ISSN 2278-3091
Volume 9, No.4, July – August 2020
International Journal of Advanced Trends in Computer Science and Engineering
Available Online at http://www.warse.org/IJATCSE/static/pdf/file/ijatcse165942020.pdf
https://doi.org/10.30534/ijatcse/2020/165942020

Sentiment Analysis by Novel Hybrid Method BE-CNN using Convolutional Neural Network and BERT

Prerna Mishra[1], Ranjana Rajnish[2], Pankaj Kumar[3]
[1]Amity Institute of Information Technology, Amity University Uttar Pradesh, Lucknow Campus, Lucknow (India), India. prerna21.mishra@gmail.com
[2] Amity Institute of Information Technology, Amity University Uttar Pradesh, Lucknow Campus, Lucknow (India), India. rrajnish@lko.amity.edu
[3] Department of Computer Science, Sri Ramswaroop College of Engineering & Technology, Lucknow, India. pk79jan@gmail.com

ABSTRACT

Sentiment analysis also well known as Opinion Mining is one of the important task of Natural Language processing for analyzing the mood of Customers. With the advancement in Internet and Big Data there is terabytes of data available from various Social Media platforms. Numerous models are available for text classification using deep learning techniques. Influenced by the favorable outcome of the deep learning models, we developed a method by using CNN (Convolutional Neural Network) and BERT (Bidirectional Encoder Representation from Transformers). This paper discusses about our novel method of sentiment analysis entitled as BE-CNN that uses 3 layers of Convolutional layer architecture for analysis of corpus collected from twitter related to "#Demonetization". We demonstrated by our experiments on corpus of 17000 tweets and the results are very encouraging showing that method BE-CNN, proposed by us outperforms other existing approaches of deep learning. We achieved the accuracy of 97% in comparison to other model and hence we can conclude that our method is better than other existing models.

Key words: BERT; Convolutional Neural Network; Deep Learning; Sentiment Analysis, Opinion Mining

1. INTRODUCTION

Sentiment analysis is the procedure for analyzing the sentiments (Positive, Negative and Neutral) of customers and has gained an increasing interest in the field of Natural Language processing. For growing any business or for understanding the business in a better way one should know the mood/feeling of customers with respect to any product or services. In the era of Big data, there is an enormous amount of data available for text classification. There are various social media platforms like Twitter, Facebook where customers express their sentiment regarding any topic. Nowadays Twitter is an emerging source for capturing the sentiments of public associated to any topic[1]. We used the twitter as our dataset source for this model and extracted 17000 tweets related to "Demonetization" topic.

Like various other fields, text classification has grabbed much attention of researchers for sentiment analysis. There have been various studies done by machine learning techniques using BOW (Bag-Of-Words) and TF-IDF (Term Frequency-Inverse Document Frequency) and have achieved better performance, but model needs to be pre-trained from the datasets by defining the features manually. Advancement in deep learning methods has made an impressive result in the field of text classification. Deep learning, a wider subfield from Machine learning and build by implementing artificial intelligence. Our model can learn and performs well by making intelligent decision on its own by using deep learning.

In this paper, we performed a sentiment classification on the corpus of 17000 records using CNN architecture. We proposed a method for sentiment classification by implementing 5-layer architecture using BERT embeddings. We assessed the execution of our proposed method (BE-CNN) with existing models through experimental results.

This paper is categorized in various sections as -
Section 2 reviews the corresponding work performed in the domain of sentiment analysis by using deep learning. Section 3 presents a detailed study about deep learning techniques, Section 4 describe about the proposed method and various steps involved in methodology. Results and performance evaluation of method is demonstrated in section 5.Paper is concluded in Section 6 by discussing conclusions of method.

2. RELATED WORK

Yoon Kim did a series of experiments by using a single layer of CNN by tuning various hyper parameters. Author proposed a modified architecture that performs well by the use of Static vectors and task-specific [2]. Proposed a model [3] by implementation of 2 neural networks with various parameters

like distinct architectural size and Word vector representation. For comparison of result, corpus used is BIG yelp dataset and small Yelp dataset and attained accuracy of 95.6%. Sentiment classification was done on corpus of Movie reviews by Hannah Kim[4].Compared result of model with other existing models, achieved accuracy of 68.3% for ternary classification and 28.1% for binary classification.

Xi Ouyang [5] proposed a hybrid framework by implementing 7 layers architecture of CNN model with Word2Vec for sentence classification. Performance results were evaluated on movie corpus having 5 labels positive, somewhat positive, negative, somewhat negative and neutral. Model achieved accuracy of 45.4% in comparison to other models.

Another deep learning model for sentiment analysis of tweets is proposed by Severyn [6]. In this author predicted the polarities of tweets at both message Level and Phrase level. Kalchbrenner [7] described a dynamic CNN architecture for semantic modeling of the sentences. Model is implemented by using dynamic K-Max pooling over linear sequences. Zhengjie [8] introduced a method using BERT for target-dependent sentiment classification, concluded results that method achieved almost similar performance in aspect-based sentiment analysis. V. Ganesh [9] proposed a method for intensity based sentiment classification using CNN and name it as CC-EISC. In this author classified the sentiments by four intensities which are Sad, Joy, Fear and Anger. Proposed method achieved accuracy of 82.7%, Precision of 77.08% and Recall of 62.52%.

3. DEEP LEARNING FOR SENTIMENT ANALYSIS

3.1 Bidirectional Encoder Representation from Transformers (BERT)

BERT is a technique for NLP developed by Google. It was developed and created by Jacob Develin in 2018.BERT is pre-trained model architecture. It is a deeply bidirectional, by bidirectional we mean it reads and interprets the meaning of text from both (left and right) sides. In generic models the text is read in sequential l (Left - right and Right - left) manner.

Left Context

We went to watch the movie
I want to buy a watch for my mother.

Right Context

In above example we can correctly understand the meaning of "watch", they both interpret different meaning with respect to left and right context. To deal with this type of problem, BERT provides a bidirectional approach of reading text. Thus, we can say that BERT embedding is a new way to achieve a pre-trained language model for text representation.

As shown in Figure1, BERT architecture has two variants BERT base and BERT large.

BERT Base: 12 Layers (transformer blocks), 12 Attention Heads, 768 Hidden units and 110 million parameters[10].

BERT Large: 24 Layers (transformer blocks), 16 Attention Heads, 1024 Hidden units and 340 million parameters[10].

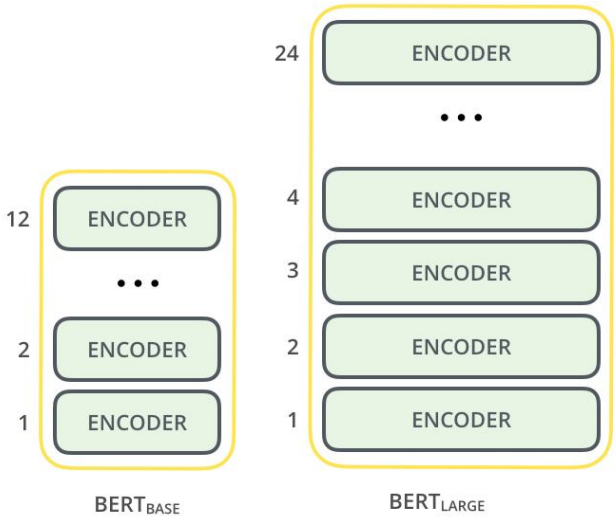


Figure 1: BERT Architecture [11]

BERT is mainly dependent on the mechanism of transformers which means a transformer interprets the contextual relationship between words in any text. Transformer basically comprises of two main parts: Encoder and Decoder. Task of an encoder is to read the text for input and decoder to build a prediction for the task. Input provided to the encoder in BERT is a sequence of tokens. In our method we used the BERT embedding for input to CNN model and hence [CLS] and [SEP] is added at the beginning and end of the sentence respectively for Token Embeddings.

3.2 Convolutional Neural Network (CNN)

Neural network is a huge innovation in the field of deep learning. It is a most popular machine learning algorithm.

Convolutional Neural Network (CNN) is a sub class of neural network. CNN is broadly used in various areas like Image recognition, text classification, etc. Neural network is formed by the basic unit of Perceptrons. Perceptron is the basic unit which consists of single neurons with various numbers of adjustable weights.

Perceptrons can be clearly understood as an entity which gets

holds of several inputs and generates a one output. A Single layer Perceptron is portrayed below in Figure 2, in which it takes 3 inputs x1, x2, x3 and generates one output 'Y'. An Output layer generates the output by weighted sum of all the inputs where w1, w2, w3 represents the weight of inputs x1, x2 and x3 respectively. Output can be either 0 or 1.

