International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 06 | June - 2025 SJIF Rating: 8.586 ISSN: 2582-3930

Enhancing Fake Review Detection Using BERT Transfer Learning Algorithm in Natural Language Processing

Karthick P[1], Janarthanan P[1], Bora Nikhil Sai[1], Manish Kumar N[1*]
[1]*Department of Computer Science,*
Indian Institute of Industry Interaction Education and Research, Chennai, Tamil Nadu 600066

Abstract - In recent years one of the biggest world concerns is the proliferation of fake news through social networks. Due to their intended purpose as means to sway the beliefs of large crowds, fake news has been causing a great effect on the world. Much attention has been paid to this area by researchers as the process of manual confirmation of the news' authenticity is practically impossible and very costly. Explorations into the detection of false news targeted content based approaches, social context based approaches, image based approaches, sentiment based approaches and hybrid context based classification systems. As a consequence of utilizing the content-based classification approach, this work will propose a model for False news Classification utilizing the headlines of the news. The model used to solve the current problem is a BERT model and the output layer of which is connected to an LSTM layer. The FakeNewsNet dataset was used in both the training and evaluation processes, which consists of two sub-datasets: PolitiFact and GossipCop. Comparison has been made between the model and basic classification model. The suggested model, which utilizes an LSTM layer for the evaluation of impact, is almost analogous to the vanilla BERT model that was trained on the dataset on the same terms and conditions. The findings showed that there was a 2 percent increase in the level of accuracy that has been achieved. For PolitiFact, the recall was equal to 45 % and for training a 1. The performance boost achieved on the GossipCop dataset is about 11% as compared to the vanilla pre-trained BERT model.

Key Words: Fake news, Social networks, Detection approaches, BERT model, FakeNewsNet dataset, Classification accuracy

1.INTRODUCTION

By Fake news, we address the disinformation problem that continues to plague popular social platforms such as Facebook and twitter. Pep talk, as a rule, is a very bad thing: convincing the people to look at the world as you do, is a very, very bad idea. For example, as Paskin noted, specific news articles; news that started in the mainstream media either online or offline but social media also and have no backing with facts but it is presented as facts not satire. There are many examples here how this is effecting our lives in a variety of ways meaning it must be an area of focus. Research work underscore that the election of Donald Trump in the December, 2016 US presidency was informed by fake news. And will be the people of the United Kingdom will ever regret the that made in connection with the Brexit. In these events, social media users were deceived by directions that flooded them with false details based on their political stance to influence their voting. For instance, after realizing the high level of fake news being spread through its social platform, the Indian version of Whatsapp had to embark on an anti-fake-news awareness campaigns. During the COVID-19 crisis there was the reporting of fake news on social media as well as mainstream media sources where people keep sharing the misconceptions. It is also attributed to a large number of people's deaths and civil unrest and problems.

Disrupting the foundations of human behavior, the fake news environment a) Naked Realist: the concept that individual's reality is reality, and any other not min and me is fake. b) This is the reason why there exists a kind of information bias, especially confirmation bias, in as much as there is an inclination towards confirmation bias in prejudices that are already ingrained in an individual or a group. The problem of fake news has entered the world stage in the last few years, largely due to the advancements of technology. Since it has become a popular platform on which people depend on for breaking news, these issues have continued to gain momentum. This concept has been rightly more popular in recent times because the content shows on the social media feeds are biased towards what one feels. When a user follows the similar interactions with other users or the material consumed that supports his/her current beliefs and the cognitive patterns, it is categorized under echo chamber effect [2]. For this reason, most people never attempt to do a follow-up to see that the news is factual.

Taking into consideration, all the aforementioned facts pointing to the fact that false news has become more widespread over the last decade, and it has been used in various fields, it becomes crucial to look for a solution to this problem. The effects of these fake news are notable since the larger population of the social media users interacts with messages and news items with high frequency without necessarily determining the truthfulness of the items. For instance, there are private media sites that only seek to replicate those facts reported and help to determine the veracity of the news, taking into consideration that fake news is today's world is real and frequently. Besides, even with the help of all three types, it is impossible to process the amount of news content which has increased through social media. As a result of the constant influx of large amounts of data every day, this procedure is gradually evolving into a time-consuming and costly process. It means that the media and the information it provides must be verified twice – at the source – given the existence of DeepFake [3]. Thus, computer scientists have been attempting to perform the procedure automatically recently. Because of this, the authors were forced to come up with a model that could help in achieving this and identify false information found in media and news stories.

A myriad of authors in this field have employed both the support vector machines (SVMs) and naive bayes (NB) classifiers. These models were used as baselines because, although the model is structurally and functionally different from the other, both of them yielded similar results. The available literature employs decision trees and other clustering methods to purposeful experiments. Notably, among these models, Recurrent Neural Networks (RNNs) have received a great deal of attention in this field, mainly LSTM. Similarly, unlike LSTMs, RNNs are

prone to the vanishing gradient problem which restricts the model's ability to capture long sequences of data. To make a model that can resolve NLP issues, word embedding should be considered as a priority guideline. This was avoided in the suggested study (Gu et al. , 2020) by using contextual word embedding with the BERT model. BERT takes advantage of a vast volume of unlabeled text corpus for pre-training a contextualized word embeddings. As seen, due to its intricate architecture and its ability to perform nonlinear representation learning, BERT has faired well on the NLP tasks. Sophisticated by the pattern of important data and enable the LSTMs to remember these patterns, the performance is enhanced. Due to their ability to provide semantics and long dependencies inherent in headlines of news articles, contextual pre-trained representations from BERT are employed in LSTM to improve the ability of false news classification.

•The proposed technique categorises the news articles based on syntactical, grammatical, and semantic features.

To predict whether a news piece is real or not, thereby making it easier to identify false news stories, the current study proposes an approach that combines the BERT and LSTM models.

To make the suggested methodology more sound and reliable, four evaluation parameters including Accuracy, Precision, Recall, and F1 Score were used.

•By the use of the adopted testing and training phases, it has been demonstrated that the proposed approach is effective compared to other traditional methods such as the TCNN-URG, LIWC, CSI, HAN, and SAFE among others.

2. METHODOLOGY

In this part, the author explains in detail the framework of the model, which is proposed in the paper. Additional details of the dataset on which the model's training and testing was done are also mentioned. Here, we contributed some background information about BERT and LSTM architectures, respective to the utilized dataset and preprocessing steps. The suggested methodology for classifying false news based on its content is shown in Figure 1: BERT followed by LSTM layer was used for categorizing the news tiles into the mentioned sets. To this end, the classification model that is utilized is BERT, which is enhanced with an LSTM layer. To do so, it obtain word embeddings that are infused with context from a vast amount of unannotated text datasets. The capacity to perform representation learning and the extensive structure of the model proved useful for BERT in the NLP tests. LSTMs hence boosted performance competence by being capable of learning and memorizing vital information patterns.

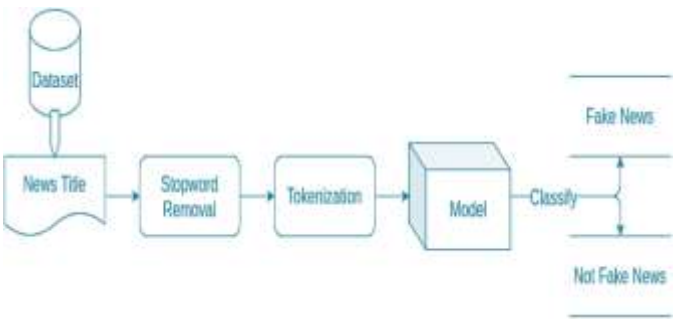

Fig 1: below illustrates the general approach that will be used in training and testing the false news classifier.

2.1 Data Preprocessing

The news items in the dataset have been preprocessed for the same reason, because data preparation is an important stage in preparing the models.

•Evgalized all the words in the phrase converting all of them to lower case •Replaced the word "'t "with "not". I cannot change it to "can't" since, for example,

Removed the "@name"I removed all the characters excluding '?' by using the regular expression. I also was going through and deleting any extra special characters that were not necessary. I stripped off those words considered as stop words but I kept "not" and "can" Finally I dropped the trailing whitespace.

