Project Report: Fake News & Sentiment Detection using NLP & Transformers

Abstract

The unprecedented speed at which misinformation and fabricated news propagate across digital platforms such as Twitter and Reddit has emerged as one of the most pressing challenges in today's information-driven world. These deceptive narratives have the potential to distort public perception, shape political outcomes, and amplify societal divisions. Addressing this issue requires more than manual detection or simplistic keyword-based strategies, as these approaches are neither scalable nor capable of handling the complexity of modern online discourse.

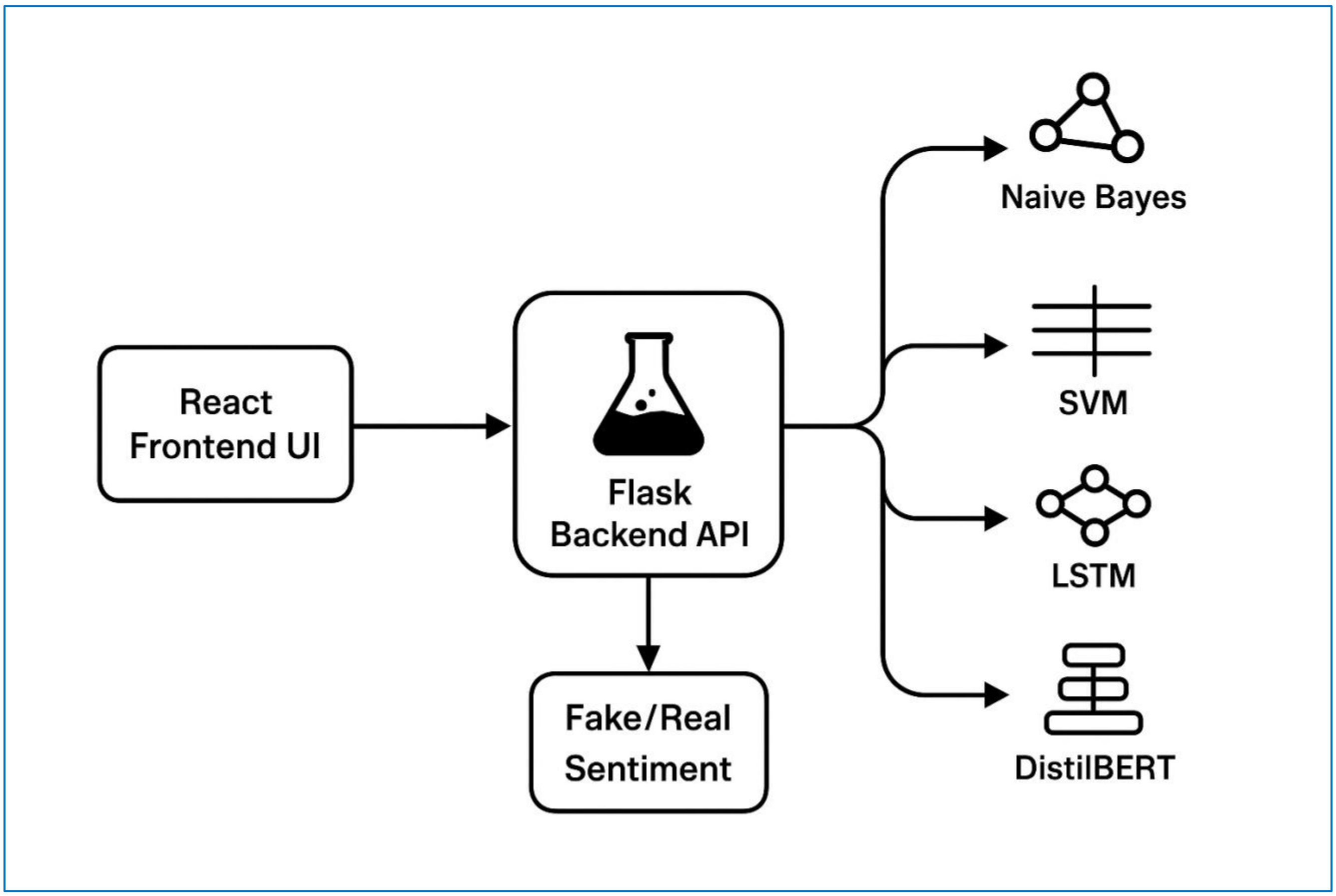
This project presents a comprehensive Natural Language Processing (NLP) pipeline designed to both **detect fake news** and **analyze sentiment polarity** (positive, negative, or neutral) within news headlines and social media text. The pipeline progresses through several layers of sophistication: beginning with classical machine learning models such as Naïve Bayes, Logistic Regression, and Support Vector Machines; advancing to sequential deep learning approaches including Recurrent Neural Networks (RNNs), Long Short-Term Memory networks (LSTMs), and Gated Recurrent Units (GRUs); and culminating with modern transformer-based architectures like BERT and DistilBERT, which represent the current state of the art in NLP.

The final system is deployed through an end-to-end solution that integrates a Flask-based backend with a React frontend branded as "NEWSCHECK.". This deployment allows users to input a piece of text

(headline, article excerpt, or tweet) and receive two outputs simultaneously: a classification of whether the text is **Fake or Real**, and its corresponding **Sentiment category**.

By combining fake news detection with sentiment analysis, the project not only identifies misinformation but also reveals the emotional undertone behind it, offering deeper insight into the influence of deceptive content on public opinion.

flowchart illustrating the end-to-end system



1. Introduction

In the digital age, social media platforms have become a dominant source of news and information. While this accessibility offers convenience, it has also created serious challenges. The rapid spread of misinformation and fake news can shape public opinion, influence political outcomes, and cause widespread confusion. Unlike traditional journalism, which follows editorial checks, online platforms allow content to be shared instantly and without verification, enabling false narratives to gain traction quickly.

Traditional detection techniques such as manual fact-checking or keyword-based filtering cannot scale to the size and speed of today's online information flow. As a result, there is a growing need for automated systems that can analyze language more deeply, detect patterns of deception, and classify news with high accuracy.

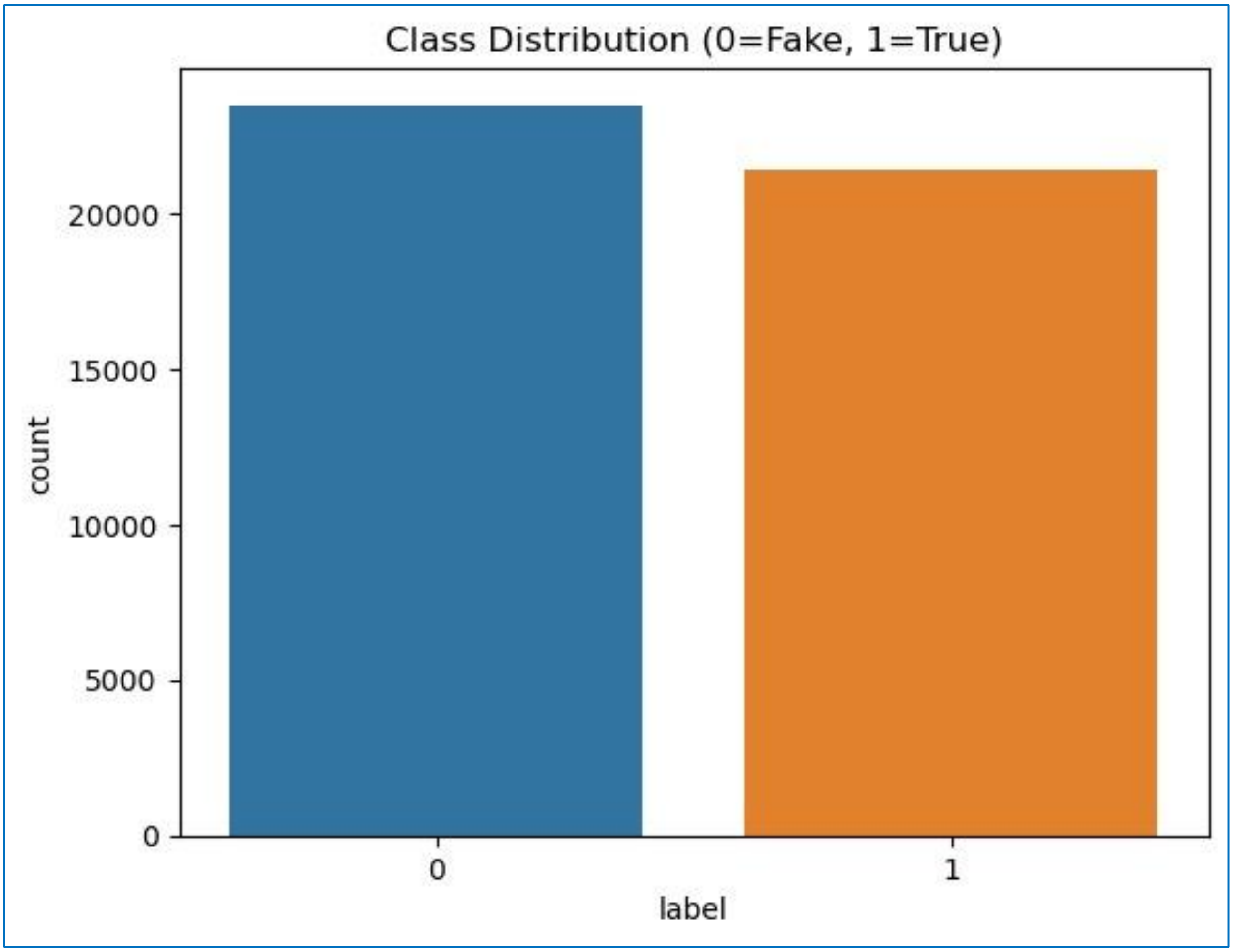
This project addresses the problem by developing a **dual-purpose NLP pipeline** that combines **Fake News Detection** with **Sentiment Analysis**. The system integrates several stages of experimentation, from preprocessing and classical machine learning models to advanced deep learning and transformer-based approaches, ensuring both breadth and depth in tackling the problem.