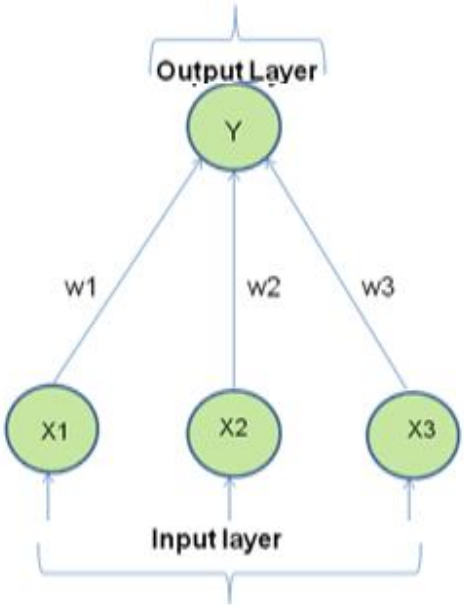


Figure 2: Single Layer Perceptron

Another type of perceptron is multilayer Perceptron (MLP) also named as feed forward neural networks CNN are basically based on multilayer perceptron as it has the great processing speed. It helps in solving complex problems. Figure 3 depicts the multilayer perceptron with 3 layers which are Input Layer, hidden layer and Output layer. Perceptron rom the first layer (Input layer) sends output to all the perceptrons in the hidden layer and then generates the output in Output Layer.

MLP with three layers is called as Shallow Neural Network and MLP with four or more than four layers is called as Deep Neural Network [12]

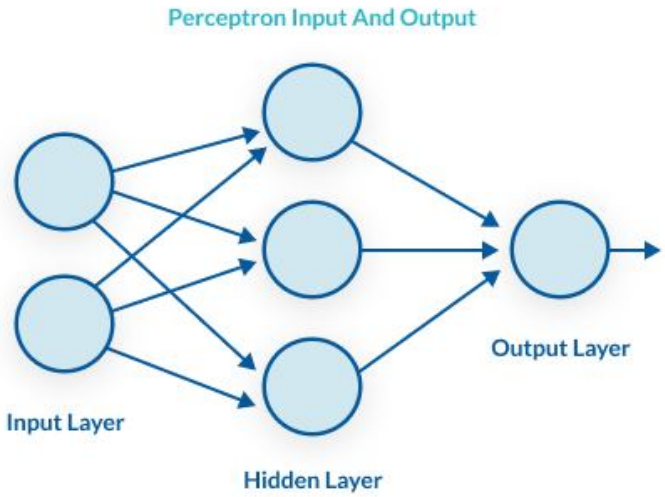


Figure 3: MultiLayer Perception [12]

Figure 4 shows the basic architecture of CNN which consists of 5 different layers Convolutional Layer, Pooling layer, fully connected layer and an Output layer. Role of input layer, Convolutional layer and pooling layer is feature extraction. Whereas Classification is done by the fully connected layer and output layer[13].

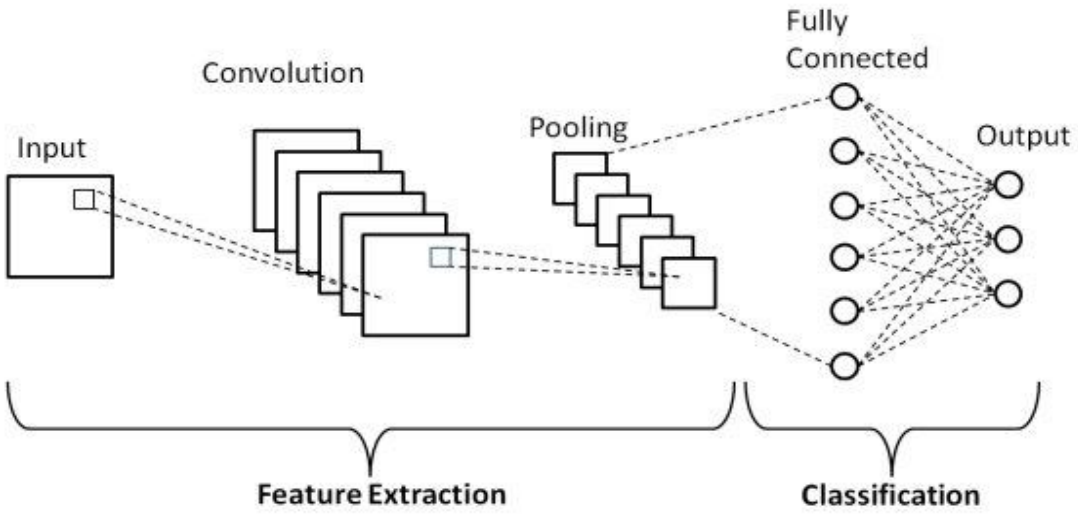


Figure 4: CNN Architecture Schematic Diagram [14]

