IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 12, April 2026
Impact Factor: 8.2

Fake News Detection Using AI

Manoharan T[1], Dharanishwaran V[2], Thirugnanam B[3], Chandru M[4], Suresh T[5]
Final Year Students, Department of Computer Science Engineering[1234]
Associate Professor, Department of Computer Science and Engineering[5]
Annamalai University, Annamalai Nagar, Tamil Nadu, India

Abstract: *The rapid proliferation of misinformation in digital media has created a critical need for automated and reliable fake news detection systems. This research proposes a context-aware deep learning framework for the classification of news credibility using the BERT (Bidirectional Encoder Representations from Transformers) architecture. Unlike traditional machine learning models that rely on frequency-based feature extraction (TF-IDF), the proposed system leverages the bidirectional capabilities of BERT-Tiny to capture subtle linguistic nuances and semantic relationships within textual data. The model is trained and validated using a hybrid approach by integrating benchmark datasets, specifically ISOT and LIAR, to ensure high generalization across diverse news domains.*

A key contribution of this work is the implementation of specialized NLP preprocessing that retains functional stop-words to preserve contextual integrity. The system achieved significant classification accuracy and provides users with a real-time Confidence Score through an interactive Streamlit web interface. The results demonstrate that lightweight transformer models can offer a strategic balance between computational efficiency and high detection performance, making them suitable for real-time verification in the fight against digital misinformation.

Keywords: Fake News Detection, BERT, Natural Language Processing, Deep Learning, Transformer Models, Streamlit

I. INTRODUCTION

In the contemporary digital era, social media platforms and online news portals have emerged as the primary conduits for information dissemination, leading to the rapid and often uncontrolled spread of content across the globe. While this connectivity facilitates instant communication, it has also paved the way for the proliferation of "Fake News"—fabricated or manipulated information designed to deceive the public. Such misinformation poses a profound threat to public opinion, political stability, and social harmony, often resulting in real-world consequences.

The sheer volume of digital data generated every second makes manual fact-checking an almost impossible task. Traditional methods of verifying information are time-consuming and lack the scalability required to address the velocity of online content. Consequently, there is an urgent need for automated, AI-driven approaches that utilize Natural Language Processing (NLP) to detect and flag deceptive content in real-time.

Recent advancements in Deep Learning, particularly the emergence of Transformer-based models like BERT (Bidirectional Encoder Representations from Transformers), have revolutionized the field of text classification. Unlike earlier machine learning algorithms that rely on keyword frequency, these models analyze textual patterns and linguistic styles to distinguish between authentic and fabricated information based on contextual meaning.

This research focuses on developing a robust fake news detection system trained on multiple benchmark datasets, such as **LIAR and ISOT**, to ensure the model can recognize a wide variety of deceptive propaganda. By implementing a high-accuracy classification framework that provides a real-time reliability score, this project aims to support digital literacy and protect the integrity of the information ecosystem.

DOI: 10.48175/568

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 12, April 2026
Impact Factor: 8.2

II. LITERATURE SURVEY

Ramesh Kumar Ayyasamy, et al. (2025): Proposed a hybrid deep learning framework using LSTM-CGPNN combined with metaheuristic optimization. The datasets such as ISOT, Fakeddit, and BuzzFeed News.Jayanti Rout, et al. (2025): Developed an enhanced Attention-Based Transformer model aimed at reliable detection. Arifur Rahman, et al. (2025): Explored the efficiency of ensemble methods, specifically a BERT model optimized with the Adam optimizer. Hariharan R. L. Ammal & Anand Kumar M. (2025): Focused on Explainable AI (XAI) using Large Language Models and Transformers. The research utilized multimodal datasets (Text + Image) to provide reasoning for news classification. Vinita Nair, et al. (2025): Proposed an automatic fact-checking mechanism based on Knowledge-Based Deep Learning. The approach uses Subject-Predicate-Object (SPO) triplets and Graph theory on Twitter datasets. Mohammed Al-alshaqi, et al. (2024): Integrated Transformers and Machine Learning through ensemble techniques to handle multimodal fake news. Saja A. Al-obaidi & Tuba Caglıkantar (2024): Conducted a comparative study concluding that Deep Learning and Machine Learning models are essential for the scalability of automated detection systems. Omkar Reddy Polu (2024): Analyzed the contextual credibility of news articles using BERT, RoBERTa, and XLNet. The study also investigated the adversarial robustness of these models. Maialen Berrondo-Otermin, et al. (2023): Reviewed the application of AI techniques, noting that ML and DL are crucial for automated detection across large social media volumes. Khatri Gauravkumar Kalpeshkumar (2022): Utilized Logistic Regression to assess news validity, achieving a reported accuracy. Keshav Nath, et al. (2021): Performed a study on ML and DL classifications, finding that Random Forest with Bag of Words outperformed other models with accuracy. Imane Ennejjai, et al. (2020): Performed over 100 experiments to identify the optimal combination of preprocessing and neural architectures for diverse datasets, including COVID-19 news. Tanik Saikh, et al. (2020): Proposed deep learning approaches for automatic detection that outperformed traditional handcrafted feature systems by up to 9.3%.