The project objectives are:

- **Preprocess and clean textual data** through normalization, tokenization, and lemmatization.
- **Train and evaluate classical machine learning models** such as Naïve Bayes, Logistic Regression, and SVM.
- **Develop deep sequence models** including RNN, GRU, and LSTM to capture contextual dependencies.
- **Fine-tune transformer-based architectures** such as BERT and DistilBERT for optimal classification performance.

- **Deploy the final solution** as a web-based application with a Flask backend and React frontend for real-time predictions.

By integrating both authenticity detection and sentiment analysis, the project not only determines whether a news item is real or fake but also reveals its emotional tone, offering valuable insights into the broader impact of misinformation on digital communities.



2. Dataset Description

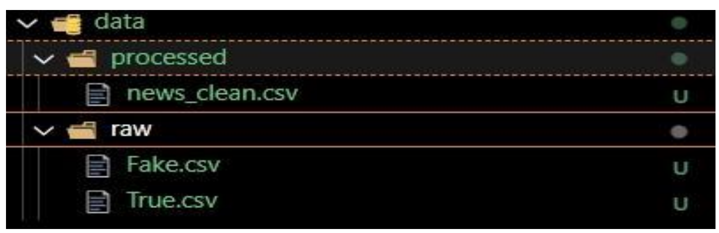
2.1 Dataset Source

The project uses the **Kaggle Fake News Dataset**, containing labeled real and fake articles. Additional sentiment annotations are applied using pre-trained sentiment models and lexicons.

- Format: CSV
- Fields: Headline, Article Body, Label (Real/Fake)
- Size: **30,000 articles**

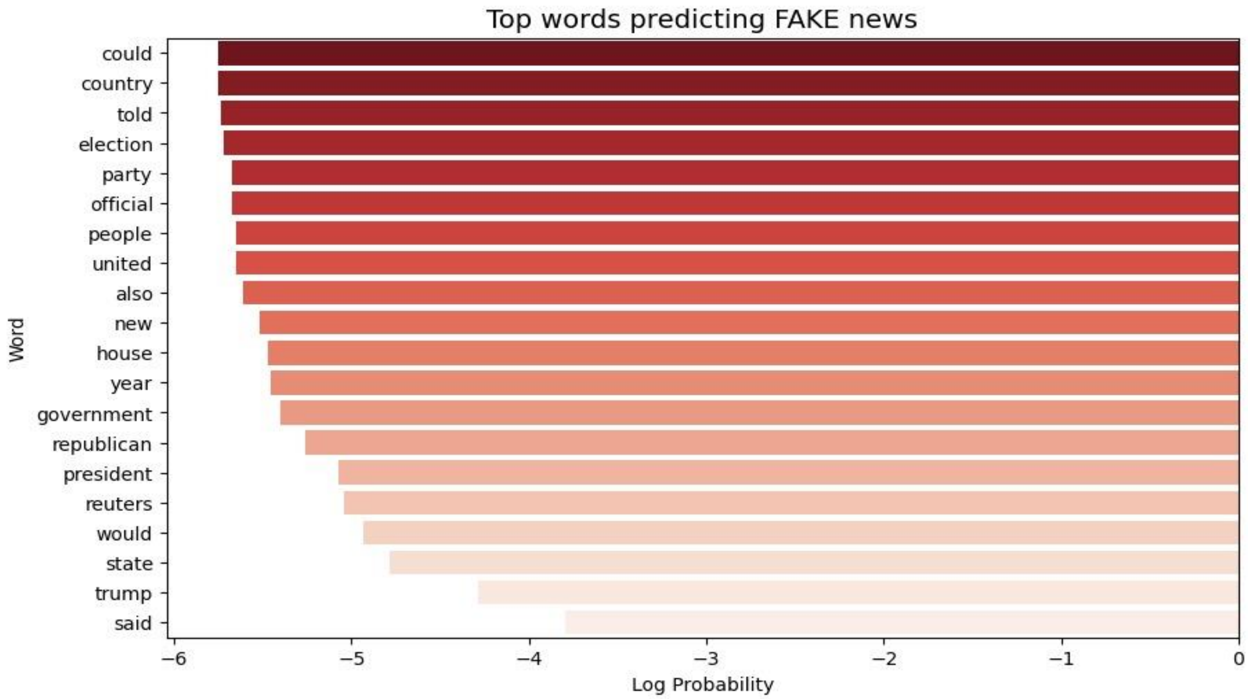
2.2 Dataset Statistics

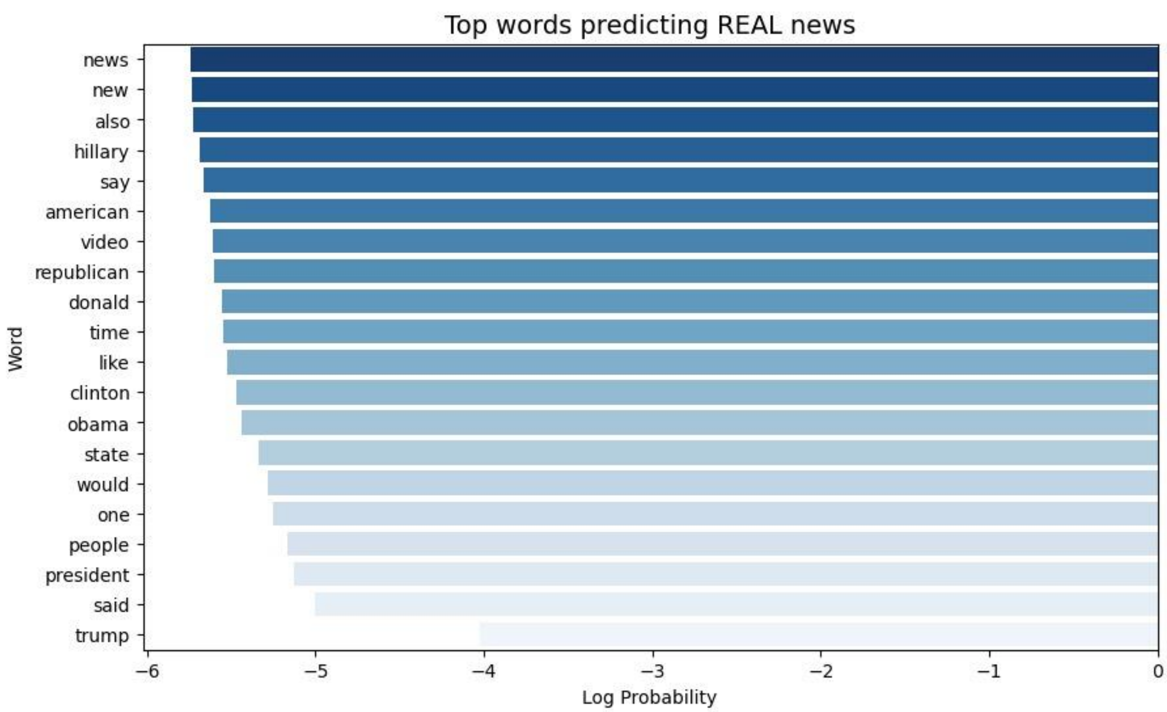
- Total articles: **30,000 articles**
- Fake articles: **14,998**
- Real articles: **15,06**



2.3 Exploratory Data Analysis (EDA)

Analysis revealed key insights into the language of misinformation. As shown in the log probability charts, words like "trump," "said," and "state" are common in both real and fake news, highlighting the political nature of the dataset. However, fake news tends to use more words like "republican," "government," and "election," while real news more frequently uses words like "new," "news," "video," and "american."


Top words predicting FAKE news
could
country
told
election
party
official
people
united
also
new
house
year
government
republican
president
reuters
would
state
trump
said
Word
−6
−5
−4
−3
−2
−1
0
Log Probability


Top words predicting REAL news
news
new
also
hillary
say
american
video
republican
donald
time
like
clinton
obama
state
would
one
people
president
said
trump
Word
−6
−5
−4
−3
−2
−1
0
Log Probability