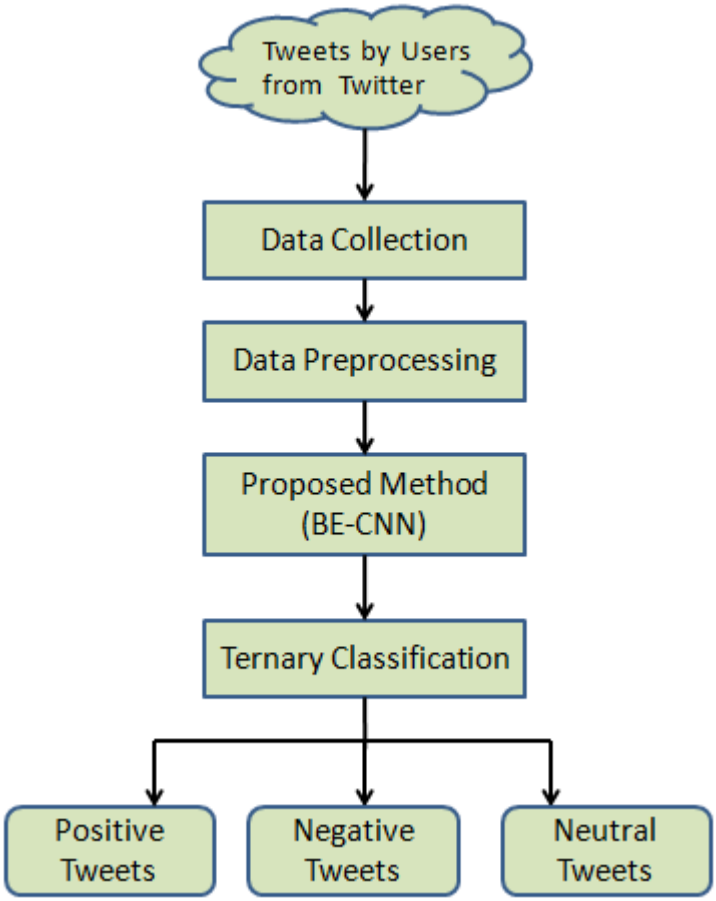


Figure 5: Process Flow diagram for Sentiment Classification

4. EXPERIMENTS

For analyzing sentiments of tweets extracted from twitter, we trained our method and achieved a better accuracy. Flow of our process is depicted above in figure 5. We used a Python 3.4 with NLTK library for our experiments.

4.1 Data Collection

We extracted tweets from twitter using Tweepy API related to "#Demonetization". So we have corpus of 17000 tweets which are stored in '.csv' file for further processing on our proposed method. CSV file contains two columns: Tweets and Polarity. Polarity column consists of three labels Positive, Negative and Neutral which makes our classification task as Ternary classification problem. We assigned some numerical values to Sentiment polarity as Positive: 1, Negative: 0 and Neutral: 2.

Some of the sample tweets are shown below in figure 6.

proffes10223951 i so wanted to feel like shit in life modi and his demonetisation did helped me feel so at personal level..
sambitswaraj mitron give me 50 days to fight against black money result rbi said demonetization failed mitron
they didn t cancel the project worth 1 126cr given to stec .. govt looted money from their citizens through demonitization
ndtv so please take bold step by banning chinese products in india like announcement of demonetization lock down
he denied even when the super one appreciated modi s demonetisation implement. more than the word of mouth or press
shefvaidya these people cried during demonetisation.
modi syndrome demonetisation currency of human life of economy what is not demonetised is it projected emi cum loss

Figure 6: Sample Tweets "#Demonetization"

4.2 Data Preprocessing

For further processing of our method, data should be in structured format. For this, cleaning of data is the preliminary step so that we can remove noisy data which is of no use in sentiment analysis. We removed punctuation marks, single spaces, html tags and any number present in the text. We used "RegEx" module for data cleaning. Thus, our data is now in structured format and ready for experiments on BE-CNN.

