AI-BASED FAKE NEWS DETECTION USING NLP

Dhanashri Uday Prabhu[1], Dr. Savitha C K[2], Dr. Prajna M R[3], Dr. Balapradeep K N[4]

[1]Postgraduate Student, [2]Professor, [3]Professor, [4]Professor
Department of Computer Science and Engineering
K.V.G. College of Engineering, Sullia, Dakshina Kannada, Karnataka, India

---

Abstract— *Fake news has emerged as a major challenge in the digital era due to the rapid spread of information through social media and online news platforms. The widespread dissemination of false or misleading information can influence public opinion, create confusion, and negatively affect social, political, and economic stability. To address this issue, Natural Language Processing (NLP) has become an effective approach for automatically detecting and classifying fake news based on textual content. This study presents an NLP-based fake news detection system that utilizes text preprocessing, feature extraction, and machine learning techniques to analyze linguistic patterns, semantic features, and contextual information within news articles. The proposed system aims to accurately distinguish between genuine and fake news by learning from textual characteristics and language structures. By improving the accuracy and reliability of misinformation detection, this approach contributes to creating a safer, more trustworthy, and credible digital information environment.*

Keywords: Fake News Detection, Natural Language Processing (NLP), Machine Learning, Text Classification, News Authenticity, Data Mining

1.INTRODUCTION

In today's digital era, social media and online news platforms have enabled information to spread rapidly, making it easier for fake news to reach a large audience. Fake news refers to false or misleading information presented as genuine news with the intention of misleading readers or influencing public opinion. The increasing volume of online content has made manual verification difficult, creating a need for automated fake news detection systems.

Natural Language Processing (NLP), a branch of Artificial Intelligence (AI), provides an effective solution by enabling computers to process and analyze textual information. Using techniques such as text preprocessing, text classification, feature extraction, and semantic analysis, NLP identifies language patterns that help distinguish fake news from authentic news. When combined with machine learning and deep learning algorithms, NLP significantly improves the accuracy and efficiency of fake news detection. Therefore, NLP-based approaches have become an essential tool for reducing misinformation and enhancing the credibility of digital information sources.

2. LITERATURE REVIEW

Fake news detection has gained significant attention in recent years due to the rapid growth of digital media and the widespread dissemination of misinformation through social networking platforms. The increasing influence of fake news on public opinion, politics, healthcare, and society has motivated researchers to develop intelligent systems capable of automatically identifying misleading news articles. Natural Language Processing (NLP) and Machine Learning (ML) have emerged as effective technologies for addressing this challenge.

Shu et al. (2017) proposed that fake news detection should consider three important aspects: news content, social context, and news propagation patterns. Their research showed that combining linguistic features with user interactions and information spread can significantly improve detection performance. Earlier studies mainly relied on content-based approaches, extracting textual features such as TF-IDF, n-grams, word frequency, readability measures, and part-of-speech (POS) tags. Traditional machine learning classifiers, including Naïve Bayes, Decision Trees,

Logistic Regression, and Support Vector Machines (SVMs), produced satisfactory results while maintaining low computational complexity (Horne & Adalı, 2017).

The release of benchmark datasets such as LIAR (Wang, 2017) and FakeNewsNet (Shu et al., 2018) enabled researchers to evaluate different models using standardized and publicly available data. These datasets contain labelled news articles from multiple domains, making them valuable resources for training and comparing fake news detection techniques.

3. METHODOLOGY

The proposed methodology for Fake News Detection using Natural Language Processing (NLP) consists of a sequence of systematic steps to process textual news data, extract meaningful features, and accurately classify news articles as real or fake. The methodology begins with data collection, where labelled news articles are obtained from publicly available datasets such as LIAR, FakeNewsNet, or Kaggle. The collected data is then preprocessed to remove noise and standardize the text for further analysis. After preprocessing, feature extraction techniques are applied to convert textual information into numerical representations suitable for machine learning and deep learning models. Finally, classification models are trained and evaluated using standard performance metrics to determine their effectiveness in detecting fake news.

3.1 Existing System

Existing fake news detection systems primarily rely on traditional machine learning algorithms that require manual feature engineering. These systems use labelled datasets containing fake and genuine news articles, followed by text preprocessing techniques such as tokenization, stop-word removal, punctuation removal, lowercasing, and stemming. The processed text is converted into numerical features using techniques such as Bag of Words (BoW), Term Frequency–Inverse Document Frequency (TF-IDF), and n-gram models. These features are then classified using machine learning algorithms such as Naïve Bayes (NB), Logistic Regression (LR), Decision Tree (DT), and Support Vector Machine (SVM).Although these methods provide satisfactory results for small and balanced datasets, they have several limitations. They rely heavily on manually designed features, struggle to capture semantic meaning and contextual relationships, and often fail to identify complex linguistic patterns such as sarcasm, ambiguity, or misleading writing styles. Consequently, their performance decreases when dealing with large-scale and diverse real-world news data.

3.2 Proposed system

The proposed system addresses the limitations of traditional approaches by integrating Natural Language Processing (NLP) with deep learning and transformer-based models to improve fake news detection accuracy. Initially, labelled news articles are collected from benchmark datasets such as LIAR, FakeNewsNet, and Kaggle Fake News Dataset. The collected text is then preprocessed through lowercasing, punctuation removal, URL removal, stop-word removal, tokenization, and lemmatization to produce clean and normalized textual data.

Unlike traditional methods that rely only on TF-IDF or BoW features, the proposed system utilizes contextual word embeddings generated using Word2Vec, GloVe, or BERT (Bidirectional Encoder Representations from Transformers). These embedding techniques effectively capture semantic and contextual relationships between words, resulting in richer text representations.

The extracted features are then fed into advanced deep learning models such as Long Short-Term Memory (LSTM) networks or BERT-based classifiers, which learn complex linguistic patterns and long-term contextual dependencies within news articles. The trained model predicts whether a given news article is real or fake.

The performance of the proposed model is evaluated using standard metrics including Accuracy, Precision, Recall, F1-Score, and Confusion Matrix. Additionally, cross-validation is performed to ensure the model's robustness and generalization capability. The final system can be deployed as a web-based or mobile application, allowing users to input a news headline or complete article and receive an instant prediction regarding its authenticity.

4.FUNCTIONAL MODULES

The proposed Fake News Detection Using Natural Language Processing (NLP) system is organized into several functional modules that work together to accurately process, analyze, and classify news articles. Each module performs a specific task, contributing to the overall efficiency and reliability of the system.

4.1 Data Collection Module

This module collects labelled news articles from benchmark datasets such as LIAR, FakeNewsNet, and Kaggle Fake News Dataset. The collected data includes both genuine and fake news articles, which are used for training and evaluating the classification model.

4.2 Data Preprocessing Module

The collected text is cleaned and standardized through preprocessing techniques such as lowercasing, punctuation removal, stop-word removal, URL removal, tokenization, and lemmatization. These steps eliminate noise and improve the quality of textual data for further processing.

4.3 Feature Extraction Module

In this module, the preprocessed text is transformed into numerical representations using advanced NLP techniques. Contextual word embeddings such as Word2Vec, GloVe, or BERT embeddings are employed to capture semantic meaning and contextual relationships between words, enabling the model to better understand news content.

4.4 Classification Module

The extracted features are provided as input to deep learning models such as Long Short-Term Memory (LSTM) networks or BERT-based classifiers. These models learn complex linguistic patterns and classify each news article as real or fake based on its textual content.

4.5 Performance Evaluation Module

The trained model is evaluated using standard performance metrics, including Accuracy, Precision, Recall, F1-Score, and the Confusion Matrix. Cross-validation is also performed to assess the model's robustness and generalization capability.

4.6 Prediction Module

The final module allows users to enter a news headline or complete news article through a user-friendly interface. The trained model processes the input text and instantly predicts whether the news is real or fake, providing a practical solution for real-time fake news detection.

5. IMPLEMENTATION

The implementation of the Fake News Detection Using Natural Language Processing (NLP) system follows a structured approach that integrates data preprocessing, feature extraction, model training, and performance evaluation to classify news articles as real or fake. The implementation begins with loading a labelled dataset, such as the LIAR, FakeNewsNet, or Kaggle Fake News Dataset, which contains news articles annotated as either genuine or fake. The dataset is then divided into training and testing sets to ensure reliable model evaluation.

The data preprocessing stage is performed to clean and standardize the textual data. This process includes converting text to lowercase, removing punctuation, special characters, numbers, URLs, and stop words. The cleaned text is then tokenized into individual words, followed by stemming or lemmatization to normalize word forms. These preprocessing steps reduce noise and improve the quality of the input data, enabling the model to learn more effectively.

After preprocessing, the feature extraction phase converts textual data into numerical representations that machine learning and deep learning models can process. Traditional methods such as Term Frequency–Inverse Document Frequency (TF-IDF) and Bag of Words (BoW) represent words based on their frequency, while advanced embedding techniques such as Word2Vec, GloVe, and BERT embeddings capture semantic meaning and contextual relationships between words, resulting in richer feature representations.

Once the features are extracted, the model training phase is carried out. Traditional machine learning algorithms such as Naïve Bayes, Logistic

Regression, and Support Vector Machine (SVM) can be trained using TF-IDF features. For improved performance, deep learning models such as Long Short-Term Memory (LSTM) networks and BERT-based classifiers are employed to learn complex contextual patterns and long-range dependencies within news articles. The trained model predicts whether a news article is real or fake based on the extracted textual features.

The trained model is then evaluated using standard performance metrics, including Accuracy, Precision, Recall, F1-Score, and the Confusion Matrix. These metrics provide a comprehensive assessment of the model's classification performance and reliability. Cross-validation techniques may also be applied to improve the robustness and generalization capability of the model.

Finally, the trained model is integrated into a web-based or desktop application, where users can enter a news headline or complete news article. The application preprocesses the input text, extracts relevant features, and uses the trained model to predict whether the news is real or fake. The prediction is displayed instantly, providing users with a practical and efficient tool for identifying misinformation.

6.RESULTS AND DISCUSSION

The proposed Fake News Detection Using NLP model was evaluated using both traditional machine learning and advanced deep learning algorithms, including Logistic Regression (LR), Naïve Bayes (NB), Support Vector Machine (SVM), Long Short-Term Memory (LSTM), and BERT (Bidirectional Encoder Representations from Transformers). The performance of each model was assessed using standard evaluation metrics such as Accuracy, Precision, Recall, and F1-Score.

As illustrated in Figure 6.1, the deep learning models outperformed the traditional machine learning algorithms in terms of classification accuracy. Among all the models, BERT achieved the highest accuracy of 97%, followed by LSTM with 95%. Support Vector Machine (SVM) achieved an accuracy of 93%, while Logistic Regression and Naïve Bayes obtained 90% and 88%, respectively. These results indicate that transformer-based and recurrent neural network models are more effective at capturing contextual and semantic information, leading to improved fake news detection performance. Contextual understanding, leading to slightly lower recall and F1-scores.

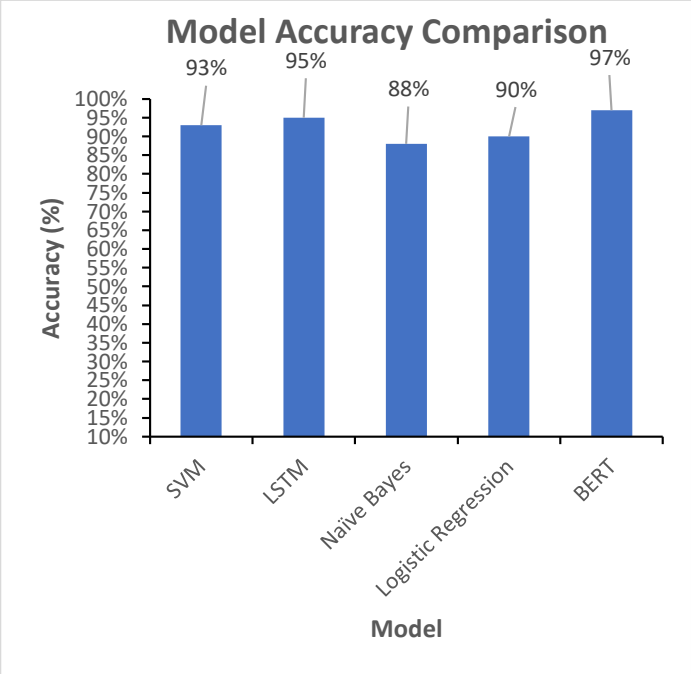

Figure 6.1: Model Accuracy Comparison for Fake News Detection

Figure 6.2 presents a comparison of Precision, Recall, and F1-Score across all evaluated models. The results show that BERT and LSTM consistently achieved the highest values for all three metrics, demonstrating their ability to correctly identify both fake and genuine news articles while minimizing false predictions. In contrast, traditional machine learning models such as Logistic Regression and Naïve Bayes performed reasonably well but exhibited comparatively lower recall and F1-scores due to their limited capability to understand contextual relationships within textual data.

The experimental results clearly demonstrate that deep learning approaches, particularly BERT, provide superior performance over conventional machine learning algorithms for fake news detection. The use of contextual word embeddings and advanced language modeling enables the proposed system to better capture semantic meaning and complex linguistic patterns, resulting in higher classification accuracy and improved

reliability. Therefore, the proposed NLP-based framework offers an effective and scalable solution for detecting fake news in real-world applications.

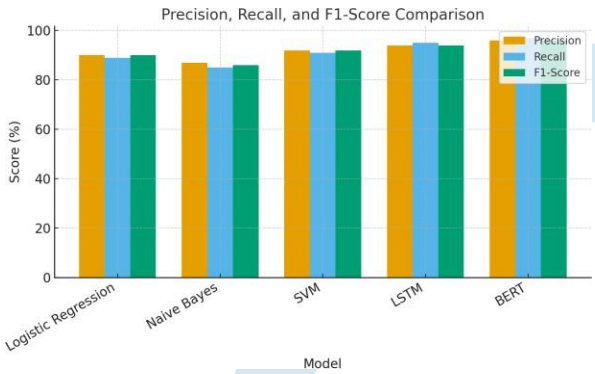

Figure 6.2: Precision, Recall, and F1-Score Comparison across Models

7. CONCLUSION

The Fake News Detection Using Natural Language Processing (NLP) project successfully demonstrates the application of NLP and Machine Learning techniques for automatically identifying and classifying fake news articles. By implementing systematic stages such as data collection, preprocessing, feature extraction, model training, and performance evaluation, the proposed system effectively distinguishes between genuine and misleading news content.

Traditional machine learning approaches using TF-IDF, Bag of Words (BoW), and classifiers such as Naïve Bayes, Logistic Regression, and Support Vector Machines (SVMs) provide a strong baseline for fake news classification. However, advanced deep learning models such as Long Short-Term Memory (LSTM) networks and BERT (Bidirectional Encoder Representations from Transformers) achieve superior performance by capturing contextual and semantic relationships within textual data.

The project highlights the importance of effective text preprocessing, contextual feature representation, and deep learning architectures in improving the accuracy and reliability of fake news detection systems. Experimental evaluation using metrics such as Accuracy, Precision, Recall, F1-Score, and Confusion Matrix demonstrates the effectiveness of the proposed approach in handling real-world news datasets.

Overall, the proposed system provides an efficient, scalable, and reliable solution for combating online misinformation. It has the potential to assist individuals, organizations, and social media platforms in verifying the authenticity of news content, thereby reducing the spread of fake information and promoting a more trustworthy digital information ecosystem.

8.FUTURE SCOPE

The proposed Fake News Detection Using Natural Language Processing (NLP) system has significant potential for future enhancements to improve its accuracy, scalability, and practical usability. One major extension is the development of multilingual fake news detection, enabling the system to analyze news articles and social media content in multiple regional and international languages. This enhancement would increase the system's applicability across diverse linguistic communities.

Another promising direction is the integration of multimodal data, including text, images, videos, and audio. Since fake news often contains manipulated visual or multimedia content, combining multiple data sources can improve the detection of misleading information and increase overall classification accuracy.

The system can also be enhanced with real-time fake news detection, allowing continuous monitoring of online news portals and social media platforms to identify and flag suspicious content as it spreads. This capability would support early detection and help reduce the rapid dissemination of misinformation.

To improve transparency and user trust, Explainable Artificial Intelligence (XAI) techniques can be incorporated. These methods provide explanations for the model's predictions, enabling users to understand why a particular news article has been classified as real or fake.

Furthermore, advanced deep learning architectures such as Graph Neural Networks (GNNs) and next-generation transformer-based models (including improved BERT variants and other large language models) can be employed to better capture contextual information, semantic relationships, and propagation patterns. These models can also

improve robustness against adversarial attacks and AI-generated fake news.

For wider accessibility, the proposed system can be deployed as a browser extension, mobile application, or web-based platform, enabling users to verify the authenticity of news articles before sharing them. Future versions may also integrate fact-checking APIs and trusted news sources to provide supporting evidence along with the prediction results.

Overall, the future scope of this project lies in developing an intelligent, scalable, transparent, and real-time fake news detection framework capable of effectively combating misinformation and promoting trust in digital information ecosystems.

REFERENCES

[1] Polu, Omkar Reddy. "AI-Based Fake News Detection Using NLP." International Journal of Artificial Intelligence & Machine Learning (IJAIML), vol. 3, no. 2, IAEME Publication, 2024

[2] Ali, Shahwan Younis, and Ibrahim Mahmood. "The Role of NLP in Fake News Detection and Misinformation Mitigation." International Journal of Scientific Research in Computer Science, Engineering and Information Technology (IJSRCSEIT), vol. 11, no. 3, 2025.

[3] Wang, Yuxiang, Y. Zhang, X. Li, and Xinyao Yu. "COVID-19 Fake News Detection Using BERT-based Models." Proceedings of the International Conference on Artificial Intelligence and Big Data (ICAIBD), IEEE, 2021.

[4] Mayank, M., S. Sharma, and R. Sharma. "DEAP-FAKED: Knowledge Graph-Based Approach for Fake News Detection." Proceedings of the International Conference on Computing, Communication and Intelligent Systems (ICCCIS), IEEE, 2021.

[5] Jadhav, P., and R. K. Shukla. "Deep Learning Analysis for Revealing Fake News Using Linguistic Complexity and Semantic Signatures." Journal of Information and Optimization Sciences, vol. 44, no. 8, Taylor & Francis, 2023.

[6] Al-Zubaidi, F., et al. "Stylometric Fake News Detection Based on NLP Using Named Entity Recognition: In-Domain and Cross-Domain Analysis." IEEE Access, vol. 11, IEEE, 2023.