iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 14 Issue: IV Month of publication: April 2026

DOI: https://doi.org/10.22214/ijraset.2026.78248

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue IV Apr 2026- Available at www.ijraset.com

Bangla Fake News Detection using DistilBERT

Pikul Das Joy[1], Md Simon Alam Sakib[2], Syed Md. Minhaz Hossain[3]
Department of Computer Science and Engineering, Premier University, Chittagong

Abstract: Living in the era of social media exposes people to vast amounts of information, much of which is unreliable. The widespread circulation of fake news poses serious risks, particularly when it mimics credible reporting. While fake news detection has been extensively studied in high-resource languages like English, low-resource languages such as Bangla remain underexplored due to limited linguistic tools and datasets. This study focuses on detecting fake news in Bengali by developing a DistilBERT-based model, achieving an accuracy of 86% on an online Bangla Fake News dataset. Model performance was evaluated using accuracy, precision, recall, and F1-score, highlighting the importance of dataset balance for effective classification. The proposed approach aims to rapidly identify fake news using natural language processing, reducing users' exposure to misinformation. A comparative analysis with an LSTM model further demonstrates the effectiveness of DistilBERT, underscoring the promise of advanced NLP techniques in combating misinformation within Bengali-speaking digital communities.

Keywords: Bangla Fake News Detection; DistilBERT; Transformer; Natural Language Processing; pattern recognition;

I. INTRODUCTION

Fake news is one kind of information that has been intentionally created to mislead or manipulate the readers. The reason for fake news is either commercial or political to make money or change opinion. Fake news manipulates people's minds to think a certain way in addition makes people support a particular opinion. Fake news is not always just lies; more often it is a mixture of lies and truths. As information spreads very quickly in addition freely on digital technologies in addition social media, fake news moreover spread much more quickly and to many more people. So much research has been done already to detect fake news by using English data. However, Using Bangla data detecting Bengali fake news is still in the early stage.

In order to solve the problem of fake news identification in the Bengali language, this report focuses on the application of advanced natural language processing (NLP) techniques, particularly using the DistilBERT model. DistilBERT is a simplified version of the BERT (Bidirectional Encoder Representations from Transformers) model, has demonstrated good performance in addition computing efficiency in a range of natural language processing applications. It is an excellent option for the difficult task of recognizing false information in the Bengali language because of its capacity to collect complex syntax and context-related data.

By exploring details of Bangla fake news identification with DistilBERT, we hope to add valuable information to the growing corpus of research on false information detection in many different cultural and linguistic contexts.

Although the Internet is a fantastic source of news and information, not everything found there is reliable. Any article or video that presents false information under the appearance of a reliable news source is considered fake news. Although fake news is not exclusive to the Internet, it has lately grown to be a significant issue in the modern digital world. Websites that specialize in fabricated or sensationalized tales are usually the source of fake news. Provocative headlines such as "Politician selling toxic waste on the black market" or "Celebrity endorses not brushing teeth" are frequently used. These headlines may appear dubious or even incredible to the extent of being foolish, leading one to consider false news as innocuous.

While certain individuals who produce and distribute false information may have political or social motives, others adopt a more business-oriented approach by utilizing emotionally appealing fake news to generate revenue from digital advertisements displayed alongside the content.

The utilization of false information to disseminate propaganda might pose a significant threat. Furthermore, beyond from influencing public perception and conduct, it can also foster skepticism, promote disagreement, and divert focus away from genuine news. Although it takes effort to ensure that an article is truthful and correct, those who create fake news hope that no one will examine their work for veracity. Recall that extraordinary claims necessitate extraordinary proof. Before sharing the articles in online, use caution because Rumors have been the cause of numerous terrible situations in Bangladesh during the past few years. In Bangladesh, numerous pagodas were set on fire in 2012 when a picture appeared that depicted the Quran in an offensive way. The young Buddhist who was tagged in the picture was not in any way associated with the picture. Even if the identity of the photograph could not be confirmed, a lot of rumors circulated because of it [2].

A fake rumor regarding the Padma Bridge authorities risking people's lives at the building site became viral online in 2019. This rumor then gives rise to suspicions that unidentified people are abducting children. An example of this can be found in Bangladesh in the Ramu incident of 2012, when over 25,000 people joined forces to destroy Buddhist temples based on a Facebook post from a fictitious account [3]. One of the factors that may have affected the outcome of the 2016 US election is the 25% of Americans who visited a fake news website during the six-week election period (Grave et al., 2018) [5]. Rumors have been the cause of numerous terrible situations in Bangladesh during the past few years. Rumors that human sacrifice will be required for the construction of the Padma Bridge caused crowds to beat five persons to death and wound ten more in July 2019[8]. The cost of salt increased in Bangladesh on November 19, 2019 [16], particularly in Dhaka. People have been going to stores all around the country to buy and stockpile salt in reaction to rumors that the price of salt may increase. People bought into the story of the price increase when onions became much more expensive. Instead of their regular one kilogram of salt, they purchased and kept four to five kilograms. Deepfakes have raised concerns since they can be used to produce deceptive videos and fake news. In response to a purportedly disparaging social media post by a fictitious Hindu fisherman against Islam on October 30, 2016, Islamic extremists attacked the minority Hindu population in Nasir Nagar Upazila, Bangladesh, resulting in the Nasir Nagar riots of 2016. Around 300 homes were vandalized in addition to 19 temples, and more than 100 individuals were hurt in the incident.

Regarding bogus news on Facebook A communal attack occurred on June 2, 2017, targeting the Chakma group residing in Langadu, an Upazila in Rangamati. Three villages' worth of homes were reduced to ashes as a result, and over 400 families now face the possibility of losing both their personal belongings and their lives.

False information circulated during Comilla's Durga Puja, suggesting that Hindus were honoring the goddess with the sacred Quran. But later on, it was discovered that Iqbal was the guy who completed this task. But prior to that, there were other individuals who caused violence against the Hindu population there. As a result, we are putting out a methodology that will more precisely identify bogus news by utilizing distinct natural language processing techniques.

One of the largest problems any nation has is the rise in violent crimes, including murders, as a result of the spread of fake news. Social media and the internet environment have an impact on people. As a result, they spread the word about any story they come across with an attention-grabbing headline even though they are unsure of its veracity. With the use of social media, anyone with a keyboard can write and distribute fake news quickly, all while keeping their websites profitable. These individuals don't give a damn if the false information leads to bloodshed. Some people don't care whether incorrect information hurts other people because they are preoccupied with the number of Likes and Reactions they receive after sharing anything on social media.

In Bangladesh, a sizable portion of internet users lack adequate digital literacy. Bangladeshi internet users searched social media for health-related information during the COVID-19 pandemic. The initial false material about COVID-19 that surfaced online in this nation was religious in nature and stated that eating Thankuni leaves, also known as Indian pennywort, and frequently reciting the blessing "in the name of Allah" would keep one from contracting the virus. Additionally, there is a story circulating on WhatsApp and Facebook that in order to construct the Padma Bridge, human sacrifices must also be made, which is why people are attempting to kidnap youngsters. Bangladesh is a multiethnic, multireligious, and multicultural nation. These groups could become tense and violent as a result of fake news. Maintaining societal harmony can be aided by identifying and refuting misleading information. By disseminating untrue information about companies, financial regulations, or trade partnerships, false information can have an effect on the economy. In addition to being essential for the nation's general economic stability, trustworthy information is also critical for businesses to make educated judgments. As a result, we are putting out a methodology that will more precisely identify bogus news by utilizing distinct natural language processing techniques.

II. RELATED WORK

Bangla article classification has already benefited greatly from the contributions of many outstanding people. In order to comprehend the emotions of these articles, the authors of paper [1] employ reputable Bangla news texts from numerous newspapers as well as internet journals. Where the texts will be divided into ten categories. As long as their model can be successfully learnt, these ten sorts of data will be recognized. Together, they are completing this task on Bangla messages. They've created a mechanism that makes it simple to distinguish between the data presented in Bangla. The ten categories are separated into the following groups: "Entertainment," "National," "Kolkata," "State," "International," "Sport," "Nation," "World," and "Travel." Long Short-Term Memory and Convolutional Neural Networks are used in the implementation of this study. Their working accuracy is significantly higher when they utilize the hybrid model when they employ a single model. Additionally, they preprocess Bangla texts for model training using techniques including word embedding, punctuation removal, and stop word removal. The main objective of emotion research is to divide the task into positive and negative amplitude in order to differentiate parental details or attitudes.

