Fake News Detection in Social Media using a Novel FakeBERT Approach

M.Sudhakar*[1], K.P.Kaliyamurthie[2]

[1]Department of Computer Science and Engineering, Bharath Institute of Higher Education and Research, Chennai, India

[2]Department of Computer Science and Engineering, Bharath Institute of Higher Education and Research, Chennai, India

[1]sudhakarmtech@gmail.com, [2]kpkaliyamurthie@gmail.com

Abstract

Today, the news media has changed from offline to online, and this transformation will help the public to get information quickly and efficiently; in the same way, this media will spread phoney details rapidly. In recent research, many valuable methods were used to detect counterfeit information and analyse it unidirectional. In this research, we used bidirectional training approaches. We proposed two methods in this research. The first method is the deep learning approach as a Bidirectional Encoder Representation from Transformers (FakeBert) and a combination of Convolutional Neural Networks. This combination will help us manage the quality of detecting fake news. The proposed classification model FakeBert will provide better performance when compared to the existing model, and the accuracy is 99.90%.

Keywords: Bidirectional Encoder Representations for Transformers, Fake News, Prediction, CNN, Deep Learning.

1. Introduction

Today, the Internet is one of its most influential inventions, and many people use it to share their thoughts. Many people use the Internet for various purposes and can access multiple social media platforms from anywhere. This platform does not verify the user post and the identity of the users, and some people use this platform to share counterfeit information (Ahmed et al., 2021) [1]. Counterfeit information is an essential issue for the community and has a negative impact. It has created attention for researchers to develop a better solution for fake news (Gorrell G et al., 2019).

The phoney information can be written and shared to mislead the people and damage the company's integrity, either for financial or political benefits (Zhou X et al., 2019) . Figure 1. Examples of fake news that were trending during the Covid-19 pandemic will be shown. It is challenging for the public to detect fake information from people. To identify fake news, algorithms must be used. The main achievement of this research paper is to construct a machine-learning model that can predict which Tweets are about natural disasters and which are not. The dataset contains 10,000 tweets that have been manually classified (Ayub Ahmed et al., 2021). To optimize revenue, the application

helps organizations predict which articles will be popular so that their targeted advertising campaigns can be optimized. (An overview of RF algorithm in ML, 2020). The Random Forest Algorithm is a classification technique using a bundle of decision trees. It is also considered an effective technique as well as avoiding overfitting. It reduces overfitting and helps improve accuracy in decision trees.

This one applied to both problems of classification and regression. Continuous and categorical data can both be used. Automatically fills in any missing values present in the data. As a rule-based approach is used, there is no need to normalize data. While random forest algorithms have many advantages, they also have some disadvantages. These algorithms require time and resources with high computational power to build numerous trees to combine their outputs. As many decision trees are used to determine the class, training takes more time. Many decision trees are available, so it is also challenging to interpret and fails to determine the significance of each variable.

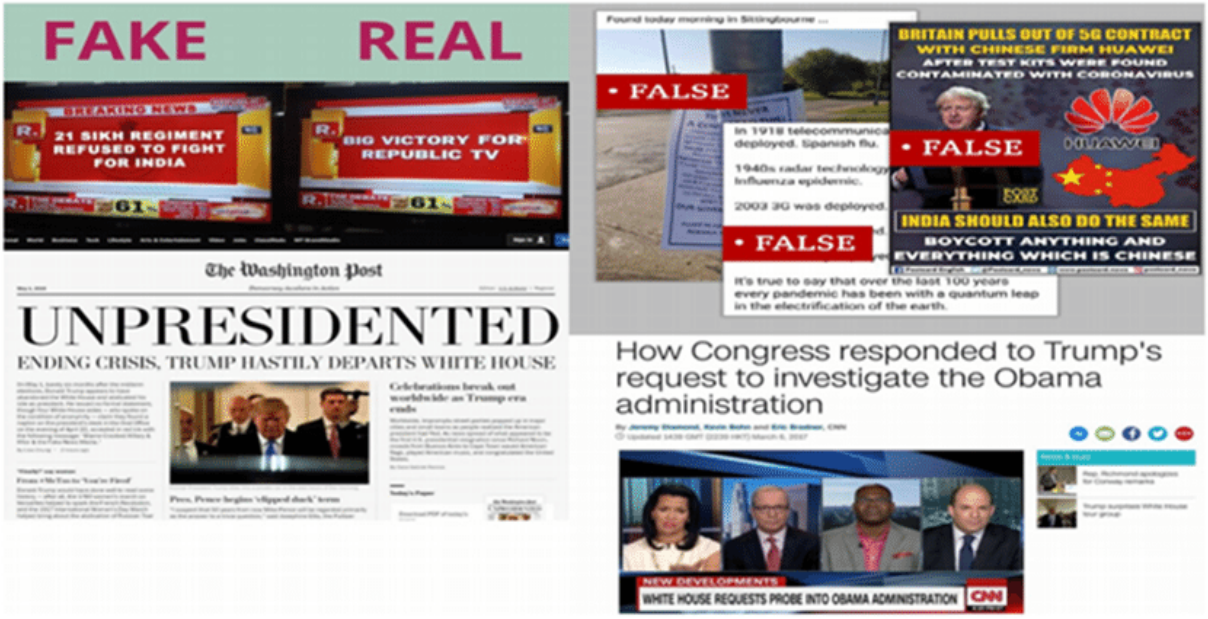

Figure 1. Social media fake news spreads.

2. Existing fake news detection approaches

The existing fake news detection methods are classified into content-based learning (News) and context-based learning (Social). The content-based approaches only deal with the different writing styles for published news articles (Shah C et al., 2018). In these techniques, our primary focus is extracting several features in fake news articles related to information. The fake news spreader regularly has the plan to spread the fake information to the public. In these learnings, style-based methodologies help capture manipulators' writing styles using linguistic features to identify phoney news. It is very complicated to detect fake news more accurately using only news content-based features (Abulaish et al., 2018).

