Global Journal of Engineering Innovations & Interdisciplinary Research

Original Article

Fake News Detection Using Distilbert

M. RamaRaju[1], K. Sanjay[2], Ch.Mrudula[2], G. Akhila[2], G. Adithya[2]

[1]Assistant Professor, Department of CSE, Christu Jyothi Institute of Technology & Science, Jangaon, Telangana, India
[2]UG Students, Department of CSE, Christu Jyothi Institute of Technology & Science, Jangaon, Telangana, India

Correspondence

M Rama Raju

Assistant Professor, Department of CSE, Christu Jyothi Institute of Technology & Science, Jangaon, Telangana, India

- Received Date: 25 Jan 2026
- Accepted Date: 14 Mar 2026
- Publication Date: 20 Mar 2026

Keywords

Fake News Detection, DistilBERT, Natural Language Processing, Transformer Models, Deep Learning, Streamlit.

Abstract

The rapid spread of fake news and misinformation across digital platforms has emerged as a serious challenge to public trust and informed decision-making. This project focuses on addressing the problem of automated fake news detection by designing an efficient text classification system. The proposed approach examines a given news article and determines whether it is classified as fake or true based solely on its textual information. The foundation of the system is a pre-trained transformer-based language model, DistilBERT, which is capable of producing rich contextual and semantic representations of text. Before processing the input through the model, the text undergoes essential Natural Language Processing (NLP) preprocessing to ensure standardized and noise-free input. DistilBERT, a compact and computationally optimized variant of BERT, is fine-tuned for binary classification, achieving a balance between high predictive accuracy and fast inference. The model effectively identifies subtle linguistic patterns associated with misinformation. The final application offers a scalable and reliable solution for fake news detection, contributing to efforts aimed at reducing the impact of digital misinformation.

Introduction

The rapid growth of the internet and social media platforms has made information easily accessible to people worldwide. News articles and posts can be shared instantly through websites, blogs, and social networking platforms. While this has improved communication and information sharing, it has also led to the rapid spread of fake news and misinformation. Detecting fake news manually is difficult because of the large volume of information generated every day on the internet, requiring significant time and human effort from journalists and fact-checking organizations. Therefore, automated fake news detection systems using machine learning and Natural Language Processing (NLP) techniques have become an important area of research. This project focuses on developing a Fake News Detection system using DistilBERT, a transformer-based deep learning model, integrated with a Streamlit-based interface for classifying textual content as Fake News or Real News.

Background

The widespread use of the internet and digital media has transformed how people consume news and information, allowing it to spread rapidly across the globe. However, this rapid distribution also increases the chances of misinformation and fake news spreading quickly before it can be verified. Earlier methods of fake news detection relied mainly on manual fact-checking, which is reliable but time-consuming and unable to keep up with the large volume of content. With the advancement of deep learning, transformer-based language models like BERT and DistilBERT are capable of understanding contextual relationships between words in a sentence. DistilBERT provides strong language understanding while requiring fewer computational resources, making it suitable for efficient fake news detection systems.

Problem Statement

The rapid growth of digital media platforms and social networking sites has significantly increased the volume of information shared online, becoming a major source for spreading fake news. Fake news refers to fabricated or misleading information intended to manipulate public opinion or create confusion. Traditional approaches to identifying fake news rely heavily on manual verification, which is time-consuming and cannot effectively handle the massive amount of content generated daily. As a result, there is a strong need for automated systems using NLP and deep learning techniques that can quickly analyze news content, leverage contextual language understanding, and identify whether the information is genuine or misleading.

Citation: Rama Raju M, Sanjay K, Mrudula Ch, Akhila G, Adithya G. Fake News Detection Using Distilbert. GJEIIR. 2026;6(3):0183.

Literature Survey

Existing Fake News Detection (FND) systems are primarily based on traditional Natural Language Processing (NLP) and classical machine learning approaches. These systems typically rely on manually engineered lexical features extracted from textual data using techniques such as Bag-of-Words (BoW) and Term Frequency–Inverse Document Frequency (TF-IDF). After feature extraction, classification models such as Naïve Bayes, Logistic Regression, and Support Vector Machines (SVM) are used to categorize news articles.

However, because traditional models process words independently, they often fail to capture deeper semantic relationships and contextual meaning within the text. As fake news frequently involves complex linguistic structures or contextual manipulation, traditional machine learning approaches struggle to achieve high accuracy. Furthermore, models trained on fixed lexical features suffer from poor generalization, failing to detect new or evolving misinformation patterns. They also heavily depend on manual feature engineering, requiring extensive domain knowledge.

Proposed System

The proposed system introduces a transformer-based deep learning approach for Fake News Detection using a fine-tuned DistilBERT model. Unlike traditional machine learning methods, the proposed system directly learns contextual and semantic representations from raw textual data without relying on manual feature engineering. User-provided news articles are processed through the DistilBERT tokenizer, which performs subword tokenization and generates input IDs and attention masks required by transformer architectures. The tokenized input is then passed to a fine-tuned DistilBERT classification model trained on labeled fake and real news datasets. DistilBERT retains strong language understanding capabilities while significantly reducing model size and inference time.

