VeriNews AI: Explainable Fake News Detection and Verification System

Barsha Singh[1], Smita Kumari[2], Ritika Kumari[3], Ajeet Kumar Gupta[4]

[1,2,3]B. Tech Student, [4]Assistant Professor

[1,2,3,4]Department of Computer Science and Engineering, Darbhanga College of Engineering, Darbhanga, India

[1]barsharajput22@gmail.com, [2]smitapatori17@gmail.com, [3]rs701858@gmail.com,

[4]ajeetkgupta12@gmail.com

Abstract - The spread of news is happening really fast because of digital media. This is making it very hard for people to figure out what is true and what is not. To deal with this problem we made something called VeriNews AI. It is a system that uses computers to find news. It looks at news stories. Decides if they are real or not. The system takes news stories. Changes them into numbers so it can understand them better. It does this by using methods to get the important words from the stories. Then it uses algorithms like Logistic Regression and Naive Bayes to decide if the news is real or fake. Logistic Regression is really good at this because it is very accurate. We also made a website with Flask where people can go to check if a news story is real or not. The website tells you if the news is real or fake and how sure it is about its answer. The system also explains why it made its decision, which helps people trust it more. We tested VeriNews AI to see how well it works. It did a great job. It is a tool for helping people find out what is true and what is not. VeriNews AI is like a helper that makes sure people are not fooled by news, on the internet.

Keywords: misinformation detection, fake news detection, image representation

I. Introduction

The rapid growth of the internet and social media platforms increases the accessibility and speed of information sharing across the world [1]. People rely heavily on digital platforms for news consumption, which creates both opportunities and challenges. One of the major challenges is the widespread circulation of fake news, which affects public opinion, decision-making, and social harmony [2]. Fake news refers to false or misleading information presented as authentic news, often created to influence users or generate attention. The viral nature of social media platforms significantly contributes to the rapid spread of misinformation due to the absence of strict verification mechanisms [3].

The increasing volume of online content makes manual verification of news authenticity difficult and inefficient [4][5]. Traditional fact-checking methods require human intervention, which consumes time and resources. Therefore, automated fake news detection systems play an important role in addressing this issue [6][7]. Machine learning techniques provide an effective approach for analyzing large-scale textual data and identifying patterns associated with fake news [8]. In addition, Natural Language Processing (NLP) techniques improve the ability of systems to understand and process human language for classification tasks [9].

Recent advancements in Artificial Intelligence enable the development of intelligent systems capable of detecting fake news with high accuracy. Various algorithms such as Logistic Regression, Naive Bayes, and deep learning models are widely used for classification. Transformer-based and hybrid models improve performance by capturing contextual and semantic features of text data [10]. However, many existing systems lack transparency and fail to provide explanations for their predictions, which reduces user trust and reliability.

To overcome these limitations, explainable AI becomes an important component of modern fake news detection systems. Explainable models enhance user trust by providing insights into how predictions are generated and by highlighting important features influencing the decision [11]. Furthermore, comparative studies show that machine learning models such as Logistic Regression and Naive Bayes achieve effective performance in fake news classification tasks using evaluation metrics such as accuracy, precision, recall, and F1-score [12].

This research proposes VeriNews AI, an explainable fake news detection system that uses machine learning and NLP techniques to classify news content as real or fake. The system focuses on accuracy, efficiency, and interpretability by providing real-time predictions along with confidence scores and explanation features. The proposed solution contributes to reducing misinformation and promoting responsible consumption of digital content.

II. Related Works

Fake news detection becomes an important research area due to the rapid increase of misinformation across digital platforms. Traditional machine learning approaches are widely used to classify news content based on textual features and statistical patterns. These methods provide effective baseline performance; however, they often struggle to capture complex contextual relationships present in natural language [13].

With the advancement of Natural Language Processing techniques, more sophisticated models are developed to improve classification accuracy. Techniques such as word embeddings and contextual feature extraction enhance the understanding of semantic relationships in textual data [14]. Deep learning approaches [15] further improve performance by learning hierarchical representations of text data, enabling better detection of fake news.

