BERT Model for Classification of Fake News using the Cloud Processing Capacity

Athiya Marium
Information Science and Engineering
RV College of Engineering
Bengaluru, India
athiya2895@gmail.com

Dr. G S Mamatha.
Information Science and Engineering
RV College of Engineering
Bengaluru, India
mamathags@rvce.edu.in

***Abstract*—This paper aims at conducting a predictive analysis on news articles in order to find if they are fake or real. After conducting an extensive research on the topic, various Machine Learning and Deep Learning models for the purpose of evaluating news articles were discovered. A new transfer learning model, Bi-directional Encoder Representation for Transformers (BERT), is tested using the Google Cloud GPU capacity for the purpose of detection. The first step in this direction will be to pre-process the data to clean out the garbage and missing values. After this, all the news articles collected will be tokenized, according to the BERT tokenizer. The tokenized corpus will be converted into tensors for the model to be trained. The data will be trained in batches with each batch having 32 articles. The final layer for training will consist of a five layered neural network. The model with the least validation loss will be tested for accuracy. The predictions for news articles will be made on this model. The paper will also explore the best cloud platform to host such a model and performance of the hosted model as well.**

Keywords— BERT, NLP, ML, Fake news, Google Colab

I. INTRODUCTION

The proposed framework is used to detect the fake news from e-newspapers and social media platforms and to detect whether the content could be potentially fake or harmful. More and more news is being shared on Social Media Platforms and websites, since it is cheaper and easier than traditional newspapers and reaches far more demographic, and some of these sites like Facebook and Twitter have become active news sharing mediums. Sadly, this has given rise to a lot of fake news being generated, maybe to spread propaganda or to generate tensions and spread maliciousness.

The framework will use a transfer learning model, BERT, which stands for Bi-directional Encoder Representation for Transformers. Bi-directional means that each word will derive its context from the words coming before and after it. For example, "I will run a marathon" and "I will run for the elections", has two different meanings for the word run. BERT will be able to understand the two varying meanings of the word and assign two different tokens for the two words.

Encoder Representation means that each word in the corpus will be assigned a different token according to the BERT tokenizer. A transformer model takes the learnings of a pre-trained model, in this case BERT, and trains its own data according to the pre-trained model. In this regard, the BERT model has been trained with an extensive vocabulary, including 1500 million words from Wikipedia and 800 million from bokus corpus.

The paper will cover a literature survey conducted on the topic, a brief description of the dataset considered, a description of the models considered after performing basic NLP techniques on the dataset followed by the architecture of the BERT model with the description of the processing techniques and finally the results as comparative study between the two methodology and conclusions.

II. LITERATURE SURVEY

The paper by Akash Junnarkar et al. [1] studies the spam classification problem and proposes a solution which uses NLP and a URL based differentiation method. In order to create a model which has higher accuracy and efficiency, different ML algorithms are used. The study talks about bettering the accuracy scores by using algorithms like XGBoost.

A study for detecting fake news from new articles which are written in the ALovak language has been given in the paper by Klaudia Ivancova et al. [2]. The solutions for fake news detection mostly have data which is written in English for the purpose of training. In order to use the same solutions in a different language, the models just need to be trained with the datasets collected and labelled in the Slovak language.

A recent paper helps in understanding the virus that caused the pandemic on COVID-19 [3]. NLP techniques were used to understand some important factors of the virus by extracting information from newspaper articles about the pandemic. An extensive dataset containing all the articles relating to the information about the pandemic can make for an even better and improved model result.

The next paper studied for the research, assigned individual scores for influencers from all fields, be it entertainment, politics or athletics. This generation is done by using a bias calculator [4]. The bias calculator is built using techniques of sentiment analysis, which will give the sentiment - whether positive or negative - of a tweet. The calculator also uses natural language methods to check the amount of bad words which were used by a person in their twitter.

The Transfer Learning model of BERT is used for classifying emotions and behavior modelling given by Austin Hembree et al. [7]; using this a model is created which can predict emergencies from Twitter with a good accuracy score.