3. Methodology

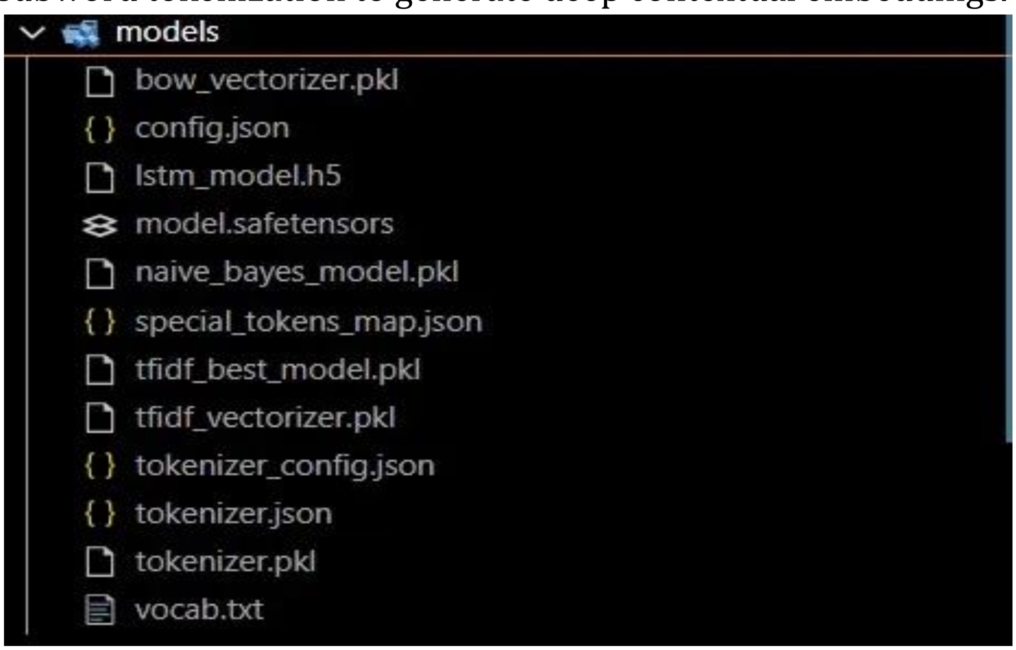
3.1 Text Preprocessing

- Lowercasing
- Removing punctuation and stopwords
- Tokenization
- Stemming and Lemmatization

This standardized the input for all models.

3.2 Feature Representations

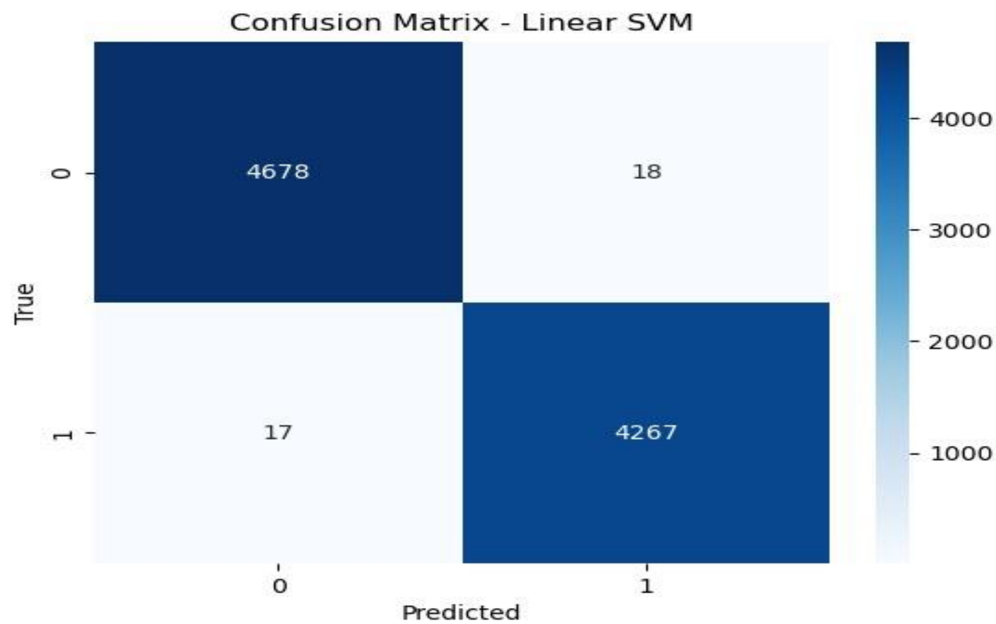
- Bag of Words (BoW): Used for the Naïve Bayes baseline model.
- TF-IDF: Used for the high-performing Linear SVM model.
- Word Embeddings: GloVe embeddings were used for the LSTM model to capture semantic meaning.
- Contextual Embeddings: The DistilBERT transformer uses its own subword tokenization to generate deep contextual embeddings.

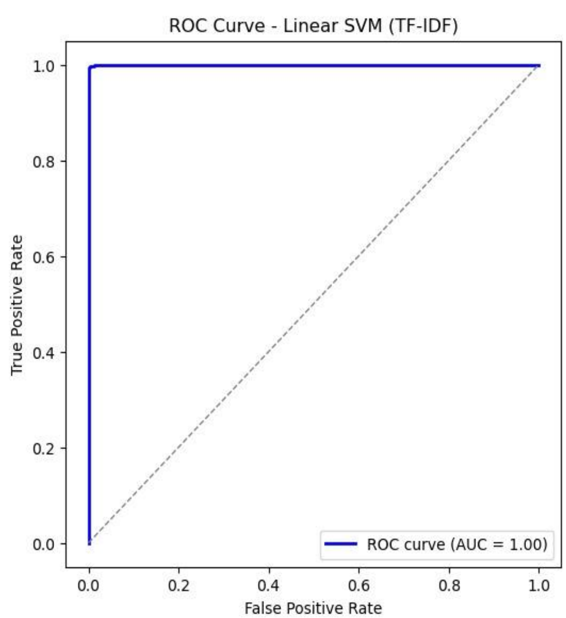


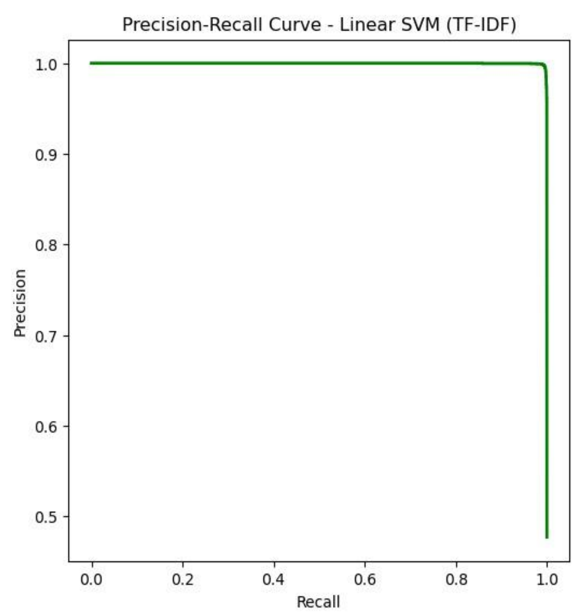
3.3 Classical Machine Learning Models

A Linear SVM model using TF-IDF features proved to be an exceptionally strong baseline, nearly perfect on the test set.

- Linear SVM (TF-IDF) Accuracy: 99.61%



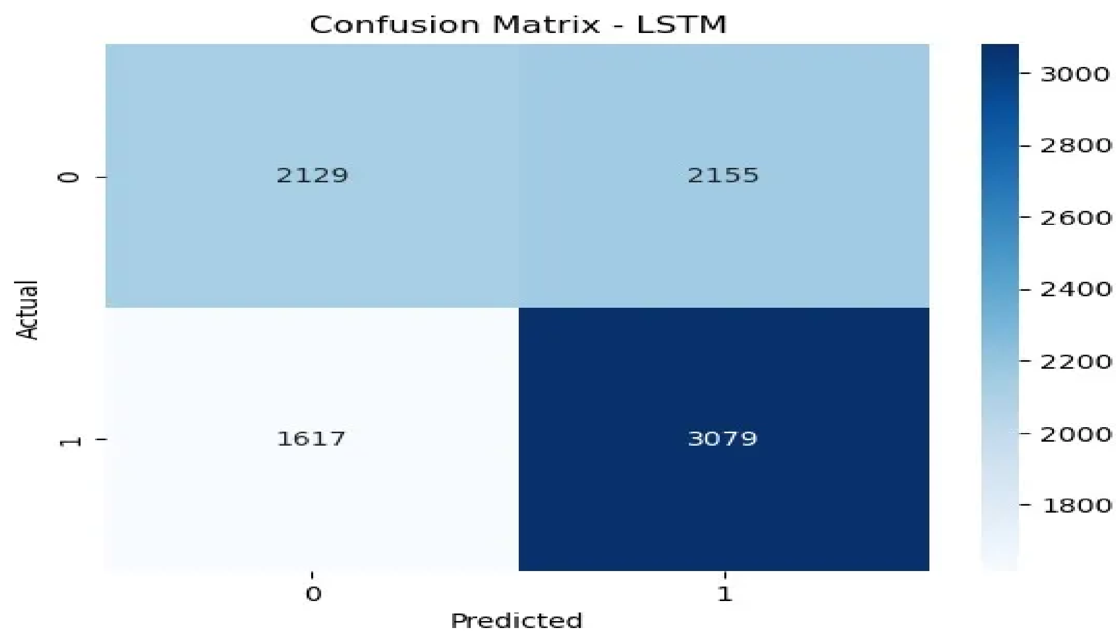


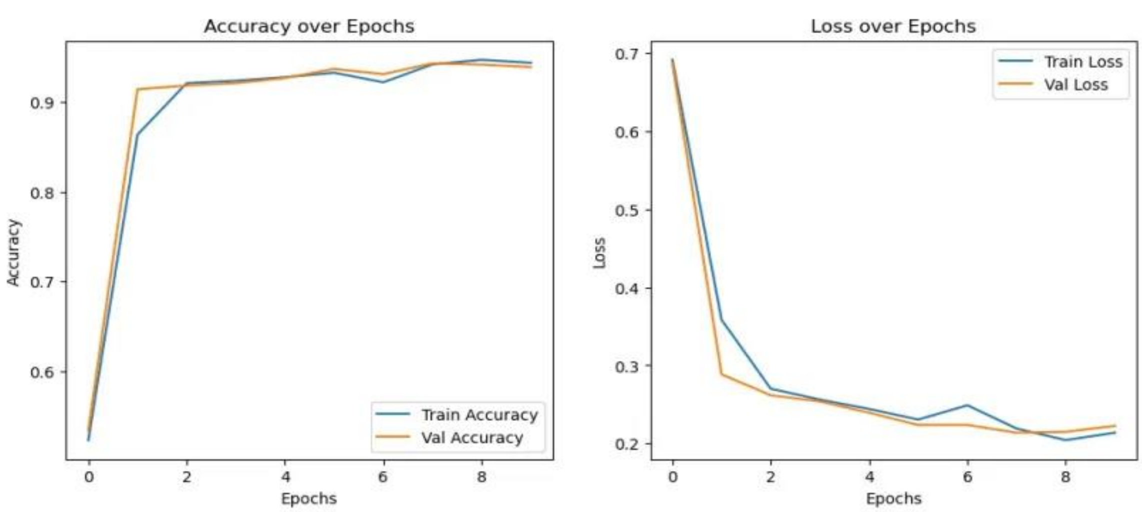


3.4 Sequential Deep Learning Models

An LSTM network with embedding layers was implemented. The model showed signs of overfitting, with validation accuracy dropping significantly in the final epoch, indicating a need for better regularization or a larger dataset.

- LSTM Final Val Accuracy: 59.79%
- LSTM Test Accuracy: 58%

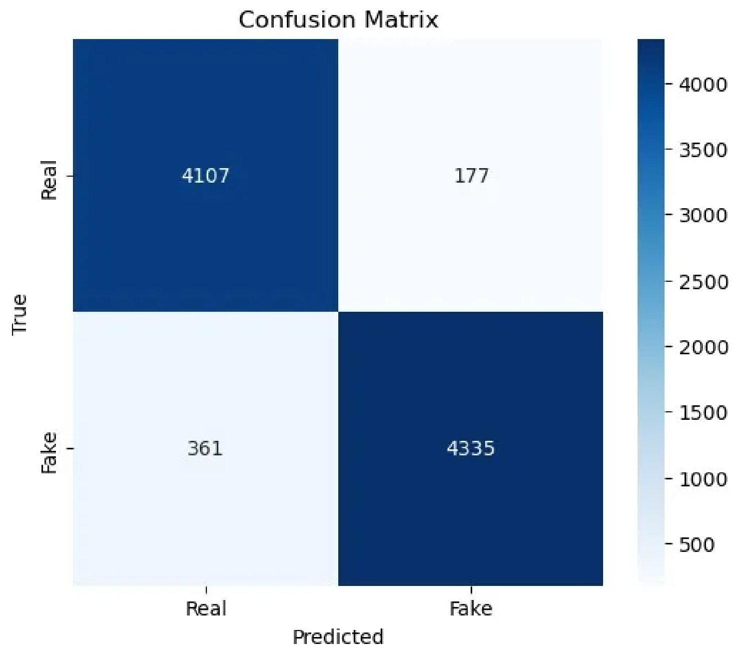


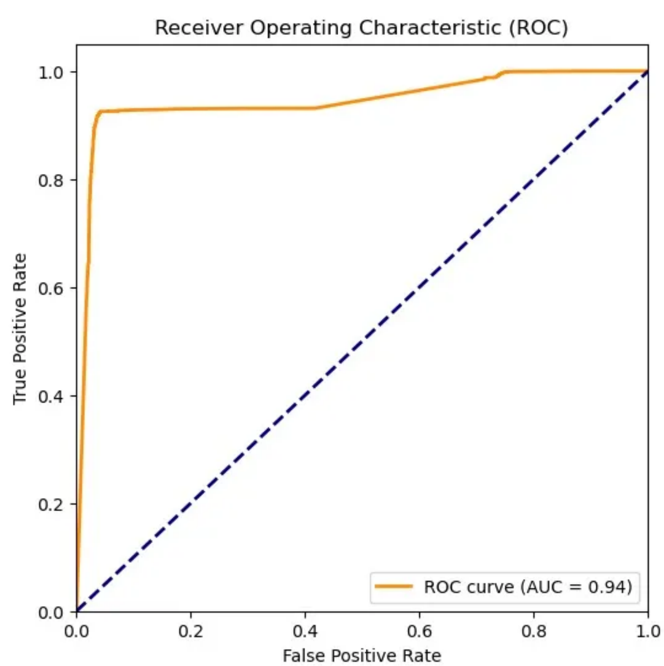


3.5 Transformer-based Models

The DistilBERT model was fine-tuned on the dataset, achieving high accuracy with much more stable performance than the LSTM.

