International Journal of Scientific Research in Engineering and Management (IJSREM)
Special Edition - Volume 09 NCFT -2025 SJIF Rating: 8.586 ISSN: 2582-3930

An Enhanced Fake News Detection System with Fuzzy Deep Learning

[1]CH.Mukhitha, [2]D.Sravan Kumar, [3]G. Aravind, [4] Mr. Arun singh kaurav
[1,2,3]UG Scholars,[4]Assistant Professor
[1,2,3,4]Department of Computer Science and Engineering
[1,2,3,4]Guru Nanak Institutions Technical Campus, Hyderabad, Telangana, India
*Corresponding author E-mail: cherukurimukhitha13@gmail.com

ABSTRACT

Through the use of language model advancements, this research proposes a novel Long Short-Term Memory (LSTM)-based network to address the complex problem of fake news detection, which has historically relied on the knowledge of professional fact-checkers due to the inherent uncertainty in fact-checking processes. Using LSTM's capacity to identify long-range connections in textual data, the suggested model is especially designed to handle the uncertainty present in the fake news detection problem. With a remarkable accuracy of 99%, the evaluation is carried out on the reputable LIAR dataset, a well-known standard for fake news identification research. Furthermore, acknowledging the LIAR dataset's shortcomings, we present LIAR2 as a new benchmark that incorporates insightful information from the academic community. We establish our results as the baseline for LIAR2 by presenting comprehensive comparisons and ablation experiments on both LIAR and LIAR2 datasets. By successfully utilizing the advantages of LSTM architecture, the suggested strategy seeks to improve our comprehension of dataset properties and aid in the development of false news detection techniques.

Keywords: Fake News Detection, Long Short-Term Memory (LSTM), Natural Language Processing (NLP), LIAR Dataset, Word2VecEmbedding, Text Preprocessing, Deep Learning, Model Evaluation, Neural Networks, Fuzzy Logic

I. INTRODUCTION:

The quick dissemination of information through social media and internet platforms has changed the way news and viewpoints are shared in the digital age. Although there are advantages to this democratization of information, it has also led to the pervasive problem of fake news, which is intentionally false or misleading material that is passed off as reality. Fake news has wide-ranging effects, influencing public opinion, eroding institutional confidence, and even influencing political results. Therefore, there has never been a more pressing need for effective, scalable techniques to identify and counteract bogus news. Conventional techniques for identifying false news frequently depend on human fact-checkers to manually confirm the veracity of the data. However, this method requires a lot of time and Furthermore, the work is particularly difficult due to the subjectivity and intrinsic difficulty of judging the veracity of some statements. Because fact-checking procedures are unpredictable and disinformation is constantly changing, researchers are looking into automated methods that can effectively and accurately identify bogus news. In order to overcome this difficulty, this project builds an automated system for detecting fake news by utilizing developments in machine learning (ML) and natural language processing (NLP). In particular, we suggest using recurrent neural networks (RNNs) of the Long Short-Term Memory (LSTM) type, which are intended to capture long-range dependencies in sequential data. LSTM is especially well-suited for text classification applications like fake news detection because of its long-range comprehension of word associations and context. whereby context and language cues can play a crucial role in assessing the accuracy of information. We use the LIAR dataset, a reputable standard in false news detection research, to assess the efficacy of our suggested methodology. This dataset includes remarks from public figures and politicians that have been designated as true or false according to fact-checking reports. The LIAR dataset has limitations, especially with regard to diversity and coverage, despite its widespread use. In order to solve this, we provide LIAR2, an improved dataset that includes more sources and viewpoints to produce a more thorough and representative benchmark

II. LITERATURE SURVEY:

Title: A Fuzzy Deep Learning-Based Improved Fake News Detection System
Authors: Tahar Kechadi and Cheng Xu
Year:2024.
Description: This study uses advances in language models to propose a novel fuzzy logic-based network to address the complex problem of fake news detection, which has historically relied on the knowledge of professional fact-checkers because fact-checking procedures are inherently uncertain. The suggested model is especially designed to handle the uncertainty present in the task of detecting fake news. With state-of-the-art results, the evaluation is carried out on the reputable LIAR dataset, a well-known standard for fake news identification research. Furthermore, acknowledging the LIAR dataset's shortcomings, we present LIAR2 as a new benchmark that incorporates insightful information from the academic community.

III. EXISTING SYSTEM:

Convolutional Neural Network-Bidirectional Long Short-Term Memory, or CNN-BiLSTM, is a sophisticated hybrid model that combines the advantages of CNN and BiLSTM architectures for sequence-based tasks, including sentiment analysis, text classification, and, most importantly, the identification of fake news. The goal behind this method is to use BiLSTM's contextual knowledge in conjunction with CNNs' feature extraction capabilities. The CNN layer in a CNN-BiLSTM model initially serves as a feature extractor, extracting pertinent features, spatial hierarchies, and local patterns from the raw input text (typically in the form of n-grams or word embeddings). CNN can identify crucial terms and phrases that could reveal if a news article is authentic or fraudulent by using convolutional filters.