•Tying out non requisite words from cleaned text

After removing the unnecessary words and characters from the text contents of the given datasets, the training and classification phases involved the extraction of tokenization vectors, attention masks, and the binary category using Bert Tokenizer.

2.2 BERT

BERT is a system that learns word-to-word contextual relations through an attention mechanism known as the transformer. The transformer is an encoder that has text input. It also contains a decoder that predicts based on the nature of the job to be done. The transformer encoder is non-directional because unlike the directional devices it scans the whole text at once as oppose to scanning it word by word. What this implies is that the model is able to decode the meaning of the term from the context. The manner in which sentence-level categorization has been done is known as bidirectional BERT architecture. Fig. 2 illustrates the design of the BERT sentence-level categorization model. Different applications are catered to by the multiple pre-trained models of BERT. There are two models that are commonly used: BERT-base and multi-head attention heads. BERT-base uses 12 encoder stack levels, contains 767 hidden units, and 109 million parameters.

BERT-large has 333 million parameters: 24 layers of the encoder stack, 1016 hidden units, and 15 multi-head attention.

Before using the pre-trained model, the input data has to be properly preprocessed. We have been able to obtain suitable embeddings for all the sentences. These models include encoder layers that take in a list of token embeddings and the attention masks corresponding to the list. The output is the same number of embeddings of the same hidden size.

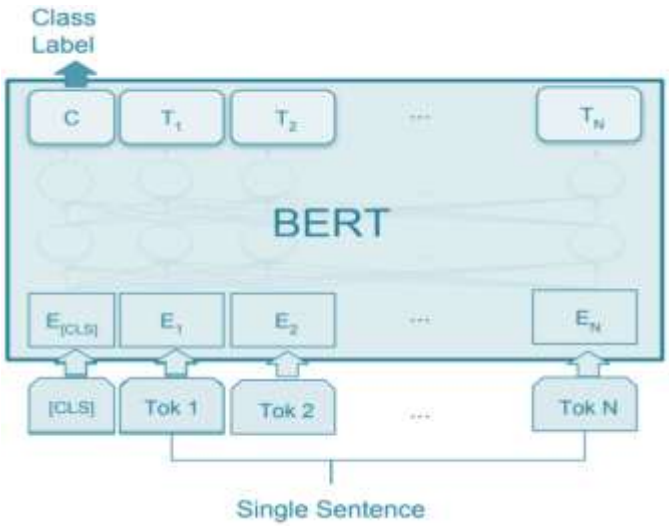

Fig 2: presents the BERT structure for classifying sentences according to [4].

2.3 LSTM

To that end, to train the feature classifier, we provided to it only one vector which was equal to the entire input phrase. Next, we used the first token isolated from the output of the model, called [CLS] as the representation of the whole sentence in classification.

There is a specific type of RNN that can learn long-term dependency; it is called long short-term memory. While LSTMs has a chain unlike structure with RNNs, their fundamental module different architecturally from other RNNs. It is observed that the RNNs are more appropriate for learning short sequences of data [5]. A problem with RNNs is that they are not very effective for learning and understanding context and Chain-Generalization due to the vanishing gradient problem. There are another type of RNNs called Long short-term memories (LSTMs) that are improved on for learning long-term dependencies and hence not affected by this problem. Thus, while it is possible to find RNN with fairly simple architecture, such as a single tanh layer, they are always given by stacked modules. On the other hand, long short-term memories are made up of cells which are loops and each cell of LSTM has four neural networks wired in a particular manner as shown in fig 3. Only the current cell state and the cell's hidden state (c t and h t) pass from one cell to another. Memory cells have three ports for managing information flow: the forgetting port, the input port, and the output port. Forget Gate removes data which is no longer useful in predicting LSTM, which has a sigmoid layer. The output gate applies a sigmoid layer to shows the pertinent data from the current cell. Here Fig. 3 is presented to illustrate the LSTM architecture.

Fig 3: LSTM design

3. PROPOSED ARCHITECTURE

The technique being proposed uses a feed-forward network with 768 hidden size nodes and BERT-base-uncased. This BERT model takes two inputs — the input sentence and the attention mask. This has been done using the BERT tokenizer where the input sequences combined with [CLS] in front of the sentences and end of the sequence [CLS] to generate input identifiers and attention mask as outputs. These inputs are then passed to the BERT model and what comes out is an embedding vector for each token for the size of the vector being 768. AID helps LSTM by understanding the semantics of the given sentences since BERT provides contextualized representations at the sentence level. The studies carried out in the literature reveal that, utilization of LSTM in combination with word embedding models leads enhanced results [6, 7]. Thus, the error can be reduced even further by combining LSTM and BERT, which indicates that the suggested model better comprehends semantic meaning. Since BERT is birectional, the encoding process, need to pass [CLS] through multiple layers of encoding. This procedure contains all general features of all tokens and [CLS] acts as a "summary representation" for classification tasks. Therefore, to finish the classification job, the classifier could be provided entire phrase which is encoded in the sets of embeddings corresponding to the [CLS] token. In the most basic sense, the classifier has been built from scratch. The classifier contains a 128-node feed forward linear layer with 128 inputs. Due to the inconsistency in inputs, batch normalization layer was implemented for normalization of inputs. Next, a 0. To combat overfitting, a dropout of 6-rate layer was used. A dense layer of two feed-forward layers with an output size of three helps decide whether the incoming news is true or false. To enable the separation of the cells' output in a way that one is favored over the other, a cutoff of 0 is applied to the second cell. 8 is used. The structure of this approach is depicted in figure four below.

4. RESULTS

For the purpose of testing and training of the model, the FakeNewsNet dataset is used. Additional to the primary dataset, there are the Politi-Fact Sub-corpus and the GossipCop Sub-corpus. Created from news articles of Politifact and GossipCop websites that analyze claims, the resource was developed. You can ascertain if the political news stories are true or fake on Politifact. From entertainment news, GossipCop ensures they determine whether or not the news is true, and then rate the story based on its credibility. Often, there are series of word occurrences particular to disinformation in many publications available. It has been decided that the news titles should be divided into categories according to the nature of the news. Titles of news stories were collected from both the sets of data as well. The PolitiFact dataset contains 427 labeled data and 617 real data altogether. Out of all the articles in the GossipCop dataset, there are fake data points equal to 5217 while the real data points are equal to 15617. The details of training the models on the given dataset and the comparison of the proposed model with other classification models are presented in this section. Our choice of the training and testing data sets as contained 79% of the data points and used 17% for testing. For the performance analysis of the models, four measures such as recall or precision, accuracy, and F1 score were employed. We have combined them, constructed comparative figures and tables, have estimated and illustrated the proposed model's confusion indicators.

The differences between the baseline models of Fake News Detection have been discussed and the performance comparison made between the current technique proposed.

URG-TCNN [8]: The model employs covariant feature detector through an annealed conditional variational autoencoder on user comments and two stage convolutional neural network on news articles.

The psycholinguistic categories that can be identified to make up the lexicons may be derived via tools from which LIWC, which is an abbreviation for Linguistic Inquiry and Word Count. Consequently, from a psychological and deceitful perspective, it possesses a feature vector [9].

Based on the principle of deep learning, CSI integrates source information, text information, and response information [10]. As for the news, an LSTM neural network is used to encode the news, where the Doc2Vec for news content and comments from users are utilized as the inputs. Figure 5 shows Relative comparison of the accuracy of Politifact statements in the form of Line Graph

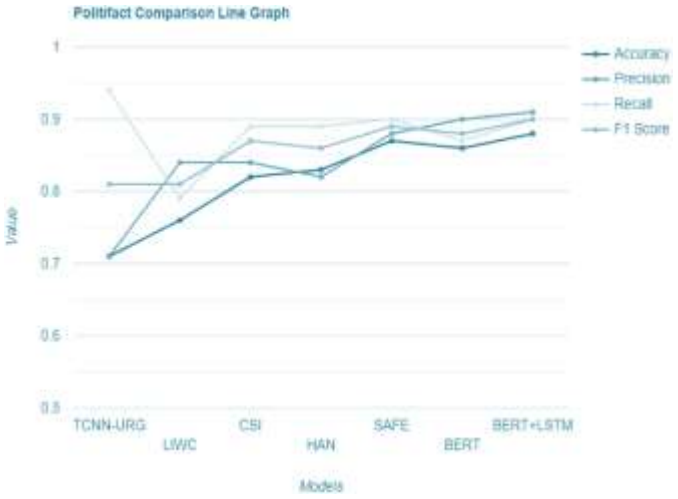

Fig 5: Relative comparison of the accuracy of Politifact statements in the form of Line Graph

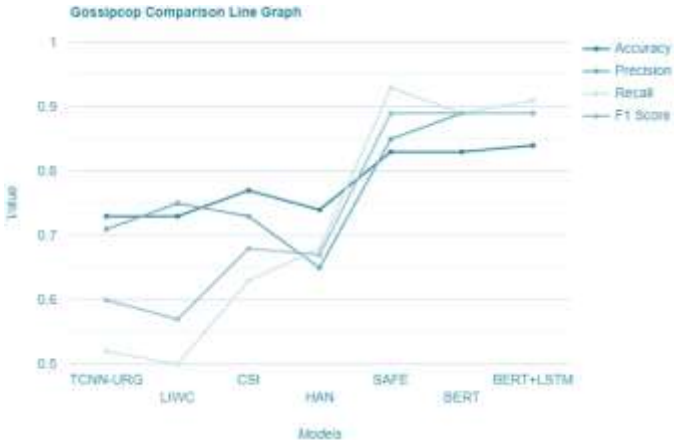

Fig 6: The chart showing the comparison between GossipCop and other different platforms

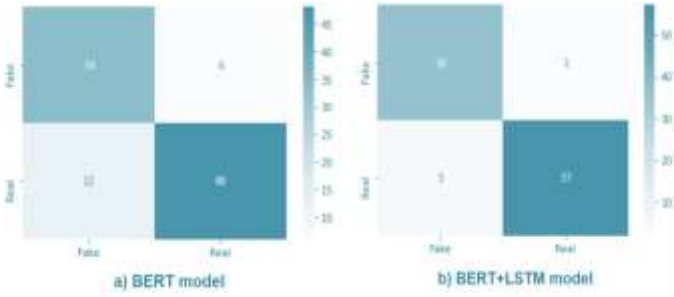

Fig 7: In this paper, the confusion matrix compiled by Politifact will be reviewed as this tool helps to distinguish between the true positives, false positives, false negatives or the false negatives, all of which are crucial in creating an accurate index.

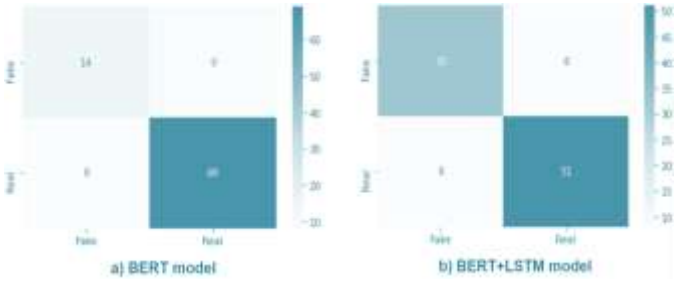

Fig 8: That is why in order to compare GossipCop it is useful to construct the Confusion Matrix.

We thus compared the outcome of the proposed model to that of the baseline models including TCNN-URG, LIWC, CSI, HAN, SAFE(Multimodal), and the most recent and highly efficient one, BERT after its' application to the FakeNewsNet dataset using (BERT + LSTM). Compared to other models, the proposed one was 88% accurate at its utmost peak. 75%. In the PolitiFact dataset, the indicated model offered additional accuracy to the baseline models to be at least 1. Up to 35% and by a full 15 percent depending on the circumstances. 67 percent. The performance increases compared to the baseline models in the GossipCop dataset are as follows: The accuracy increase is from 0. 2 percent to 9. 5 percent. Over PolitiFact and GossipCop, the above assessed measures are visually illustrated in Figure 6 and Figure 7 a-b. The suggested model for the PolitiFact and GossipCop datasets' confusion matrix is provided in Figure 8 a-b, respectively.

5. CONCLUSION

The opponents propagate wrong information to the public through Fake News with an aim of affecting their decision through social media, television channels, etc. The attempts of classifying false news as well as, importance of the topic in today's world are illustrated in the work. Machine learning is used to develop the classification model with the titles of news stories to distinguish between legitimate and fake news. This was followed by the classification of the news headlines using BERT with an LSTM layer as the model of choice to sort the news. To accomplish this, there is the use of the BERT classification model that is fitted with an LSTM layer. BERT might use a large variety of unlabeled text datasets to learn about the contextualized word representations. Testing, BERT was prominent in natural language processing tests due to its complex structure and its capability of representations learning in a non-linear index. They also increase performance by the ability of slowly learning and

memorizing critical data patterns. Now that BERT has been proved to use its contextualized word representations to capture the semantics and long-distance relevance in news headlines surely it can be incorporated into LSTM to help in classifying fake news. Other categorization approaches such as a Plain Old BERT Model has been done for comparison with the above. Although the suggested model showcased a slight enhancement, this indicates that the model has been able to comprehend the patterns of the language of the headlines and how it is connected to false news. There is one drawback that the model faced: it is difficult to distinguish a headline that belongs to fake news from that of actual news. When owners of fake news outlets begin to use language that somewhat resembles real news, it gets complicated to distinguish them. One way of addressing this problem is by personally going through the facts of the news headline in question. More data augments deep learning models making them efficient and faster in their work. In this regard, increasing the set of data would enable us to conduct a deeper analysis of the language of headlines of false news. At the same time, misinformation is likely to circulate within a short time on social media platforms such as the Twitter and the Face book. The trend in the lexical and linguistic patterns of false news found in social media platforms may help in future work in establishing this model. In the future, this design can be compared with other applications in every domain in which it is going to be used and it is a great addition to the benchmarks which are already present. The primary idea of the further model investigation and tests with respect to different settings is to achieve the performance improvement in comparison with the state approaches. However, we also need to explore how communities might change if we alter the hyperparameters of BERT and subsequent layers.

REFERENCES

1. Paskin, Danny. "Real or fake news: who knows?." The Journal of Social Media in Society 7, no. 2 (2018): 252-273.
2. Törnberg, Petter. "Echo chambers and viral misinformation: Modeling fake news as complex contagion." PLoS one 13, no. 9 (2018): e0203958.
3. Singh, Amritpal, Amanpreet Singh Saimbhi, Navjot Singh, and Mamta Mittal. "DeepFake video detection: a time-distributed approach." SN Computer Science 1, no. 4 (2020): 212.
4. Moravec, Patricia, Randall Minas, and Alan R. Dennis. "Fake news on social media: People believe what they want to believe when it makes no sense at all." Kelley School of Business research paper 18-87 (2018).
5. Luengo, María, and David García-Marín. "The performance of truth: politicians, fact-checking journalism, and the struggle to tackle COVID-19 misinformation." American Journal of Cultural Sociology 8, no. 3 (2020): 405-427.
6. Deepak, S., and Bhadrachalam Chitturi. "Deep neural approach to Fake-News identification." Procedia Computer Science 167 (2020): 2236-2243.
7. Anand, Ishu, Himani Negi, Deepika Kumar, Mamta Mittal, T. H. Kim, and Sudipta Roy. "Residual u-network for breast tumor segmentation from magnetic resonance images." Comput. Mater. Contin 67 (2021): 3107-3127.
8. Qian, Feng, Chengyue Gong, Karishma Sharma, and Yan Liu. "Neural User Response Generator: Fake News Detection with Collective User Intelligence." In IJCAI, vol. 18, pp. 3834-3840. 2018.
9. Pennebaker, James W., Ryan L. Boyd, Kayla Jordan, and Kate Blackburn. "The development and psychometric properties of LIWC2015." (2015).
10. Ruchansky, Natali, Sungyong Seo, and Yan Liu. "Csi: A hybrid deep model for fake news detection." In Proceedings of the 2017 ACM on Conference on Information and Knowledge Management, pp. 797-806. 2017.