4.3 Proposed method & Implementation

In our proposed method (BE-CNN) for sentiment analysis, we have developed a hybrid method in which we have used features CNN with BERT embeddings to train the model on our text data for better accuracy. On the basis of proposed method BE-CNN, we have developed an algorithm that comprises of 3 Convolutional layers, 3 pooling layers, 1 Dense layer and 1 output layer. Flow of BE-CNN method is represented below in figure 7.

For implementation of algorithm based on our proposed method, we need to install BERT for Tensor Flow 2.0 and now to train any model, data should be in numerical format and it leads to conversion of data for numerical representation. In order to use BERT embeddings for our method, we first need to tokenize our tweets. For tokenization of tweets we used BERT Tokenizer and with the help of convert_tokens_to_ids () function of tokenizer object will get the ids related to tokens.For training any model, input of tweets should be of same length. We set a Batch size of 32 words for every tweet which means if any tweet is less than 32 words then we append it with 0 and if it is more than 32 words we take the average of extra words and assign it at 32 index number. Now we pass the numerical ids into BERT embedding pre trained model which generates an embedding value for each word. We used a pre-trained vectors model on Google News dataset which contains about 100 billion words for sentiment analysis. We can freely download file Googlenews-vectors-negative300.bin file from Google which contains approx. 300 dimensional vectors for 3 million words and phrases [4]. With this vector representation of all tweets we can train our CNN model.

Convolutional layers accept the input from BERT embeddings and start processing the input with certain parameters. Convolutional layer1 start processing the input from embedding layer with certain parameters like Kernel size of 2, Relu activation and Dropout. Main functionality of ReLu (Rectified Linear Unit) is to generate certain active neurons having value either 0 or 1, and it picks only active neurons which are represented by 1 to generate output. Furthermore, the first layer conveys information to second Convolutional layer having kernel size 3 for classifying the sentiments. Before passing information to next layer we applied MaxPool 1D which helps in reducing the size of input. Similarly, output from second layer is passed as input to 3 layer having kernel size 4. Now the outputs from the entire Convolutional network are concatenated together and are sent as input to Dense Layer. Dense layer converts the output into n*1 matrix so in case of ternary classification (Positive, Negative and Neutral) it will be represented as 3*1 matrix. Now we apply Sigmoid function to dense layer which predicts the probability as output.

So, processing our data through 3 Convolutional layers in BE-CNN method is successful in analyzing sentiments efficiently. This declaration is exhibited by our results in next section.

