Check for updates

Opinion Mining and Tweet Analysis Using Topic Modeling by LDA with BERT and GLOVE Embedding

Ashwini Pachlore(✉) and Vrishali Chakkarwar

Government College of Engineering, Aurangabad, Maharashtra, India
ashwinipachlore@gmail.com, vrush.a143@gmail.com

Abstract. An example of natural language processing is opinion mining. A system is created to gather and process opinions about a product expressed in blog posts, comments, reviews, or tweets in order to assess public sentiment. Data preprocessing is used in this study to clean tweets and remove any punctuation, special symbols, hashtags, and URLs. Topic modelling and LDA are used to extract the themes from the corpus of gathered topics. The Principal Components Analysis (PCA) techniques are described in this article as dimensional reduction strategies with the goal of identifying the fewest possible Principal Components (PCs) that can help achieve the best classification performance. K-means clustering is a technique used to group together comparable words in tweets, along with cluster analysis. Utilizing the glove model, words are represented as vectors. The tweets can be grouped using k-means. Text input is sequentially read using the BERT paradigm. LSTM (long short-term memory) is used to anticipate sequences. In order to maintain the semantic association between words in a low-dimensional embedding space, Word2vec is a potent and effective word embedding technique. It is capable of handling tiny text corpora with a few million unique words (also called as vocabulary). "T- SNE" that places each datapoint on a two- or three-dimensional map to view high-dimensional data. In order to more effectively validate models and outcomes, other performance metrics like accuracy, F1-score, and a confusion matrix were also used.

Keywords: Opinion Mining · Glove Model · BERT Model · LSTM Model · Principal Component Analysis · K-means

1 Introduction

Topic modelling is a probabilistic statistical approach that reveals the hidden theme organization in document collections and makes analyzing massive amounts of unlabeled text simple. By examining distinct patterns present in documents, the fundamental purpose of modelling is to identify patterns of words in text and discover hidden structural terms that run across the corpus [2]. Probabilistic clustering is another term for topic modelling. It's more reliable, and the results are usually more accurate. Grouping that is particularly difficult (e.g. k-mean clustering). Topic modelling differs from traditional

The original version of this chapter was revised: The typo in the first author name has been corrected. The correction to this chapter is available at
https://doi.org/10.2991/978-94-6463-196-8_55

© The Author(s) 2023, corrected publication 2023
R. Manza et al. (Eds.): ACVAIT 2022, AISR 176, pp. 660–673, 2023.
https://doi.org/10.2991/978-94-6463-196-8_50

clustering algorithms in that it assumes a distance measure between topics and allocates one topic to each text. Allocates a document to a group of subjects with varying degrees of importance Without making any. Assumptions about the distance, weights or probabilities are used of Compared to other subjects.

There are many topic models available, but the latent Dirichlet allocation (LDA) model proposed by David Blei, Andrew Ng and it is the most extensively utilized topic models [1]. Opinion mining is a type of natural language processing that is used to track how people feel about a product. Opinion mining, sometimes referred as sentiment analysis, is the process of developing a system for gathering and analyzing client opinion expressed in blog posts, comments, reviews, or tweets [3] Typically sentiment polarity is classified into three groups. Good, bad, and neutrality. Our text analysis based on the meaning of words in a given context [4]. The model in this work is built on time consuming studies using LSTM and Global Vector's 300-dimensional word embedding features (Glove) [7]. Using the distributional features learned from a large sample of language corpora and creating low dimensional vectors, word embedding allows for identical representation for words with similar meaning. Word2vec and Glove are the most well-known approaches for creating word embeddings in our setting [5]. The first step in the analysis of a tweet is to preprocess the data by removing stop words, special symbols, punctuation, hashtags, and URL, among other things. Next, we create word embeddings and vector forms of textual data from the dataset. Many Natural Language Processing (NLP) tasks benefit from these vector representations of words. So far, Word2Vec and Glove are the two most well-known representation models in this area. Glove is an unlabeled data tool for obtaining word vector representations. The resulting representations highlight intriguing linear substructures of the word vector space, which are trained using aggregated global word to word co-occurrence information from a corpus. Glove word embedding is a global log to bilinear regression model that uses co-occurrence and matrix factorization to produce vectors. The classification of texts using LSTM in this analysis because the structure of LSTM is a sequence in which an integrated whole or cannot be cut, similar to the structure of text documents, which if cut would change the meaning of the phrase.