The paper by Acheampong Francisca Adoma et al. [8] talks about a comparative study between different transformer-based models. The different models considered are RoBERTa, XLNet, BERT and DistilBERT. Emotions of anger, sadness, shame, joy, guilt, surprise, and fear are checked using these models on the ISEAR (International Survey on Emotion Antecedents and Reactions) dataset of 7666 sentences. The BERT model performed third best from the lot with an accuracy of 0.7009. The model RoBERTa performed the best with an accuracy, which is marginally better than BERT, of 0.7431. This paper goes to prove that the

2021 IEEE 9th Region 10 Humanitarian Technology Conference (R10-HTC) | 978-1-6654-3240-5/21/$31.00 ©2021 IEEE | DOI: 10.1109/R10-HTC53172.2021.9641632

978-1-6654-3240-5/21/$31.00 ©2021 IEEE

transformer model for Natural Language Processing performs much better.

The paper by Jaideep Yadav et al. [9] tries detecting cyberbullying using the BERT model. They use two different datasets, Formspring and Wikipedia, with conversations and labels of whether the comment is bullying or not. The data is tokenized using the BERT model for each sentence with the toekn_id and attention mask of each sentence in the dataset. Using the BERT-base-model and a deep neural network for the classification, the model is trained with the dataset. The accuracy of 96% was reached for the Formspring dataset and 98% for Wikipedia dataset.

The paper by Bhavesh Pariyani et al. [10]. uses classification algorithms for determining hate speech among social media content. The content is represented using tf-idf vectors and bag of words vectors. Algorithms like logistic regression are applied next. Accuracy of the classification can be improved by trying more machine learning algorithms.

A classification of COVID-19 misinformation using the BERT model gave an accuracy of 72%. BERT based models like RoBERTa performed 79%, with the model AlBERTa performing the best among them with accuracy of 88%[4]. An LDA topic modeling technique is used which collects all the tweets that are talking about Covid-19. A sentiment analysis is carried out to check if the tweet is positive, negative or neutral. This analysis is done using the BERT, RoBERTa and AlBERTa models and the accuracies of each are compared

A total of 18 papers were considered for the literature survey. Fake news detection techniques involve sentiment analysis, article abstraction and using techniques which are used to detect spam emails. Along with this Machine Learning and Deep Learning techniques were also used. The algorithms that performed the best were: Naive Bayes, NLP techniques and Neural network.

III. Data

The dataset taken from Kaggle Fake News Detection challenge has about 20,000 articles with their authors, title and text along with the label of whether the article is real or fake. For training we will consider the text and label fields.[12]

	text	label
House Dem Aide: We Didn't Even See Comey's Let...		1
Ever get the feeling your life circles the rou...		0
Why the Truth Might Get You Fired October 29, ...		1
Videos 15 Civilians Killed In Single US Airstr...		1
Print \nAn Iranian woman has been sentenced to...		1

Fig. 1. Dataset with text and label fields for training

From the 20761 articles in the training dataset, 30% of the data is kept aside for testing. Out of this 30%, about 5% data is used for validation. The fig 1 shows the top 5 rows in the dataset.

A. Dataset Distribution

The dataset has equal distribution for real and fake articles, as shown in fig 2, which means we will not be needing a weight parameter for the training data.

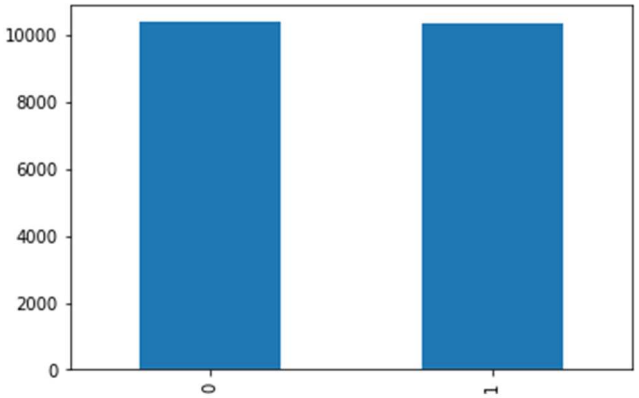

Fig. 2. The Bar chart gives a representation of real(1) and fake(0) data

B. Word Count

Each article in the corpus is checked for its length by counting the number of words in the text field. From fig 3, we see that the number of words are maximum in the range of 0-100. We will take this as the value of the maximum length in order to pad the other articles with the same number of words.

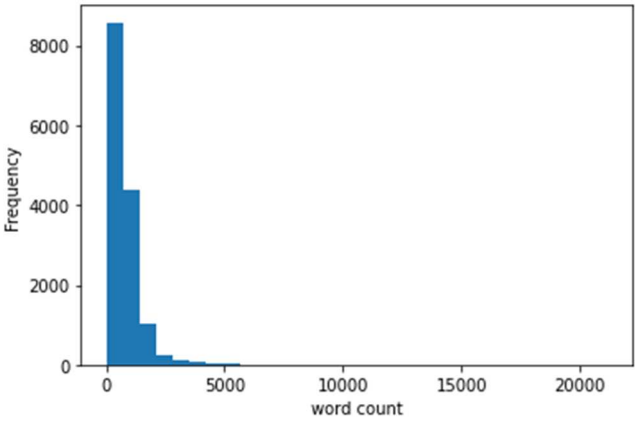

Fig. 3. A histogram of word count distribution for text field

IV. Methodology using Natural Language Processing

As described in the fig 4, following steps are carried out for detecting fake news:

Collection of data: News articles and social media posts which are categorised into real and fake. The data has to be relevant to the area under consideration and the categories have to be verified. For this paper the dataset collected was from Kaggle.

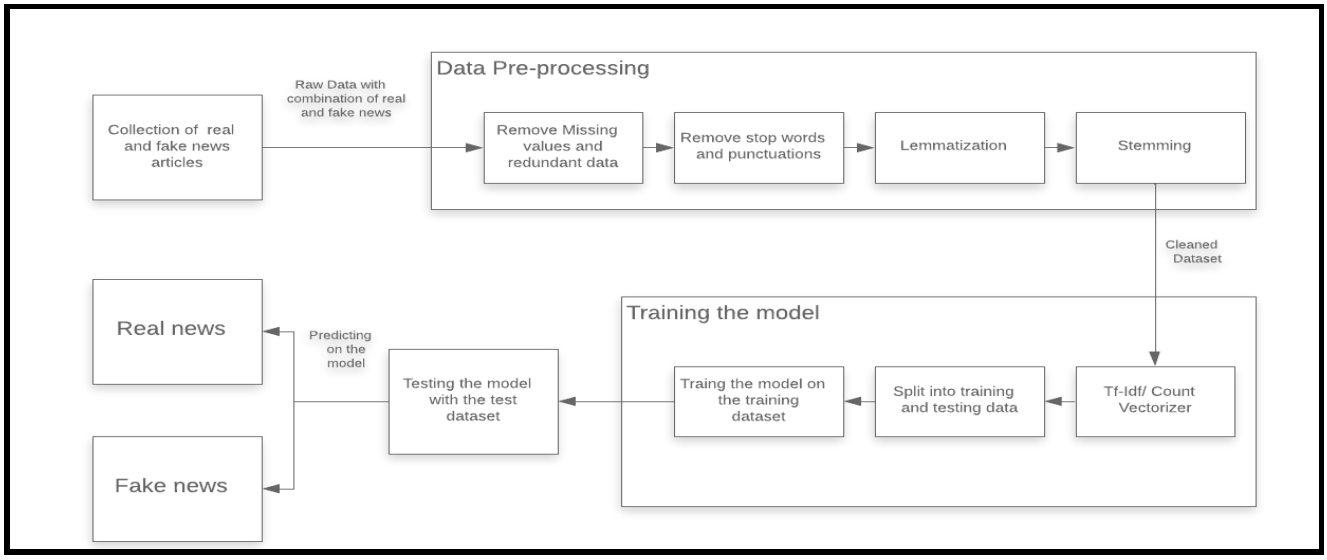

Fig. 4. Architecture for Fake news Detection model

Data pre-processing: The step requires the data to be cleaned of missing values, redundant values. Following that the next step would be to remove punctuations and stop words like a, the, and etc. from the data corpus. The following are the steps carried out:

- Removing the missing values - Some of the text fields in the dataset are empty. This throws an error when training. We take care of these values by substituting empty strings in these rows.
- Removing Unnecessary Data - Stops words from english for example, a, the, an, them, what, is, where etc which do not add to the context of the article along with punctuations are removed.
- Lemmatization - Lemmatization, reduces the inflected words properly ensuring that the root word belongs to the language. This is done so that words like election, elect, electing are all reduced to the root word elect
- Stemming - Stemming is the process of reducing inflection in words to their root forms such as mapping a group of words to the same stem even if the stem itself is not a valid word in the Language.

The fig 5 shows the output of each stage in pre-processing and the final dataset is the rightmost column which will be used in the next steps.

Vectorizer: A vectorizer creates a vocabulary of all words in all the tuples of the dataset and assigns each word a number or a token. The corpus of data is then transformed such that each document is represented by the vectors of the words in it. This makes it easier to train the model which will come next. The algorithms that can be used include Count Vectorization, Tf-idf Vectorization etc.

Training: The dataset is split into training and test sets. The training data is fed into the machine learning model. The model is trained with about 80% of the data in the dataset. This step will ready our model to detect fake news from new data with a confidence score associated with it.

Testing: Testing is done on 20% of the dataset. The step will categorize the test set into real and fake which will be compared to the original labels. We will be able to get the accuracy of the model from this step.

Predicting: New articles can now be fed into the model to predict whether they are fake or real with a certain level of accuracy

	id	title	author	text	label	body_text_clean	body_text_tokenized	body_text_nostop	body_text_lemmatized	body_text_cleaned
0	0	House Dem Aide: We Didn't Even See Comey's Let...	Darrell Lucus	House Dem Aide: We Didn't Even See Comey's Let...	1	House Dem Aide We Didn't Even See Comey's Lett...	[house, dem, aide, we, didn, t, even, see, com...	[house, dem, aide, even, see, comey, letter, j...	[house, dem, aide, even, see, comey, letter, j...	[hous, dem, aid, even, see, comey, letter, jas...
1	1	FLYNN: Hillary Clinton, Big Woman on Campus - ...	Daniel J. Flynn	Ever get the feeling your life circles the rou...	0	Ever get the feeling your life circles the rou...	[ever, get, the, feeling, your, life, circles,...	[ever, get, feeling, life, circles, roundabout...	[ever, get, feeling, life, circle, roundabout,...	[ever, get, feel, life, circl, roundabout, rat...
2	2	Why the Truth Might Get You Fired	Consortiumnews.com	Why the Truth Might Get You Fired October 29, ...	1	Why the Truth Might Get You Fired October 29 2...	[why, the, truth, might, get, you, fired, octo...	[truth, might, get, fired, october, 29, 2016, ...	[truth, might, get, fired, october, 29, 2016, ...	[truth, might, get, fire, octob, 29, 2016, ten...
3	3	15 Civilians Killed In Single US Airstrike Hav...	Jessica Purkiss	Videos 15 Civilians Killed In Single US Airstr...	1	Videos 15 Civilians Killed In Single US Airstr...	[videos, 15, civilians, killed, in, single, us...	[videos, 15, civilians, killed, single, us, ai...	[video, 15, civilian, killed, single, u, airst...	[video, 15, civilian, kill, singl, u, airstrik...
4	4	Iranian woman jailed for fictional unpublished...	Howard Portnoy	Print \nAn Iranian woman has been sentenced to...	1	Print \nAn Iranian woman has been sentenced to...	[print, an, iranian, woman, has, been, sentenc...	[print, iranian, woman, sentenced, six, years,...	[print, iranian, woman, sentenced, six, year, ...	[print, iranian, woman, sentenc, six, year, pr...

Fig. 5. Dataset after pre-processing stage

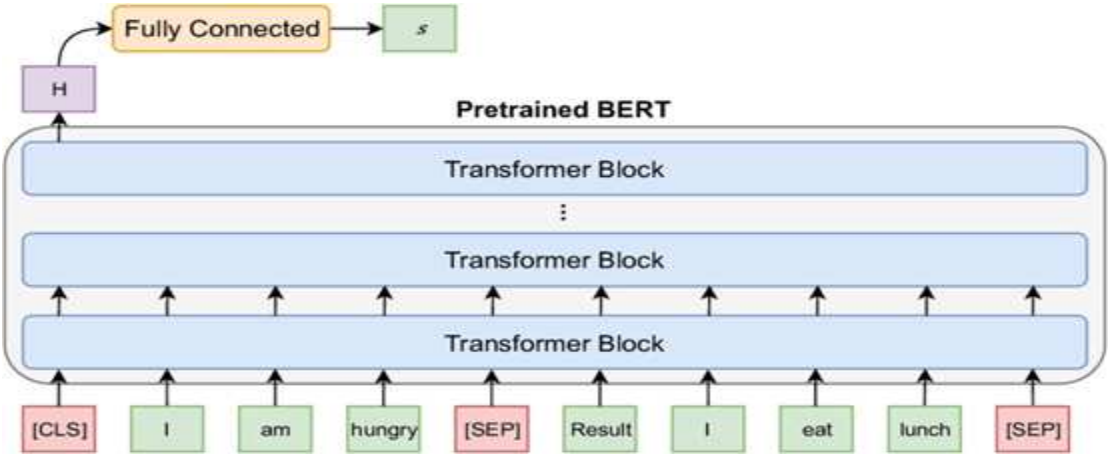

Fig. 6. The architecture of the fake news detection tool using the BERT base model [15]

V. BERT Model for Fake News Detection

A. Architecture

From the fig 6, we see that model using BERT architecture has the following components:

A pre-trained BERT model, which will have tokenizers to convert each incoming article of the corpus into a vector of tokens. Each token will have the contextual information of a given word which will be ideal in training the model. The BERT model we use for this tool is the BERT base model which has 12 encoders. The other model is the BERT large with 24 encoders.

The output of the tokenizer from the BERT model will be a set of mathematical tokens or integers ideal for training a neural network. This is represented by the H in the Figure 6.

The tokenized data after being split into training, testing and validation sets is passed on to a fully connected neural network. This network will be trained with the training set for 5 epochs and the validation set will give the training losses. The model with the least loss will be taken as the final model for prediction and accuracy purposes.

B. Tokenization

Tokenization is the process of converting English words into integers or tokens that can be understood by the neural network for training. For example, we us consider two sentences: 1."A BERT model will give good accuracy", 2. "any data can be transformed using a BERT model".

Tokenized output for these sentences will be:

```
{
    'input_ids': [[101, 1037, 14324, 2944, 2097, 2507,
    2204, 10640, 102, 0, 0], [101, 2151, 2951, 2064,
    2022, 8590, 2478, 1037, 14324, 2944, 102]],

    'attention_mask': [[1, 1, 1, 1, 1, 1, 1, 1, 1, 0, 0], [1,
    1, 1, 1, 1, 1, 1, 1, 1, 1, 1]]
}
```

The input_ids fields give the tokens or integers corresponding to each word. The fields with 0 entry are the padding inserted in order to make the two sentences of the same length. The attention_mask field indicates which of the words in the sentence have to be taken as priority and which can be ignored. An attention mask of 1 indicates an important word, whereas an attention mask of 0 indicates the word can be ignored.

C. Neural Network Layer

A fully connected neural network layer will make the final output layer of our model.

This model has a dropout layer with a RelU activation function, followed by two dense layers with SoftMax activation function. The training of the model is done in batches of 32 articles. After every batch the gradient is shifted.

D. Processing

In order to train the 20,000 odd articles in a neural network, we use a GPU or a Graphical Processing Unit provided by the Google Colab platform. This GPU is utilized better by using parallel processing. CUDA which is a parallel processing platform was developed by Nvidia for compute intrinsic processing's. Using CUDA platform with GPU on the Colab makes for a high speed efficient way to train the neural network with the large dataset that is given.

VI. Results

A. Accuracy and Loss

Using the 30% test data, the model is evaluated for accuracy and loss. The final model has a Training Loss of 0.419 and a Validation Loss of 0.345. The loss is calculated using a cross entropy function.

B. Precision and Recall

The precision and recall are calculated for each label of the output. From this confusion matrix other metrics are calculated using the equations (1), (2), (3) and (4) [13].

The values in the formulas are explained below with reference to the dataset collected in the project:

- TP - True Positives - articles which were correctly predicted as their actual result
- FP - False Positives - articles that were wrongly predicted as correct labels.
- FN - False Negatives - articles that were wrongly predicted as wrong labels.
- TN - True Negatives - articles which are actually Fake and are correctly predicted as Fake.

$$Precision = \frac{TP}{TP + FP} \quad (1)$$

$$Recall = \frac{TP}{TP + FN} \quad (2)$$

$$Accuracy = \frac{TP}{TP + FP + TN + FN} \quad (3)$$

$$F1\ Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (4)$$

C. Performance with Earlier NLP Techniques

After carrying out the NLP techniques before performing ML algorithms, as described in the first methodology, the prediction on the test data was compared with the actual values for accuracy.

The accuracies for some of the models that were used on the data for detecting fake news is summarized in Table I. The table concludes that the Neural Network model works best for fake news detection after the necessary NLP steps are carried out.

Compared to the BERT model, the ML models after performing basic NLP pre-processing works better, but the BERT model can be fine-tuned to work better for any dataset including the one described in this paper. The next section describes the performance of the BERT model.

TABLE I. THE COMPARISON TABLE FOR FAKE NEWS PREDICTION MODELS USING EARLIER NLP TECHNIQUES

Models and Algorithms used	Accuracy Score	Precision Score	Recall Score
Neural Network	95.51	95.26	94.35
Support Vector Machine	94.25	94.55	94.15
Naive Bayes	90.02	97.81	82.28

D. Performance of the BERT Model

The fig. 7 gives the confusion matrix for the BERT model. The confusion matrix describes the comparative output for the labels of real (0) and fake (1). The precision, recall and F1 scores for the BERT model predictions are defined in table II.

Fig. 7. Confusion matrix for the test set using BERT

TABLE II. PRECISION, RECALL AND F1 SCORES OF MODEL USING BERT MODEL

Output	Precision	Recall	F1 Score
Fake	0.83	0.88	0.85
Real	0.87	0.83	0.85

VII. CONCLUSION

The BERT model is one of the best pre-trained models for Natural Language Processing, as it is trained with a large corpus of data. The vocabulary of the BERT model is exhaustive for English language. There is also context based tokenization which makes it easier for the model to differentiate between two different meanings of the same word. For the current Fake News use case, the BERT model makes it easier for the Neural Network layer to predict if news articles could potentially be fake.

REFERENCES

[1] A. Junnarkar, S. Adhikari, J. Fagania, P. Chimurkar and D. Karia, "E-Mail Spam Classification via Machine Learning and Natural Language Processing," 2021 Third International Conference on Intelligent Communication Technologies and Virtual Mobile Networks (ICICV), Tirunelveli, India, 2021, pp. 693-699, doi: 10.1109/ICICV50876.2021.9388530.

[2] K. Ivancová, M. Sarnovský and V. Maslej-Krčšňáková, "Fake news detection in Slovak language using deep learning techniques," 2021 IEEE 19th World Symposium on Applied Machine Intelligence and Informatics (SAMI), Herl'any, Slovakia, 2021, pp. 000255-000260, doi: 10.1109/SAMI50585.2021.9378650.

[3] N. Sadman, N. Anjum, K. Datta Gupta and M. A. Parvez Mahmud, "Understanding the Pandemic Through Mining Covid News Using Natural Language Processing," 2021 IEEE 11th Annual Computing and Communication Workshop and Conference (CCWC), NV, USA, 2021, pp. 0362-0367, doi: 10.1109/CCWC51732.2021.9376002.

[4] R. Raju, S. Bhandari, S. A. Mohamud and E. N. Ceesay, "Transfer Learning Model for Disrupting Misinformation During a COVID-19 Pandemic," 2021 IEEE 11th Annual Computing and Communication Workshop and Conference (CCWC), NV, USA, 2021, pp. 0245-0250, doi: 10.1109/CCWC51732.2021.9376066.

[5] E. Tankard, C. Flowers, J. Li and D. B. Rawat, "Toward Bias Analysis Using Tweets and Natural Language Processing," 2021 IEEE 18th Annual Consumer Communications & Networking Conference (CCNC), Las Vegas, NV, USA, 2021, pp. 1-3, doi: 10.1109/CCNC49032.2021.9369461.

[6] A. Hembree, A. Beggs, T. Marshall and E. N. Ceesay, "Decoding Linguistic Ambiguity in Times of Emergency based on Twitter Disaster Datasets," 2021 IEEE 11th Annual Computing and Communication Workshop and Conference (CCWC), NV, USA, 2021, pp. 1239-1244, doi: 10.1109/CCWC51732.2021.9375916.

[7] W. Li, S. Gao, H. Zhou, Z. Huang, K. Zhang and W. Li, "The Automatic Text Classification Method Based on BERT and Feature Union," 2019 IEEE 25th International Conference on Parallel and Distributed Systems (ICPADS), 2019, pp. 774-777, doi: 10.1109/ICPADS47876.2019.00114.

[8] S. Mohammadi and M. Chapon, "Investigating the Performance of Fine-tuned Text Classification Models Based-on Bert," 2020 IEEE 22nd International Conference on High Performance Computing and Communications; IEEE 18th International Conference on Smart City; IEEE 6th International Conference on Data Science and Systems (HPCC/SmartCity/DSS), 2020, pp. 1252-1257, doi: 10.1109/HPCC-SmartCity-DSS50907.2020.00162.

[9] B. Pariyani, K. Shah, M. Shah, T. Vyas and S. Degadwala, "Hate Speech Detection in Twitter using Natural Language Processing," 2021

Third International Conference on Intelligent Communication Technologies and Virtual Mobile Networks (ICICV), Tirunelveli, India, 2021, pp. 1146-1152, doi: 10.1109/ICICV50876.2021.938849

[10] Bhavika Bhutani Neha Rastogi Priyanshu Sehgal Archana Purwar; Fake News Detection Using Sentiment Analysis; Department of Computer Science Engineering and Information Technology Jaypee Institute of Information Technology

[11] Mykhailo Granik, Volodymyr Mesyura, Fake News Detection Using Naive Bayes Classifier, 2017 IEEE First Ukraine Conference on Electrical and Computer Engineering (UKRCON)

[12] Fake News: Build a system to identify unreliable news articles: Dataset Fake news https://www.kaggle.com/c/fake-news/data

[13] Davis, Jesse & Goadrich, Mark. (2006). The Relationship Between Precision-Recall and ROC Curves. Proceedings of the 23rd International Conference on Machine Learning, ACM. 06. 10.1145/1143844.1143874. .

[14] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, “Attention Is All You Need”, arXiv:1706.03762v5 [cs.CL] 6 Dec 2017, 31st Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA.

[15] F. Demirkıran, A. Çayır, U. Ünal and H. Dağ, "Website Category Classification Using Fine-tuned BERT Language Model," 2020 5th International Conference on Computer Science and Engineering (UBMK), 2020, pp. 333-336, doi: 10.1109/UBMK50275.2020.9219384.