RESEARCH ARTICLE OPEN ACCESS

NLP BASED NEWS CREDIBILITY ANALYSIS

Mrs. G. Meena *, Deepika M**, Elavarasi S**, Dineshkumar G**, Atchaya M**

* Assistant Professor - Computer Science and Engineering, R P Sarathy Institute of Technology, Salem

** UG Students -Computer Science and Engineering, R P Sarathy Institute of Technology, Salem

Tamil Nadu, India

Email: deepikamanikandan032@gmail.com

Abstract:

The rapid growth of social media and online news platforms has made it easier for fake news to spread. This trend threatens public trust, journalism, and democratic processes. Traditional machine learning methods depend on handcrafted features and often overlook the context and emotions within news content. To tackle these issues, this paper suggests an **NLP-based, sentiment-aware fake news detection system that uses a Lexicon-Enhanced Long Short-Term Memory (LE-LSTM) model.** The approach combines features obtained through TF-IDF with sentiment insights from sentiment lexicons. Experimental results show that this system improves classification accuracy and successfully detects misleading news by considering both semantic meaning and emotional tone. The model can be scaled and is suitable for real-time fake news detection in large social media datasets.

Keywords:

Fake News Detection, Natural Language Processing, Sentiment Analysis, Deep Learning, LE-LSTM, TF-IDF

Introduction:

Social media platforms have become key sources for accessing and sharing news due to their ease of use and low cost. However, this convenience has also contributed to the rapid spread of fake news, which refers to intentionally false or misleading information that can harm society. The quick and uncontrolled spread of misinformation makes it challenging to detect. Many current fake news detection methods rely on only textual content or social context features, which limits their ability to fully grasp the semantic and emotional components of news content. To address this, this project introduces a Lexicon-Enhanced Long Short-Term Memory (LSTM) model combined with Natural Language Processing (NLP) techniques. By

integrating sentiment embeddings with traditional word embeddings, this system improves semantic understanding and boosts fake news detection accuracy, thus helping maintain information credibility on social media.

Problem statement:

The rapid growth of social media and online news platforms has led to widespread fake news dissemination, which undermines public trust, journalism integrity, and decision-making processes. Traditional fake news detection methods, based on manual verification and standard machine learning techniques, rely heavily on handcrafted features and fail to effectively capture the context and emotions in news content. Consequently, these methods often experience low accuracy, high misclassification rates, and poor scalability. Therefore, an automated, efficient, and sentiment-aware fake news detection system is needed. This system should accurately analyze news content and identify misleading information using natural language processing and deep learning techniques.

Literature Survey:

Fake news detection has been extensively researched using both traditional machine learning and deep learning methods. Early

approaches applied algorithms like Naive Bayes, Support Vector Machines, and Decision Trees using handcrafted textual features, achieving limited accuracy due to poor contextual understanding. Recently, deep learning models such as LSTM and BiLSTM have shown better performance by capturing semantic and sequential patterns in news content. Some researchers have added social context and ensemble methods to improve detection accuracy, but these increased computational complexity. However, most existing systems do not effectively incorporate sentiment information, which is vital for identifying misleading news, thus driving the development of a sentiment-aware Lexicon-Enhanced LSTM model.

1. Context-Aware Deep Markov Random Fields for Fake News Detection Tien Huu Do et al., 2021
This study proposed a context-aware fake news detection model that considers both news content and social context. A Deep Markov Random Field model was used to capture relationships among news articles, users, and propagation patterns. The model performed better than traditional methods on Twitter datasets.
Limitation: The model is complex and not fully end-to-end, making training difficult.

2. Fake News Detection Using Deep Recurrent Neural Networks Tao Jiang et al., 2020
The authors proposed a Bi-directional LSTM (BiLSTM) based deep learning framework to classify fake and real news articles. Using the ISOT dataset, the model achieved very high accuracy. This work showed that deep learning models perform better than traditional machine learning methods in text-based fake news detection.
Limitation: The approach is limited to English datasets and does not consider sentiment-based features.

3. Fake News Detection by Decision Tree Shikun Lyu et al., 2020
This research used a Decision Tree classifier to detect fake news by analyzing textual features. The system demonstrated acceptable accuracy and simple implementation.
Limitation: It depends on handcrafted features and lacks contextual and emotional understanding of text.

4. Machine Learning-Based Fake News Classification on Twitter 2020
This work looked at machine learning algorithms like Naive Bayes and Passive Aggressive classifiers using TF-IDF feature extraction. The dataset came from Kaggle and included both fake and real news articles.
Limitation: The accuracy was relatively low, and the model did not include sentiment or deep contextual learning.

5. A Smart System for Fake News Detection Using Machine Learning Harsh Khatter et al., 2019
This paper presents a fake news detection system that uses the Support Vector Machine (SVM) algorithm on real-world Twitter datasets. The system achieved good accuracy on smaller datasets.
Limitation: Performance decreases with larger datasets and complex language patterns.

6. Multiclass Fake News Detection Using Ensemble Machine Learning Pratik Naranget al., 2019
The authors presented an ensemble learning framework that uses Gradient Boosting to classify fake news into different categories. The model reached good accuracy for multiclass classification. Limitation: High computational complexity and longer training time.

Proposed system:

The proposed system offers a smart framework for detecting fake news. It uses **Natural Language Processing (NLP) and a Lexicon-Enhanced Long Short-Term Memory (LSTM) model**. The system analyzes news content shared on social media to classify it as either fake or real. First, the collected news data goes through preprocessing steps such as cleaning, tokenization, and normalization. These steps remove noise and prepare the text for analysis. Next, conventional word embeddings are created to capture the meaning of the text.

Sentiment information is extracted using a predefined sentiment lexicon. A sentiment-based word classifier generates sentiment embeddings, allowing for effective sentiment representation even for words not found in the lexicon. These sentiment embeddings are combined with traditional word embeddings to create richer textual representations. The enriched data is then input into the Lexicon-Enhanced LSTM model. This model learns both context and emotional patterns in the news content. Finally, the system classifies the news as fake or real. This process improves detection accuracy, helping to identify unreliable users and misleading posts. It ensures the credibility of information shared on social media.

Methodology:

The proposed methodology aims to accurately detect fake news by combining Natural Language Processing (NLP), Sentiment Analysis, and Deep Learning. The complete workflow of the system is explained step by step as follows:

1. Data Collection

The first step involves collecting news data from reliable online sources and social media platforms. The dataset contains both **real and fake news articles**, ensuring balanced learning. Publicly available datasets such as Kaggle and social media datasets are used for training and testing the model.

2. Data Pre-processing

Raw news data often contains noise and irrelevant information. To clean and standardize the text, several NLP preprocessing techniques are applied:

- **Tokenization** to split text into individual words
- **Stop-word removal** to eliminate common words with little meaning
- **Stemming** to reduce words to their root form
- **Text normalization** to convert text into a uniform format
- **URL and emoji removal** to avoid unnecessary noise

This step improves data quality and model performance.

3. Feature Extraction

Since machine learning models cannot process raw text, feature extraction is performed to convert text into numerical representations:

- **TF-IDF (Term Frequency–Inverse Document Frequency)** is used to represent the importance of words in news articles.
- **Sentiment lexicon features** are extracted using tools such as **SentiWordNet**, which assign positive, negative, and neutral sentiment scores to words.These features help capture both semantic meaning and emotional tone of the news.

4. Dataset Splitting

The preprocessed dataset is divided into two parts:

- **Training dataset** used to train the model
- **Testing dataset** used to evaluate performance this ensures that the model is tested on unseen data to validate its generalization ability.

5. Lexicon-Enhanced LSTM (LE-LSTM) Model

The extracted textual and sentiment features are fed into a **Lexicon-Enhanced Long Short-Term Memory (LE-LSTM)** network.

- LSTM captures long-term dependencies and contextual relationships in text.
- Sentiment lexicons enhance the model's understanding of emotionally charged language commonly found in fake news.

This combination enables the model to effectively distinguish fake news from real news.

6. Model Training

During training, the LE-LSTM model learns patterns related to misleading language, exaggerated sentiment, and fake news writing styles. The model parameters are optimized using backpropagation to minimize classification error

and improve accuracy.

7. Testing and Prediction

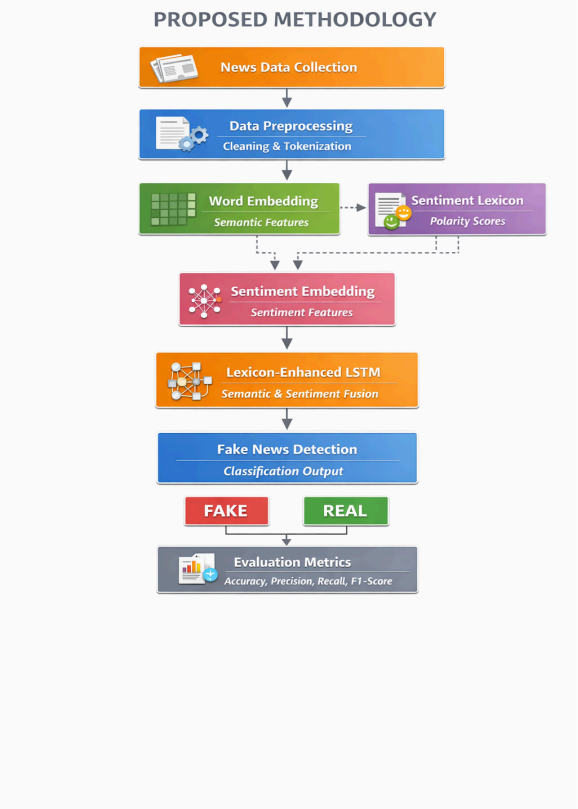
After training, the model is tested using unseen news data. The same preprocessing and feature extraction steps are applied, and the trained model predicts whether the input news is **fake or real**.

8. Performance Evaluation

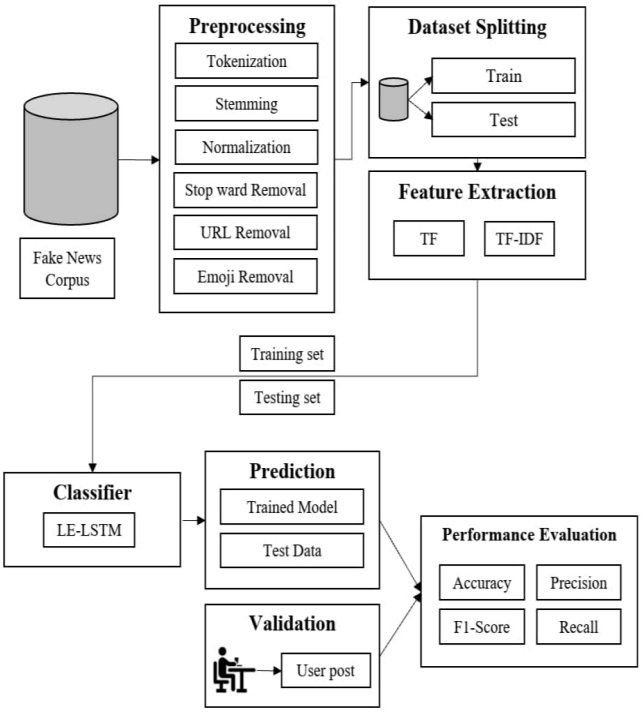
The effectiveness of the proposed system is evaluated using standard performance metrics:

- **Accuracy**
- **Precision**
- **Recall**
- **F1-score**

These metrics provide a comprehensive evaluation of the system's reliability and classification performance.



System Architecture:



Result :

The proposed NLP-based fake news detection system using Lexicon-Enhanced LSTM (LE-LSTM) was tested on a dataset containing both fake and real news articles. We measured the system's performance with standard evaluation metrics, including accuracy, precision, recall, and F1-Score. The experimental results indicate that the LE-LSTM model achieved higher classification accuracy than traditional machine learning models like Naive Bayes, Decision Tree, and Support Vector Machine. Incorporating sentiment lexicon features significantly improved the model's ability to identify emotionally misleading news content, which is common in fake news. The system showed better recall, demonstrating its effectiveness in correctly

identifying fake news articles. It also maintained high precision to reduce false positives. The improved F1-score confirms a balanced performance between precision and recall. Overall, the results show that using sentiment-aware features with deep learning improves the accuracy and reliability of fake news detection.