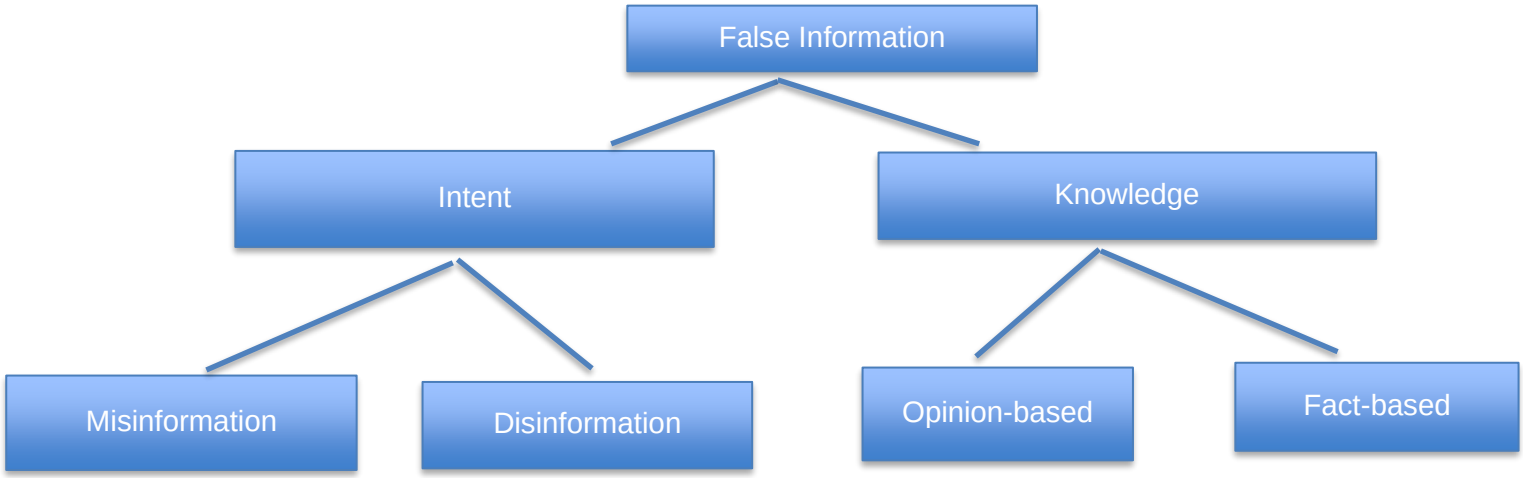

Figure 2. False information categorization

The context-based approach will deal with the information between the user and the articles, and social engagement is the significant relationship going to be used as a significant feature for fake news detection (Shu K et al., 2019). There are two methodologies used in this approach, the first methodology is Instance-based, and the second one is propagation-based. The Instance-based methodology is going to deal with the behavior of the user and the user's social media posts to scrap the integrity of the news. The propagation-based methodology will propagate the values between the users, posts and the news. Figure2 will shows the fake news detection approaches.

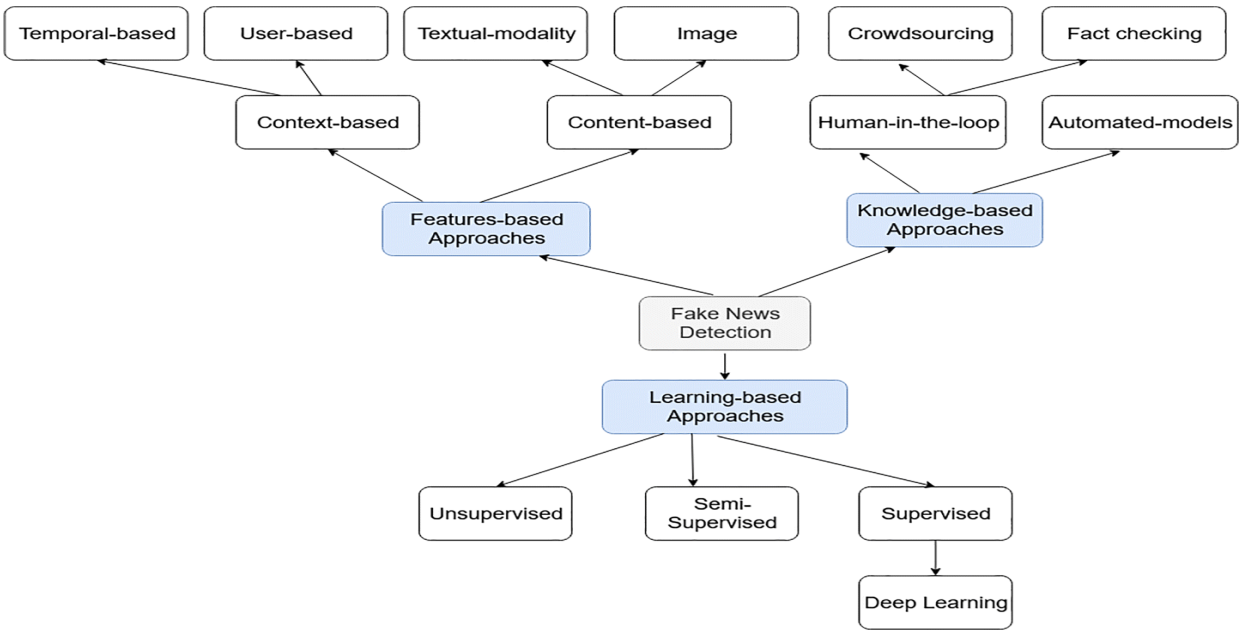

Figure 3. Fake news detection approaches

3. Related research works

This section will summarize the fake news discovery disquisition factory. (Kumar et al., 2018). Have explored a comprehensive check of different aspects of fake news. Different orders of fake news, algorithms for fake news discovery, and future aspects have been explored in this disquisition composition.

