Fake News Detection Using Machine Learning Methods

1st S M Samin
Computer Science and Engineering
Ahsanullah University of Science and Technology
Dhaka, Bangladesh
samin.cse.200104112@aust.edu

2nd Asad Chowdhury
Computer Science and Engineering
Ahsanullah University of Science and Technology
Dhaka, Bangladesh
asad.cse.200104127@aust.edu

3rd Aman Bhuiyan
Computer Science and Engineering
Ahsanullah University of Science and Technology
Dhaka, Bangladesh
aman.cse.200104102@aust.edu

4th Tanzilur Rahman
Computer Science and Engineering
Ahsanullah University of Science and Technology
Dhaka, Bangladesh
tanzilur.cse.200104103@aust.edu

I. ABSTRACT

The prevalence of fake news circulating on social media platforms has become a significant and concerning issue due to its potential for causing widespread harm and disseminating misinformation. Researchers have extensively explored the application of various machine learning approaches to combat this problem. This paper addresses the challenge of fake news detection by leveraging a dataset comprising 4047 instances sourced from Kaggle. The study employs four different models, namely Logistic Regression (LR), Convolutional Neural Network (CNN), Recurrent Neural Network (RNN), and Bidirectional Long Short-Term Memory (Bi-LSTM). Among these models, the CNN model emerges as the most effective, achieving the highest accuracy of 97.7%. This research sheds light on the importance of diversifying the focus of fake news detection methodologies to enhance their applicability across various types of misinformation in social media.

II. INTRODUCTION

In today's digital age, the widespread use of social media for news consumption has revolutionized how information is disseminated and accessed. With the convenience of online platforms like Facebook and Twitter, people find news readily available at their fingertips. However, this shift comes with a significant challenge - the rise of fake news. Fake news, or deliberately false information, has emerged as a formidable threat to democracy, journalism, and public trust.The emergence of fake news as a global issue is rooted in its rapid and cost-effective production and dissemination compared to traditional news media. Social media, with its vast reach and echo chamber effect, accelerates the spread of biased information, creating an ideal environment for fake news to flourish. Fake news has become a global concern, particularly due to its potential influence on major events like the "Brexit" referendum and the 2016 U.S. presidential election. During the critical months of the election campaign, fake news stories on social media outperformed those from reputable news websites, garnering millions of shares and interactions. The consequences of fake news extend beyond politics, affecting economies by causing stock market fluctuations and substantial financial losses.Fake news exploits social and psychological factors, taking advantage of human vulnerability in distinguishing truth from falsehood.Individuals tend to trust fake news if it aligns with their preexisting beliefs, confirming biases and further amplifying the spread. Peer pressure, confirmation bias, and desirability bias contribute to the challenge of fake news gaining public trust. The explosion of the World Wide Web since the mid-1990s has transformed communication patterns, with online social media platforms becoming primary channels for real-time information sharing. Approximately one-thirds of people now access news via social media. However, this shift has created an ideal breeding ground for fake news, encompassing misleading information, fake reviews, advertisements, rumors, and political statements.

The characteristics of fake news encompass its volume, variety, and velocity. The ease of creating fake news without verification procedures results in a massive influx of misleading content online. Fake news takes various forms, including rumors, satire, misinformation, and false advertisements, affecting every aspect of individuals' lives. The transient nature of fake news creators, coupled with their focus on current events, adds complexity to real-time detection. Efforts to identify valuable information amidst the vast amount of online content have led to extensive research on establishing automatic frameworks for fake news detection. The challenges lie in distinguishing truthful signals from fake and anomalous information, considering the dynamic and heterogeneous nature of online social communication. The increasing prevalence of fake news poses a serious threat to the fabric of democracy, public trust, and the integrity of information dissemination. This research seeks to unravel the complexities surrounding fake news detection, offering insights into its origins, motivations, and characteristics.

III. RELATED WORK

Fake news has become a pervasive issue in the age of information, especially on social media platforms where misinformation can spread rapidly. Detecting and mitigating the impact of fake news is crucial to maintaining the integrity of information ecosystems. In this literature review, we explore three key studies that contribute to the field of fake news detection, each employing distinct datasets, methodologies, and models. [1] ”Fake News Detection on Social Media: A Data Mining Perspective”: This seminal work addresses the challenge of manually determining the veracity of news and highlights four datasets: BuzzFeedNews, LIAR, BS Detector, and CREDBANK. The study emphasizes the diversity of data sources, including news agency homepages, search engines, and social media websites. The authors identify four primary methods for gathering annotated news data: Expert journalists, Fact-checking websites, Industry detectors, and Crowd-sourced workers. This foundational work sets the stage for subsequent research by emphasizing the importance of comprehensive datasets and diverse annotation methods. [2] ”A benchmark study of machine learning models for online fake news detection”: This study builds upon the LIAR Fake or real news Combined corpus dataset and employs machine learning models such as Support Vector Machines (SVM), Logistic Regression (LR), Convolutional Neural Networks (CNN), and Long Short-Term Memory (LSTM). The results demonstrate the effectiveness of CNN and LSTM models, achieving impressive accuracies of 93.6% and 94.2%, respectively. This work underscores the significance of leveraging advanced machine learning techniques for fake news detection, showcasing the potential for high accuracy in distinguishing between fake and real news.

[3] ”Beyond News Contents: The Role of Social Context for Fake News Detection”: Focusing on the BuzzFeed dataset, this study explores the role of social context in fake news detection. The authors utilize models such as Recursive Suppression Tree (RST), Linguistic Inquiry and Word Count (LIWC), and Castillo. Notably, the study goes beyond news contents to consider the broader social context. Results indicate promising accuracies, with Castillo achieving an accuracy of 80.34%. This research highlights the importance of incorporating social context features and alternative models for enhanced fake news detection. These three key studies collectively contribute valuable insights to the evolving landscape of fake news detection. They underscore the importance of diverse datasets, advanced machine learning models, and the incorporation of social context for more effective detection strategies. As the field continues to evolve, future research may benefit from combining these approaches, exploring new datasets, and developing more sophisticated models to address the persistent challenge of fake news proliferation in our information-driven society.

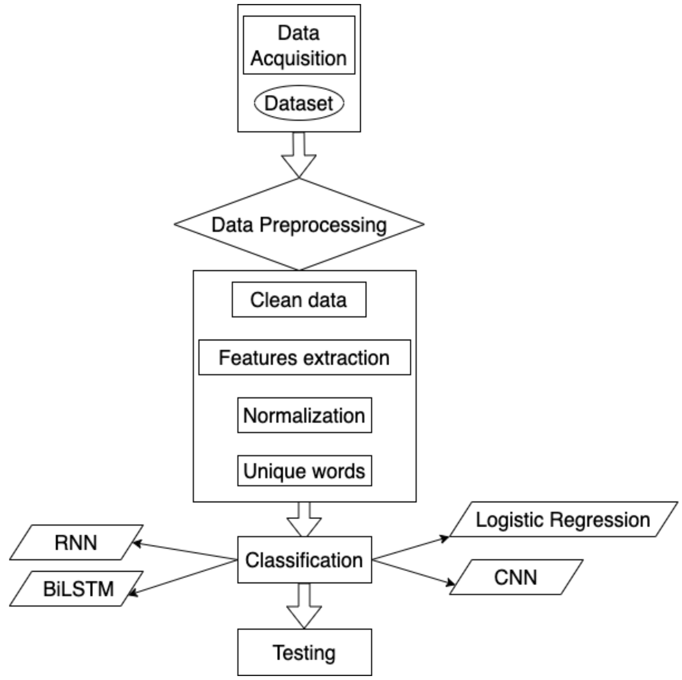

Fig. 1. Enter Caption

IV. DATASET REVIEW

Our dataset name is ”Fake News detection” dataset, we collected it from Kaggle, encompasses 4047 data points with four key features: URLs, Headline, Body, and Label. The Labels are binary, with 0 denoting real news and 1 indicating fake news. The URL feature presumably contains the web addresses of the news articles, while the Headline and Body features provide concise summaries and detailed information, respectively. Given the binary nature of the Label, the dataset likely serves a binary classification task, aiming to train a model to discern between real and fake news based on the provided features.

V. METHODOLOGY

Numerous computational models have been explored to formulate a system for detecting fake news. In this section, we elaborate on our proposed approach, encom- passing aspects such as data acquisition, data preprocessing, classification, and evaluation. The graphical representation of our proposed methodology is depicted in Figure 1.

A. *Data Acquisition*

The dataset employed in this study is referred to as the ”Fake news dataset”. This dataset categorizes observations into tow classes real news and fake news. Approximately 80% of the data has been utilized for training our models, while the remaining 20% has been reserved for testing these models.

B. *Data Preprocessing*

Preprocessing plays a pivotal role in every Natural Language Processing (NLP) study. Its primary objective is to ensure that the preprocessed data does not introduce any bias or skewness into the experiments. To achieve this, we initiated the preprocessing by cleaning our dataset, wherein we removed stop words and punctuations. This step significantly reduced the size of our dataset. Subsequently, for feature extraction, we employed the TF-IDF technique. Following feature extraction, we normalized the data to ensure that all features, specifically words in this context, have a similar scale. This normalization helps prevent any skewness in the performance of our models.

C. *Classification*

In our dataset classification process, we utilized two distinct classifiers to categorize the data into fake news or . The employed models are as follows:

i. Logistic Regression: Logistic Regression is a statistical method used for binary and multiclass classification. It models the probability of a certain class using the logistic function, which maps any real-valued number into the range of [0,1]. Logistic Regression classifies our dataset by calculating the weighted sum of input features and applying the logistic function to obtain the probability of each class. The class with the highest probability is then assigned to the data point.

ii CNN: CNNs are well-known for their success in image classification tasks, but they can also be applied to sequential data like text. In the context of processing textual information, 1D convolutions are often employed to capture local patterns and dependencies within sequences. Each convolutional layer uses filters to detect specific features in the input data. In the case of text classification, these filters may identify important n-grams or patterns of words. The resulting feature maps are then passed through pooling layers to reduce dimensionality and extract the most salient information. CNNs are particularly effective in capturing local structures and patterns in sequential data, making them suitable for tasks such as hate speech detection, where identifying specific linguistic patterns is crucial for accurate classification.

iii. RNN: : RNN is a type of recurrent neural network (RNN) designed to address the vanishing gradient problem in traditional RNNs. It is well-suited for processing and classifying sequences of data. RNN processes sequential data by learning long-term dependencies and relationships between words in sentences. It captures context and dependencies, making it effective for classifying text data like our hate speech dataset.

iv. BiLSTM (Bidirectional Long Short-Term Memory): BiLSTM is an extension of LSTM that processes sequences in both forward and backward directions. This bidirectional processing allows the model to capture context from both past and future data points. BiLSTM, by considering information from both directions, enhances the model's understanding of context, leading to improved classification accuracy for our dataset.

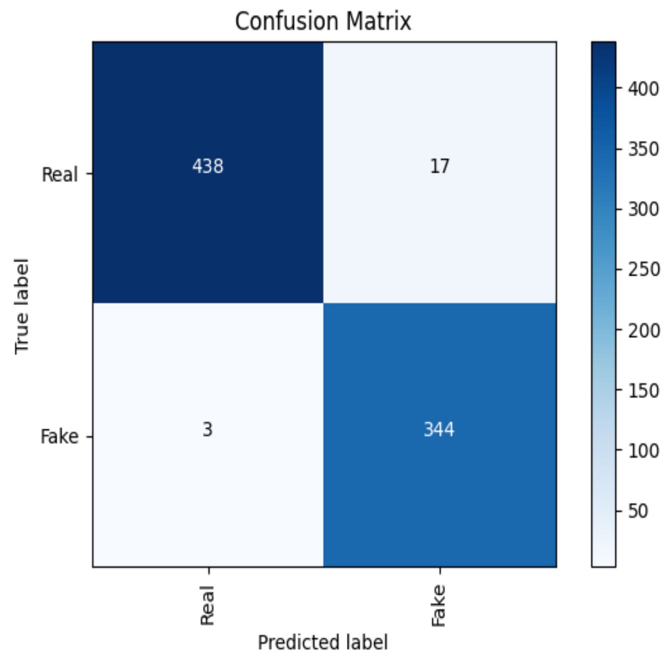

Fig. 2. Confusion Matrix of Logistic Regression

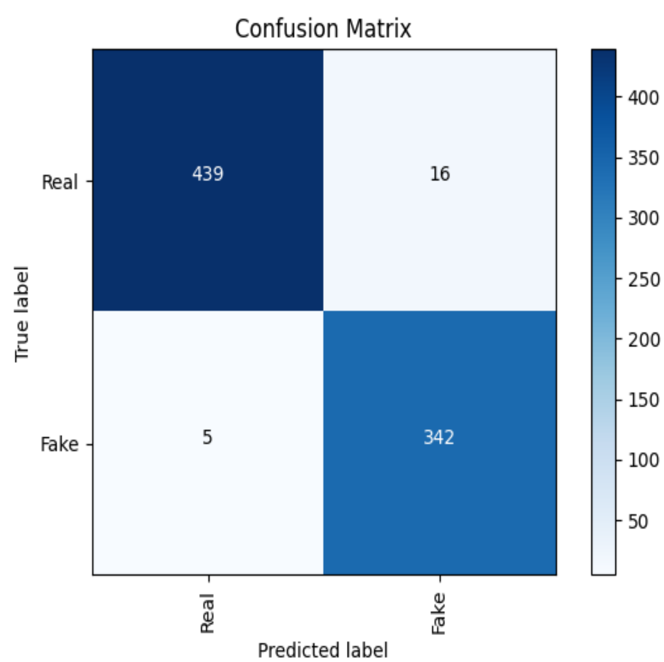

Fig. 3. Confusion Matrix of CNN

VI. RESULTS ANALYSIS

This section provides a comprehensive overview of the outcomes resulting from experimentation with four distinct models: Logistic Regression, LSTM (Long Short-Term Memory), BiLSTM (Bidirectional Long Short-Term Memory), and Conv-BiLSTM (Convolutional Bidirectional Long Short-Term Memory). The subsequent subsections delve into a detailed presentation of the results obtained, shedding light on the performance and efficacy of each model in addressing the specific objectives of the experiments.

A. *Logistic Regression:*

Upon executing the Logistic Regression model on our dataset, remarkable performance metrics were achieved. The model exhibited an accuracy of 97.51%, showcasing its ability to correctly classify instances. Precision, recall, and F1 score were equally impressive, standing at 97.58%, 97.51%, and 97.51%, respectively. This reinforces the efficacy of Logistic Regression in accurately predicting categories within the hate speech dataset. Its graphical representation is shown in Figure 2.

B. *CNN):*

The Convolutional Neural Network demonstrated exceptional performance in classifying hate speech. With an accuracy of 97.75%, the model showcased its robustness in discerning patterns within the dataset. Precision, recall, and F1 score aligned closely with high values of 97.79%, 97.75%, and 97.75%, affirming the CNN's proficiency in capturing complex features and relationships.

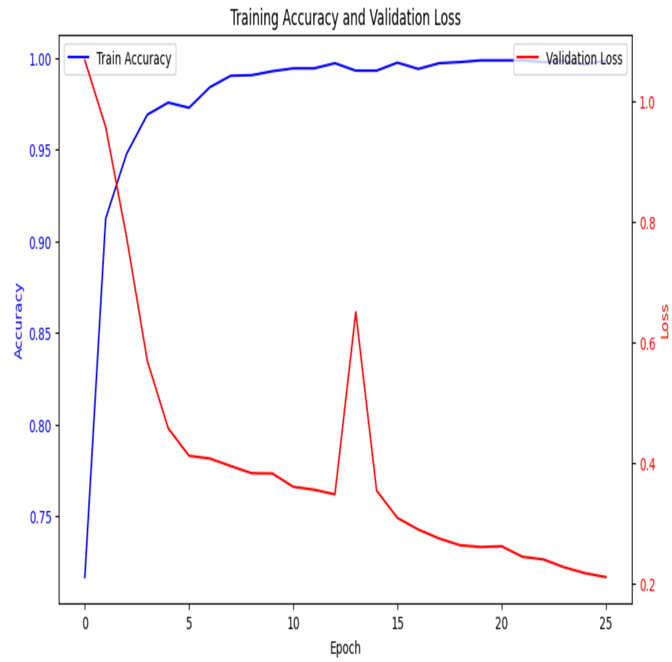

Fig. 4. RNN: Accuracy VS Loss

C. *RNN):*

Transitioning to the Recurrent Neural Network, the model's performance during testing revealed promising outcomes. Achieving a minimal test loss of 0.3087 and an impressive test accuracy of 97.38%, the RNN displayed its capability in understanding sequential dependencies and effectively classifying hate speech instances.

D. *BiLSTM (Bidirectional Long Short-Term Memory)):*

The Bidirectional Long Short-Term Memory model emerged as a strong performer, with a test loss of 0.4774 and a test accuracy of 97.01%. This underscores the model's ability to capture bidirectional context and long-term dependencies, contributing to its effectiveness in accurately classifying hate speech instances..

In summary, the experiments with Logistic Regression, Convolutional Neural Network, Recurrent Neural Network, and Bidirectional Long Short-Term Memory models highlight the diversity of approaches employed to address hate speech classification, each demonstrating noteworthy strengths in different aspects of the task.

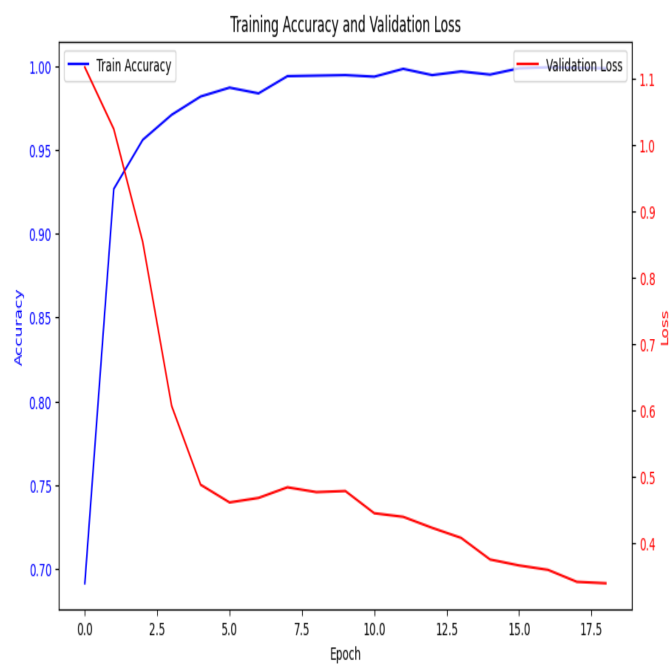

Fig. 5. BiLSTM: Accuracy VS Loss

VII. CONCLUSION

In summary, this paper underscores the significance of identifying and preventing fake news on social media plat- forms. We propose a methodology for fake news detection using a dataset that categorizes observations into two classes: fake news and real news. The proposed approach encompasses data ac- quisition, data preprocessing, classification, and evaluation utilizing four distinct classifiers— LR (Logistic Regression), CNN(Convolutional Neural Network), RNN(Recurrent Neural Network), BiLSTM (Bidirectional Long Short-Term MemoryAmong these, the algorithm that yielded the highest accuracy is 97.97% CNN provide this accuracy. We also stress the necessity for ongoing research and the development of effective methodologies to detect fake news in diverse languages and across various social media platforms. In conclusion, the paper provides valuable insights into the challenges and opportunities associated with addressing fake news in the contemporary digital era.

VIII. REFERENCE

[1] Fake News Detection on Social Media: A Data Mining Perspective (Kai Shu, Amy Sliva, Suhang Wang, Jiliang Tang, and Huan Liu)

[2] A benchmark study of machine learning models for online fake news detection (Junaed Younus Khan, Md. Tawkat Islam Khondaker, Sadia Afroz , Gias Uddin, Anindya Iqbal)

[3] Beyond News Contents: The Role of Social Context for Fake News Detection (Kai Shu, Suhang Wang, Huan Liu)