# FAKE NEWS DETECTION USING BIDIRECTIONAL LSTM

**Dr. Suma Sira Jacob [1], Jayesh Narayanan S [2], Kamalesh A R [3], Kaviya S [4]**

[1] Associate Professor, Department of IT, Sri Krishna College of Technology, Coimbatore, India

[2,3,4] Students, Department of IT, Sri Krishna College of Technology, Coimbatore, India

sumasirajacob@skct.edu.in[1], 19tuit033@skct.edu.in[2] , 19tuit034@skct.edu.in[3] , 19tuit039@skct.edu.in[4].

*Abstract:* **Social media sites like Facebook, WhatsApp, Twitter, and Telegram have captured people's attention all around the world by disseminating false information and offering a convenient, affordable, and easy way to exchange information. In order to get personal or financial advantage from society, fake news is frequently produced in order to fool the general audience. So, several types of studies are conducted to detect false news with high accuracy in order to prevent it in order to lessen the harmful effects. We present a full evaluation of this false news detection technique during this research because of its disastrous effect, which is driven by the aforementioned problems. Also, the implementation of improvisation and the limitations of such approaches are discussed. Existing methods - K-Means, CNN, Naives Bayes, etc. Here we have attempted to provide a more accurate classification using Bidirectional LSTM. This model marches toward the path of early detection to flag the propagators before propagation.**

***Keywords:*** *Twitter, False news detection, Bidirectional LSTM.*

## I. INTRODUCTION

The emergence of the internet and the quick adoption of social media platforms like Twitter and Facebook have enabled the dissemination of information on an unprecedented scale never seen before in human history. This has been beneficial for news outlets, who can now provide real-time updates to subscribers through digital platforms like online news websites, blogs, and social media feeds. With 70% of traffic to news websites coming from Facebook referrals, it has become much easier for consumers to stay up-to-date with the latest news. Social media platforms have also become a powerful tool for users to share ideas, discuss and debate important issues such as democracy, education, and health. However, some entities use these platforms negatively, for monetary gain or to spread biased opinions, manipulate mindsets, or spread fake news and satire. This negative phenomenon is commonly known as fake news.

To spot and monitor false information, ML and DL techniques such as SVM, logistic regression, NB, DT, CNN, DNN, LSTM, and BiLSTM have been utilized by researchers worldwide. These methods have yielded highly precise outcomes. Therefore, we conduct a comprehensive examination of the most advanced approaches for detecting fake news.

### 1.1 Problem Formulation

In order to articulate the problem in a formal manner, we aim to find a function *'f'* that can be applied to a news article described by specific features (e.g., title, text, photos, newspaper, author, etc.). The function *'f'* should output a binary value *'a'*, where '0' indicates that the article is fake and '1' indicates that the article is true.

$$f(a) = \begin{cases} 0 & \text{if } a \text{ is fake} \\ 1 & \text{if } a \text{ is true} \end{cases}$$

With importance on Bi-directional LSTM , the perception of unofficial news is collected and monitored accordingly on the daily basis of usage and of an individual in a social network.

### 1.2 Our Advances

The existing corpus of fake news contains several instances where supervised and unsupervised learning algorithms have been used to classify text. Algorithms trained on a specific domain perform better on articles from that domain, but may not produce optimal results when applied to articles from other domains. This is because articles from different domains have unique textual structures, making it challenging to train a generic algorithm that performs well across all news domains.

To address this problem, our study proposes a machine learning approach to detecting fake news. We train the machine learning algorithm using techniques such as bagging and boosting, with live news feeds, which have been proven to reduce error rates and improve training effectiveness and efficiency. Our results demonstrate improved performance compared to existing techniques, and show the potential for our approach to be applied across a wide range of domains.

## II. RELATED WORK

Following the 2016 U.S. presidential election, the machine learning research community devoted significant attention to identifying fake news, which became a major topic in the media. One study conducted an extensive examination of false news identification by leveraging information from various sources. The study identified several characteristics that are utilized in [20] content-based classification, including lexical, syntactic, visual, statistical, user, post, and network features. Deep neural network models were employed to detect fake news, and the experiments were conducted on real-world datasets from Buzzfeed and PolitiFact. The classification technique involved three distinct sections, with the first two being text-based categorization systems and the third utilizing both social context and social network data. The deep neural network model used in [8], was found to perform the best among the evaluated models, outperforming the prior machine learning-based approach by over eight percent, and improving on four language aspects.

In other studies [20], [19], [18], several authors employed various machine learning methods to identify fake news and obtained the best results using Support Vector Machine (SVM), achieving a result of 95.70%. Additionally, a capsule neural network was utilized to detect fake news by another researcher.

## III. METHODOLOGY

This task involves six distinct steps, which can be briefly outlined in this paragraph. The first step is to collect data from multiple online news sources to construct a comprehensive dataset. Next, the data will be processed live in preparation for analysis. Bi-LSTM classifiers will then be employed to classify the dataset and develop the model. Once the model is trained and tested on the dataset, its performance will be evaluated using measures such as accuracy, recall, and precision.

The flow of inputs in Bidirectional LSTM is shown in Fig 3.1.

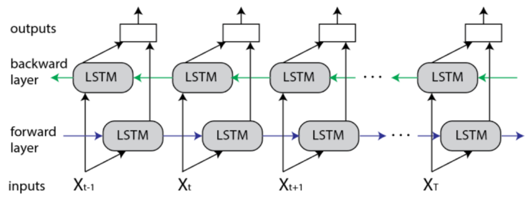


**Fig 3.1 Bidirectional - LSTM**

The proposed model workflow for this research work is shown in Fig 3.2.

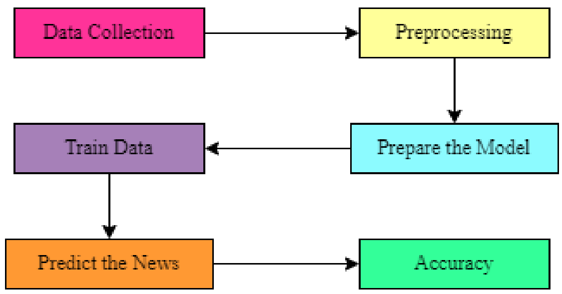


**Fig 3.2: Proposed Model Workflow**

The most crucial aspect of conducting research is the specific topic and field being investigated. Our study focuses primarily on news-related matters. In addition to simply reporting news, we aim to present it to our audience in a unique and informative way. It is not sufficient to merely have a general understanding of the research; a detailed explanation is necessary to truly comprehend the subject matter. Therefore, we have covered all relevant subfields in our study to ensure a comprehensive understanding of the topic. Researchers utilize various instruments and tools to aid in their investigations. As depicted in Fig 3.3, we have developed a proposed flowchart for our study.

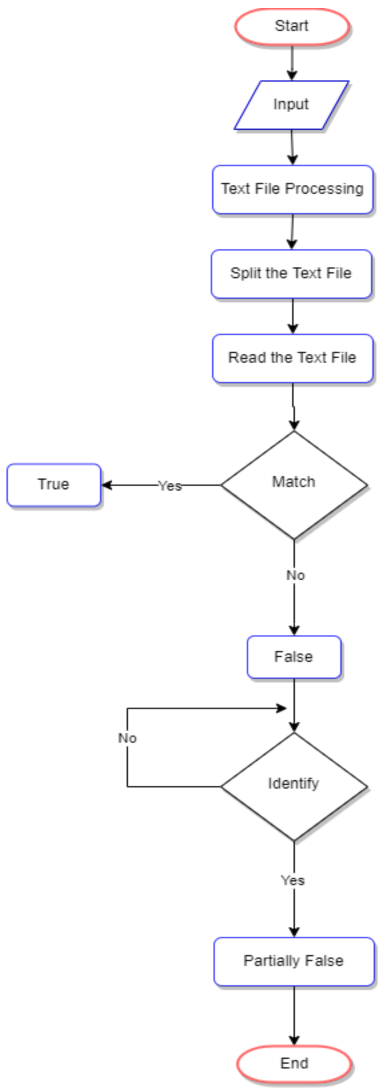


**Fig 3.3: Proposed Flowchart of our Work**

## IV. DATA OVERVIEW

In this study, the data was sourced from Media Stack, which comprises 10000+ news articles that are categorized as either true or fake. Each article in the dataset includes various attributes such as titles, contents, publisher details, authors, and image URLs. However, for the purpose of this research, only the titles and contents were utilized. The reason for selecting this particular dataset was that it has been employed in previous research with satisfactory outcomes, indicating that the news data's quality is suitable for training neural network models, such as the one suggested in this study.

The initial version of the data involves converting the article titles and bodies into integer vectors using a dictionary (Word2Vec) where the vector position corresponds to the token ID. However, as neural networks require a fixed input size, a maximum number of words needed to be determined for each title and body. To determine this maximum number, the distribution of tokens in both titles and bodies was analyzed in order to find a value that is both small enough to be computationally efficient and large enough to minimize text truncation.

Table 3.1 shows the quantity of data collections along with platforms.

| **Platform (MediaStack)** | **Data** |
|---|---|
| CNN | 2343 |
| Sky News | 2235 |
| Euro News | 567 |
| ABC News | 2786 |
| USA Today | 1359 |
| The Guardian | 841 |
| **Total** | 10131 |

**Table 3.1 Quantity of Data with Platforms (*Live news data collection count may vary)**

## V. WORKING PROCEDURE

The data is collected manually by our researcher and saved in a text document file. These files contain various unwanted elements such as html tags and redundant information. It has been realized that the data needs to be cleaned or preprocessed to be suitable for use in the model.

*I) Bi-LSTM model:*

The Bi-Directional Long Short-Term Memory (BiLSTM) outperforms Long Short-Term Memory (LSTM) in sequence categorization. It consists of two LSTMs, one for processing the forward input and the other for processing the backward input. Concealed states are utilized to enable the exchange of data in both directions. At each time step, the outputs of the two LSTMs are merged into a single output. By overcoming the limitations of standard RNN, the BiLSTM technique enhances the comprehension of context due to its high level of precision.

Fig 5.1 shows the common architecture of Bi-directional LSTM.

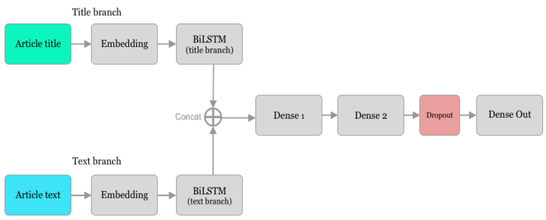


**Fig 5.1: Bidirectional LSTM Architecture Diagram**

*II) Data Cleaning:*

We simplify our data pre-processing task by employing a python script that eliminates html tags and removes any unnecessary gaps or new lines in the text. Additionally, each news category is assigned a specific integer number for pre-definition. This approach allows us to obtain categorized news in separate files, but the data returned has already undergone pre-processing and classification.

*III) Recurrent Neural Network (RNN):*

An RNN, or recurrent neural network, is an artificial neural network that is designed to handle sequential or time series data. These types of deep learning algorithms are often used for problems that involve ordering or time-based considerations, such as language translation, natural language processing, speech recognition, and image captioning. RNNs are utilized in popular applications like Siri, voice search, and Google Translate. Like feedforward and convolutional neural networks, RNNs are trained using data. However, they are unique in that they have a "memory" that allows them to incorporate information from previous inputs to influence the current input and output. Unlike traditional deep neural networks, which assume that inputs and outputs are independent, the output of an RNN depends on prior elements within the sequence. Although knowing about future events could be useful in determining the output of a sequence, unidirectional RNNs are unable to consider them in their predictions.

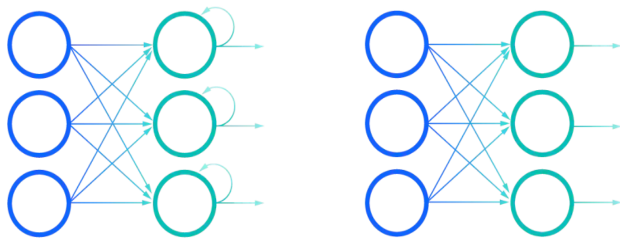

**Fig 5.2 Comparison of Recurrent Neural Networks (on the left) and Feedforward Neural Networks (on the right)**

### 5.1 Extracting Features & Dataset Fitting

The classification stage of this approach primarily involves feature selection and extraction, which plays a critical role in the final decision-making process for categorization. Due to the massive volume of data, manual processing is not feasible, and feature extraction becomes important. To begin the modeling process, we divide our dataset into two parts: the training dataset and the testing dataset. The training dataset accounts for 80% of the data, while the testing dataset accounts for 20%. We expect this ratio to be reflected in our model. Since we are working with Bi-LSTM classifiers, we import the sklearn package. This classifier generates an integer that indicates the predicted news category.

The data ratio for this research work is shown in Fig 5.2.

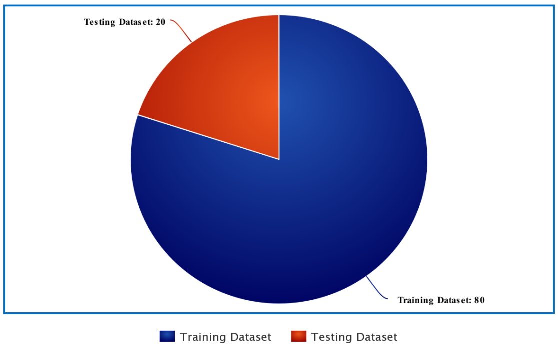


**Fig 5.2 Pie Chart comparison of the Data**

## VI. RESULTS AND DISCUSSION

The approach uses soft-max activation for the output layer, as it is a multiclass dataset, ensuring that neither the training dataset nor the test dataset are used. During the training process, the model utilizes 128 batch sizes of training data over 10 epochs. The model's accuracy was found to be 91.84%, and its F1-macro score was 79.0. The model's output identifies the news item shown, which can be either real or fake. In order to determine the identification of the news, True is assigned as 1 and False as 0, as the RNN model cannot process text. We trained our model for 10 epochs and achieved an accuracy of 91.84% and an F1-macro score of 79%.

Table 6.1 presents the model's accuracy and classification.

| Class | Precession |
|---|---|
| True | 0.92 |
| False | 0.91 |
| Partially False | 0.17 |

**Table 6.1 Classification Report of the Model**

Table 6.2 represents the Method and its Accuracy.

| Method | Accuracy |
|---|---|
| Bi LSTM | 91.84% |

**Table 6.2 Accuracy Report**

Figure 6.1 below illustrates a graphical representation of the accuracy of our proposed model's training data compared to its validation data. The image shows that our model is effectively learning from its previous data.

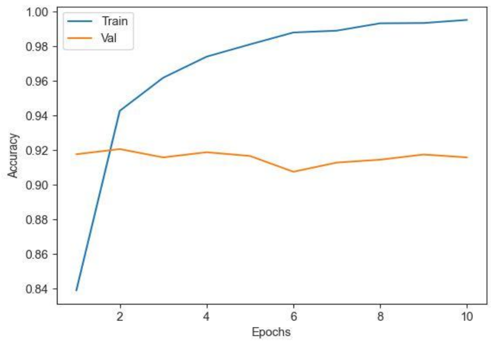


**Fig 6.1 Epoch vs Accuracy on Training data and Validation data.**

## VII. MATRIX REPORT OF TRUE vs PREDICTED LABELS

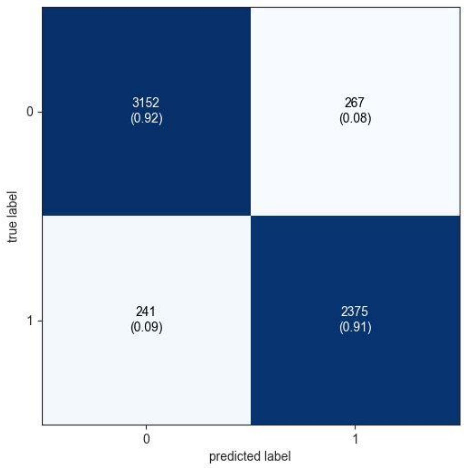


**Fig 7.1 True Label vs Predicted Label Matrix**

## VIII. CONCLUSION

This study proposed a BiLSTM model for detecting fake news, which involved using filter techniques to remove unnecessary attributes from the dataset for faster training. Word2Vec tokenizer were also utilized to make features machine-readable. To speed up the training process, various data preparation activities like stopword removal and stemming were performed. The model achieved an accuracy rate of 91.84% and a macro score of 79.0, which is considered satisfactory. However, there is room for improvement, such as combining BiLSTM with CNN and investing in high-quality hardware components to enhance model performance. In the future, we can plan to test their approach on a larger and more diverse dataset to overcome the current limitations.

## ACKNOWLEDGEMENT

We want to express our profound gratitude to Dr S Siamala Devi, the department's head. We also would like to thank our project guide Associate Professor Dr. Suma Sira Jacob.

## REFERENCES

[1] Albright, J.: Welcome to the era of fake news. Media Commun. 5(2), 87 (2017).
[2] J. Weedon, W. Nuland, and A. Stamos. Information Operations and Facebook. Retrieved from Facebook (2017).
[3] Ahmed, S., Hinkelmann, K., Corradini, F.: Combining machine learning with knowledge engineering to detect fake news in social networks - a survey. In: Proceedings of the AAAI Spring Symposium, vol. 12 (2019).
[4] Atodiresei, C.-S., Tănăselea, A., Iftene, A.: Identifying fake news and fake users on Twitter. Procedia Comput. Sci. 126, 451–461 (2018).
[5] P. Thakkar, K. Varma, V. Ukani, S. Mankad, and S. Tanwar, "Combining user-based and item-based collaborative filtering using machine learning," in Information and Communication Technology for Intelligent Systems, S. C. Satapathy and A. Joshi, Eds. Singapore: Springer, 2019, pp. 173–180 (2019).
[6] B. D. Horne and S. Adali, "This just in: Fake news packs a lot in title, uses simpler, repetitive content in text body, more similar to satire than real news," in Proc. 11th Int. AAAI Conf. Web Social Media, 2017, pp. 759–766 (2017).
[7] B. D. Horne, J. Nørregaard, and S. Adali, "Robust fake news detection over time and attack,"ACM Trans. Intell. Syst. Technol., vol. 11, no. 1, pp. 1–23, Feb. (2020).
[8] Thota, A., Tilak, P., Ahluwalia, S., Lohia, N.: Fake news detection: a deep learning approach. SMU Data Sci. Rev. 1(3), 21 (2018).
[9] R. Varma, Y. Verma, P. Vijayvargiya, and P. P. Churi, "A systematic survey on deep learning and machine learning approaches of fake news detection in the pre- and post-COVID-19 pandemic," Int. J. Intell. Computer Cybern., vol. 14, no. 4, pp. 617–646, Oct. (2021).
[10] J. Golbeck, M. Mauriello, B. Auxier, K. H. Bhanushali, C. Bonk, M. A. Bouzaghrane, C. Buntain, R. Chanduka, P. Cheakalos, J. B. Everett, and W. Falak, "Fake news vs satire: A dataset and analysis," in Proc. 10th ACM Conf. Web Sci., 2018, pp. 17–21 (2018).
[11] Horne, B.D., Adali, S.: This just in: fake news packs a lot in title, uses simpler, repetitive content in text body, more similar to satire than real news. In: International AAAI Conference on Web and Social Media, vol. 8 (2017).
[12] ADS—Accertamenti Diffusione Stampa, ADS, (2017).
[13] History of Fake News—Fake News—Libguides at Newcastle University. Accessed: Feb. 7, (2020).
[14] Collins 2017 Word of the Year Shortlist, Nov. (2020).
[15] C. Shao, G. L. Ciampaglia, O. Varol, K. Yang, A. Flammini, and F. Menczer, "The spread of low-credibility content by social bots," 2017, arXiv:1707.07592 (2017).
[16] K. Nagi, "New social media and impact of fake news on society," in Proc. ICSSM, Jul. 2018, pp. 77–96 (2018).
[17] V. Vapnik, The Nature of Statistical Learning Theory. New York, NY, USA: Springer, (2013).
[18] J. C. S. Reis, A. Correia, and F. Murai, A. Veloso, and F. Benevenuto, "Supervised learning for fake news detection," IEEE Intell. Syst., vol. 34, no. 2, pp. 76–81, Mar./Apr. (2019).
[19] H. Ahmed, I. Traore, and S. Saad, "Detecting opinion spams and fake news using text classification," Secur. Privacy, vol. 1, no. 1, p. e9, Jan. 2018, doi: 10.1002/spy2.9.
[20] S. B. Deokate, "Fake news detection using support vector machine learning algorithm," Int. J. for Res. Appl. Sci. Eng. Technol., vol. 7, no. 7, pp. 438–444, Jul. 2019