In one of the disquisitions, (Shin et al., 2018) excavated fundamental propositions across various disciplines to enhance the interdisciplinary study of fake news. The authors have mainly excavated the problem of fake news from four perspectives false knowledge it carries, writing styles, propagation patterns, and the credibility of its creators and spreaders. (Bondielli et al., 2019) have presented a crossbred approach for detecting automated spammers by integrating community-predicated features with other point orders, videlicet meta-content and commerce-predicated features.

In another disquisition, (Ahmed et al., 2017) have concentrated on automatically detecting fake content using online fake reviews. Authors have also explored two different point birth styles for classifying fake news. They have examined six machine knowledge models and shown bettered accomplishments compared to state-of-the-art marks. In one disquisition, (Allcott et al., 2017) concentrated on the fake news on the 2016U.S. Presidential General Election and its effect on U.S. Pickers. Authors have excavated the genuine and spurious BuzzFeed dataset used for the fake news from the URLs.

(Shu et al., 2019) Excavated a way for the robotization process through hashtag rush in one of the studies. In this disquisition composition, they have also done a comprehensive review of detecting fake news on social media, newsgroups on psychology and social generalities, and algorithms from a data mining perspective.

(Ghosh et al., 2018) Have excavated the impact of web-predicated social networking on political opinions. Authors have also explored the Twitter-predicated data of six Venezuelan government officers with a specific end thing to collaboration.

In numerous studies, the experimenters explored the problem of fake news. Their exploration (Ahmed et al., 2017) employed TF- IDF as a point birth system with different machine-literacy models. Expansive trials have been performed with LR and attained an accuracy of 89. Latterly, they have shown an accuracy of 92 using their LSVM. (Liu et al., 2018) Have delved into the styles for feting false tweets. In their disquisition, authors have employed a corpus of further than 8 million tweets gathered from the sympathizers of the presidential campaigners in the general election in the U.S. In their disquisition, they have employed deep CNNs for fake news discovery. In their approach, they employed the conception of subjectivity analysis and attained an accuracy of 92.10. (O'Brien et al., 2018) Have applied deep literacy strategies for classifying fake news. Their study achieved an accuracy of 93.50 using the black-box system. (Ghanem et al., 2018) Have espoused different word embedding, including n-gram features, to descry the stations in fake papers. They attained an accuracy of 48.80.

4. Methodology

In this section, we will discuss our proposed model and deep learning model architecture.

The word embedding will reduce the training time and improve classification performance. This embedding model will be used in the machine learning and deep learning model. Two-word embedding models convert the word into meaningful information. The GloVe is a weighted least square model that trains the model using co-occurrence counts of the words in the input vectors. It effectively leverages the statistical information's benefits by introducing the non-zero elements in a word-to-word co-occurrence matrix. The GloVe is an unsupervised training helpful model for finding the correlation between two words with their distance in a vector space (Qi y et al, 2018).

Word embedding will work with machine learning, and deep learning and this model will reduce the training time and improve overall classification performance. Pre-trained representations either be static or contextual. Contextual models generate a picture of each word based on the sentence (Peters et al., 2018).

4.1. BERT Model

BERT is one of the most recent prolific advances in natural language processing.BERT is an advanced-trained word embedding model predicated on motor-decrypted architecture. We use BERT as a judgment encoder, which can directly get the terrain representation of a judgment. BERT removes the unidirectional constraint using a mask language model. It erratically masks some of the commemoratives from the input and only predicts the original vocabulary id of the masked word predicated.

MLM has increased the capability of BERT to outperform as compared to former embedding styles. In this discourse, we have pulled embeddings for a judgment or a set of words or pooled the sequence of retired countries for the whole input sequence. A deep bidirectional model is more potent than a shallow left-section-to-right and right-section-to-left model.