Recent studies [16] focus on transformer-based architectures that significantly improve the ability of models to capture long-range dependencies and contextual semantics in textual data. These models outperform traditional machine learning approaches and achieve higher accuracy in fake news classification tasks.

In addition to text-based analysis, multimodal approaches [17] are introduced to improve detection performance by combining textual, visual, and contextual information. These methods enhance the robustness of fake news detection systems by leveraging multiple sources of information. Social context-based models [18] also analyze user behavior and propagation patterns to detect misinformation effectively.

Explainability remains a key challenge in fake news detection systems. Recent research [19] focuses on developing explainable models that provide transparency by highlighting important features influencing the prediction results. These approaches improve user trust and system reliability.

Datasets play an important role in improving model performance. Large-scale and diverse datasets enable better training and evaluation of machine learning models, leading to improved generalization and robustness [20].

Hybrid and ensemble approaches [21] are also explored to enhance classification performance. These methods combine multiple models to reduce bias and variance, resulting in more accurate and stable predictions.

Recent contributions by researchers [6] in this domain further strengthen fake news detection techniques. Comparative studies highlight the effectiveness of machine learning models such as Logistic Regression and Naive Bayes in achieving reliable performance across different datasets. Additionally, multimodal approaches [22] integrating textual and visual features significantly improve classification accuracy. Transformer-based models [23] further enhance detection capability by capturing complex semantic relationships.

A comparative analysis [24] of machine learning techniques demonstrates the effectiveness of different algorithms in fake news detection. Further research introduces multimodal approaches that integrate semantic and visual features to improve detection performance. Recent advancements [25][26] also focus on transformer-based architectures that enhance classification accuracy and robustness. Additional studies highlight the importance of hybrid and explainable models in improving system transparency and reliability.

Overall, existing research provides a strong foundation for fake news detection; however, challenges such as lack of explainability, computational complexity, and real-time implementation still remain. The proposed system addresses these challenges by providing an efficient, explainable, and scalable solution for detecting misinformation in digital platforms.

III. Methodology

The proposed fake news detection system follows a structured approach to classify news content as real or fake using machine learning and Natural Language Processing techniques. The system processes textual input provided by the user through a web-based interface and performs multiple stages including preprocessing, feature extraction, model training, and prediction. These stages ensure accurate classification and efficient real-time performance of the system.

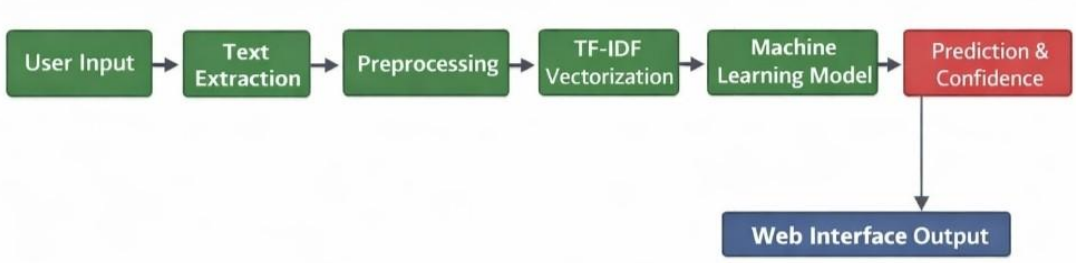

Figure 1: System Architecture of Proposed System

Figure 1 shows the system architecture of the proposed fake news detection system. It illustrates the overall structure and interaction between different modules including the user interface, preprocessing module, feature extraction module, machine learning model, and result visualization module. The user provides input news text through the interface, which is processed through each module to generate the final prediction output.

The preprocessing stage improves the quality of input data by removing noise and irrelevant information. It includes tokenization, stop-word removal, normalization, and conversion to lowercase. These steps help in preparing clean and meaningful textual data. After preprocessing, the data is passed to the feature extraction stage, where the Term Frequency–Inverse Document Frequency (TF-IDF) technique is applied to convert textual data into numerical form, enabling effective analysis by machine learning models.