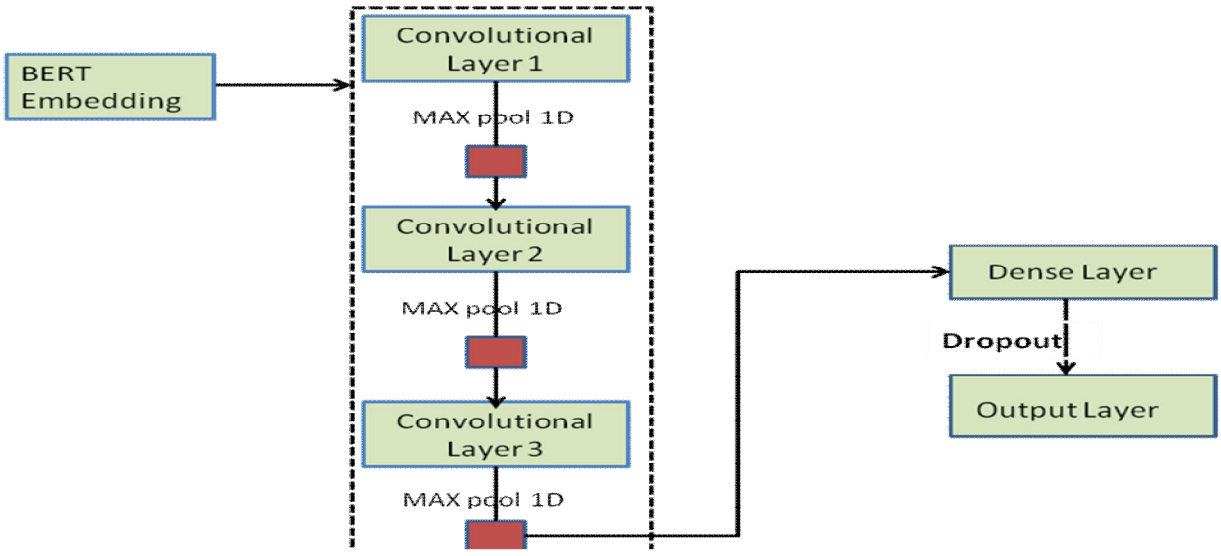


Figure 7: Flow diagram of BE-CNN Method

Table 1: Efficiency Measure of Algorithms

Model	Accuracy	Precision	Recall
BOW Gaussian Naïve Bayes	43.23%	49.15%	41.31%
TF-IDF Gaussian Naïve Bayes	96.48%	94.80%	96.00%
SVC Word Embeddings	53.49%	58.56%	51.33%
BE-CNN	97%	97.30%	97.51%

5. RESULTS AND DISCUSSION

For measuring the efficiency of BE-CNN method, we used classification accuracy metrics to measure the performance of our algorithm with other existing methods. We measured the performance of our algorithm by measuring Accuracy, Precision and Recall.

Table 1 depicts the results of all models for text classification on 17000 tweets related to demonetization. We can simply see that Accuracy measure of our method of approx 97% is much better in comparison to other models, although TF-IDF also performed well by achieving accuracy of 96.48%. BOW (Bag of Words) and SVC (Support Vector Classifier) attained an accuracy of 43.23% and 53.49% respectively which is very less in performance for text classification on such a huge amount of data. Graphical representation of Accuracy, Precision and Recall is depicted below in Figure 8, 9 and 10 by Bar Charts.

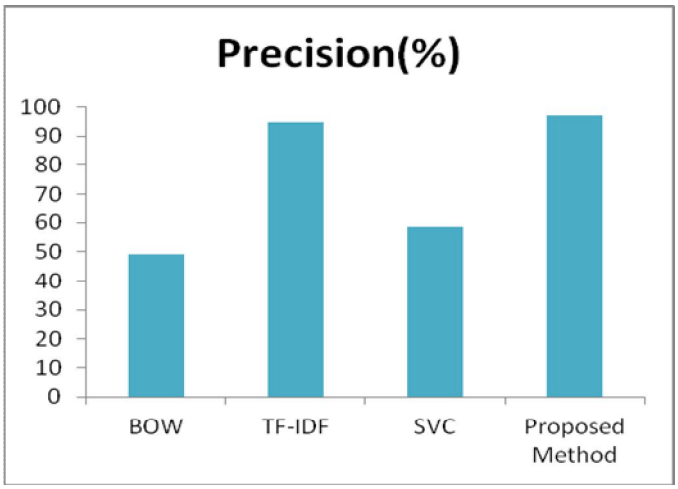
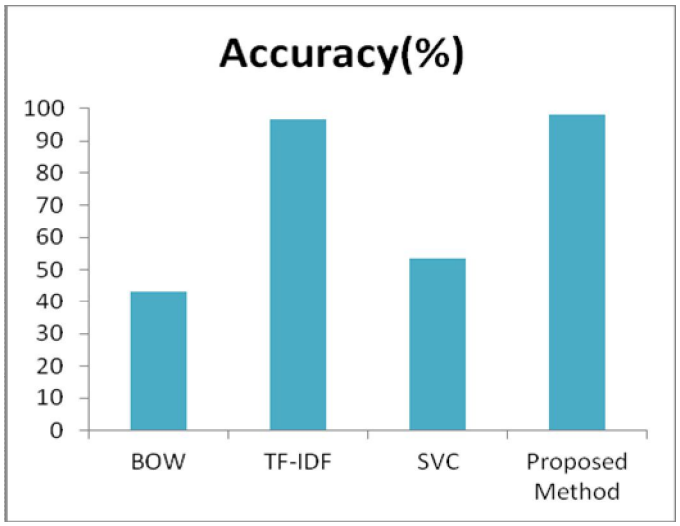