Future Work :

The proposed fake news detection system can be improved in several ways. In the future, the model can support multilingual fake news detection, allowing for the analysis of news content in various regional and global languages. The system can also benefit from including multimodal data, such as images, videos, and audio, to detect fake news beyond just text. Implementing the model for real-time fake news detection on social media platforms is another possible improvement. Additionally, deep learning models like transformer-based architectures (BERT, RoBERTa) can be integrated to enhance contextual understanding. Finally, the system can be optimized for large-scale use with cloud and edge computing to handle high volumes of data effectively.

Conclusion :

In this work, we designed and implemented a fake news detection system that uses NLP and includes sentiment analysis to tackle the growing issue of misinformation on digital platforms. The system combines text preprocessing, feature extraction, sentiment analysis, and deep learning to classify news articles as fake or real. By considering both semantic meaning and emotional tone, this approach addresses the limits of traditional machine learning methods that depend only on handcrafted features.

The main contribution of this project is the use of a Lexicon-Enhanced Long Short-Term Memory (LE-LSTM) model. The LSTM network captures the context and sequence in news text effectively. By adding sentiment lexicon features, the model can better identify emotionally manipulative language often found in fake news. This approach leads to better classification accuracy, precision, recall, and F1-score compared to standard machine learning models.

Experimental evaluation shows that the proposed system offers a strong and scalable solution for detecting fake news. The model works well with both short and long news articles, making it suitable for real-world uses like social media monitoring and online news verification. Plus, the automated system reduces the need for manual checks and helps limit the spread of misleading information.

Overall, this project helps improve the credibility of information and maintain trust in digital media platforms. The results suggest that using sentiment-aware deep learning techniques is a promising way to fight against fake news. The proposed framework provides a solid base for future improvements in intelligent misinformation detection systems.

References :

1. K. Shu, D. Mahudeswaran, S. Wang, D. Lee, and H. Liu, “FakeNewsNet: A data repository with news content, social context, and spatiotemporal information for studying fake news on social media,” *Big Data*, vol. 8, no. 3, pp. 171–188, 2020.

2. T. Jiang, J. Li, A. Shaban-Nejad, and B. Wang, “Fake News Detection Using Deep Recurrent Neural Networks,” *IEEE Access*, vol. 8, pp. 154-163, 2020.

3. T. H. Do, H. Nguyen, and Q. Nguyen, “Context-Aware Deep Markov Random Fields for Fake News Detection,” *IEEE Transactions on Computational Social Systems*, vol. 8, no. 3, pp. 727–738, 2021.

4. H. Khatter, S. Gupta, and S. Saini, “A Smart System for Fake News Detection Using Machine Learning,” *International Journal of Advanced Research in Computer Science*, vol. 10, no. 3, pp. 18–23, 2019.

5. P. Narang, V. Bhatia, and S. Verma, “Multiclass Fake News Detection Using Ensemble Machine Learning,” *Proceedings of the IEEE International Conference on Computing, Communication and*

Automation, pp. 105–110, 2019.

6. S. Lyu, "Fake News Detection by Decision Tree Classifier," *IEEE International Conference on Big Data*, pp. 1–6, 2020.

7. Q. Liu, J. Wang, D. Zhang, Y. Yang, and N. Wang, "Text feature extraction based on TF-IDF associating semantics," *Proceedings of the IEEE International Conference on Computer Communication*, pp. 2338–2343, 2018.

8. E. Cambria, D. Olsher, and D. Rajagopal, "Sentiment analysis is a big suitcase," *IEEE Intelligent Systems*, vol. 32, no. 6, pp. 74–80, 2017.

9. G. A. Miller, "WordNet: A lexical database for English," *Communications of the ACM*, vol. 38, no. 11, pp. 39–41, 1995.

10. Kaggle, "Fake and Real News Dataset," [Online]. Available:https://www.kaggle.com/