Indonesian Journal of Electrical Engineering and Computer Science
Vol. 29, No. 3, March 2023, pp. 1817~1826
ISSN: 2502-4752, DOI: 10.11591/ijeecs.v29.i3.pp1817-1826

Sentiment analysis of Malayalam tweets using bidirectional encoder representations from transformers: a study

Syam Mohan Elankath, Sunitha Ramamirtham
Department of Computer Science, Pondicherry University, Puducherry, India

Article Info

Article history:

Received May 25, 2022
Revised Nov 18, 2022
Accepted Nov 24, 2022

Keywords:

BERT
Deep learning
Machine learning
Malayalam tweets
Sentiment analysis

ABSTRACT

Sentiment analysis on views and opinions expressed in Indian regional languages has become the current focus of research. But, compared to a globally accepted language like English, research on sentiment analysis in Indian regional languages like Malayalam are very low. One of the major hindrances is the lack of publicly available Malayalam datasets. This work focuses on building a Malayalam dataset for facilitating sentiment analysis on Malayalam texts and studying the efficiency of a pre-trained deep learning model in analyzing the sentiments latent in Malayalam texts. In this work, a Malayalam dataset has been created by extracting 2,000 tweets from Twitter. The bidirectional encoder representations from transformers (BERT) is a pre-trained model that has been used for various natural language processing tasks. This work employs a transformer-based BERT model for Malayalam sentiment analysis. The efficacy of BERT in analyzing the sentiments latent in Malayalam texts has been studied by comparing the performance of BERT with various machine learning models as well as deep learning models. By analyzing the results, it is found that a substantial increase in accuracy of 5% for BERT when compared with that of Bi-GRU, which is the next best-performing model.

Corresponding Author:

Syam Mohan Elankath
Department of Computer Science, Pondicherry University
Puducherry, India
Email: syammohane@gmail.com

1. INTRODUCTION

Massive amount of textual data is uploaded to the Internet every day through various social media platforms by users globally. According to the Twitter statistics of 2018 [1], a stunning statement reveals that in a year 500 million tweets are posted, which means 6,000 tweets are posted each second. This unstructured data comprises plenty of intrinsic subjective information, the analysis of this subjective information can be beneficial in countless spaces. Sentiment analysis (SA) helps to extract this latent information from such data by analyzing and processing it. Sometimes, SA is also referred as sentiment mining, and opinion mining. wherein the expressed opinions or sentiments are identified and its polarity is classified as negative, positive, or neutral. Given the unstructured textual data, SA is performed at different granularities viz. document level, sentence level, and aspect-level [2]. In document level SA, the overall sentiment orientation of the text document is evaluated. Whereas in sentence-level SA, the sentiment orientation of all sentences in the document are individually evaluated. The most fine-grained level SA is the Aspect level SA where aspects are the attributes that characterizes the entities. In this type of SA, all aspects present in the text are determined and later their related sentiments are evaluated. Machine learning (ML) and deep learning (DL) approaches have shown promising results in the area of SA, just like all other areas where it accomplished excellently [3].

Handling the language component of a text data is the challenging aspect of SA [4]. Present-day online entertainment platforms empower individuals to communicate their perspectives in various worldwide languages. Hence analyzing sentiments present in different languages has become an aspect of research. Being a universal language, lot of research has happened in SA on English text. But very few researches have happened on SA with languages other than English, especially Indian regional languages. Malayalam is one among the 22 official languages in India and is spoken by 38 million people across the world. It is a south Indian language that comes under the Dravidian family and is also a morphologically rich agglutinative language, where comparably few research happened in SA [5]-[8]. One of the major reasons for this gap, is the lack of proper dataset and corpus in Malayalam language to facilitate SA.

Rakshitha *et al.* [9] extracted tweets using Twitter API of five different Indian languages including Kannada, Hindi, Telugu, Malayalam, and Tamil. They have used Python package TextBlob for finding the sentiment polarity. Rohini *et al.* [10] created a dataset consisting of movie reviews in the Kannada language from various websites. The authors have used decision tree (DT) classifier for finding the sentiment of reviews and furthermore they have compared the results with machine-translated English reviews of the same. Vrunda Joshi and Vekariya [11] compiled reviews in the Gujarati language from different social networking websites like Facebook, Twitter, and so on. Document level SA is done on the dataset using five different ML algorithms such as support vector machine (SVM), Naïve Bayes (NB), k-nearest neighbors (KNN), multi-layer perceptron and found that SVM is performing better than other ML algorithms with their dataset. Shrivastava and Kumar [12] proposed an approach with genetic algorithm to select the hyperparameter setting on the gated recurrent unit (GRU) model for SA in the Hindi language. Here, the authors have manually created a dataset consisting of 1,352 reviews. Mathews and Abraham [5] proposed a rule-based approach for SA on the Malayalam language. The authors have collected 136 tweets from Twitter and manually annotated them.

In an earlier work of Kumar *et al.* [6], DL models convolutional neural networks (CNN) and long short term memory (LSTM) were compared for SA on Malayalam tweets. Later on, in another work [7], the authors considered SVM and regularized least-squares classification (RLSC) as baseline models and compared them with DL models CNN and LSTM. They have used a manually created Malayalam Twitter dataset consisting of 13,000 tweets for their work. Soumya and Pramod [8] did a binary classification SA on Malayalam tweets using three ML models consisting of SVM, NB, and random forest (RF). They have used Unigram, term frequency-inverse document frequency (TF-IDF) and bag of words (BoW), with SentiWordNet as feature selection algorithms. A dataset consisting 3,184 tweets in Malayalam language has been constructed. Bayhaqy *et al.* [13] have done Hindi SA over movie reviews by creating a small datasets containing 250 reviews. The authors have utilized Hindi SentiWordNet and machine translation method for doing the SA. Soumya and Pramod [14] have done the same work as in [8], by replacing the ML models with various DL models like recurrent neural network (RNN), LSTM, and GRU. Thavareesan and Mahesan [15] used 5 different corpora's that contain a total of 2,691 reviews in the Indian language Tamil, which are collected from different social media platforms. Authors also compared different feature selection algorithms like BoW, TF, and TF-IDF with different ML techniques such as SVM, RF, NB, and KNN. Prasad *et al.* [16] have done SA on Indian languages Bengali and Tamil by creating datasets that consist of 999 and 1,103 tweets respectively. ML models NB and DT are compared by training on their datasets. Naidu *et al.* [17] created a dataset with newspaper sentences in the Indian language Telugu, which contained 1400 labeled sentences. The authors used Telugu SentiWordNet to classify the sentiments in this work.

Sharif *et al.* [18] have done SA with restaurant reviews in the Indian language Bengali, where the authors created a dataset and trained a model using the ML technique multinomial NB. Li *et al.* [19] proposed a model that dismisses the necessity for additional training in the bidirectional encoder representations from transformers (BERT) model. They devised two simple modules called Hierarchical Aggregation and Parallel Aggregation to use in conjunction with BERT. Karimi *et al.* [20] analyzed the BERT embedding component for the task of end-to-end aspect based sentiment analysis (ABSA). Abdelguad [21] used pre-trained BERT on Arabic hotel reviews dataset and found that multilingual BERT performs very well and is robust to overfitting on Arabic language. Safaya *et al.* [22] used BERT with CNN for multilingual offensive language classification with the SemEval 2020 dataset. Their results indicate that combining Convolutional Neural Network (CNN) with BERT improves the performance than using BERT alone. Jafrian *et al.* [23] used sentence pair input for BERT, which showed better results for Persian ABSA. Horne *et al.* [24] proposed a method which combines BERT hidden layers with GRU so that it improves the performance on Twitter SA. Moubtahij *et al.* [25] have done Arabic SA using Arabic BERT (AraBERT), a transformer-based model for the Arabic language. They have used ARev dataset, which holds more than 40,000 reviews on the tourism domain, and it is found that AraBERT performs competently with the existing works in the Arabic language.

From the literature, it is clear that a major hindrance to SA research over Indian languages is the lack of good datasets. Table 1 shows various manually created datasets for SA in different languages in India. Even though manual creation of a dataset is a challenging task, the majority of SA on Indian regional languages were done on their own manually created datasets by researchers. Moreover, the majority of these languages are morphologically complex and agglutinative, making the SA process considerably more challenging.

The objective of this study is to assess the performance of the transformer-based BERT model in the Malayalam language, as there are no works on SA using BERT on the Malayalam language. The problem of SA is portrayed as a binary classification of Malayalam tweets' overall polarity as positive or negative. In this paper, the authors have done SA on Malayalam tweets utilizing BERT [26] which is a powerful pre-trained language model. It is pre-trained on millions of textual documents, which enables the BERT model to understand the language and domain when compared to other ML and DL models. Moreover, BERT supports 104 languages which in turns helps to understand and resolve diverse problems in languages other than English including SA. This work aims at studying the performance of BERT model in carryout the SA of Malayalam tweets. As there are no publicly available datasets on Malayalam text, a dataset is manually created by extracting Malayalam tweets from Twitter using Twitters API. A total of 2,000 tweets were extracted which had explicit sentiment words as hashtags. The manually created twitter dataset is used for training the multilingual BERT (mBERT) model [26] for Malayalam language. Furthermore, the results of BERT are compared with various ML and DL models such as SVM, NB, DT, KNN, RF, logistic regression (LR), GRU, Bi-directional GRU (Bi-GRU), LSTM, and Bi-directional LSTM (Bi-LSTM), in order to evaluate the performance of BERT against legacy methods. Results shows that BERT outperforms all other ML and DL models with a highest accuracy of 88.61% followed by Bi-GRU with an accuracy of 83%. The ML model KNN achieved the lowest accuracy of 62.94%.

The remaining sections of this paper are organized as follows: section 2 describes the proposed work and briefly explains different ML and DL approaches employed in this work. Section 3 discusses the results and comparative analysis of different ML and DL models. Finally, section 4 concludes the paper.

Table 1. Manually created datasets in Indian languages

Dataset	Language	Size
Mathews and Abraham [5]	Malayalam	136
Kumar *et al.* [6]	Malayalam	13000
Kumar *et al.* [7]	Malayalam	12922
Soumya and Pramod [8]	Malayalam	3184
Rohini *et al.* [10]	Kannada	100
Joshi and Vekariya [11]	Gujarati	40
Shrivastava and Kumar [12]	Hindi	8352
Bayhaqy *et al.* [13]	Hindi	230
Soumya and Pramod [14]	Malayalam	5468
Thavareesan and Mahesan [15]	Tamil	2691
Prasad *et al.* [16]	Bengali, Tamil	999, 1103
Naidu *et al.* [17]	Telugu	1400
Sharif *et al.* [18]	Bengali	1427

2. METHOD

The objective of this study is to assess the performance of the transformer-based BERT model in the Malayalam language, as there are no works on SA using BERT on the Malayalam language. Furthermore, the results of BERT are compared with various ML and DL models such as SVM, NB, DT, KNN, RF, LR, GRU, Bi-GRU, LSTM, and Bi-LSTM, in order to evaluate the performance of BERT against legacy methods. Each of these approaches are briefly explained.

2.1. BERT

BERT is a powerful DL-based state-of-the-art language model for numerous tasks in NLP natural language inference, question answering, and text classification [8]. It is built on encoders of transformers and is also pre-trained on millions of text documents. Pre-training of BERT is done using two methods, namely masked language modeling (MLM) and next sentence prediction (NSP). The overall architecture of the BERT model is given in Figure 1. Contextual bi-directional embedding is being supplied by BERT, where contextualization means that the same words can have a different meaning with respect to the domains. For that, unlike LSTM, BERT acquires the input sentence as a whole input, and therefore it is bi-directional. The higher layers of BERT extract the language semantics, and the lower layers extract the syntactic information. The first input to the model is the CLS token, which is used as a classification token. It is followed by the sequence of words in the input. This input is then given to a stack of encoders, where it passes through self-attention and feedforward networks. The primary objective of self-attention is to provide contextual information to terms in the sentence. 12 transformer-based encoders are there in a BERT base model. The output of this model will be a vector of size 768, which can be given to a classification layer for the task of classification [25]-[31].