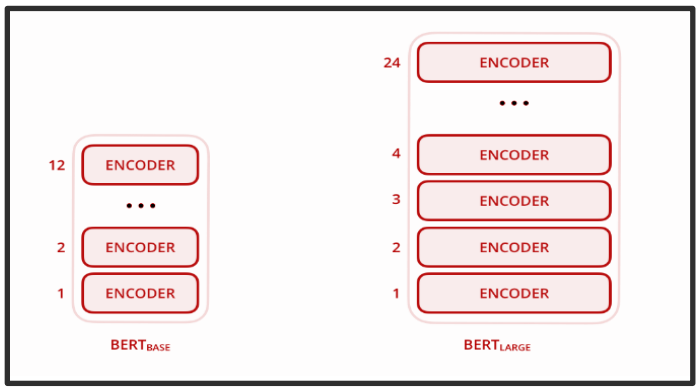

Figure 4. Architecture of BERT

This one has a Transformer encoder of multi-layer bidirectional. It has been implemented in BERT using 12 base layers, attention heads of 12, and parameters of 110 million. Figure 2 represents BERT architecture; figure 3 shows the combination of 3 embedding: 1. Token, 2. Segment, and 3. Position.

Name of the Parameter	Value of the parameter
Total Number of Layers	12
Size of Hidden Layers	768
Attention Heads	12
Total Number of Parameters	110M

Table.1 BERT Parameters

4.2. Convolutional Neural Network Model

Deep learning models are well-known for achieving state-of-the-art results in a wide range of artificial intelligence applications. This section provides an overview of the deep learning models used in our research with their architectures to achieve the end goal. Experiments have been conducted using deep learning-based models CNN and LSTM and our proposed model FakeBert with different pre-trained word embeddings. Figure 5 indicates the fake & real news from the word, the news has no title indicated as a no title and the words on the top-left are frequently used in fake news. The words on the bottom right indicate real news. The top fake words contain capital characters and special characters represent meaningless numbers. The top real words contain more names and verbs (Who, what, when, why, who).

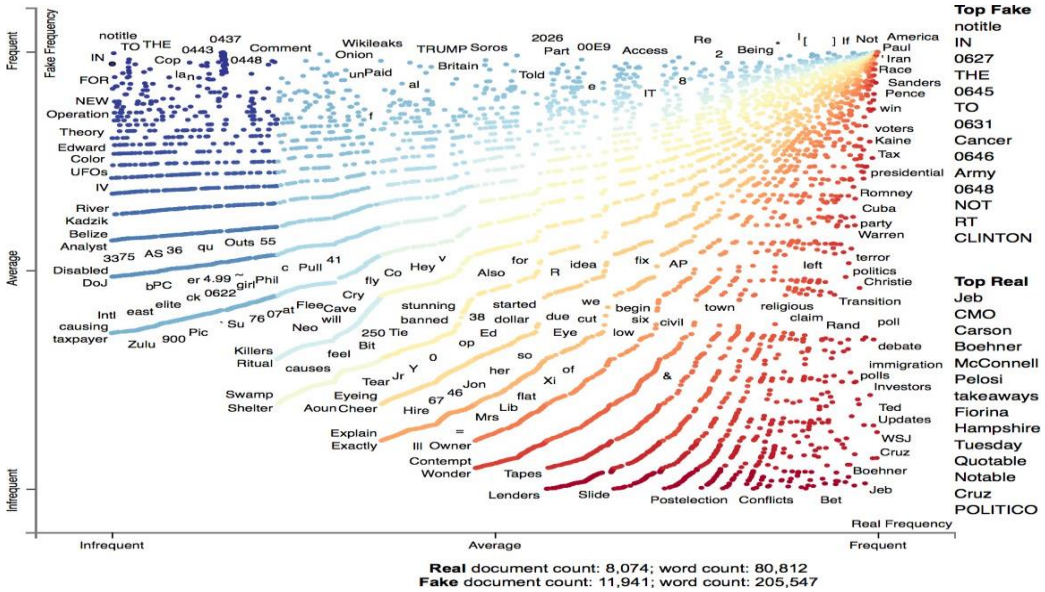

Figure 5. Real & fake news from word frequency in titles

4.3. Datasets

The dataset in this research article contains 20800 unique IDs, and the total number of titles for the particular news is 20242.

Attributes	Total number of Instances
Unique value to the new article	20800
Main heading related to the particular news	20242
Name of the news creator	18843

Full news article text	20761
Information about the article (fake or real)	20800

Table.2 Attributes of the fake news dataset.

Class Labels	Number of Instances
True	10540
False	10260

Table.3 Class label (Fake news dataset)

4.4. Proposed model Architecture

In this section, we introduce the proposed architecture CNN model and the explicit features. We used two parallel CNNs to extract latent features from both textual and visual information. And then explicit and latent features into the same feature space to form new representations of texts and images. At last, we propose tofuse textual and visual representations for fake news detection. As shown in Fig. 6, the overall model contains two major branches, first one is the text branch and the second one is the image branch. For each branch, taking textual or visual data as inputs, explicit and latent features are extracted for final predictions.

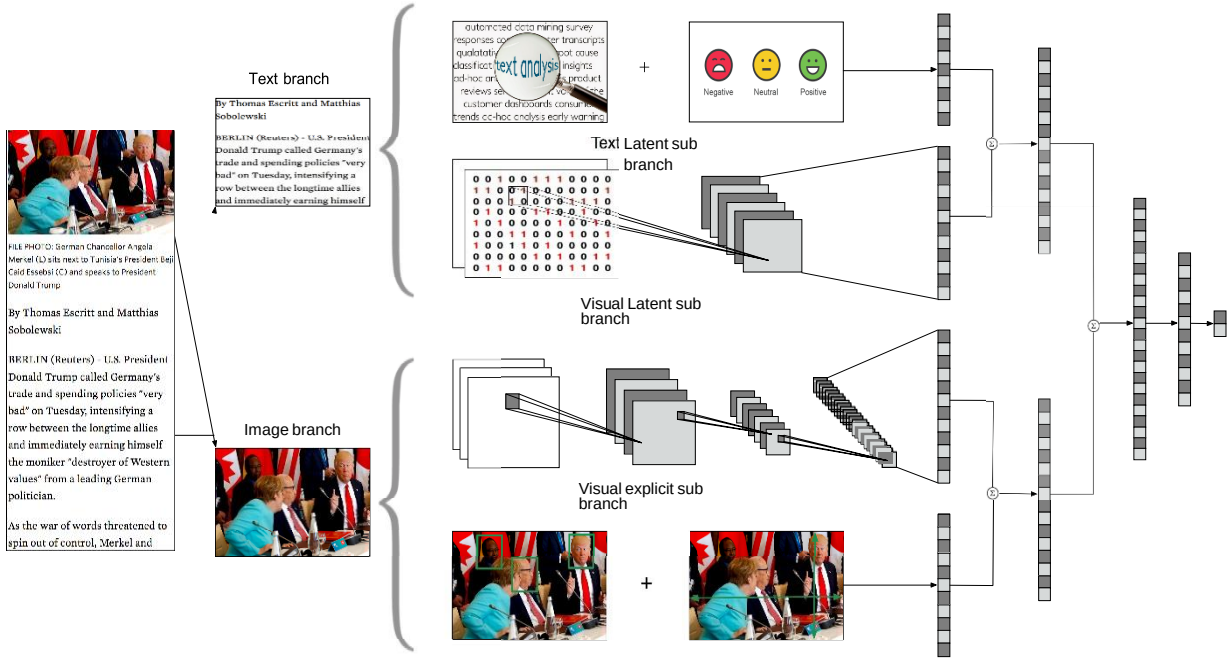

Figure 6. The architecture of the proposed model

5. Algorithms

Step 1: Data Cleaning

(a) "Stop words" usually refers to the most common words in a language. for example, 'a', 'the' etc. These words are essential parts of any language but do not add anything significant to the meaning of a word.