Figure 9: Precision Measure of Algorithms

Figure 8: Accuracy Measure of Algorithms



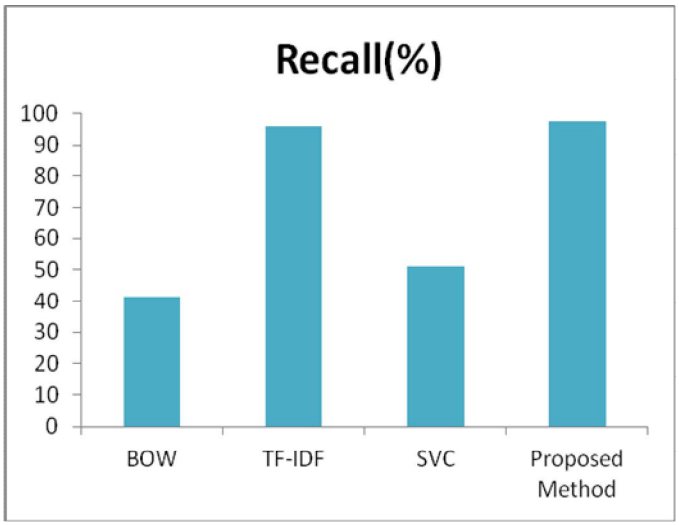


Figure 10: Recall Measure of Algorithms

6. CONCLUSION

Sentiment analysis is a demanding task for text classification. Nowadays, Data scientist & researchers are doing significant work for text classification by using various Machine Learning techniques and Deep Learning models. In this paper, we have proposed a hybrid method BE-CNN by implementing BERT with CNN. We have implemented the model in the form of algorithm on 17000 tweets, and thus from the results we can conclude that our proposed method is performing better with with other existing Machine Learning models like BOW, TF-Idf, and SVC. Our model outperforms existing models in all the measures i.e. Accuracy, Recall and Precision.

In the recent years, Deep learning has accomplished great results in field of Image Classification, text Classification and medical. We feel that our method BE-CNN can give better imagination to other researchers for better efficient model for the task of Sentiment analysis.

REFERENCES

1. J. P. Pinto, T. Vijaya Murari, and S. Kelur, "**Twitter sentiment analysis: A political view**," *Int. J. Adv. Trends Comput. Sci. Eng.*, 2020 doi: 10.30534/ijatcse/2020/103912020.
2. Y. Kim, "**Convolutional Neural Networks for Sentence Classification**," 2014, Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing ({EMNLP}),doi: 10.3115/v1/D14-1181.
3. A. Salinca, "**Convolutional Neural Networks for Sentiment Classification on Business Reviews**," 2017, Proceedings of IJCAI Workshop on Semantic Machine Learning, Melbourne, 2017.
4. H. Kim and Y. S. Jeong, "**Sentiment classification using Convolutional Neural Networks**," *Appl. Sci.*, vol. 9, no. 11, pp. 1–14, 2019, doi: 10.3390/app9112347.
5. X. Ouyang, P. Zhou, C. H. Li, and L. Liu, "**Sentiment analysis using convolutional neural network**," 2015, doi: 10.1109/CIT/IUCC/DASC/PICOM.2015.349.
6. A. Severyn and A. Moschitti, "**Twitter Sentiment Analysis with deep convolutional neural networks**," 2015, doi: 10.1145/2766462.2767830.
7. E. Grefenstette and P. Blunsom, "**A Convolutional Neural Network for Modelling Sentences,**" pp. 655–665, 2014.
8. 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.
9. V. Ganesh and M. Kamarason, "**Multi-labelled emotion with intensity based sentiment classification model in tweets using convolution neural networks**," *Int. J. Adv. Trends Comput. Sci. Eng.*, 2020, doi: 10.30534/ijatcse/2020/114922020.
10. M. S. Z. Rizvi, "**Demystifying bert the groundbreaking nlp framework**," https://medium.com/analytics-vidhya/demystifying-bert-the-groundbreaking-nlp-framework-8e3142b3d366.
11. J. Alammar, "**The Illustrated BERT , ELMo , and co . (How NLP Cracked Transfer Learning**)," *Blog*, 2018.
12. "**Core of deep Learning**." https://missinglink.ai/guides/neural-network-concepts/perceptrons-and-multi-layer-perceptrons-the-artificial-neuron-at-the-core-of-deep-learning/.
13. V. H. Phung and E. J. Rhee, "**A deep learning approach for classification of cloud image patches on small datasets**," *J. Inf. Commun. Converg. Eng.*, 2018, doi: 0.6109/jicce.2018.16.3.173.
14. V. H. Phung and E. J. Rhee, "**A High-accuracy model average ensemble of convolutional neural networks for classification of cloud image patches on small datasets**," *Appl. Sci.*, vol. 9, no. 21, 2019, doi: 10.3390/app9214500.