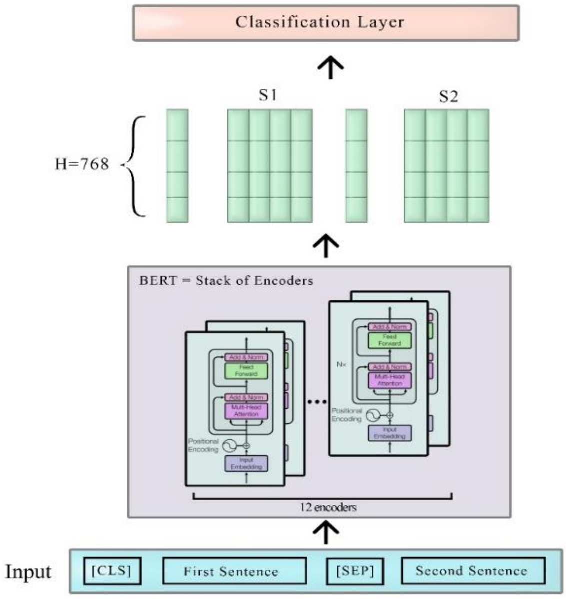

Figure 1. Overall architecture of BERT model

2.2. ML approaches

2.2.1. Decision tree (DT)

The DT is a logic-based algorithm where the whole complex decision is divided into various uncomplicated, simpler decisions. In other words, we can say that it is a mathematical model used to represent a decision-making process. In this technique, a logical tree is constructed with different levels of logical conditions and options which helps to derive the desired solution [31]-[36].

2.2.2. Logistic regression (LR)

LR algorithm uses the logistic sigmoid function to calculate the probability of the target variable. This supervised algorithm is an updated version of linear regression for classification tasks where the sigmoid function is oblique to map the original value between 0 and 1. Unlike other classification models, LR not only classifies the data but also gives the probability of that data in its particular category [37]-[41].

2.2.3. Support vector machine (SVM)

SVM is a supervised ML algorithm used for classification and regression tasks. On an n-dimensional graph, all the data items will be plotted and a line will be drawn around the support vectors separating different classes. This line is called a hyperplane and there will be many hyperplanes. Among them, one hyperplane is chosen when it satisfies the highest distance from the support vectors. In the background, SVM solves the complex optimization problem which helps to maximize the distance from support vectors to the hyperplane [34], [36], [42]-[47].

2.2.4. Random forest (RF)

RF is an ensemble approach that contains an extensive number of decision trees. The result of the RF algorithm is calculated by taking the average of outputs of individual decision trees. So as the number of trees increases, the accuracy of the RF also improves. Also, the problem of overfitting found in DT algorithm is resolved in RF approach. [35], [36], [45], [47].

2.2.5. Naïve bayes (NB)

NB is a straightforward but powerful statistics-based approach for predictive modeling. It depends on the Bayesian theorem of likelihood which makes the probabilities for every event. NB assumes that each feature is independent and hence the most elevated probability output is predicted. The advantage of the NB classifier is that it needs less training data and still gives promising results. The drawback of this method is that it is also known as a bad estimator since it assumes each feature as independent [48]-[51].

2.2.6. K-nearest neighbors (KNN)

KNN is lazy learner algorithm which employs a simple classification technique. The dataset will be stored in the initial phase and when new data arrives, based on the similarity of the new data with stored data, KNN determines its categories. Here Euclidean distance is calculated to find the K nearest neighbors. The KNN algorithm works well with noisy training data as well as the implementation is simple. The disadvantage is that when new data comes, K neighbors have to be recalculated again, which in turns increase the computational time consumption [45], [50], [52], [53].

2.3. DL approaches

2.3.1. Long short-term memory (LSTM)

LSTM is a type of RNN and LSTM overcomes the RNNs problem of long-term dependency. Also, vanishing gradients and exploding gradients problems that arise while the training process is also solved in LSTM. Unlike most ML models, LSTM can memorize information for a prolonged amount of period. This is facilitated by an explicit memory unit named cell in its architecture. The variant, Bi-LSTM contains two LSTMs where one takes input in the forward direction and the other in the opposite direction. This arrangement enables the Bi-LSTM model to include more context knowledge [54]-[61].

2.3.2. Gated recurrent unit (GRU)

GRU is an advanced type of RNN and it is a variant of LSTM. Instead of having a separate memory unit called cell, GRU have hidden states to store information. Just like LSTM, GRU also uses gated mechanism to control the flow and here there are only two gates, namely, update gate and forget gate. Update gate makes sure of the amount of information flowing to the future and forget gate removes the irrelevant information. This makes the model less complex and hence it is much faster than the LSTM. Also, GRU performs well when the training data is comparably small. Bi-GRU comprises of two GRUs, where one takes input in the backward direction and the other one takes input in the forward direction. In language processing, Bi-GRU gives better results as it can understand the underlying information in the languages compared to other models [62]-[67].

The objective of this work can be divided into three; assessing the multilingual BERT model's effectiveness with the Malayalam language, creating the Malayalam dataset on tweets, and finally, comparing the results of the BERT model with aforementioned legacy methods. Figure 2 shows the architecture of SA for Malayalam tweets.

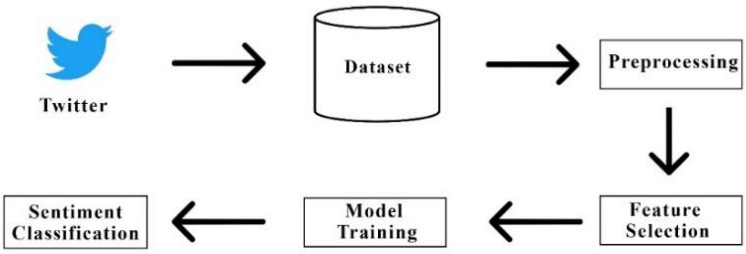

Figure 2. Architecture of sentiment analysis

2.3.3. Construction of Malayalam dataset

The lack of datasets for SA on Malayalam is main hindrance to research in this field. As of now, there are no publicly available datasets for Malayalam SA. Therefore, the authors have created a dataset by extracting Malayalam tweets from Twitter. For the extraction of tweets, a set of Malayalam sentiment words as hashtags are used. Table 2 shows the list of positive and negative Malayalam hashtags used for extracting the tweets. With the help of Twitter API, these hashtags were used to extract tweets from Twitter. Further, these tweets were manually labeled based on their sentiment polarity into two classes, viz. negative and positive. A total of 2,000 tweets are labeled, where 50% are positive tweets and the other 50% are negative sentiment oriented. Table 3 and Table 4 expresses the sample positive and negative tweets along with their English translations.

2.3.4. Preprocessing

After the creation of the dataset, preprocessing is done on the data to make sure it is suitable for the further processing. The extracted tweets in the dataset contain a lot of irrelevant details for SA like hyperlinks, user id’s, and whitespaces. In the preprocessing stage, white spaces, punctuations, mentions, and URLs are removed from the extracted tweets.

Table 2. Malayalam hashtags

Positive	Negative
സന്തോഷം (Happy)	സങ്കടം (Sad)
ഇഷ്ടം (Love)	ദുഃഖം (Sad)
ഗംഭീരം (Great)	നഷ്ടപ്പെട്ടു (Lost)
വിജയം (Success)	പരാജയം (Failure)
അഭിനന്ദനം (Appreciation)	ഭീഷണി (Threat)
അംഗീകാരം (Approval)	ഭയം (Fear)
ആനന്ദം (Happiness)	ആശങ്ക (Suspicion)
സമാധാനം (Peace)	വഞ്ചന (Cheat)
ജയം (Victory)	പേടി (Fear)
നന്മ (Goodness)	വേദന (Pain)

Table 3. Positive tweets

Positive Tweet
അജു നല്ല ഭാഗ്യവാൻ ആണ്. നല്ല ജീവിതം നേരുന്നു എന്നും ഈ സന്തോഷം ഉണ്ടക്കട്ടെ ജീവിതത്തിൽ (*Aju is very lucky. I wish you a good life and may you have this happiness in life*)
പാട്ട് ഇഷ്ടപ്പെട്ടു. ഹിഷാം അബ്ദുൽ വഹാമിന്റെ ശബ്ദവും പ്രണവിന്റെയും ദർശനയുടെയും അഭിനയവും ഇഷ്ടപ്പെട്ടു. (*Loved the song. Loved Hisham Abdul Waham's voice and acting of Pranav and Darshana*)

Table 4. Negative tweets

Negative Tweet
എന്നാലും ഇത്ര ദയനീയമായ പരാജയം സ്വപ്നത്തിൽ പോലും വിചാരിച്ചില്ല. (*Yet such a miserable failure was not even imagined in a dream.*)
സഹിക്കാനാവാത്ത വേദന..ലെജ്ജിക്കു കേരളമേ (*Unbearable pain .. Kerala to shame*)

2.3.5. Feature selection

In this work, for the task of Malayalam SA, both ML and DL approaches are used. ML models like DT, LR, SVM, RF, NB, KNN are considered and BERT, LSTM, Bi-LSTM, GRU, and Bi-GRU are considered from DL models. In ML, it is required to explicitly mention the feature selection method to extract the relevant features. But in the DL approaches, feature selection is automatically done. In this work, TF-IDF feature selection method is adopted for feature selection, because from the review of literature [15], [16] it is understood that TF-IDF is known for better extraction of features in SA. The statistical measure TF-IDF expressed in (1) is used to evaluate the significance of a distinct word in a corpus.

$$tf - idf(t, d) = tf(t, d) * idf(t) \tag{1}$$

where idf is the inverse document frequency and tf is the term frequency, and t is term (word) and d is document (set of words). Unlike other DL models like LSTM, BERT has its own embedding and it uses the concept of word-piece tokenization which means that the words will be broken into sub words. BERT embedding starts with the tag [CLS] and each sentence will be separated with [SEP] tags. For example, consider the sentence,

അപ്പവും മുട്ടക്കറിക്കും ഫാൻസ് ഇല്ല എന്നുള്ളതിന് അതിയായ സങ്കടം രേഖപ്പെടുത്തി കൊള്ളുന്നു .

(It is very sad that Appam and Muttakari has no fans)

The BERT tokenized form will be:

['[CLS]', 'അ', '##പ്', '##പ', '##വും', 'മ', '##ു', '##ട്', '##ട', '##ക്', '##ക', '##റി', '##ക്കും', 'ഫ', '##ാൻ', '##സ്', 'ഇ', '##ല്ല', 'എന്നു', '##ള്ള', '##തിന്', 'അ', '##തി', '##യായ', 'സ', '##ങ്', '##ക', '##ടം', 'ര', '##േ', '##ഖ', '##പ്', '##പ', '##െ', '##ടുത്ത', '##ി', '[SEP]']

2.3.6. Model training

The dataset is split in a 70-30 ratio to form training and testing data. BERT and other ML and DL models are used to train on the training data and are also tested with the test data of the dataset. The embedding layer will convert the tweets into meaningful vectors. The embedding vector dimension of BERT is 768 and that of other DL models is set to 128. BERT uses 12 layers of transformer encoders with a hidden size of 768. For DL models, the number of neurons in the hidden layer is fixed to 60, 80, and 100. The regularization parameter value of 0.3 is set at both embedding and hidden layers and also to minimize overfitting problems, the dropout layer is added. To classify the tweets into either positive or negative sentiments, the sigmoid activation function is employed in the final layer. Furthermore, during training, Adam optimization is used and for loss function, binary cross-entropy is used. BERT used 10 epochs for training the dataset. For DL models, the values 50 and 45 are set as the number of training epochs and batch size respectively.