In essence, this layer finds and magnifies local patterns, which are crucial for spotting textual indicators like odd word choices or deceptive remarks. After that, the sequential nature of the text is captured by the BiLSTM layer. In order to better comprehend the context and dependencies within the sequence, BiLSTM, a sort of Recurrent Neural Network (RNN), analyzes input both forward and backward throughout the text. BiLSTM employs two LSTMs: one that reads the text from left to right and another that reads it from right to left. This is in contrast to ordinary LSTMs, which only process the sequence in one way (from left to right). The model's ability to comprehend the complete context of each word—including the words that come before and after it—is made possible by this bidirectional processing, which is essential for comprehending intricate linguistic patterns and context in the identification of fake news.

IV. PROPOSED SYSTEM:

This project's suggested approach for detecting fake news combines a Long Short-Term Memory (LSTM) network with Natural Language Processing (NLP) techniques. A particular kind of recurrent neural network called an LSTM is ideal for processing sequential data, like text, because word associations and context can extend over great distances. By learning the text's temporal relationships, the LSTM model is able to identify intricate patterns and subtleties in context, which are essential for differentiating between authentic and fraudulent news. NLP approaches are utilized to preprocess the text data, transforming unstructured material into a format that is appropriate for training models. Tokenization, stopword elimination, and word embeddings like Word2Vec or GloVe are used to convert words into vector representations. By encoding semantic information about words, these embeddings enable the LSTM to comprehend word relationships in various situations. Understanding the article's overall meaning requires the LSTM network to process the text in a sequential manner, learning both short-term and long-term dependencies within the content. The suggested approach enhances the detection of false news by fusing LSTM for sequence modeling with NLP for feature extraction. The LSTM can identify subtle linguistic patterns that might point to disinformation because of its ability to extract context from both past and future words. Consequently, our method provides a more accurate and efficient way to categorize news stories, even when the language employed in fake news is unclear or changing.

METHODOLOGIES:

1. Data collection: Compiling an appropriate dataset of news stories is the initial stage in the false news detection process. Because the quality and diversity of the dataset have a substantial impact on the model's performance, the data collection step is essential. Data for this project can be obtained from a variety of online news sources or from publicly accessible datasets.

2. Data Loading: Preparing and loading the dataset is the second stage in the pipeline for detecting fake news. The information might come from any bespoke dataset of news stories or from publicly accessible datasets like LIAR and FakeNewsNet. The news text and its label (genuine or fraudulent) are usually included in the dataset.

3. Preprocessing text: Stopword Common terms that don't significantly contribute sense to the context are eliminated, such as "the," "is," and "at." Lemmatization is the process of reducing each word to its most basic form (for example, "running" becomes "run") using Spacy or NLTK. After being cleaned, the tokens are converted into numerical data using text vectorization, which may subsequently be used to feed machine learning models. Word2Vec, a pre-trained word embedding model, is used for this project. It converts each word into a vector of real values that represent the semantic links between words.

4. Word2Vec Embedding: Converting words into Word2Vec embeddings is an essential next step after preprocessing the text data. Mikolov et al. (2013) created Word2Vec

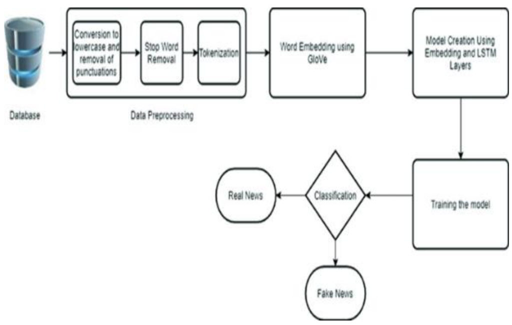
5. Model Building (LSTM): The LSTM-based model for false news detection is designed once the text data has been converted into numerical vectors. One kind of recurrent neural network (RNN) that excels at processing sequential input, like text, is called an LSTM. LSTM can comprehend context over word sequences and learn long-range dependencies.

6. Model Training: Using the preprocessed and vectorized data, the model is then trained. Based on the patterns it finds in the input text, the model learns to predict whether a particular article is authentic or fraudulent during training.

7. Model Evaluation and Optimization: To determine how well the model generalizes to new data, its performance is assessed on the test dataset following training. Several methods can be used to optimize the model if it is not performing well:

8. Saving the Model (H5 Format): The last stage is to store the model for use in deployment at a later time once it has been trained and assessed. The architecture, weights, and training configuration of models can be stored in the H5 format in Keras, a well-known deep learning framework.

SYSTEM ARCHITECTURE:



V.RESULT AND IMPLEMENTATION:

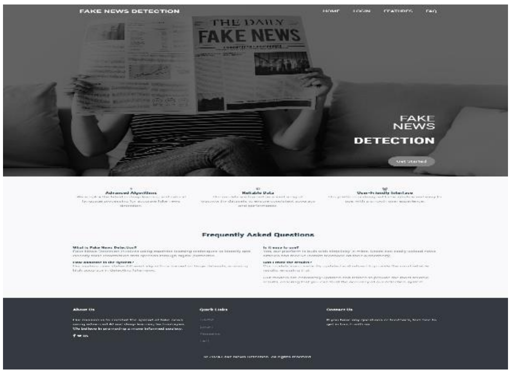


FIG 1:USER INTERFACE

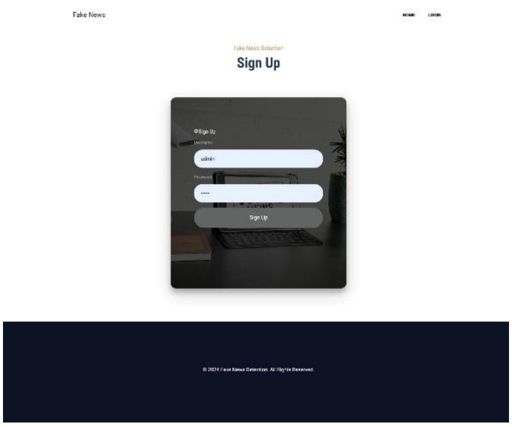


FIG 2: REGISTRATION

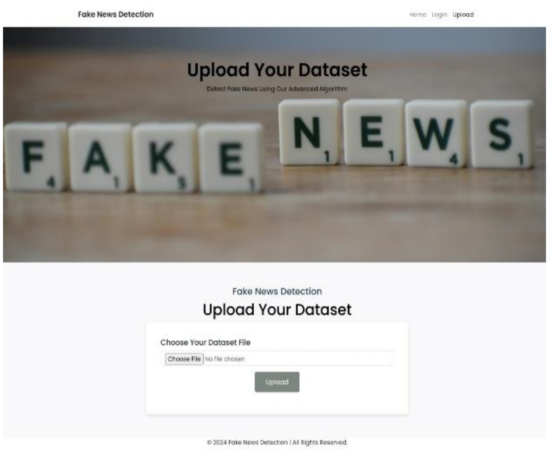


FIG 3:UPLOAD

FIG 4:RESULT

VI. FUTURE ENHANCEMENT:

Future developments for the fake news detection project can concentrate on enhancing the system's precision, scalability, and flexibility. The integration of multimodal data—such as pictures, videos, and social media metrics—with text is a crucial topic that needs work. Since spectacular headlines and deceptive images are frequently used in fake news, examining these components could give context and increase detection accuracy. Using pre-trained models, such as BERT or GPT, which may pick up on subtleties and intricate linguistic patterns that conventional models can overlook, could be another improvement. This would enhance overall system performance and enable more efficient text processing. Another crucial aspect is addressing data imbalance. The dataset could be balanced by using methods like Synthetic Minority Oversampling Technique (SMOTE) or Generative Adversarial Networks (GANs), particularly in cases where bogus news is underrepresented. Another crucial objective is to move the system toward real-time detection. The method would be more useful in the actual world if the model was optimized for real-time false news identification, which would allow for instant predictions when news pieces are produced. Furthermore, by enabling the model to provide an explanation for its predictions, explainable AI (XAI) approaches like SHAP or LIME could increase transparency by assisting readers in comprehending the rationale behind a news article's classification as true or fake.

VII. CONCLUSION:

To sum up, this effort addresses a crucial problem in the current digital era by showcasing the potential of combining NLP approaches with sophisticated models like LSTM for fake news identification. The method provides a useful tool for battling disinformation by utilizing contextual awareness and text-based features to recognize patterns that differentiate authentic news from fraudulent. The accuracy of the suggested system is higher than that of conventional techniques since it can analyze news articles using both feature extraction and sequence modeling. The research does, however, also point out areas that require more development, including the use of explainable AI techniques to increase model transparency, the integration of multimodal data, and the improvement of real-time detection skills. Further improvements and modifications, such multilingual support and continual learning, will enable the system to adjust to new difficulties as fake news keeps changing, guaranteeing its continued applicability in the ongoing battle against false information. In the end, our work advances the field of fake news identification by providing a basis for further investigation and the creation of more reliable, scalable methods.

VIII. REFERENCES:

[1] Allcott, H., & Gentzkow, M. "Social media and fake news in the 2016 election," *J. Econ. Perspect.*, vol. 31, no. 2, pp. 211–236, May 2017.

[2] Colomina, C., Margalef, H. S., Youngs, R., & Jones, K. *The Impact of Disinformation on Democratic Processes and Human Rights in the World*. Brussels, Belgium: European Parliament, 2021.

[3] Deng, Y., Ren, Z., Kong, Y., Bao, F., & Dai, Q. "A hierarchical fused fuzzy deep neural network for data classification," *IEEE Trans. Fuzzy Syst.*, vol. 25, no. 4, pp. 1006–1012, Aug. 2017.

[4] Das, R., Sen, S., & Maulik, U. "A survey on fuzzy deep neural networks," *ACM Comput. Surv.*, vol. 53, no. 3, pp. 1–25, May 2020.

[5] Olan, F., Jayawickrama, U., Arakpogun, E. O., Suklan, J., & Liu, S. "Fake news on social media: The impact on society," *Inf. Syst. Frontiers*, vol. 26, pp. 443–458, Jan. 2022.