INTERNATIONAL JOURNAL
FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Fake News Classifier

Amaram Divija[1], Kurkuri Smitha Kiran[2], Telukuntla Priyanka[3], Dr. Mantha Shailaja[4]

[1, 2, 3]*Student, Department of Electronics and Computer Engineering, Sreenidhi Institute of Science and Technology, Telangana, India*

[4]*Associate Professor, Department of Electronics and Computer Engineering, Sreenidhi Institute of Science and Technology, Telangana, India*

Abstract: Fake news is incorrect information that is spread through a social network to harm individuals, authorities or organizations. The spread of fake news poses great challenges to society. Fake news is difficult to detect, but it is easy to spread and have widespread effects. Automated analysis of the reliability of articles is the subject of ongoing research. To address this issue, we offer a model that detects fake information and communications using deep learning and natural language processing. This paper presents a fake news detection model based on LSTM (Long Short-Term Memory) and Bi-LSTM (Bidirectional Long Short-Term Memory). In the first place, we want to present a dataset containing both fake news and genuine news, and perform various tests to sort out a fake news detector. The model was prepared and assessed utilizing a fake news dataset got from Kaggle.
Keywords: deep learning, long short-term memory, natural language processing, fake news

I.   INTRODUCTION

The media plays an important role in the public dissemination of event information. Because of the rapid development of the Internet, information may be disseminated quickly via social media and websites. In recent years, social media has become increasingly crucial in everyone's lives. Fake news is mostly spread through social media. Politics, finance, education, democracy, and business are all at risk as a result of fake news. Although fake news is not a new issue, individuals today place a higher focus on social media, which contributes to the acceptance and dissemination of incorrect information. It is getting increasingly difficult to distinguish between accurate and misleading news, resulting in confusion and problems. Manually recognising fake news is tough; it is only achievable when the individual identifying the news has extensive expertise of the subject. It is now much easier to create and spread fake news thanks to recent developments in computer science. Unverified or fake news is disseminated over social media without concern for its veracity, and it reaches tens of thousands of people. For commercial and political advantage, fake news is frequently generated to fool readers. Fake news and hoaxes have the potential to affect online user's personal beliefs and decisions.

An important approach to distinguishing trusted users, extracting useful message features, and developing genuine information distribution systems is some useful research areas that need further investigation. There are many techniques for dealing with the issue of disinformation on social media. In the field of fake news detection, many supervised learning models have been applied to classification tasks. This challenge has recently been addressed using neural networks and deep learning techniques. Topic modelling can be an important part of fake news detection. This is because the detection algorithm can behave differently for messages from different domains

This paper presents multiple deep learning neural network model applications for detecting fake news. To detect fake news, the paper's main objective is to use only the document's title content, without taking into account the author's features or other characteristics. The model was trained and tested on Kaggle's fake news dataset. The basic goal is to process the data in text format and determine if it can be classified as fake news.

We used LSTM and Bi-LSTM algorithms to solve the fake news detection problem in this project. The algorithms performed well when tested with examples after being effectively trained on the given dataset. The accuracy of the algorithms is practically same. We also used Streamlit to construct a web interface that allows the user's input to be taken while the model does the background prediction task and returns the outcome to the user via the web interface.

We provided a system to discern between true and fraudulent news pieces in this paper. The objective of classification is to use long short-term memory. Data is pre-processed before using NLP. Stop word deletion, part-of-speech marking, and embedding are all done by NLTK. This neural network was built and updated using an RNN with an LSTM unit. Train the model using the specified dataset. The loss is also displayed.

The basic loss was calculated using the binary cross entropy loss. Adam Optimizer was used to mitigate this. Model results for epoch needs, confusion matrix results, and classification reports are satisfactory. The system has demonstrated up to 93% accuracy. The following is how the paper is structured: Section 2 conducts a literature review in the field of fake news identification. Section 3 provides a theoretical summary of deep-learning models. The proposed approach is described in Section 4. Section 5 discusses the dataset utilised in the project, as well as its pre-processing, model construction, and the web interface. Section 6 describes the findings. Section 7 contains the project's conclusion, limitations, and future scope.

II. LITERATURE SURVEY

This section outlines current research in the field of machine learning / deep learning to identify fake news.

A. Viera Maslei, Martin and Peter Butka: "Deep learning methods for Fake News detection"

"The study presented in this paper deals with the detection of fake news from the textual data using deep learning techniques. Our main idea was to train different types of neural network models using both entire texts from the articles and to use just the title text. The models were trained and evaluated on the Fake News dataset obtained from the Kaggle competition." [1]

B. Rohit Kumar Kaliyar: "Fake News Detection Using a Deep Neural Network"

"In this project we explored different Machine learning models like Naïve Bayes, K nearest neighbors, Decision tree, Random Forest and Deep Learning networks like Shallow Convolutional Neural Networks (CNN), Very Deep Convolutional Neural Network (VDCNN), Long Short-Term Memory Network (LSTM), Gated Recurrent Unit Network (GRU), Combination of Convolutional Neural Network with Long Short-Term Memory (CNN-LSTM) and Convolutional Neural Network with Gated Recurrent Unit (CNN-LSTM). We also explored the benefit of feature extraction, features like n-gram, TF-IDF features were extracted and used in our model. We also explored the effective of word embedding's and word2vec features in Deep Neural networks. We also explored the use of select best and chi2 for feature extraction in Machine learning model." [2]

C. Tejaswini Yesugade, Shrikant Kokate, Sarjana Patil, Ritik Varma, Sejal Pawar: "Fake News Detection using LSTM"

In this paper they carried out Machine Learning calculations and procedures like NLTK, LSTM for the discovery of fake news. [3]

D. Pritika Bahad, Preeti Saxena, Raj Kamal: "Fake News Detection using Bi-directional LSTM-Recurrent Neural Network"

"The paper presents a fake news detection model based on Bi-directional LSTM-recurrent neural network. Two publicly available unstructured news articles datasets are used to assess the performance of the model. The result shows the superiority in terms of accuracy of Bi-directional LSTM model over other methods namely CNN, vanilla RNN and unidirectional LSTM for fake news detection." [4]

E. Syed Ishfaq Manzoor, Dr Jimmy Singla, Nikita: "Fake News Detection Using Machine Learning approaches: A systematic Review"

"This paper reviews various Machine learning approaches in detection of fake and fabricated news. The limitation of such and approaches and improvisation by way of implementing deep learning is also reviewed." [5]

F. Supanya, Prabhas: "Detecting Fake News with Machine Learning Method"

"This work proposes the use of machine learning techniques to detect fake news. Three popular methods are used in the experiments: Naive Bayes, Neural Network and Support Vector Machine. The result show that Naive Bayes to detect Fake news has accuracy 96.08%. Two other more advance methods which are Neural Network and Support Vector Machine achieve the accuracy of 99.90%." [6]

III. DEEP LEARNING

A. Long Short-Term Memory

Long Short-Term Memory organizations, known as "LSTMs," are a sort of RNN that can learn long term dependencies. LSTMs are explicitly intended to avoid the problem of long-term dependency. It is fundamentally their default conduct to recollect data for significant stretches of time. The LSTM has a chain-like construction, and the repeating module has four brain network layers that collaborate in a one-of-a-kind style. The cell state, addressed by the level line going through the highest point of the image, is essential to LSTMs.

The LSTM can erase or add data to the cell state, which is constrained by structures known as gates. Gates are a strategy for alternatively permitting data to go through. They are built with a sigmoid brain net layer and pointwise multiplication. The sigmoid layer produces numbers going from zero to one, showing the amount of every part ought to be permitted through. A worth of zero specifies "let nothing through," while a worth of one designates "permit everything through!"
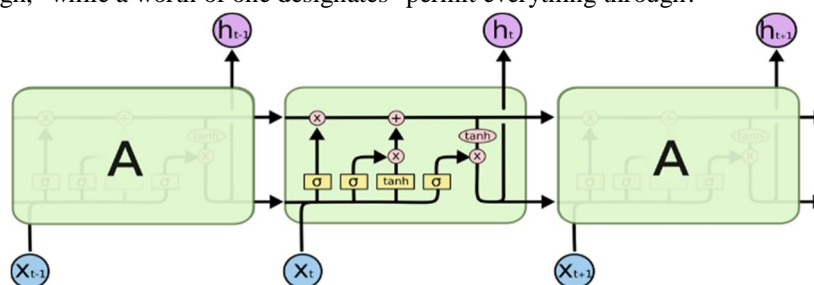

Fig. 1. Architecture of LSTM

B. Bidirectional Long Short-Term Memory

Bidirectional long short-term memory (bi-lstm) is the method of permitting any brain organization to store grouping data in the two directions, forward and in reverse (future to past) (past to future). Our feedback runs in two directions in bidirectional, recognizing a bi-lstm from a regular LSTM. We can make input stream in just a single course utilizing a regular LSTM, either in reverse or advances. Nonetheless, utilizing bi-directional, we can make the info stream in the two directions, saving both the future and the past.
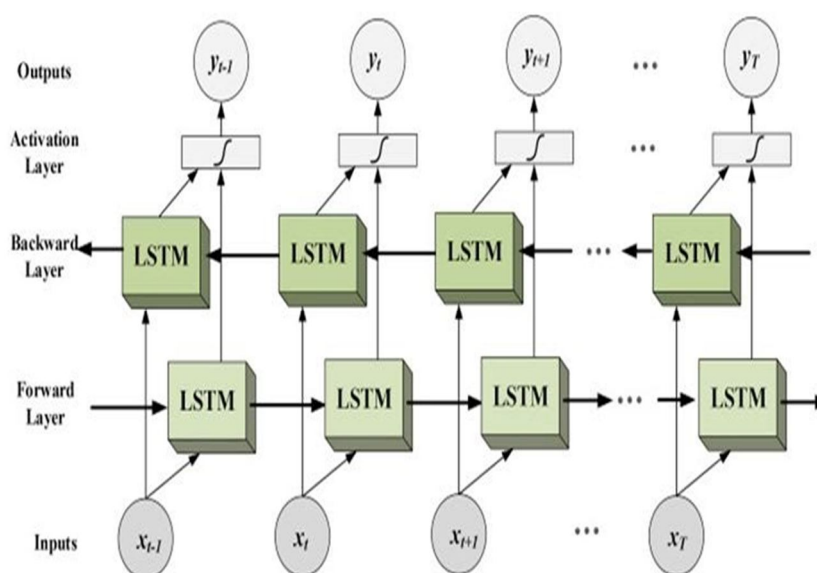

Fig. 2. Architecture of Bi-LSTM

IV. PROPOSED METHOD

For this project, we used the dataset from Kaggle. The dataset contains attributes such as id, title, author text, and label. We employed attributes such as title and label in this project. Our intention was to build a fake news classifier that detects fake news based on the given title. The data was first pre-processed.

The data was divided into an 80:20 ratio, with 80 percent used to train the model and 20 percent used to test the model. We used lstm and bi-lstm to generate two models. Using the training dataset, we trained the model.

We saved the model and used Streamlit to create a web page and take user input from a text box. In the background, the user input is provided to the model which makes predictions on the provided data and the output is displayed on the web page whether it is reliable news or not. As both the models have the same accuracy, we used the LSTM model to perform the prediction.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
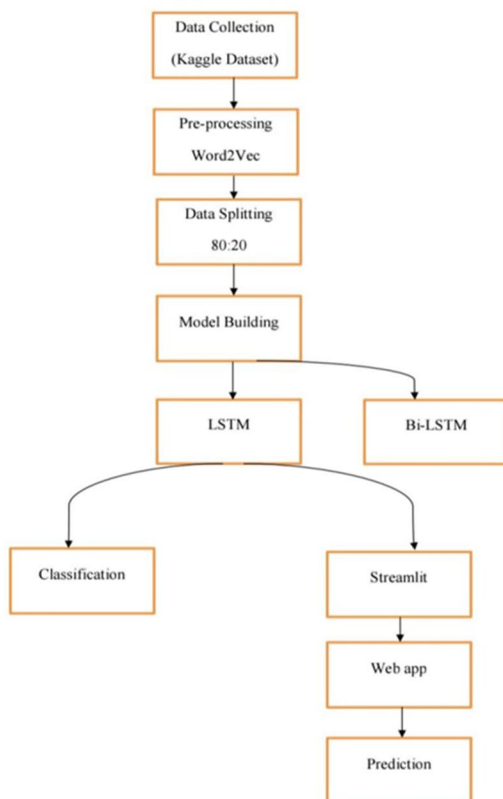*Volume 10 Issue VI June 2022- Available at www.ijraset.com*

Fig. 3. Proposed Method

V. FAKE NEWS DETECTOR

A. Data Collection
We used a Kaggle Fake News dataset for this research. The dataset contains 20800 articles. Each record was specified using the following attributes:
1) id: a unique identifier for a news story
2) title: the title of a news article
3) author: the news article's author
4) text: the article's text
5) label: a label that indicates whether an article is potentially reliable or not.

                 1: Untrustworthy
                 0: trustworthy

B. Data Pre-processing
The data pre-processing was done in the following steps:
1) Creating word tokens
2) Converting to lower case
3) Removing punctuation
4) Removing tokens that are not alphabetic.
5) Getting rid of stop words

The textual data is converted into an integer representation based on the vocabulary size using one hot representation. We then used pad sequences to make all the vectors the same length. We then applied an embedded layer on the integer encoded data.

C. *Model Building*

We separated the dataset into an 80/20 proportion, with 80% being utilized to prepare the model and 20 percent being utilized to test it. We created two models using LSTM and using bi-LSTM for title classification. The performance metric calculated in this project is accuracy. Several hyper parameters were modified throughout the training phase to improve results. The testing dataset is utilized to test models. To avoid data overfitting, the validation loss and accuracy are examined. To ensure learning variation, we fixed the parameters and ran the model numerous times. We discovered that the accuracy of both models is nearly identical. As per the results obtained from the models on title data, we were able to identify fake news based on the title data.

D. *Web Page*

We use the save() function to save the model and weights in the same file. This makes it possible to save a model's whole state in a single file. keras.models.load_model() can be used to re-instantiate saved models . The model given by load_model () is a ready-to-use built model. predict_classes(x) are used to generate prediction for the input sample. predict_classes() will return the index of the class having maximum value. Predict_classes() return a numpy array of class predictions. We designed a web interface that includes a text section where the user can enter the article title and a button for making predictions. The data is pre-processed and turned into an embedded vector before being delivered to the model when the user taps the predict button. The model will make predictions based on the input data, and the results will be presented to the user via a web interface. If the model's prediction value is zero, we can conclude that it is trustworthy. If the model's forecast is one, we will establish that it is untrustworthy.

VI.   RESULTS

A. *Accuracy Score*

Accuracy is characterized as the quantity of right expectations separated by the all-out number of estimates. The accuracy of our model is the number of right forecasts.

Model	Accuracy Score
LSTM	0.9319
Bi-LSTM	0.9308

Table-1: Results on title data

B. *Confusion Matrix*

A confusion matrix is a table that is commonly used to show how well a classification model performs on a set of test data with known true properties.

	Predicted: No	Predicted: Yes
Actual: No	1943	133
Actual: Yes	116	1465

Table-2: Confusion matrix for LSTM

	Predicted: No	Predicted: Yes
Actual: No	1918	158
Actual: Yes	95	1486

Table-3: Confusion matrix for Bi-LSTM

C. *Web Interface*

We used streamlit to create a web page and integrated the lstm model. The data from the web page is pre-processed and provided to the model, which makes the corresponding prediction and outcome is displayed to the user via the web page. In this case, we predicted the title corresponding to index 3 and index 5 in our dataset.

```
df['title'][3]
'15 Civilians Killed In Single US Airstrike Have Been Identified'

df['label'][3]
1
```

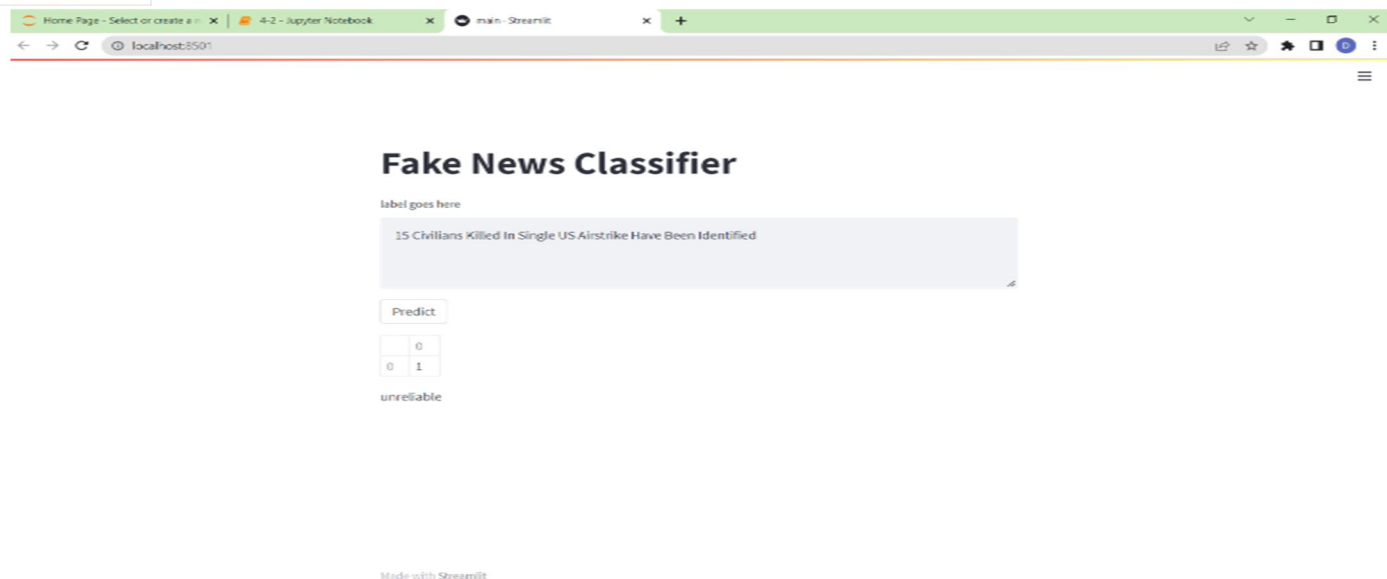Fig. 6. Title corresponding to index 3 in our dataset

Fig. 7. Prediction made on data corresponding to index 3

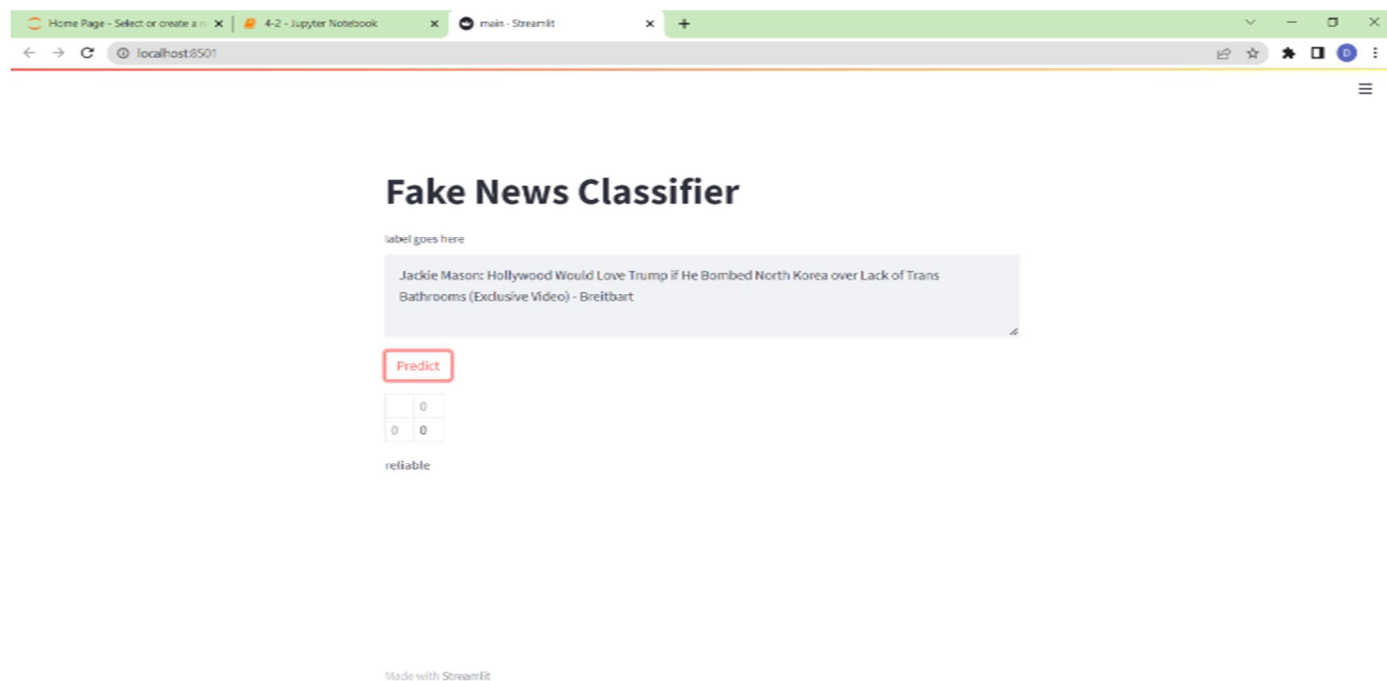Fig. 8. Title corresponding to index 5 in our dataset

Fig. 9. Prediction made on data corresponding to index 5

VII. CONCLUSION

We have devised a computational approach for detecting fake news. The goal of the work presented in this project is to apply deep learning techniques to the challenge of detecting fake news from the headline. On data containing the title of the article, we trained LSTM and Bi-LSTM models. The models were trained on a labelled dataset consisting of fake and real news, and they were successful in this job. The LSTM and Bi-LSTM algorithms proved to be effective with an accuracy of 93%.

While the findings suggest that various external features such as the source of the news, author of the news, place of origin of the news, time stamp of the news, and text of the news were not taken into account in our model, which may have influenced the model's output. Fake News training time grows as the number of layers of module increases.

Text, author names, and other attributes available in the dataset can be used to develop models. In the future, further in-depth studies will be required to better understand how a deep learning model may aid in the automatic credibility analysis of news

.

REFERENCES

[1] Viera Maslei, Martin and Peter Butka, "Deep learning methods for Fake News detection", CINTI-MACRo 2019, IEEE Joint 19th International Symposium on Computational Intelligence and Informatics and 7th IEEE International Conference on Recent Achievements in Mechatronics, Automation, Computer Sciences and Robotics, November 14-16, 2019, Szeged, Hungary

[2] Rohit Kumar Kaliyar, "Fake News Detection Using a Deep Neural Network", 2018 4th International Conference on Computing Communication and Automation (ICCCA)

[3] Tejaswini Yesugade, Shrikant Kokate, Sarjana Patil, Ritik Varma, Sejal Pawar," Fake News Detection using LSTM", International Research Journal of Engineering and Technology (IRJET), Volume: 08 Issue: 04 | Apr 2021.

[4] Pritika Bahad, Preeti Saxena, Raj Kamal, "Fake News Detection using Bi-directional LSTM-Recurrent Neural Network", INTERNATIONAL CONFERENCE ON RECENT TRENDS IN ADVANCED COMPUTING 2019, ICRTAC 2019

[5] Syed Ishfaq Manzoor, Dr Jimmy Singla, Nikita, "Fake News Detection Using Machine Learning approaches: A systematic Review", Proceedings of the Third International Conference on Trends in Electronics and Informatics (ICOEI 2019) IEEE Xplore Part Number: CFP19J32-ART; ISBN: 978-1-5386-9439-8

[6] Supanya, Prabhas, "Detecting Fake News with Machine Learning Method",2015 15th International Conference on Electrical Engineering/Electronics, Computer, Telecommunication and Information Technology

INTERNATIONAL JOURNAL
FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY