Crossian Resonance

A Multidisciplinary Research Journal **ISSN 0976-5417**
(A refereed Biannual Published in June and December)
Vol. 15, No.2, December 2024
https://research.holycrossngl.edu.in/

Lexicon Based LSTM Framework for Context-Aware Fake News Detection on Social Media Platforms

***[1]Harilakshmi V.S., [2]Sahaya Rose Vigita E., Konsia M., Ditty Mol S., Shunmuga Priya E., Sahaya Jelastin Mary A.**

[1]Department of Artificial Intelligence and Data Science, Holy Cross College (Autonomous), Nagercoil - 629 004

[2]Department of Computer Science, Holy Cross College (Autonomous), Affiliated to Manonmaniam Sundaranar University, Tirunelveli - 629 012

ABSTRACT

The rise of fake news in social media platforms threatens information accuracy and public trust. In order to enhance the identification of fake news, this study presents a Lexicon-Based Long ShortTerm Memory (LSTM) framework that integrates social context and content analysis. The model uses a lexicon of keywords, sentiment cues, and deceptive language to enhance feature analysis. It also incorporates social context, such as the reliability of sources, user engagement, and how news spreads, to better understand misinformation patterns. The LSTM processes data in sequence to identify patterns in both content and its dissemination. Test on publicly available standard dataset shows that this method outperforms traditional approaches with higher accuracy. Adding lexicon features and social context makes the model better at detecting fake news. This framework offers an effective and scalable way to reduce the spread of misinformation and supports efforts in content moderation and social media regulation.

Keywords: *Fake News Detection, Lexicon-Based Approach, Long Short-Term Memory (LSTM), Social Media Platforms*

1. Introduction

The rapid rise of social media platforms has completely transformed how information is created, shared, and consumed. While these platforms enable instant communication and global connectivity, they have also transformed into breeding grounds for the spread of fake news. The dissemination of misinformation on social media undermines public trust, influences opinions, and, in extreme cases, leads to significant societal and political consequences. Addressing these challenges requires effective and scalable tools capable of distinguishing between genuine and false information [1]. Traditional models of fake news detection, primarily focused on textual content analysis [2], often overlook the social context in which the news is shared. However, the propagation patterns, user interactions, and

Corresponding Author:** **harilakshmi@holycrossngl.edu.in

credibility of sources have a critical role in distinguishing authentic news from that of fake news [3]. Additionally, the nuanced use of language in misinformation calls for advanced methods that go beyond basic text processing.

This paper proposes a Lexicon-Based Long Short-Term Memory (LSTM) framework that integrates both the content and the social context for the detection of fake news. The model uses a curated lexicon containing sentiment markers, domain-specific keywords, and deceptive language indicators to enhance the analysis of textual features. Furthermore, the framework incorporates social context factors, such as user engagement metrics, source credibility, and news-sharing patterns, that provides a holistic view of misinformation dynamics.

Long Short-Term Memory (LSTM) networks have demonstrated significant improvements in capturing semantic and temporal dependencies in textual data. However, many of these models focus solely on textual content, overlooking the critical role of social context—such as source credibility, user engagement patterns, and propagation dynamics—in distinguishing fake news from genuine content. To address these limitations, hybrid approaches have emerged. For example, works by [4] [5] have highlighted the importance of incorporating social context into fake news detection models. These studies emphasize that user behavior, source characteristics, and dissemination patterns provide valuable cues for identifying misinformation. Similarly, lexicon-based techniques have been employed to enhance feature extraction by leveraging domain-specific keywords and sentiment markers, as seen in the works of [6].

By combining the sequential processing capabilities of LSTM networks with enriched feature inputs [7] from the lexicon and social context, this framework offers a robust solution to fake news identification on social networking sites. This study demonstrates the efficacy of the proposed framework in identifying fake news with greater accuracy and reliability, contributing initiatives to combat misinformation in this digital age. Through this work, we aim to develop scalable, data-driven solutions to combat misinformation on social networking platforms.

2. Proposed Methodology

The proposed methodology centers around developing a Lexicon based Long Short-Term Memory (LSTM) based model for identifying fake news. The proposed system aims to enhance false news detection by incorporating sentiment analysis into a Long Short-Term Memory (LSTM) model. This system uses sentiment embeddings derived from a sentiment lexicon to improve word representation and overall classification accuracy.

The figure 1 illustrates the fake information detection model using a lexicon based Long Short-Term Memory (LSTM) model. The Fake news Web User Interface is the primary interface module for both system administrators and users. System Administrators can log in to access the training pipeline for detecting fake news. Users can interact with the system to check if a news post is fake or genuine after logging in or registering. Load News Dataset is the module where a system administrator loads a news dataset with labeled data (e.g., fake or genuine news). This data serves as input for the training phase in the fake news detection system model.

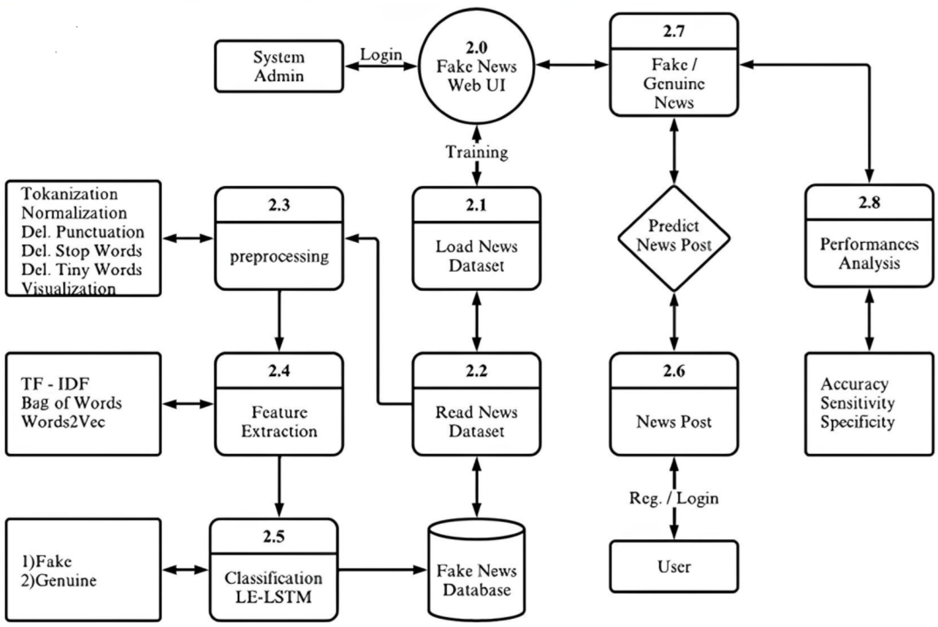

Figure 1. Lexicon Based LSTM Fake News Detection System

The Read News Dataset module processes the uploaded dataset for further analysis. It ensures that the data is correctly formatted for preprocessing and feature extraction. Fake News Database module stores processed news data, training datasets, and the outputs of classifications for reference and improvement of the detection model.

3. Preprocessing Module

In this module, the raw text data is prepared for feature extraction by applying the following:

Tokenization: Breaking of the text into individual words or tokens. Normalization: Standardizing text input, such as converting all words to lowercase. Removal of Punctuation: Eliminating punctuation marks to focus on meaningful content. Stop Word Removal: Removing frequently used words (e.g., "is," "and," "the") that do not add significant meaning. Removal of Tiny Words: Eliminating words with very few characters to reduce noise. Visualization: Visualizing the data (e.g., word clouds, frequency distribution) for better understanding.

4. Feature Extraction

Extracts numerical representations of the text data using methods such as: TF-IDF (Term Frequency-Inverse Document Frequency): Measures the importance of a term in a document relative to the dataset. Bag of Words: Represents text data as a collection of word occurrences. Word2Vec: Converts words into continuous vector representations based on their context.

Classification Using Lexicon Based LSTM

This module uses a Lexicon Based Long Short-Term Memory (LE-LSTM) neural network for classification. The Lexicon Enhancement involves adding sentiment analysis or semantic information from a lexicon to improve accuracy. The model through user engagement and keywords incorporates social context along with textual features in its analysis of fake news. The LSTM model processes sequential data to classify news posts as either Fake or Genuine. In module News Post Submission, users submit a news post and the system processes it and predicts whether the post is fake or genuine based on the trained model. In the Prediction Output module, after processing the input, the system outputs the prediction: as Fake News or as Genuine News. Figure 2 shows the output screen when genuine and fake messages are posted.

