**Fake News Detection using Machine Learning**

Saydunnahar Jerin, [Saydunnaharjerin007@gmail.com](mailto:Saydunnaharjerin007@gmail.com)

Zannatul Ferdous Jui, [zannatuljui69@gmail.com](mailto:zannatuljui69@gmail.com)

Shahana Alam, [shahanaalamantora32@gmail.com](mailto:shahanaalamantora32@gmail.com)

**INTRODUCTION**

Fake news is increasingly becoming a menace to our society. It is a form of news consisting of deliberate [disinformation](https://en.wikipedia.org/wiki/Disinformation) spread via traditional [news media](https://en.wikipedia.org/wiki/News_media) or online [social media](https://en.wikipedia.org/wiki/Social_media) which is generated for commercial interests—to attract viewers and collect advertising revenue. Clickbait, Propaganda, Satire/parody, Sloppy journalism, Misleading headings, Biased or slanted news are some of the examples of fake news.

Due to low cost, simple access and fast dissemination of information, the majority of people search and consume news from social media rather than traditional news organizations these days. On one side, where social media has become a powerful source of information and bringing people together, on the other side it also puts a negative impact on society. So, Fake news detection is one of the vital research domains in this modern era that can enlighten people with proper information. This can reduce cybercrimes, can be a help to the police force in their investigation as well as explosive growth in fake news and its erosion to democracy, justice, and public trust has increased the demand for fake news detection and intervention.

Fake news identification faces several challenges. Firstly, it is difficult to collect fake news data. Furthermore, it is difficult to label fake news manually. Since they are intentionally written to mislead readers, it is difficult to detect them simply on the basis of news content. Furthermore, the misinformation disseminated by trusted news outlets or their friends and family is therefore difficult to be considered as fake. It is not easy to verify the credibility of newly emerging and time-bound news as they are not sufficient to train the application dataset.

**RELATED WORK**

To detect whether a news is fake or not multiple different models of deep learning can be implemented on different datasets such as this paper used Kaggle dataset and Signal Media News dataset [1]. The Logistic Regression model was implemented to serve as a baseline for comparison. The Recurrent Neural Network was implemented to process the entire length of every news article. Long Shoot-Term Memories method maintains a memory to determine the output. Besides they implemented Two-layer Feedforward Neural Network, Gated Recurrent Units, Bidirectional RNN with LSTMs, Convolutional Neural Network with Max Pooling, Attention-Augmented Convolutional Neural Network and they found Gated Recurrent Units model performed better among all models.

Hybrid convolutional neural networks and long-shortterm recurrent neural network models were used to detect fake news on twitter which consist of 5,800 tweetswith five rumor stories. The dataset was labeled as rumorand non-rumors [2]. Here first of the twofold approach is theautomatic identification of features within Twitter postswithout prior knowledge of the subject domain and second is thedetermination and classification of fake news posts onTwitter using both text and images. Three deep neural network variants were implemented and they are Long-Short Term Memory (LSTM), LSTM with dropout regularization, LSTM with convolutional neural networks. They tried to improve model prediction accuracy and got 82% accuracy. LSTM performed better among all methods.

Another technique to detect fake news is using Bi-directional LSTM-recurrent neural network. This method detects fake news by determining whether the information is correct by analyzing the relationship between the news article headline and article body. Besides Bi-directional LSTM-RNN here they performed CNN, Vanilla RNN, Unidirectional LSTM-RNN model. Global Vectors is used for obtaining vector representations for words. The datasets are partitioned as train, validate and test partitioning. Here they used two datasets which are obtained from open Machine Learning Repository accessible at Kaggle. The result shows the triumph of Bi-directional LSTM model. Test accuracy for DS1 is 0.9108% and for DS2 is 0.9875% using Bi-directional LSTM-RNN (Long Short Term Memory-Recurrent Neural Network) model.

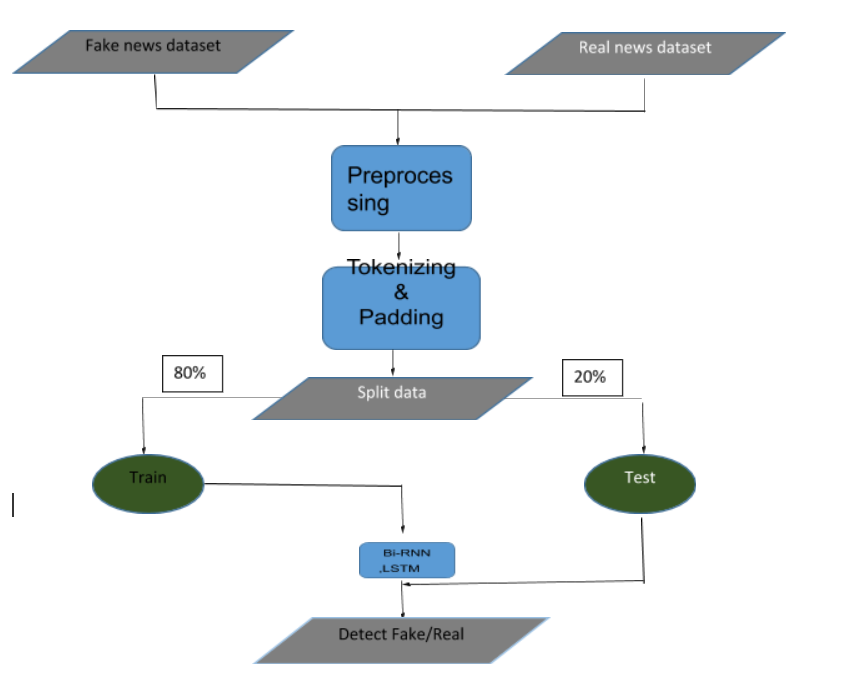
In our project we re-implement LSTM, RNN( Long Short-Term Memory , Recurrent Neural Network) model and obtain 0.9979% accuracy.

**PROJECT OBJECTIVES**

In our project, we have done the following steps:

1. Apply python libraries to import and visualize datasets
2. Perform exploratory data analysis and plot word-cloud
3. Perform text data cleaning such as removing punctuation and stop words
4. Perform tokenizing and padding on text corpus to feed the machine learning model
5. Build and train the machine learning model
6. Access the performance of the trained model

**Flow Chart of the System**

****

Some dummy input and output of the system

**Input :**

The wife of British aristocrat Lord Lucan, who vanished without trace 43 years ago after the murder of his child nunny, has been found dead.

**Output :** True

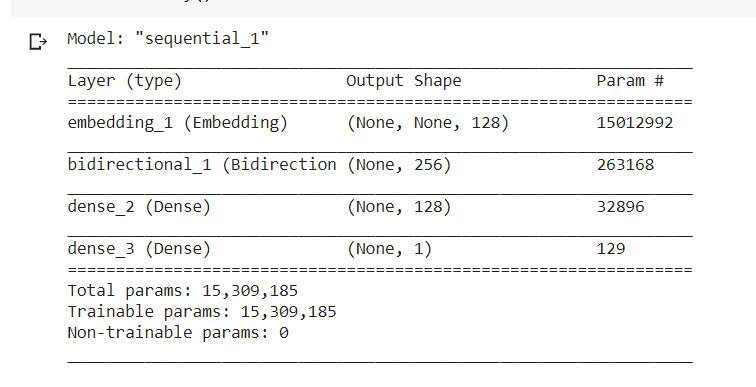
**Input:**

Donald Trump sends out embarrassing New Year Message

**Output :** False

**METHODOLOGIES**

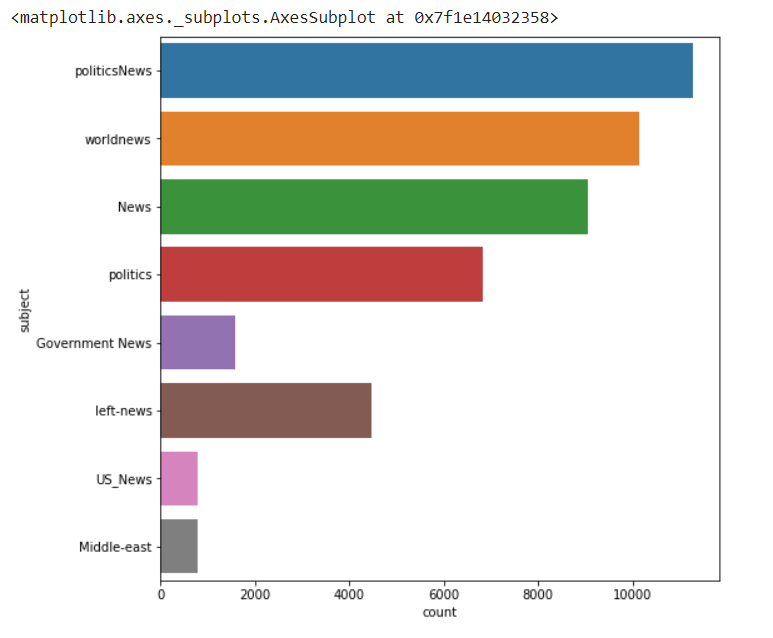
Architecture of the Model



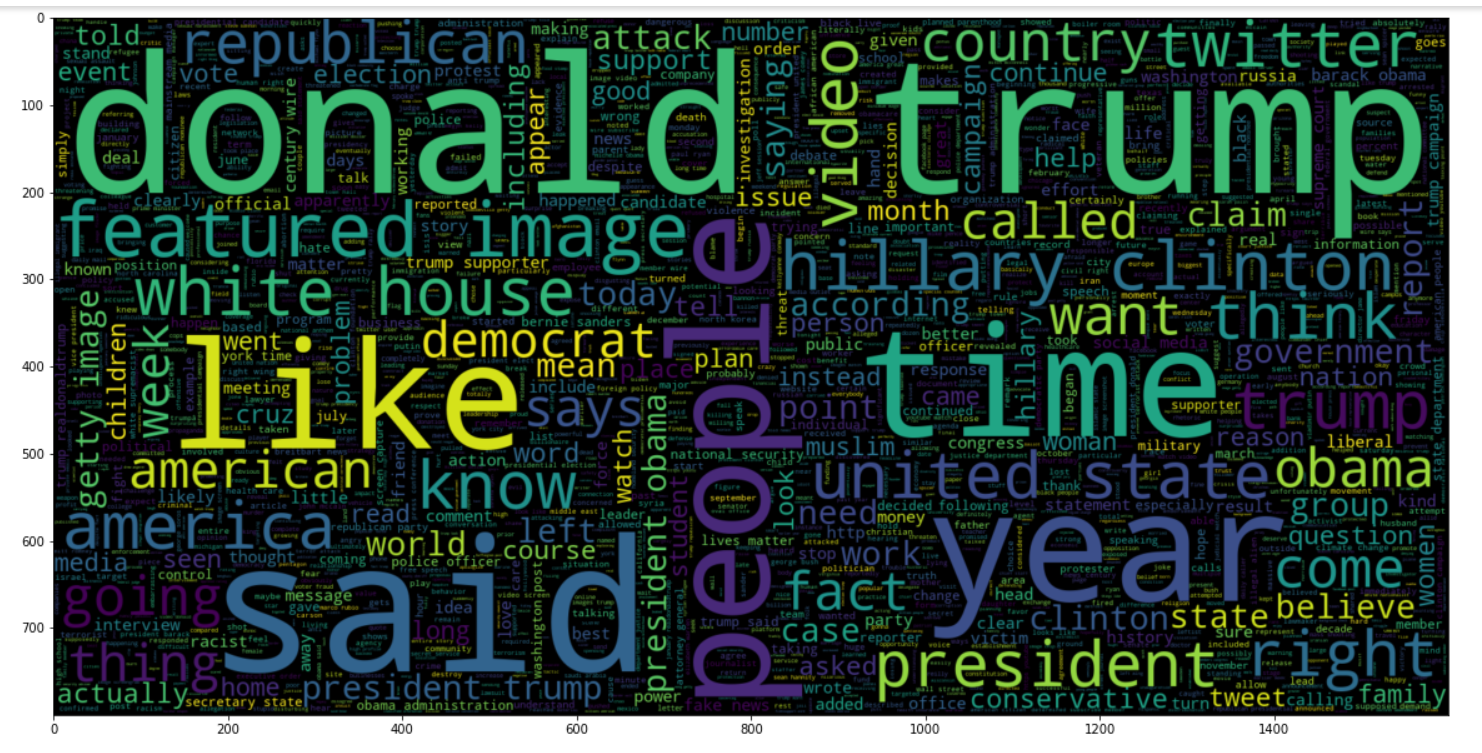
**EXPERIMENTS**

**Dataset**

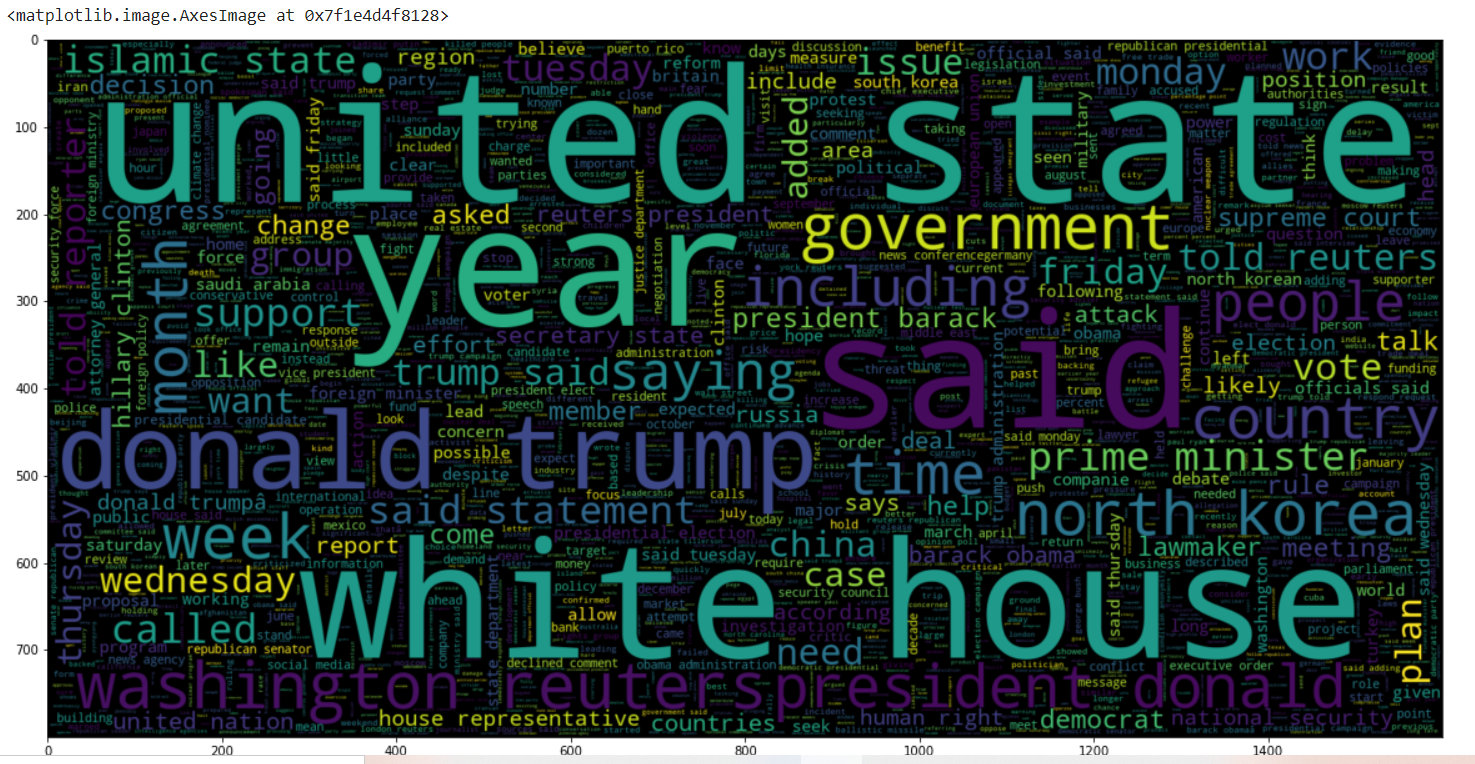
1. Two datasets had been used. Fake news dataset contains 23503 rows and True news dataset contains 21418 rows.
2. Datasets had been split on Train on 32326 samples, Test on 12571 samples, Validate on 3592 samples
3. Datasets had different types of news like-



1. Fake words which had been used-

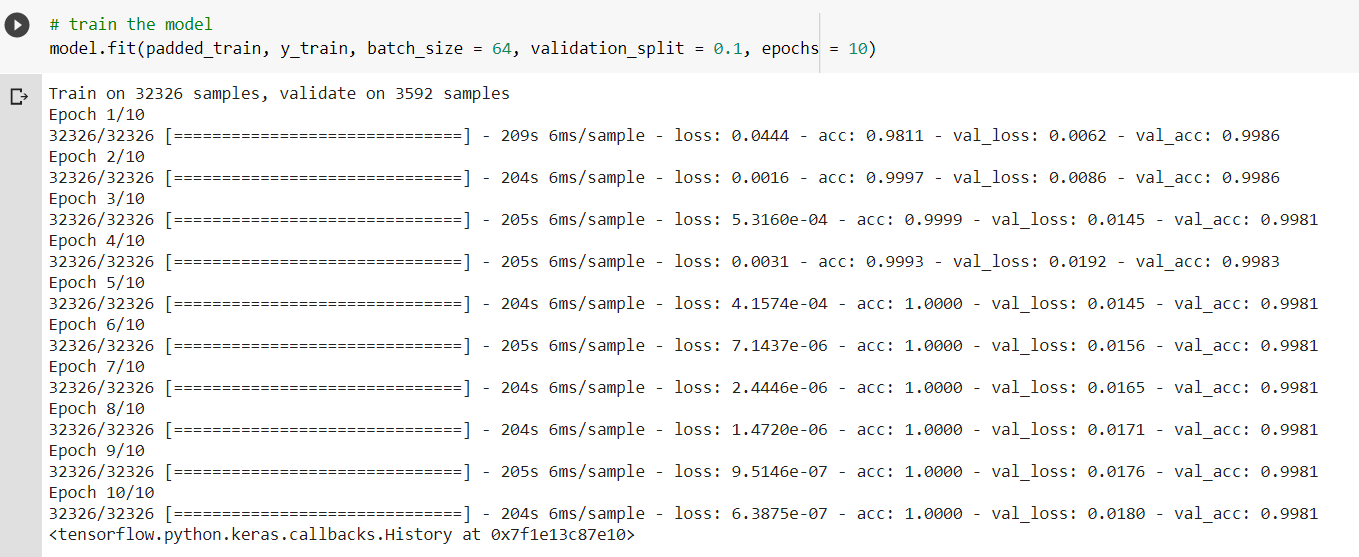


1. True words which had been used-



**Evaluation Metric**

Accuracy for the training model-



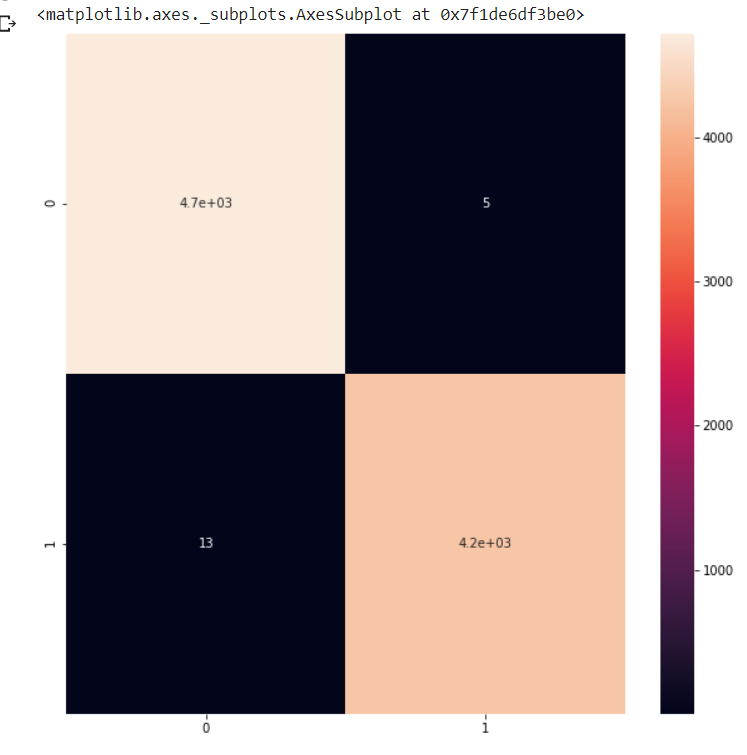
Training accuracy after 10 epoch 0.998.

Accuracy for the testing model-

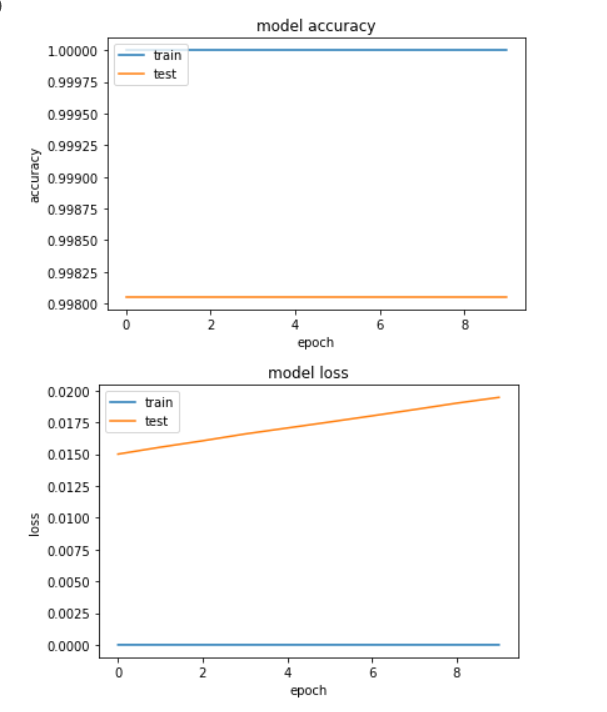


Testing accuracy 0.997.

**Confusion Matrix**

****

**RESULTS**

****

**CONCLUSION**

Fake news means presenting untrue information as news for advertising revenue or damaging the reputation of an entity or a person. It can mislead people. Fake news detection is a time demand. In our proposed project we used such a model which can detect fake news automatically. Here we merged two datasets one is Fake News Dataset another is Real News Dataset. Then we preprocessed that merged dataset. After tokenizing and padding the dataset is partitioned as 80% : 20% to train and test data. Then we applied Bi-directional RNN, LSTM model on train data. The final output will be either fake or real. If the news people are getting is fake, then they will establish their logic on lies. Fake news can cast a huge loss of a person as well as of a society. Fake news detection can dispel this damage. Also it will bring people out from so much harassment and will reduce crime, corruption from society.

**REFERENCES-**

1. <https://web.stanford.edu/class/archive/cs/cs224n/cs224n.1174/reports/2710385.pdf?fbclid=IwAR1Q9D2CQRWlbAlNz3VmqxdxDA6DiLdMYE_nNIp2QLU3W_Rd6ZMGJWjnRMM>
2. <https://arxiv.org/ftp/arxiv/papers/1806/1806.11316.pdf?fbclid=IwAR1Q9D2CQRWlbAlNz3VmqxdxDA6DiLdMYE_nNIp2QLU3W_Rd6ZMGJWjnRMM>
3. <https://www.researchgate.net/publication/339543567_Fake_News_Detection_using_Bi-directional_LSTM-Recurrent_Neural_Network?fbclid=IwAR1Q9D2CQRWlbAlNz3VmqxdxDA6DiLdMYE_nNIp2QLU3W_Rd6ZMGJWjnRMM>
4. <https://www.kaggle.com/clmentbisaillon/fake-and-real-news-dataset?select=True.csv>