2 Literature Survey

On the basis of their characteristics, Shrivatava, A., Mayor, S. & Pant, performed a poll of user opinions regarding tweets from Twitter and classified those tweets as positive, negative, or neutral [3]. The glove model, which depicts word representations in glove vector space, was discussed as a current technique for learning matrix factorization representations of words by. Pennington, J., Socher, R., & Manning, C. D. By focusing on the nonzero elements of a verb cooccurrence matrix rather than learning the entire sparse matrix or specific context windows in a huge dataset, the result is seen in global matrix factorization and statistical information [6]. Document clustering is a technique used to group similar documents, and processing and dimensionality reduction techniques (pca and)which aid in document clustering with the use of k-means algorithm were examined in the method provided in [8] Kumar, A. A, & Chandrasekhar, S. The model BERT was developed in by Geetha, M. P., and Renuka, D. K. It is distinctive and different from previous machine learning models in that it is deeply bidirectional, both from left to right and right to left text representation, and pre-trained on a sizable unlabeled text corpus.

3 System Design

Figure 1 depicts the layout of our experiment. To begin, collect tweets from the dataset and pre-process the information so that it may be analysed. Then, using LDA, we execute topic modelling on it, employing glove, PCA, and BERT models, and comparing their accuracy in a graphical approach.

3.1 Dataset

The "omicron tweets" dataset from Kaggle is utilized in this work. The user id, name, user location, user description, user created, tweet, user followers, user friends, user likes, user verified, date, text, hashtags, source, retweets, favorites, and is retweet are the 16 fields that make up each row in the dataset. Omicrons tweet data from 10/2/2022 to 3/3/2022 is the time range of the dataset. The dataset contains 15000 tweets, of which 80% and 20% are used for training and testing, respectively. in order to display the results using graphs and a confusion matrix.

3.2 Data Pre-processing

The reduction of stop - word and punctuations is an important step in the pre-processing of data mining activities involving natural language processing. We utilize the pre-processing module to clean the data after extracting tweets from the dataset. The goal of this module is to remove hashtags, user names, URLs, RT symbols, punctuations, and non-English characters. It also includes data normalization, noise reduction. The tweets returned from Twitter must be pre-processed before we can utilize them to train our

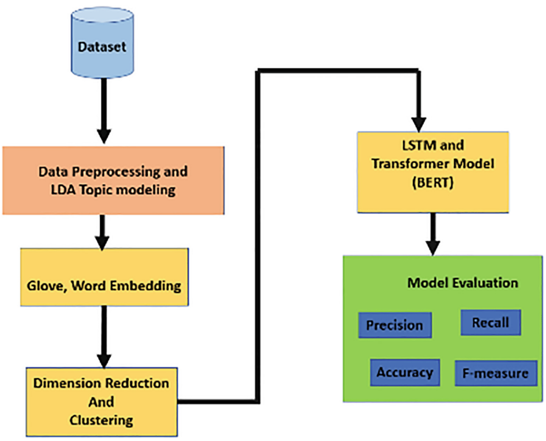

Fig. 1. System design for tweet Analysis

subject modelling algorithms A sample of tweets from each dataset at random URLs, emojis, and other special characters and sequences, such as "RT," are included in the tweets. When you see RT in a tweet, it means it was retweeted by another person [2].

3.2.1 Removing Stop Words

Stop words are words that recur frequently in writings but are irrelevant to the issue at hand. Stop words in the English language include "is," "and," "at," "the," and "it" which do not contribute much to text classification. To remove stop words from our tweets, we are using the NLTK python module's stop word component.

3.2.2 Removing Special Characters and Punctuation

The tweets contain punctuation, standard emoticons, and consumer emoticon built from sequences of special characters. The punctuation field in Python's string class is used to remove punctuation marks. Regular expressions are used to eliminate both user-created and standard emojis.

3.2.3 Tokenization

Tokenization is the process of breaking a text down into words, phrases, or other significant parts, or tokens. Thus, text segmentation includes tokenization. Tweets in the dataset have all been tokenized.