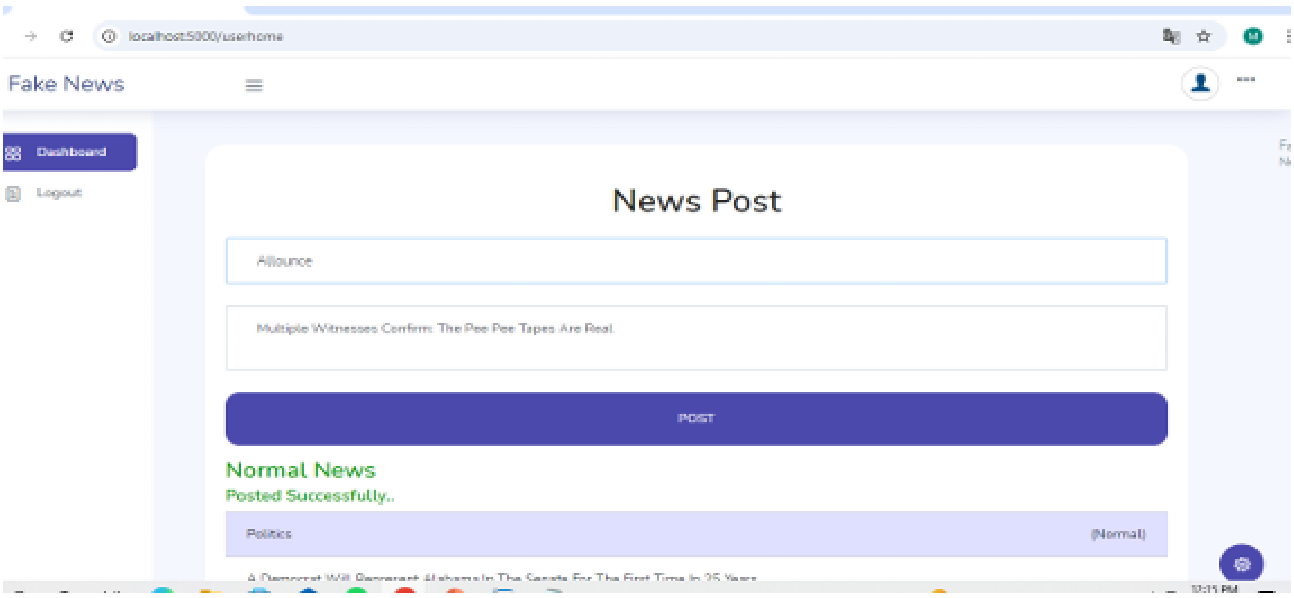

Fig. 2. Output Screen of a Genuine User Post

5. Conclusion

This paper introduces Lexicon-Based LSTM architecture for social media fake news detection. The model combines a lexicon of key phrases and deceptive language with the power of LSTM networks to analyze both the content and the social context in which news spreads. By considering factors like source credibility, user engagement, and how news is shared, our approach improves the accuracy of false news detection. Through its combination of semantic and contextual features, the proposed model exhibits superior output. Also, the proposed model outperforms traditional methods by better capturing the nuances of both language and social dynamics. This approach offers a scalable solution for applications like content moderation and fact-checking. Future work could expand the model to include other forms of media, like images and videos, and refine the lexicon to stay update with emerging trends in online content.

References

1. Gupta, S., Arora, A., & Kumaraguru, P. Analyzing the Impact of Feature Extraction Techniques on Fake News Detection. Applied Artificial Intelligence. 2021; 35(1): 58 - 80.
2. Wu, L., Morstatter, F., Hu, X., & Liu, H. Mining Misinformation in Social Media. Proceedings of the 25th International Conference on World Wide Web. 2016; 641 - 647.
3. Zhou, X., & Zafarani, R. Fake News: A Survey of Research, Detection Methods, and Opportunities. ACM Computing Surveys, 2020; 53(5): 1 - 40.
4. Shu, K., Sliva, A., Wang, S., Tang, J., & Liu, H. Fake news detection on social media: A data mining perspective. ACM SIGKDD Explorations Newsletter. 2019; 19(1): 22 - 36.
5. Wang, W. Y. Fake News Detection on Social Media: Challenges and Opportunities. IEEE Transactions on Computational Social Systems. 2020; 7(1): 56 - 67.
6. Zhou, X., & Zafarani, R. Fake News: A Survey of Research, Detection Methods, and Opportunities. arXiv preprint arXiv:1812.00315. 2018
7. Rashkin, H., Choi, E., Jang, J. Y., Volkova, S., & Choi, Y. Truth of Varying Shades: Analyzing Language in Fake News and Political Fact-Checking. Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing (EMNLP). 2017; 2931 - 2937.