

MOTIVATION

Existing System

- Many earlier models rely on surface-level features such as word frequencies, sentiment scores, or predefined rules to detect fake news. While simple to implement, these models often fall short when the statements are contextually complex or nuanced.
- Crowdsourced fact-checking platforms like Snopes and Politifact exist but are timeconsuming and may not scale effectively. They also depend heavily on human subjectivity.
- Some academic studies have used the LIAR dataset but with limited exploration of advanced models like BERT.

Proposed System

- We propose a hybrid approach combining traditional ML techniques with state of the art deep learning models. Specifically, we utilize Logistic Regression and Random Forest for baseline comparisons, and BERT for deeper language comprehension.
- Text data will be preprocessed using NLP techniques and transformed into numerical features using TF-IDF and BERT embeddings.
- The performance of each model will be assessed using standard evaluation metrics and visualized for comparative analysis.
- This comprehensive evaluation will help identify the strengths and limitations of each model, aiding in the development of an efficient and scalable fake news detection system.

Key points

- **Hybrid Approach:**
Combines traditional machine learning models (Logistic Regression, Random Forest) with advanced deep learning models (BERT).
- **Text Preprocessing:**
Uses Natural Language Processing (NLP) techniques to clean and process input data.
- **Feature Extraction:**
Converts text into numerical features using TF-IDF and BERT embeddings for model training.
- **Model Evaluation:**

Each model is evaluated using standard metrics like accuracy, precision, recall, and F1-score.

- **Comparative Analysis:**

Performance results are visualized and compared to identify the most effective model.

- **Scalability Focus:**

The system is designed to be scalable and efficient, suitable for real-time deployment in detecting fake news.