Detection of Fake News Using Machine Learning and Natural Language Processing Algorithms

Noshin Nirvana Prachi, Md. Habibullah, Md. Emanul Haque Rafi, Evan Alam, and Riasat Khan
Electrical and Computer Engineering, North South University, Dhaka, Bangladesh
Email: {noshin.nirvana, md.habibullah, emanul.haque, evan.alam, riasat.khan}@northsouth.edu

***Abstract*—The amount of information shared on the internet, primarily via web-based networking media, is regularly increasing. Because of the easy availability and exponential expansion of data through social media networks, distinguishing between fake and real information is not straightforward. Most smartphone users tend to read news on social media rather than on the internet. The information published on news websites often needs to be authenticated. The simple spread of information and news by instant sharing has included the exponential growth of its misrepresentation. So, fake news has been a major issue ever since the growth and expansion of the internet for the general mass. This paper employs several machine learning, deep learning and natural language processing techniques for detecting false news, such as logistic regression, decision tree, naive bayes, support vector machine, long short-term memory, and bidirectional encoder representation from transformers. Initially, the machine learning and deep learning approaches are trained using an open-source fake news detection dataset to determine if the information is authentic or counterfeit. In this work, the corresponding feature vectors are generated from various feature engineering methods such as regex, tokenization, stop words, lemmatization and term frequency-inverse document frequency. All the machine learning and natural language processing models' performance were evaluated in terms of accuracy, precision, recall, F-1 score, ROC curve, etc. For the machine learning models, logistic regression, decision tree, naive bayes, and SVM achieved classification accuracies of 73.75%, 89.66%, 74.19%, and 76.65%, respectively. Finally, the LSTM attained 95% accuracy, and the NLP-based BERT technique obtained the highest accuracy of 98%.**

***Index Terms*—bidirectional encoder representation from transformers, fake news detection, lemmatization, long short-term memory, naive Bayes, support vector machine, tokenization**

I. Introduction

Information is significant for human dynamics and affects life practices. In earlier days, the daily news or information was presented through print media, newspapers, and electronic media such as television and radio. The data from these publishing technologies are more credible as it is either self-screened or constrained by specialists [1]. These days, individuals are presented with an extreme amount of data through various sources, particularly with the prominence of the internet and web-based media stages. The ease of internet access has caused the hazardous development of a wide range of falsehoods like malicious discussion, double-dealing, fabrications, fake news, spam assessment, which diffuses quickly and widely in the human culture. The misinformation of online social media has become a global problem in public trust and society as it has become an essential mode of communication and networking nowadays.

Nowadays, online social platforms and blogs contain a significant amount of fake and fabricated news, negatively affecting society [2]. This news is embellished with dubious facts and misleading information, causing interpersonal anxiety and detrimental social panic. This unreliable information destroys people's trust and adversely influences the economy and major political processes, such as the stock market, elections, etc. The proliferation of fake and fabricated news is generally detected manually by human verification. This manual fact-checking process is subjective in nature, laborious, time-consuming, and inefficient. In recent years, automatic systems based on machine learning and natural language processing algorithms have been utilized to tackle the issue of fake news detection [3], [4]. With the advancement of technology and artificial intelligence, these automatic systems efficiently restrain misleading and false news propagation. Thus, these techniques have created deep interest among researchers in detecting fake news for a better future endeavor.

This paper has designed a fake news detection and classification system using different types of machine learning techniques. The open-source fake news datasets of the proposed artificial news detection system contain the information of various articles' authors, captions, and main descriptions. Initially, the dataset is preprocessed using conventional techniques, e.g., regex, tokenization, stop words, lemmatization, and then applied NLP techniques, count vectorizer, TF-IDF vectorizer. The major contributions of this work are as follows:

- In this paper, an automatic fake news detection system has been developed using various machine learning and natural language processing algorithms. This work uses logistic regression, decision tree, naive bayes, and SVM machine learning techniques.

Manuscript received April 4, 2022; revised June 13, 2022; accepted July 4, 2022.

doi: 10.12720/jait.13.6.652-661

- Additionally, Long Short-Term Memory (LSTM), deep learning model and natural language processing algorithm, Bidirectional Encoder Representations from Transformers (BERT) are also implemented.
- Next, the efficiency of all the machine learning and natural language processing models are compared in terms of classification accuracy, precision, recall, F-1 score, and ROC curve.
- Finally, the performance of the proposed fake news detection system is compared with previous relevant works in terms of classification accuracy. The nobility of this work is to utilize the BERT-based NLP model for detecting fake news.

The other part of the paper is constructed as follows. In Section III, the proposed system has been discussed with appropriate equations. The actual results of the research have been shown in Section IV. Lastly, Section V concludes the paper with some directions for the future improvement of this work.

II. Related Works

Some of the recent works implemented to detect fake and fabricated news have been discussed in the following section. Machine learning and deep learning-based neural network models execute the identification and classification of real and fake news. For instance, in [5], the authors worked on fake news detection with the help of a mixed deep learning technique, CNN-LSTM. This paper has used the Fake News Challenge (FNC) dataset, which was created in 2017. They matched the claim with the news article body whether the claim matches with the article body or not. The authors have developed four data models. First, they use data without preprocessing, second with preprocessing. The authors obtained different results when they preprocessed data and when they did not. The third and fourth models are built on dimensionality reduction techniques by using PCA and Chi-square approaches. Finally, they trained on forty-nine thousand and nine hundred seventy-two samples and tested on twenty-five thousand and four hundred thirteen headlines and articles by CNN-LSTM. On their model with no knowledge of cleanup or preprocessing, the achieved accuracy was 78%. When preprocessing was done, the accuracy increased up to 93%. Next, the application of Chi-square raises the accuracy by 95%. Lastly, they conclude that using PCA with CNN and LSTM design resulted in the highest accuracy of 96%, significantly reducing the prediction time. In [6], T. Jiang *et al.* used baseline fake news identification techniques to locate the baseline methods' flaws and provide a viable alternative. First, the authors performed five completely different conventional machine learning models and three deep learning models to compare their efficiency. The authors used two datasets (ISOT and KDnugget) of various sizes to test the corresponding models' performance in this work. Finally, they take advantage of an adaptation of modified McNemar's check to decide if they are square measure contrasts between these two models' presentation, then determine the simplest model for detecting the fake news. The authors obtained accuracies of approximately 99.94% and 96.05% on the ISOT dataset and KDnugget dataset, respectively.

In a recent work [7], the authors designed a system for detecting fake news using various machine learning techniques. First, each tweet/post has been categorized as a binary categorization result by the authors. They collected data manually from their own research sets by using Twitter API and the DMOZ directory. The authors ran a test of their proposed system on the Twitter dataset. The results show that fifteen percent of fabricated tweets and forty-five percent of the actual tweets were adequately classified, and the remainder of the posts were not decided. In this paper, the author proposed the detection of deception using the labeled benchmark dataset “LIAR”. They have also improved efficiency in the detection of fake posts/news with evidence. The authors have introduced the need for hoax detection in their system. They used the ML approach by combining news content and social content. Finally, the authors claim their proposed system's performance is good compared to other works described in the literature. In [8], A. Jain *et al.* design an automated system that detects the news as false or true. Sometimes, social media like Facebook, YouTube, Twitter, and other online platforms spread the news, creating anxiety and unrest in society. In this paper, the author applied several machine learning-based fake news detection systems. The author utilized naive bayes classifier, SVM algorithm and logistic regression in their proposed detection system. They implement their model and classify the authentic and fabricated news. Finally, the proposed SVM model achieved an accuracy of 93.5%. Machine learning ensemble techniques have been used in [9] to detect and classify fake news automatically. In this regard, the textual features have been applied in different machine learning approaches. This paper used ISOT and two open-source datasets to build the proposed system. In the data preprocessing step, documents containing less than 20 words are filtered out. Next, the dialectal mechanism LIWC is employed to convert the textual features into numerical values. Next, various machine learning algorithms, logistic regression, SVM, KNN, random forest and boosting classifies have been used. Finally, the decision tree approach with 10-fold cross-validation obtained the highest accuracy of 94%.

III. Methodology

In this section, we have discussed the methodology of our work in great detail. We have explained all the regular machine learning, neural network and NLP methods that we have used in our dataset.

A. Dataset

In this work, an open-source fake news dataset from Kaggle [10] has been used. The public dataset has been created by web scrapping of different search engines. Lots of fake news and agenda always take place around us, so the whole data was curated with the help of automated data science technologies. It was posted on the data science community as a challenge to use those data to implement efficient fake news detection architecture. This specific database of fake news has been utilized in this work because it involves a diverse dataset from a wide variety of news portals and social sites. The dataset comprises 26,000 unique sample documents and has been used successfully in some papers to identify fake news [11], [12]. The original dataset has four columns, viz. id, title, author, text. The id column represents a particular numerical label for a news article; the title holds the heading of a news article; the author column contains the information about the writer of the news item; and finally, under the text column, the text of the report has been described. The training dataset has the label column, which marks the news item as potentially unreliable or reliable. It is worth mentioning that, the dataset has 20,822 unique values in the text column.

B. Data Preprocessing

We need to transform the text data using preprocessing techniques, NLP, tokenization, and lemmatization before feeding them through the ML and DL models [13]. Data preprocessing helps to remove the noises and inconsistency of data, which increases the performance and efficiency of the model. In this work, we have used traditional techniques, regex, tokenization, stopwords, lemmatization, NLP technique, and TF-IDF for data preprocessing. The implemented data preprocessing techniques are explained briefly in the subsequent paragraphs.

1) Regex

We use regex to remove punctuations from the text data. Often in the sentences, there may have extra punctuations like exclamatory signs. We use regex to remove those additional punctuations to make the dataset noise-free. Regex is based on context-free grammar.

2) Tokenization

Tokenization, preprocessing tool is used to break the sentences into words [14].

3) Stopwords

We use the English stopwords library in our preprocessing technique because our model data is English. We need to use the stopwords preprocessing technique to remove noises, make the model faster and more efficient, and save memory space.

4) Lemmatization

Lemmatization is used to transform the words into root words. We can resolve data ambiguity and inflection with lemmatization.

5) NLP technique

NLP techniques have been applied to convert the texts into meaningful numbers to feed these numbers into our proposed machine learning algorithm.

6) Bag of words

The bag of words technique converts texts into machine-understandable numbers, which is expressed as:

$$TF - IDF = TF_{td}.IDF_t \tag{1}$$

where t is a term, and d denotes the documents. TF stands for term frequency, which is a measurement of how frequently a term appears in a document. Consequently, term frequency TF is measured as:

$$TF = \frac{q_{td}}{Number\ of\ terms\ in\ the\ document} \tag{2}$$

where q is the number of times the term, t appears in the document, d .IDF denotes inverse document frequency, which indicates the importance of a particular term. IDF is calculated as:

$$IDF = \frac{\log(1+n)}{(1+df)_{dt}} + 1 \tag{3}$$

where n means the number of documents and the denominator indicates the document frequency of the term, t.

C. Machine Learning Algorithms

To detect and classify real and fake news, we have used different machine learning algorithms: logistic regression, naive bayes, decision tree, and support vector machine.

1) Logistic regression

Logistic regression is a statistical ML classification model [15]. The basis of the proposed system consists of the binary classification problem. Logistic regression is manipulated to model the probability of a certain existing event, such as true/false, reliable/unreliable, win/lose, etc. Hence, the logistic model is one of the most appropriate models for the fake news detection system. The condition for predicting logistic model is:

$$0 \leq h\theta(x) \leq 1 \tag{4}$$

The logistic regression sigmoid function is expressed as:

$$h\theta(x) = g(\theta^T X) \tag{5}$$

where,

$$g(z) = \frac{1}{(1+x^{-z})} \tag{6}$$

and the cost function of logistic regression is:

$$J(\theta) = 1/m \sum_{i=1}^{m} cost(h\theta(x^i, y^i)) \tag{7}$$

2) Naive Bayes

The naïve Bayes method is at the basis of Bayesian classifiers. It is a strategy for looking at possible outcomes that allow flipping the state around straightforwardly [16]. A conditional probability is a probability that incident X will happen provided information Y. The typical notation for this is $P(X|Y)$. We can use the naive bayes rule to compute this probability when we only have the probability of the opposite result and the two components separately.

$$P(X|Y) = \frac{P(X)\ P(Y|X)}{P(Y|X)\ P(Y)} \tag{8}$$

This restatement can be extremely useful when we are trying to predict the likelihood of something based on examples of it is happening.

In this research, we are attempting to determine if an article is false or genuine based on its contents. We may rephrase it in terms of the likelihood of that document being real or fake if it has been predetermined to be real or fake. This condition is useful since we already have instances of real and fake articles in our data collection.

Generally, a large assumption is considered for computing the likelihood of the article happening; it is equal to the product of the probabilities of each word inside its occurrence, making this procedure a “naive” Bayesian one [17]. This assumption suggests that there is no connection between the two words. It is also known as the assumption of independence. We can estimate the likelihood of a term occurring by looking at a set of real and fake article samples and noting how many times it appears in each class. The necessity for training the pre-classified samples distinguishes this method from the typical supervised learning.

3) Decision tree

The conventional J48 method is one of the most widely used classification algorithms [18]. It is based on the C4.5 algorithm, which requires all data to be studied quantitatively and categorically. As a result, continuous data will not be investigated [19]. J48 technique employs two distinct pruning techniques.

Algorithm 1. Algorithm of the proposed decision tree classification model

Input: Predefined classes with 17,000 number of features.
Output: decision tree construction.
Begin
Step 1: Make the tree's root node.
Step 2:
Return leaf node 'positive' if all instances are positive.
Return leaf node 'negative' if all instances are negative.
Step 3: Determine the current state's entropy $H.(S)$
Step 4: Calculate the entropy for each characteristic.
Step 5: Choose the attribute with the highest IG value (S, x)
Step 6: From the list of attributes, remove the attribute with the greatest IG.
Step 7: Continue until all characteristics have been exhausted or the decision tree has all leaf nodes.
End

Algorithm 1 briefly explains the building steps of the decision tree classification technique. The first approach is subtree replacement, which refers to replacing nodes in a decision tree's leaves to reduce the number of tests in the convinced route. In most cases, subtree raising has a minor influence on decision tree models. Usually, there is no accurate method to forecast an option's usefulness. However, turning it off may be advisable if the induction operation takes longer than expected because the subtree's raising is computationally complex. Next, the current state's entropy and its corresponding characteristics are determined. Consequently, the attribute with the maximum information gain is computed and removed. This process is continued until all features have been exhausted or the decision tree has all leaf nodes.

4) Support Vector Machine (SVM)

SVM, which is also known as support vector machine network, is a supervised learning method [20]. SVMs are trained using particular data that has previously been divided into two groups [21]. As a result, once the model has been trained, it is created. Moreover, the goal of the support vector machine technique is to decide any new information belongs to which group and to increase the class label [22]. The final goal of the SVM is to locate a subspace that divides the data into two parts. As Radial Basis Function (RBF) is suitable for large systems like a collection of media articles, it was chosen as the kernel for this proposed system. On two samples x and x', the radial basis function is expressed as:

$$K(x, x') = e^{-\frac{\|x-x'\|^2}{2\sigma^2}} \tag{9}$$

where $\|x - x'\|^2$ is a free parameter that denotes the squared Euclidean distance.

D. Deep Learning and Natural Language Processing Algorithms

In this work, we have used a deep learning technique, LSTM, and an NLP algorithm, BERT, to classify fake news, and both of them are dynamic.

1) Long Short-Term Memory (LSTM)

Long Short-Term Memory (LSTM) is an exclusive type of recurrent neural network, which allows information to endure. Typical RNN networks encounter short-term memory, which is solved by the cell states of the LSTM. Separate hidden motors are used in LSTMs, and their nature is to recall inputs for a long time [23]. A memory cell, also known as a gated leaky neuron or an accumulator, has a relationship in the following stages with its weight of 1. It mimics its genuine position and inserts an external signal, but this signal is multiplied by another unit that determines when to wipe or keep information from memory. Finally, the sigmoid layer-based forget gates control the transfer of data to the following hidden networks. Fig. 1 shows a generic LSTM based neural network architecture.

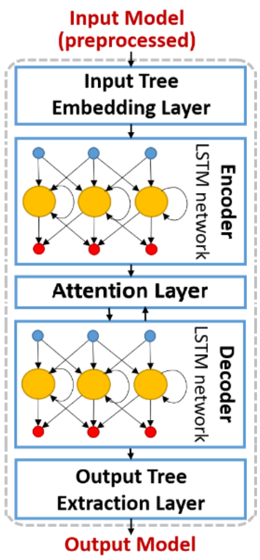

Figure 1. Generic LSTM architecture.

For our classification, we used an LSTM model with an input layer that takes the input titles and article body and an embedding layer that turns every word into a 300-pixel vector. As there are 256 features, this layer will produce a 256×300 matrix. The weights we obtain from matrix multiplication will be in the output matrix, which will generate a vector for every word. These vectors are input through an LSTM, which is subsequently transferred to a fully linked dense layer, resulting in a single final output.

Table I shows the model layers and parameters, which were trained on batches of size 256.

TABLE I. LAYERS AND PARAMETERS OF THE PROPOSED LSTM MODEL

Layer	Output Shape	Number of Parameters
Input	(None, 256)	0
Embedding	(None, 256, 300)	60,974,100
Spatial Dropout	(None, 256, 300)	0
Bidirectional	(None, 256)	439,296
Dense	(None, 64)	16,448
Dropout	(None, 64)	0
Total parameters: 61,429,909 Trainable parameters: 61,429,909 Non-trainable parameters: 0		

2) Bidirectional Encoder Representation from Transformers (BERT)

BERT is described to be pre-trained bidirectional representations from an unlabeled text by conditioning both right and left backgrounds in all levels [24]. As a result, the BERT model might suffice with only one additional output layer to produce advanced models for various tasks, including query answers. BERT is composed of two components, encoder and decoder. In this pre-training phase, this model learns about the language and its corresponding contexts. As this technique learns contexts from both directions simultaneously, the contexts of words are better learned.

For tokenizing sentences into words, converting token strings to ids and back, and encoding/decoding, BertTokenizer from the pretrained 'bert-base-uncased' model, was utilized in this study. The max sentence length is 60 characters, and we utilized the encode plus technique to encode each one of them. This technique will tokenize the phrase, prep the [CLS] (classification) token at the beginning, and append the [SEP], which tells BERT where to start the next phrase. In most cases, it is inserted after each phrase's token. Tokens should be mapped to their ids; the phrase should be padded to the attention masks, and the maximum length for [PAD] (padding) tokens should be created. The BERT model uses the argument of attention mask, which specifies which tokens should be dealt with and which can be ignored. Finally, in this step, the model is notified whether tokens include valid data or not. The architecture of the BERT model employed in this proposed detection system has been depicted in Fig. 2.

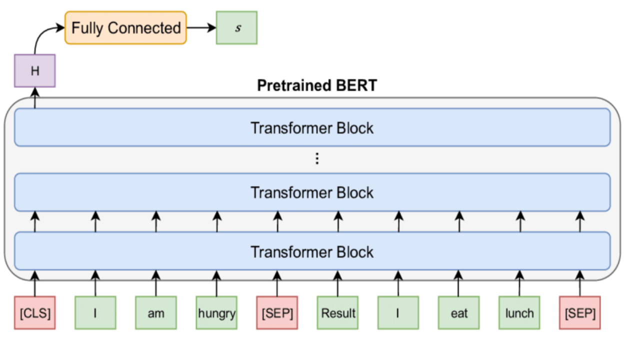

Figure 2. Architecture of BERT technique.

Table II shows the layers and parameters of the proposed BERT model input ids and attention masks used as the input layer. After that, the output of the input layers goes to the transformer BERT model, which is subsequently transferred to a fully linked dense layer, resulting in a single final output.

TABLE II. LAYERS AND PARAMETERS OF THE PROPOSED BERT-BASED NLP MODEL

Layer	Number of Parameters	Connected to
Input	0	
Attention masks	0	
TF BERT model	109,482,240	Input [0][0] Attention masks [0][0]
Dense	24,608	TF BERT model [0][1]
Dropout	0	Dense [0][0]
Total parameters: 109,506,881 Trainable parameters: 109,506,881 Non-trainable parameters: 0		

IV. RESULT AND ANALYSIS

This section discusses the numerical results of the proposed fake news detection system with the applied regular ML, DL, and NLP approaches. The employed fake news dataset has been divided into 8:2 training and testing samples. After completion of the necessary processing and training of the dataset, all the models are assessed. In this work, all the models are evaluated in various ways by checking their accuracy, confusion matrix, recall, precision, F1-score, ROC curve, and other metrics.

A. Performance of Logistic Regression Model

In Fig. 3, the confusion matrix for the logistic regression model of the proposed system has been shown. The real news class has 862 right predictions and 170 wrong predictions from 1032 test samples of real news. Therefore, the accuracy for real news prediction is 83.52%, and for the fake news class, it has 487 correct predictions but a significant number of the wrong predictions of 310 from 797 test samples. So, the accuracy for fake news is 61% and finally, the overall accuracy is 74%.

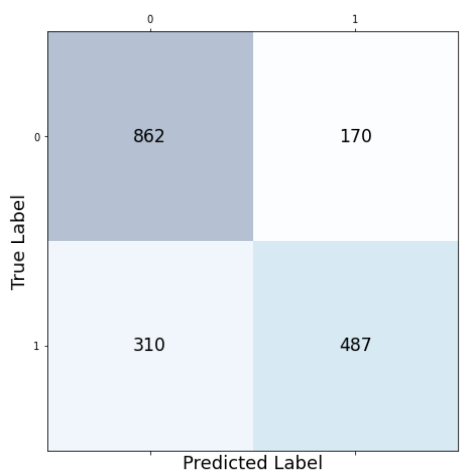

Figure 3. Confusion matrix for logistic regression.

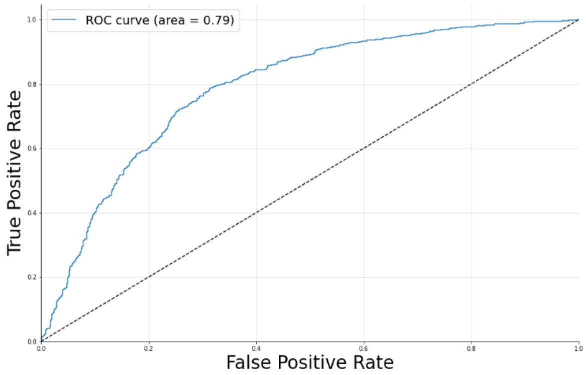

Figure 4. ROC curve for logistic regression.

According to Fig. 4, the area under the curve (AUC) score of the ROC curve of the proposed logistic regression algorithm is 0.79. The rest of the performance metrics for the logistic regression model are demonstrated in Table III. The proposed logistic regression model's precision, recall, and F1-score are 74%, 72%, and 73%, respectively.

TABLE III. LOGISTIC REGRESSION MODEL'S PERFORMANCE METRICS

	Precision	Recall	F1-score
0 (Not Fake)	0.74	0.84	0.78
1 (Fake)	0.74	0.61	0.67
Accuracy			0.74
Weighted Average	0.74	0.74	0.73

B. Performance of Naive Bayes Model

The confusion matrix for the naive bayes model of the proposed system has been shown in Fig. 5. The authentic news class has 830 right predictions and 202 wrong predictions from the total 1032 test samples. So, the accuracy for real news prediction is 80%, and for the fake news class, it has a significant number of wrong classifications similar to the logistic regression model. Finally, the accuracy for fake news is 66%, and the overall accuracy is 74%.

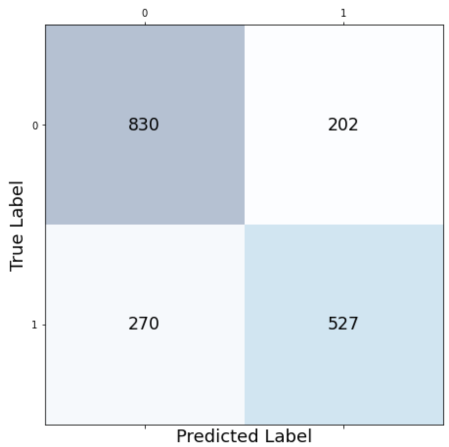

Figure 5. Confusion matrix for naive bayes.

The true and false positive rates of the proposed naive bayes approach are depicted in Fig. 6. According to Fig. 6, the naive bayes model has an ROC AUC score of 0.79. In Table IV, the rest of the performance metrics for the naive bayes model are demonstrated. The precision, recall, and F1-score of the proposed naive bayes model are 74%, 73%, and 73%, respectively.

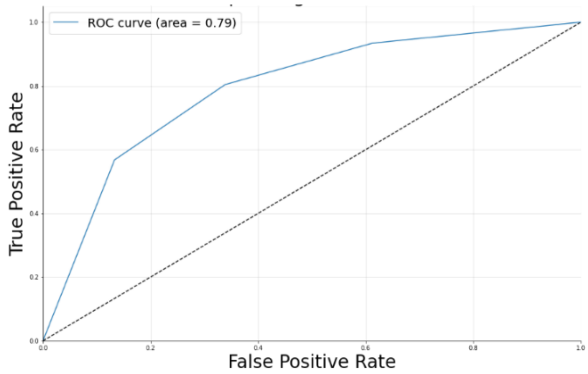

Figure 6. ROC curve for naive bayes.

TABLE IV. VARIOUS EVALUATION METRICS OF THE NAIVE BAYES MODEL

	Precision	Recall	F1-score
0 (Not Fake)	0.75	0.80	0.78
1 (Fake)	0.72	0.66	0.69
Accuracy			0.74
Weighted Average	0.74	0.74	0.74

C. Performance of Decision Tree Model

In Fig. 7, the confusion matrix for the decision tree model of the proposed system has been demonstrated. The real news class has 940 right predictions and 92 wrong predictions from 1032 test samples of real news. So, the accuracy for real news prediction is 91%, and for the fake news class, it has 700 correct predictions but an acceptable number of the wrong predictions of 97 from 797 test samples of fake news. So, the accuracy for fake news is 88%. Finally, the decision tree technique achieved an overall accuracy of 90%.

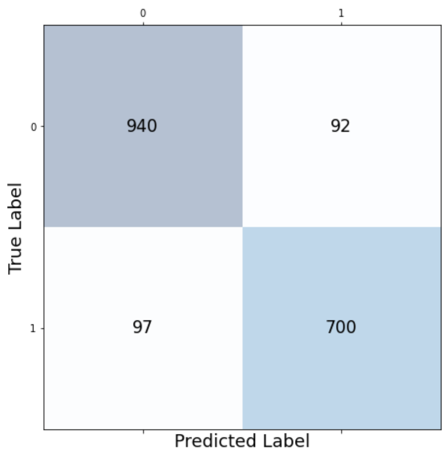

Figure 7. Confusion matrix for decision tree.

According to Fig. 8, the ROC AUC value of the proposed decision tree algorithm is 0.89. In Table V, the rest of the performance metrics for the decision tree model are demonstrated. The precision, recall, and F1-score of the proposed decision tree model are 90%, 89%, and 89%, respectively.

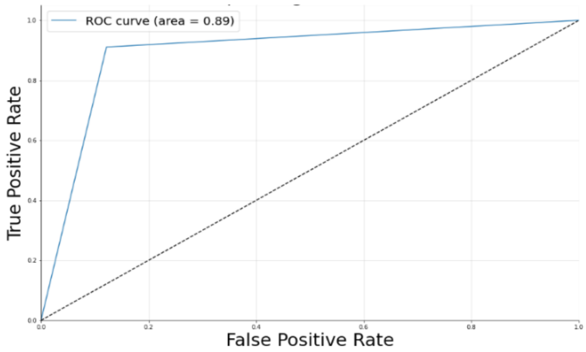

Figure 8. ROC curve for decision tree.

TABLE V. DECISION TREE MODEL ACCURACY METRICS

	Precision	Recall	F1-score
0 (Not Fake)	0.91	0.91	0.91
1 (Fake)	0.88	0.88	0.88
Accuracy			0.90
Weighted Average	0.90	0.90	0.90

D. Performance of Support Vector Machine Model

In Fig. 9, the confusion matrix for the SVM model with the RBF kernel of the proposed system has been shown. The accuracies of the real and fake news are 82% and 70%, respectively. Finally, the overall accuracy of the SVM classifier model is 77%.

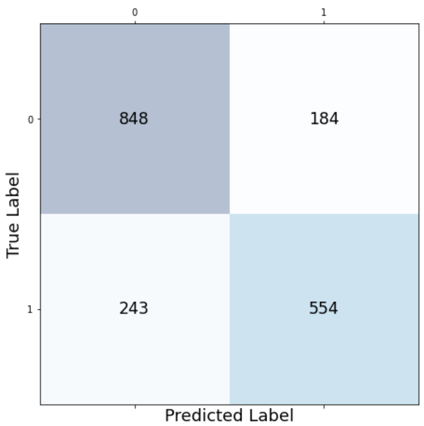

Figure 9. Confusion matrix for SVM.

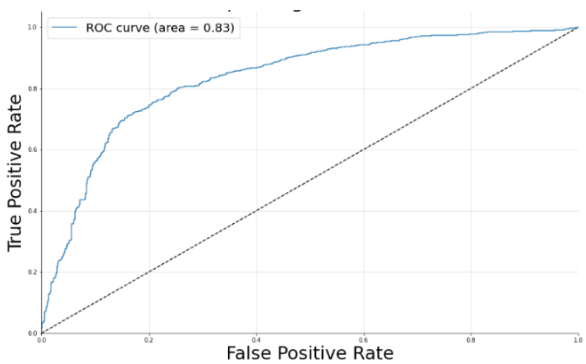

Figure 10. ROC curve for SVM.

According to Fig. 10, the ROC AUC coefficient of the proposed SVM algorithm is 0.83. Table VI depicts the rest of the performance metrics for the SVM model.

TABLE VI. SVM MODEL ACCURACY METRICS

	Precision	Recall	F1-score
0 (Not Fake)	0.78	0.82	0.80
1 (Fake)	0.75	0.70	0.72
Accuracy			0.77
Weighted Average	0.77	0.77	0.77

E. Performance of LSTM Model

Fig. 11 illustrates the confusion matrix for the deep learning-based LSTM model of the proposed system. The real news class has 1920 right predictions and 157 wrong predictions. So, the accuracy for real news prediction is 92%, and for the fake news class, the prediction is significantly improved compared to other ML techniques. Finally, the overall accuracy of the LSTM technique is 95%. The total number of test samples for each class is different from the ML approaches because of the better preprocessing for NLP methods which helps to decrease the chances of removing samples.

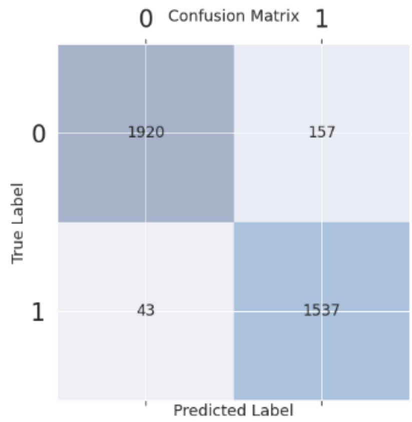

Figure 11. Confusion matrix for LSTM.

According to Table VII, other performance metrics for the LSTM model demonstrated better results. The precision, recall, and F1-score of the proposed LSTM model are 94%, 95%, and 94%, respectively.

TABLE VII. PERFORMANCE METRICS OF THE LSTM APPROACH

	Precision	Recall	F1-score
0 (Not Fake)	0.98	0.92	0.95
1 (Fake)	0.91	0.97	0.94
Accuracy			0.95
Weighted Average	0.95	0.95	0.95

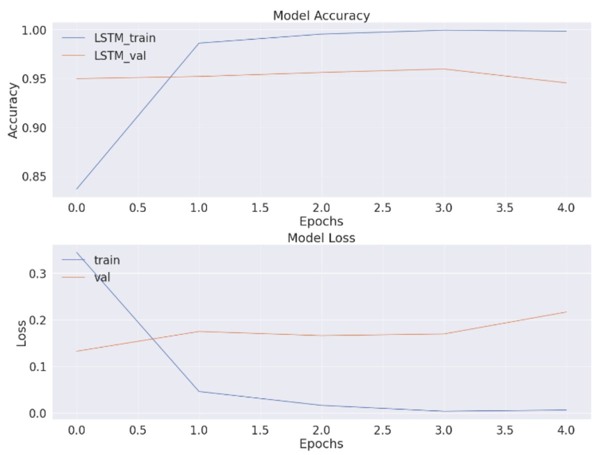

Figure 12. Accuracy and loss vs. epochs graph of LSTM.

Fig. 12 shows the accuracy and loss vs. epochs graphs of LSTM with respect to epoch. For the LSTM model, initially, the model's validation accuracy was 95%, which did not vary significantly with the change of the epochs.

F. Performance of BERT Model

In this section, the results for the proposed fake news detection system implemented on the BERT technique will be discussed. Table VIII shows the encoder and decoder result on an example sentence. The purpose of this result is to show how all of the input sentences are encoded and decoded. Here the input for the encoding is "Hi nice meet you!". After encoding, we can see that all the words and symbols represent a value, i.e., "hi" is assigned a numerical value of 7632. If we decode it, we will get the exact output given to the encoder as an input. There are two new words after decoding. One is at the beginning of the sentence, which is CLS, which represents classification. Another one is SEP at the end of the sentence, which tells BERT where to start the following sentence.

TABLE VIII. ENCODER AND DECODER EXAMPLE RESULTS

Input	encode = bert_tokenizer.encode ("Hi nice meet you !") decode = bert_tokenizer.decode (encode)
Command	print ("Encode: X", encode) print ("Decode: X", decode)
Output	Encode: [101, 7632, 3835, 3113, 2017, 999, 102] Decode: [CLS] hi nice meet you! [SEP]

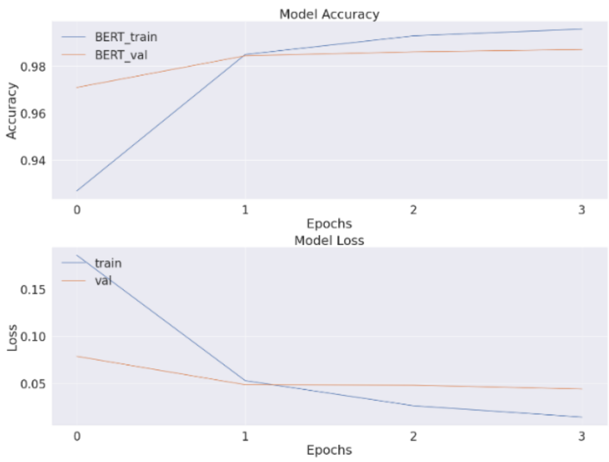

Figure 13. Accuracy and loss vs. epochs graphs of BERT framework.

Fig. 13 shows the accuracy and loss graph of BERT with respect to epoch. For the BERT model, at the initial stages of training, the model's validation starts from 97%, which did not change remarkably, and after three epochs, it increased only by 1% and achieved 98%.

G. Model Comparison of Our Paper

In Table IX, comparison for all the applied detection models have been demonstrated that we have trained in this work. For the machine learning-based techniques, the fake news detection performs well for the decision tree classifier, but the naive bayes and logistic regression approaches perform unsatisfactorily. The highest accuracy from machine learning models is 90% for the decision tree approach. The deep learning LSTM approach achieved the second-highest accuracy of 95%. Finally, the best detection performance is offered by the NLP-based BERT technique, with 98% accuracy.

TABLE IX. ACCURACY COMPARISON OF DIFFERENT APPLIED TECHNIQUES

Models	Precision	Recall	F1-Score	Accuracy
Logistic Regression	74%	72%	73%	74%
Naive Bayes	74%	73%	73%	74%
Decision Tree	90%	89%	89%	90%
SVM	76%	76%	76%	77%
LSTM	94%	95%	94%	95%
BERT				98%

H. Model Comparison with Others Work

Finally, the proposed fake news detection system with the BERT technique has been compared with other related works. According to Table X, the implemented BERT approach outperformed all the other works in terms of accuracy.

TABLE X. ROPOSED MODEL'S ACCURACY COMPARISON WITH RELATED WORKS

Reference	Applied Method	Accuracy
[3]	Random forest	95%
[4]	Decision tree	96.8%
[5]	CNN+LSTM with PCA	96%
[8]	SVM	93.5%
[9]	Decision tree	94%
[25]	Deep neural network	94%
Our study	BERT	98%

V. CONCLUSION

Finding the accuracy and credibility of information and news that is available on the internet is critical nowadays. It has recently been discovered that various online platforms significantly influence disseminating misleading information and spreading fake news to serve several dreadful purposes and benefit many people. Because of the plethora of spreading and sharing data on the internet, there is a growing demand for automated false news identification systems that are accurate and efficient. This paper proposes an automatic fake news detection system that utilizes various regular machine learning, deep learning, and natural language processing techniques. Various feature extraction methods, such as regex, tokenization, stopwords, lemmatization, NLP, TF-IDF, were used to preprocess the data in this suggested system. Next, several models, logistic regression, decision tree, naive bayes, support vector machine, long short-term memory, bidirectional encoder representation from transformers have been employed to classify the fabricated news. For the machine learning model logistic regression, decision tree, naive bayes, and SVM, we got 73.75%, 89.66%, 74.19%, and 76.65% accuracies, respectively. Finally, substantial better performance was achieved by the neural network LSTM and NLP-based BERT techniques. In the future, the proposed system can be extended to detect more specific false news with various categories, e.g., religious, political, COVID-19, etc. The word2vec approach can be applied to deal with and classify images and video-related visual datasets. News data from diverse languages can be utilized to identify false news from different nations and countries. A future extension of this work can be to employ attention-based deep learning approaches.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

AUTHOR CONTRIBUTIONS

M. E. H. Rafi proposed the research idea; N. N. Prachi and MH conducted the research; E. Alam and R. Khan analyzed the data; N. N. Prachi, M. Habibullah and M. E. H. Rafi wrote the paper; R. Khan helped to draft the final manuscript; all authors had approved the final version.

REFERENCES

[1] J. Strömbäck, Y. Tsfati, H. Boomgaarden, *et al.*, "News media trust and its impact on media use: Toward a framework for future research," *Annals of the International Communication Association*, vol. 44, pp. 139-156, 2020.

[2] E. Mitchelstein and P. J. Boczkowski, "Online news consumption research: An assessment of past work and an agenda for the future," *New Media & Society*, vol. 12, pp. 1085-1102, 2010.

[3] P. Henrique, A. Faustini and T. F. Covões, "Fake news detection in multiple platforms and languages," *Expert Systems with Applications*, vol. 158, pp. 1-9, 2020.

[4] F. A. Ozbay and B. Alatas, "Fake news detection within online social media using supervised artificial intelligence algorithms," *Physica A: Statistical Mechanics and its Applications*, vol. 540, pp. 1-19, 2020.

[5] M. Umer, "Fake news stance detection using deep learning architecture (CNN-LSTM)," *IEEE Access*, vol. 8, pp. 156695-156706, 2020.

[6] T. Jiang, J. P. Li, A. U. Haq, *et al.*, "A novel stacking approach for accurate detection of fake news," *IEEE Access*, vol. 9, pp. 22626-22639, 2021.

[7] S. I. Manzoor, J. Singla, and Nikita, "Fake news detection using machine learning approaches: A systematic review," in *Proc. International Conference on Trends in Electronics and Informatics*, 2019, pp. 230-234.

[8] A. Jain, A. Shakya, H. Khatter, *et al.*, "A smart system for fake news detection using machine learning," in *Proc. International Conference on Issues and Challenges in Intelligent Computing Techniques*, 2019, pp. 1-4.

[9] I. Ahmad, M. Yousaf, S. Yousaf, *et al.*, "Fake news detection using machine learning ensemble methods," *Complexity*, pp. 1-11, 2020.

[10] UTK machine learning club. (July 2017). Fake news, version 1. [Online]. Available: https://www.kaggle.com/c/fake-news/data

[11] H. Ali, M. S. Khan, A. AlGhadhban, *et al.*, "All your fake detector are belong to us: Evaluating adversarial robustness of fake-news detectors under black-box settings," *IEEE Access*, vol. 9, pp. 81678-81692, 2021.

[12] I. K, Sastrawan, I. P. A. Bayupati, and D. M. S. Arsa, "Detection of fake news using deep learning CNN-RNN based methods," *ICT Express*, pp. 1-13, 2021.

[13] Y. A. Solangi, Z. A. Solangi, S. Aarain, *et al.*, "Review on Natural Language Processing (NLP) and its toolkits for opinion mining and sentiment analysis," in *Proc. International Conference on Engineering Technologies and Applied Sciences*, 2018, pp. 1-4.

[14] G. Kim and S. H. Lee, "Comparison of Korean preprocessing performance according to Tokenizer in NMT transformer model," *Journal of Advances in Information Technology*, vol. 11, pp. 228-232, 2020.

[15] T. Daghistani and R. Alshammari, "Comparison of statistical logistic regression and random forest machine learning techniques in predicting diabetes," *Journal of Advances in Information Technology*, vol. 11, pp. 78-83, 2020.

[16] W. He, Y. He, B. Li, *et al.*, "A naive-Bayes-based fault diagnosis approach for analog circuit by using image-oriented feature extraction and selection technique," *IEEE Access*, vol. 8, pp. 5065-5079, 2020.

[17] Q. Xue, Y. Zhu, and J. Wang, "Joint distribution estimation and naïve bayes classification under local differential privacy," *IEEE Transactions on Emerging Topics in Computing*, vol. 9, pp. 2053-2063, 2021.

[18] H. A. Maddah, "Decision trees based performance analysis for influence of sensitizers characteristics in dye-sensitized solar cells," *Journal of Advances in Information Technology*, vol. 13, pp. 271-276, 2022.

[19] I. D. Mienye, Y. Sun, and Z. Wang, "Prediction performance of improved decision tree-based algorithms: A review," *Procedia Manufacturing*, vol. 35, pp. 698-703, 2019.

[20] J. A. C. Moreano and N. B. L. S. Palomino, "Global facial recognition using gabor wavelet, support vector machines and 3D face models," *Journal of Advances in Information Technology*, vol. 11, pp. 143-148, 2020.

[21] A. B. Gumelar, A. Yogatama, D. P. Adi, *et al.*, "Forward feature selection for toxic speech classification using support vector machine and random forest," *International Journal of Artificial Intelligence*, vol. 11, pp. 717-726, 2022.

[22] J. Cervantes, F. García-Lamont, L. Rodríguez, *et al.*, "A comprehensive survey on support vector machine classification: Applications, challenges and trends," *Neurocomputing*, vol. 408, pp. 189-215, 2020.

[23] I. Benchaji, S. Douzi, and B. E. Ouahidi, "Credit card fraud detection model based on LSTM recurrent neural networks," *Journal of Advances in Information Technology*, vol. 12, pp. 113-118, 2021.

[24] N. Yadav and A. K. Singh, "Bi-directional encoder representation of transformer model for sequential music recommender system," in *Proc. Forum for Information Retrieval Evaluation*, 2020, pp. 49-53.

[25] S. Ni, J. Li, and H. Y. Kao, "MVAN: Multi-view attention networks for fake news detection on social media," *IEEE Access*, vol. 9, pp. 106907-106917, 2021.

Noshin Nirvana Prachi obtained her bachelor's degree in computer science and engineering in July 2021 from North South University, Bangladesh. Noshin was born in Dhaka, Bangladesh. One of her research work on deep learning-based speaker recognition system was published at Interdisciplinary Research in Technology and Management (IRTM) conference. She is working on data science, machine learning, computer vision and software engineering.

Md. Habibullah completed his B.Sc. degree in computer science and engineering in 2021 from North South University, Bangladesh's electrical and computer engineering department. Recently he has published a manuscript on a deep learning-based speaker recognition system at an IEEE conference. Currently, he is doing research on data science, machine learning, cryptography and cyber security.

Md. Emanul Haque Rafi received his bachelor of science degree in computer science and engineering from the electrical and computer engineering department of North South University, Bangladesh. Emanul was born in Dhaka, captial city of Bangladesh. His primary research interest includes data science and management, machine learning, deep learning, and natural language processing.

Evan Alam has a bachelor's degree in computer science and engineering from electrical and computer engineering department of North South University, Bangladesh. He was an active member of the Computer & Engineering Club of North South University during his undergraduate study. Currently, his primary research interests are computer vision, data science, machine learning, and computer network security.

Riasat Khan received a B.Sc. degree in Electrical and Electronic Engineering from the Islamic University of Technology, Bangladesh, in 2010. He completed his M.Sc. and Ph.D. degrees in Electrical Engineering from New Mexico State University, Las Cruces, USA, in 2018. Currently, Dr. Khan is working as an Assistant Professor in the Department of Electrical and Computer Engineering at North South University, Dhaka, Bangladesh. His research interests include biomedical engineering, cardiac electrophysiology and computational bioelectromagnetics.