3. RESULTS AND DISCUSSION

Six ML and five DL approaches have used on training manually created dataset on Malayalam tweets. BERT has shown superior results over all other ML and DL models with an accuracy of 88.06%, followed by Bi-GRU with 83%. The training and validation loss graph is depicted in Figure 3 and the BERT model's confusion matrix is illustrated in figure 4. Table 5 shows the detailed results of various ML and DL approaches. Bi-GRU is getting better results with respect to other DL and ML models is because of its ability to work efficiently in smaller datasets. Even though dropout regularization is used in DL models, still there is a problem of overfitting. This is due to the size of the dataset, which is comparably small. It is clear from the results that DL approaches are giving better accuracy over ML approaches and the pre-trained model BERT is performing better than all other models in SA on Malayalam tweets.

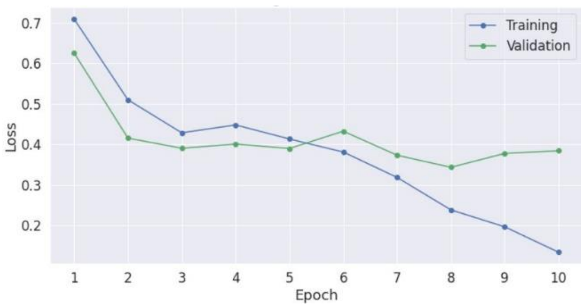

Figure 2. Training and validation loss graph of BERT model

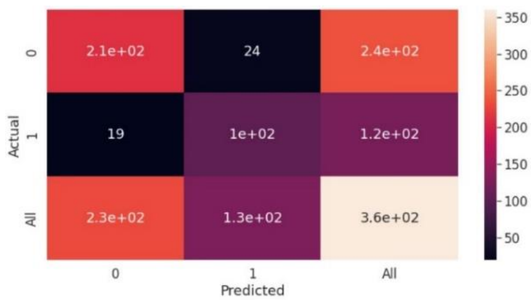

Figure 3. Confusion matrix of BERT model

Table 5. Results

Model	Precision	F Score	Recall	Accuracy
DT	0.75	0.74	0.73	0.74
LR	0.92	0.90	0.89	0.73
SVM	0.92	0.89	0.88	0.74
RF	0.87	0.86	0.86	0.78
NB	0.66	0.64	0.63	0.73
KNN	0.45	0.44	0.44	0.62
LSTM	0.81	0.81	0.81	0.80
Bi-LSTM	0.83	0.83	0.83	0.82
GRU	0.83	0.83	0.82	0.82
Bi-GRU	0.84	0.84	0.84	0.83
BERT	0.86	0.86	0.87	0.88

4. CONCLUSION

SA on Indian regional languages is one of the less explored areas of research. In this paper, BERT which is a transformer-based pre-trained model is used for SA on Indian language Malayalam tweets. Since there aren't any publicly accessible datasets, authors have created a dataset on Malayalam by extracting tweets from Twitter. For this, Twitter API has been used and later the tweets are labeled manually according to their sentiment polarity. Total of 2,000 tweets were extracted and 50% of them are positive sentiment oriented and other 50% is negative sentiment oriented. Along with BERT, ten ML and DL models are also used for the same dataset and compared their results of SA on Malayalam tweets. The BERT model achieved highest accuracy of 88.61%. Among the other ML and DL approaches, Bi-GRU achieved the next highest test accuracy of 83.0% and KNN achieved lowest accuracy of 62.94%. Due to the size of dataset, proposed models suffer overfitting problem even after using dropout regularization. The proposed methodologies will be tested on a wider corpus in the future, avoiding the problem of overfitting and increasing model efficiency.

REFERENCES

[1] R. Bose, R. K. Dey, S. Roy, and D. Sarddar, “Analyzing Political Sentiment Using Twitter Data,” *Information and Communication Technology for Intelligent Systems*, pp. 427–436, Dec. 15, 2018, doi: 10.1007/978-981-13-1747-7_41.

[2] K. Ravi and V. Ravi, “A survey on opinion mining and sentiment analysis: Tasks, approaches and applications,” *Knowledge-Based Syst.*, vol. 89, pp. 14–46, 2015.

[3] D. Riswantini and E. Nugraheni, “Machine learning in handling disease outbreaks: a comprehensive review,” *Bulletin of Electrical Engineering and Informatics*, vol. 11, no. 4, pp. 2160–2168, Aug. 2022, doi: 10.11591/eei.v11i4.3612.

[4] M. V. D. Kauter, D. Breesch, and V. Hoste, “Fine-grained analysis of explicit and implicit sentiment in financial news articles,” *Expert Syst. Appl.*, vol. 42, no. 11, pp. 4999–5010, 2015.

[5] D. M. Mathews and S. Abraham, “Twitter Data Sentiment Analysis on a Malayalam Dataset Using Rule-Based Approach,” *Emerging Research in Computing, Information, Communication and Applications*, pp. 407–415, 2019. doi: 10.1007/978-981-13-6001-5_33.

[6] S. S. Kumar, M. A. Kumar, and K. P. Soman, “Identifying Sentiment of Malayalam Tweets Using Deep Learning,” *Digital Business*, pp. 391–408, Jul. 27, 2018, doi: 10.1007/978-3-319-93940-7_16.

[7] S. S. Kumar, M. A. Kumar, and K. P. Soman, “Sentiment Analysis of Tweets in Malayalam Using Long Short-Term Memory Units and Convolutional Neural Nets,” *Mining Intelligence and Knowledge Exploration*, pp. 320–334, 2017, doi: 10.1007/978-3-319-71928-3_31.

[8] S. Soumya, and K. V. Pramod, “Sentiment analysis of malayalam tweets using machine learning techniques,” *ICT Express*, vol. 6, no. 4, pp. 300–305, Dec. 2020. doi: 10.1016/j.icte.2020.04.003.

[9] K. Rakshitha, H. M. Ramalingam, M. Pavithra, H. D. Advi, and M. Hegde, “Sentimental Analysis of Indian Regional Languages on Social Media,” *Global Transitions Proceedings*, vol. 2, no. 2, pp. 414–420, 2021, doi: 10.1016/j.gltp.2021.08.039.

[10] V. Rohini, M. Thomas, and C. A. Latha, “Domain based sentiment analysis in regional Language-Kannada using machine learning algorithm,” in *2016 IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology (RTEICT)*, May 2016, doi: 10.1109/rteict.2016.7807872.

[11] V. C. Joshi and V. M. Vekariya, “An Approach to Sentiment Analysis on Gujarati Tweets,” *Advances in Computational Sciences and Technology*, vol. 10, no. 5, pp. 1487-1493, 2017.

[12] K. Shrivastava and S. Kumar, “A Sentiment Analysis System for the Hindi Language by Integrating Gated Recurrent Unit with Genetic Algorithm,” *The International Arab Journal of Information Technology*, vol. 17, no. 6, pp. 954–964, Nov. 2020, doi: 10.34028/iajit/17/6/14.

[13] A. Bayhaqy, S. Sfenrianto, K. Nainggolan, and E. R. Kaburuan, “Sentiment Analysis about E-Commerce from Tweets Using Decision Tree, K-Nearest Neighbor, and Naïve Bayes,” in *2018 International Conference on Orange Technologies (ICOT)*, 2018, pp. 1-6, doi: 10.1109/ICOT.2018.8705796.

[14] S. Soumya and K. V. Pramod, “Sentiment Analysis of Malayalam Tweets using Different Deep Neural Network Models-Case Study,” in *2019 9th International Conference on Advances in Computing and Communication (ICACC)*, Nov. 2019, doi: 10.1109/icacc48162.2019.8986192.

[15] S. Thavareesan and S. Mahesan, “Sentiment Analysis in Tamil Texts: A Study on Machine Learning Techniques and Feature Representation,” in *2019 14th Conference on Industrial and Information Systems (ICIIS)*, Dec. 2019, doi: 10.1109/iciis47346.2019.9063341.

[16] S. S. Prasad, J. Kumar, D. K. Prabhakar, and S. Tripathi, “Sentiment mining: An approach for Bengali and Tamil tweets,” in *2016 Ninth International Conference on Contemporary Computing (IC3)*, Aug. 2016, doi: 10.1109/ic3.2016.7880246.

[17] R. Naidu, S. K. Bharti, K. S. Babu, and R. K. Mohapatra, “Sentiment analysis using Telugu SentiWordNet,” in *2017 International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET)*, Mar. 2017, doi: 10.1109/wispnet.2017.8299844.

[18] O. Sharif, M. M. Hoque, and E. Hossain, “Sentiment Analysis of Bengali Texts on Online Restaurant Reviews Using Multinomial Naïve Bayes,” in *2019 1st International Conference on Advances in Science, Engineering and Robotics Technology (ICASERT)*. May 2019, doi: 10.1109/icasert.2019.8934655.

[19] X. Li, L. Bing, W. Zhang, and W. Lam, “Exploiting BERT for End-to-End Aspect-based Sentiment Analysis,” in *Proceedings of the 5th Workshop on Noisy User-generated Text (W-NUT 2019)*, 2019, doi: 10.18653/v1/d19-5505.

[20] A. Karimi, L. Rossi, and A. Prati, “Improving BERT Performance for Aspect-Based Sentiment Analysis,” 2019, *arXiv:* 2010.11731.

[21] M. M. Abdelgwad, “Arabic aspect based sentiment analysis using BERT,” 2021, *arXiv:* 2107.13290, 2021

[22] A. Safaya, M. Abdullatif, and D. Yuret, “KUISAIL at SemEval-2020 Task 12: BERT-CNN for Offensive Speech Identification in Social Media,” in *Proceedings of the Fourteenth Workshop on Semantic Evaluation*, 2020, doi: 10.18653/v1/2020.semeval-1.271.

[23] H. Jafarian, A. H. Taghavi, A. Javaheri and R. Rawassizadeh, “Exploiting BERT to Improve Aspect-Based Sentiment Analysis Performance on Persian Language,” in *2021 7th International Conference on Web Research (ICWR)*, 2021, pp. 5-8, doi: 10.1109/ICWR51868.2021.9443131.

[24] L. Horne, M. Matti, P. Pourjafar, Z. Wang, “GRUBERT: A GRU-Based Method to Fuse BERT Hidden Layers for Twitter Sentiment Analysis,” in *Proceedings of the 1st Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics and the 10th International Joint Conference on Natural Language Processing: Student Research Workshop*, 2020, 130–138.

[25] H. E. Moubtahij, H. Abdelali, and E. B. Tazi, “AraBERT transformer model for Arabic comments and reviews analysis,” *IAES International Journal of Artificial Intelligence (IJ-AI)*, vol. 11, no. 1, p. 379, Mar. 2022, doi: 10.11591/ijai.v11.i1.pp379-387.

[26] A. Joshi, A. R. Balamurali, and P. Bhattacharyya, “A Fall-Back Strategy for Sentiment Analysis in Hindi: A Case Study,” in *Proceedings of the Fifth International Conference on Systems (ICONS)*, Apr. 2010, 1-6.

[27] J. Devlin, M. W. Chang, K. Lee, and K. Toutanova, “BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. CoRR, 2018, *arXiv:* 1810.04805.

[28] H. Xu, B. Liu, L. Shu, and P. S. Yu, “BERT Post-Training for Review Reading Comprehension and Aspect-based Sentiment Analysis.” 2019, *arXiv*: 10.48550/ARXIV.1904.02232.

