INTERNATIONAL JOURNAL OF ADVANCE RESEARCH, IDEAS AND INNOVATIONS IN TECHNOLOGY

ISSN: 2454-132X
Impact Factor: 6.078
(Volume 12, Issue 2 - V12I2-1258)
Available online at: https://www.ijariit.com

Detecting Misinformation in News Using BERT and Natural Language Processing

Ankannagari Harshith Reddy
harshithreddy12377@gmail.com
Gokaraju Rangaraju Institute of Engineering and Technology, Telangana

Tabitha Indupalli
tabitha1669@grietcollege.com
Gokaraju Rangaraju Institute of Engineering and Technology, Telangana

Dinesh Ragipani
dineshragipani143@gmail.com
Gokaraju Rangaraju Institute of Engineering and Technology, Telangana

T.Dheeraj
dt7768181@gmail.com
Gokaraju Rangaraju Institute of Engineering and Technology, Telangana

Ch.Bhanu uday
bhanudaynaidu980@gmail.com
Gokaraju Rangaraju Institute of Engineering and Technology, Telangana

ABSTRACT

The widespread use of social media and online news platforms has made it easier for misinformation and fake news to spread rapidly. This creates serious challenges for individuals and organizations that rely on accurate information. To address this problem, this study proposes a fake news detection system that combines Natural Language Processing (NLP) techniques with both traditional machine learning and transformer-based models. The dataset used for the study is derived from the WELFake dataset, containing labeled news articles categorized as real or fake. Text preprocessing techniques such as tokenization, removal of noise, and normalization are applied to prepare the data. Traditional models like Support Vector Machine (SVM) and LightGBM use TF-IDF features to capture important word patterns, while DistilBERT is used to understand contextual meaning in text. The results show that transformer-based models achieve higher accuracy, while traditional models remain efficient and reliable. This hybrid approach improves the overall effectiveness of fake news detection systems.

Keywords: *Fake News, WELFake dataset, DistilBERT, Support Vector Machine, LightGBM, TF-IDF.*

INTRODUCTION

In the modern times, most of the news is disseminated through social media and the internet. Even though these formats give individuals access to more information, they may be the cause of the rapid spread of fake news. Such misinformation may give the audience the wrong conclusions, make them confused, and even influence the social structures and political and economical systems. The fake news detection is a valuable strata of research because of these issues. Simple and efficient machine learning approaches such as Logistic Regression and Naive Bayes are usually applied in classification of text. However, they do not tend to get the deeper meaning of text. BERT and its smaller framework, DistilBERT are both transformer-based models that are highly adept at sufficient language understanding, at the sacrifice of higher costs of operation. The aim of this project was to develop a hybrid model, which employed both the traditional machine learning and DistilBERT. The ensemble learning is a combination of the advantages of two models into a system. On top of that, to your particular problem you are keen on creating a fake news detection model that would not only be accurate but also have broad results.

LITERATURE REVIEW

The features of the text were known and trained with the help of SVM and LightGBM, which effectively find various linguistic patterns and are computationally efficient and understandable through NLP-based preprocessing such as tokenization, stop-word removal, and vectorization. DistilBERT was employed in order to enhance situational understanding - a smaller variant of the BERT model, but nonetheless very efficient and able to identify misleading information at different levels of abstraction due to its deep semantic validation capability. The outcome was a fruitful tradeoff between speed and semantic relevance - SVM with LightGBM were the fast interpretable baselines to compare them with, and DistilBERT gave them the ability to flexibly add vocabulary generalization to those sets. The hybrid framework was very scalable, interpretable, and rich in the context. The last system is a powerful and robust model of automated fake news detection, which can be extended in the future, to the multilingual and multimodal systems.

In *Fake News Detection Using Logistic Regression and Decision Tree Algorithm* (2025), the authors developed a fake news detection system using basic machine learning models on a Kaggle dataset. The study applied preprocessing techniques such as tokenization and stop-word removal before training. The results showed that simple models can effectively classify fake news. However, the use of a limited and domain-specific dataset restricts the model's generalization capability. Future improvements could involve larger and more diverse datasets to enhance performance.

In *A Comprehensive Survey on Machine Learning Approaches for Fake News Detection* (2025), Jawaher Alghamdi et al. presented an extensive review of fake news detection techniques using machine learning and deep learning approaches. The study explored different definitions, feature types, datasets, and psychological aspects influencing fake news spread. It also categorized detection methods into content-based, context-based, and hybrid approaches. Despite its comprehensive coverage, the study highlights challenges such as evolving misinformation strategies and the need for more generalized and scalable solutions.

In *Romanian Fake News Detection Using Machine Learning and Transformer-Based Approaches* (2024), Elisa Valentina Moisi et al. conducted a comparative analysis of machine learning and transformer models on Romanian datasets. The study used models such as Naive Bayes, Logistic Regression, SVM, BERT, and RoBERTa, evaluated using standard metrics. The results showed that BERT achieved the highest accuracy, while SVM also performed competitively. However, the study is limited to Romanian-language datasets, which may reduce its applicability to other domains.

In *Ensemble Techniques for Robust Fake News Detection* (2024), Mohammed Al-alshaqi et al. proposed a multimodal fake news detection framework integrating text, image, and video data. The approach combined machine learning models with BERT and a modified CNN for enhanced performance. Experimental results showed that the multimodal approach outperformed traditional methods. However, the increased complexity and computational cost of multimodal systems may limit their real-world deployment.

In *Constructing a User-Centered Fake News Detection Model* (2023), Minjung Park and Sangmi Chai proposed a model that incorporates user behavior, content, and social network features. The study used XGBoost for feature importance and evaluated multiple classifiers including SVM, Random Forest, and Neural Networks. The results showed that Random Forest achieved the highest accuracy. However, the reliance on user and social data may limit applicability where such information is unavailable.

In *A Comparative Study of Machine Learning and Deep Learning Techniques for Fake News Detection* (2022), Jawaher Alghamdi et al. compared various machine learning, deep learning, and transformer-based models across multiple datasets. The study evaluated models such as SVM, Logistic Regression, CNN, BiLSTM, and BERT. Results indicated that transformer-based models achieved superior performance due to better contextual understanding. However, these models come with higher computational costs and scalability challenges.

In *Fake News Detection Using Majority Voting Technique* (2022), Dharmaraj R. Patil proposed a multi-model approach using a majority voting ensemble of several machine learning classifiers, including Decision Tree, Logistic Regression, SVM, Random Forest, and XGBoost. The study utilized a publicly available dataset of over 20,000 news articles and extracted textual features for classification. The results showed that the ensemble method achieved high performance, with an accuracy of 96.38%, outperforming individual models. However, the approach relies primarily on textual features and may not capture deeper contextual or semantic information.

In *Evaluating BERT-based Pre-training Language Models for Detecting Misinformation* (2022), Rini Anggrainingsih, Ghulam Mubashar Hassan, and Amitava Datta proposed a transformer-based approach using BERT, RoBERTa, and DistilBERT for misinformation detection. The study evaluated these models on both short and long text datasets and compared their performance with traditional methods. Results showed that pre-trained language models outperformed existing techniques, and even simple classifiers achieved strong results due to rich contextual embeddings. However, the approach depends heavily on dataset size and proper data splitting, which significantly influence model performance.

The study *"Machine Learning-Based Approach for Fake News Detection"* (2024) by H. L. Gururaj, H. Lakshmi, B. C. Soundarya applies multiple machine learning algorithms to analyze textual features and detect fake news from news article datasets. The paper provides a comparative evaluation of algorithms based on accuracy and interpretability. It also discusses dataset characteristics, model explainability, and trade-offs between performance and transparency.

In Fake Media Detection Based on Natural Language Processing and Blockchain Approaches (2021), Zeinab Shahbazi and Yung-Cheol Byun proposed an integrated framework combining NLP, machine learning, reinforcement learning, and blockchain technology. The approach aims to detect fake news as well as identify fake user accounts and posts while ensuring data authenticity through a decentralized blockchain system. The study highlights improved security and trust in information sharing platforms. However, the integration of multiple complex technologies may increase system complexity and affect scalability in real-world applications.

In Fake News Stance Detection Using Deep Learning Architecture (CNN-LSTM) (2020), Muhammad Umer et al. proposed a hybrid deep learning model combining CNN and LSTM with dimensionality reduction techniques such as PCA and Chi-square. The model was evaluated on the Fake News Challenge (FNC) dataset to analyze the stance between news headlines and articles. Results showed that PCA-based features achieved superior performance with an accuracy of 97.8%, outperforming existing methods. However, the approach focuses mainly on stance detection and may not generalize well to broader fake news classification tasks.

From the reviewed studies, several important limitations can be identified in existing fake news detection approaches. Many works rely on limited, language-specific, or domain-specific datasets, which restrict the generalization of models across different platforms and real-world scenarios. A significant number of studies focus either on traditional machine learning models or advanced deep learning techniques, but fail to effectively combine the strengths of both approaches. While ensemble and multimodal methods improve performance, they often introduce increased computational complexity and are difficult to deploy in practical environments. Some approaches also depend heavily on additional data such as user behavior or social network information, which may not always be available. Furthermore, several models primarily focus on textual features or specific tasks like stance detection, limiting their applicability to broader fake news classification problems. These limitations highlight the need for a balanced, scalable, and efficient hybrid framework that can leverage both contextual understanding and computational

efficiency for real-world fake news detection.

The literature on fake news detection demonstrates a clear shift from traditional machine learning techniques to more advanced deep learning and transformer-based approaches. Early methods such as Support Vector Machines (SVM), Logistic Regression, and Decision Trees were widely used due to their efficiency, simplicity, and interpretability. However, these models often struggled to capture the contextual and semantic relationships present in textual data. With the advancement of deep learning, models such as Long Short-Term Memory (LSTM) networks and transformer-based architectures like BERT, ALBERT, and XLNet have been introduced, significantly improving the ability to understand semantic context and achieve higher classification accuracy. Despite these improvements, challenges still remain, particularly in balancing interpretability, computational complexity, and contextual understanding. These limitations highlight the need for more scalable, robust, and generalized approaches to fake news detection.

Table-1: Review of Literature

s.no	Title & Authors (Year)	Techniques Used	Research Gap / Limitation
1	*Fake News Detection Using Logistic Regression and Decision Tree Algorithm* — Arihara Suthan, Sri Devi (2025)	Logistic Regression, Decision Tree, NLP preprocessing	Limited dataset size reduces generalization and robustness
2	*A Comprehensive Survey on Machine Learning Approaches for Fake News Detection* — Jawaher Alghamdi et al. (2025)	ML, DL, Transformer models (survey)	Lacks implementation; highlights need for scalable and generalized solutions
3	*Romanian Fake News Detection Using Machine Learning and Transformer-Based Approaches* — Elisa Valentina Moisi et al. (2024)	NB, LR, SVM, BERT, RoBERTa	Language-specific dataset limits applicability across domains
4	*Ensemble Techniques for Robust Fake News Detection* — Mohammed Al-alshaqi et al. (2024)	BERT, CNN, Random Forest, Multimodal learning	High computational complexity; difficult real-world deployment
5	*Constructing a User-Centered Fake News Detection*	XGBoost, SVM, RF, LR, Neural Networks	Requires user/social data, limiting practical usage

	Model — Minjung Park, Sangmi Chai (2023)		
6	*A Comparative Study of Machine Learning and Deep Learning Techniques for Fake News Detection* — Jawaher Alghamdi et al. (2022)	SVM, LR, CNN, BiLSTM, BERT, RoBERTa	High computational cost of deep learning models
7	*Fake News Detection Using Majority Voting Technique* — Dharmaraj R. Patil (2022)	Ensemble ML (DT, LR, RF, SVM, XGBoost, NB)	Focuses only on textual features; lacks contextual understanding
8	*Evaluating BERT-based Pre-training Language Models for Detecting Misinformation* — Rini Anggrainingsih et al. (2022)	BERT, RoBERTa, DistilBERT	Performance depends heavily on dataset size and data splitting
9	*Fake Media Detection Based on Natural Language Processing and Blockchain Approaches* — Zeinab Shahbazi, Yung-Cheol Byun (2021)	NLP, Blockchain, Reinforcement Learning	High complexity and scalability issues
10	*Fake News Stance Detection Using Deep Learning Architecture (CNN-LSTM)* — Muhammad Umer et al. (2020)	CNN, LSTM, PCA, Chi-square	Focused on stance detection, not full classification

DATA COLLECTION AND PREPROCESSING

The dataset used in this study is derived from the publicly available **WELFake dataset**, which is widely used as a benchmark for fake news detection research. The dataset contains news articles collected from various online sources, including political news websites and fact-checking platforms such as PolitiFact. Each record in the dataset consists of the textual content of a news article along with a binary label indicating whether the news is real or fake. This labeled structure allows machine learning models to learn patterns that help distinguish between authentic and misleading news content.

The dataset was refined to include only the **title and the main textual content** of each news article. These two fields were combined into a single input sequence to preserve the contextual flow of the information. Data cleaning procedures were performed to improve dataset quality by removing duplicate records and incomplete entries. The dataset was then encoded using **binary labels**, where **0 represents real news and 1 represents fake news**.

Before training the machine learning models, several preprocessing steps were applied to prepare the textual data. All text was converted to **lowercase** to maintain consistency in representation. URLs and hyperlinks were removed to reduce source-related bias. Non-informative phrases such as “read more” and “click here” were also eliminated since they do not contribute meaningful information to the classification task. In addition, user mentions and hashtags were removed to minimize noise in the dataset and improve the quality of the textual input.

Additional preprocessing steps were applied to further standardize the dataset. Emojis were handled differently depending on the model architecture used. For traditional machine learning models, emojis were removed to prevent the introduction of sparse and noisy features. In contrast, for transformer-based models, emojis were converted into their textual descriptions so that their semantic meaning could be preserved. Furthermore, punctuation marks, non-alphabetic characters, and unnecessary whitespace were removed to ensure cleaner textual input.

After preprocessing, the cleaned text was converted into numerical representations suitable for model training. For traditional machine learning models such as Support Vector Machine (SVM) and LightGBM, **Term Frequency–Inverse Document Frequency (TF-IDF)** vectorization was used to capture important lexical patterns and word importance within the text. In contrast, the transformer-based model **DistilBERT** directly processes tokenized text and learns contextual embeddings without requiring manual feature engineering. This preprocessing and feature representation pipeline ensures that the dataset is well-structured and optimized for both traditional machine learning algorithms and deep learning models used in the study.

This preprocessing pipeline made sure that the data was normalized, de-noised and both traditional machine learning models and deep learning models applied in this study were optimized.

METHODOLOGY

The proposed system introduces a hybrid framework for fake news detection that combines traditional machine learning models with a transformer-based deep learning model. Support Vector Machine (SVM) and LightGBM classifiers are employed to analyze long-form textual content using TF-IDF features. For shorter texts such as tweets and headlines, the transformer-based model DistilBERT is utilized due to its strong ability to capture contextual and semantic relationships within text.

In this framework, textual data for classical models is converted into numerical form using **Term Frequency–Inverse Document Frequency (TF-IDF)** vectorization, which effectively captures significant lexical patterns in the data. In contrast, DistilBERT processes tokenized text directly and learns contextual representations without requiring manual feature engineering. Each model is trained independently using the same preprocessed dataset with consistent label encoding to ensure reliable and comparable predictions.

A rule-based ensemble strategy is applied during the inference stage to combine the predictions from different models. DistilBERT is used for texts containing **forty words or fewer**, as it is well suited for short-form content such as tweets and headlines. For longer texts, the SVM and LightGBM models are used for classification. In cases involving long-form content, the final prediction is determined based on the agreement between the **SVM and LightGBM** models. If the two models produce different predictions, the final decision is made using the model with the higher confidence score. This ensemble strategy improves the overall robustness and accuracy of the system by effectively handling texts of varying lengths.

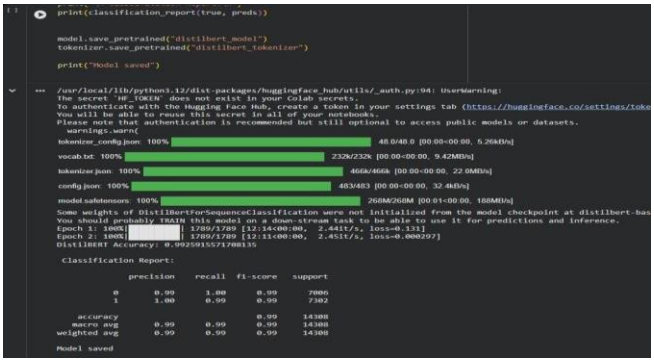
EXPERIMENTS AND RESULTS

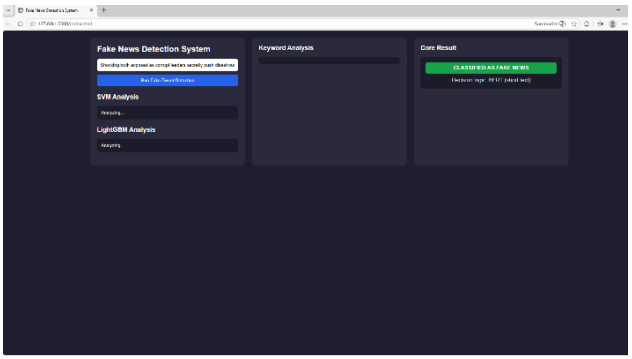
Model	Accuracy
SVM (TF-IDF)	97.48%
LightGBM (TF-IDF)	97.94%
DistilBERT (Transformer)	99.26%

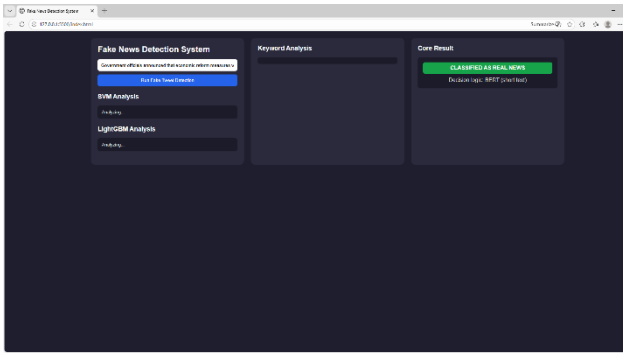
Experiments were conducted to evaluate the effectiveness of three models—**SVM, LightGBM, and DistilBERT**—for domain-specific tweet classification. The dataset was divided into **80% training data and 20% testing data** using stratified sampling to maintain class balance between real and fake news. For the SVM and LightGBM models, tweets were first converted into numerical representations using **TF-IDF vectorization**. In contrast, the DistilBERT model directly processed the raw tweet text using the pre-trained **distilbert-base-uncased tokenizer**, and the model was fine-tuned using **GPU acceleration**. The performance of each model was evaluated using standard classification metrics, including **accuracy, precision, recall, and F1-score**.

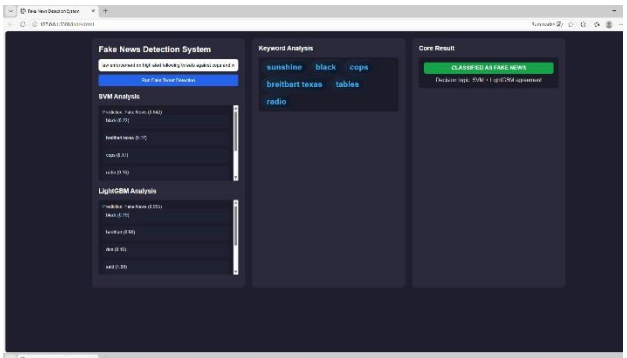
The **SVM model** achieved an accuracy of **97.48%,** demonstrating strong performance in identifying domain-specific tweets based on keyword-oriented features extracted through TF-IDF representation. The **LightGBM model** slightly outperformed SVM with an accuracy of **97.94%,** benefiting from its gradient boosting mechanism, which can capture more complex feature interactions within the TF-IDF feature space. Both traditional machine learning models showed stable and efficient performance in the tweet classification task.

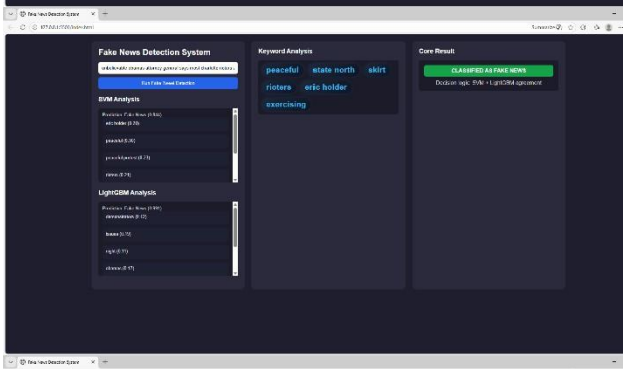
The **DistilBERT model** achieved the highest accuracy of 99.26%, significantly outperforming the TF-IDF–based models. Its transformer architecture enables a deeper understanding of contextual and semantic relationships within short and informal tweet text, leading to superior classification performance. These results indicate that transformer-based approaches are more effective for tweet classification tasks, while traditional machine learning models remain useful in scenarios where computational resources are limited.

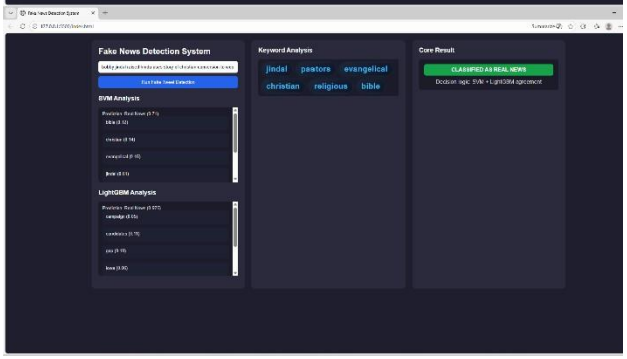


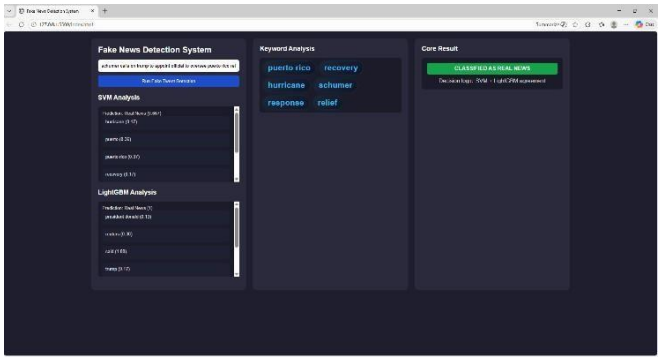












DISCUSSION

The experimental results demonstrate that combining transformer-based models with classical machine learning techniques improves the effectiveness of fake news detection across texts of varying lengths. DistilBERT performed particularly well on short texts such as tweets and headlines, as its architecture enables it to capture contextual and semantic information more effectively. In contrast, traditional models like SVM and LightGBM showed better performance when handling longer articles, where lexical patterns and term distributions are more prominent.

The rule-based ensemble strategy also played an important role in improving prediction reliability. By directing short texts to DistilBERT and longer texts to classical machine learning models, the system reduced misclassification caused by model bias toward specific text formats. Agreement between SVM and LightGBM increased confidence in predictions for long-form content, while a confidence-based fallback mechanism helped stabilize results when the two models produced different predictions. This approach improved the overall robustness and consistency of the fake news detection system.

Despite achieving strong performance, several limitations were observed. The models showed sensitivity to stylistic variations such as emojis, informal language, and non-standard expressions, which occasionally influenced prediction accuracy. Additionally, the effectiveness of the system largely depends on the characteristics of the training dataset, indicating that the use of more diverse and representative data sources could further improve model generalization. Incorporating datasets from different domains and platforms may help address this limitation in future work. Overall, the proposed hybrid framework provides a balanced and interpretable approach for detecting fake news across heterogeneous textual data while maintaining reliable performance.

CONCLUSION

In conclusion, the proposed fake news detection system addresses the growing challenge of misinformation spreading across digital platforms by leveraging advanced Natural Language Processing (NLP) techniques along with both traditional and deep learning models. The framework combines classical machine learning algorithms such as Support Vector Machine (SVM) and LightGBM with the transformer-based model DistilBERT to create a robust and effective detection mechanism. The NLP pipeline includes important preprocessing steps such as tokenization, stop-word removal, noise filtering, and transforming raw textual data into numerical representations suitable for machine learning models. These preprocessing steps help standardize the data and enable the models to focus on meaningful linguistic patterns that distinguish fake news from legitimate information.

The traditional models, SVM and LightGBM, are effective in capturing surface-level textual patterns such as specific keywords, stylistic indicators, and term distributions. Their computational efficiency and interpretability make them suitable for real-time or resource-constrained environments. On the other hand, DistilBERT, a lightweight transformer-based model, is capable of understanding deeper contextual and semantic relationships within text. This enables the system to detect subtle indicators of misinformation, such as biased language, misleading phrasing, or disguised false claims that may not be easily captured by traditional models.

By integrating these approaches through a hybrid and ensemble-based framework, the system balances efficiency, interpretability, and contextual understanding. As a result, the proposed model improves the overall accuracy and robustness of fake news detection. Such systems can play an important role in reducing the spread of misinformation and promoting a more reliable and trustworthy digital information ecosystem.

REFERENCES

[1] Essa, E., Omar, K., & Alqahtani, A. (2023). Fake News Detection based on a Hybrid BERT and LightGBM Model.
[2] Arihara Suthan, R., & Sri Devi, R. (2025). Fake News Detection Using Logistic Regression and Decision Tree Algorithm.
[3] Verma, P. K., Agrawal, P., & Amorim, I. (2022). Word Embedding Over Linguistic Features for Fake News Detection.
[4] Krishna, N. L. S. R., & Adimoolam, M. (2022). Fake News Detection System Using Decision Tree Algorithm and Comparison with SVM Based on Textual Features.
[5] Ashok Kumar, J., Trueman, T. E., & Cambria, E. (2021). Fake News Detection Using XLNet Fine-Tuning Model. https://doi.org/10.1109/ICACCS51430.2021.944175 2
[6] Aldwairi, M., & Alwahedi, A. (2021). Characterization, Classification and Detection of Fake News in Online Social Media Networks. IEEE Access, 9, 140100–140117. https://doi.org/10.1109/ACCESS.2021.3118402
[7] Chowdhury, S., & Noori, S. R. H. (2022). Multi-Class Fake News Detection Using the LSTM Approach. International Journal of Advanced Computer Science and Applications (IJACSA), 13(6). https://doi.org/10.14569/IJACSA.2022.0130624
[8] Agarwal, S., & Sharma, S. (2022). Classifying Fake News Using SVM, Naïve Bayes, and LSTM. https://www.researchgate.net/publication/36 5456789_Classifying_Fake_News_Using_SVM_Nai ve_Bayes_and_LSTM
[9] Gururaj, H. L., Lakshmi, H., & Soundarya, B. C. (2024). Machine Learning-Based Approach for Fake News Detection.
[10] Mimura, M., & Ishimaru, T. (2024). Analyzing Common Lexical Features of Fake News Using Multi-Head Attention Weights. Journal of Information & Data Science. https://doi.org/10.1007/s42488-024-000XX