(b) To write a separate sentence we can use punctuation marks. It's also will help us to clarify the meaning of sentences.

(c) Convert all the messages in lowercase so that words.
(d) Convert the words to their lemma form.
(e) embedded special characters, "URLs" and finally digits are removed from the tweets

Step 2: Apply this vocab on our train and test datasets.

Step 3: Apply N-gram analysis.

Step 4: To know the real value information we can implement the method of embedding representation.

Step 5: The embedding layer is applied to the neural network with a Backpropagation algorithm.

Step 6: Efficiently applied Word2Vec.

Step 7: Continuous Bag-of-Words or CBOW mode is applied. The reason applies to this is to learn and to know how the current world is predicting its context.

Step 8: For predicting the surrounding words for a given current word we used CSG (Continuous Skip Gram) model.

Step 9: To calculate analogies, we used the GloVe algorithm is applied and classical vector space model representation of words using matrix factorization techniques.

Step 10: Apply CNN with Word Embeddings.

(a) Integers have been prepared to map the words. encode the tweets in the training dataset and make sure that all documents have the same length
(b) To find the longest review we are using the max () function. We will take the length from a training dataset and truncate tweets to the smallest size or zero-pad.
(c) Define the neural network model, the model with the embedding layer as the first hidden layer and specify the size of the real-valued vector space, and the maximum length of input documents.
(d) Maximum document length was calculated.

Step 11: To predict the Tweet analysis problem we develop a multi-channel CNN model.

(a) CNN configuration with 32 filters, linear activation function with the size of kernel 8.
(b) To interpret the CNN application in the back end we used SMPL
(c) The o/p value is in-between o to 1. Zero is negative and 1 is positive
(d) Fit a network on the training data having the parameters of stochastic gradient descent optimizer and training epochs as 100, to obtain the accuracy and loss of the metric.

Step 12: Make predictions on test data.

Step 13: Evaluate and compare the model.

7. Results

The performance evaluation of the BERT algorithm and Random forest classifiers on the disaster tweets dataset are discussed. Figure 7 shows that the accuracy rate is 99.90% when we use the BERT classifier and the Random forest classifiers. Figure 7

represents the ROC curve with an AUC of 99% obtained by the proposed algorithm.

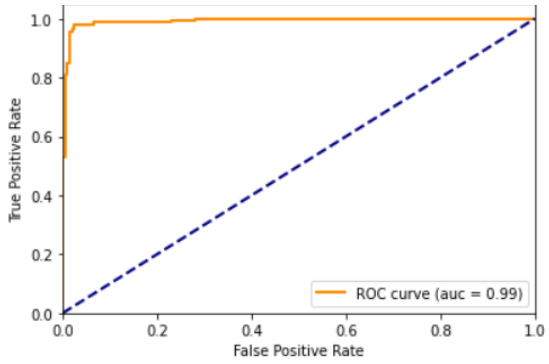

Figure 7: ROC curve proposed FakeBERT

The Training loss and validation loss obtained by the proposed FakeBERT algorithm in each epoch are represented in figure 8. It proves that the curve starts from 25% and gradually decreases and reaches below 0.05 in 30 epochs.

Figure 8: Training and validation loss of the proposed FakeBERT algorithm for 30 epochs

The Training accuracy and validation accuracy obtained by the proposed BERT algorithm in each epoch is represented in figure 9. It proves that the cure starts from 85% and gradually increases and reaches 99% in 30 epochs. Also, the Proposed BERT algorithm obtains a precision rate of 98%.