[29] M. Hoang, O. A. Bihorac, and J. Rouces, "Aspect-based sentiment analysis using BERT," in *Proceedings of the 22nd nordic conference on computational linguistics*, 2019, pp. 187-196.
[30] Z. Gao, A. Feng, X. Song and X. Wu, "Target-Dependent Sentiment Classification With BERT," *IEEE Access*, vol. 7, pp. 154290-154299, 2019, doi: 10.1109/ACCESS.2019.2946594.
[31] M. Hammad, M. Al-Smadi, Q. B. Baker, and S. A. Al-Zboon, "Using deep learning models for learning semantic text similarity of Arabic questions," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 11, no. 4, p. 3519, Aug. 2021, doi: 10.11591/ijece.v11i4.pp3519-3528.
[32] P. Mookdarsanit and L. Mookdarsanit, "The COVID-19 fake news detection in Thai social texts," *Bulletin of Electrical Engineering and Informatics*, vol. 10, no. 2, pp. 988–998, Apr. 2021, doi: 10.11591/eei.v10i2.2745.
[33] H. T. Phan, V. C. Tran, N. T. Nguyen, and D. Hwang, "Decision-Making Support Method Based on Sentiment Analysis of Objects and Binary Decision Tree Mining," *Lecture Notes in Computer Science*, pp. 753–767, 2019, doi: 10.1007/978-3-030-22999-364.
[34] M. Rathi, A. Malik, D. Varshney, R. Sharma, and S. Mendiratta, "Sentiment Analysis of Tweets Using Machine Learning Approach," in *2018 Eleventh International Conference on Contemporary Computing (IC3)*, 2018, pp. 1-3, doi: 10.1109/IC3.2018.8530517.
[35] X. Huang *et al.* "LSTM based sentiment analysis for cryptocurrency prediction," *International Conference on Database Systems for Advanced Applications*, pp. 617-621, 2021, doi: https://doi.org/10.48550/arXiv.2103.14804
[36] M. Aufar, R. Andreswari and D. Pramesti, "Sentiment Analysis on Youtube Social Media Using Decision Tree and Random Forest Algorithm: A Case Study," in *2020 International Conference on Data Science and Its Applications (ICoDSA)*, 2020, pp. 1-7, doi: 10.1109/ICoDSA50139.2020.9213078.
[37] H. Parveen and S. Pandey, "Sentiment analysis on Twitter Data-set using Naive Bayes algorithm," in *2016 2nd International Conference on Applied and Theoretical Computing and Communication Technology (iCATccT)*, 2016, pp. 416-419, doi: 10.1109/ICATCCT.2016.7912034.
[38] W. P. Ramadhan, S. T. M. T. A. Novianty, and S. T. M. T. C. Setianingsih, "Sentiment analysis using multinomial logistic regression," in *2017 International Conference on Control, Electronics, Renewable Energy and Communications (ICCREC)*, 2017, pp. 46-49, doi: 10.1109/ICCEREC.2017.8226700.
[39] A. Tyagi and N. Sharma, "Sentiment Analysis using Logistic Regression and Effective Word Score Heuristic," *International Journal of Engineering & Technology*, vol. 7, no. 2.24, p. 20, Apr. 2018, doi: 10.14419/ijet.v7i2.24.11991.
[40] A. Prabhat and V. Khullar, "Sentiment classification on big data using Naïve bayes and logistic regression," in *2017 International Conference on Computer Communication and Informatics (ICCCI)*, 2017, pp. 1-5, doi: 10.1109/ICCCI.2017.8117734.
[41] M. Z. Ansari, T. Ahmad, M. M. S. Beg, and N. Bari, "Language lexicons for Hindi-English multilingual text processing," *IAES International Journal of Artificial Intelligence (IJ-AI)*, vol. 11, no. 2, p. 641, Jun. 2022, doi: 10.11591/ijai.v11.i2.pp641-648.
[42] M. Al Omari, M. Al-Hajj, N. Hammami, and A. Sabra, "Sentiment Classifier: Logistic Regression for Arabic Services' Reviews in Lebanon," in *2019 International Conference on Computer and Information Sciences (ICCIS)*, 2019, pp. 1-5, doi: 10.1109/ICCISci.2019.8716394.
[43] M. Ahmad, S. Aftab, M. Salman, N. Hameed, I. Ali, and Z. Nawaz, "SVM Optimization for Sentiment Analysis," *International Journal of Advanced Computer Science and Applications*, vol. 9, no. 4, 2018, doi: 10.14569/ijacsa.2018.090455.
[44] F. Luo, C. Li and Z. Cao, "Affective-feature-based sentiment analysis using SVM classifier," in *2016 IEEE 20th International Conference on Computer Supported Cooperative Work in Design (CSCWD)*, 2016, pp. 276-281, doi: 10.1109/CSCWD.2016.7566001.
[45] J. Singh and P. Tripathi, "Sentiment analysis of Twitter data by making use of SVM, Random Forest and Decision Tree algorithm," in *2021 10th IEEE International Conference on Communication Systems and Network Technologies (CSNT)*, 2021, pp. 193-198, doi: 10.1109/CSNT51715.2021.9509679.
[46] J. Govindappa and K. Channegowda, "Analyzing sentiment dynamics from sparse text coronavirus disease-19 vaccination using natural language processing model," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 12, no. 4, p. 4054, Aug, 2022, doi: 10.11591/ijece.v12i4.pp4054-4066.
[47] A. B. Gumelar, A. Yogatama, D. P. Adi, F. Frismanda, and I. Sugiarto, "Forward feature selection for toxic speech classification using support vector machine and random forest," *IAES International Journal of Artificial Intelligence (IJ-AI)*, vol. 11, no. 2, p. 717, Jun. 2022, doi: 10.11591/ijai.v11.i2.pp717-726.
[48] F. M. J. M. Shamrat *et al.*, "Sentiment analysis on twitter tweets about COVID-19 vaccines using NLP and supervised KNN classification algorithm," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 23, no. 1, p. 463, Jul. 2021, doi: 10.11591/ijeecs.v23.i1.pp463-470.
[49] P. Gamallo and M. Garcia, "Citius: A Naive-Bayes Strategy for Sentiment Analysis on English Tweets," in *Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014)*, 2014, doi: 10.3115/v1/s14-2026.
[50] S. Tan, X. Cheng, Y. Wang, and H. Xu, "Adapting Naive Bayes to Domain Adaptation for Sentiment Analysis," *Lecture Notes in Computer Science*, 2009, pp. 337–349, doi: 10.1007/978-3-642-00958-7_31.
[51] H. Kang, S. J. Yoo, and D. Han, "Senti-lexicon and improved Naïve Bayes algorithms for sentiment analysis of restaurant reviews," *Expert Systems with Applications*, vol. 39, no. 5, pp. 6000–6010, Apr. 2012, doi: 10.1016/j.eswa.2011.11.107.
[52] A. Alshamsi, R. Bayari, and S. Salloum, "Sentiment Analysis in English Texts," *Advances in Science, Technology and Engineering Systems Journal*, vol. 5, no. 6, pp. 1683–1689, Dec. 2020, doi: 10.25046/aj0506200.
[53] M. Rezwanul, A. Ali, and A. Rahman, "Sentiment Analysis on Twitter Data using KNN and SVM," *International Journal of Advanced Computer Science and Applications*, vol. 8, no. 6, 2017, doi: 10.14569/ijacsa.2017.080603.
[54] Y. Ma, H. Peng, T. Khan, E. Cambria, and A. Hussain, "Sentic LSTM: a Hybrid Network for Targeted Aspect-Based Sentiment Analysis," *Cognitive Computation*, vol. 10, no. 4, pp. 639–650, Mar. 2018, doi: 10.1007/s12559-018-9549-x.
[55] Z. Jin, Y. Yang, and Y. Liu, "Stock closing price prediction based on sentiment analysis and LSTM," *Neural Computing and Applications*, vol. 32, no. 13, pp. 9713–9729, Sep. 2019, doi: 10.1007/s00521-019-04504-2.
[56] S. Minaee, E. Azimi, and A. Abdolrashidi, "Deep-Sentiment: Sentiment Analysis Using Ensemble of CNN and Bi-LSTM Models," 2019, *arXiv*: 10.48550/ARXIV.1904.04206.
[57] S. Wen *et al.*, "Memristive LSTM Network for Sentiment Analysis," in *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 51, no. 3, pp. 1794-1804, Mar. 2021, doi: 10.1109/TSMC.2019.2906098.
[58] J. Zhou, Y. Lu, H. N. Dai, H. Wang, and H. Xiao, "Sentiment Analysis of Chinese Microblog Based on Stacked Bidirectional LSTM," *IEEE Access*, vol. 7, pp. 38856-38866, 2019, doi: 10.1109/ACCESS.2019.2905048.
[59] F. Long, K. Zhou, and W. Ou, "Sentiment Analysis of Text Based on Bidirectional LSTM With Multi-Head Attention," *IEEE Access*, vol. 7, pp. 141960-141969, 2019, doi: 10.1109/ACCESS.2019.2942614.
[60] H. Elfaik and E. H. Nfaoui, "Deep Bidirectional LSTM Network Learning-Based Sentiment Analysis for Arabic Text," *Journal of Intelligent Systems*, vol. 30, no. 1, pp. 395–412, Dec. 31, 2020, doi: 10.1515/jisys-2020-0021.

[61] B. Talafha, A. Abuammar, and M. Al-Ayyoub, "Atar: Attention-based LSTM for Arabizi transliteration," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 11, no. 3, p. 2327, Jun. 2021, doi: 10.11591/ijece.v11i3.pp2327-2334.
[62] M. M. Abdelgwad, T. H. A Soliman, A. I.Taloba, and M. F. Farghaly, "Arabic aspect based sentiment analysis using bidirectional GRU based models," *Journal of King Saud University - Computer and Information Sciences*, Sep. 2021, doi: 10.1016/j.jksuci.2021.08.030.
[63] F. Liu, J. Zheng, L. Zheng, and C. Chen, "Combining attention-based bidirectional gated recurrent neural network and two-dimensional convolutional neural network for document-level sentiment classification," *Neurocomputing*, vol. 371, pp. 39–50, Jan. 2020, doi: 10.1016/j.neucom.2019.09.012.
[64] M. Al-Smadi, M. M. Hammad, S. A. Al-Zboon, S. AL-Tawalbeh, and E. Cambria, "Gated Recurrent Unit with Multilingual Universal Sentence Encoder for Arabic Aspect-Based Sentiment Analysis," *Knowledge-Based Systems*, p. 107540, Oct. 2021, doi: 10.1016/j.knosys.2021.107540.
[65] Q. Lu, Z. Zhu, F. Xu, D. Zhang, W. Wu, and Q. Guo, "Bi-GRU Sentiment Classification for Chinese Based on Grammar Rules and BERT," *International Journal of Computational Intelligence Systems*, vol. 13, no. 1, p. 538, 2020, doi: 10.2991/ijcis.d.200423.001.
[66] S. Sachin, A. Tripathi, N. Mahajan, S. Aggarwal, and P. Nagrath, "Sentiment Analysis Using Gated Recurrent Neural Networks," *SN Computer Science*, vol. 1, no. 2, Mar. 2020, doi: 10.1007/s42979-020-0076-y.
[67] J. S. Lee, D. Zuba and Y. Pang, "Sentiment Analysis of Chinese Product Reviews using Gated Recurrent Unit," in *2019 IEEE Fifth International Conference on Big Data Computing Service and Applications (BigDataService)*, 2019, pp. 173-181, doi: 10.1109/BigDataService.2019.00030.

BIOGRAPHIES OF AUTHORS

Syam Mohan Elankath is a Research Scholar in Department of Computer Science, Pondicherry University, India. He received his master's degree from St. Thomas' college, Thrissur, Kerala, India. His areas of interest include sentiment analysis, natural language processing, machine learning, and image processing. He can be contacted at email: syammohane@gmail.com.

Dr. Sunitha Ramamirtham is an Assistant Professor in Department of Computer Science, Pondicherry University, India. She has received her Ph.D. in Computer Science and Engineering. Her area of research includes E-learning, social network analysis, Spatio-temporal data analysis and Sentiment analysis. She has published her research articles in various reputed Journals and Conferences. She has co-authored a book. She is twice a recipient of the Best Teacher award of Pondicherry University. She can be contacted at email: sunitha.pondiuni@gmail.com.