Md. Elias Hossain et all, [2] aims to detect false news articles in Bangla. To train their corpus, 57000 news items related to authenticity and falsity are used. This work applies Glove and Fast Text models to the Bi-LSTM using K-fold cross-validation. The accuracy of the study, which also examined state-of-the-art techniques like the Gated Recurrent Unit (GRU), was 77%. In the striking contrast, they saw the accuracy of the Bi-LSTM, which also suggests their proposed model.

Md Zobaer Hossain et all, [4] a about 50K news article annotated dataset that can be utilized to create automated systems for spotting fake news in languages with limited resources, like Bangla. Furthermore, we offer an examination of the dataset and create a reference system utilizing cuttingedge natural language processing methods to detect false information in Bangla. They use both neural network-based techniques and conventional linguistic aspects to develop this system. They anticipate that this dataset will be a useful tool for developing tools that stop the spread of false information and aid in research involving low-resource languages. They achieve 99% accuracy for CNN, BERT, and LSTM in their test set.

Risul Islam Rasel et all, [6] In order to achieve some state-of-the-art results, they also constructed a fake news dataset with 4678 distinct news stories. They experimented with the data using multiple machine learning (LR, SVM, KNN, MNB, Adaboost, and DT), deep neural networks (LSTM, BiLSTM, CNN, LSTM-CNN, BiLSTM-CNN), and transformer (BanglaBERT, m-BERT) models. In terms of performance, CNN, CNN-LSTM, and BiLSTM are the top models. In addition to improving accuracy, their models demonstrate a noteworthy rise in memory of fake news information when compared to earlier research. The accuracy of the suggested model is 99.8%.

One of the work Md Gulzar Hussain et all, [7] illustrates the exploratory study of identifying false information in Bangla social media, as much work needs to be done in this field. Throughout this research project, they have used two supervised machine learning techniques—Support Vector Machine (SVM) and Multinomial Naive Bayes (MNB) classifiers—to identify bogus news in Bangla. Inverse Document Frequency - Term Frequency For feature extraction, Vectorizer and Count Vectorizer have both been utilized. The proposed approach identifies bogus news based on the connected post's polarity. In the end, this study recommends MNB (95%) and SVM (98%) with linear kernels for good accuracy.

Another work Tasnuba Sraboni et all, [9] they have pre-processed their dataset in addition to using feature extraction methods. Experimental investigation on real-world data shows that, in comparison to other machine learning classifiers, both the Passive Aggressive Classifier and the Support Vector Machine obtain good accuracy.

The author Iftikhar Ahmad et all, [10] suggest classifying news articles automatically using an ensemble machine learning method. Their research examines many textual characteristics that can be utilized to differentiate authentic content from counterfeit. They train a variety of machine learning algorithms employing those properties in conjunction with different ensemble approaches, and then assess the algorithms' performance on four real-world datasets. Their suggested ensemble learner strategy outperforms individual learners, according on experimental evaluation.

In this paper the author Sadik Al Jarif et all, [11] aims to construct a tree with pieces from Bengali and English. They suggest four models that have good detection rates for both English and Bangla fake news. The BiLSTM model yields a 97% accuracy increase.

In this work [12] they have suggested using an automated method based on deep learning to verify the veracity of Bangla news. The primary advantage of the suggested model above other models already in use is its extremely competitive performance. They created a deep learning model, which our chosen dataset was used to train and validate. The collection includes 1,299 fake news stories and 48,678 real news stories for educational purposes. They employed random under sampling and ensemble to produce the combined output in order to deal with the unbalanced data. The suggested model produced a 99% recall and 98.29% accuracy in Bangla text processing.

In this paper [13] They offered an evaluation of the various approaches put forth and compared them in order to address the imbalance problem in Bangla fake news identification. They also suggest a method for enhancing performance in the event that the dataset is unbalanced.

In this piece of work,[14] Their multichannel mixed CNN LSTM model works remarkably well in identifying fake content, particularly bogus news in Bangla. Additionally, this model can be used to issues in this field. Additionally, the data set is useful for tackling challenges similar to this one, which is the classification of Bangla text. They collected a set of web data, comprising roughly 50,000 news stories, and tried to apply deep learning techniques to it. The accuracy of the suggested Multichannel hybrid CNN-LSTM architecture model increased to 75.05%.

Shawly Rohman et all,[15] unveiled the IBFND, or Improved Bangla Fake News Dataset. They have addressed the issue of an unbalanced dataset in their dataset and have employed web scraping to augment the quantity of false news. Furthermore, they have resorted to translating a few bogus English news articles using Google Translate. Subsequently, they have applied the multinomial NB, BERT, XGBoost, and linear SVC models to IBFND.

The maximum F1-score achieved when all models were run on the BanFakeNews dataset was 81%. They have used the BERT deep neural network model to obtain the highest F1-score of 97% that can be attained using the IBFND.

In this study [17] given the paucity of research in this area, it presents an experimental benchmark inquiry into the identification of bogus news on a Bengali news website. In order to establish a standard for identifying Bangla false news, this study examines 11434 instances of both fake and genuine news written in the Bengali language. It also assesses how well machine learning and deep learning algorithms perform in these scenarios. The effectiveness of the model is compared in this study to several linguistic traits and word vectorizers.

In this paper [18] A deep hybrid model that uses traditional machine learning approaches for classification along with 1D convolutional neural networks (CNNs) for feature extraction is offered as a means of detecting fake news in Bangla. The suggested model successfully separates false news from real news in the BanFakeNews dataset, achieving similar performance to the majority of other state-of-the-art models for the total and fake only datasets, respectively, of about 99% and 82% in the F1 measure.

MD. RAFI-UR-RASHID et all [19], apply a subset of the majority samples to create an ensembling of deep learning models. Moreover, impose the focus loss function while classifying data that is imbalanced. To raise the minorityf1 score, they also use data resampling, hidden feature extraction, and the outlier identification technique. Our experiments are all solely concerned with textual content analysis.

In this paper [21] They produce a dataset known as BanMANI that contains social media content that has been flagged for information manipulation in relation to reference papers. The methodology for gathering datasets outlined here overcomes the constraints of the Bangla NLP tools that are currently accessible. They anticipate that these methods will translate to the creation of comparable datasets in additional low-resource languages. BanMANI serves as the foundation for both training or finetuning new models expressly for this job as well as assessing the performance of current NLP systems.

III.METHODOLOGY

A. Overview

DistilBERT is a smaller, faster, cheaper, and light Transformer model trained by distilling BERT base. It is a pre-trained model. It works with GRU layer, some dense layers, and a dropout layer. The model is trained in such a way that if a input data is given to model it generates captions which shows whether the news is real or fake. A bidirectional encoder is integrated into DistilBERT to define word representations that are eventually utilized for a number of downstream tasks, such as the detection of false news. Using DistilBERT's tokenizer, Bangla news articles are tokenized into sub-word units in addition transformed into numerical representations. The DistilBERT model receives these numerical representations as input, including token IDs and attention masks.