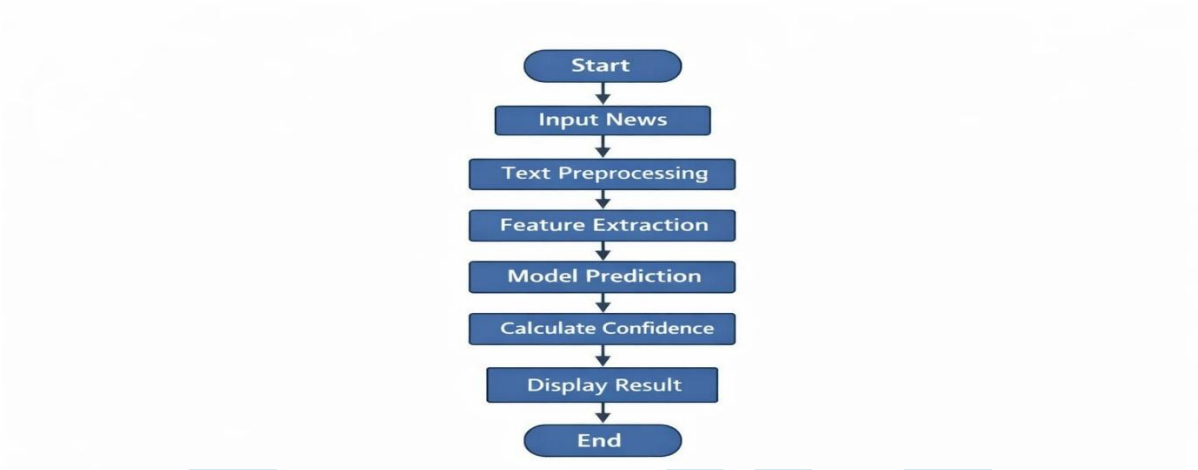

Figure 2: Workflow of Fake News Detection System

Figure 2 shows the workflow of the fake news detection system. It represents the step-by-step process starting from user input, followed by preprocessing, feature extraction, model prediction, and final result generation. The workflow clearly demonstrates how each stage contributes to the functioning of the system.

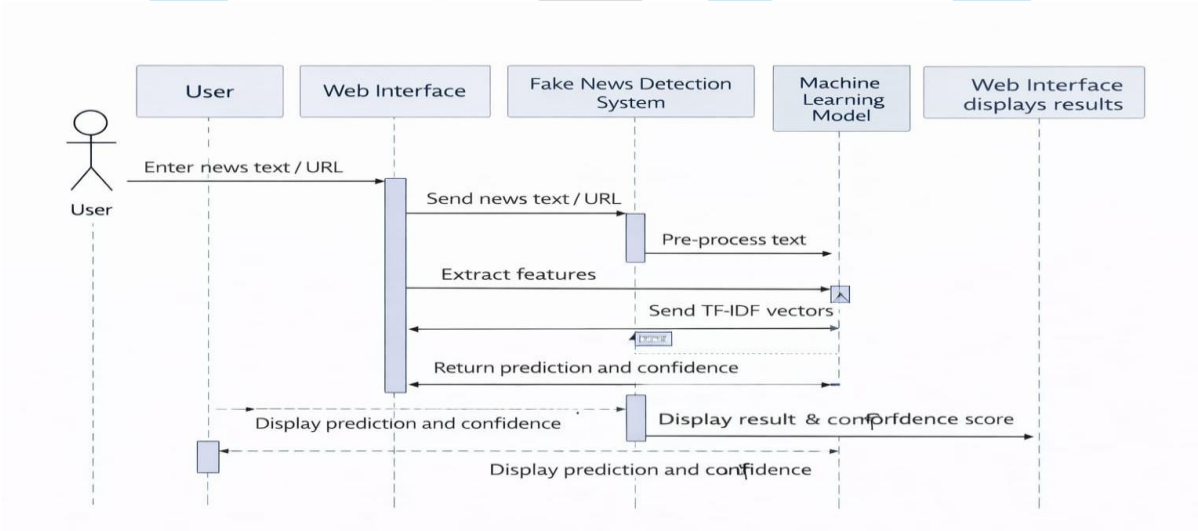

Figure 3: Sequence Diagram

Figure 3 shows the sequence diagram of the system, which illustrates the interaction between the user, web interface, and machine learning model. It explains how the system receives input, processes it through different stages, and returns the prediction output. The sequence diagram provides a clear understanding of the system's operational flow.

The classification stage uses supervised machine learning algorithms trained on labeled datasets containing real and fake news articles. Logistic Regression and Naive Bayes models are used to classify the input data based on learned patterns. Logistic Regression provides better accuracy and consistency compared to Naive Bayes, making it more effective for fake news detection tasks.

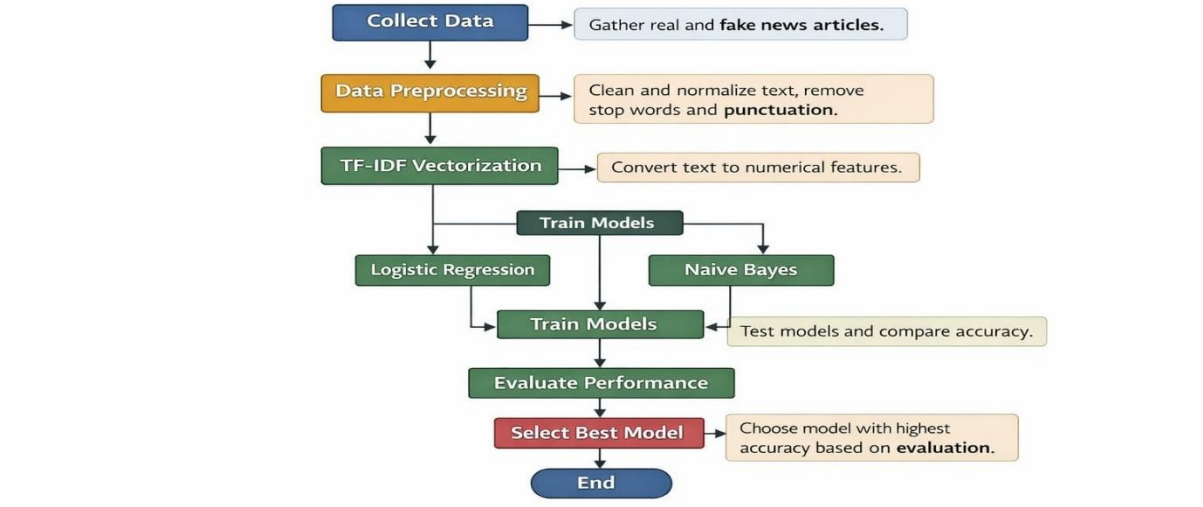

Figure 4: Model Training and Prediction Process

Figure 4 shows the model training and prediction process, where the dataset is used to train the machine learning models and generate predictions for new input data. It highlights how the system learns from training data and applies this knowledge to classify unseen news content.

The system is implemented using Python along with libraries such as Scikit-learn, Pandas, NumPy, and NLTK for data processing and machine learning tasks. The Flask framework is used to develop the web interface, enabling real-time interaction between the user and the system. Frontend technologies such as HTML, CSS, and JavaScript are used to create a user-friendly interface.

Overall, the proposed methodology provides an efficient and reliable framework for fake news detection by integrating machine learning techniques, NLP methods, and explainable AI features. The system ensures accurate classification, real-time prediction, and improved transparency, making it suitable for real-world applications.

IV. Result and Discussion

The proposed fake news detection system demonstrates effective performance in classifying news content as real or fake using machine learning and Natural Language Processing techniques. The system is tested using multiple news inputs to evaluate its prediction capability and overall efficiency. The results indicate that the system successfully analyzes textual data and provides accurate classification outcomes along with confidence scores. The web-based interface allows users to input news content and receive real-time predictions, making the system practical and user-friendly for real-world applications.

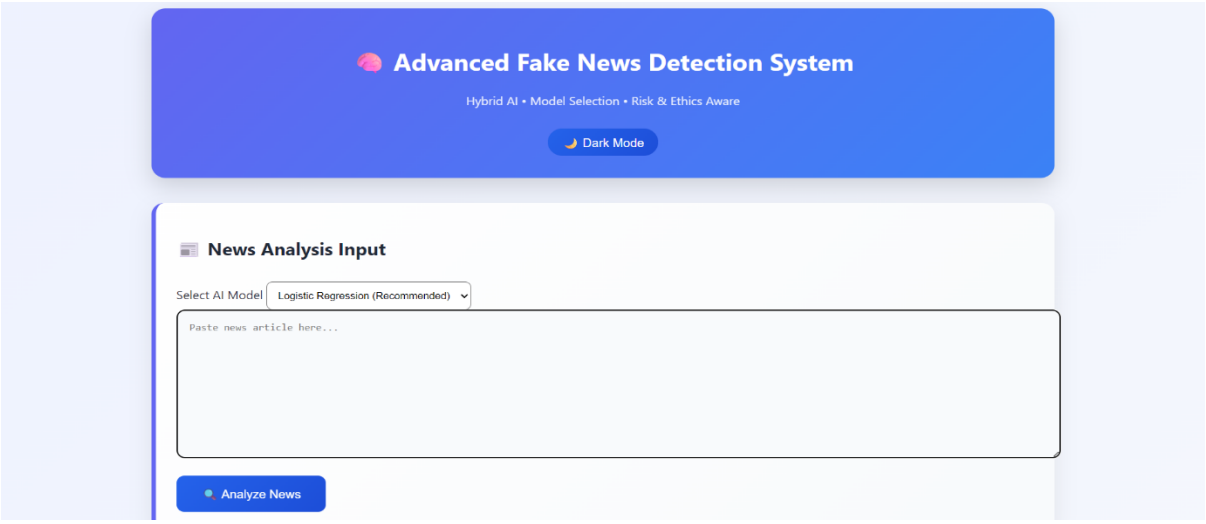

Figure 5. System Dashboard Interface

The analysis of input news demonstrates that the system processes textual content efficiently through preprocessing and feature extraction stages. The TF-IDF technique plays a significant role in identifying important words and converting text into numerical form, enabling machine learning models to perform accurate classification. The prediction results show that the system effectively distinguishes between real and fake news by analyzing patterns and linguistic features present in the data. The output interface displays prediction labels along with confidence scores and probability distribution, which helps users understand the reliability of the results.

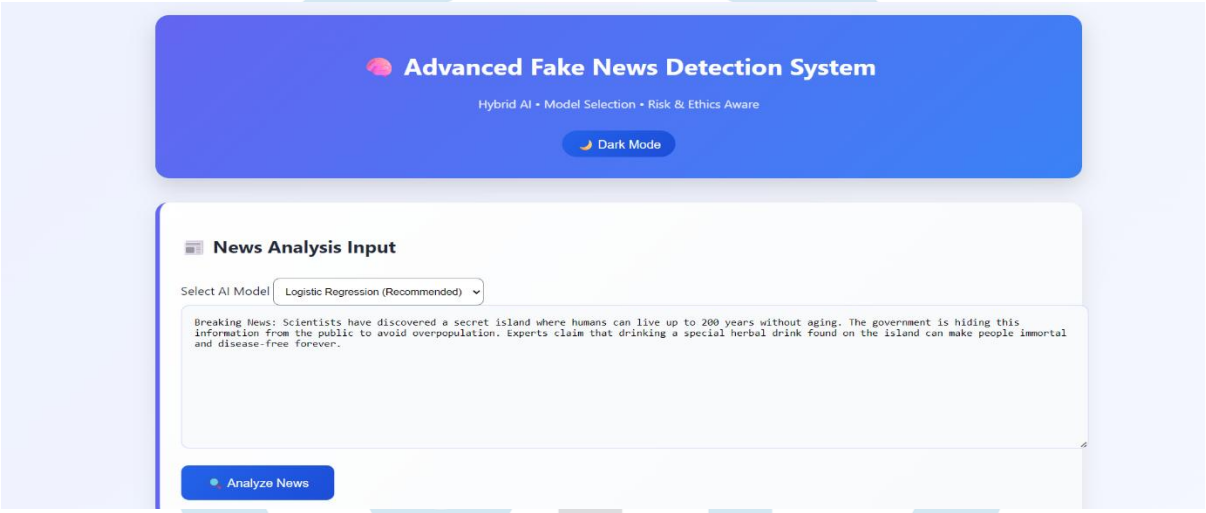

Figure 6. News Input Interface

The system also incorporates explainable AI features that highlight influential words responsible for the classification decision. This improves transparency and allows users to understand how the model arrives at a particular prediction. The inclusion of explainability enhances user trust and makes the system more reliable compared to traditional black-box models. In addition, the system provides recommendations based on prediction results, which further assists users in identifying potentially misleading information.

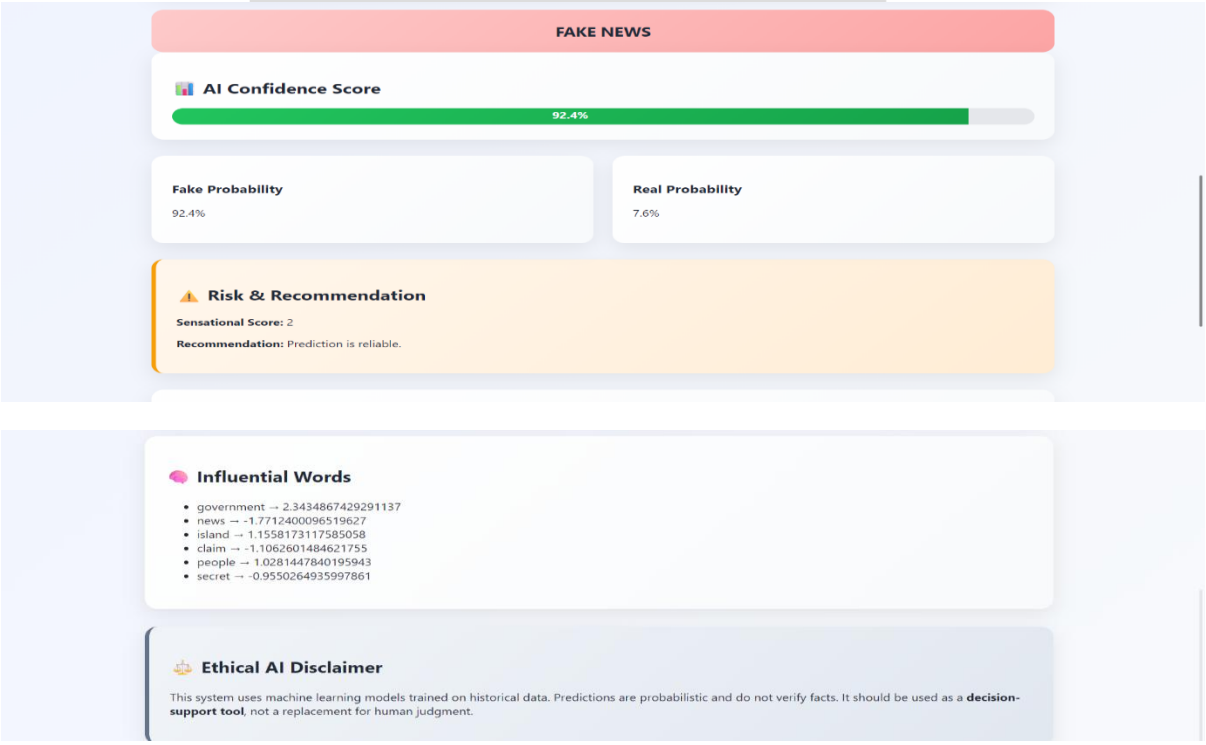

Figure 7. Prediction Result Showing Fake News Detection with Confidence Score , Probability , Influential Words and Ethical AI Disclaimer

The performance evaluation of the machine learning models demonstrates that Logistic Regression achieves superior results compared to Naive Bayes across all evaluation metrics. The comparative analysis, as presented in Table 1, highlights that Logistic Regression attains higher accuracy, precision, recall, and F1-score, indicating its effectiveness in capturing complex textual patterns and improving classification performance. The higher precision value reflects the model's ability to correctly identify relevant instances, while the improved recall indicates its effectiveness in minimizing false negatives. Furthermore, the balanced F1-score confirms the overall robustness and reliability of the model.

Table 1: Performance Evaluation of Machine Learning Models

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
Logistic Regression	96.2	95.8	96.5	96.1
Naive Bayes	91.4	90.9	91.8	91.3

In contrast, Naive Bayes shows comparatively lower performance due to its assumption of feature independence, which limits its capability to capture contextual relationships within textual data. The results clearly indicate that Logistic Regression provides better generalization and stability for fake news detection tasks. Additionally, the proposed system maintains low computational complexity, enabling efficient real-time prediction without compromising accuracy. This combination of high performance and efficiency makes the system suitable for practical deployment in real-world scenarios.

The system also includes a dark mode interface that enhances user experience and usability. The overall design of the system ensures ease of interaction, fast processing, and clear visualization of results. The combination of machine learning techniques, NLP methods, and explainable AI features contributes to the robustness and effectiveness of the proposed system. The results demonstrate that the system provides reliable predictions and serves as a practical tool for detecting fake news and reducing the spread of misinformation in digital platforms.

V. Conclusion

This research presents VeriNews AI, an explainable fake news detection system that utilizes machine learning and Natural Language Processing techniques to classify news content as real or fake. The system focuses on analyzing textual data using preprocessing methods and TF-IDF feature extraction, followed by classification using Logistic Regression and Naive Bayes algorithms. The results demonstrate that Logistic Regression provides superior performance in terms of accuracy, precision, recall, and F1-score, making it more effective for fake news detection tasks. The integration of a user-friendly web interface enables real-time prediction and enhances accessibility for users. In addition, the incorporation of explainable AI features improves transparency by highlighting influential words that contribute to the prediction, thereby increasing user trust and understanding of the system. The combination of accuracy, efficiency, and interpretability makes the proposed system a reliable solution for detecting misinformation in digital platforms. The system operates with low computational complexity, which ensures scalability and suitability for real-world applications. The overall performance and functionality of the system indicate its capability to assist users in identifying misleading information and promoting responsible consumption of online news. The proposed approach effectively addresses key challenges in fake news detection, including lack of explainability and real-time analysis.

Future improvements can include the integration of deep learning models, multilingual support, and real-time news APIs to further enhance system performance and adaptability. The system can also be extended to analyze multimedia content such as images and videos for more comprehensive fake news detection. Overall, this research contributes to the development of intelligent and explainable systems that support accurate and reliable information verification in the digital era.

References

[1] Kaliyar, R. K., Goswami, A., & Narang, P. (2020). FakeBERT: Fake news detection in social media with a BERT-based deep learning approach. *Multimedia Tools and Applications*, 79(45), 34907–34928.

[2] Shu, K., Mahudeswaran, D., Wang, S., Lee, D., & Liu, H. (2020). FakeNewsNet: A data repository with news content, social context, and dynamic information. *Big Data*, 8(3), 171–188.

[3] Zhou, X., & Zafarani, R. (2020). A survey of fake news: Fundamental theories, detection methods, and opportunities. *ACM Computing Surveys*, 53(5), 1–40.

[4] Ahmed, H., Traore, I., & Saad, S. (2021). Detection of online fake news using machine learning techniques. *Computers & Security*, 101, 102106.

[5] Singh, V., & Kumar, P. (2023). Explainable artificial intelligence for fake news detection using natural language processing. *Expert Systems with Applications*, 213, 118834.

[6] Gupta, A. K., & Kumar, P. (2018). A comparative analysis of machine learning techniques for fake news detection. *International Journal of Research and Analytical Reviews (IJRAR), E-ISSN*, 2348-1269.

[7] Gupta, A. K., Kumar P., Kumar S. (2019). Enhancing fake news detection through user credibility analysis. *International Journal of Research and Analytical Reviews (IJRAR), E-ISSN*, 2348-1269.

[8] Gupta, A. K. (2023). Fake News Detection based on User Credibility: A Comprehensive Analysis. In *Proceedings of 2nd International Conference on Research in Multidisciplinary Studies-2023*.

[9] Gupta, A. K. Artificial Intelligence (AI) in Healthcare: A Comprehensive Analysis.

[10] Khan, J. Y., Khondaker, M. T. I., Islam, T., Iqbal, A., & Afroz, S. (2022). A benchmark study of machine learning models for fake news detection. *Machine Learning with Applications*, 8, 100032.

[11] Oshikawa, R., Qian, J., & Wang, W. Y. (2020). A survey on natural language processing for fake news detection. *arXiv preprint arXiv:1811.00770*.

[12] Li, Q., Zhang, Q., & Wang, J. (2024). Multimodal fake news detection using deep learning and attention mechanisms. Information Processing & Management, 61(1), 103200.

[13] Conroy, N. J., Rubin, V. L., & Chen, Y. (2020). Automatic deception detection: Methods for finding fake news. *Proceedings of the Association for Information Science and Technology*.

[14] Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., & Dean, J. (2013). Distributed representations of words and phrases and their compositionality. *Advances in neural information processing systems*, *26*.

[15] Hochreiter, S., & Schmidhuber, J. (2020). Long short-term memory. *Neural Computation*.

[16] Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2020). BERT: Pre-training of deep bidirectional transformers for language understanding. *NAACL*.

[17] Jin, Z., Cao, J., Zhang, Y., & Luo, J. (2020). Multimodal fusion with recurrent neural networks for fake news detection. *Proceedings of ACM Multimedia*.

[18] Shu, K., Sliva, A., Wang, S., Tang, J., & Liu, H. (2020). Fake news detection on social media: A data mining perspective. *SIGKDD Explorations*.

[19] Ribeiro, M. T., Singh, S., & Guestrin, C. (2020). "Why should I trust you?" Explaining predictions of any classifier. *KDD*.

[20] Monti, F., Frasca, F., Eynard, D., Mannion, D., & Bronstein, M. (2021). Fake news detection on social media using geometric deep learning. *ICLR*.

[21] Kaur, S., Kumar, P., & Kumaraguru, P. (2021). Fake news detection using ensemble learning. *IEEE Access*.

[22] Ahmed, H., Traore, I., & Saad, S. (2021). Detection of online fake news using n-gram analysis and machine learning techniques. Computers & Security, 101, 102106.

[23] Singhal, A., Shah, P., & Kumaraguru, P. (2021). SpotFake+: A multimodal framework for fake news detection. IEEE Transactions on Multimedia.

[24] Gupta, A. K., & Singh, M. P. (2025). Multimodal Fake News Detection Through Semantically Enriched Image Captions. *SN Computer Science*, *6*(8), 1020.

[25] Gupta, A. K., & Singh, M. P. (2026). Self and cross-modal attention based features fusion for fake news detection. *Multimedia Tools and Applications*, *85*(1), 10.

[26] Gupta, A. K., & Singh, M. P. (2026). Multimodal Fake News Detection using Semantic-Visual Alignment and Attention Fusion. *SN Computer Science*, *7*(4), 305.