Kathmandu University Journal of Science, Engineering and Technology, Vol. 16, No. 2, December 2022

Kathmandu University
Journal of Science, Engineering and Technology

Detection of fake news using deep neural networks

Neha Shakya and Prakash Poudyal*

Department of Computer Science and Engineering, Kathmandu University, Dhulikhel, Nepal.

Abstract

With the increased use of the Internet and social networking sites, information can now disseminate at a rapid rate. It is an age of information where any information can be accessed with a single click. This has increased the risk of the spread of misleading false information. These fake news have negatively impacted people and society. So, a strong mechanism is needed to detect false news and stop its propagation. The content of the news, the source of the news, and the response to the news are the main features that can help to detect the credibility and authenticity of the news. This paper aims to implement deep neural networks to accurately detect fake news. It aims to evaluate the various deep learning models to determine the model that can accurately and efficiently distinguish fake news. The experiment uses the Source Based Fake News (SBFN) dataset, a publicly available dataset for Fake News Detection. Various deep learning models such as Long Short Term Memory (LSTM) and Bidirectional Encoder Representations from Transformers (BERT) along with a hybrid model have been trained and evaluated on this SBFN dataset.

Keywords: Fake news detection; Deep Learning; LSTM; BERT

1. Introduction

Today, the Internet has converted the world into an interconnected global village. So, news can now propagate rapidly, especially with the use of Social Networking Sites (SNS) like Facebook, Twitter, and Instagram. It has become increasingly difficult to detect the authenticity of the shared news on these social platforms. Many kinds of fabricated information are also being shared along with the real news to deceive people. These misinformation and disinformation are deceiving people and creating a chaotic environment, i.e., an infodemic or information epidemic.

Fake news can be defined as false or misleading information presented as news [1]. It is usually created to be widely distributed to discredit public figures, political movements, or companies. However, the term has an ambiguous definition and also includes stories that are unintentional and unconscious as well. The spread of fake news started with the start of journalism. Fake news has since then been used to defame individuals and trick the public about any propaganda. The increased use of the Internet as a source of news has helped in the spread of any news i.e. both real and fake ones. News travels from one part of the world to another in an instant and people are not given enough time to process the authenticity of the news before it spreads rapidly.

Primarily, political sectors are the main targets for fake news. but it is not limited to this. The US presidential election in 2016 saw a surge in manipulation through fake news using social media [2]. Lately, with the outbreak of the COVID-19 pandemic, lots of bogus news and myths regarding the disease have gone viral on the Internet [3]. This has affected the mental well-being of the people during this difficult time. According to a recent report published by Media Action (`https://mediaactionnepal.org/`), 3.70% of 49,051 news collected from major newspapers and online news portals in Nepal during the first 3 months of the COVID-19 lockdown were fake or misleading. The study included 23,291 news published in 10 daily newspapers and 25,760 news published in 10 online news portals. They studied the 8 different indicators of misleading information and reported 95.71% of misleading information had misleading news sources. 95.57% of the misleading information had no source mentioned, 2.42% of them had fictitious sources and the remainder had anonymous sources.

People are often deceived by the fake news circulating on the Internet mainly due to three reasons [4]. First, the information confirming their preexisting attitudes is preferred (selective exposure). Second, the information consistent with their preexisting beliefs is more persuasive (confirmation bias). Third, people are more inclined to accept the information that pleases them (desirability bias). Unless the information violates the preconception of the individual, the credibility of the information is not questioned.

The spread of fake news can have a serious impact on people and society [5]. It breaks the authenticity balance of the news ecosystem and changes the way individuals view the real news as it makes them accept biased or false beliefs. This emphasizes the importance of some mechanism to detect fake news in its early stages to minimize the damage it can cause to human psychology and well-being.

Machine learning has a wide range of applications. This research aims to use machine learning techniques specifically deep learning to differentiate fake news from real ones. Various Natural Language Processing (NLP) approaches can be used to analyze the content and style of the news to detect the context and facts in the article. But only the content of the news article is not enough to detect the credibility of the news. The characteristics such as the source of the news and how the readers perceive that piece of information through social media must also be considered [6].

2. Related Works

There are many forms of fact-checking websites that evaluate factual claims of the reported news such as PolitiFact (`https:`

*Corresponding author. Email: prakash@ku.edu.np

//www.politifact.com/) and Snopes (https://www.snopes.com/). But, the effectiveness of fact-checking is mixed because people tend not to question the credibility of information unless it violates their preconception and when any misinformation is repeated, they are perceived as truth [4]. Additionally, with the use of the internet and social media, Google, Facebook and Twitter have become our mediators of news media as well as our friends and families. Thus, a proper mechanism to detect fake news has become crucial and artificial intelligence can be used to address this problem. In this section, different approaches used for fake news detection have been discussed.

2.1. Stance Detection

To explore how artificial intelligence technologies could be used to combat fake news, Fake News Challenge Stage 1 (FNC-I) was introduced [7]. FNC-I focuses on Stance Detection which was believed to be a useful building block in the AI-assisted fact-checking pipeline. In the context of FNC, Stance Detection is estimating the relationship between an article's body with its headline/claim. The body of an article may agree, disagree, discuss, or is completely unrelated to its headline. Thus, the output of the stance detection system was 'agree', 'disagree', 'discuss', and 'unrelated.'

The model that performed with the highest score of 82.02 in the FNC-I used an ensemble model based on a 50-50 weighted average between gradient-boosted decision trees and a deep convolution neural network to perform stance detection on several news headlines and article text [6]. Another model that obtained the highest accuracy of 90.05% and 78.04% for validation and test data respectively, used a stacked ensemble of five independent weaker slave classifiers that fed a strong master classifier [8]. Here, the master classifier is a Gradient boosted decision tree classifier which uses the prediction from the five weak classifiers as features along with the original data.

An automatic stance evaluation that facilitated the fact-checking process used a common bag-of-words (BOW) i.e. Term Frequency (TF) and Inverse Document Frequency (IDF) for text inputs [9]. They extracted TF vector of the headline and body along with its cosine similarity. These lexical and similarity features passed through a Multi-Layer Perceptron (MLP) with one hidden layer of 100 units and a softmax layer on the output of the final linear layer. This model achieved a score of 81.72% and was placed third in FNC-I.

A Stance Detection (StD) benchmark combined the benefits of Transfer Learning (TL) and Multi-Dataset Learning (MDL), which have emphasized the need to focus on robustness and de-biasing strategies in multi-task learning approaches [10]. An automatic hoax detection system that classifies hoaxes and non-hoaxes based on the likes on Facebook was based on logistic regression and harmonic boolean crowdsourcing algorithms [11].

2.2. Machine Learning Approaches

Various machine learning algorithms like Support Vector Machine (SVM), Stochastic Gradient Descent, Gradient Boosting, Bounded Decision Tree and Random Forest, for fake news detection were evaluated by [12]. Bigram Term Frequency-Inverse Document Frequency (TF-IDF) and Probabilistic Context Free Grammar (PCFG) were applied to the data before training the various machine learning algorithms. The stochastic Gradient Descent model fed with bigram TF-IDF best identified the non-reliable sources with an accuracy of 77.2% when evaluated on a probabilistic threshold of 70%. It was observed that TF-IDF showed good predictive power even when named entities were ignored whereas PCFGs did not add much to the predictive value.

An innovative hybrid approach that combines linguistic cues with machine learning approaches and network analysis approaches was proposed by Conroy et al. [13]. Linguistic processing includes multiple layers of word/lexical and semantic analysis. Network behavior is combined with this linguistic approach to add a trust dimension by identifying the credibility of the source.

Bharadwaj et al. [14] presents detailed and comprehensive results of experiments on various Machine Learning models for fake news detection. Source-based fake news dataset has been used for the experiment. Various machine learning models like SVM, Naive /Bayes, Random Forest, Logistic Regression, AdaBoost, Decision Tree, and Neural Networks have been used along with TF/IF, GloVe, and Word2Vec embeddings. The results show that AdaBoost with TF/IDF had the highest accuracy of 96.91%.

2.3. Deep Learning Approaches

A benchmark dataset for Fake News Detection was proposed by Wang [15], which contains short statements from politifact.com called the LIAR dataset. This dataset was tested against baseline models like logistic regression, Support Vector Machine (SVM), LSTM, and CNN along with a proposed hybrid CNN model. An accuracy of 27.4% was obtained using the proposed hybrid CNN model incorporating all the metadata.

A multi-source multi-class Fake news detection (MMFD) framework on LIAR dataset consists of three parts: 1) automated feature extraction, which extracts features from textual sources based on CNN and LSTM, 2) interpretable multi-source fusion, which combines the features from different sources, and 3) fakeness discrimination multi-class discriminative function [16]. The accuracy obtained by this model while using all the statements. metadata, history, and report was 38.81%.

An attention-based LSTM model was used for fake news detection using LIAR dataset [17]. This paper obtained an accuracy of 41.5% by considering the speaker's profile which provides an additional input and attention factor for learning of news text. Long [18] proposed a deep ensemble framework using Bi-LSTM to capture the sequential information and CNN to capture the hidden features efficiently and obtained an accuracy of 44.87%.

The CSI model incorporated the three characteristics of news i..e. text of the article, the response received by the article, and source promoting it [19]. The CSI model comprises three parts, namely Capture, Score, and Integrate. Capture uses Recurrent Neural Network (RNN) to capture the temporal engagement of users with an article in terms of frequency and distribution. Score utilizes the source characteristic present in the behavior of users. Integrate takes the output from the previous models and combines them to produce labeled predictions i.e. real or fake. Two publicly available real-world datasets collected from Twitter and Weibo [20] were fed to the proposed model and results indicated that the model outperformed other state-of-the-art models with an accuracy of 89.2% and 95.3% respectively.

A hybrid of Convolution Neural Network (CNN) and Long Short Term Memory (LSTM) Recurrent Neural Network (RNN) for fake news identification on Twitter posts used three deep neural network variants i.e. LSTM, LSTM with dropout regularization and LSTM with CNN on Twitter [21] dataset [22]. The LSTM model outperformed the remaining two models in terms of precision, recall, and F1-score with an accuracy of 82.29%.

Deepak and Chitturi [23] proposed a deep neural network i.e Feed Forward (FF) Neural Network and Long Short Term Memory (LSTM) to identify fake news with a live mining stage to fetch auxiliary features like source domains, author names, etc. from the web using python and Beautiful Soup HTML Parser. The live mining stage mimics the fact-checking process and adds to the effectiveness of the model. The results showed that there was a significant improvement in performance from 82-84% to 91-94% after the addition of mined features.

Kula et al. [24] used a hybrid architecture mostly based on Bidirectional Encoder Representation from Transformers (BERT) for word embeddings and RNN network for document embedding. ISOT (Information Security and Object Technology) dataset [25] was used to train the proposed hybrid model and obtained results with accuracy above 90% which guarantees the reliability of the proposed architecture. BERT and its modifications have a deep impact on Natural Language Processing (NLP).

The progress in NLP has also aided the generation of neural fake news which closely mimics real news. Text generators such as Grover and GPT-2 can be used to generate fake news that can fool readers and spread disinformation. So defenses against this kind of news are also necessary. According to Zeller et al. [26], Grover performs better at detecting Grover's fake news than other NLP approaches like BERT and so it is critical to release these models for effective detection of neural fake news.

The review of this literature has underlined the importance of both linguistic approaches as well as network analysis approaches for fake news detection. A model that incorporates the content of the article, the source of the article, and the response received by the article on social platforms is essential to detect the authenticity and credibility of the news.

There has been a lot of progress in the field of machine learning, deep learning approaches and natural language processing approaches. So, state-of-the-art deep learning models such as LSTM, transformers, attention mechanisms, and BERT can be used to detect fake news in addition to machine learning approaches like logistic regression, decision tree, and SVM. Also, these deep learning models could be combined to build a hybrid model [15, 22, 24], which could be a better detector for fake news.

3. Fake News Detection Models

RNN, LSTM, and BERT are some of the state-of-the-art deep learning algorithms for NLP tasks. In this section, the deep learning models used for the experiment are discussed along with the proposed hybrid model.

3.1. LSTM

The first model is a simple LSTM model (Fig. 1). In this model, the processed input is first passed to a hidden Embedding layer which learns the word embeddings in the training dataset. Here the embedding layer has a vocabulary size of 10000 words, The output from the embedding layer is then passed on to an LSTM layer with 100 neurons. Finally, the result is passed on to a dense layer with a sigmoid activation function to give the final classification result. A dropout of 0.3 is used after the Embedding and LSTM layer to reduce overfitting. The model is optimized using an Adam optimizer and its efficiency is measured by binary cross entropy.

