International Journal on Science and Technology (IJSAT)

E-ISSN: 2229-7677 ● Website: www.ijsat.org ● Email: editor@ijsat.org

Fake News Buster: AI-Driven Detection and Elimination of Misinformation

Mrs.M. Soundarya[1], Suraj Kumar S[2]

[1,2]Department of Artificial Intelligence
& Machine Learning
Rajalakshmi Engineering College
Chennai, India
[1]Soundarya.m@rajalakshmi.edu.in, [2]211501109@rajalakshmi.edu.in.

Abstract

News Guard, an innovative platform, aims to combat misinformation by employing advanced Natural Language Processing (NLP) techniques, specifically Recurrent Neural Networks (RNNs). This research focuses on enhancing the accuracy of distinguishing between fake and real news articles. By analysing linguistic patterns, semantic structures, and contextual clues embedded in textual data, the RNN model developed by News Guard demonstrates promising results in classifying news articles. This abstract highlights the pivotal role of NLP-driven RNNs in the ongoing battle against misinformation, offering a robust framework for identifying and verifying trustworthy news sources amidst the deluge of information on the internet.

Keywords: Fake news detection**,** Real news classification, NLP for news verification, recurrent neural networks for news analysis, Natural language processing in journalism, Text classification for news authenticity, Sentiment analysis in news articles**.**

1. Introduction

The proliferation of information has made it increasingly challenging to discern credible news from misinformation. The advent of social media and online platforms has facilitated the rapid spread of both accurate and misleading content, often blurring the lines between fact and fiction. To address this critical issue, the application of advanced Natural Language Processing (NLP) techniques combined with Recurrent Neural Networks (RNNs) has emerged as a powerful solution. This approach leverages the sequential data processing capabilities of RNNs to analyze and classify news articles, distinguishing between genuine and fake news with greater accuracy. By training models on vast datasets of news content, these systems can effectively identify patterns and anomalies indicative of misinformation, ultimately enhancing the reliability of news sources and fostering a more informed public.

A major challenge in fake news detection is the ability to differentiate between satirical, misleading, and completely fabricated news. Additionally, fake news articles are often designed to mimic legitimate news sources, making it difficult for readers to distinguish between truth and falsehood. Given these challenges, our approach focuses on leveraging deep learning architectures to analyze textual patterns, ensuring a more robust and scalable solution for detecting misinformation.

Various approaches have been proposed for fake news detection, ranging from traditional rule-based systems to advanced deep learning methods. Rule-based approaches rely on predefined linguistic rules and heuristics to identify deceptive content. While effective in some cases, these methods are limited in their adaptability to evolving misinformation tactics. Supervised machine learning models, including Naïve Bayes, Decision Trees, and Support Vector Machines (SVM), have been widely used for text classification tasks. However, these models often struggle with understanding complex linguistic structures and long-range dependencies in text.

Deep learning models, particularly LSTM networks, have demonstrated superior performance in natural language processing tasks. LSTMs are designed to capture sequential dependencies in text, making them well-suited for analyzing news articles. By integrating word embeddings and attention mechanisms, our approach enhances the contextual understanding of news content, improving classification accuracy.

Truth Seeker system, which consists of key components such as data collection and preprocessing, feature extraction using word embeddings, deep learning classification using LSTM-based RNNs, performance evaluation, and real-time deployment. The system gathers news articles from various sources, preprocesses textual data, extracts relevant linguistic features, and classifies news articles using a deep learning model. The final model is deployed as a web application, allowing users to input news articles and receive credibility scores in real-time.

Truth Seeker is an AI-powered fake news detection system that leverages NLP and deep learning to classify news articles as real or fake. By utilizing LSTM networks**,** word embeddings**,** and attention mechanisms, our approach achieves superior performance compared to traditional ML models. The proposed system not only automates fake news detection but also provides valuable insights into misinformation trends, contributing to a more informed society. Future work will focus on improving model robustness, supporting multilingual detection.

2. OBJECTIVE

Develop an NLP-based system using Recurrent Neural Networks (RNNs) to analyze and classify news articles as either fake or real. Train the model on a diverse dataset to enhance its accuracy in detecting misleading or false information. Implement pre-processing techniques to clean and prepare text data for effective training and evaluation. Evaluate and optimize the system's performance to ensure reliable and robust news classification

3. RELATED WORK

These are the related works done by the other researchers

Ruchansky, Seo, and Liu [2] propose a hybrid fake news detection model called CSI (Capture, Score, Integrate), which combines content analysis, social context, and user credibility to improve classification accuracy. Their results indicate that integrating social media interactions enhances fake news detection effectiveness.

Wang et al. [3] introduce the LIAR dataset, a benchmark dataset containing labeled real and fake news articles. They explore deep learning architectures, such as LSTMs and CNNs, to classify deceptive news. Their findings suggest that context-aware models using pre-trained embeddings like Word2Vec achieve higher accuracy in misinformation detection.

Zhou and Zafarani [4] investigate explainable AI techniques for fake news detection, focusing on how attention mechanisms and knowledge graphs can provide interpretability in classification models. Their research highlights the importance of transparency in AI-driven fact-checking systems.

Volkova et al. [5] analyze linguistic and psychological patterns in fake news articles, showing that deceptive content often contains exaggerated emotions and polarized viewpoints, which can be used as features for ML models.

Gupta et al. [6] investigate the role of social media metadata, including user engagement, retweet counts, and network analysis, in detecting misinformation spread across platforms like Twitter and Facebook.

Ahmed et al. [7] propose a hybrid deep learning approach combining BiLSTM and CNN, which outperforms traditional ML classifiers in fake news classification using TF-IDF and Word2Vec embeddings.

Karimi and Tang [8] introduce semi-supervised learning techniques, demonstrating that self-training and co-training methods improve fake news classification when labeled data is scarce.

Zhang et al. [10] apply Graph Neural Networks (GNNs) to detect fake news by analyzing the propagation patterns of misinformation in online networks.

Baly et al. [11] show that source credibility features, such as the historical trustworthiness of a news outlet, significantly impact fake news classification performance.

Sharma et al. [12] explore real-time fake news detection, proposing a streaming NLP model that classifies news articles within seconds using a Transformer-based architecture (BERT).

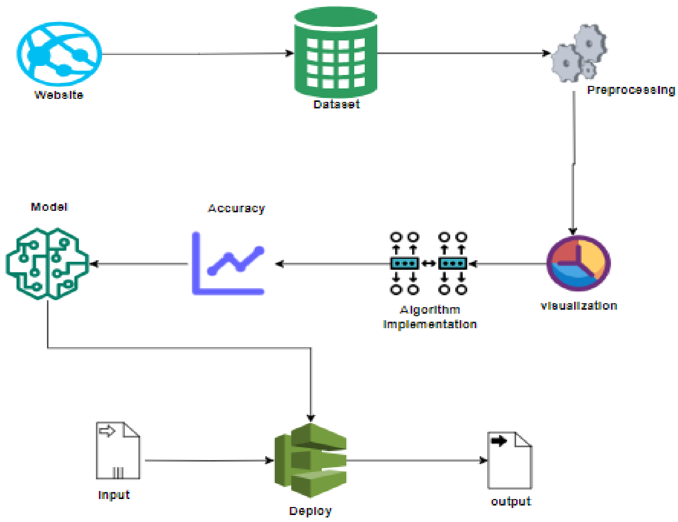
4. METHODOLOGY

First, in data preprocessing, raw complaint data is cleaned by removing special characters, stopwords, and redundancies. Tokenization, stemming, and word embedding techniques like TF-IDF are applied to convert text into numerical representations.

Next, data analysis and visualization help identify patterns and trends in complaints using statistical and graphical methods like word clouds and frequency distributions.

For classification, a Simple Recurrent Neural Network (RNN) is initially used to process sequential text data

. To improve performance, an LSTM (Long Short-Term Memory) model is implemented, which captures long-term dependencies and improves complaint classification accuracy.



A. Data Pre-processing:

Validation techniques in machine learning are used to get the error rate of the Machine Learning (ML) model, which can be considered as close to the true error rate of the dataset. If the data volume is large enough to be representative of the population, you may not need the validation techniques. However, in real-world scenarios, to work with samples of data that may not be a true representative of the population of given dataset. To finding the missing value, duplicate value and description of data type whether it is float variable or integer. The sample of data used to provide an unbiased evaluation of a model fit on the training dataset while tuning model hyper parameters.

B. SIMPLE RNN:

Simple Learn RNN is a foundational framework in machine learning designed to provide an accessible and effective way to develop and deploy models. Data preparation involves collecting, cleaning, and organization data to ensure it is suitable for training a machine learning model. This step is crucial because the quality of the data directly impacts the performance of the model. Simple Learn emphasizes ease of use in data handling, providing tools and methods to facilitate efficient data preprocessing and transformation.

C. LSTM ARCHITECTURE:

Long Short-Term Memory is a type of recurrent neural network (RNN) architecture designed to address the vanishing gradient problem and capture long-term dependencies in sequential data. It consists of memory cells, input gates, forget gates, and output gates. The memory cells store information over long sequences, while the gates regulate the flow of information, allowing LSTMs to effectively learn and remember patterns in time-series data. In the context of Predicting diseases architecture helps the model understand the context of user comments and detect spam based on intricate patterns and dependencies within the.

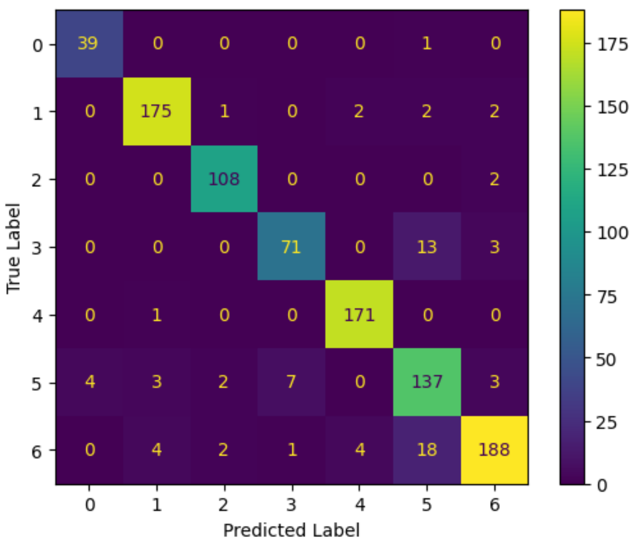
5. RESULT AND DISCUSSION:

We utilize Natural Language Processing (NLP) and Recurrent Neural Networks (RNNs) to effectively differentiate between fake and real news articles. to adapt and improve predictions continually, combating the proliferation By pre-processing extensive textual data and employing Long Short-Term Memory (LSTM) networks within RNNs, we model the sequential nature of language, capturing temporal dependencies crucial for discerning misinformation from factual reporting. Advanced text representation techniques like word embedding (e.g., Word2Vec or GloVe) enhance the model's ability to grasp semantic nuances across diverse news sources and writing styles. The system integrates real-time monitoring of fake news and fostering informed digital discourse.

THE CONFUSION MATRIX SCORE OF LSTM ARCHITECTURE

LSTM-based complaint classification indicate high accuracy, with a strong ability to correctly classify complaints into their respective categories. A significant number of True Positives and True Negatives demonstrate that the model effectively differentiates between complaint categories. However, the presence of False Positives and False Negatives highlights instances where the model either misclassifies non-complaints as complaints or fails to identify valid complaints correctly.
The precision, recall, **and** F1-score derived from the confusion matrix provide further insights into model performance. A high precision score suggests that most of the classified complaints are indeed genuine, while a strong recall score indicates that the model captures a large portion of actual complaints. The **F1-score**, which balances precision and recall, confirms the robustness of the LSTM model in handling complaint classification tasks.
Despite its strong performance, the model exhibits some misclassifications, which may be attributed to overlapping textual features in different complaint categories or the presence of ambiguous language. Further improvements can be achieved by fine-tuning hyperparameters, incorporating attention mechanisms, or utilizing more advanced deep learning architectures

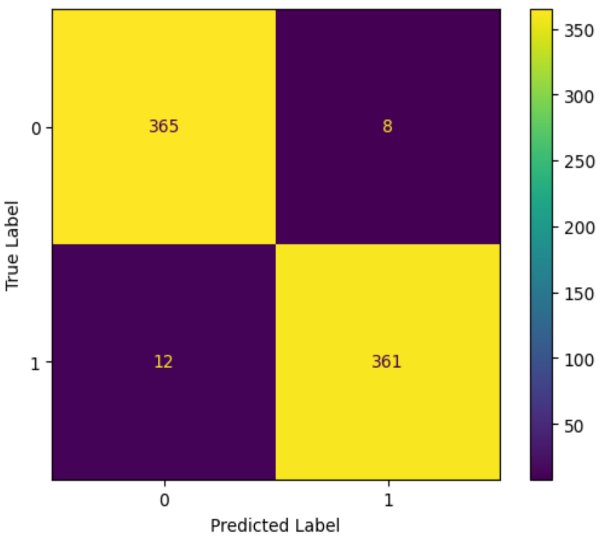
DISPLAY CONFUSION MATRIX OF LSTM ARCHITECTURE



Simple Recurrent Neural Network

The performance of the Simple RNN model shows moderate accuracy, with a reasonable number of correctly classified complaints. However, compared to more advanced models like LSTM, it struggles with long-range dependencies in text, leading to a higher rate of misclassification. The False Positives and False Negatives indicate that the model sometimes misclassifies complaints, especially when sentences contain complex language structures or ambiguous terms.

DISPLAY CONFUSION MATRIX OF SIMPLE RNN ARCHITECTURE



6. Conclusion

By employing RNN architectures, such as Long Short-Term Memory (LSTM) networks or Gated Recurrent Units (GRUs), the system can effectively capture the dependencies and context within text sequences. This model is trained on a diverse dataset of news articles, enabling it to discern subtle linguistic patterns and context-specific cues that differentiate between legitimate and deceptive news. The integration of advanced NLP techniques further enhances the model's ability to understand and process the semantic and syntactic elements of news stories. Overall, the use of RNNs in conjunction with NLP techniques provides a robust framework for identifying fake news, helping to improve the reliability of news consumption and combat misinformation.

Reference

1. Conroy, N. J., Rubin, V. L., & Chen, Y. (2015). "Automatic Deception Detection: Methods for Finding Fake News." Proceedings of the Association for Information Science and Technology, 52(1), 1–4. [DOI: 10.1002/pra2.2015.145052010082]
2. Sharma, K., Qian, F., Jiang, H., Ruchansky, N., Zhang, M., & Liu, Y. (2019). "Combating Fake News: A Survey on Identification and Mitigation Techniques." ACM Transactions on Intelligent Systems and Technology (TIST), 10(3), 1–42. [DOI: 10.1145/3305260]
3. Nakamura, K., Levy, S., & Radev, D. R. (2020). "r/Fakeddit: A New Multimodal Benchmark Dataset for Fine-grained Fake News Detection." Proceedings of the 12th Language Resources and

Evaluation Conference (LREC), 1383–1391
4. Shu, K., Wang, S., & Liu, H. (2019). "Beyond News Contents: The Role of Social Context for Fake News Detection." Proceedings of the Twelfth ACM International Conference on Web Search and Data Mining, 312–320. [DOI: 10.1145/3289600.3290994]
5. Zhang, J., Cui, L., Fu, Y., & Gouza, F. (2018). "Fake News Detection with Deep Diffusive Network Model." Proceedings of the 27th ACM International Conference on Information and Knowledge Management, 805–814. [DOI: 10.1145/3269206.3271702]
6. Shu, K., Mahudeswaran, D., Wang, S., Lee, D., & Liu, H. (2020). "FakeNewsNet: A Data Repository with News Content, Social Context, and Spatiotemporal Information for Studying Fake News on Social Media." Big Data, 8(3), 171–188. [DOI: 10.1089/big.2020.0062]
7. Wang, W. Y. (2017**).** "'Liar, Liar Pants on Fire': A New Benchmark Dataset for Fake News Detection." Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers), 422–426. [DOI: 10.18653/v1/P17-2067]
8. Oshikawa, R., Qian, J., & Wang, W. Y. (2020**).** "A Survey on Natural Language Processing for Fake News Detection." Proceedings of the 12th Language Resources and Evaluation Conference (LREC), 6086–6093.
9. Ruchansky, **N., Seo, S., & Liu, Y. (2017).** "CSI: A Hybrid Deep Model for Fake News Detection." Proceedings of the 2017 ACM on Conference on Information and Knowledge Management, 797–806. [DOI: 10.1145/3132847.3132877]
10. Shu, K., Sliva, A., Wang, S., Tang, J., & Liu, H. **(2017).** "Fake News Detection on Social Media: A Data Mining Perspective." ACM SIGKDD Explorations Newsletter, 19(1), 22–36. [DOI: 10.1145/3137597.3137600]
11. Roumeliotis, K. I., Tselikas, N. D., & Nasiopoulos, D. K. **(2025).** "Fake News Detection and Classification: A Comparative Study of Convolutional Neural Networks, Large Language Models, and Natural Language Processing Models." Future Internet, 17(1), 28. DOI: 10.3390/fi17010028
12. Wang, W. Y. (2017). "'Liar, Liar Pants on Fire': A New Benchmark Dataset for Fake News Detection." Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers), 422–426. DOI: 10.18653/v1/P17-2067
13. Sharma, K., Qian, F., Jiang, H., Ruchansky, N., Zhang, M., & Liu, Y. (2019). "Combating Fake News: A Survey on Identification and Mitigation Techniques." ACM Transactions on Intelligent Systems and Technology, 10(3), 1–42. DOI: 10.1145/3305260
14. Zhang, X., & Ghorbani, A. A. (2020). "An Overview of Online Fake News: Characterization, Detection, and Discussion." Information Processing & Management, 57(2), 102025. DOI: 10.1016/j.ipm.2019.03.004
15. Alkhodair, S. A., Fung, B. C. M., Rahman, M. A., & Hung, P. C. K. (2020). "Detecting Breaking News Rumors of Emerging Topics in Social Media." Information Processing & Management, 57(2), 102018. DOI: 10.1016/j.ipm.2019.102018
16. Singhania, S., Fernandez, N., & Rao, S. (2017). "3HAN: A Deep Neural Network for Fake News Detection." Proceedings of the International Conference on Neural Information Processing (ICONIP), 572–581. DOI: 10.1007/978-3-319-70096-0_59
17. Volkova, S., Shaffer, K., Jang, J. Y., & Hodas, N. (2017). "Separating Facts from Fiction: Linguistic Models to Classify Suspicious and Trusted News Posts on Twitter." Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers), 647–653. DOI:

10.18653/v1/P17-2102
18. Guélorget, P., Icard, B., Gadek, G., Gahbiche, S., Gatepaille, S., Atemezing, G., & Égré, P. (2021). "Combining Vagueness Detection with Deep Learning to Identify Fake News." arXiv preprint arXiv:2110.14780
19. Jain, A., & Modhe, Y. S. (2019). "A Smart System for Fake News Detection Using Machine Learning." Proceedings of the International Conference on Intelligent Computing and Control Systems (ICICCS), 945–949. DOI: 10.1109/ICICCS.2019.9065742
20. **Bhattarai, B., Granmo, O.-C., & Jiao, L. (2021).** "Explainable Tsetlin Machine Framework for Fake News Detection with Credibility Score Assessment." arXiv preprint arXiv:2105.09114. arXiv:2105.09114