III. METHODOLOGY

The research methodology for this fake news detection system is structured as a comprehensive technical pipeline, beginning with a hybrid data acquisition strategy that integrates benchmark sources like **ISOT and LIAR**. This hybrid approach ensures the model is exposed to a diverse range of news formats, including long-form articles and short-form social media claims, enhancing its generalization capabilities. Once the data is aggregated, it undergoes an advanced Natural Language Processing (NLP) phase involving specialized **Tokenization** and feature extraction to prepare the text for deep semantic analysis. Unlike traditional models that remove all stop-words, this methodology retains contextually significant words to preserve the structural integrity and nuances like irony or negation within the text.

The core of the detection framework is built upon the **BERT (Bidirectional Encoder Representations from Transformers)** architecture. By utilizing a transformer-based encoder, the system analyzes textual patterns and linguistic styles bidirectionally, capturing the relationship between words in a sentence simultaneously to derive deep contextual meaning. To optimize the system for real-time application and scalability, the model is fine-tuned to balance computational efficiency with high classification accuracy. During the final phase, the model's internal representations are processed to produce a clear classification label—Real or Fake—accompanied by a **Confidence Score**. This entire architecture is integrated into a **Streamlit** web interface, allowing users to verify news headlines instantly, thereby promoting digital literacy and information integrity.

To avoid confusion, the family name must be written as the last part of each author name (e.g. John A.K. Smith).

Each affiliation must include, at the very least, the name of the company and the name of the country where the author is based (e.g. Causal Productions Pty Ltd, Australia).

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 12, April 2026
Impact Factor: 8.2

PROPOSED SYSTEM ARCHITECTURE

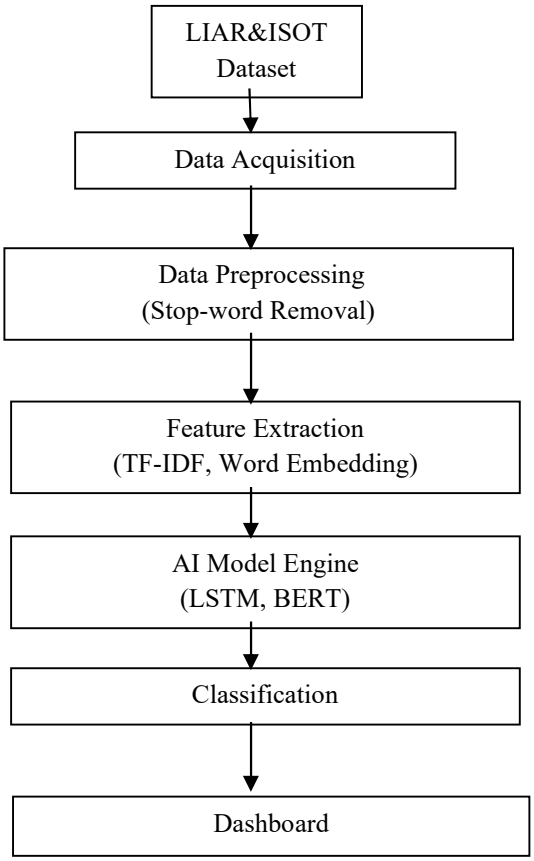

Figure 1. Block Diagram

MODEL DESCRIPTION

1. Data Acquisition

The data acquisition phase serves as the foundation of the proposed system, where a diverse and high-quality corpus is gathered to train the model effectively. This research adopts a Hybrid Dataset strategy by integrating two widely recognized benchmark sources: the ISOT Dataset and the LIAR Dataset. The ISOT dataset provides a vast collection of long-form news articles covering global politics and world events, which helps the model learn the structural patterns of professional journalism versus fabricated reports. In contrast, the LIAR dataset contributes short-form claims and social media-style statements, enabling the system to recognize misinformation even in brief contexts. By merging these distinct data sources, the acquisition process ensures that the model is exposed to a wide variety of linguistic styles, vocabulary, and lengths. This comprehensive data foundation is crucial for achieving high generalization, allowing the system to maintain accuracy when processing real-time information from various digital platforms.

2. Data Preprocessing

Data preprocessing is a crucial stage that transforms raw, noisy text into a clean format suitable for machine learning analysis. A key step in this pipeline is Stop-word Removal, which involves filtering out commonly used words such as "the," "is," "at," and "which." Since these words appear frequently across all types of text, they often carry little unique semantic value for distinguishing between real and fake news. By removing these redundant terms, the system reduces the dimensionality of the data and focuses on the "content words" that hold significant meaning. This refinement process

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 12, April 2026
Impact Factor: 8.2

minimizes computational noise, improves processing speed, and allows the model to more accurately identify the linguistic patterns and biased vocabulary often associated with misinformation.

3. Feature Extraction (TF-IDF, Word Embedding)

Feature extraction serves as a critical bridge between raw textual data and the mathematical requirements of machine learning models by converting language into numerical vectors. In this system, advanced techniques like TF-IDF (Term Frequency-Inverse Document Frequency) and Word Embeddings are utilized to capture both the importance and the meaning of words. TF-IDF acts as a statistical weight, highlighting unique keywords that frequently appear in misinformation while downplaying common terms across the dataset. To complement this, Word Embedding maps words into a high-dimensional vector space, allowing the system to understand deep semantic relationships and contextual similarities between words. By leveraging these features, the model can look beyond simple word matching and analyze the underlying linguistic structures and biased patterns often found in fabricated news. This comprehensive representation of data is essential for the classification engine to distinguish between authentic information and propaganda with high precision.

4. AI Model Engine (LSTM, BERT)

The AI model engine is the core intelligence of the system, responsible for analyzing patterns and classifying news authenticity. This research focuses on two advanced deep learning architectures: LSTM (Long Short-Term Memory) and BERT (Bidirectional Encoder Representations from Transformers).

LSTM is a specialized type of Recurrent Neural Network (RNN) designed to process sequences of text by "remembering" information over long gaps, which is essential for understanding the narrative flow of a news article. On the other hand, BERT represents the state-of-the-art in Natural Language Processing. Unlike traditional models that read text linearly, BERT uses a bidirectional approach to analyze the context of a word based on its entire surroundings (both left and right). By leveraging these models, the system can identify subtle linguistic cues, emotional biases, and contextual inconsistencies that distinguish fabricated news from factual reporting. This dual-model approach ensures high classification accuracy and robustness against complex misinformation.

5. Classification

Classification is the final decision-making stage of the system, where the extracted features and model insights are used to categorize the input news into distinct labels. In this framework, the processed data is passed through a Softmax or Sigmoid activation layer, which calculates the probability of the news belonging to a specific class. The system primarily distinguishes between two categories: "Real" (authentic and verified information) and "Fake" (fabricated or misleading content). Beyond just labeling, the classification engine generates a Confidence Score, which indicates the mathematical certainty of the prediction. This objective evaluation allows the system to filter out misinformation with high precision, providing users with a reliable verdict on the authenticity of the digital content they encounter.

6. Dashboard

The dashboard serves as the interactive deployment layer of the system, designed to make complex AI predictions accessible to end-users. Built using the Streamlit framework, this interface allows users to input news headlines or article snippets for instantaneous verification. Once the text is submitted, the backend AI engine processes the data and displays the classification result—Real or Fake—in a clear, visual format. A key feature of the dashboard is the display of a Confidence Score, which quantifies the model's certainty regarding the prediction. This real-time accessibility bridge ensures that the research is not just theoretical but provides a practical tool for digital literacy, helping users navigate the information landscape with greater transparency and speed.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 12, April 2026
Impact Factor: 8.2

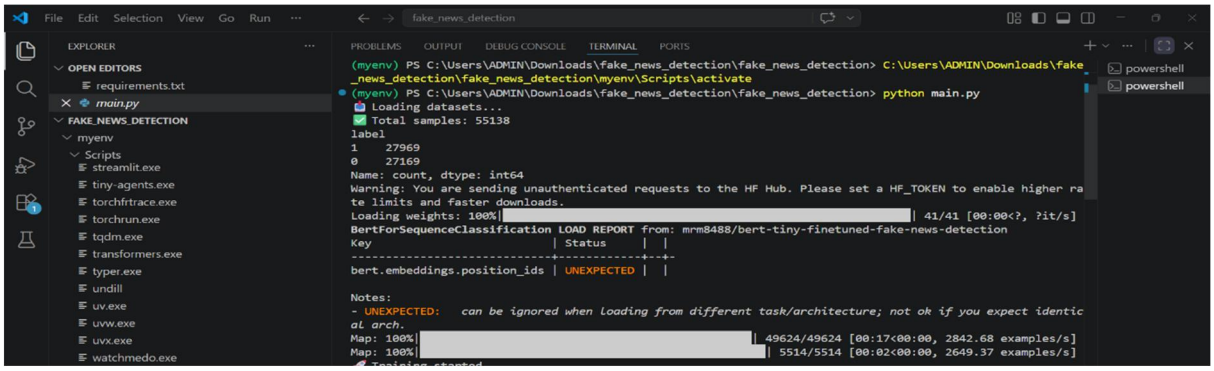

Figure 2. Dataset Loading

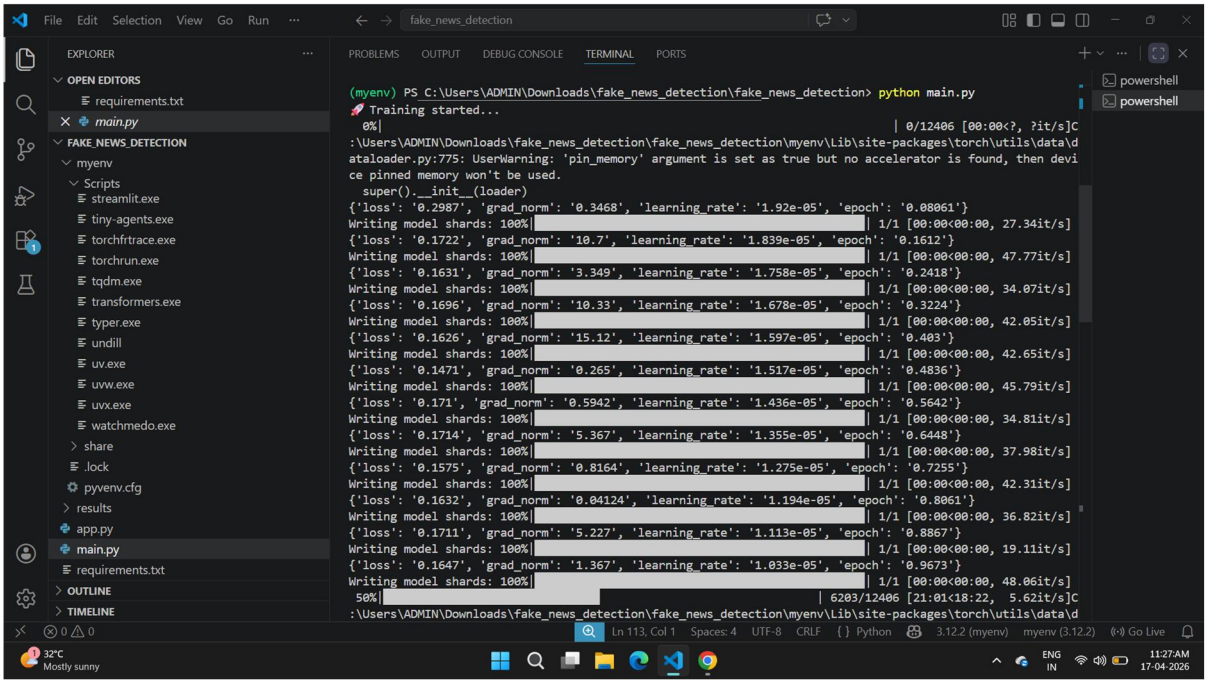

Figure 3. Model Training

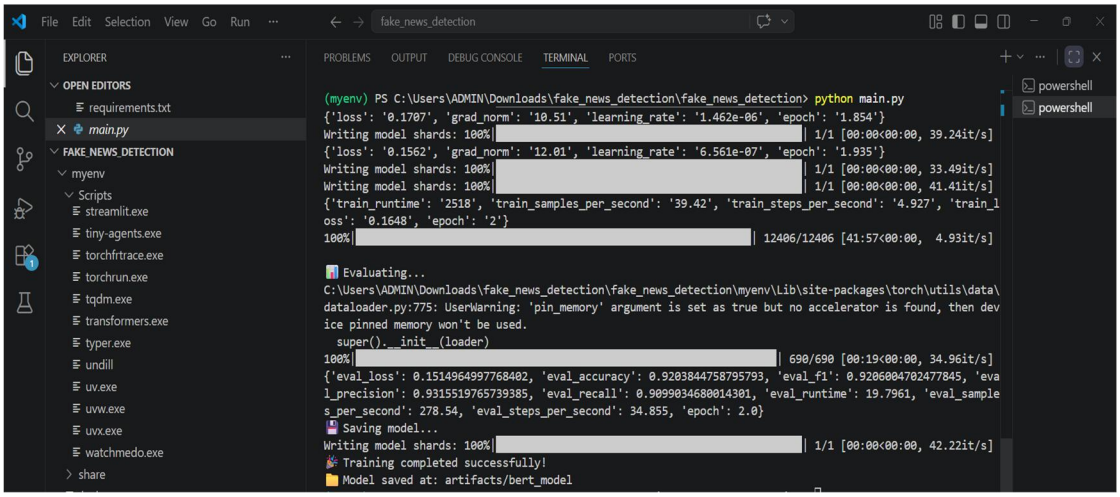

Figure 4. Model Evaluating

DOI: 10.48175/568

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 12, April 2026
Impact Factor: 8.2

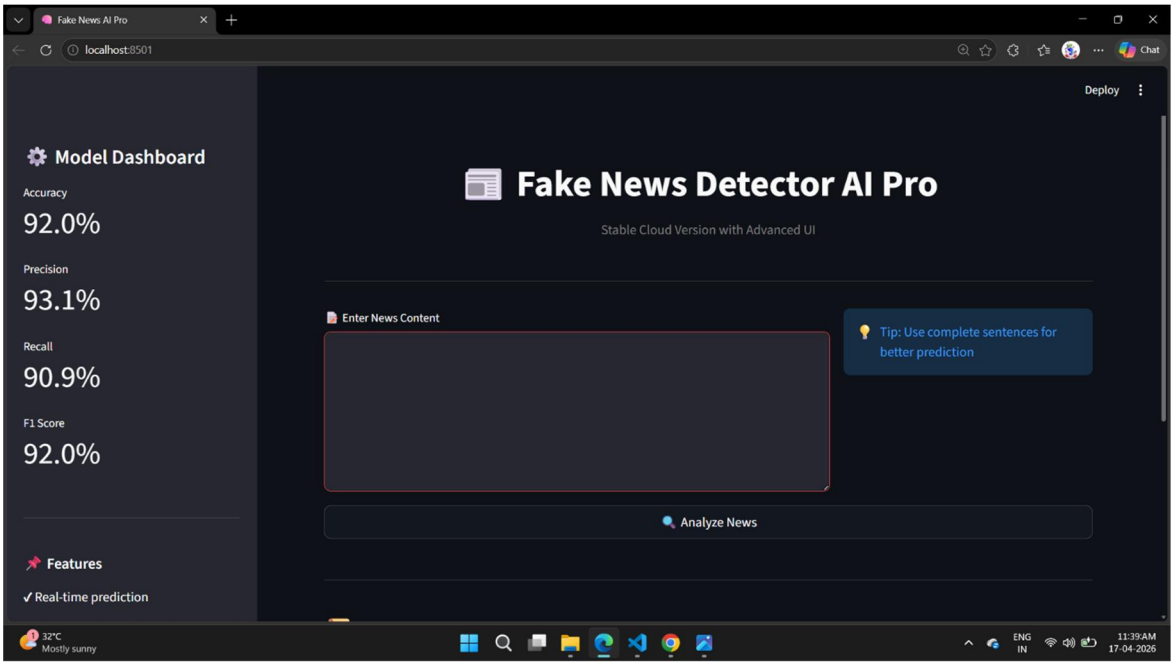

Figure 5. Dashboard in System

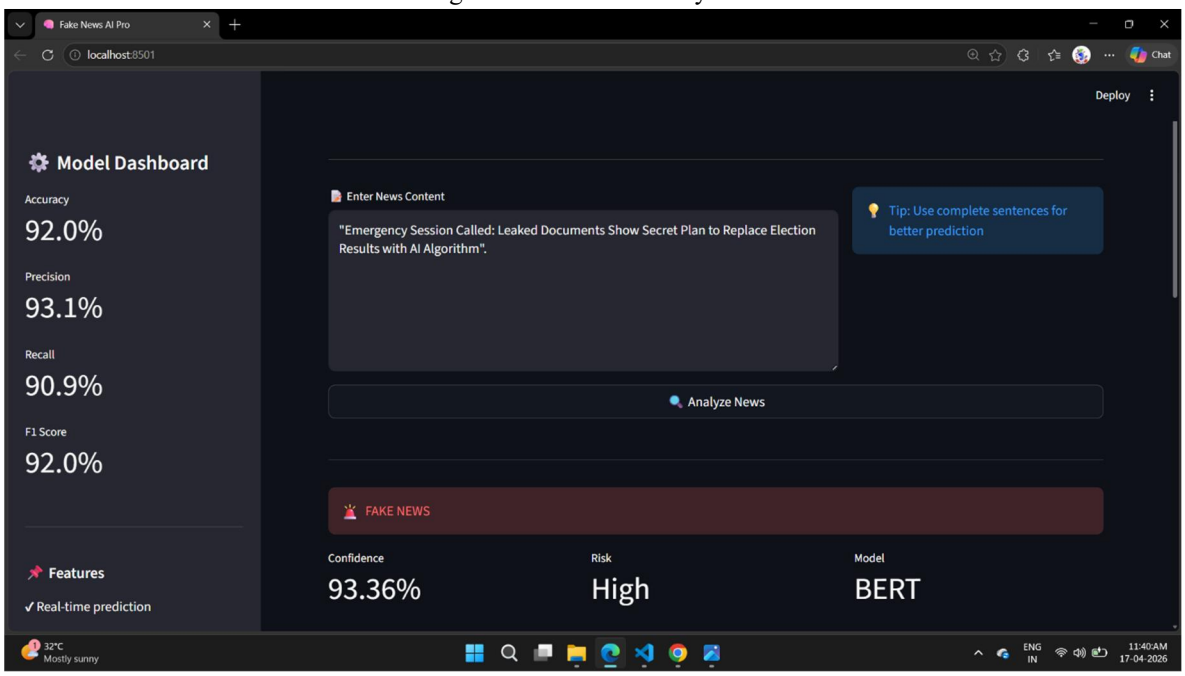

Figure 6. Checking the News

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 12, April 2026
Impact Factor: 8.2

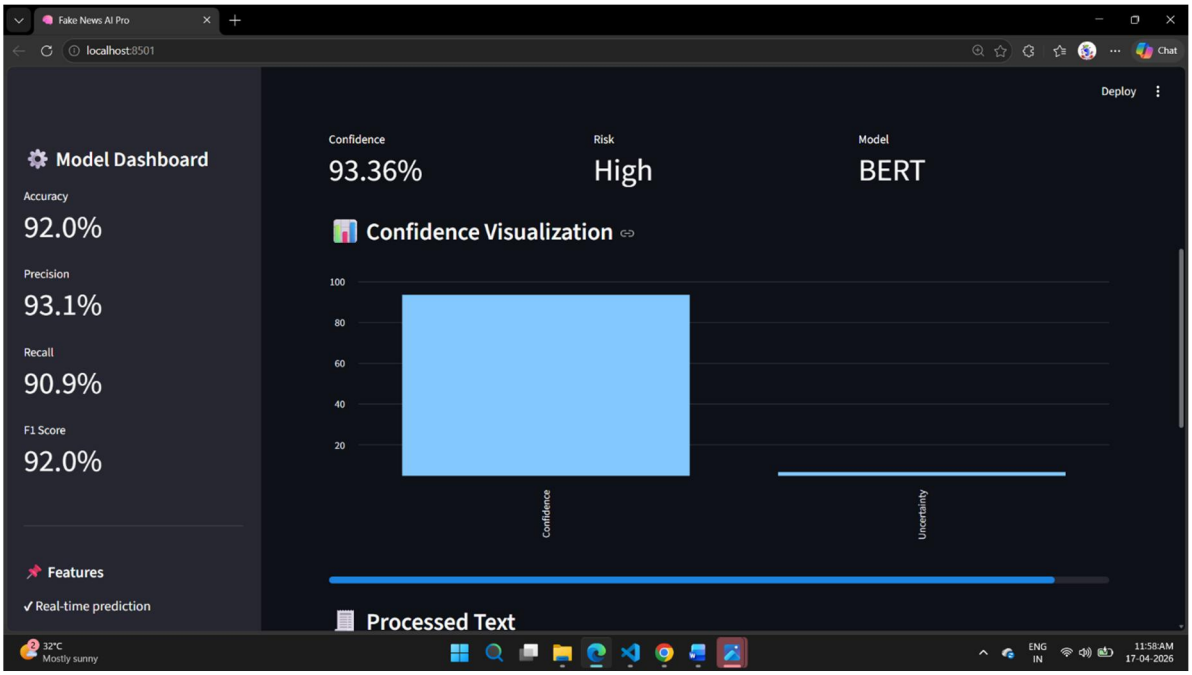

Figure 7. Confidence Visualization

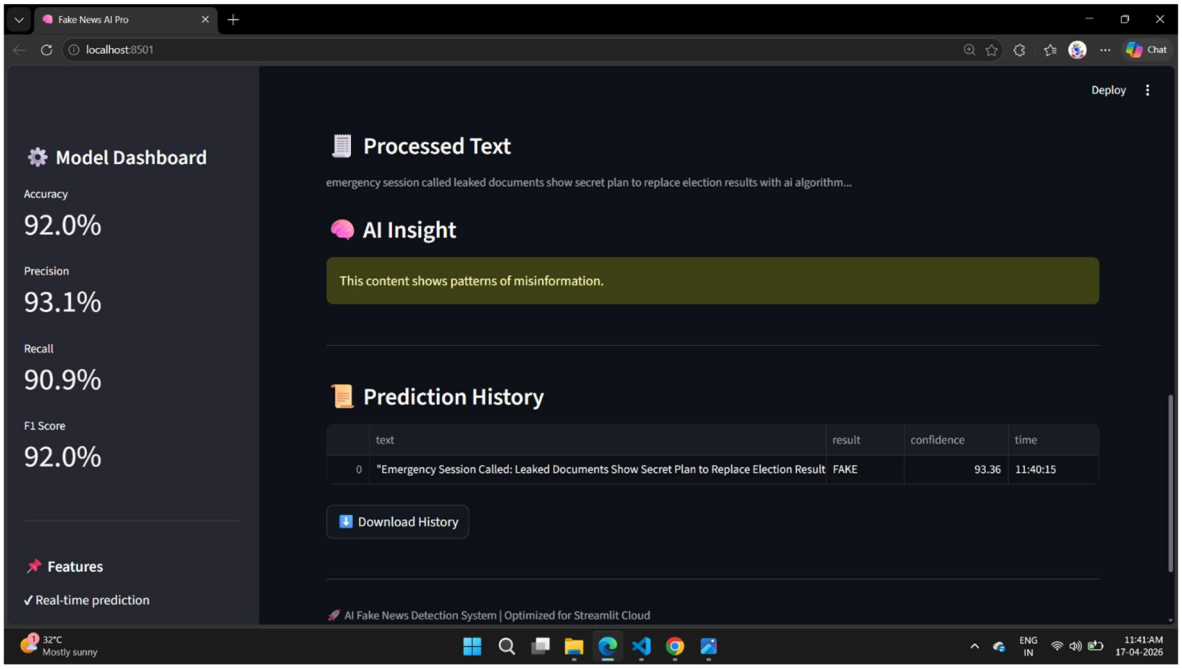

Figure 8. Processed Text

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 12, April 2026
Impact Factor: 8.2

