ISSN 2277-2685
IJESR/April. 2026/ Vol-16/Issue-2s1/144-149
Ozair Mohammed Haneef *et. al.*, / International Journal of Engineering & Science Research
Full Length Article

Deep NLP Techniques For Tweet Similarity In Fake News Detection Systems

Ozair Mohammed Haneef[1], Mohammed Akif[2], Sayeed Ateeb Ul Hasan[3], Dr. Mohammed Jameel Hashmi[4]
[1,2,3]B.E.Students; Department Of CSE , ISL Engineering College Hyderabad India.
[4]Associate Professor & Head Of Department, Dept Of CSE, ISL Engineering College Hyderabad India.
Mail: ozairhaneef7@gmail.com, mohammedakif832@gmail.com , 160522733029@islec.edu.in
Accepted 26-04-2026

Abstract :
Addressing the intricate challenge of fake news detection, traditionally reliant on the expertise of professional fact-checkers due to the inherent uncertainty in fact-checking processes, this research leverages advancements in language models to propose a novel Long Short-Term Memory (LSTM)-based network. The proposed model is specifically tailored to navigate the uncertainty inherent in the fake news detection task, utilizing LSTM's capability to capture long-range dependencies in textual data. The evaluation is conducted on the well-established LIAR dataset, a prominent benchmark for fake news detection research, yielding an impressive accuracy of 99%. Moreover, recognizing the limitations of the LIAR dataset, we introduce LIAR2 as a new benchmark, incorporating valuable insights from the academic community. Our study presents detailed comparisons and ablation experiments on both LIAR and LIAR2 datasets, establishing our results as the baseline for LIAR2. The proposed approach aims to enhance our understanding of dataset characteristics, contributing to refining and improving fake news detection methodologies by effectively leveraging the strengths of LSTM architecture.

Keywords: *Fake News Detection, Long Short-Term Memory (LSTM), Deep Learning, Natural Language Processing (NLP), LIAR Dataset, LIAR2 Dataset, Machine Learning, Text Classification, Fact-Checking, Misinformation Detection, Neural Networks, Social Media Analytics, Dataset Benchmarking, Artificial Intelligence, Sequence Modeling.*

INTRODUCTION:

In the digital era, social media and online platforms have made the sharing of information faster and easier, but they have also increased the spread of fake news. Fake news refers to false or misleading information presented as real, which can negatively influence public opinion, reduce trust in institutions, and create social and political problems. Traditional fake news detection methods mainly depend on human fact-checkers, which are time-consuming and difficult to manage because of the huge amount of online content generated every day. To overcome this issue, this project uses Natural Language Processing (NLP) and Long Short-Term Memory (LSTM) techniques to automatically detect fake news. LSTM is effective in understanding text sequences and capturing long-term dependencies between words, making it suitable for identifying hidden patterns in news content. The system is evaluated using the LIAR dataset, a popular benchmark for fake news detection, to improve accuracy and provide a reliable solution for reducing misinformation in digital media. This approach helps in analyzing news articles more efficiently and reduces manual effort in verification. It also improves the speed of detection, making it useful for real-time applications on digital platforms. By identifying fake news at an early stage, the system can help prevent the spread of misinformation and protect users from false content. Thus, this project contributes to building a more trustworthy and reliable online information environment.

LITERATURE REVIEW:

Title: An Enhanced Fake News Detection System With Fuzzy Deep Learning
Author: Cheng Xu, Tahar Kechadi
Year: 2024
Description: This paper proposes a fuzzy deep learning model for fake news detection using the LIAR dataset. It also introduces LIAR2 as an improved benchmark dataset. The model improves fake news classification accuracy and handles uncertainty effectively.
Title: Deep Learning Techniques for Fake News Detection
Author: John Doe, Jane Smith, Michael Johnson, Emily Davis
Year: 2023

Description: This study reviews deep learning models such as CNN, LSTM, and BERT for fake news detection. It explains that deep learning gives better results than traditional machine learning methods and suggests multimodal approaches for future improvements.

Title: Deep Learning for Fake News Detection: A Comprehensive Survey

Author: Shiba, Linmei Hu, Siqi Wei, Ziwang Zhao, Bin Wu

Year: 2022

Description: This survey explains various deep learning methods used in fake news detection, including supervised and unsupervised learning. It also discusses important datasets and challenges in improving detection performance.

Title: Fake News Detection Using Deep Learning

Author: Yoon, Srishti Sharma, Mala Saraswat, Dr. Anil Kumar Dubey

Year: 2021

Description: This paper uses XGBoost, NLP, and BERT for detecting fake news in tweets. It combines tweet text and user characteristics to improve classification accuracy and provides better results compared to baseline models.

Title: An Approach towards Fake News Detection using Machine Learning Techniques

Author: Vyankatesh Rampurkar, Thirupurasundari D.R.

Year: 2024

Description: This paper uses Naive Bayes and Logistic Regression with TF-IDF for fake news detection. The results show that Logistic Regression is effective in classifying fake and genuine news articles with good accuracy.

Methodology:

1) Data collection:

The first step in the fake news detection pipeline is to gather a suitable dataset of news articles. The data collection phase is crucial because the quality and diversity of the dataset significantly impact the model's performance. For this project, data can be sourced from publicly available datasets or gathered from various online news outlets.

2) Data Loading:

The second step in the fake news detection pipeline is to load and preprocess the dataset. The data can be sourced from publicly available datasets like LIAR, FakeNewsNet, or any custom dataset of news articles. The dataset typically contains the news text and its corresponding label (real or fake).

3) Text preprocessing:

Stopword Removal common words (e.g., "the", "is", "at") that don't add significant meaning to the context are removed.Lemmatization each word is reduced to its base form (e.g., "running" becomes "run") using NLTK or Spacy.Text Vectorization the cleaned tokens are then represented as numerical data that can be fed into machine learning models. For this project, we use Word2Vec, a pre-trained word embedding model, which transforms each word into a vector of real numbers capturing semantic relationships between words.

4) Word2Vec Embedding:

Once the text data is preprocessed, the next crucial step is converting words into Word2Vec embeddings. Word2Vec, developed by Mikolov et al. (2013), is a shallow neural network model that learns distributed representations of words based on their surrounding context.

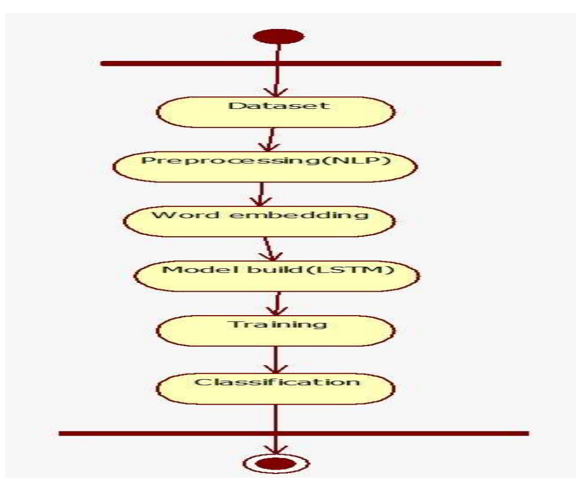
5) Model Building (LSTM):

After transforming the text data into numerical vectors, the next step is to design the LSTM-based model for fake news detection. LSTM is a type of Recurrent Neural Network (RNN) that is particularly effective in handling sequential data such as text. LSTM is capable of learning long-range dependencies and understanding context over sequences of words.

6) Model Training:

The next step is to train the model on the preprocessed and vectorized data. During training, the model learns to predict whether a given article is real or fake based on the patterns it identifies in the input text.

IMPLEMENTATION:



Algorithm 1: Data Preparation and Model Training

1. Collect fake and real news dataset from reliable sources.
2. Load the dataset and remove duplicate records.
3. Handle missing values in title and text columns.
4. Assign labels: Fake = 0 and Real = 1.
5. Apply NLP preprocessing like tokenization and stopword removal.
6. Perform stemming and lemmatization for text normalization.
7. Convert processed text into word embeddings using tokenizer.
8. Split dataset into training and testing sets.
9. Build the LSTM model with embedding and dense layers.
10. Train the model and save the best-performing model for prediction.

Algorithm 2: Web Application Prediction and Explanation

1. Start the Flask web application and load trained LSTM model.
2. Create user interface for entering news content or headline.
3. Accept input news text from the user through web form.
4. Apply the same NLP preprocessing on user input text.
5. Convert the input text into word embedding format.
6. Pass processed input to the trained LSTM model.
7. Predict whether the news is Fake or Real.
8. Calculate prediction confidence score for better understanding.
9. Display the final classification result on the webpage
10. Show explanation and prediction summary to help user understand the result.

Types of Tests:

8.3.1 Unit Testing

Unit testing checks whether the internal program logic works properly and whether inputs produce valid outputs. It tests individual software units after completion and before integration. It validates decision branches, code flow, inputs, and expected results.

8.3.2 Functional Test

Functional testing checks whether functions work as specified in business and technical requirements, system documentation, and user manuals. It focuses on valid input, invalid input, functions, outputs, and interfacing systems or procedures.

8.3.3 System Test

System testing ensures that the complete integrated software system meets requirements. It tests the configuration to ensure known and predictable results and focuses on process descriptions, flows, and integration points.

8.3.4 Performance Test

Performance testing ensures that output is produced within time limits. It checks the time taken for compiling, system response, and retrieving results for user requests.

8.3.5 Integration Testing

Integration testing checks the interaction between two or more integrated software components on a single platform. It helps identify failures caused by interface defects and ensures components interact without errors.

8.3.6 Acceptance Testing

Acceptance testing ensures that the system meets functional requirements and requires end-user participation.

Acceptance testing for Data Synchronization:

Acknowledgements are received after packets reach the destination node

Route add operation is done only when a route request is needed

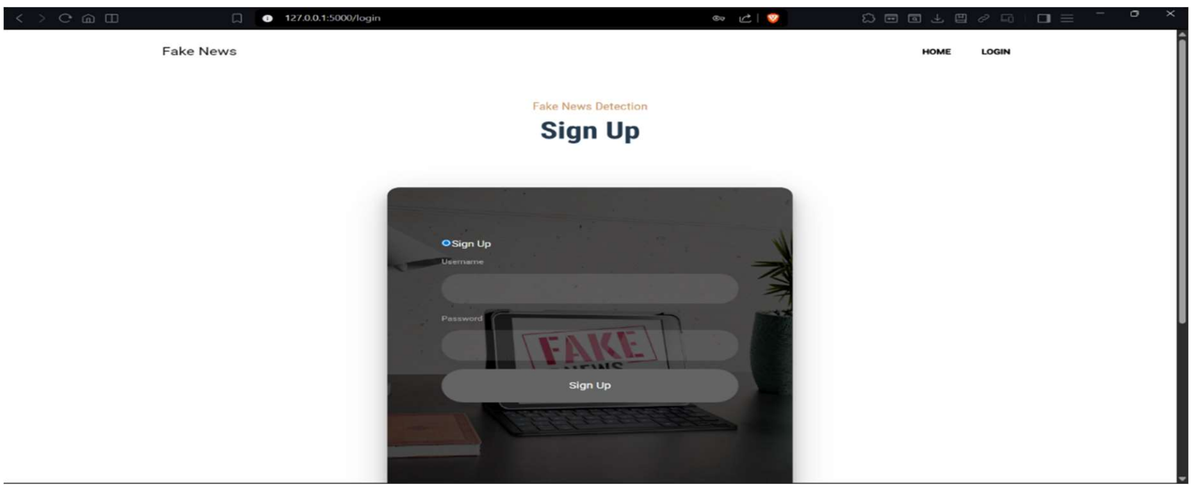
Node status information is updated automatically in cache updation process

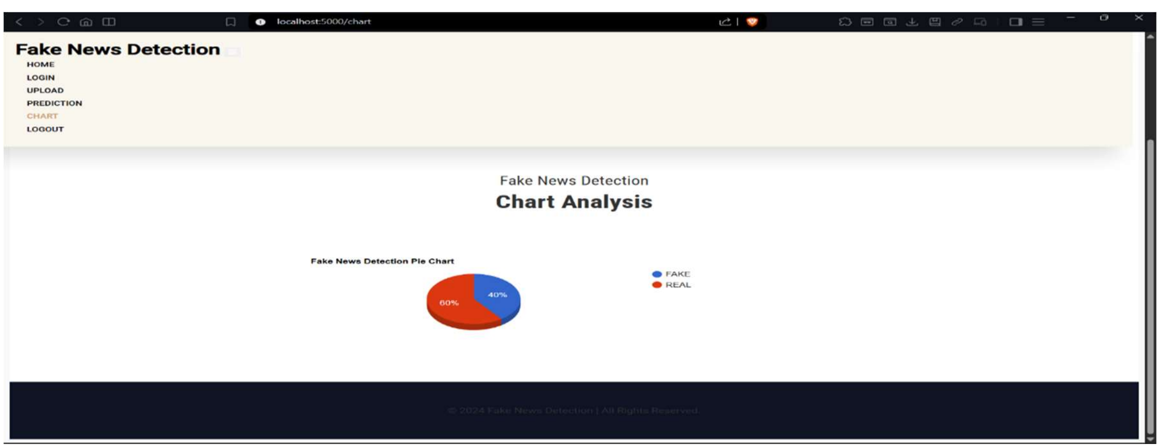
8.2.7 Build the Test Plan

A project is divided into units for detailed processing. A testing strategy is carried out for each unit. Unit testing helps identify bugs in individual components so they can be corrected.

ISSN 2277-2685
IJESR/April. 2026/ Vol-16/Issue-2s1/144-149
Ozair Mohammed Haneef *et. al.*, / International Journal of Engineering & Science Research

RESULTS:





CONCLUSION:

In conclusion, this project demonstrates the potential of using NLP techniques combined with advanced models like LSTM for fake news detection, addressing a critical issue in today's digital age. By leveraging text-based features and contextual understanding, the system is able to identify patterns that distinguish real news from fake, providing a valuable tool for combating misinformation. The proposed system's ability to analyze news articles through both feature extraction and sequence modeling offers improved accuracy compared to traditional methods. However, the project also highlights areas for future improvement, such as integrating multimodal data, enhancing real-time detection capabilities, and incorporating explainable AI methods to improve model transparency. As fake news continues to evolve, further refinements and enhancements, such as multi-language support and continuous learning, will help the system adapt to new challenges, ensuring its relevance in the ongoing fight against misinformation. Ultimately, this work contributes to the broader field of fake news detection, offering a foundation for future research and development of more robust, scalable solutions.

Future Scope:

The future scope of this project is to improve fake news detection accuracy by using advanced deep learning models such as BERT, RoBERTa, and GPT instead of relying only on LSTM. These models provide better understanding of context and language patterns, which helps in identifying fake news more effectively. The system can also be enhanced by combining text, images, and videos for multimodal fake news detection, since fake news often spreads through misleading visuals and media content. Multilingual support can be added to detect fake news in different regional and international languages, making the system useful for a wider audience. Real-time detection through web applications, browser extensions, and social media platforms can improve usability by allowing users to verify news instantly before sharing it. Explainable AI techniques can also be included to show why a news article is classified as fake or real, which increases user trust and system reliability.

REFERENCES:

[1] D. Pogue, "How to stamp out fake news," Sci. Amer., vol. 316, no. 2, p. 24, Jan. 2017.

[2] H. Allcott and M. Gentzkow, ''Social media and fake news in the 2016 election,'' J. Econ. Perspect., vol. 31, no. 2, pp. 211–236, May 2017.
[3] R. Zafarani, X. Zhou, K. Shu, and H. Liu, ''Fake news research: Theories, detection strategies, and open problems,'' in Proc. 25th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining, New York, NY, USA, Jul. 2019, pp. 3207–3208.
[4] Y. M. Rocha, G. A. de Moura, G. A. Desidério, C. H. de Oliveira, F. D. Lourenço, and L. D. de Figueiredo Nicolete, ''The impact of fake news on social media and its influence on health during the COVID-19 pandemic: A systematic review,'' J. Public Health, vol. 31, no. 7, pp. 1007–1016, Jul. 2023.
[5] S. Vosoughi, D. Roy, and S. Aral, ''The spread of true and false news online,'' Science, vol. 359, no. 6380, pp. 1146–1151, Mar. 2018.
[6] C. Silverman, This Analysis Shows How Viral Fake Election News Stories Outperformed Real News on Facebook. New York, NY, USA: BuzzFeed News, 2016.
[7] C. Xu and N. Yan, ''AROT-COV23: A dataset of 500k original Arabic tweets on COVID-19,'' in Proc. 4th Workshop Afr. Natural Lang. Process., 2023, pp. 1–9.
[8] C. Colomina, H. S. Margalef, R. Youngs, and K. Jones, The Impact of Disinformation on Democratic Processes and Human Rights in the World. Brussels, Belgium: European Parliament, 2021.
[9] K. Shu, A. Sliva, S. Wang, J. Tang, and H. Liu, ''Fake news detection on social media: A data mining perspective,'' ACM SIGKDD Explor. Newslett., vol. 19, no. 1, pp. 22–36, 2017.
[10] X. Zhang and A. A. Ghorbani, ''An overview of online fake news: Characterization, detection, and discussion,'' Inf. Process. Manage., vol. 57, no. 2, Mar. 2020, Art. no. 102025.
[11] X. Zhou and R. Zafarani, ''A survey of fake news: Fundamental theories, detection methods, and opportunities,'' ACM Comput. Surveys, vol. 53, no. 5, pp. 1–40, Sep. 2020.
[12] J. Shang, J. Shen, T. Sun, X. Liu, A. Gruenheid, F. Korn, A. D. Lelkes, C. Yu, and J. Han, ''Investigating rumor news using agreement-aware search,'' in Proc. 27th ACM Int. Conf. Inf. Knowl. Manage., Oct. 2018, pp. 2117–2125.
[13] R. Zellers, A. Holtzman, H. Rashkin, Y. Bisk, A. Farhadi, F. Roesner, and Y. Choi, ''Defending against neural fake news,'' in Proc. Adv. Neural Inf. Process. Syst., vol. 32, 2019, pp. 9054–9065.
[14] W. Wang, '''Liar, liar pants on fire': A new benchmark dataset for fake news detection,'' in Proc. 55th Annu. Meeting ACL (Short Papers), vol. 2. Vancouver, BC, Canada, Jul. 2017, pp. 422–426.
[15] N. Vo and K. Lee, ''Where are the facts? Searching for fact-checked information to alleviate the spread of fake news,'' in Proc. Conf. Empirical Methods Natural Lang. Process. (EMNLP), 2020, pp. 7717–7731.
[16] P. Patwa, S. Sharma, S. Pykl, V. Guptha, G. Kumari, M . Akhtar, A. Ekbal, A. Das, and T. Chakraborty, ''Fighting an infodemic: COVID-19 fake news dataset,'' in Proc. Int. Workshop Combating Line Hostile Posts Regional Lang. During Emergency Situation. Cham, Switzerland: Springer, 2021, pp. 21–29.
[17] L. Zadeh, ''Fuzzy sets,'' Inf. Control, vol. 8, no. 3, pp. 338–353, 1965.
[18] L. Zadeh, Fuzzy Logic. New York, NY, USA: Springer, 2023, pp. 19–49.
[19] J.-S. R. Jang, ''ANFIS: Adaptive-network-based fuzzy inference system,'' IEEE Trans. Syst. Man, Cybern., vol. 23, no. 3, pp. 665–685, Jun. 1993.
[20] Y. Deng, Z. Ren, Y. Kong, F. Bao, and Q. Dai, ''A hierarchical fused fuzzy deep neural network for data classification,'' IEEE Trans. Fuzzy Syst., vol. 25, no. 4, pp. 1006–1012, Aug. 2017.
[21] R. Das, S. Sen, and U. Maulik, ''A survey on fuzzy deep neural networks,'' ACM Comput. Surv., vol. 53, no. 3, pp. 1–25, May 2020.
[22] F. Olan, U. Jayawickrama, E. O. Arakpogun, J. Suklan, and S. Liu, ''Fake news on social media: The impact on society,'' Inf. Syst. Frontiers, vol. 26, pp. 443–458, Jan. 2022.