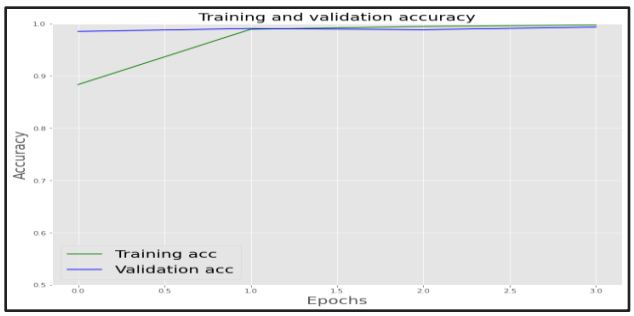

Figure 9: Training and validation Accuracy obtained by FakeBERT algorithm for 30 epochs

CONCLUSION

This paper it is explained and demonstrated the proposed FakeBERT algorithm performs better with the highest accuracy in the prediction of fake news. This work has extraordinary potential and can be effective in holding, improving and identifying fake news, hence it tends to be carried out on social networking sites like Twitter, Facebook and other social media.

REFERENCES:

1. Al Ayub Ahmed Et al., A. (2021). Detecting Fake News using Machine Learning: A Systematic Literature Review. In Psychology and Education Journal (Vol. 58, Issue 1, pp. 1932–1939). https://doi.org/10.17762/pae.v58i1.1046.

2. Gorrell G, Kochkina E, Liakata M, Aker A, Zubiaga A, Bontcheva K, Derczynski L (2019) SemEval-2019 task 7: RumourEval, determining rumour veracity and support for rumours. In: Proceedings of the 13th International Workshop on Semantic Evaluation, pp. 845–854.

3. Zhou X, Zafarani R (2018). Fake news: a survey of research, detection methods, and opportunities. arXiv:arXiv-1812

4. Ghosh S, Shah C (2018) Towards automatic fake news classification. Proc Assoc Inf Sci Technol 55(1):805–807.

5. Fazil M, Abulaish M (2018). A hybrid approach for detecting automated spammers in twitter. IEEE Trans Inf Forensics Secur 13(11):2707–2719.

6. Shu K, Cui L, Wang S, Lee D, Liu H (2019). defend: Explainable fake news detection. In: Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery and data mining, pp 395–405.

7. Kumar S, Shah N (2018). False information on web and social media: a survey. arXiv:arXiv-1804.

8. Shin J, Jian L, Driscoll K, Bar F (2018) The diffusion of misinformation on social media: Temporal pattern, message, and source. Comput Hum Behav 8:278–287.

9. Bondielli A, Marcelloni F (2019). A survey on fake news and rumour detection techniques. Inform Sci 497:38–55.

10. Ahmed H, Traore I, Saad S (2017). Detection of online fake news using N-gram analysis and machine learning techniques. In: International conference on intelligent, secure, and dependable systems in distributed and cloud environments. Springer, Cham, pp 127–138.

11. Allcott H, Gentzkow M (2017). Social media and fake news in the 2016 election. J Econ Perspect 31(2):211–36.

12. Shu K, Cui L, Wang S, Lee D, Liu H (2019). defend: Explainable fake news detection. In: Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery and data mining, pp 395–405.

13. Ghosh S, Shah C (2018). towards automatic fake news classification. Proc Assoc Inf Sci Technol 55(1):805–807.

14. Liu Y, Yi-Fang BW (2018). Early detection of fake news on social media through propagation path classification with recurrent and convolutional networks. In: Thirty-second AAAI conference on artificial intelligence.

15. O'Brien N, Latessa S, Evangelopoulos G, Boix X (2018). The language of fake news: Opening the black-box of deep learning based detectorsReturn.

16. Ghanem B, Rosso P, Rangel F (2018). Stance detection in fake news a combined feature representation. In: Proceedings of the first workshop on fact extraction and Verification (FEVER), pp 66–71.

17. Qi Y, Sachan D, Felix M, Padmanabhan S, Neubig G (2018) When and why are pre-trained word embeddings useful for neural machine translation? In:

Proceedings of the 2018 conference of the north american chapter of the association for computational linguistics: human language technologies, vol 2 (short papers), pp 529–535.

18. Peters ME, Neumann M, Iyyer M, Gardner M, Clark C, Lee K, Zettlemoyer L (2018) Deep contextualized word representations. In: Proceedings of NAACL-HLT, pp 2227–2237