14

Lexicon-based sentiment analysis using Twitter data: a case of COVID-19 outbreak in India and abroad

H.S. Hota[1], Dinesh K. Sharma[2], Nilesh Verma[1]

[1]ATAL BIHARI VAJPAYEE UNIVERSITY, BILASPUR, CHHATTISGARH, INDIA; [2]UNIVERSITY OF MARYLAND EASTERN SHORE, PRINCESS ANNE, MD, UNITED STATES

1. Introduction

Emerging infectious diseases (EIDs) refer to those infectious diseases appearing in the last 20 years like severe acute respiratory syndrome. Many EIDs emerged in the last decades, e.g., H_1N_1 outbreak in 2009, Ebola in 2013 and 2016 [1]. When the entire world was ready to welcome the New Year at the end of 2019, Wuhan, a famous city in China, faced an outbreak of a novel coronavirus. The novel was named as 2019 novel coronavirus (2019-nCov) by the Chinese researchers. The international committee on the taxonomy of virus named the virus COVID-19 [2]. COVID-19 is a kind of EID that is more dangerous than earlier EIDs and spreading rapidly as compared to other EIDs. As per the initial report published by the World Health Organization [3] on January 21, 2020, a new virus of corona family emerged in December 2019 in Wuhan, Hubei province of China, and has spread globally all over the world at a rapid rate. Till the date of writing this article, more than 32 lakhs persons are affected in 215 countries of the world because of this EID, and more than 2.3 lakhs deaths were recorded, and number is increasing day by day. WHO has declared it a global pandemic.

Sentiment Analysis (SA) is a kind of opinion mining (OM) that is also known as emotion. Artificial Intelligence a series of methods, techniques, and tools about a detective and extracting personal information such as opinion and attitude from language [4] and also is a subfield of natural language processing (NLP). The rise of social media has fueled the interest of SA because people post their tweets every day to put their opinion, views without any restrictions or fear. Through Twitter, users can send a short 280 characters message, including emoji, as well as can read and follow other users.

Data Science for COVID-19. https://doi.org/10.1016/B978-0-12-824536-1.00015-0

Twitter currently ranks as one of the leading social networks worldwide based on active users. As of the fourth quarter of 2019, Twitter had 152 million monetizable daily active users worldwide [5]. This is one of the popularly used social media not only for ordinary people but also for politicians, media persons, economists, and many more. Twitter data can, therefore, be used for OM for any issue. It has been observed that people often use social media so frequently in case of any global issue or EID like COVID-19. Currently, the world is facing global crises because of the COVID-19 virus, and most of the countries are in lockdown, people are isolated in their home to save their lives. During this lockdown, they are frequently using many social media to post their views and opinions along with twitter, and hence it becomes the largest platform for COVID-19. EID-related issue is a global issue which is being considered by the researchers. Many authors [1,6,7] have established a relationship of Twitter data to predict the sentiment of any EIDs. They also found that Twitter is the most reliable source of data to analyze public opinion, emotions, and views on any EIDs.

Many approaches of SA are in practice, but the two most reliable approaches: Lexicon-based and machine learning (ML)-based have been applied frequently along with a hybrid of Lexicon-based and ML-based. Lexicon-based approaches have been used and proposed by many authors, and are being used by the researchers in different domains to analyze their models. D'Andrea et al. [8] have described different approaches, methods, and tools of SA. Augustyniak et al. [9] have done a comparative study on Lexicon-based ensemble classification SA. SA is not only used for English language but also being used for other languages also. Syed et al. [10] and Rehman and Bajwa [11] have used the Lexicon-based approach for SA of Urdu text. In another research, Ref. [12] have used Arabic sentiment lexicon built through automatic lexicon expansion. In a similar paper Ref. [13], a sentiment lexicon was developed for SA of Saudi dialect tweets. Authors [14] have proposed to assess various methods for supporting an additional target language in Lexicon-based SA which automatically translate text into a reference language. Another variation of Lexicon-based approach is N-Gram lexicon on the basis of this authors [15] have proposed Senti-N-Gram lexicon. Another new method of SA has been proposed by Ref. [16] which is based on semantic similarity metric between text word and lexicon vocabulary that uses semantic similarity measure in combination with embedding representations. They have tested their approach with the help of eight different lexicon datasets.

On the other hand, authors have contributed toward applying ML approaches in various domains for SA. In the financial domain for stock market prediction and correlation [17–19], for a product review [20], for government decision making [21], election monitoring and prediction [22], forecasting cryptocurrency prices [23] list a few. A hybrid approach of ML for SA has also become popular among the researchers. Hassonah et al. [24] have used a hybrid of support vector machine (SVM) and an evolutionary approach for SA of various topics of Twitter. Pandey et al. [25] proposed a metaheuristic-based method based on K-means and Cuckoo search to find optimum cluster-heads from the sentiment contents of the Twitter dataset. Ensemble technique is another approach that combines more than one classifier into a single classifier to improve the performance of the classifier. Ankit [26] present ensemble classification for SA by combining three classifiers, namely Random Forest, Linear Regression, and SVM. They have tested this ensemble model with four different benchmark datasets and found better results as compared to each classifier. In one more papers, Ref. [27] have used a combination of ML and Lexicon-based approach for SA. Mukhtar et al. [28] have compared these two approaches of SA in their paper on Urdu SA in multiple domains and found that Lexicon-based approach outperforms ML approach. Based on reviewed papers, they have concluded that both approaches have their strength and drawbacks. In the case of Lexicon-based approach, coverage as important lexical features are ignored, acquiring a new set of lexical features along with their valence scores is a labor-intensive and time-consuming process. In contrast, the ML-based approach requires a massive amount of training-testing data, several features required in a training sample are sometimes challenging to obtain, and it is computationally expensive.

Because of COVID-19 pandemic, people are posting their emotions and opinions on Twitter and hence Twitter is flushing with COVID-19—related texts and therefore authors are started using Twitter data for SA of COVID-19 pandemic both in context of India and abroad. In a recent paper, Barkur et al. [29] have analyzed the sentiment of Indians because of the COVID-19 pandemic based on Twitter data during the lockdown. They found that there is negative sentiment in terms of many emotions like fear, disgust, and sadness about the lockdown; however, still, Indians are more positive toward the situation that arises because of COVID-19. The conclusion was made based on a total of 24,000 tweets collected from March 25th to March 28th, 2020. In one more recent paper, authors [30] have discussed sentiment of Indians because of COVID-19 pandemic in terms of many emotions and found that Indians are still positive and lockdown or other restrictions do not impact the psychological well-being of the population. In another similar research, Ref. [31] have found that sentiment are mostly positive indicating that Twitter users are at home but are not feeling nervousness because of COVID-19 pandemic and experiencing a new social opportunity.

This research work explores the sentiment of people of six most affected countries because of COVID-19 along with India as a particular case of lockdown using Lexicon-based approaches. These approaches were used to identify the human sentiment of these countries in terms of Negative, Neutral, or Positive. Comparative study of the status of sentiment among all these countries shows that the UK and France are having the highest negativity, followed by the USA. In addition, the sentiment study of India with particular reference to lockdown was done. The result shows that the negativity of Indians increased after lockdown and then decreased during lockdown 2.0. This research work also replicates the panic situation in terms of negativity worldwide because of COVID-19, and on the other hand, majority of population of the country are positive and might be experiencing new challenges of life [30,31].

The rest of the chapter is divided into the following sections along with subsections: Section 2 explains about proposed methodology and explains the phases involved in this while Section 3 explores discussion, and finally, the chapter has been concluded in Section 4.

2. Proposed methodology

This section will explain the overall process of the current research work, as shown in Fig. 14.1. In the first phase, we start with Twitter data collection of five most affected countries of the world as well as of India. Next, in phase II, data are preprocessed so that data can be used for model development. Phase III describes applying methods for SA, and the subsequent phase (Phase IV) compares the results of two different approaches used for SA in this research work. Details of phases are explained in more detail as below:

2.1 Data collection

There are many online social media platforms like Twitter, Facebook, blogs, WhatsApp, Instagram, etc., which are used widely by people to express their opinion on any issue. This research work uses Twitter data for SA on the global pandemic because of the COVID-19 virus. As compared to other social media platforms, twitter provides a simple way to extract tweets through its application program interface (API) and is known to be very practical and straightforward. Data from twitter has been extracted using TWINT API known as a twitter intelligence tool, which is an advanced scraping tool written in Python, and allows for scraping tweets from tweeter without using twitter API and is more reliable than other API. To analyze the sentiment of various countries, Twitter data of six countries have been downloaded, and details of which are as follows.

2.1.1 Global Twitter data

A balanced Tweeted data of six countries of duration March 15 to April 15, 2020, as shown in Fig. 14.2 have been extracted. Data are based on #COVID-19 and #coronavirus hashtags. These two hashtags consist of most of the tweets related to COVID-19 in all the countries. The frequency of tweets on the basis of the various words is shown in Fig. 14.3, and this figure shows that coronavirus has the highest frequency, followed by the word COVID-19, lockdown, and so on. On the other hand, Fig. 14.4 shows the word cloud showing most frequently used words in tweets.

2.1.2 Indian Twitter data

Data from the Twitter social media platform have also been collected from March 15, 2020, on the basis of two hashtags: #COVID-19 and #Coronavirus. A total of 1,19,495 tweets were extracted, which consists of 34,666, 39,176, and 45,673 tweets, respectively, from before the lockdown, during March 15 to March 24, 2020, after lockdown during March 25 to April 4, 2020, and lockdown 2.0 during April 15 to April 24, 2020. The frequency of tweets during all three periods is shown in Figs. 14.5–14.7, respectively. From Fig. 14.5, it can be clearly observed that frequency of tweets on March 22nd, 2020 has been increased to its highest value because of Janata Curfew announced by Indian Prime Minister Shri Narendra Modi on March 19, 2020, he urged Indians to remain indoors from morning 7 a.m. to 9 p.m. on April 22, 2020 to fight against COVID-19.

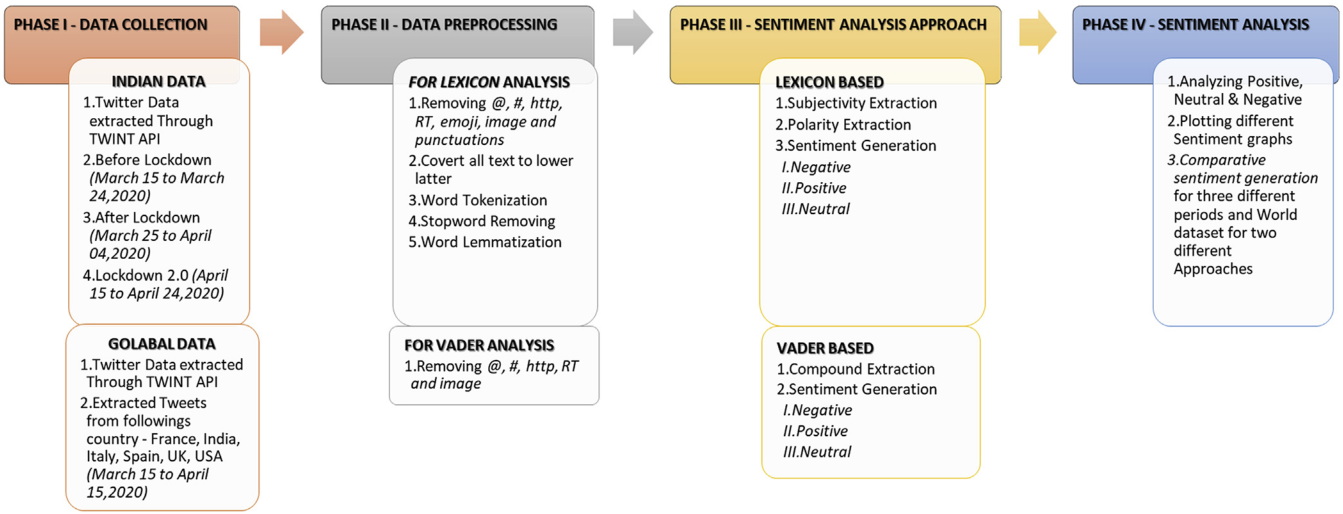

FIGURE 14.1 An architecture of proposed work for sentiment analysis.

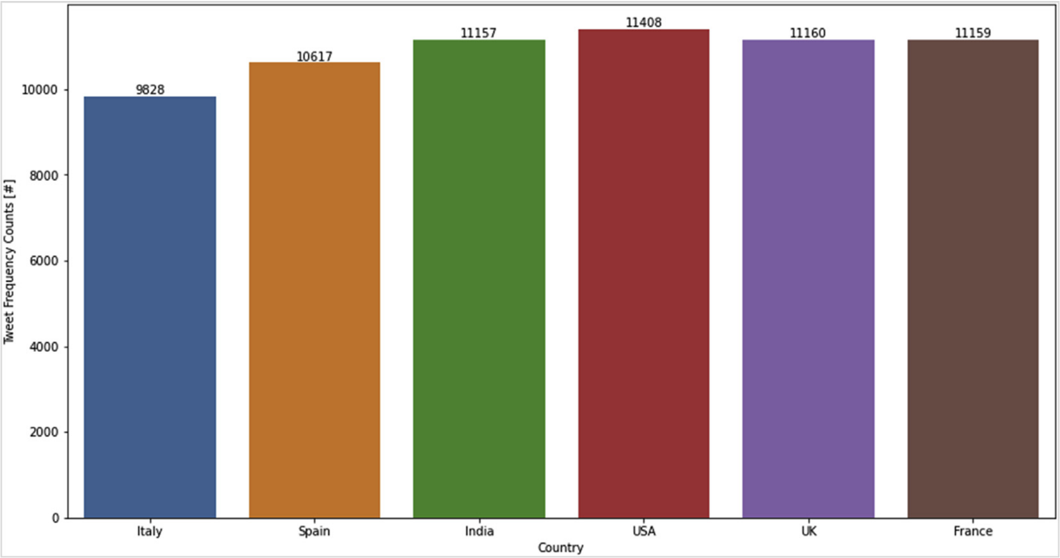

FIGURE 14.2 World wise tweets frequency count.

This was the first time when Indians are little panic about the negative impact of COVID-19 in India. Similarly, tweets are at their highest peak on March 03, 2020, after lockdown and on April 22, 2020, during lockdown 2.0, as shown in Figs. 14.6 and 14.7, respectively.

2.2 Data preprocessing

The unstructured nature of Twitter makes tweets so complicated and hence is a challenging task to remove these and to preprocess it before using. This research work has also applied data preprocessing to remove many irrelevant contents from Twitter data. In general, the following steps [32] involved in preprocessing of textual data like Twitter:

i. **Removing noise:** This step removes all noises like a hashtag, profile picture, retweet, emoji, URLs, user names, numbers, punctuations, white space, tags, @ sign, etc.
ii. **Change in lower case:** Tweets are then changed to lowercase letters.
iii. **Tokenization:** Tokenization is the process of NLP by which broad textual data is divided into smaller parts called tokens. It divides the sentence into a group of the word. It may be of two types: word tokenization and sentence tokenization. We have used word tokenization in this research work.
iv. **Removing stop words:** All articles, prepositions, and conjunctions like am, are, an, the, is, etc., are an example of stop words and are frequently used words in NLP. They do not carry much discriminative content. Also does not play an important role in sentiment classification process and is useless pieces of information found in any text like Tweet hence need to be removed [33].

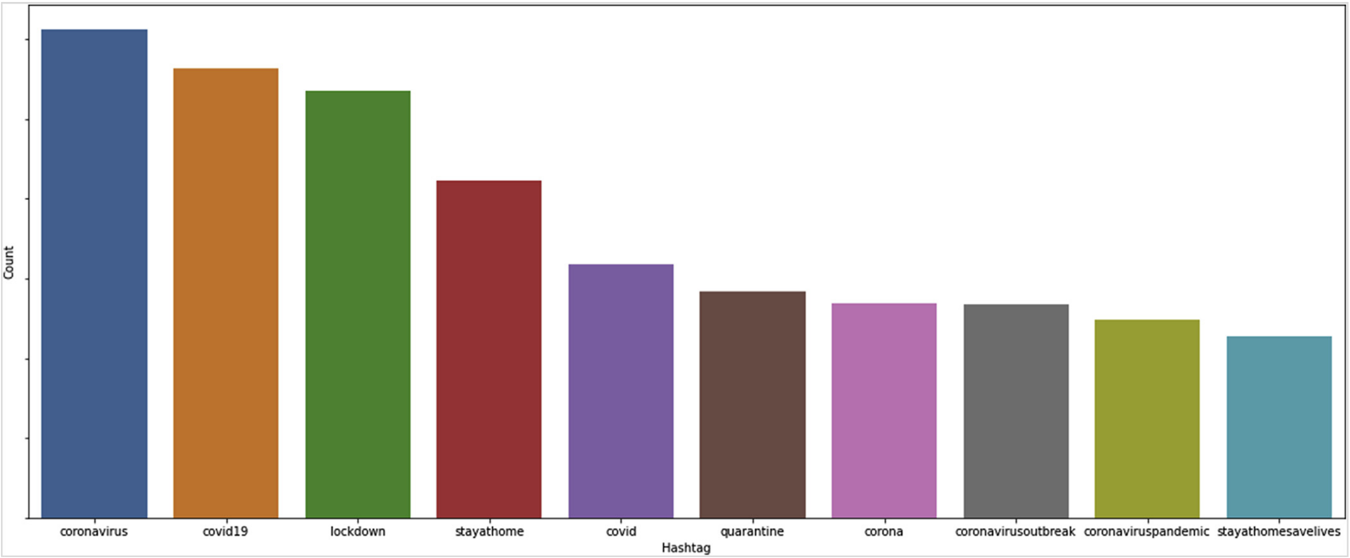

FIGURE 14.3 Top #Tag used in Twitter dataset.