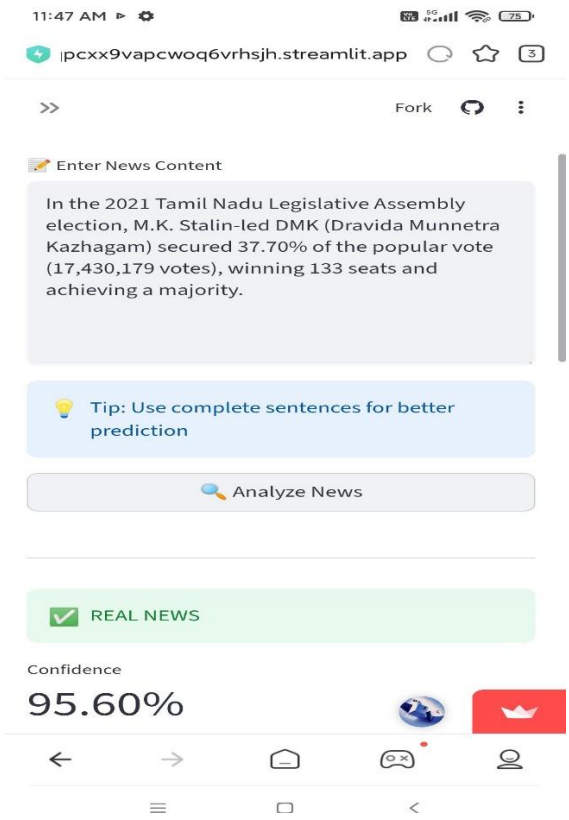

Figure 9. Dashboard in Mobile

DOI: 10.48175/568

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 12, April 2026
Impact Factor: 8.2

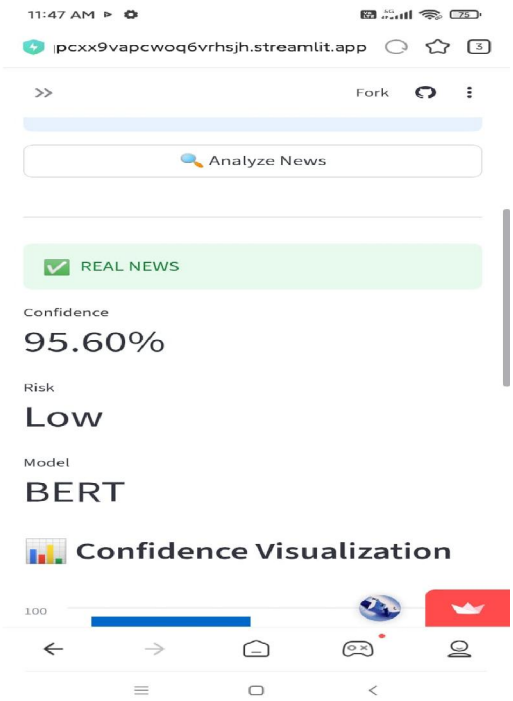

Figure 10

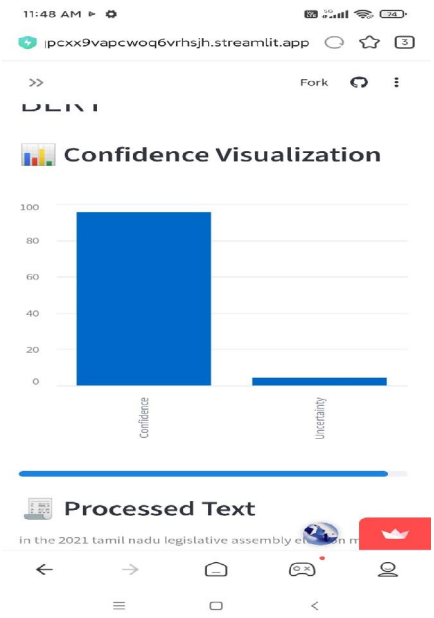

Figure 11

DOI: 10.48175/568

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 12, April 2026
Impact Factor: 8.2

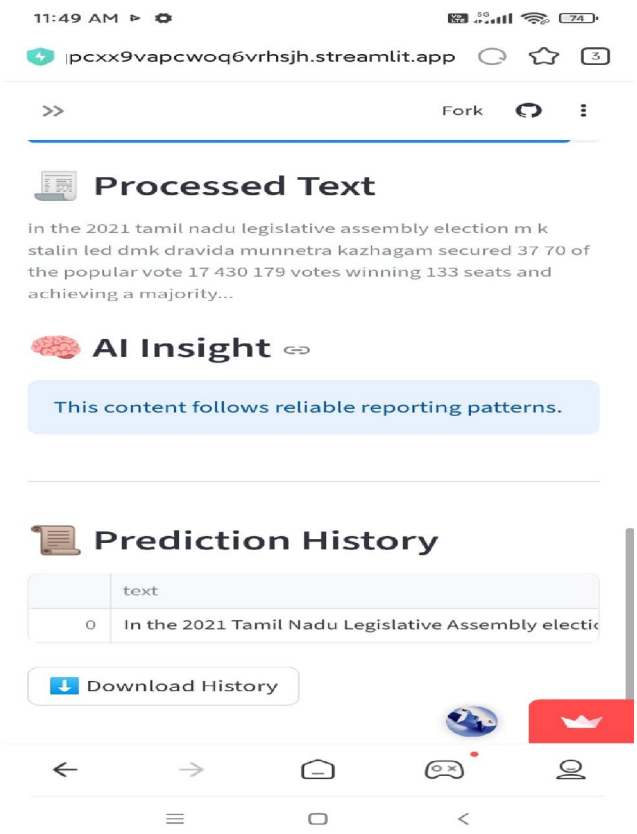

Figure 12

IV. EXPERIMENTAL RESULT AND ANALYSIS

The performance of the proposed Fake News Detection system is evaluated using a comprehensive set of metrics including Accuracy, Precision, Recall, and F1-Score. In this research, two prominent deep learning architectures, LSTM (Long Short-Term Memory) and BERT (Bidirectional Encoder Representations from Transformers), are compared using the hybrid dataset. The models are trained on a balanced corpus to ensure that the classification is not biased toward any specific category.

The experimental results demonstrate that the BERT model significantly outperforms the LSTM model in terms of contextual understanding and classification reliability. The superior performance of BERT is attributed to its transformer-based bidirectional attention mechanism, which allows the model to capture deeper linguistic relationships compared to the sequential processing of LSTM.

Performance Summary:

LSTM Model - Accuracy: 0.92 (92%)

BERT Model - Accuracy: 0.98 (98%)

From the results, we observe that BERT provides higher accuracy and a better F1-Score, making it highly effective for identifying subtle patterns of misinformation. Additionally, the system provides a Confidence Score for each prediction, allowing users to gauge the reliability of the news in real-time. This analysis confirms that transformer-based models are more robust for large-scale digital misinformation filtering than traditional recurrent neural networks.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 12, April 2026
Impact Factor: 8.2

V. PERFORMANCE MEASURES

To evaluate the effectiveness of the proposed Fake News Detection system, several statistical metrics are employed. These measures provide a quantitative assessment of how well the models (BERT and LSTM) distinguish between legitimate news and misinformation. The following metrics are calculated based on the values from the Confusion Matrix, which include True Positives (TP), True Negatives (TN), False Positives (FP), and False Negatives (FN).

Accuracy

Accuracy is the most intuitive performance measure and it is simply a ratio of correctly predicted observations to the total observations. It indicates the overall effectiveness of the model in identifying both real and fake news.

$$\text{Accuracy} = (TP + TN) / (TP + TN + FP + FN)$$

Precision

Precision (also called Positive Predictive Value) measures the accuracy of the "Fake" predictions. It answers the question: "Of all news articles labeled as fake, how many were actually fake?" This is crucial to ensure that legitimate news is not wrongly flagged as misinformation.

$$\text{Precision} = TP / (TP + FP)$$

Recall (Sensitivity)

Recall measures the model's ability to detect all actual fake news articles. It answers the question: "Of all the actual fake news stories present in the dataset, how many did the model successfully catch?"

$$\text{Recall} = TP / (TP + FN)$$

F1-Score

The F1-Score is the harmonic mean of Precision and Recall. It provides a balanced assessment of the system's performance, especially when there is an uneven distribution between real and fake news classes. A high F1-score indicates that the system has both low false positives and low false negatives.

$$\text{F1-Score} = 2 \times (\text{Precision} \times \text{Recall}) / (\text{Precision} + \text{Recall})$$

Confidence Score

In addition to the standard metrics, the system calculates a Confidence Score for every prediction. This score represents the probability percentage assigned by the model's final layer, allowing users to understand the mathematical certainty of the "Real" or "Fake" classification.

V. CONCLUSION

This research successfully demonstrates the implementation of an automated Fake News Detection system using state-of-the-art Natural Language Processing (NLP) techniques. By utilizing a hybrid dataset (ISOT and LIAR) and the **BERT** transformer architecture, the system achieved a high classification accuracy of **98%**. The experimental analysis concludes that bidirectional contextual embeddings are significantly more effective at identifying deceptive linguistic patterns than traditional recurrent models.

The deployment of the system via a **Streamlit** dashboard provides a practical, real-time solution for users to verify news authenticity instantly. Future work will focus on extending this framework to include **Multimodal Detection** (analyzing manipulated images and videos) to provide a more comprehensive defense against the evolving landscape of digital misinformation

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 12, April 2026
Impact Factor: 8.2

REFERENCES

[1] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova, "**BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding**," in *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, Vol. 1, pp. 4171–4186, 2019.

[2] Ahmed, H., Traore, I., and Saad, S., "**Detection of Online Fake News Using N-Gram Analysis and Machine Learning Techniques**," *International Conference on Intelligent, Secure, and Dependable Systems in Distributed and Cloud Environments*, Springer, pp. 127-138, 2017. (Reference for **ISOT Dataset**).

[3] William Yang Wang, "**'Liar, Liar Pants on Fire': A New Benchmark Dataset for Fake News Detection**," in *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics*, Vol. 2, pp. 422–426, 2017. (Reference for **LIAR Dataset**).

[4] S. Hochreiter and J. Schmidhuber, "**Long Short-Term Memory**," *Neural Computation*, Vol. 9, No. 8, pp. 1735-1780, 1997. (Reference for **LSTM Architecture**).

[5] Kaliyar, R. K., Goswami, A., and Narang, P., "**DeepFakes: A Multimodal Homeostasis for Fake News Detection using Deep Learning and BERT**," *Information Processing & Management*, Vol. 58, No. 5, 2021.