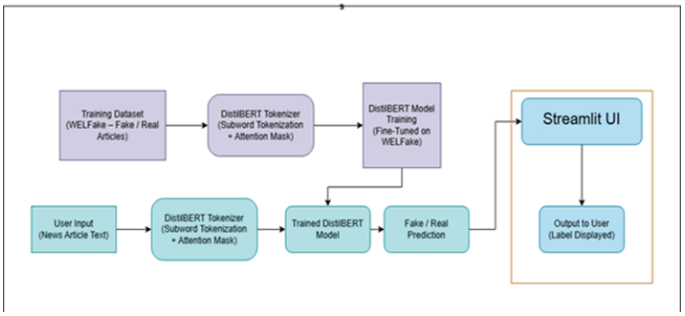
Modules used

- **User Interface Module:** This module provides a web-based interface developed using Streamlit. It allows users to input news articles and view classification results in a simple and interactive environment, offering both Single Article and Batch Analysis modes.
- **Text Processing Module:** This module processes the input text using the DistilBERT tokenizer. The tokenizer performs subword tokenization and generates input IDs and attention masks required by the transformer model.
- **Classification Module:** The classification module loads the fine-tuned DistilBERT model and performs inference on the tokenized text. The model analyzes contextual and semantic patterns within the text to determine whether the news article is fake or real.

Advantages of Proposed System

- **Context-Aware Language Understanding:** Transformer architecture captures contextual and semantic relationships within text.
- **Higher Prediction Accuracy:** Identifies complex linguistic patterns associated with misinformation.
- **Improved Generalization:** Performs better on unseen news articles and evolving misinformation patterns.
- **Reduced Feature Engineering:** Automatically learns relevant features directly from textual data without manual extraction.
- **Computational Efficiency:** DistilBERT offers faster inference while maintaining strong performance compared to larger transformer models.
- **Real-Time Prediction Capability:** Streamlit interface enables instant classification of user-provided news articles.

System architecture



The system architecture of the Fake News Detection system is designed to process textual news articles and classify them as Fake or Real using a two-stage approach: the model training stage and the prediction (inference) stage. In the training stage, a labeled dataset containing fake and real news articles (WELFake dataset) is used. The dataset is processed using the DistilBERT tokenizer, and the tokenized text is provided as input to the DistilBERT model, which learns contextual and semantic patterns. In the prediction stage, the user enters a news article through the Streamlit interface. The input text is passed to the tokenizer, converted into tokenized representations, and processed by the trained DistilBERT model to generate a prediction. Finally, the output label is displayed to the user via the interface, ensuring an efficient pipeline for real-time fake news classification.

Results

The output of the model classifies the submitted text as "Real News" or "Fake News". The testing process evaluated the system under various scenarios, including valid articles, fake news content, mixed content, short articles, and multiple article batch submissions. The DistilBERT model accurately classified both real and fake news articles, while the Streamlit interface effectively handled user inputs, batch processing, and prediction visualization via Plotly charts. No major functional issues were observed during testing, indicating that the system provides stable performance for real-time fake news classification.

Conclusion

This project successfully designed and implemented an automated Fake News Detection system using a fine-tuned DistilBERT transformer model. By leveraging DistilBERT's ability to capture deep contextual and semantic information, the system outperforms traditional manual feature engineering approaches, resulting in improved generalization and robustness. The lightweight nature of DistilBERT enables faster inference while maintaining strong classification performance. The integration of a Streamlit-based web interface allows users to perform both single-article and batch analysis intuitively, providing an effective and scalable solution for combating digital misinformation..

Future scope

RESTful API Deployment: Deploying the trained DistilBERT model as an API to allow external applications, news websites,

and platforms to integrate real-time article verification.

Explainable Artificial Intelligence (XAI): Integrating XAI techniques to highlight the important words or phrases influencing the model's prediction, providing better transparency.

Multilingual Support: Fine-tuning multilingual transformer models to analyze news content written in different languages for global information environments.

Model Optimization: Applying quantization or pruning techniques to reduce model size and improve inference speed on devices with limited computational resources.

References

1. J. Devlin, M. W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," in Proc. NAACL-HLT, 2019, pp. 4171–4186.
2. V. Sanh, L. Debut, J. Chaumond, and T. Wolf, "DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter," arXiv preprint arXiv:1910.01108, 2019.
3. H. Ahmed, I. Traore, and S. Saad, "Detection of online fake news using machine learning," in Proc. IEEE Int. Conf. Smart Cloud, 2017, pp. 145–150.
4. A. Vaswani et al., "Attention is all you need," in Proc. Adv. Neural Inf. Process. Syst. (NeurIPS), 2017, pp. 5998–6008.