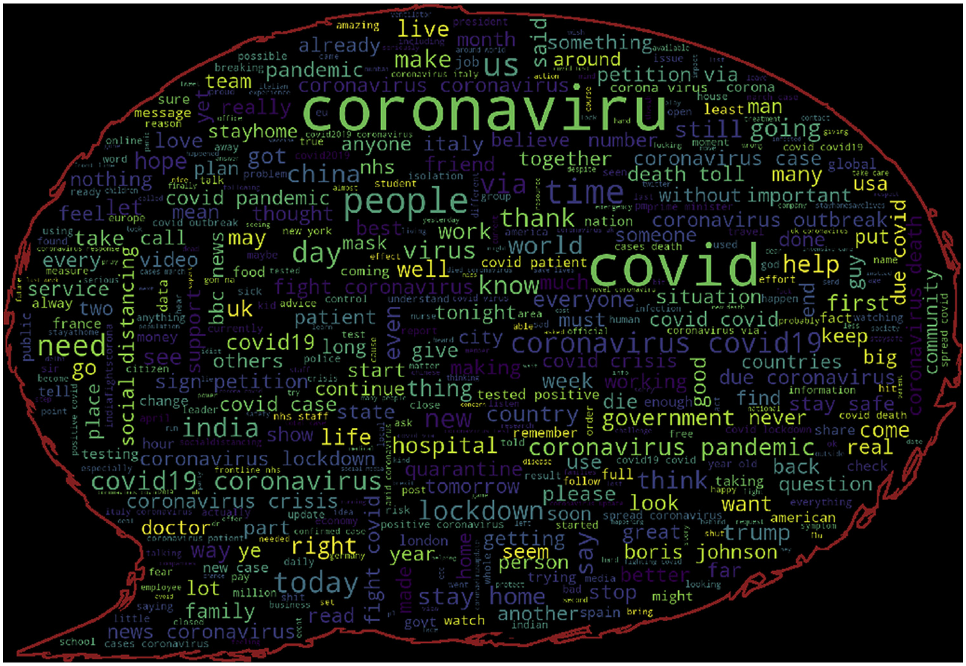

FIGURE 14.4 Word cloud of tweets.

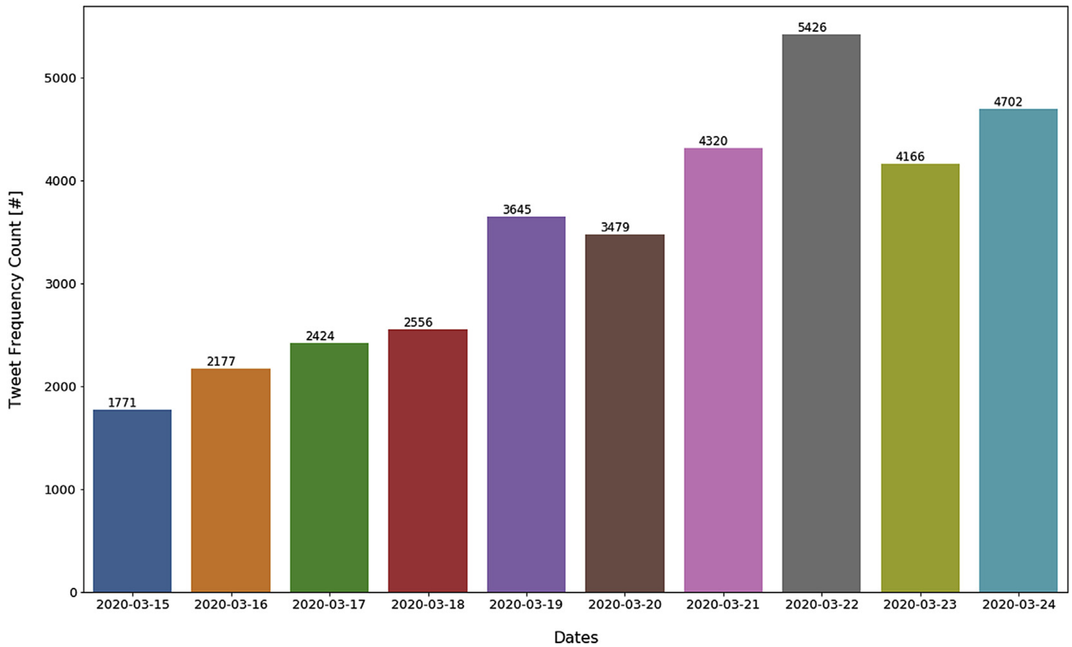

FIGURE 14.5 Frequency of tweet on the basis of dates (before lockdown).

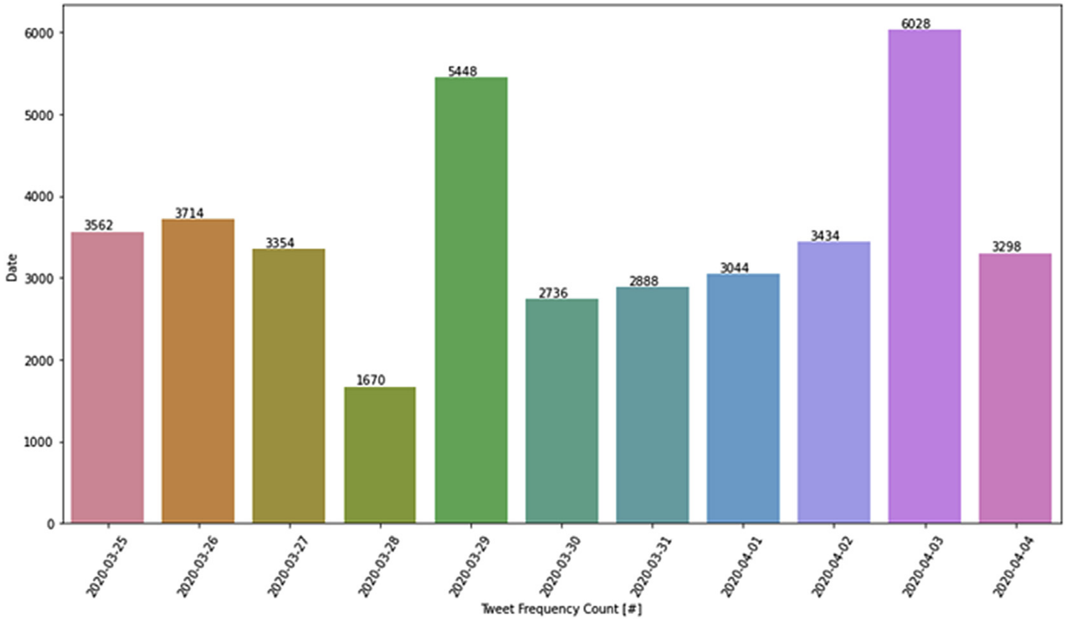

FIGURE 14.6 Frequency of tweet on the basis of dates (after lockdown).

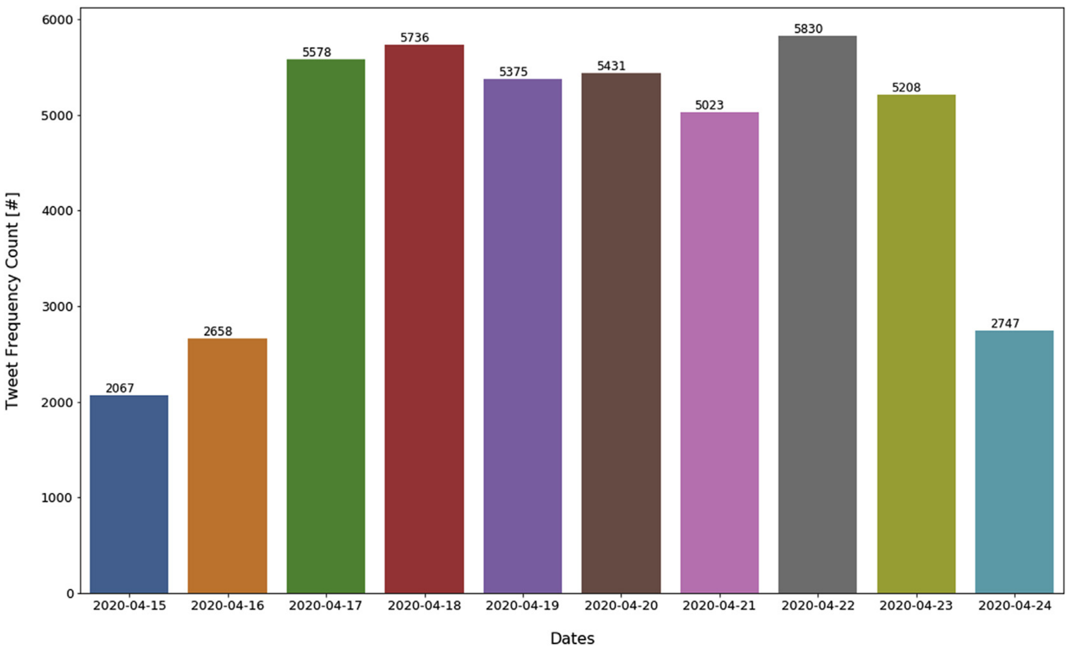

FIGURE 14.7 Frequency of tweet on the basis of dates (during lockdown 2.0).

- **v. Lemmatization:** It is a steaming which consolidates related words with the associated root. For example, a text document containing singular, plural, various tense, and other variations like reads, reading belonging to the same stem will be converted to the corresponding word read. The tweet might contain the singular or plural form of the same word, various tenses, and others. Lemmatization is a type of steaming which uses the specific part of speech to determine the root form of a word. The normalization rules depend on the part of speech, and therefore, they are highly language-specific. It is sometimes considered different from stemming, in that it goes beyond the simple stripping rules and uses the morphological roots of the words [33].

2.3 Sentiment analysis approach

There are three popular approaches widely used for SA: Lexicon-based, machine learning–based, and hybrid based. All these approaches involve various computational steps. The Lexicon-based approach determines polarity or sentiment to classify tweets in three different categories: Negative, Neutral, and Positive (Negative and Positive if desired). On the other hand, the ML-based approach is also popular among the researchers, especially when a large number of labeled data samples are available and is used in a supervised manner. This section describes two Lexicon-based approaches: Simple lexicon and VADER-based lexicon. A simple Lexicon is simple one which is based on polarity score of the word and classify the sentiment as negative, neutral, and positive while on the other hand VADER is also a Lexicon-based approach but with little different approach. VADER-based approach scores each word along with emoji and uses combination of a sentiment lexicon [34].

2.3.1 Lexicon-based approach

The Lexicon-based approach [9] assumes that sentiment is related to the presence of certain words or phrases in the document: raw text in a processed structural representation. A lexicon is a set of features that have an assigned sentiment value. It is basically used as a predefined list of words called the dictionary, and each word has many synonyms through which word is associated with. Popular lexicons are WordNet, SenticNet, etc. The tweet also contains emoji called emoticons and is the easiest way to identify sentiment associated with a tweet, is also an example of the lexicon.

In this research work, we have used polarized lexicon [9], in which the word "w" is assigned with a numeric value 1,0,−1 for Positive, Neutral, or Negative sentiment. The polarity of the document $d = \{w_1, w_2, \text{———}, w_n\}$ is calculated based on detecting occurrences of the sentiment words "w" from the lexicon "l" in the document d. Pos(l,d) and neg(l,d) are the positive and negative words from "l" that occur in "d." And sum(l,d) = pos(l,d)-neg(l,d) then sentiment orientation $s_1(d)$ of a document "d" under a polarized lexicon "l" is assigned using the following formulas:

$$s_1(d) = 1 \quad \text{if } \mathrm{sum}(l, d) > 0 = 0 \quad \text{if } \mathrm{sum}(l, d) = 0 = -1 \quad \text{if } \mathrm{sum}(l, d) < 0$$

Researchers first create a sentiment lexicon through compiling sentiment word lists such as manual approaches, lexical approaches, and corpus-based approaches, then determine the polarity score of the given review based on the Negative, Neutral, or Positive indicators which are identified in the lexicon [33].

2.3.2 VADER-based approach

Valence Aware Dictionary and Entiment Reasoner (VADER) is a lexicon and rule-based SA tool used for general SA [34] that is specifically attuned to sentiments expressed in social media. It is fully open-sourced under the (MIT License). In this approach, the compound score is computed by summing the valence scores of each word in the lexicon, adjusted according to the rules, and then normalized to be between -1 (most extreme negative) and $+1$ (most extreme positive). On the basis of compound score calculated, sentiment can be classified as follow: positive sentiment: compound score ≥ 0.05, neutral sentiment: (compound score > -0.05) and (compound score < 0.05) and negative sentiment: compound score ≤ -0.05.

2.4 Sentiment analysis

The experiment was carried out with open source Python, which is an interpreted, object-oriented, and high-level programming language that provides rich libraries. In this research work, we have used NLTK for NLP, Pandas for Data analytics, Textblob for text processing, Matplotlib, and Seaborn for data visualization. Vyas and Uma [35] compared various tools like Rapid Miner, SAS text miner, and others for SA. However, authors have not considered Python an open-source ML tool for SA. It is a widely used tool, and popular among the researchers nowadays.

2.4.1 Result of Lexicon-based approach

Preprocessed twitter data are used to find out polarity based on lexicon and sentiment in form of Negative, Neutral, or Positive calculated as explained in Section 2.3.1. The calculated subjectivity and polarity score based on lexicon have been shown in the form of line and bar graphs country wise in Fig. 14.8. This figure shows that the polarity score is highest in the case of India followed by Italy and Spain. The overall sentiment score of all the countries has been shown in Fig. 14.9 while country wise sentiment score is shown in Fig. 14.10.

2.4.2 Result of VADER-based approach

Similarly, VADER-based approach was applied to the same data. While feeding data to the VADER-based approach, we did not remove emoji as this approach has a lexicon of emoji. This approach is able to identify sentiment based on word as well as emoji, and hence sentiment calculated using this approach may be more reliable. As stated in Section 2.3.2, this approach identifies sentiment based on the compound score. However, it also calculates Negative, Neutral, as well as Positive score for in-depth analysis of sentiment. Fig. 14.11 shows all the above scores in graphical form, while Fig. 14.12 depicts the overall score of three sentiments. Country-wise sentiment scores in form Negative, Neutral, and Positive are also shown in Fig. 14.13.

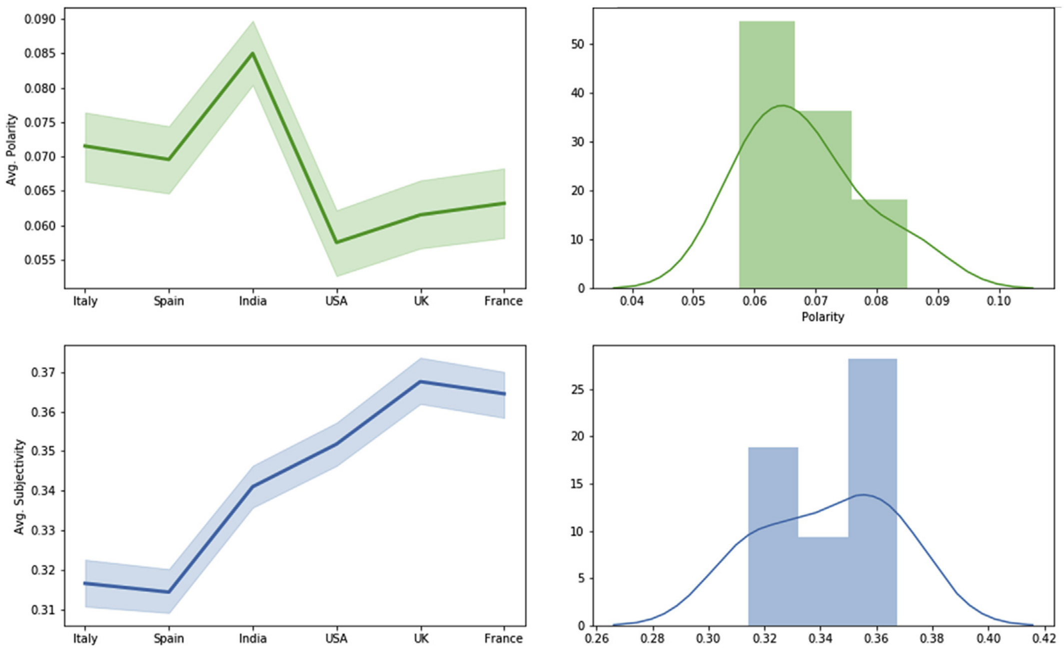

FIGURE 14.8 Country wise score of polarity and subjectivity and its corresponding bar graph using Lexicon-based approach.

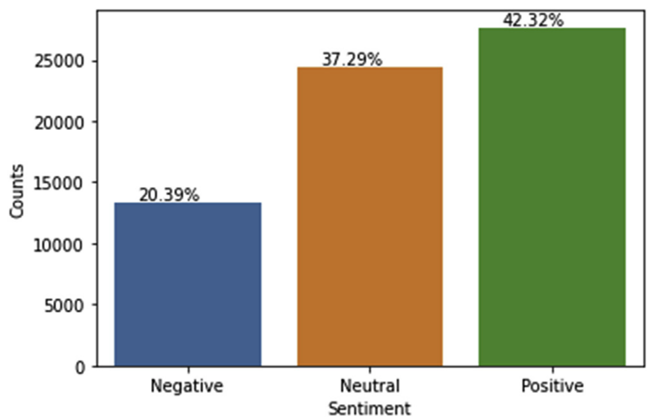

FIGURE 14.9 Sentiment of all countries calculated on the basis of polarity using Lexicon-based approach.

2.4.3 Sentiment analysis of India during and before lockdown

Preprocessed data, as explained in Section 2.2, are used to analyze Indian sentiment before and after lockdown. Data of three different periods are analyzed separately using the Lexicon-based approach, and the results are shown in Fig. 14.14. A negative score is helpful and used to understand the sentiment of humans because of COVID-19 and can

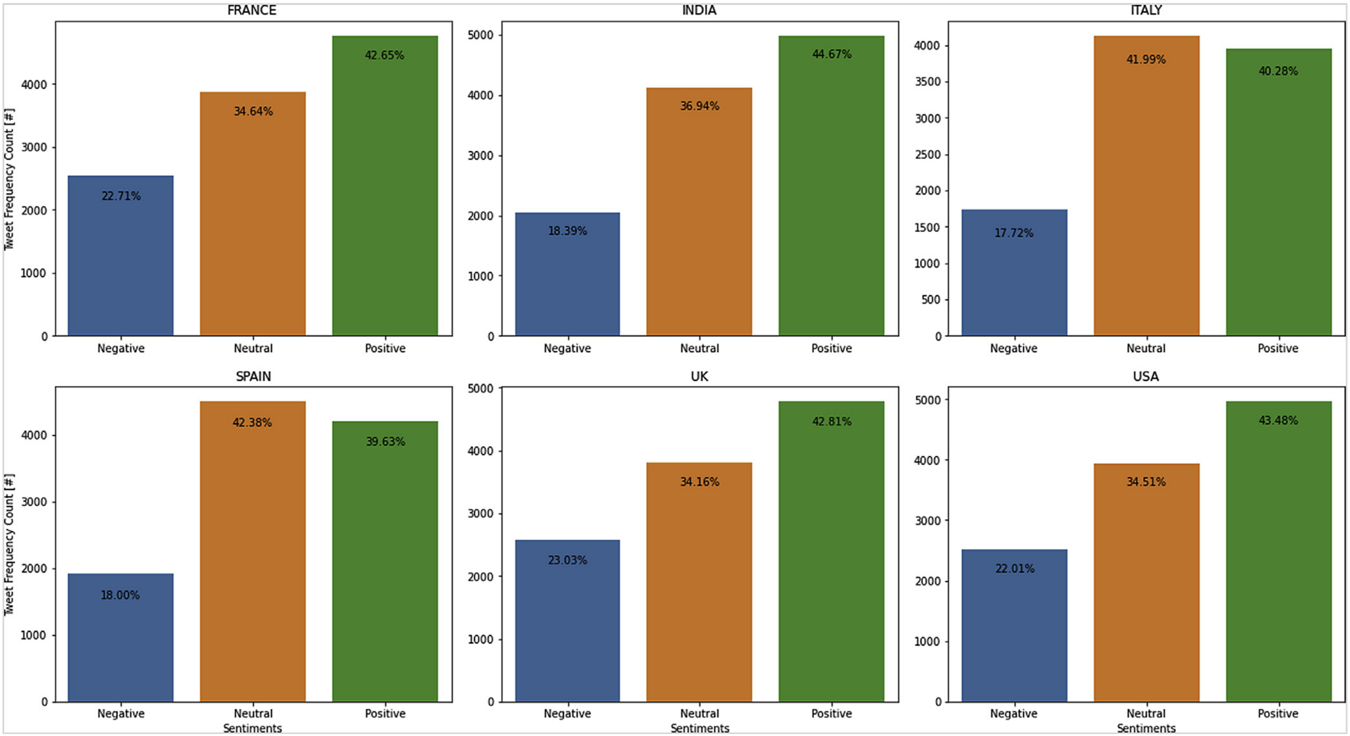

FIGURE 14.10 Country wise sentiment calculated on the basis of polarity using Lexicon-based approach.

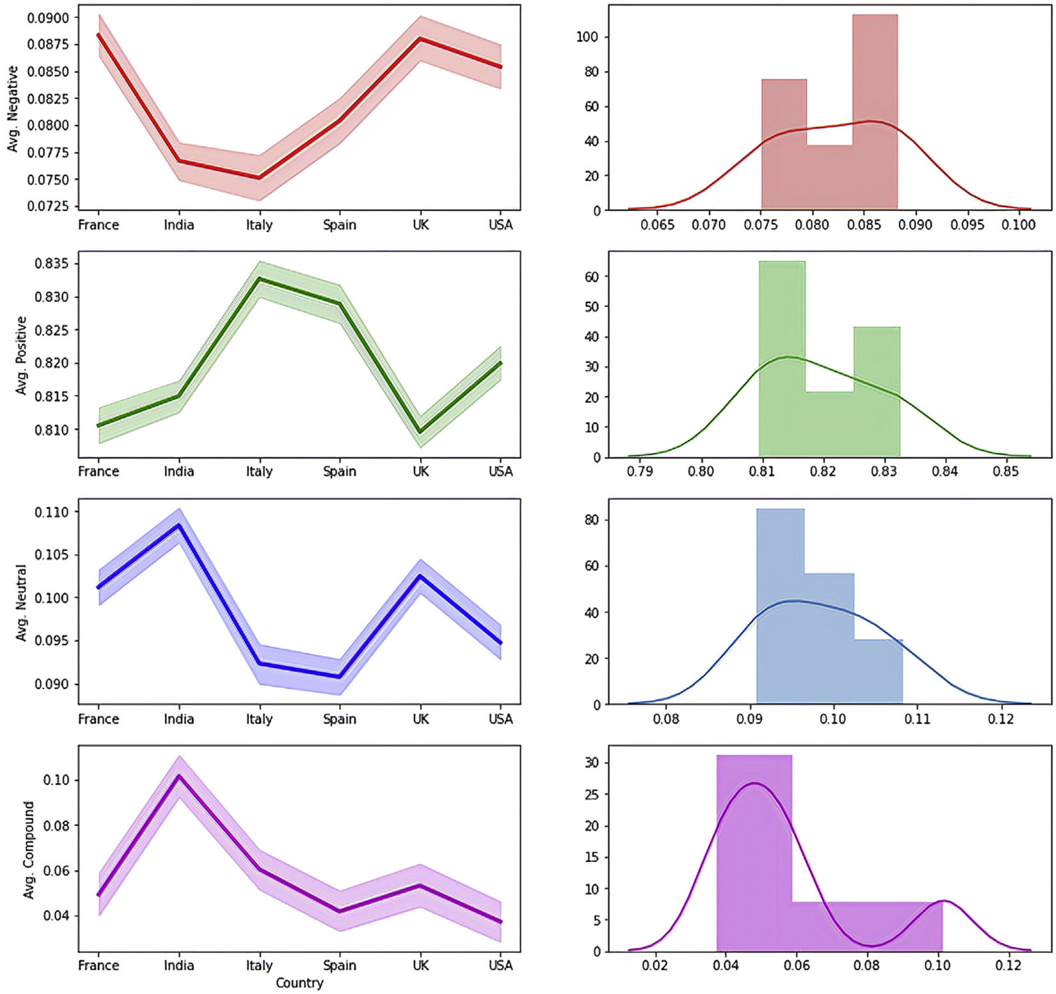

FIGURE 14.11 Country wise score of positive, negative, neutral, and average using VADER approach.

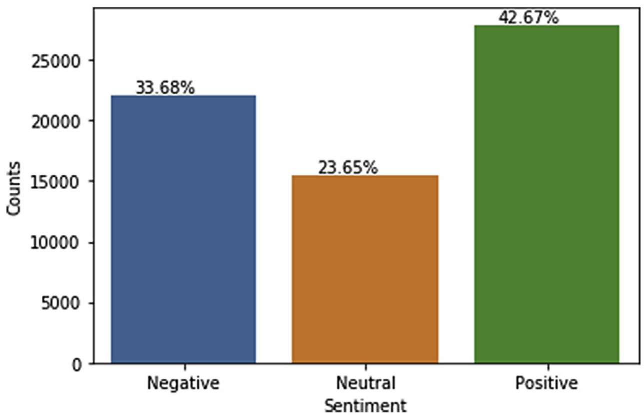

FIGURE 14.12 Sentiment of all countries calculated on the basis of compound score using VADER-based approach.

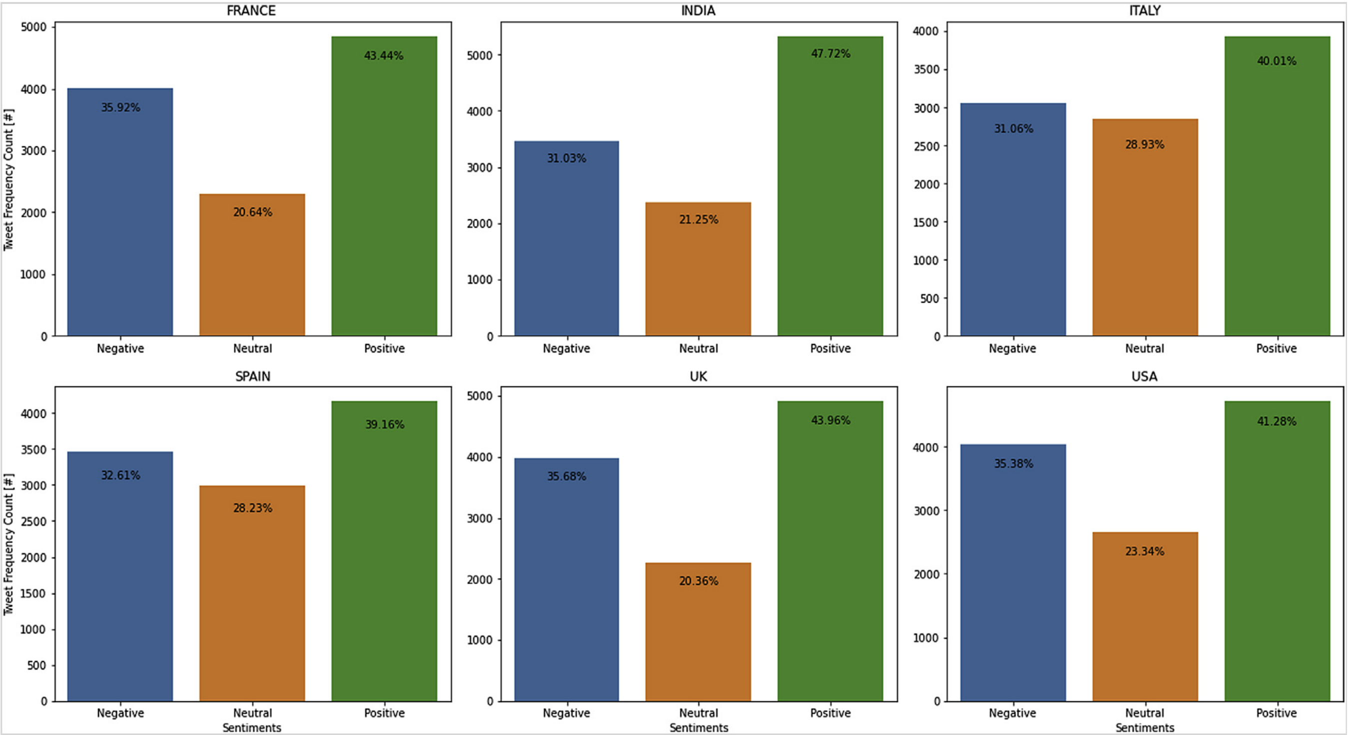

FIGURE 14.13 Country wise sentiment calculated on the basis of compound score using VADER-based approach.

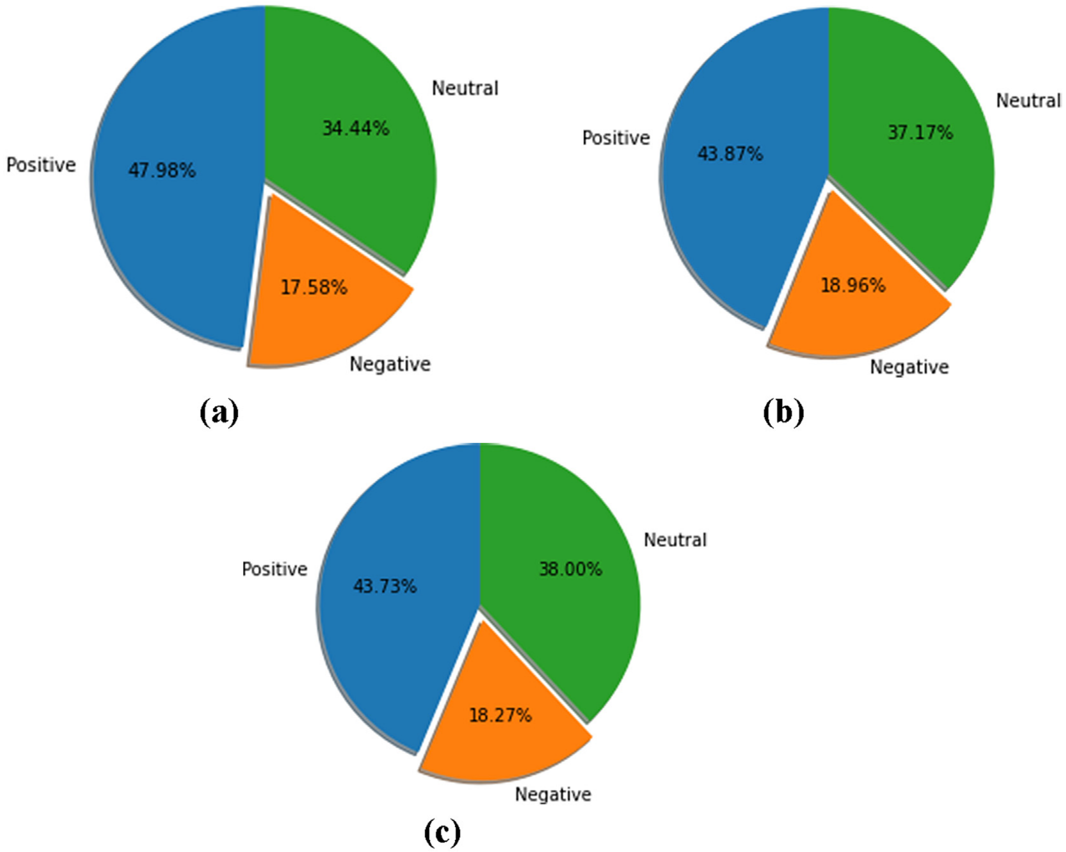

FIGURE 14.14 Sentiment analysis score (A) before lockdown (B) after lockdown (C) lockdown 2.0.

impact human psychological parameters. A higher percentage of Positive sentiment is better for the better health status of any country. However, it was observed during data analysis that there is enough percentage of people who are neutral during this period; these tweets actually do not contain any negative or positive word to be utilized for data SA. The results of Indian sentiment revealed that negative sentiment is increasing after lockdown from 17.58% to 18.96% (Fig. 14.14A and B) and decreasing from 18.96% to 18.27% (Fig. 14.14B and C) during lockdown 2.0. These trends are just indicative that Indians are little sentiment just after lockdown announced by Indian Prime Minister Shri Narendra Modi due to worldwide critical situation and slightly decreased afterward during lockdown 2.0.

3. Discussion

Sentiment score may vary as per the lexicon data used for calculating polarity or any other numeric score and hence will impact on the total percentage of sentiments. In this

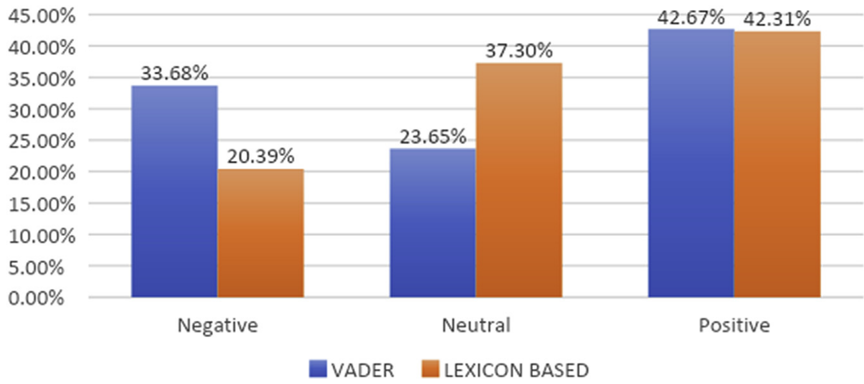

FIGURE 14.15 Comparative sentiment (In %) of all countries together.

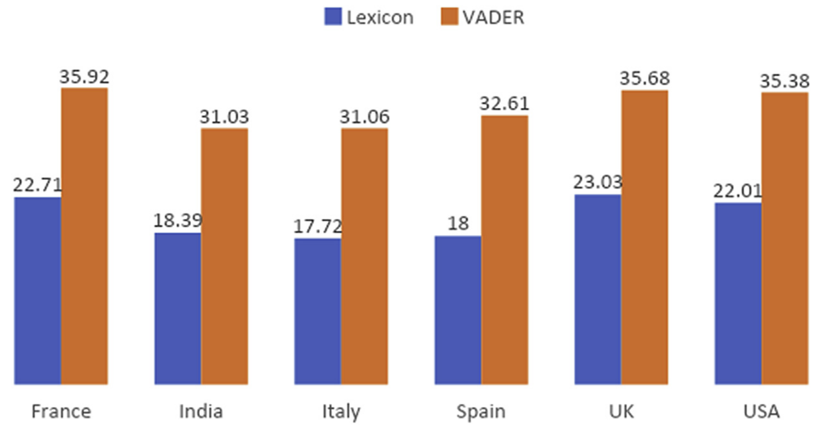

FIGURE 14.16 Country wise comparative negative sentiment (In %).

research work, two different approaches: Lexicon-based and VADER-based have been used on different lexicons (Dictionary). As stated, the Lexicon-based approach calculates polarity based on words of the dictionary while the VADER-based approach uses its lexicon with emoji as well. Tweet extracted for SA also contains emoji along with other noises. Lexicon words, along with emoji, may produce better sentiment scores. This can be seen in Fig. 14.15, where the sentiment percentage is higher in the case of the VADER-based approach as compared to the Lexicon-based approach for tweets of all the countries together. For example, negative sentiment is 20.39% in the case of the Lexicon-based approach. In comparison, it is 33.68% in the case of the VADER-based approach; similar higher results were found for Neutral and Positive sentiments also. A country-wise comparison of Negative sentiment was also made and depicted in Fig. 14.16, which reflects the negativity score using two approaches. Based on the Lexicon-based approach, the UK is having the highest negativity (23.03%), followed by France (22.71%), and the USA (22.01%) and France have the highest negativity (35.92%),

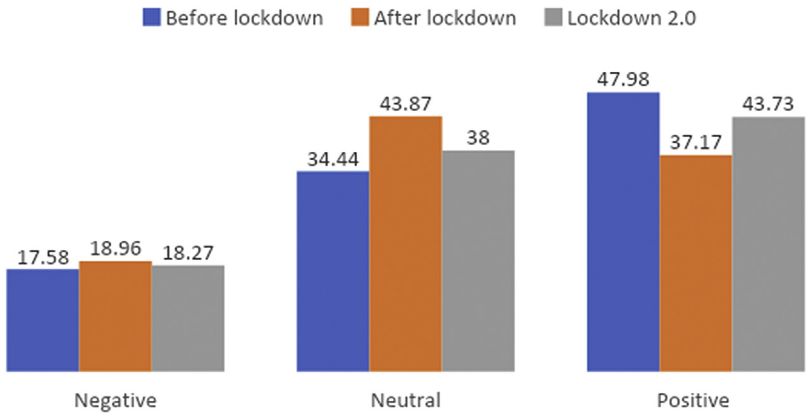

FIGURE 14.17 Comparative Indian sentiment (In%) for three different segments (before lockdown, after lockdown, and lockdown 2.0) using Lexicon-based approach.

followed by the UK (35.68%) and USA (35.38%) based on the VADER-based approach. However, there are slight variations at rank four in different approaches; based on the Lexicon-based approach, India has the highest negative sentiment of 18.39%, followed by Spain (18%) and Italy (17.72%). In comparison, Spain stood first with 32.61% of negativity, followed by Italy (31.06%) and India (31.03%) based on the VADER-based approach.

As stated, SA was also done with special reference to lockdown in India. A comparative result of all three different periods: before lockdown, after lockdown, and lockdown 2.0, has been elaborated in Section 2.1.2 and is depicted in Fig. 14.17. One can analyze through this comparative bar chart that negative sentiment of Indians is increasing from before lockdown (17.58%) to after lockdown (18.96%) and then decreasing (18.27%) by 0.69% during lockdown 2.0. On the other hand, the Neutral sentiment of people is decreasing (38%), and positive sentiment is increasing (43.73%) during lockdown 2.0. It might be because of many reasons; one of the reasons may be comparatively a smaller number of cases of COVID-19 in India and also a continuation of the strict countrywide implementation of lockdown.

4. Conclusion

Twitter data has a lot of potential and features to analyze the sentiment of people during global pandemic like COVID-19. SA because of COVID-19 needs to be analyzed in the current panic situation so that the necessary action of psychological treatment may be taken, appropriate precaution measures can be implemented, and misinformation can be reduced.

This research work compares the sentiment of six countries that are most affected because of COVID-19 along with India as a special case. Twitter data of all six countries, namely India, the USA, Spain, Italy, France, and the UK from March 15, to April 15, 2020, were extracted from Twitter social network and used to identify sentiment as Negative, Neutral, or Positive using Lexicon-based and VADER-based approaches. Simulated results show that negativity exists in all the countries because of COVID-19. The percentage of sentiment is higher in the case of the VADER-based approach as compared to the Lexicon-based approach; this is because of a more sophisticated lexicon database of VADER along with emoji. It has been observed uniformly that Negative, Neutral, and Positive sentiments are almost the same in order in case of both the approaches except few variations at different positions. Out of six countries considered for the SA, the UK has the highest negativity of 23.03%, followed by France with 22.71%, and the USA with 22.01%. India is having a negativity of 18.39% using a simple Lexicon-based approach while it is 35.92% in the case of France, 35.68% in the case of UK, and 35.38% in the case of the USA with least negativity of 31.03% of India. Furthermore, comparative detail analysis of India has also been done based on Twitter data collected before and after lockdown using a simple Lexicon-based approach, and it has been observed that negativity is increasing after lockdown and slightly decreased during lockdown 2.0. Furthermore, it has been observed that majority of the population is having positive sentiment which indicates that people are having positive feeling and facing challenges as well as opportunity during this restricted time of COVID-19.

Deep Learning (DL) methods are widely used in machine vision along with SA. Due to its inherent capability of learning behavior of high-dimensionality data and automatic feature extraction, and is well suited for SA based on emoji. DL methods like CNN and LSTM are widely implemented by the researchers nowadays, so, in future, individual DL methods or hybrid methods can be used to asses sentiment in global scenarios, especially for the top 6 to 10 most affected countries because of COVID-19. Also, in more depth, sentiment can be analyzed on the basis of many emotions such as joy, sadness, hate, etc., known as emotion analysis. Furthermore, text data is high dimensional, and many words are irrelevant and do not play any vital role in the classification process. So, to improve the performance of the sentiment classifier, advanced feature selection techniques may be incorporated after extracting features from text data.

References

[1] L. Tang, B. Bie, S.E. Park, D. Zhi, Social media and outbreaks of emerging infectious diseases: a systematic review of literature, Am. J. Infect. Contr. 46 (2018) 962−972. Available from: https://doi.org/10.1016/j.ajic.2018.02.010.

[2] M.A. Shereen, S. Khan, A. Kazmi, N. Bashir, R. Siddique, COVID-19 infection: origin, transmission, and characteristics of human coronaviruses, J. Adv. Res. 24 (2020) 91−98. Available from: https://doi.org/10.1016/j.jare.2020.03.005.

[3] reportCOVID-19 Situation Reports [Internet], n.d. Available from: https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports/. (Accessed 14 May 2020).

[4] M.V. Mäntylä, D. Graziotin, M. Kuutila, The evolution of sentiment analysis—a review of research topics, venues, and top cited papers, Comput. Sci. Rev. 27 (2018) 16−32. Available from: https://doi.org/10.1016/j.cosrev.2017.10.002.

[5] Twitter-Statistics & Facts Statista [Internet], n.d. Available from: https://www.statista.com/topics/737/twitter/. (Accessed 16 May 2020).

[6] G.A. Ruz, P.A. Henríquez, A. Mascareño, Sentiment analysis of Twitter data during critical events through Bayesian networks classifiers, Future Generat. Comput. Syst. 106 (2020) 92–104. Available from: https://doi.org/10.1016/j.future.2020.01.005.

[7] P.S. Devi, S. Karthika, #CycloneGaja-rank based credibility analysis system in social media during the crisis, Procedia Comput. Sci. 165 (2019) 684–690. Available from: https://doi.org/10.1016/j.procs.2020.01.064.

[8] A. D'Andrea, F. Ferri, P. Grifoni, T. Guzzo, Approaches, tools and applications for sentiment analysis implementation, Int. J. Comput. Appl. 125 (2015) 26–33. Available from: https://doi.org/10.5120/ijca2015905866.

[9] L. Augustyniak, P. Szymánski, T. Kajdanowicz, W. Tuliglowicz, Comprehensive study on lexicon-based ensemble classification sentiment analysis, Entropy 18 (2016) 4. Available from: https://doi.org/10.3390/e18010004.

[10] A.Z. Syed, M. Aslam, A.M. Martinez-Enriquez, Lexicon based sentiment analysis of Urdu text using SentiUnits, in: Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), vol. 6437, Springer-Verlag, LNAI, Berlin, Heidelberg, 2010, pp. 32–43. Available from: https://doi.org/10.1007/978-3-642-16761-4_4.

[11] Z.U. Rehman, I.S. Bajwa, Lexicon-based sentiment analysis for Urdu language, in: 6th International Conference on Innovative Computing Technology, INTECH 2016 2017, 2016, pp. 497–501. Available from: https://doi.org/10.1109/INTECH.2016.7845095.

[12] M. Youssef, S.R. El-Beltagy, MoArLex: an Arabic sentiment lexicon built through automatic lexicon expansion, Procedia Comput. Sci. 142 (2018) 94–103. Available from: https://doi.org/10.1016/j.procs.2018.10.464.

[13] A. Al-Thubaity, Q. Alqahtani, A. Aljandal, Sentiment lexicon for sentiment analysis of Saudi dialect tweets, Procedia Comput. Sci. 142 (2018) 301–307. Available from: https://doi.org/10.1016/j.procs.2018.10.494.

[14] A. Hogenboom, B. Heerschop, F. Frasincar, U. Kaymak, F. De Jong, Multi-lingual support for lexicon-based sentiment analysis guided by semantics, Decis. Support Syst. 62 (2014) 43–53. Available from: https://doi.org/10.1016/j.dss.2014.03.004.

[15] A. Dey, M. Jenamani, J.J. Thakkar, Senti-N-Gram: an n-gram lexicon for sentiment analysis, Expert Syst. Appl. 103 (2018) 92–105. Available from: https://doi.org/10.1016/j.eswa.2018.03.004.

[16] O. Araque, G. Zhu, C.A. Iglesias, A semantic similarity-based perspective of affect lexicons for sentiment analysis, Knowl. Base Syst. 165 (2019) 346–359. Available from: https://doi.org/10.1016/j.knosys.2018.12.005.

[17] T.M. Nisar, M. Yeung, Twitter as a tool for forecasting stock market movements: a short-window event study, J. Finance & Data Sci. 4 (2018) 101–119. Available from: https://doi.org/10.1016/j.jfds.2017.11.002.

[18] Y. Ni, Z. Su, W. Wang, Y. Ying, A novel stock evaluation index based on public opinion analysis, Procedia Comput. Sci. 147 (2019) 581–587. Available from: https://doi.org/10.1016/j.procs.2019.01.212.

[19] A. Groß-Klußmann, S. König, M. Ebner, Buzzwords build momentum: global financial Twitter sentiment and the aggregate stock market, Expert Syst. Appl. 136 (2019) 171–186. Available from: https://doi.org/10.1016/j.eswa.2019.06.027.

[20] P.R. Kanna, P. Pandiaraja, An efficient sentiment analysis approach for product review using Turney algorithm, Procedia Comput. Sci. 165 (2019) 356–362. Available from: https://doi.org/10.1016/j.procs.2020.01.038.

[21] E. Georgiadou, S. Angelopoulos, H. Drake, Big data analytics and international negotiations: sentiment analysis of Brexit negotiating outcomes, Int. J. Inf. Manag. 51 (2020) 102048. Available from: https://doi.org/10.1016/j.ijinfomgt.2019.102048.

[22] B. Bansal, S. Srivastava, On predicting elections with hybrid topic based sentiment analysis of tweets, Procedia Comput. Sci. 135 (2018) 346–353. Available from: https://doi.org/10.1016/j.procs.2018.08.183.

[23] O. Kraaijeveld, J. De Smedt, The predictive power of public Twitter sentiment for forecasting cryptocurrency prices, J. Int. Financ. Mark. Inst. Money 65 (C) (2020). Available from: https://doi.org/10.1016/j.intfin.2020.101188.

[24] M.A. Hassonah, R. Al-Sayyed, A. Rodan, A.M. Al-Zoubi, I. Aljarah, H. Faris, An efficient hybrid filter and evolutionary wrapper approach for sentiment analysis of various topics on Twitter, Knowl. Base Syst. 192 (2020) 105353. Available from: https://doi.org/10.1016/j.knosys.2019.105353.

[25] A. Chandra Pandey, D. Singh Rajpoot, M. Saraswat, Twitter sentiment analysis using hybrid cuckoo search method, Inf. Process. Manag. 53 (2017) 764–779. Available from: https://doi.org/10.1016/j.ipm.2017.02.004.

[26] S.N. Ankit, An ensemble classification system for Twitter sentiment analysis, Procedia Comput. Sci. 132 (2018) 937–946. Available from: https://doi.org/10.1016/j.procs.2018.05.109.

[27] L. Zhang, R. Ghosh, M. Dekhil, M. Hsu, B. Liu, Combining Lexicon-Based and Learning-Based Methods for Twitter Sentiment Analysis, Havlett packord development company, 2011. L.P. Report Number: HPL-2011-89.

[28] N. Mukhtar, M.A. Khan, N. Chiragh, Lexicon-based approach outperforms supervised machine learning approach for Urdu sentiment analysis in multiple domains, Telematics Inf. 35 (2018) 2173–2183. Available from: https://doi.org/10.1016/j.tele.2018.08.003.

[29] G. Barkur, Vibha, G.B. Kamath, Sentiment analysis of nationwide lockdown due to COVID 19 outbreak: evidence from India, Asian J. Psychiatry 51 (2020) 102089. Available from: https://doi.org/10.1016/j.ajp.2020.102089.

[30] Vibha, A.N. Prabhu, G.B. Kamath, D.V. Pai, Keeping the country positive during the COVID 19 pandemic: evidence from India, Asian J. Psychiatry 51 (2020) 102118. Available from: https://doi.org/10.1016/j.ajp.2020.102118.

[31] M. Bhat, M. Qadri, N.-A. Beg, M. Kundroo, N. Ahanger, B. Agarwal, Sentiment analysis of social media response on the Covid19 outbreak, Brain Behav. Immun. 87 (2020) 136–137. Available from: https://doi.org/10.1016/j.bbi.2020.05.006.

[32] M. Pandey, R. Williams, N. Jindal, A. Batra, Sentiment analysis using lexicon based approach, IITM J. Manag. & IT 10 (1) (2019).

[33] C.C. Aggarwal, Machine Learning for Text, Springer International Publishing, 2018. Available from: https://doi.org/10.1007/978-3-319-73531-3.

[34] C.J. Hutto, E. Gilbert, VADER: a parsimonious rule-based model for sentiment analysis of social media text, in: Proceedings of the 8th International Conference on Weblogs and Social Media, ICWSM 2014, 2014, pp. 216–225.

[35] V. Vyas, V. Uma, An extensive study of sentiment analysis tools and binary classification of tweets using Rapid Miner, Procedia Comput. Sci. 125 (2018) 329–335. Available from: https://doi.org/10.1016/j.procs.2017.12.044.