3.2. LSTM with Attention Mechanism

The attention mechanism is widely used to decide which part of the input should be given more importance and which should be ignored so that the model can concentrate only on the most relevant things (Fig. 2). Similar to the previous model, the vectorized input is first passed to an Embedding layer before passing on to the LSTM layer. An attention layer is added which takes the output of the LSTM layer and gives a context vector. It consists of dense layers with a tanh activation function to calculate the weighted sum. Finally, the output is passed to a dense layer with a sigmoid activation function to give the final result.

3.3. BERT

BERT [27] is another widely used NLP transfer learning model (Fig. 3). Hugging Face provides a ton of pre-trained models, among which 'bert-base-cased' has been used for the BERT models. It consists of 12-layer, 768-hidden, 12-heads, 109M parameters and is trained on cased English text. BERT uses a special BERT tokenizer to tokenize the input words. The tokenized input is used in the BERT model. The output of the BERT model is in its raw state, so it is then passed to a linear layer with a softmax activation function to give the final output.

3.4. BERT with LSTM and Attention

The proposed model uses both BERT as well as LSTM models with an attention layer as shown in Fig. 4. The BERT model uses a special tokenizer called the WordPiece tokenizer to tokenize the input sentences and represent it to vector space. The tokenized input is passed on to the BERT model. Then, the raw output is passed on to a LSTM layer. The output from the LSTM layer is then passed to an attention layer. The attention layer is the same as in the second model (section 3.2) and consists of dense layers to get the attention weights and context vector. Finally, the output from the attention layer is passed to a linear layer with a softmax activation function that gives the final output.

4. Experiment

Google Colab's (`https://colab.research.google.com/`) GPU has been used to train and test the model written in Python. Google colab is a free online cloud-based Jupyter Notebook environment that allows training machine learning and deep learning models on CPUs, GPUs and TPUs.

4.1. Dataset

Since machine learning algorithms are fuelled by data that is fed to them, data collection is an important step for any machine learning problem. There are various publicly available datasets for Fake News Detection. Among them, the SBFN dataset is the most popular and widely used dataset.

Source Based Fake News (SBFN) dataset [28] is a preprocessed dataset from Getting Real about Fake News (KaggleFN) Dataset, which consists of text and metadata scrapped from 244 websites that were tagged as 'bullshit' by the BS Detector Chrome Extension. The skewness in the KaggleFN dataset has been removed in the SBFN dataset. The SBFN dataset consists of 2096 data with various attributes. These attributes include the various aspects of the news. First is the content of the news which is given by the attributes title and text. Second is the source of the news which is given by the attributes author and site_url. The content of the news is the actual news on any given topic and the source gives information about who has written the article and which website has published the article.

The dataset contains news articles that were collected from 68 unique websites, among which 50 sites published only real news, 8 sites published only fake news, and 10 sites published both fake and real news. This is an important aspect to be considered while detecting if the news article is real or fake. So, this factor was examined while performing the experiment. The dataset consists of 801 real and 1294 fake news, which is classified into various types such as bs (bullshit), bias, conspiracy, hate, satire, state, junksci (junk science) and fake. The bs, conspiracy, satire, junksci and fake news articles are labelled as fake news and bias, hate and state news articles are labelled as real news.

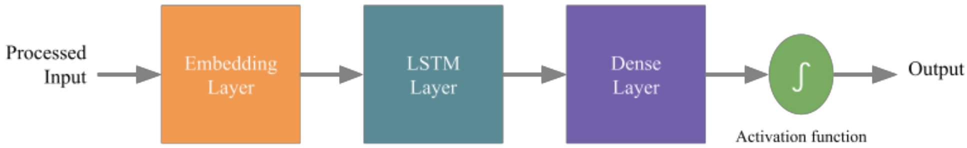

Figure 1: LSTM Model

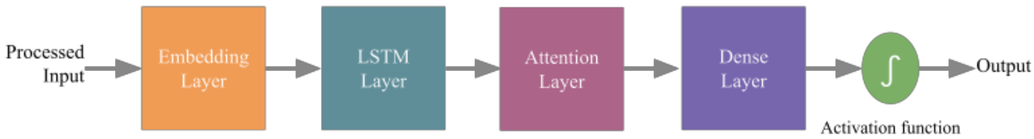

Figure 2: LSTM with attention mechanism.