Accuracy, precision, recall, F1-score, ROC-AUC score are examples of common evaluation metrices that defines how well the model can identify news articles. When the model performs well enough, it can be used for in real-world applications like social media platforms and news websites. Overview of the Model is shown in figure 3.1

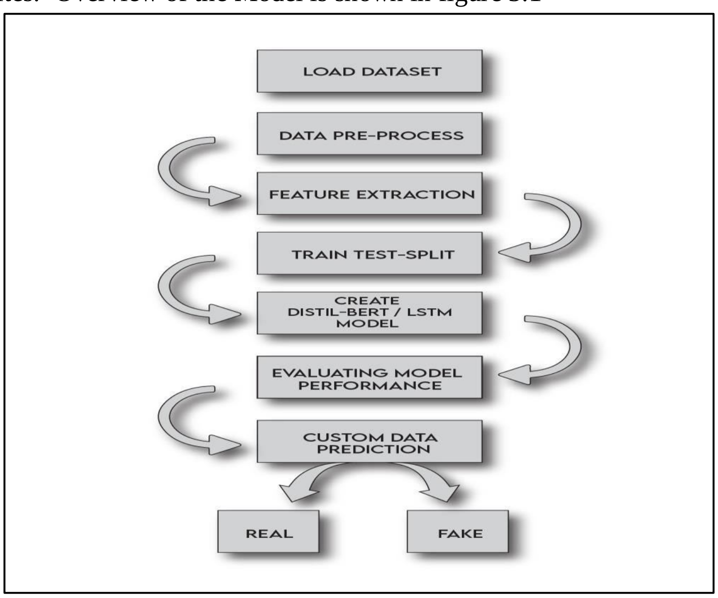


Fig 3.1: Overview of The Model

B. Data Pre-processing

Before evaluating the overall function of the model in addition analyzing the manual hyperparameters, the raw data must be processed. There are several Data preprocessing steps including stop-words removal, tokenization, stemming, lemmatization etc. Data preprocessing step are shown in figure 3.2, 3.3, 3.4 and 3.5

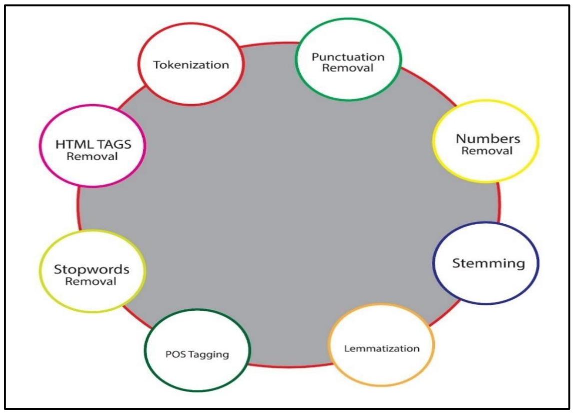


Fig 3.2: Data Preprocessing Steps.

Below we discuss preprocessing steps on "headline" and "content" columns.

1) Stop Word Removal: Stop words are that words used as a common word and don't have any context. These words have no meaningful information; therefore, we remove these words from both headline and content. On the other side, Bengali languages have moreover same common word. These are -['অতএব', 'অথচ', 'অথবা', 'অনুযায়ী', 'অননক', 'অনননক', 'অনননকই', 'অন্তত', 'অনয', 'অবধি', 'অবশ্য', 'অথাতথ', 'আই', 'আগামী', 'আনগ', 'আনগই', 'আনে', 'আজ', 'আদ্যভানগ', 'আপনার', 'আপধন', 'আবার', 'আমরা', 'আমানক', 'আমানদ্র', 'আমার', 'একটি', 'একবার', 'এনক', 'এক্', 'এখন', 'এখনও', 'এখানন', 'এখাননই', 'ত্রিা', 'ত্রিাই', 'এটি', 'ধেল'] etc.

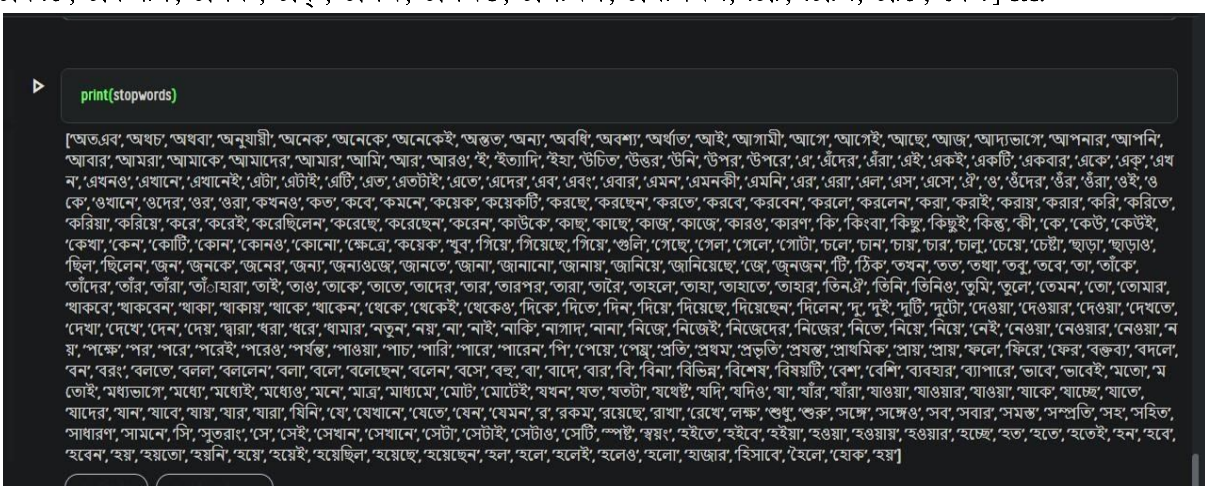


Fig 3.3: List of Stop Words Form Own Data

2) Punctuation Removal: Punctuation in natural language provides the grammatical context of the sentence. Punctuation like a comma that cannot add much value to the understanding of the sentence.

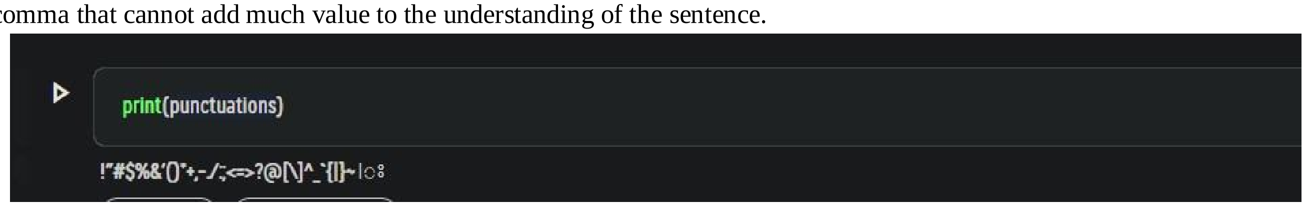


Fig 3.4: List of Punctuation Form Own Data

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue IV Apr 2026- Available at www.ijraset.com

3) Lemmatization and Stemming: Lemmatization and steaming both process the word to reach into its root. Stemming remove prefixes and suffixes from the word to reach the root. Therefore, sometimes the stemming word has no meaning and no actual word. On the other hand, lemmatization build the actual word.

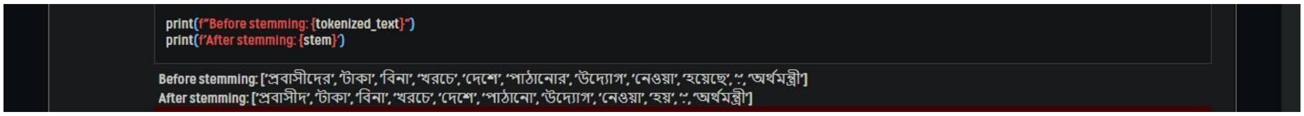


Fig 3.5: Stemming Own Data

4) Tokenization: Tokenization is a method of dividing a piece of text into smaller units. Here, the tags can be words, letters, or sub-worlds. Therefore, tokenization is roughly divided into 3 types of tokenization of words, letters, and sub words. The most common way to create tokens is based on space.

Assuming the space as a boundary, the tokenization result of the sentence is 4 tokens.

Bengali sentence: " আিম বাাংলায গান গাই "

After tokenization it shows: " আিম ", "বাাংলায ", "গান", "গাই"

C. Duplicate And Missing Data

The data may contain duplicate observations. We have to analyze the data set to make a decision on whether to drop duplicates or not. In our data set, there is a chance of two observations. Duplicates occur due to some data entry errors. These types of duplicates must be dropped. Similarly, missing or null data were also dropped.

Top frequent words: Using python libraries, we found top 30 words from real news dataset.

Therefore,we can see that real news have much "ও","করা","এবাং" etc words in addition fake news have much " করনত", "ধননয", "এক" etc.

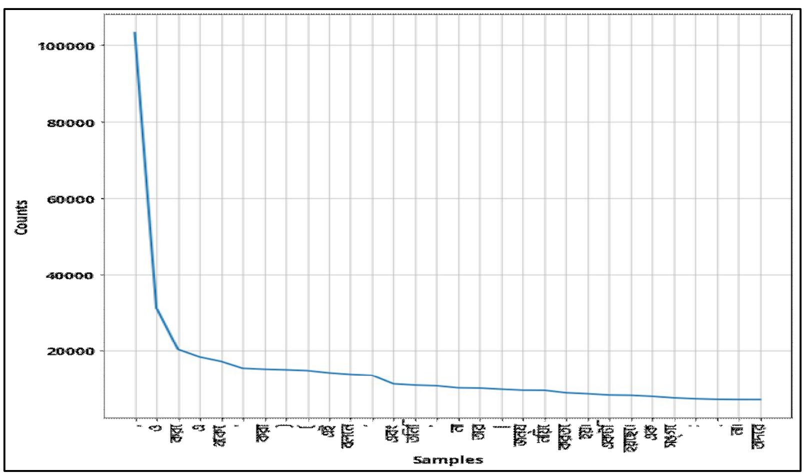


Fig 3.6 : Top word in true bangla news

D. Train-Test Split

The train-test split is a technique for evaluating the performance**.** The procedure involves taking a dataset and dividing it into two subsets. The first subset is used to fit the model in addition is referred to as the training dataset. The second subset is not used to train the model; instead, the input element of the dataset is provided to the model, then predictions are made and compared to the expected values. This second dataset is referred to as the test dataset. The scikit-learn Python machine learning library provides an implementation of the train-test split evaluation procedure via the train_test_split() function. The function takes a loaded dataset as input and returns the dataset split into two subsets. Train, test = train_test_split(dataset, ...). Ideally, split original dataset into input (X) and output (y) columns, then call the function passing both arrays in addition have them split appropriately into train and test subsets. The size of the split can be specified via the "test_size" argument that takes a number of rows (integer) or a percentage (float) of the size of the dataset between 0 and 1..Here the spilt percentage is Train: 70%, Test: 30% .

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue IV Apr 2026- Available at www.ijraset.com

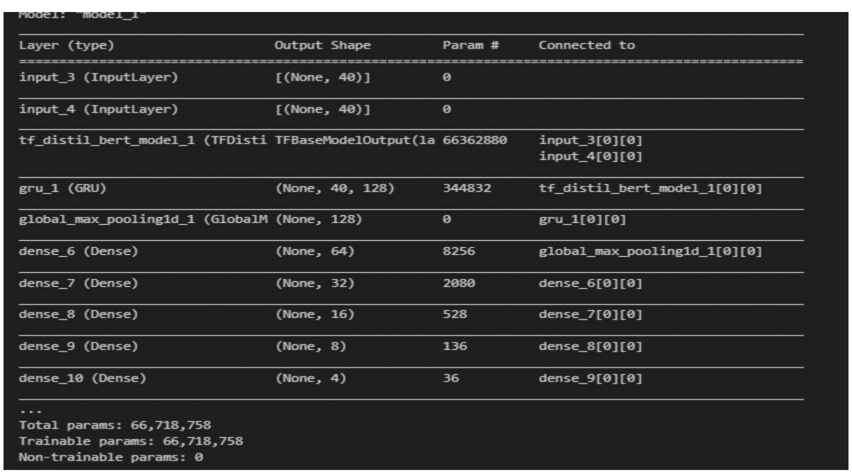
IV. IMPLEMENTATION

A. DistilBERT Model Summary

In this model architecture, the input sequences undergo contextualized embedding extraction through a DistilBERT model, capturing the nuanced meanings of each word/token within the context of the entire sequence. Subsequently, the contextualized embeddings are processed by a Gated Recurrent Unit (GRU) layer with 128 units, facilitating sequential data modeling by capturing dependencies among tokens. To mitigate overfitting and enhance generalization, a dropout layer with a dropout rate of 0.3 is applied after the GRU layer, randomly nullifying a portion of activations during training to introduce regularization. Finally, a dense layer with SoftMax activation produces probabilities for each class in the classification task, enabling the model to assign likelihood scores to different classes based on the input sequence. This architecture, commonly utilized in natural language processing tasks such as text classification or sentiment analysis, effectively combines advanced embedding techniques with recurrent neural networks and regularization mechanisms to achieve accurate classification results. Figure 4.1 shows the Summary of The DistilBERT Model.

Here's a summary of the model architecture we used for detecting fake news:

1) Input Layers: Two input layers named input_1 and input_2 for input sequences and masks respectively, both with a shape of (None, 40), where None indicates a variable batch size and 40 represents the max_len.
2) TF DistilBERT Layer: This layer represents a pre-trained DistilBERT model. It takes the two input sequences in addition produces contextual embeddings for each input sequence.
3) GRU Layer: A Gated Recurrent Unit (GRU) layer with 128 units. This layer processes the sequence of embeddings and returns sequences with hidden states.
4) Global Max Pooling1D: Global max pooling1d: Global Max Pooling1D layer collapses the sequence dimension, resulting in a tensor with shape (None, 128).
5) Dense Layers: Dense layer with 64 units and tanh activation function. These dense layers are fully connected layers. Each layer receives the output of the previous layer as input dense_1: Dense layer with 64 units. dense_2: Dense layer with 32units. dense_3: Dense layer with 16 units. dense_4: Dense layer with 8 units. dense_5: Dense layer with 4 units.
6) Dropout Layer: Dropout layer with a dropout rate of 0.3, applied after the last dense layer.
7) The output layer: The output layer is a dense layer with 2 units in addition SoftMax activation, indicating a classification task with two classes.


Model: "model_1"

Layer (type)	Output Shape	Param #	Connected to
input_3 (InputLayer)	[(None, 40)]	0	
input_4 (InputLayer)	[(None, 40)]	0	
tf_distil_bert_model_1 (TFDisti	TFBaseModelOutput(la	66362880	input_3[0][0] input_4[0][0]
gru_1 (GRU)	(None, 40, 128)	344832	tf_distil_bert_model_1[0][0]
global_max_pooling1d_1 (GlobalM	(None, 128)	0	gru_1[0][0]
dense_6 (Dense)	(None, 64)	8256	global_max_pooling1d_1[0][0]
dense_7 (Dense)	(None, 32)	2080	dense_6[0][0]
dense_8 (Dense)	(None, 16)	528	dense_7[0][0]
dense_9 (Dense)	(None, 8)	136	dense_8[0][0]
dense_10 (Dense)	(None, 4)	36	dense_9[0][0]

...
Total params: 66,718,758
Trainable params: 66,718,758
Non-trainable params: 0

Fig 4.1: Summary of The DistilBERT Model

B. LSTM Model Summary

The LSTM (Long Short-Term Memory) layer within the sequential model architecture is a critical component for processing sequential data, particularly in tasks like natural language processing or time series analysis. With 128 units, this LSTM layer is capable of learning and retaining long-term dependencies within the input sequences. It excels in capturing intricate patterns and dependencies across various time steps, making it highly effective in understanding the context and semantics of the input data. The Dropout layer, following the LSTM, serves to prevent overfitting by randomly dropping a fraction of the connections between neurons during training, thereby encouraging the network to learn more robust and generalizable representations. It enhances the model's ability to generalize to unseen data by reducing reliance on specific features or patterns present only in the training data.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue IV Apr 2026- Available at www.ijraset.com

Finally, the Dense layer with a single neuron outputs the prediction or classification result. With only one neuron, it is suitable for binary classification tasks, where the model predicts a single output indicating the presence or absence of a particular feature or class. The parameters of this layer are fine-tuned through backpropagation during training, enabling the model to make accurate predictions based on the learned representations from the preceding layers.

1) Embedding layer: Output shape (None, 1400, 100) **-** This layer converts integer indices representing words into dense vectors of fixed size. The input shape (None, 1400) indicates that the model accepts sequences of integers, where each integer represents a word index, and each sequence has a length of 1400. The output shape (None, 1400, 100) indicates that the layer outputs dense vectors of size 100 for each word in the input sequence.
2) LSTM layer: Output shape (None, 128) **-** This layer is a Long Short-Term Memory (LSTM) layer with 128 units. It processes the input sequences and produces output sequences. The output shape (None, 128) indicates that the layer outputs sequences of vectors of size 128.
3) Dropout layer: Output shape (None, 128) **-** This layer randomly sets a fraction of input units to 0 at each update during training time to prevent overfitting. The output shape remains unchanged (None, 128).
4) Dense layer: Output shape **(None, 1) -** This layer is a fully connected dense layer with 1 unit and sigmoid activation function, used for binary classification. The output shape **(None, 1)** indicates that the layer outputs a single scalar value for each input sequence.

V. EVALUATION MATRICES

A. Software and Hardware Configuration

Python 3.x installation in addition several Python libraries, including TensorFlow, Keras, NumPy, Pandas, WordCloud, Matplotlib, Stop words, PIL, TensorFlow, BanglaNLTK, Transformers. The code uses a recurrent neural network and a pre – trained DistilBERT model which works with a GRU layer and several multiple layers to detect fake news. The hardware requirements include a CPU or GPU with sufficient processing power to train the model, at least 8 GB of RAM, In addition sufficient storage space for the dataset and model checkpoints. All hardware was made available by Kaggle Notebook, a free python environment that runs entirely in the cloud.

B. Accuracy Graph

The accuracy graph demonstrates how a model's accuracy varies during training epochs. It plots two lines. One is the training accuracy and another one is the validation accuracy. Accuracy graph of DistilBERT and LSTM model are shown in figure 5.1 and 5.2 respectively.

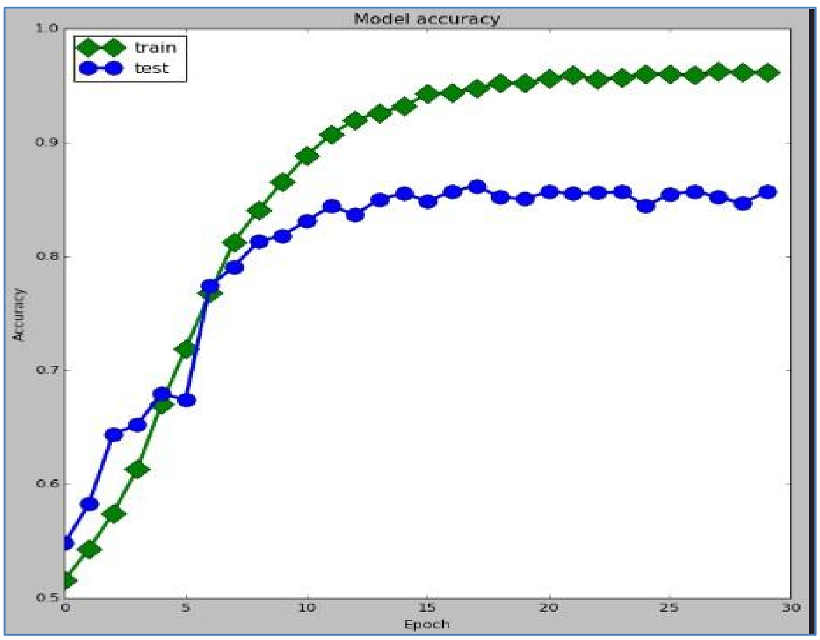


Fig 5.1: Accuracy Graph of DistilBERT Model.

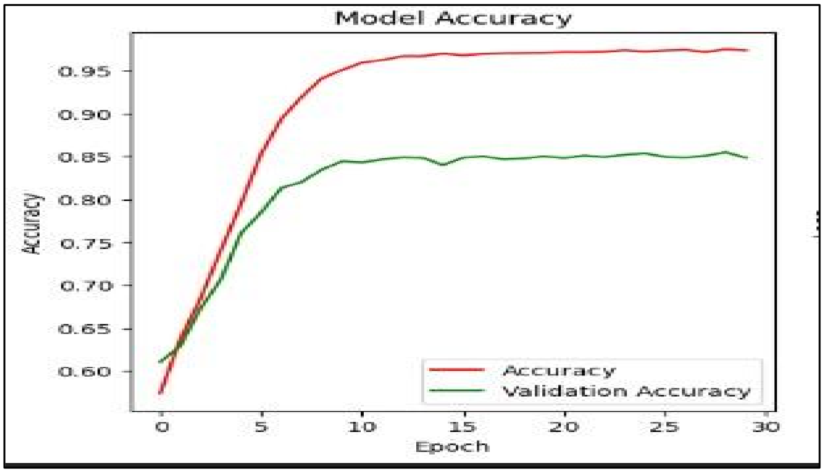


Fig 6.2: Accuracy Graph of LSTM Model.

C. *Loss Graph*

The loss graph demonstrates how a model's loss varies during training epochs. It plots two lines. One is the training loss and another one is the validation loss. Model loss graph of DistilBERT and LSTM model are shown in figure 6.3 and 6.4 respectively

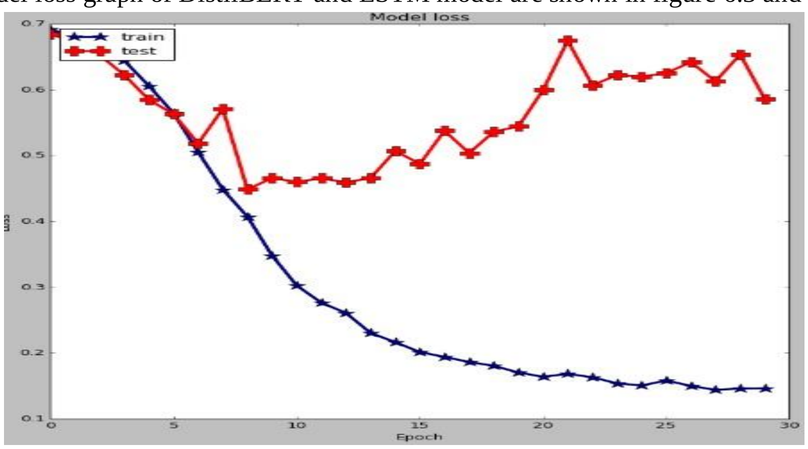


Fig 5.3: Loss Graph of DistilBERT model.

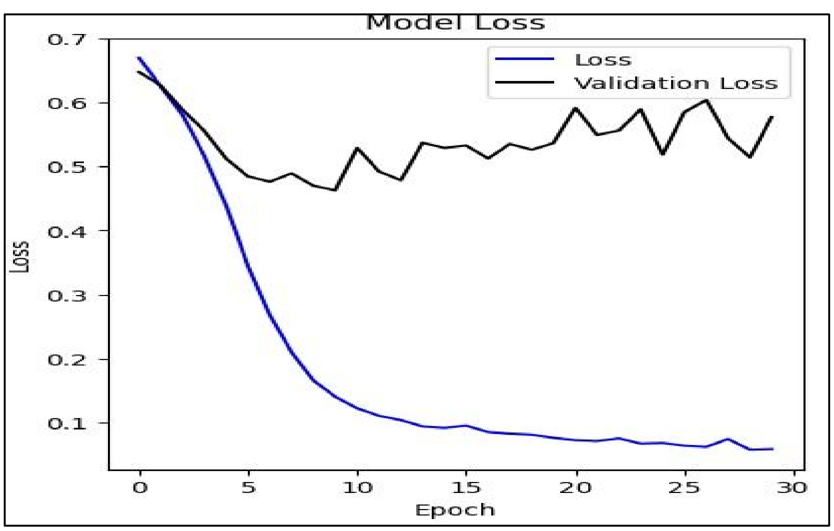


Fig 6.4: Loss Graph of LSTM model.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue IV Apr 2026- Available at www.ijraset.com

D. *Evaluation Metrices*

1) *(ROC – AUC) Curve*

ROC Curve (Receiver Operating Characteristic Curve) is a graph that shows the performance of a classification model at all classification thresholds. The AUC (Area Under the Curve) represents the area under the ROC curve. It measures the overall performance of the binary classification model. This curve plots two parameters: True Positive Rate and False Positive Rate. ROC-AUC curve of DistilBERT and LSTM model are shown in figure 6.5 and 6.6

True Positive Rate (TPR) is defined as follows: (TPR) = TP / (TP+FN)…………..eq^{n} (1)

False Positive Rate (FPR) is defined as follows:

(FPR) = FP / (TN+FP)…………...eq^{n} (2)

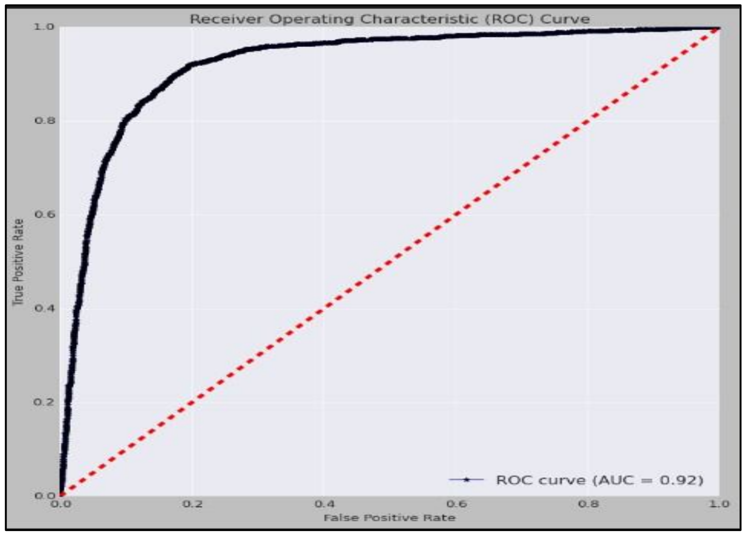


Fig 5.5: (ROC-AUC) Curve of DistilBERT Model.

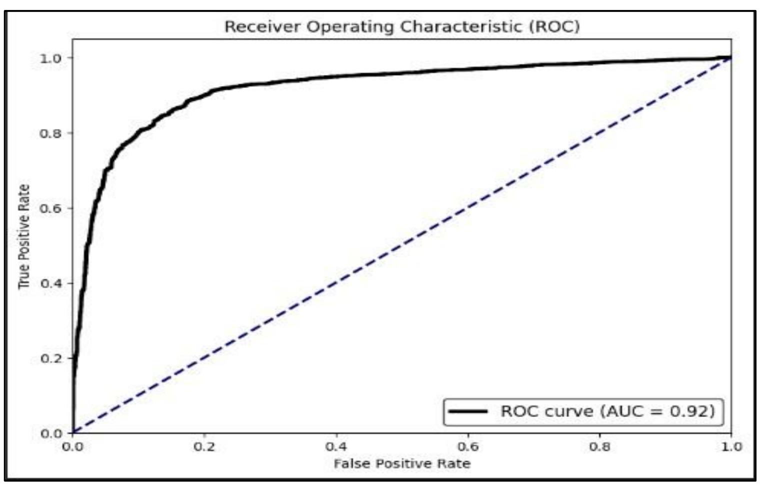


Fig 5.6: (ROC-AUC) Curve of LSTM Model

2) *Confusion Matrix*

Confusion matrix measures the performance of a machine learning classification algorithm. It is a table with 4 different combinations of predicted and actual values.

- True Positives (TP): Occurs when the model accurately predicts a positive data.
- True Negatives (TN): Occurs when the model accurately predicts a negative data.
- True Positives (FP): Occurs when the model predicts a positive data incorrectly.
- False Negatives (FN): Occurs when the model mis predict a negative data incorrectly.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue IV Apr 2026- Available at www.ijraset.com

Confusion Matrix of DistilBERT and LSTM model are shown in figure 5.7 ..

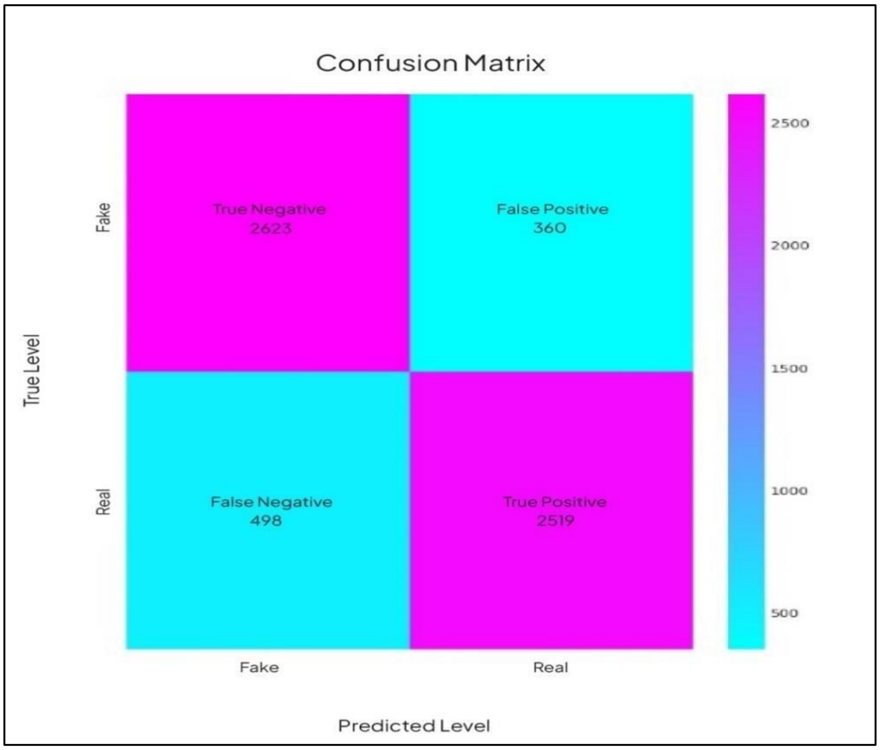


Fig 5.7: Confusion Matrix of DistilBERT.

3) *Classification Report*

It is one of the performance evaluation metrics of a classification-based machine learning model. It shows model's precision, recall, F1 score and support. It gives a better understanding of the overall performance of the trained model. To understand the classification report of a machine learning model, we need to know all of the metrics given below:

- Precision: Precision is defined as the ratio of true positives to the sum of true and false positives.
- Recall: Recall is defined as the ratio of true positives to the sum of true positives and false negatives.
- F1 Score: The F1 score is the weighted harmonic mean of precision and recall. The closer the value of the F1 score is to 1.0, the better the expected performance of the model is ideal.
- Support: Support is the number of actual occurrences of the class in the dataset. It doesn't vary between models; it just diagnoses the performance evaluation process.

Classification Report of DistilBERT and LSTM model are shown in figure 5.8 and 5.9

Classification Report

	precision	recall	f1-score	support
0	0.84	0.88	0.86	2983
1	0.87	0.83	0.85	3017
accuracy			0.86	6000
macro avg	0.86	0.86	0.86	6000
weighted avg	0.86	0.86	0.86	6000

Fig 5.8: Classification Report of DistilBERT.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue IV Apr 2026- Available at www.ijraset.com

	precision	recall	f1-score	support
Fake	0.83	0.88	0.85	2983
Real	0.87	0.82	0.84	3017
accuracy			0.85	6000
macro avg	0.85	0.85	0.85	6000
weighted avg	0.85	0.85	0.85	6000

Fig 5.9: Classification Report of LSTM.

VI. RESULTS

A. Custom Data Prediction

The Bangla news article input is tokenized, transformed into input features that are compatible with the DistilBERT model, in addition a prediction is then made using the model. The pre-trained DistilBERT model is used to predict the label (fake or real) of the input news article. It uses the tokenizer to encrypt the text, runs it through the model to make predictions, in addition outputs whether the news is considered to be fake or real. The news is categorized as true if the predicted label is 1, and as fake if it is 0. The prediction is made using the tokenized input and attention mask. The predicted label is determined by taking the index of the maximum value in the prediction array. If the index corresponds to label 0, it indicates that the news is classified as fake. Otherwise, if the index corresponds to label 1, it indicates that the news is classified as real. Visualizing Result of our worshown in figure 6.1 **.**

```
x = ['নড়াইল থেকে ফিরে: গত ১০ অক্টোবর মধুমতী নদীর ওপর দেশের প্রথম ছয় লেনের মধুমতী সেতু উদ্বোধন করেন প্রধানমন্ত্রী শেখ হাসিনা। লোহাগড়া উপজেলার মধুমতী নদীর অংশে এ সেতু নির্মাণে কমেছে যশোর-নড়াইলের দূরত্ব']

test_input = distil_bert_tokenizer.batch_encode_plus(x,add_special_tokens=True, max_length=40,pad_to_max_length=True,
                                                     truncation=True)

test_input_ids = np.array(test_input['input_ids'])
test_input_mask = np.array(test_input['attention_mask'])

prediction = model.predict([test_input_ids, test_input_mask])
test_pred_label = np.argmax(prediction, axis=1)[0]

if test_pred_label == 0:
    print("News is Fake")
else:
    print("News is Real")

News is Real
```

Fig 6.1: Visualizing Result

B. Comparative Analysis

Table 6.1 and 6.2 are shown the result of DistilBERT and LSTM model.

Table 6.1: Result of DistilBERT Model

DistilBERT				
	Precision	Recall	f1-score	support
0 (Fake)	84%	88%	86%	2983
1 (Real)	87%	83%	85%	3017
accuracy			86%	6000
macro avg	86%	86%	86%	6000
weighted avg	86%	86%	86%	6000

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue IV Apr 2026- Available at www.ijraset.com

Table 6.2: Result of LSTM Model

LSTM				
	Precision	recall	f1-score	support
0 (Fake)	83%	88%	85%	2983
1 (Real)	87%	82%	84%	3017
accuracy			85%	6000
macro avg	85%	85%	85%	6000
weighted avg	85%	85%	85%	6000

DistilBERT achieves an accuracy of 86%, which is slightly higher than LSTM's accuracy of 85%. Accuracy measures the overall correctness of the model's predictions. F1-score is the harmonic mean of precision and recall, providing a balance between them. DistilBERT achieves F1-scores of 86% and 85% for classes Fake and Real respectively, while LSTM achieves F1-scores of 85% for both classes. Higher F1-scores indicate better performance in terms of both precision and recall. DistilBERT shows balanced precision and recall values for both classes (Fake and Real), indicating its ability to correctly classify instances of both classes without bias towards one class over the other. DistilBERT consistently outperforms LSTM across multiple evaluation metrics (precision, recall, F1-score, and accuracy), suggesting its overall superiority in this specific classification task.

Overall, while the performance difference between DistilBERT and LSTM is not substantial, DistilBERT exhibits slightly better performance across various metrics, making it the preferable choice in this scenario.

VII. FUTURE WORK AND CONCLUSION

A. Limitations

Detecting fake news in any language, including Bangla, using models like DistilBERT, can be effective however comes with several limitations:

1) Insufficient Training Data: Models for detecting fake news that are unique to Bangladesh may not have enough training data. Because of this, models such as DistilBERT may find it more difficult to identify the complexities and patterns specific to Bangla fake news.
2) Language Complexity: Bangla languages have its own complexities including morphology, syntax, and semantics. Pre-trained models like DistilBERT may not capture all these nuances effectively, especially if they were not adjusted specifically for Bangla.
3) Adversarial Attacks: DistilBERT is at risk of adversarial attacks, which occur when adversaries intentionally change input data in order to confuse the model, just like other machine learning models. This may make the model less accurate for detecting false news.

B. Future Works

The future work of Bangla fake news detection using DistilBERT include improving accuracy, fine-tuning its parameter, handling more complex data, better handling of long-term dependencies, contextual understanding and developing better evaluation metrics.

C. Conclusion

DistilBERT has great potential for identifying fake news in Bangla, as do other language models of a similar nature. Its capacity for textual data analysis in addition comprehension can help stop the spread of false information in places where Bangla is the primary language. Although it has great potential, there are a number of obstacles and restrictions in using DistilBERT for Bangla fake news identification. These include problems regarding adversarial attacks, linguistic complexity, cultural background, in addition dataset accessibility. In order to address these obstacles and improve the accuracy of Bangla fake news identification using DistilBERT, more study and development are obviously needed. Future research should concentrate on topics like contextual understanding, adversarial robustness, multimodal techniques, cross-lingual transfer learning, user-centric solutions, real-time detection systems, dataset expansion, and Bangla tuning. DistilBERT-based fake news detection systems can have a big positive impact on society if they are successfully implemented in Bangla. It can support people in making wise choices about the information they consume, lessen the negative effects of false information, in addition preserve the accuracy of public controversy.

REFERENCES

[1] Pall Chowdhury, P., Eumi, E. M., Sarkar, O., & Ahamed, M. F. (2021, December 4). Bangla News Classification Using GloVe Vectorization, LSTM, and CNN. Lecture Notes on Data Engineering and Communications Technologies. [Online]. Available: https://doi.org/10.1007/978-981-16-6636-0_54.

[2] Md. Elias Hossain, Md. Nadim Kaysar, Abu Zahid Md Jalal Uddin Joy, SM Mizanur Rahman, Md. Wahidur Rahman. "A Study towards Bangla Fake News Detection Using Machine Learning and Deep Learning" In book: Sentimental Analysis and Deep Learning, Proceedings of ICSADL 2021 (pp.79-95),Oct 2022.

[3] Hossain, M.Z., et al., "A dataset for detecting fake news in bangla. arXiv preprint" arXiv:2004.08789, 2020.

[4] Md Zobaer Hossain, Md Ashraful Rahman, Md Saiful Islam, Sudipta Kar, "A Dataset for Detecting Fake News in Bangla" In:12th Conference on Language Resources and Evaluation (LREC 2020).

[5] Grave, E., Bojanowski, P., Gupta, P., Joulin, A., and Mikolov, T. (2018), "Learning word vectors for 157 languages" In : International Conference on Language Resources and Evaluation (LREC 2018).

[6] Risul Islam Rasel, Anower Hossen Zihad, Nasrin Sultana1, Mohammed Moshiul Hoque, "Bangla Fake News Detection using Machine Learning, Deep Learning and Transformer Models" In: 25th International Conference on Computer and Information Technology (ICCIT) 17-19 December, Cox's Bazar, Bangladesh

[7] M. G. Hussain, M. Rashidul Hasan, M. Rahman, J. Protim and S. Al Hasan, "Detection of Bangla Fake News using MNB and SVM Classifier," 2020 International Conference on Computing, Electronics & Communications Engineering (iCCECE), Southend, UK, 2020, pp. 81-85, doi: 10.1109/iCCECE49321.2020.9231167.

[8] Mobs beat five dead for kidnapping, daily star. [Online]. Available: https://www.thedailystar.net/frontpage/news/mobsbeat-2-dead-kidnapping-1774471 (2019, July)

[9] Tasnuba Sraboni, Md. Rifat Uddin , Fahim Shahriar , Ruhit Ahmed Rizon ,Shakib Ibna Shameem Polock, "FakeDetect: Bangla Fake News Detection Model based on Different Machine Learning Classifiers Computer Science and Engineering Department of Computer Science and Engineering Brac University June 2021.

[10] Iftikhar Ahmad , Muhammad Yousaf,Suhail Yousaf , and Muhammad Ovais Ahmad , "Research Article Fake News Detection Using Machine Learning Ensemble Methods" . [Online].Available:https://doi.org/10.1155/2020/8885861(2020,October).

[11] Sadik Al Jarif, "Bangla And English Language Fake News Detection Using Deep Learning" Department of CSE Manarat International University, Oct 2020

[12] Md. Shahriar Rahman,Faisal Bin Ashraf , "An Efficient Deep Learning Technique for Bangla Fake News Detection" In: 25th International Conference on Computer and Information Technology (ICCIT), 17-19 December, 2022, Cox's Bazar, Bangladesh .

[13] Md. Muzakker Hossain, Zahin Awosaf, Md. Salman Hossan Prottoy, Abu Saleh Muhammod Alvy,Md. Kishor Morol , "Approaches for Improving the Performance of Fake News Detection in Bangla: Imbalance Handling and Model Stacking" American International University-Bangladesh, Dhaka, Bangladesh. 22 Mar 2022.

[14] Md. Zahin Hossain George,Naimul ,Md. Rafiuzzaman Bhuiyan ,Abu Kaisar Mohammad Masum, "Bangla Fake News Detection Based On Multichannel Combined CNN-LSTM" In: 12th International Conference on Computing Communication and Networking Technologies (ICCCNT).July 2021.

[15] S. Rohman, J. Ferdous, S. M. R. Ullah and M. A. Rahman, "IBFND: An Improved Dataset for Bangla Fake News Detection and Comparative Analysis of Performance of Baseline Models," 2023 International Conference on Next-Generation Computing, IoT and Machine Learning (NCIM), Gazipur, Bangladesh, 2023, pp. 1-6, doi: 10.1109/NCIM59001.2023.10212799.

[16] "Price hike rumour: People flock shops for salt." https://rb.gy/ehogkn, 2019

[17] N. M. Jakilim, S. Mahamudul Hasan and E. Hassan, "A Benchmark of Machine Learning and Deep Learning Algorithms for Detecting Fake News in Bangla Language," 2022 4th International Conference on Sustainable Technologies for Industry 4.0 (STI), Dhaka, Bangladesh, 2022, pp. 1-6, doi: 10.1109/STI56238.2022.10103235.

[18] Q. A. R. Adib, M. H. K. Mehedi, M. S. Sakib, K. K. Patwary, M. S. Hossain and A. A. Rasel, "A Deep Hybrid Learning Approach to Detect Bangla Fake News," 2021 5th International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT), Ankara, Turkey, 2021, pp. 442-447, doi: 10.1109/ISMSIT52890.2021.9604712.

[19] MD. RAFI-UR-RASHID, MAHIM MAHBUB, and MUHAMMAD ABDULLAH ADNAN, "Breaking the Curse of Class Imbalance: Bangla Text Classification" In: ACM Trans. Asian Low-Resour. Lang. Inf. Process., Vol. 21, No. 5, Article 97. Publication date: April 2022

[20] Mahammed Kamruzzaman, Md. Minul Islam Shovon and Gene Louis Kim , "BanMANI: A Dataset to Identify Manipulated Social Media News in Bangla" In:arXiv:2311.02570v1 [cs.CL] 5 Nov 2023

[21] Shafayat Bin Shabbir Mugdha, Sayeda Muntaha Ferdous, and Ahmed Fahmin. "Evaluating Machine Learning Algorithms For Bengali Fake News Detection". In: 2020 23rd International Conference on Computer and Information Technology (ICCIT). 2020, pp. 1–6. doi: 10 . 1109 / ICCIT51783 . 2020 . 9392662.

[22] Farzana Islam et al. "Bengali Fake News Detection". In: 2020 IEEE 10th International Conference on Intelligent Systems (IS). 2020, pp. 281–287. doi: 10. 1109/IS48319.2020.9199931.

[23] BanFakeNews. Available online:https://github.com/Rowan1224/FakeNews?tab=readme-ovfile (accessed on 17 February 2023)

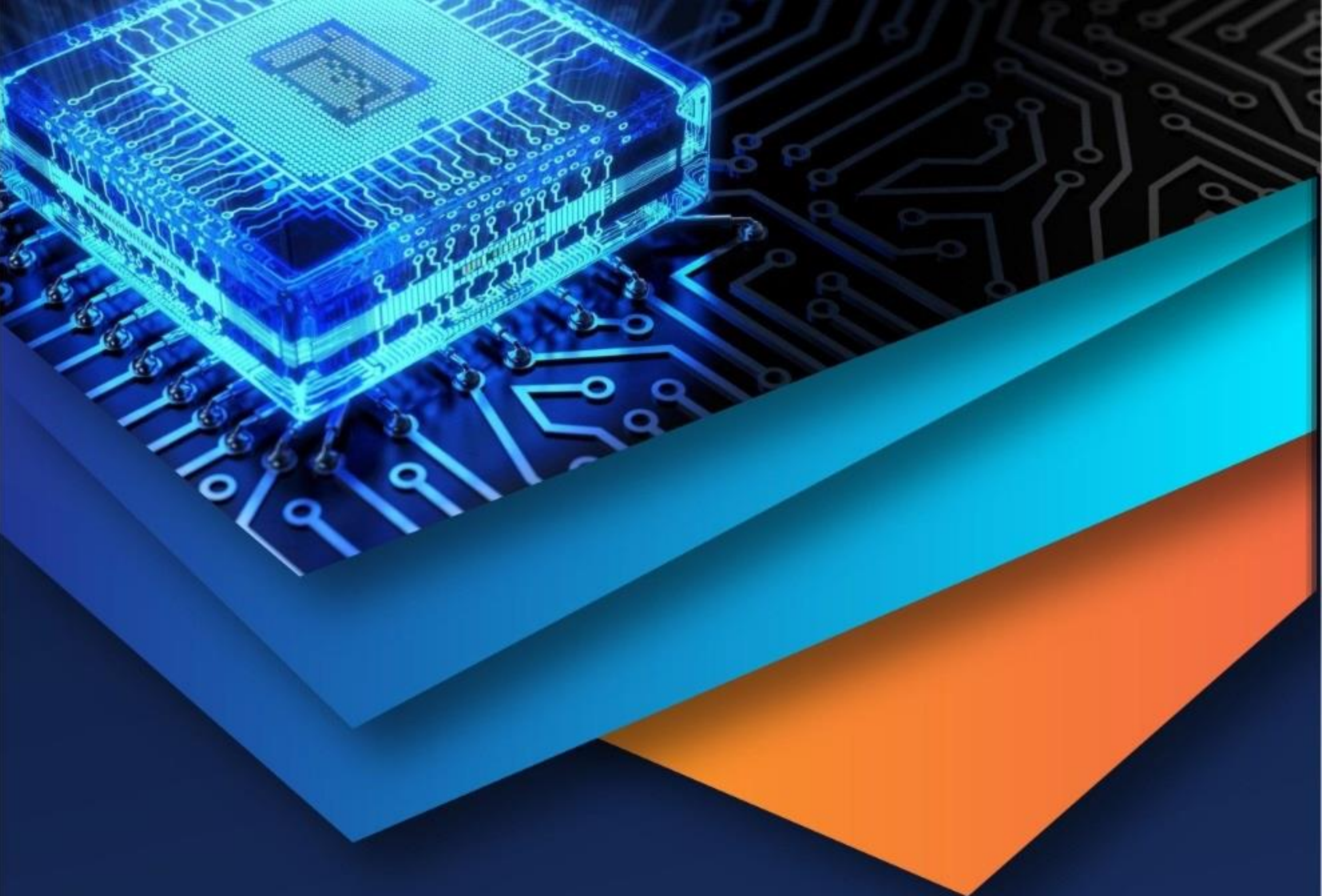
[24] BanFakeNews. Available online: https://www.kaggle.com/datasets/cryptexcode/banfakenews (accessed on 18 February 2023)

[25] BanMANI. Available online: https://www.kaggle.com/datasets/mozaman36/banmani (accessed on 20 February 2023)

[26] Bangla fake new detection. Available online: https://www.kaggle.com/datasets/mozaman36/banmani (accessed on 20 February 2023)

[27] Fake News Detection . Available online: https://www.kaggle.com/code/therealsampat/fake- news-detection (accessed on 23 February 2023)

[28] Final_bn_data. Available online: https://www.kaggle.com/datasets/pikuldasjoy/fake-newsdataset/data?select=final_bn_data.csv



10.22214/IJRASET

INDEX COPERNICUS 45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY