multinomial NB implements the Naive Bayes algorithm for multinomially distributed data, and is one of the two classic naive Bayes variant used in text classification (Rennie et al., 2003; Aggarwal, 2018.Stochastic Gradient Descent (SGD) is a simple yet very efficient approach to discriminative learning of linear classifiers under convex loss functions such as (linear) SVMs and logistic regression (Zhang, 2004). Logistic regression is a member of the family of generalised linear models, which have a natural probabilistic interpretation. Random forest is a combination of tree predictors known as a forest. Each tree depends on a random vector's values sampled independently and with the same distribution for all trees in the forest. Support vector machine (SVM) is a supervised learning method that generates input-output mapping functions from a set of labelled training data. The mapping function can be either a classification function (used to categorise the input data) or a regression function (used to estimate the desired output). Linear Support Vector classification (LSVC) implements SVM classification with the linear kernel (Olson and Delen, 2008). Ensemble voting classifier is a technique to combine more than one ML models to improve performance. Such a classifier can help set the equally well-performing model to balance out their weaknesses (Raschka, 2015) and other linear models that are used for sentiment classification are: Ridge regressor has a classifier variant of Ridge classifier. This classifier first converts binary targets to [1 –1] and then treats the problem as a regression task (Rifkin et al., 20017). Passive aggressive algorithm is a family of algorithms for large-scale learning. Perceptron is the most straightforward ANN architectures which classify linear data using threshold function (Scikit-learn, 2020).

## 3 Experimental setup

### *3.1 Performance measures*

The following measures can be formulated based on four variables. Let us assumes these variables as true positive-sentiment (TP-S), true negative-sentiment (TN-S), false positive-sentiment (FP-S), and false negative-sentiment (FN-S) as follows (Han and Kamber, 2006):

Accuracy = (TP-S + TN-S)/N, Sensitivity/Recall = TP-S/(TP-S + TN-S)

Specificity = TN-S/(TN-S + FP-S), Precision = TP-S/(TP-S + FP-S)

F-measure 2X (precision X recall)/(precision + recall). Where, N is the total number of available samples. On the other hand ROC curve is a graphical way of evaluating performance of any classifier.

## *3.2 ML-based model development*

As shown in Phase IV and V of Figure 1, sentiment classifier models were developed using the Scikit-learn ML tool of python based on 10 ML algorithms explained in Section 2.3. Pre-processed data of sentiment140 obtained from Kaggle with Tf-Idf and N-Gram features along with a 10-fold cross-validation technique was used to train the models. Results were verified based on various measures explained in Section 3.1 and presented in Tables 1–5. The highest value of each measure in each table is highlighted. From these tables, it is clear that models with Tf-Idf and $N = 1–3$ gram perform better than others. It can also be observed from this table that the Logistic regression linear model is outperforming others. A comparative bar graph of all the developed models is also shown in Figure 2.

**Table 1** Results of various machine learning methods for N-gram ($N = 1$)

| *Model* | *Accuracy* | *Sensitivity* | *Specificity* | *Precision* | *F1-Score* | *ROC* |
|---|---|---|---|---|---|---|
| Logistic regression | **79.38** | 80.76 | **78.01** | **78.60** | **79.66** | **79.38** |
| Linear SVC | 78.35 | **80.96** | 75.74 | 76.94 | 78.90 | 78.35 |
| Ensemble voting | 78.08 | 79.48 | 76.68 | 77.32 | 78.38 | 78.08 |
| Bernoulli NB | 77.52 | 77.90 | 77.13 | 77.31 | 77.60 | 77.52 |
| SGD | 77.08 | 78.32 | 75.85 | 76.43 | 77.36 | 77.08 |
| Ridge | 76.93 | 77.24 | 76.62 | 76.77 | 77.01 | 76.93 |
| Multinomial NB | 76.64 | 75.03 | 78.25 | 77.53 | 76.26 | 76.64 |
| Passive aggressive | 74.91 | 75.74 | 74.07 | 74.58 | 75.10 | 74.91 |
| Perceptron | 71.73 | 72.83 | 70.64 | 71.39 | 72.01 | 71.73 |
| Random forest | 67.91 | 75.35 | 60.47 | 65.61 | 70.11 | 67.91 |

**Table 2** Results of various machine learning methods for N-gram ($N = 2$)

| *Model* | *Accuracy* | *Sensitivity* | *Specificity* | *Precision* | *F1-Score* | *ROC* |
|---|---|---|---|---|---|---|
| Logistic regression | **77.19** | 80.31 | 74.08 | 75.60 | **77.88** | **77.19** |
| Linear SVC | 76.73 | 80.10 | 73.37 | 75.05 | 77.49 | 76.73 |
| Ensemble voting | 76.08 | 78.56 | 73.60 | 74.85 | 76.66 | 76.08 |
| Bernoulli NB | 75.86 | 72.64 | **79.08** | **77.64** | 75.06 | 75.86 |
| SGD | 75.83 | 79.66 | 71.99 | 73.99 | 76.72 | 75.83 |
| Ridge | 74.76 | 78.80 | 70.71 | 72.90 | 75.74 | 74.76 |
| Multinomial NB | 73.39 | 77.42 | 69.37 | 71.65 | 74.42 | 73.39 |
| Passive aggressive | 73.01 | 75.24 | 70.79 | 72.04 | 73.60 | 73.01 |
| Perceptron | 71.20 | 72.89 | 69.52 | 70.54 | 71.66 | 71.20 |
| Random forest | 59.94 | **89.06** | 30.83 | 56.28 | 68.97 | 59.94 |

**Table 3** Results of various machine learning methods for N-gram ($N = 3$)

| *Model* | *Accuracy* | *Sensitivity* | *Specificity* | *Precision* | *F1-Score* | *ROC* |
|---|---|---|---|---|---|---|
| Logistic regression | **68.99** | **81.05** | 56.94 | 65.30 | **72.33** | **68.99** |
| Linear SVC | 68.57 | 80.41 | 56.73 | 65.01 | 71.90 | 68.57 |
| Ensemble voting | 68.08 | 80.84 | 55.33 | 64.41 | 71.69 | 68.08 |
| Bernoulli NB | 67.92 | 79.39 | 56.46 | 64.58 | 71.22 | 67.92 |
| SGD | 67.82 | 80.91 | 54.72 | 64.12 | 71.54 | 67.82 |
| Ridge | 67.36 | 80.73 | 53.99 | 63.70 | 71.21 | 67.36 |
| Multinomial NB | 66.39 | 53.13 | **79.65** | **72.31** | 61.25 | 66.39 |
| Passive aggressive | 65.51 | 77.00 | 54.01 | 62.61 | 69.06 | 65.51 |
| Perceptron | 63.71 | 67.35 | 60.06 | 63.23 | 64.36 | 63.71 |
| Random forest | 52.74 | 79.74 | 25.73 | 57.27 | 56.18 | 52.74 |

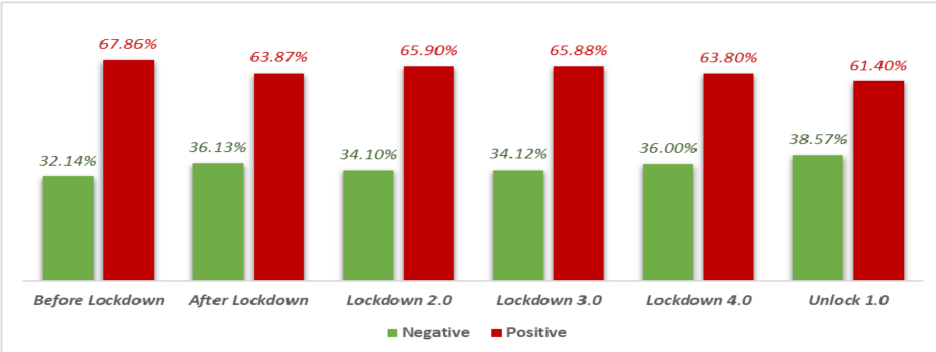
**Table 4** Results of various machine learning methods for N-gram ($N = 1–2$)

| *Model* | *Accuracy* | *Sensitivity/Recall* | *Specificity* | *Precision* | *F1-Score* | *ROC* |
|---|---|---|---|---|---|---|
| Logistic regression | **81.31** | **82.35** | 80.27 | **80.67** | **81.50** | **81.31** |
| Linear SVC | 79.93 | 81.44 | 78.42 | 79.05 | 80.23 | 79.93 |
| Ensemble voting | 79.54 | 80.65 | 78.42 | 78.89 | 79.76 | 79.54 |
| Bernoulli NB | 79.37 | 78.37 | **80.37** | 79.97 | 79.16 | 79.37 |
| SGD | 78.92 | 80.68 | 77.15 | 77.93 | 79.28 | 78.92 |
| Ridge | 78.86 | 79.35 | 78.37 | 78.58 | 78.96 | 78.86 |
| Multinomial NB | 77.36 | 77.63 | 77.09 | 77.21 | 77.42 | 77.36 |
| Passive aggressive | 76.49 | 77.11 | 75.88 | 76.18 | 76.64 | 76.49 |
| Perceptron | 74.85 | 74.09 | 75.61 | 75.27 | 74.65 | 74.85 |
| Random forest | 67.63 | 76.27 | 58.99 | 65.04 | 70.20 | 67.63 |

**Table 5** Results of various machine learning methods for N-gram ($N = 1–3$)

| *Model* | *Accuracy* | *Sensitivity/Recall* | *Specificity* | *Precision* | *F1-Score* | *ROC* |
|---|---|---|---|---|---|---|
| Logistic regression | **81.35** | **82.37** | **80.33** | **80.73** | **81.54** | **81.35** |
| Linear SVC | 79.94 | 81.51 | 78.38 | 79.03 | 80.25 | 79.94 |
| Ensemble voting | 79.67 | 81.03 | 78.30 | 78.88 | 79.94 | 79.67 |
| Bernoulli NB | 79.30 | 78.46 | 80.15 | 79.81 | 79.13 | 79.30 |
| SGD | 78.84 | 79.27 | 78.41 | 78.60 | 78.93 | 78.84 |
| Ridge | 78.62 | 80.94 | 76.30 | 77.35 | 79.10 | 78.62 |
| Multinomial NB | 77.35 | 77.38 | 77.33 | 77.34 | 77.36 | 77.35 |
| Passive aggressive | 76.71 | 77.09 | 76.33 | 76.53 | 76.80 | 76.71 |
| Perceptron | 75.07 | 74.91 | 75.22 | 75.18 | 75.02 | 75.07 |
| Random forest | 68.88 | 75.23 | 62.53 | 66.75 | 70.74 | 68.88 |

**Figure 2** Comparative sentiment (%) for six different segments (see online version for colours)




## 4 Sentiment analysis and result discussion

A comparative graph of sentiment analysis (Phase VI of Figure 1) of all six segments based on Twitter data collected for six different periods is shown in Figure 2. This graph clearly shows that positive sentiment is decreased by about 4% during first lockdown. In comparison, it has increased 2% again during lockdown 2.0 and lockdown 3.0, which is a positive indication of the better psychological status of Indians and show the optimistic and pessimistic situation. However, it has decreased by 2% during lockdown 4.0 and again by 2% during unlock 1.0. On the other hand, variations in negative sentiment from one period to another period show the negative mental status of Indians due to the COVID-19 pandemic. Negativity has increased by 4% soon during lockdown announced by Indian prime minister Shri Narendra Modi from March 25th, 2020, indicates that Indians were little nervous and then decreased by 2% during lockdown 2.0 and lockdown 3.0. Finally, it has increased by 2% and 4% during lockdown 4.0 and unlock 1.0, respectively. These analytical results indicate that unlock 1.0 was an unsafe condition due to which Indians are more negative currently.

## 5 Conclusion

This paper compares the sentiment of Indians before, during and after lockdown based on Twitter data. It has been observed that negative sentiment has increased by 4% during first lockdown and decreased by 2% during lockdown 2.0 and lockdown 3.0. The reason may be that the Indians have experienced the critical situation of COVID-19 worldwide first time more deeply during lockdown and then during lockdown 2.0 and lockdown 3.0. One reason for the increasing negativity among the Indians during lockdown 4.0 and unlock 1.0 is the increasing numbers of COVID-19 cases in India during May 2020 and unlock situations.

## Acknowledgement

The authors are very much thankful to the editor and learned referees for their valuable suggestions that improved the earlier draft.

## References

Agarwal, C.C. (2018) *Machine Learning for Text*, Springer International Publishing, Available from https://doi.org/10.1007/978-3-319-73531-3

Alamoodi, A.H., Zaidan, B.B., Zaidan, A.A., Albahri, O.S., Mohammed, K.I., Malik, R.Q., Almahdi, E.M., Chyad, M.A., Tareq, Z., Albahri, A.S., Hameed, H. and Alaa, M. (2021) 'Sentiment analysis and its applications in fighting COVID-19 and infectious diseases: a systematic review', *Expert Systems with Applications*, Vol. 167, 114155, ISSN 0957-4174, https://doi.org/10.1016/j.eswa.2020.114155

Ankit and Saleena, N. (2018) 'An ensemble classification system for twitter sentiment analysis', *Procedia Computer Science*, Vol. 132, pp.937–946, 10.1016/j.procs.2018.05.109.

Barkur, G., Vibha and Kamath, G.B. (2020) 'Sentiment analysis of nationwide lockdown due to COVID 19 outbreak: evidence from India', *Asian Journal of Psychiatry*, Vol. 51, Article 102089, https://doi.org/10.1016/j.ajp.2020.102089

Chakraborty, K., Bhatia, S. and Bhattacharyya, S. (2020) 'Sentiment analysis of COVID-19 tweets by deep learning classifiers=A study to show how popularity is affecting accuracy in social media', *Applied Soft Computing Journal*, Vol. 97, pp.1–14.

Chen, G. and Alexender, K. (2020) 'Informational role of social media: Evidence from twitter sentiment', *Journal of Banking and Finance*, Vol. 121, pp.1–13.

Han, J. and Kamber, M. (2006) *Data Mining: Concepts and Techniques*, 2nd ed., Morgan Kaufmann Publisher, San Francisco, CA.

Hassonath, A.M., Al-Sayyed, R., Rodan, A.M., Al-Zoubi, A., Aljarah, I. and Faris, H. (2020) 'An efficient hybrid filter and evolutionary wrapper approach for sentiment analysis of various topics on Twitter', *Knowledge-Based Systems*, p.192.

Olson, D.L. and Delen, D. (2008) *Advanced-Data Mining Techniques*, Springer-Verlag Berlin Heidelberg.

Pandey, M., Williams, R., Jindal, N. and Batra, A. (2019) 'Sentiment analysis using Lexicon based approach', *IITM Journal of Management and IT*, Vol. 10, No. 1, pp.68–76.

Paramanik, R.N. and Singhal, V. (2020) 'Sentiment analysis of Indian stock market volatility', *Procedia Computer Science*, Vol. 176, pp.330–338.

Pearl, J. (1988) *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*, Morgan Kaufmann, San Francisco, California.

Prabha, M.I. and Srikanth, G.U. (2019) 'Survey of Sentiment Analysis Using Deep Learning Techniques', *1st International Conference on Innovations in Information and Communication Technology (ICIICT)*, Chennai, India, pp.1–9, doi: 10.1109/ICIICT1.2019.8741438.

Ramadhani, A.M. and Goo, H.S. (2017) 'Twitter sentiment analysis using deep learning methods', *7th International Annual Engineering Seminar (InAES)*, Yogyakarta, Indonesia, pp.1–4, doi: 10.1109/INAES.2017.8068556.

Raschka, S. (2015) *Python Machine Learning*, Packt Publishing Ltd., Birmingham.

Rennie, J.D., Shih, L., Teevan, J. and Karger, D.R. (2003) 'Tackling the poor assumptions of Naive Bayes text classifiers', *ICML*, Vol. 3, pp.616–623.

Rifkin, Ryan and Lippert, R. (2007) *Notes on Regularized Least Squares*, http://cbcl.mit.edu/publications/ps/MIT-CSAIL-TR-2007-025.pdf

Ruz, G.A., Henriquenz, P.A. and Mascareno, A. (2020) 'Sentiment analysis of Twitter data during critical events through Bayesian networks classifiers', *Future Generation Computer Systems*, Vol. 106, pp.92–104.

Singh, T. and Kumari, M. (2016) 'Role of text pre-processing in twitter sentiment analysis', *Procedia Computer Science*, Vol. 89, pp.549–554.

Vyas, V. and Uma, V. (2018) 'An extensive study of sentiment analysis tools and binary classification of tweets using rapid miner. *Procedia Computer Science*, Vol. 125, pp.329–335.

Yang, P. and Chen, Y. (2017) 'A survey on sentiment analysis by using machine learning methods', *IEEE 2nd Information Technology, Networking, Electronic and Automation Control Conference* (*ITNEC*), Chengdu, pp.117–121.

Zhang (2004) 'Solving large scale linear prediction problems using stochastic gradient descent algorithms', *ICML '04: Proceedings of the Twenty-First International Conference on Machine Learning*, July, p.116, https://doi.org/10.1145/1015330.1015332

## Websites

Ministry of Health and Family Welfare, Government of India (GoI), https://www.mohfw.gov.in/ (Accessed 19 April, 2020).

Twitter Statistics and Facts, [Online], https://www.statista.com/topics/737/twitter/ (Accessed 20 April, 2020).

World Health Organization. Coronavirus Disease (COVID-2019) Situation Reports [Online] https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports (Accessed 29 April, 2020).