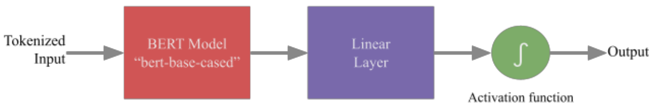

Figure 3: BERT model.

4.2. Data Preprocessing

The data is analyzed and unnecessary features are removed and the data is also rescaled, binarized and standardized to fit the model. Various text pre-processing techniques are used to detect patterns in the raw text and also to eliminate unnecessary elements.

After analyzing the data, it was found that a few of the records were missing in the SBFN dataset. Out of 2096 entries, 46 of them were missing the news content, so all of these entries have been deleted to prevent the model from getting biased. Furthermore, news content with less than 50 words have also been removed for the same reason. The remaining 2010 entries are then used to train and evaluate the various models.

For any Natural Language Processing task, before feeding the data to the model, the data must be tokenized and converted to vectors to suit the training model. The text must be the first split into tokens and mapped to their respective indices in the tokenizer vocabulary. For the experiment, the sentences were first split into word tokens. These word tokens were stemmed to their respective base words. Then, these tokens were encoded using one hot encoding which was then padded to obtain uniform vector representation for words to be fed into the embedding layers of the LSTM model.

Similarly, BertTokenizer was used for the BERT model which adds further additional special tokens for each sentence. It adds special tokens to the sentences such as CLS, SEP and PAD. [CLS] token is prepended to the beginning and [SEP] token is added at the end of each input sentence. The BERT model also requires a fixed length of sentences as input, so paddings are added to shorter sentences, which is represented by the [PAD] token. BERT tokenizer uses a WordPiece algorithm, which breaks down the unseen tokens into words with several subwords that can be represented by the model.

The dataset is then split into training, validation, and test dataset. The models are trained using training and validation dataset and the final model is evaluated using the test dataset. Firstly, the dataset is split into training and testing dataset in the ratio of 4:1. Then the obtained training dataset is again split to training and validation datasets in the ratio 4:1. The number of data in the training, validation and testing dataset is shown in Table 1. The 2010 total data were divided into 1286, 322 and 402 training, validation and test dataset. The total 2010 data consists of 745 real and 1265 fake news. There were 481 real news and 805 fake news in training data. The validation dataset had 111 real and 211 fake news. Similarly, the test data had 153 real and 249 fake news. Various deep learning models are fit using the training dataset and then its performance is evaluated using the validation dataset. Finally, the performance of the final trained model is measured using the testing dataset.

Table 1: Count of training, validation and test in SBFN dataset

Data	Count of Real News	Count of Fake News	Total Count
Training data	481	805	1286
Validation data	111	211	322
Testing data	153	249	402
Total data	745	1265	2010

4.3. Models and its Parameters

Deep learning models namely LSTM, LSTM with Attention, and BERT have been trained for the experiment. Along with that, a hybrid model i.e. BERT with LSTM and attention has been trained.

Keras library has been used to build the LSTM model. The Keras Embedding Layer is the first hidden layer of this model. The output from the embedding layer is fed to a LSTM layer. Finally, a dense layer with a softmax activation function gives the desired output.

For the second model, an attention mechanism was added in the LSTM model. The attention layer was built using dense layers with a tanh activation function to get the weighted sum or attention weights. Then, the context vector was calculated using the attention weights. This attention layer was added before the final dense layer for classification.

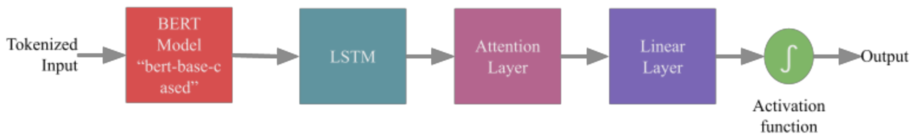

Figure 4: BERT with LSTM and Attention

Hugging Face's Transformers package has been used for the BERT model. The BertModel has been used, which is a bare Bert Model transformer which outputs raw hidden-states without any specific heads on top. Hugging Face provides a ton of pretrained models, among which 'bert-base-cased' has been used for the BERT models. It consists of 12-layer, 768-hidden, 12-heads, 109M parameters and is trained on cased English text. The raw output from BertModel is passed to a dropout layer with a probability of 0.3. Finally, a linear layer with softmax activation function gives the required output.

Finally, the proposed model is a combination of the BERT model and LSTM with an attention mechanism. The output of the BERT model is connected to a LSTM layer which is followed by an attention layer. Finally a linear layer is used for classification.

The various parameters used for different models are summarized in Table 2. Batch size defines the number of samples to work through before updating the internal model parameters and epoch defines the number of times that the learning algorithm will work though the entire training dataset. LSTM was trained with a batch size of 64 for 5 epochs. BERT models were trained with a batch size of 8 with 10 epochs.

Adam is used to optimize both the model due to its popularity and high computation capability. LSTM was optimized using Adam optimizer with learning rate of 0.001 and BERT models were optimized using Adam optimizer with learning rate of 2x10-5. The efficiency of the model is evaluated using the cross entropy loss function.

Table 2: Different Model Parameters used for the experiment

Parameter	LSTM	BERT
Batch	64	8
Epochs	5	10
Optimizer	Adam	Adam
Loss Function	Cross Entropy	Cross Entropy

5. Results

Evaluation metrics such as accuracy, precision, recall and F1-score have been used for the model evaluation during the training and testing phase. Fake News Detectors must minimize the False Negatives (FN) i.e. the number of false news that are predicted as true. So, F1-score is a better evaluation metric than accuracy for Fake News Detection.

The experiments were performed with different models as discussed in section 3 and the dataset as discussed in section 4. using different aspects of the news. The different aspects of the news i.e. the news content and source were used for the experiments. The different deep learning models like LSTM and BERT along with the proposed hybrid models were used for the experiment. The results are shown in Tables 3 and 4.

The experimental results obtained while using only the content information of the news on different models is shown in Table 3. The BERT model outperformed with the highest accuracy, precision, recall and F1-score of 0.895. It is then followed by the hybrid model i.e. BERT with LSTM and Attention, which has a precision, recall and F1-score of 0.820, 0.821 and 0.818. This shows that the BERT model works better than the combination of BERT with LSTM model. The LSTM model performed poorly with the least F1-score and recall of 0.437 and 0.294 respectively even though it could predict with 84.9% precision. The F1-score and recall improved to 0.636 and 0.641 respectively after the addition of attention layer to the LSTM model.

Table 3: Experimental result obtained by different models using only the news content

Model	Accuracy	Precision	Recall	F1-score
LSTM	0.711	0.849	0.294	0.437
LSTM with Attention	0.721	0.632	0.641	0.636
BERT	0.895	0.895	0.895	0.895
BERT with LSTM and Attention	0.819	0.820	0.821	0.818

Fig. 5 shows the confusion matrix for various models using only the news content. The false negative for the LSTM model is highest even though the false positive is the lowest due to which the precision is high even though the recall is low. It infers that the model predicted most news as fake and could not classify the real news properly. The false negatives and false positives have minimized greatly using the BERT model, which implies that the BERT model is a better classifier and could almost correctly predict the real and fake news.

The experimental results obtained while using both the content information of the news on different models is presented in Table 4. The hybrid model i.e. BERT with LSTM and attention outperformed with the highest accuracy, precision, recall and F1-score of 0.993. It is then followed by the BERT model, which has an accuracy, precision, recall and F1-score of 0.985 that is slightly less than the scores of the hybrid model. This shows that the hybrid model works better while using the source information as compared to the results obtained while using only the content information of the news. The LSTM model performed poorly with the least F1-score and recall of 0.582 and 0.536 respectively. The F1-score and recall improved to 0.608 and 0.569 respectively after the addition of attention layer to the LSTM model. However, the results have surpassed the results obtained while using only content information.

Fig. 6 shows the confusion matrix for various models using both the news content as well as the source of the news. Similar to the previous result, the false negatives and false positives have decreased for BERT models compared to the LSTM models. The false positive for the BERT model with LSTM and attention is zero which suggests that the model could perfectly classify all the fake news. The model could not classify only a few of the real news, which in-

LSTM

		Predicted Label	
		Fake	Real
Actual Label	Fake	241	8
	Real	108	45

LSTM with Attention

		Predicted Label	
		Fake	Real
Actual Label	Fake	192	57
	Real	55	98

BERT

		Predicted Label	
		Fake	Real
Actual Label	Fake	240	23
	Real	21	135

BERT with LSTM and Attention

		Predicted Label	
		Fake	Real
Actual Label	Fake	234	25
	Real	50	110

Figure 5: Confusion matrix of result obtained by different models using only the news content.

