SYSTEMATIC REVIEW ON FAKE NEWS & DISINFORMATION USING ML

[1]N. ANIL KUMAR, [2]M. BHAVYA SREE, [3]M. LAKSHMI GANESH, [4]CH.SYAM VITAL KUMAR, [5]FARHAT SULTANA

[1]ASSOCIATE PROFESSOR,[2345] B. TECH, STUDENTS

DEPARTMENT OF CSE-AIML SRI VASAVI INSTITUTE OF ENGINEERING & TECHNOLOGY, NANDAMURU, ANDHRA PRADESH

ABSTRACT

The rapid growth of social media platforms has significantly transformed the way information is produced, shared, and consumed across the world. While these platforms provide a powerful medium for communication and information exchange, they also facilitate the rapid spread of fake news and disinformation. Such misinformation can influence public opinion, disrupt political processes, and create social instability. Traditional fact-checking mechanisms rely heavily on manual verification, which is time-consuming, resource-intensive, and incapable of handling the massive volume of online content generated daily. Consequently, automated approaches using Machine Learning (ML) and Natural Language Processing (NLP) have emerged as promising solutions for identifying and mitigating fake news. This study presents a systematic review and comparative analysis of machine learning and deep learning techniques used for fake news detection. In particular, the research focuses on the performance comparison between Random Forest, a widely used machine learning algorithm, and Long Short-Term Memory (LSTM), a deep learning architecture designed for sequential text processing. Random Forest utilizes textual features extracted through TF-IDF vectorization, while LSTM leverages word embeddings to capture contextual semantics and long-term dependencies within textual data. The proposed framework includes data collection from benchmark datasets, text preprocessing, feature extraction, model training, and evaluation using performance metrics such as accuracy, precision, recall, and F1-score. The results demonstrate that deep learning models capture contextual patterns more effectively, whereas traditional machine learning models provide faster training and interpretability. The study highlights the importance of combining linguistic features with deep contextual learning to improve misinformation detection. Furthermore, it identifies research gaps and future opportunities, including multilingual detection systems, transformer-based models, and real-time deployment for social media monitoring.

Keywords: Fake News Detection, Machine Learning, LSTM, Random Forest, Natural Language Processing, Social Media Analysis.

I INTRODUCTION

The emergence of social media platforms has drastically changed the information ecosystem by enabling rapid dissemination of news and opinions across digital networks. While this transformation has improved access to information, it has also led to the proliferation of fake news and disinformation, which pose serious threats to public trust, political stability, and social harmony [1]. Fake news refers to deliberately fabricated or misleading information presented as legitimate news content [2]. With billions of users generating and sharing content every day, identifying false information has become increasingly difficult using traditional verification methods [3]. Manual fact-checking processes performed by journalists and professional organizations are often slow and incapable of addressing the scale and speed of online information flow [4]. As a result, automated detection mechanisms based on computational intelligence have gained significant attention in recent years [5]. Machine Learning (ML) techniques enable systems to learn patterns from historical data and classify news content as genuine or fake based on linguistic and contextual features [6]. Natural Language Processing (NLP) plays a crucial role in this process by analyzing textual characteristics such as word usage, sentence structure, sentiment, and semantic meaning [7]. Early research in fake news detection relied primarily on traditional machine learning algorithms such as Naïve Bayes, Support Vector Machines, and Decision Trees [8]. These models typically use feature engineering techniques such as TF-IDF or bag-of-words to convert textual data into numerical representations suitable for classification tasks [9]. Although these approaches have demonstrated reasonable performance, they often struggle to capture deeper contextual relationships within complex textual data [10]. Consequently, deep learning techniques have been increasingly explored to overcome these limitations and improve detection accuracy [11].

Deep learning models such as Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM) networks are capable of learning long-range dependencies in sequential data, making them particularly suitable for analyzing news articles and social media posts [12]. LSTM models can capture contextual semantics and temporal relationships between words, which helps identify subtle patterns associated with misleading or deceptive content [13]. In addition, the use of word embeddings such as Word2Vec and GloVe enables models to understand semantic similarities between words and phrases [14]. Despite these advancements, fake news detection remains a challenging problem due to the evolving nature of misinformation strategies and the presence of sarcasm, ambiguity, and complex narrative structures [15]. This research therefore focuses on evaluating and comparing the performance of traditional machine learning and deep learning techniques in detecting fake news. Specifically, the study analyzes Random Forest and LSTM models using benchmark datasets and standardized evaluation metrics. By systematically examining these approaches, the research aims to identify the strengths and limitations of each technique and provide insights into developing more robust and scalable misinformation detection systems.

II LITERATURE SURVEY

Researchers have extensively explored machine learning techniques for detecting fake news in digital media environments. Early studies applied traditional classification algorithms such as Naïve Bayes and Support Vector Machines to identify deceptive news content based on linguistic and statistical features [16]. These approaches primarily relied on feature engineering techniques such as n-grams, term frequency analysis, and TF-IDF representations to convert textual data into structured formats suitable for classification tasks [17]. Decision Trees and Random Forest algorithms were also widely used due to their ability to handle high-dimensional data and provide interpretable classification rules [18]. Random Forest, in particular, gained popularity because of its ensemble learning capability, which combines multiple decision trees to improve prediction accuracy and reduce overfitting [19]. Several studies demonstrated that ensemble models outperform individual classifiers in identifying fake news across different datasets [20]. In addition to textual analysis, researchers also explored social context features such as user behavior, propagation patterns, and network structures to enhance detection accuracy [21]. Graph-based models and propagation analysis techniques have been applied to study how misinformation spreads across social networks [22]. However, these traditional machine learning approaches depend heavily on manually engineered features and may fail to capture deeper semantic relationships present in textual data [23].

To address these limitations, recent research has increasingly focused on deep learning models capable of automatically learning complex representations from raw text data [24]. Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM) architectures have been widely used for sequence modeling tasks, including fake news detection [25]. These models analyze word sequences and contextual relationships within news articles to identify subtle linguistic cues associated with misinformation [26]. Convolutional Neural Networks (CNNs) have also been applied to extract hierarchical features from textual data, improving classification performance in many studies [27]. Furthermore, transformer-based architectures such as BERT have shown significant improvements in natural language understanding tasks and are increasingly used in misinformation detection research [28]. Despite the success of deep learning models, challenges remain in terms of computational complexity, dataset imbalance, and generalization across different domains [29]. Consequently, hybrid approaches that combine machine learning techniques with deep learning models are being explored to leverage the advantages of both methodologies [30]. These studies highlight the need for scalable and robust frameworks capable of detecting fake news effectively in dynamic social media environments.

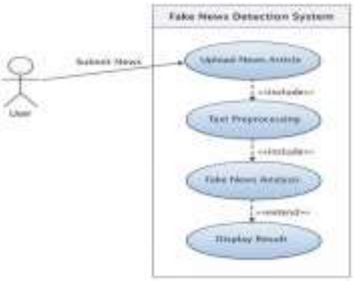
III METHODOLOGY

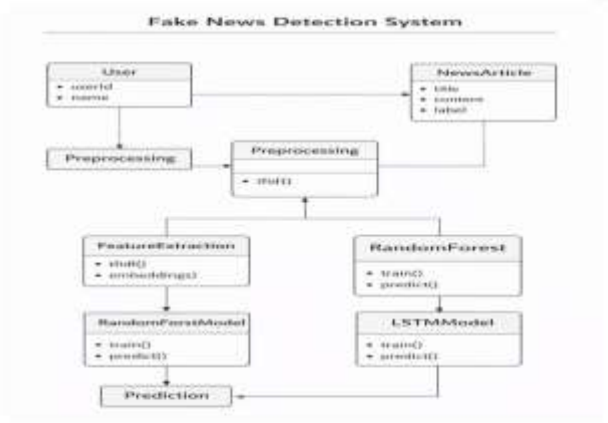
The proposed methodology for fake news detection follows a structured machine learning pipeline consisting of data collection, preprocessing, feature extraction, model training, and evaluation. Initially, datasets containing labeled fake and real news articles are collected from benchmark sources such as LIAR, ISOT, or FakeNewsNet. These datasets provide diverse textual samples that enable effective model training and testing. After data collection, preprocessing is performed to clean and standardize the textual information. This stage includes removing punctuation, converting text to lowercase, eliminating stop words, and performing tokenization to break sentences into individual words or tokens. Additional normalization techniques such as stemming or lemmatization may be applied to reduce words to their base forms and improve model generalization. Once the text is preprocessed, feature extraction techniques are applied to convert textual data into numerical representations suitable for machine learning algorithms. For the Random Forest model, the Term Frequency-Inverse Document Frequency (TF-IDF) technique is used to represent the importance of words in each document relative to the entire dataset. In contrast, the LSTM model utilizes word embedding techniques such as Word2Vec or GloVe to capture semantic relationships between words and generate dense vector representations. After feature extraction, the dataset is divided into training and testing subsets to evaluate model performance. The Random

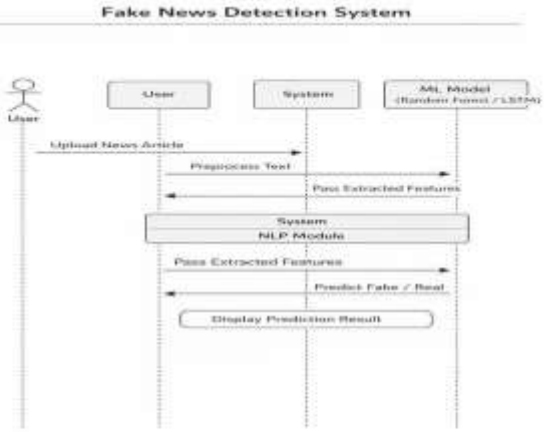
Forest classifier is trained using TF-IDF features to classify news articles as real or fake based on decision tree ensembles. Simultaneously, the LSTM model processes sequences of word embeddings to learn contextual patterns and long-term dependencies within the textual data. Both models are trained using supervised learning techniques and optimized through parameter tuning. Finally, the models are evaluated using standard performance metrics such as accuracy, precision, recall, and F1-score to determine their effectiveness in detecting fake news.

IV SYSTEM DESIGN

The system design for fake news detection is based on a modular architecture that integrates data processing, feature extraction, machine learning models, and evaluation components. The first module of the system focuses on data acquisition and preprocessing. In this stage, datasets containing labeled news articles are collected from reliable benchmark sources such as LIAR or ISOT. These datasets consist of textual news content categorized as either genuine or fake. After data collection, preprocessing is performed to clean the data and remove irrelevant information. This process includes converting text to lowercase, removing punctuation and special characters, eliminating stop words, and tokenizing sentences into individual words. The purpose of this stage is to standardize the dataset and ensure that the input data is suitable for further analysis. Once preprocessing is complete, the cleaned textual data is passed to the feature extraction module. This module converts textual information into numerical representations that can be processed by machine learning algorithms. For the Random Forest model, TF-IDF vectorization is applied to measure the importance of words across documents. The resulting vectors represent each news article as a numerical feature set that can be used for classification.







The second part of the system design focuses on model training, prediction, and evaluation. In this stage, two different models are implemented to detect fake news: a Random Forest classifier and an LSTM-based deep learning model. The Random Forest model is trained using TF-IDF features and operates as an ensemble learning method that constructs multiple decision trees and combines their predictions to improve classification accuracy. This approach is effective for handling structured feature vectors and provides interpretable results. In contrast, the LSTM model processes sequential text data using word embeddings to capture contextual relationships between words in a news article. The LSTM architecture consists of memory cells and gating mechanisms that allow the model to retain important information over long sequences. After training, both models are evaluated using testing data to measure their performance. Evaluation metrics such as accuracy, precision, recall, and F1-score are used to compare the effectiveness of the two approaches. The system ultimately identifies the model that provides the most reliable performance for fake news detection while maintaining computational efficiency.

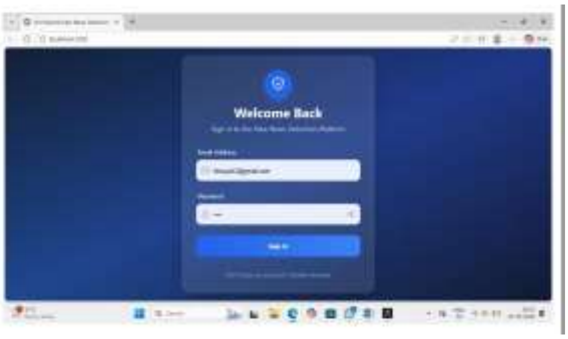
V PROPOSED SYSTEM

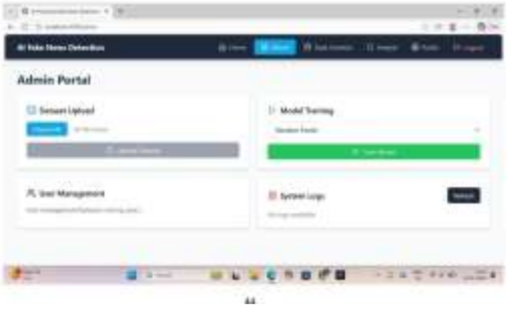
The proposed system aims to develop an efficient and scalable framework for detecting fake news using a combination of machine learning and deep learning techniques. The system integrates Random Forest and LSTM models to analyze textual data from news articles and social media posts. The main objective of the proposed framework is to improve the accuracy and reliability of misinformation detection by leveraging the strengths of both traditional machine learning algorithms and modern deep learning architectures. The system begins with a comprehensive data collection process that gathers news content from benchmark datasets containing labeled examples of real and fake news. These datasets are then processed through a preprocessing pipeline that removes noise and standardizes the text. Preprocessing steps include tokenization, stop-word removal, text normalization, and stemming. After preprocessing, the system applies feature extraction techniques to transform textual data into numerical representations suitable for classification models. For the Random Forest model, TF-IDF is used to capture the frequency and importance of words within the dataset. For the LSTM model, word embedding techniques are employed to capture semantic relationships between words and represent them as dense vectors.

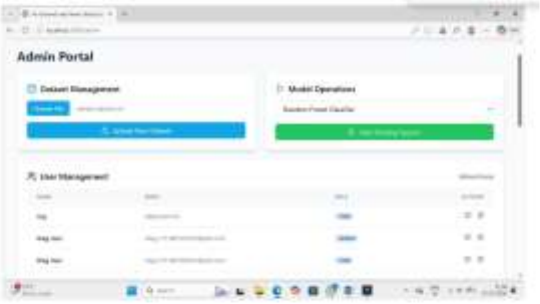
The core component of the proposed system is the classification module, which consists of two parallel models designed to detect fake news using different approaches. The Random Forest model uses ensemble learning to classify news articles based on extracted TF-IDF features. This model provides high interpretability and fast training times, making it suitable for large datasets. On the other hand, the LSTM model processes sequences of word embeddings to understand contextual relationships and long-term dependencies within textual data. This enables the model to detect subtle linguistic patterns associated with misinformation. After training both models, the system evaluates their performance using metrics such as accuracy, precision, recall, and F1-score. The results of the evaluation are compared to determine which model performs better in identifying fake news. The proposed system also supports scalability and adaptability by allowing integration with advanced models such as BERT or transformer-based architectures in the future. By combining machine learning and deep learning techniques, the proposed framework aims to provide a robust solution for combating misinformation and improving the reliability of online information ecosystems.

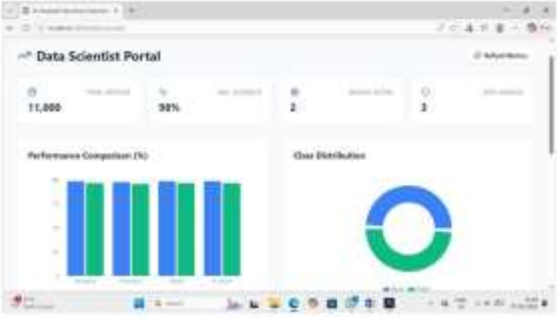
VI RESULTS & DISCUSSION

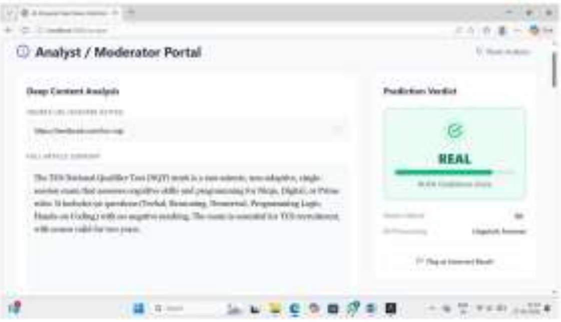
The experimental evaluation compares the performance of the Random Forest and LSTM models in detecting fake news using benchmark datasets. The Random Forest model demonstrates strong performance in terms of classification accuracy and computational efficiency due to its ensemble learning structure and ability to handle high-dimensional TF-IDF feature vectors. It also provides interpretable decision boundaries that help understand the importance of specific textual features in classification. However, the LSTM model achieves improved performance in capturing contextual semantics and sequential relationships within textual data. By analyzing word embeddings and long-term dependencies, the LSTM model is capable of identifying complex linguistic patterns that may indicate misleading information. The evaluation results show that while Random Forest provides faster training and reliable baseline performance, the LSTM model generally achieves higher recall and F1-score in detecting fake news. These findings suggest that deep learning models offer significant advantages in handling complex textual patterns present in misinformation.

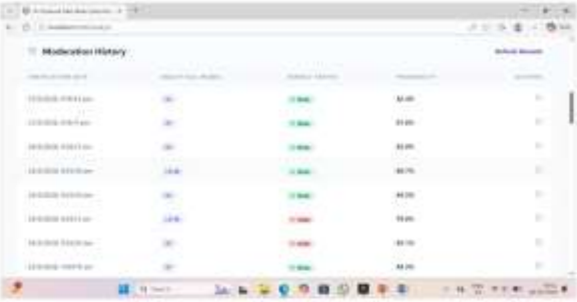


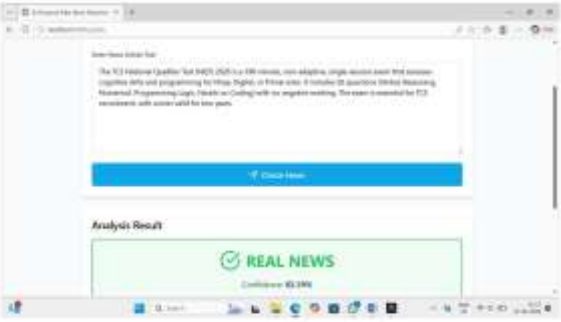












VII CONCLUSION

The rapid spread of misinformation through digital media platforms has become a major global challenge, affecting political systems, public trust, and societal stability. As social media continues to expand, the volume of information generated daily makes manual fact-checking increasingly impractical. This research study explored the application of machine learning and deep learning techniques for detecting fake news in online environments. Specifically, the study compared the effectiveness of the Random Forest algorithm and the Long Short-Term Memory (LSTM) neural network for classifying news articles as real or fake. The experimental analysis demonstrated that both models offer significant advantages in automated misinformation detection. Random Forest provides strong baseline performance, computational efficiency, and model interpretability, making it suitable for large-scale implementations. In contrast, the LSTM model excels at capturing contextual relationships and sequential dependencies in textual data, enabling it to identify subtle linguistic patterns associated with deceptive content. The results highlight the importance of combining traditional machine learning methods with advanced deep learning techniques to improve detection accuracy and system robustness. Furthermore, the study emphasizes the need for continuous research in this area due to the evolving nature of misinformation strategies. Future work can focus on integrating transformer-based models such as BERT, utilizing multilingual datasets, and developing real-time detection systems capable of monitoring social media platforms. By advancing automated fake news detection technologies, researchers and organizations can contribute to creating a more trustworthy digital information environment.

REFERENCES

1. Shu, K., Sliva, A., Wang, S., Tang, J., & Liu, H. (2017). Fake news detection on social media. SIGKDD Explorations, 19(1), 22–36.

2. Allcott, H., & Gentzkow, M. (2017). Social media and fake news in the 2016 election. Journal of Economic Perspectives, 31(2), 211–236.

3. Vosoughi, S., Roy, D., & Aral, S. (2018). The spread of true and false news online. Science, 359(6380), 1146–1151.

4. Lazer, D., et al. (2018). The science of fake news. Science, 359(6380), 1094–1096.

5. Zhou, X., & Zafarani, R. (2020). A survey of fake news detection. ACM Computing Surveys, 53(5), 1–36.

6. Castillo, C., Mendoza, M., & Poblete, B. (2011). Information credibility on Twitter. WWW Conference.

7. Bird, S., Klein, E., & Loper, E. (2009). Natural Language Processing with Python. O'Reilly.

8. Wang, W. Y. (2017). Liar dataset for fake news detection. ACL Proceedings.

9. Manning, C., Raghavan, P., & Schütze, H. (2008). Introduction to Information Retrieval. Cambridge University Press.

10. Kotsiantis, S. (2007). Supervised machine learning review. Informatica.

11. Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep Learning. MIT Press.

12. Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. Neural Computation.

13. Goldberg, Y. (2017). Neural Network Methods for NLP. Morgan & Claypool.

14. Mikolov, T., et al. (2013). Efficient estimation of word representations. ICLR.

15. Pennington, J., Socher, R., & Manning, C. (2014). GloVe word embeddings. EMNLP.

16. Conroy, N., Rubin, V., & Chen, Y. (2015). Automatic deception detection. Proceedings of the Association for Information Science.

17. Rubin, V., Chen, Y., & Conroy, N. (2016). Deception detection for news. Journal of the Association for Information Science.

18. Breiman, L. (2001). Random forests. Machine Learning.

19. Liaw, A., & Wiener, M. (2002). Classification using randomForest. R News.

20. Dietterich, T. (2000). Ensemble methods in machine learning. Multiple Classifier Systems.

21. Jin, F., et al. (2016). News propagation modeling on social networks. IEEE Transactions on Knowledge and Data Engineering.

22. Wu, L., et al. (2019). Mining misinformation in social media. ACM SIGKDD.

23. Zhang, X., & Ghorbani, A. (2020). Fake news detection review. Journal of Big Data.

24. Ruchansky, N., Seo, S., & Liu, Y. (2017). CSI: Fake news detection model. CIKM.

25. Ma, J., et al. (2016). Detecting rumors with recurrent neural networks. IJCAI.

26. Shu, K., et al. (2019). Beyond news contents. ACM WSDM.

27. Kim, Y. (2014). Convolutional neural networks for sentence classification. EMNLP.

28. Devlin, J., et al. (2019). BERT: Pre-training of deep bidirectional transformers. NAACL.

29. Zhou, X., & Zafarani, R. (2019). Network-based fake news detection. ACM SIGKDD Explorations.

30. Monti, F., et al. (2019). Fake news detection using graph neural networks. IEEE Transactions on Neural Networks.