3.2.4 Stemming and Lemmatization

To find the derived words' root forms is the aim of stemming. For instance, "retrieval," "retrieved," and "retrieves" are all stemmed to get "retrieve. "By simply erasing word endings, many stemming systems achieve this outcome in a somewhat rudimentary manner. Stemming can cause a loss of meaning because it ignores the context of the words in a sentence. Using the parts of speech, lemmatization determines whether a word is a noun, verb, pronoun, adverb, or adjective and then converts it accordingly. For instance, the word "better" lemmatizes to corpus to corpora.

3.3 Building LDA Topic Modeling

The Twitter LDA is presented in, and it assumes that each has only one topic. It simulates the tweet generation process by assuming that when a user composes a tweet, he or she first selects a topic from the topic distribution. Then, based on the selected theme or background model, we select a bag of words one at a time. When it came to assigning subjects to a group of tweets, this technique exceeded traditional LDA in terms of quality. With Omicron Tweets processed and clean tweet data, then started topic modelling. Topic modelling is used to locate hidden subjects in enormous volumes of text.

Figure 2 depicts how we construct a model with a number of topics, each of which is made up of a number of keywords, each of which contributes a certain amount of weight to the topic. And we look at the words that appear in that topic, as well as their relative importance.

Topic: 0
Words:0.035*″ govern″ +0.024*″ open″ +0.018*″ coast″ +0.017*″ tassanian″ +0.017*″ gold″ + 0.014*″ Australia″ + 0.015*″ beat″ +0.010*″ win″ +0.010*″ ahead″ +0.009*″ shark″
Topic:1
Words:0.023*″ world″ + 0.014*″ final″ +0.013*″ record″ + 0.012*″ break″ + 0.011*″ loss″ + 0.011*″ Australian″ + 0.011*″ laugh″ + 11*″test″ +0.010 *″ Australia″ + 0.010*″ hill″
Topic:2
Words:0.018*″ rural″ +0.018*″ council″ + 0.015*″ fund″ + 0.012*″ plan″ + 0.011*″ health″ +0.012*″ chang″ + 0.011″ nation″ +0.010*″ service″ +0.009*″ say″
Topic 3:
Words:0.025*″ elect″ +0.022*″ adelaid″ + 0.012*″ perth″ +0.011*″ take″ + 0.011*″ say″ + 0.010*″ labor″ + 0.010*″ turnbul″ + 0.009*″ royal″ +0.009*″ time″

Fig. 2. Words occurring in that topic and its Relative weight.

3.4 Word Embedding

A method for representing individual words in text as a matrix of numerical values or vectors is called word embedding. For words with comparable meanings, it generates similar vector representations. The word vectorization technique, also known as the word embedding approach, turns each word into a separate vector for use as a neural network input. Although the process of mapping is typically carried out in low- dimensional space, it occasionally depends on the size of the vocabulary. The two main categories of word embedding are probabilistic prediction and count- based methods. Words compiled from a corpus are used in the prediction strategy to train the model. One of the best probabilistic prediction algorithms is Word2Vec [14].

3.4.1 Glove Model

The word context matrix factorization algorithms used in Glove are based on word context matrix factorization algorithms. It begins by constructing a large matrix of co-occurrence data, in which each "word" (row) in a vast corpus is tallied how many times it appears in some "context" (column). The resulting representations highlight relevant linear substructures of the word vectorspace, and training is based on aggregated global word to word co- occurrence statistics from a corpus. The number of "contexts" is enormous because it is essentially combinatorial in scale. Glove embedding, on the other hand, employs a somewhat different working approach than word2vec and is trained using an aggregated co- occurrence matrix of words. The frequency of words occurring together is shown in a corpus [11].

Output of glove model it reads only numeric form of word and show this word id is omicron related or not. And in that Fig. 3. 0 represent means positive tweet of omicron and 1 means negative tweets of omicron.

0	5665	0
1	540	1
2	7845	0
3	4859	1
4	132	0

Fig. 3. Output of glove model

3.5 Dimensionality Reduction and Clustering Technique

3.5.1 PCA (Principal Component Analysis)

A high dimensional dataset frequently contains redundant features because the majority of its features are associated. Dimensionality reduction is the process of locating and removing certain features [9].

The number of variables that are measured for each observation is the data's dimension. The fact that many times not all the measured variables are crucial for comprehending the underlying phenomena of interest is one of the issues with high-dimensional datasets. High dimensionality reduction addresses the issue of inefficient processing and makes it more challenging to find and take advantage of connections between words. Upon receiving the relationship between the keyword relationships, clustering is done quickly and efficiently. The procedure There will be less time. Principal components analysis (PCA) is a dimensionality- reduction approach that is frequently employed to extract a lower-dimensional space from a dataset's existing features, thereby producing new brand components.

3.5.2 K-Means Clustering

A user-specified number of clusters are discovered using the cluster analysis method K-means, which is represented by the centroids of each cluster. One of the frequently employed unsupervised learning techniques for examining the context of text data in natural language form is clustering. It is a mathematical strategy that gathering and grouping documents that are comparable into clusters [10]. The dataset contains four different types of "omicron" tweets, which were separated into four clusters for this study. K-means is helps to grouping the tweets in positive and negative form. The classifications are based on the following: group 1 contains long tweets, group 2 contains tweets with emoji, group 3 contains tweets that have been retweeted, and group 4 contains short tweets.

0	6731
2	3789
1	2989
3	1491

Fig. 4. Groups of tweets

The dataset contains four different types of "omicron" tweets, which were separated into four clusters for this study.

Groups 0, 1, 2, and 3 are divided into 4 groups, and Fig. 1 displays the number of tweets in each group. It contains three groups, with group 0 including 6731 tweets, group 1 containing 2989 tweets, group 2 containing 3789 tweets, and group 3 containing 1491 tweets. And Fig. 4, shows the group 0 contains a greater number of tweets are in there.

3.5.3 Visualization of Tweet with T-SNE

"T-SNE" that places each datapoint on a two- or three- dimensional map to view high-dimensional data. By minimizing the prehensility for points to cluster together in the map's center, the method of a version of stochastic neighbor embedding is generates noticeably superior visualizations. The embedding method known as t-SN is frequently used to display high-dimensional data in scatter plots [13].

Figure 5 shows in this scatter plot, the four colors red, orange, blue, and yellow stand in for the groups of tweets that they represent. Red stands for group 0, orange for group 1, blue for group 2, and yellow for group 3. Using a scatter plot, it is simple to see which group has the most tweets and which types of tweets they belong to.

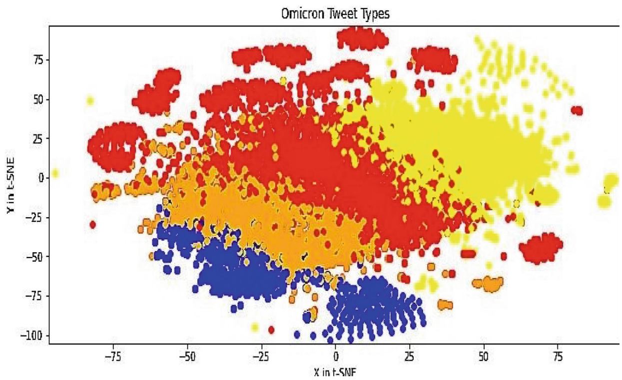

Fig. 5. Visualization of tweet using T-SNE

	id	label
0	hongkong throes worst coronavirus outbreak rec...	0
1	hi feeling slowly back life would would days o...	0
2	spread ba would believed would would times con...	0
3	covid would would covid would would covid omic...	0
4	would would day trends proportion omicronvaria...	0

Fig. 6. Output of BERT model

3.6 Transformer Model BERT and LSTM

3.6.1 BERT

A machine learning framework called Bidirectional Encoder Representations from Transformers (BERT) is intended for processing natural language. BERT is a deep learning model that is intended to pre-train deep learning systems [12]. Every output element is related to every input element. Bidirectional representations by simultaneously using unlabeled text in all layers, there is conditioning on the left and right context. Since BERT features an attention mechanism, also known as Transformer. Which enables analyzing the context of each word in a text separately and determining whether a word has already been used in a text with a similar context, it is characterized as a dynamic technique. As a result, the approach can discover contextual connections among words (or subworlds) in a text. BERT generates a dense vector representation of each word by "looking left and right numerous times." Because BERT learns two representations of each word, one on the right and one on the left, then repeats this learning n times, it is categorized as a fundamentally two- way model. BERT model is also helps to classifying the text if the text is related to omicron or not. The tweets contain word omicron and word related to omicron are in label 1 and this tweet are present in category A and the tweet not containing word related to omicron are in label 0 in label B.it means 0 represent the positive tweets of omicron.

Figure 6 depicts Which contain only positive tweets related to omicron and it contain label 0.

Figure 7 shows 7% of tweets of omicron are positive. And it gains loss is 29%.

3.6.2 LSTM

To identify long-distance dependencies in the sequential data, LSTM is developed. Using specific memory cells, it stores the contextual semantic data for dependencies in a long-range context. The input, forget, and output gates of each LSTM unit are used to coordinate and select the amount of information to hold, toss out, and proceed to the next stage. Additionally, it makes the choice on when to open gates that allow or prevent information from passing through the LSTM unit. Long-term memory in the hidden layer can be used by LSTM as the input for the activation function. Review The LSTM receives text data as input, and it categorizes the input category as output.

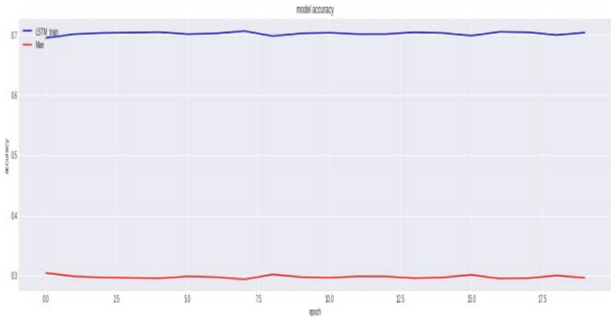

Fig. 7. Graphical representation of BERT models' output.

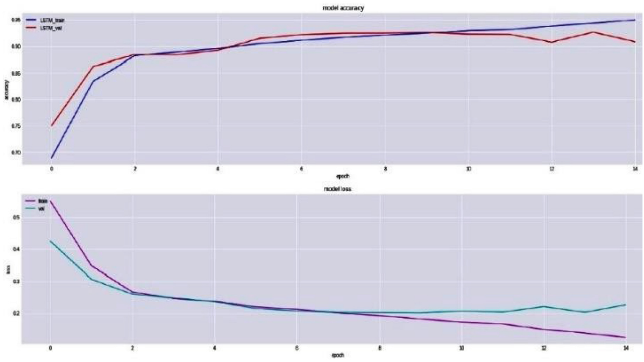

Fig. 8. Graphical representation of LSTM model

The main disadvantages of LSTM are that they take a lot longer to train than neural networks and that they are not really bidirectional because the model learns from left to right and right to left independently before concatenating the context.

Figure 8 illustrates how the LSTM model reveals that Omicron has received the most positive tweets. And it demonstrated 94% accuracy, which is higher than that of other models.

4 Results and Discussion

4.1 Evaluation Measure

The performance indicators that have been developed to gauge the results of the classification algorithms are called evaluation measures. Using the training data sets, the built-in model determines the class of unlabeled text (related to omicron and non-omicron) positive or negative tweets during the assessment phase. The prognostic the created model's performance can be assessed by using the following criteria.

Accuracy: gives the projected instances' percentage of total instances. It counts all instances that were appropriately categorized overall.

$$A = TP + TNTP + FP + TN + FN$$

Precision: in the presence of false positive cases, Precision provides model accuracy. As a result, the model's accuracy gives the overall incidence. Incidences of false positives and the rejection of positive instances.

$$\mathrm{p} = \mathrm{TPTP} + \mathrm{FP}$$

Recall: is used to gauge accuracy and displays how well the model performs when a false negative event occurs.

$$\mathrm{r} = \mathrm{TPTP} + \mathrm{FN}$$

F1- score: is a cumulative component that is tested to determine the overall influence of recall and precision in order to determine the overall impact of false negative and false positive instances across the entire accuracy.

$$\mathrm{f1} = 2(\mathrm{TP} + \mathrm{FP}) + (\mathrm{TP} + \mathrm{FN})\mathrm{TP}$$

True Positive (TP): The frequency with which real positive values coincide with predicted positive values.

False positive (FP): The number of times our model incorrectly interpreted negative values as positives.

True Negative (TN): The frequency with which our actual negative values and projected negative values coincide.

False Positive (FP): The number of times the model incorrectly interpreted negative values as positives

$$\mathrm{A} = \mathrm{TP} + \mathrm{TNTP} + \mathrm{FP} + \mathrm{TN} + \mathrm{FN}$$

The actual and anticipated favorable tweets are displayed in Fig. 9 and it illustrates the worth of TP, TN, FP, and FN.

Figure 10 illustrate that 0 represent positive tweets of omicron tweets accuracy and 1 represent negative omicron related tweets accuracy.

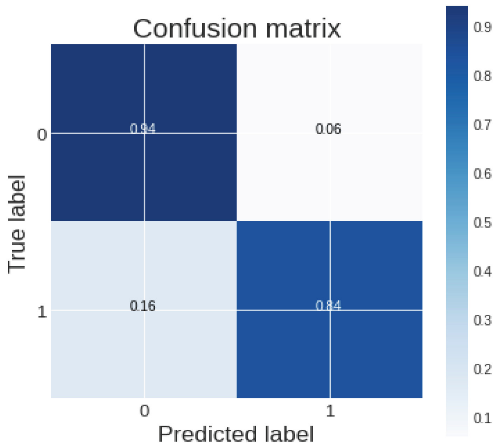

Fig. 9. Confusion matrix of analyzed tweets

	precision	recall	f1-score	support
0	0.96	0.93	0.94	2624
1	0.85	0.91	0.88	1129
accuracy			0.92	3753
macro avg	0.90	0.92	0.91	3753
weighted avg	0.92	0.92	0.92	3753

Fig. 10. Accuracy measure of omicron tweet dataset

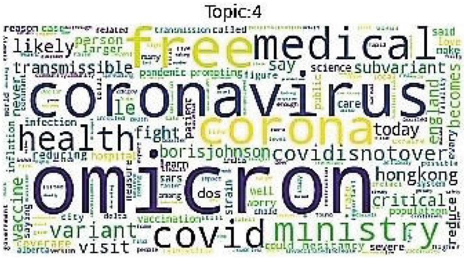

Fig. 11. Word cloud of omicron tweets.

4.2 Data Visualization Using Word Cloud

The terms that frequently appear in the document are displayed using word clouds. The output of the word cloud visualization for omicron tweets is shown below. A Word Cloud, sometimes referred as a Label Cloud, is a pictorial display of text data in the form of labels, which are typically single words whose value is indicated by their shape and color. The more commonly a term appears in a text and the more relevant it is, the greater and sharper it is.

The terms connected to the positive tweets of omicron label that appear the most are described in Fig. 11. Examples of words like omicron, coronavirus, and variation that are used frequently. The words are bolder and bigger in size it means the word most related omicron.

4.3 Sentiment Analysis

Opinion mining is another term for sentiment analysis. It's the method of assessing various lines of information to determine whether they have a positive, negative, or neutral emotional tone. Clearly define, sentiment analysis facilitates in determining an author's opinion toward a topic. Sentiment analysis software sorts writing into three categories: positive, neutral, and negative.

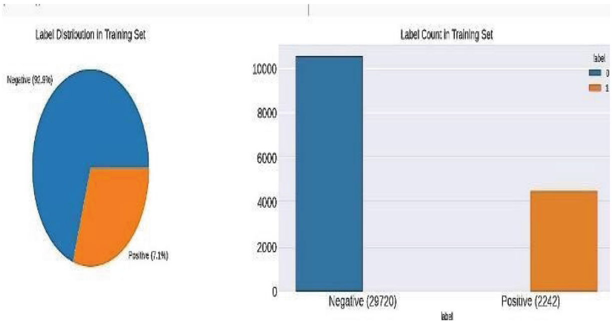

Fig. 12. Twitter Sentiment Analysis of positive and negative tweets

Figure 12 illustrates the sentiment analysis of tweets, which is shown in both positive and negative form. Positive tweets account for 7.1% of total tweets, whereas negative tweets account for 92.9%. The ratio of negative tweets is higher than the ratio of positive tweets, with label 0 indicating negative tweets in blue and label 1 indicating positive tweets in orange. The graph above illustrates that the ratio of negative tweets is the highest.

5 Conclusion

In this experiment, the first step is to clean up all of the tweets and eliminate any hashtags, URLs, special characters, and punctuation from the data. By using topic modelling to identify the abstract topics from our vast corpus after cleaning the data. Then, using the latent Dirichlet allocation (LDA) topic modelling algorithm, it displays Each document has a variety of terms, and each topic also contains a variety of words that correspond to it as well as specify their weight. On the basis of the words it contains, the LDA seeks to identify subjects to which the document belongs.

In our work, by using principal component analysis (PCA) to reduce the dimensionality of data. Based on their co-occurrences and weight, the most frequent words in the omicron language have been compiled in this study. Following the collection of tweets from the dataset, the initial stage is preprocessing, which helps to clean the tweets and make them easier to interpret and analyses. Word embedding is used to display the timing of word occurrences, and the glove represents the word's vector representation. And since the BERT transformer model is bidirectional, it enables readers to read our content from both the left and the right.

Acknowledgments. We would like Thank to Dr. Babasaheb Ambedkar Marathwada University in Aurangabad, India, publication support.

References

1. Lyu, J. C., Le Han, E., & Luli, G. K. (2021). COVID-19 vaccine–related discussion on Twitter: topic modeling and sentiment analysis. Journal of medical Internet research, 23(6), e24435.
2. Culmer, K., & Uhlmann, J. (2021). Examining LDA2Vec and Tweet Pooling for Topic Modeling on Twitter Data. Wseas Trans. Inf. Sci. Appl., 18, 102-115.
3. Shrivatava, A., Mayor, S., & Pant, B. (2014). Opinion mining of real time twitter tweets. International Journal of Computer Applications, 100(19).
4. Das, T. K., Acharjya, D. P., & Patra, M. R. (2014, January). Opinion mining about a product by analyzing public tweets in Twitter. In 2014 International Conference on Computer Communication and Informatics (pp. 1–4). IEEE.
5. Shafi, K. M., & Sheikh, S. A. (2019). Text Embedding Techniques for Sentiment Analysis: A Empirical Review. THE COMMUNICATIONS, 74.
6. Pennington, J., Socher, R., & Manning, C. D. (2014, October). Glove: Global vectors for word representation. In Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP) (pp. 1532–1543).
7. Sari, W. K., Rini, D. P., & Malik, R. F. (2019). Text Classification Using Long Short-Term Memory with GloVe. Jurnal Ilmiah Teknik Elektro Komputer dan Informatika (JITEKI), 5(1), 85-100.
8. Tam, S., Said, R. B., & Tanriöver, Ö. Ö. (2021). A ConvBiLSTM deep learning model-based approach for Twitter sentiment classification. IEEE Access, 9, 41283- 41293.
9. Noori, M. A. R., & Mehra, R. (2020, November). Fire Emergency Detection from Twitter Using Supervised Principal. In 2020 IEEE 15th International Conference on Industrial and Information Systems (ICIIS) (pp. 403–408). IEEE. Halibas, A. S., Shaffi, A. S., & Mohamed, M. A. K.V. (2018, March). Application of text classification and clustering of Twitter data for business analytics. In 2018 Majan international conference (MIC) (pp. 1–7). IEEE.
10. Es-Sabery, F., Es-Sabery, K., Qadir, J., Sainz-De- Abajo, B., Hair, A., Garcia-Zapirain, B., & De La Torre- Díez, I. (2021). A MapReduce opinion mining for COVID-19-related tweets classification using enhanced ID3 decision tree classifier. IEEE Access, 9, 58706- 58739.
11. Kilimci, Z. H., & Duvar, R. (2020). An efficient word embedding and deep learning based model to forecast the direction of stock exchange market using Twitter and financial news sites: a case of Istanbul stock exchange (BIST 100). IEEE Access, 8, 188186-188198.
12. Van der Maaten, L., & Hinton, G. (2008). Visualizing data using t-SNE. Journal of machine learning research, 9(11).

13. Fávero, E. M. D. B., & Casanova, D. (2021). BERT_SE: A Pre-trained Language Representation Model for Software Engineering. arXiv preprint arXiv:2112.00699.
14. Geetha, M. P., & Renuka, D. K. (2021). Improving the performance of aspect-based sentiment analysis using fine-tuned Bert Base Uncased model. International Journal of Intelligent Networks, 2, 64-69.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (http://creativecommons.org/licenses/by-nc/4.0/), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.