Table 4: Experimental result obtained by different models using only the news content

Model	Accuracy	Precision	Recall	F1-score
LSTM	0.706	0.636	0.536	0.582
LSTM with Attention	0.721	0.654	0.569	0.608
BERT	0.985	0.985	0.985	0.985
BERT with LSTM and Attention	0.993	0.993	0.993	0.993

dicates that the model is the best detector of fake news.

The results have improved greatly while using both content and source of the news as compared to using only the news content. It can be seen that all the scores have improved significantly after adding the source information to the model. This shows that source information is an important aspect to detect fake news. Also, the BERT models performed better than the LSTM models in both the cases. Therefore, BERT is a better model for fake news classification than LSTM.

This result obtained by the proposed model is better than the one obtained by [14], which used AdaBoost with TF-IDF and obtained 95% accuracy. Various previous studies as discussed in section 2.2 have also used different deep learning approaches. The results obtained by the proposed model exceeds that obtained by the other state-of-the-art models.

6. Conclusion

The growing use of the Internet and Social Networking sites have boosted the dissemination of information which include both true as well as the false information. The false information has an adverse effect on people and society. So, a robust system must be developed to detect these false news and protect the society from its negative impact. This paper explores the potential of deep learning approaches to classify these false news.

In this paper, various deep learning models like LSTM and pre-trained BERT models along with a hybrid model have been used to detect fake news using the SBFN dataset under different scenarios. After the experiment, the best F1-score of 99.3% was obtained using the proposed model after considering the source information of the news. However the proposed model performed slightly less than the BERT model while using only the content information of the news and the BERT model performed the best with F1-score of 89.5% in this scenario.

Therefore, the experimental results indicate that the BERT model outperforms other traditional deep learning models like LSTM model and LSTM with attention mechanisms to detect fake news. Also, the hybrid model which is a combination of BERT model and LSTM with attention model proved to be beneficial in detecting fake news especially when the source of the news is included along with the content of the news.

The experiment has been performed on only one dataset, which contains only 2010 records. Also, only textual contents of the data is used for fake news detection. Other aspects such as the images and videos in the news articles are also vital and can be used to determine if the news is real or fake. These are the limitations of this experiment.

The scores obtained during the experiment are comparable to the state-of-the-art results Further improvement experimentation of the model is needed on a more varied dataset. This paper only considers the news in textual format and the images and videos associated with the news article have not been considered. They could also be used to detect fake news. This is the future direction of the research.

References

[1] Leonhardt D & Thompson S A. President trump's lies, the definitive list (2017). URL `https://www.nytimes.com/interactive/2017/06/23/opinion/trumps-lies.html?mtrref=en.wikipedia.org&gwh=FE924CCDF28223A5B681A150FEE60A2E&gwt=pay&assetType=PAYWALL`.

[2] Rogers K & Bromwich J E. The hoaxes, fake news and misinformation we saw on election day (2016). URL `https://www.nytimes.com/2016/11/09/us/politics/debunk-fake-news-election-day.html`.

[3] Can vaccines make our body magnetic? (2021). URL `https://www.mcgill.ca/oss/article/covid-19/can-vaccines-make-our-body-magnetic`.

[4] Lazer D M, Baum M A, Benkler Y, Berinsky A J, Greenhill K M, Menczer F, Metzger M J, Nyhan B, Pennycook G, Rothschild D et al., The science of fake news, *Science*, 359(6380) (2018) 1094–1096.

[5] Shu K, Sliva A, Wang S, Tang J & Liu H, Fake news detection on social media: a data mining perspective. corr abs/1708.01967 (2017), *arXiv preprint arXiv:1708.01967*.

LSTM

Actual Label \ Predicted Label	Fake	Real
Fake	202	47
Real	71	82

LSTM with Attention

Actual Label \ Predicted Label	Fake	Real
Fake	203	46
Real	66	87

BERT

Actual Label \ Predicted Label	Fake	Real
Fake	255	4
Real	2	141

BERT with LSTM and Attention

Actual Label \ Predicted Label	Fake	Real
Fake	265	0
Real	3	134

Figure 6: Confusion matrix of result different models using the news content along with news source.

[6] Largent W. Talos targets disinformation with fake news challenge victory (2017). URL `https://blog.talosintelligence.com/2017/06/talos-fake-news-challenge.html`.

[7] Pomerleau D & D R, The fake news challenge: Exploring how artificial intelligence technologies could be leveraged to combat fake news.

[8] Thorne J, Chen M, Myrianthous G, Pu J, Wang X & Vlachos A. Fake news stance detection using stacked ensemble of classifiers. In: *Proceedings of the 2017 EMNLP Workshop: Natural Language Processing meets Journalism* (2017), pp. 80–83.

[9] Riedel B, Augenstein I, Spithourakis G P & Riedel S, A simple but tough-to-beat baseline for the fake news challenge stance detection task, *arXiv preprint arXiv:1707.03264*.

[10] Schiller B, Daxenberger J & Gurevych I, Stance detection benchmark: How robust is your stance detection?, *KI-Künstliche Intelligenz* (2021) 1–13.

[11] Tacchini E, Ballarin G, Della Vedova M L, Moret S & de Alfaro L, Some like it hoax: Automated fake news detection in social networks, *arXiv preprint arXiv:1704.07506*.

[12] Gilda S. Notice of violation of ieee publication principles: Evaluating machine learning algorithms for fake news detection. In: *2017 IEEE 15th student conference on research and development (SCOReD)*. IEEE (2017), pp. 110–115.

[13] Conroy N K, Rubin V L & Chen Y, Automatic deception detection: Methods for finding fake news, *Proceedings of the Association for Information Science and Technology*, 52(1) (2015) 1–4.

[14] Bharadwaj A, Ashar B, Barbhaya R Parshva amd Bhatia & Shaikh P Z, Source based fake news classification using machine learning, *Innovative Research in Science, Engineering and Technology*, 9.

[15] Wang W Y, " liar, liar pants on fire": A new benchmark dataset for fake news detection, *arXiv preprint arXiv:1705.00648*.

[16] Karimi H, Roy P, Saba-Sadiya S & Tang J. Multi-source multi-class fake news detection. In: *Proceedings of the 27th international conference on computational linguistics* (2018), pp. 1546–1557.

[17] Roy A, Basak K, Ekbal A & Bhattacharyya P, A deep ensemble framework for fake news detection and classification, *arXiv preprint arXiv:1811.04670*.

[18] Long Y. Fake news detection through multi-perspective speaker profiles. Association for Computational Linguistics (2017).

[19] Ruchansky N, Seo S & Liu Y. Csi. In: *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management-CIKM '17* (2017).

[20] Ma J, Gao W, Mitra P, Kwon S, Jansen B J, Wong K F & Cha M, Detecting rumors from microblogs with recurrent neural networks.

[21] Zubiaga A, Liakata M & Procter R, Learning reporting dynamics during breaking news for rumour detection in social media, *arXiv preprint arXiv:1610.07363*.

[22] Ajao O, Bhowmik D & Zargari S. Fake news identification on twitter with hybrid cnn and rnn models. In: *Proceedings of the 9th international conference on social media and society* (2018), pp. 226–230.

[23] Deepak S & Chitturi B, Deep neural approach to fake-news identification, *Procedia Computer Science*, 167 (2020) 2236–2243.

[24] Kula S, Choraś M & Kozik R. Application of the bert-based architecture in fake news detection. In: *Conference on Complex, Intelligent, and Software Intensive Systems*. Springer (2020), pp. 239–249.

[25] Ahmed H, Traore I & Saad S, Detecting opinion spams and fake news using text classification, *Security and Privacy*, 1(1) (2018) e9.

[26] Zellers R, Holtzman A, Rashkin H, Bisk Y, Farhadi A, Roesner F & Choi Y, Defending against neural fake news, *arXiv preprint arXiv:1905.12616*.

[27] Devlin J, Chang M W, Lee K & Toutanova K, Bert: Pre-training of deep bidirectional transformers for language understanding, *arXiv preprint arXiv:1810.04805*.

[28] Bharadwaj A, Ashar B, Barbhaya, Parshva B R & Zaheed S, Souce based fake news classification using machine learning, 9(6).