- DistilBERT Validation Accuracy: 96.0%
- DistilBERT Test Accuracy: 94.4%

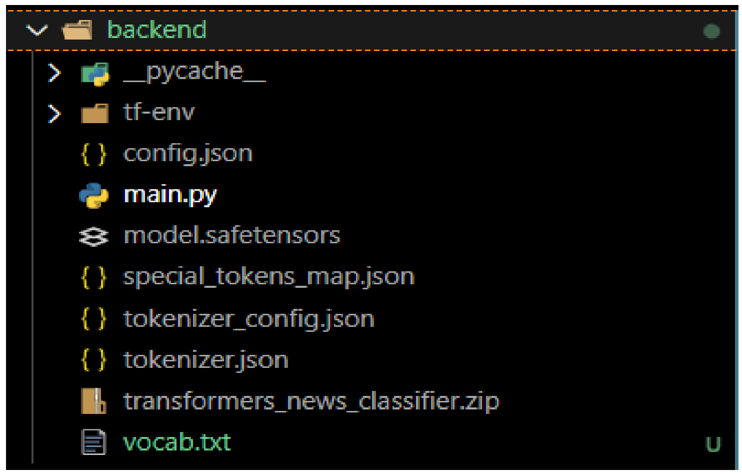




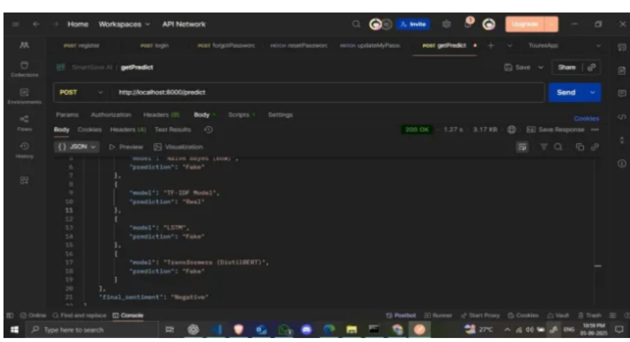
4. Deployment: The NEWSCHECK Application

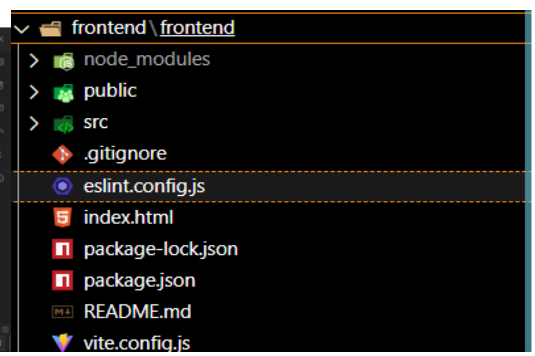
The final model was deployed as "NEWSCHECK," a web application that provides real-time analysis.

- **Backend (Flask):** Hosts the fine-tuned DistilBERT model and provides a REST API endpoint (/api/predict).



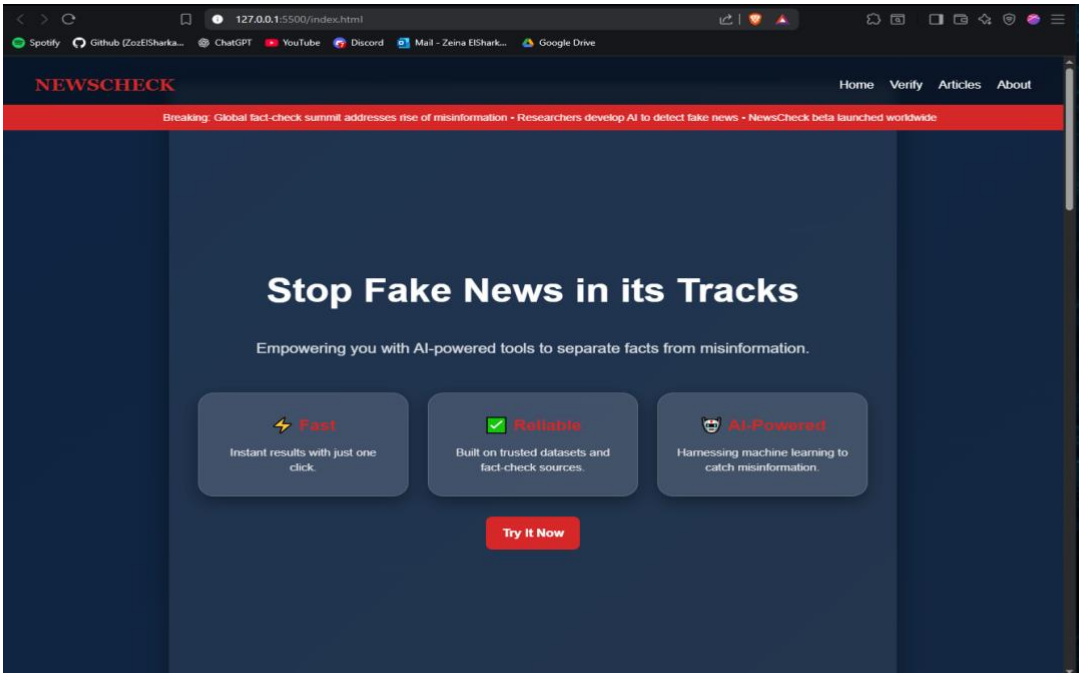
- **Frontend (React):** A user-friendly Single Page Application for input and result visualization

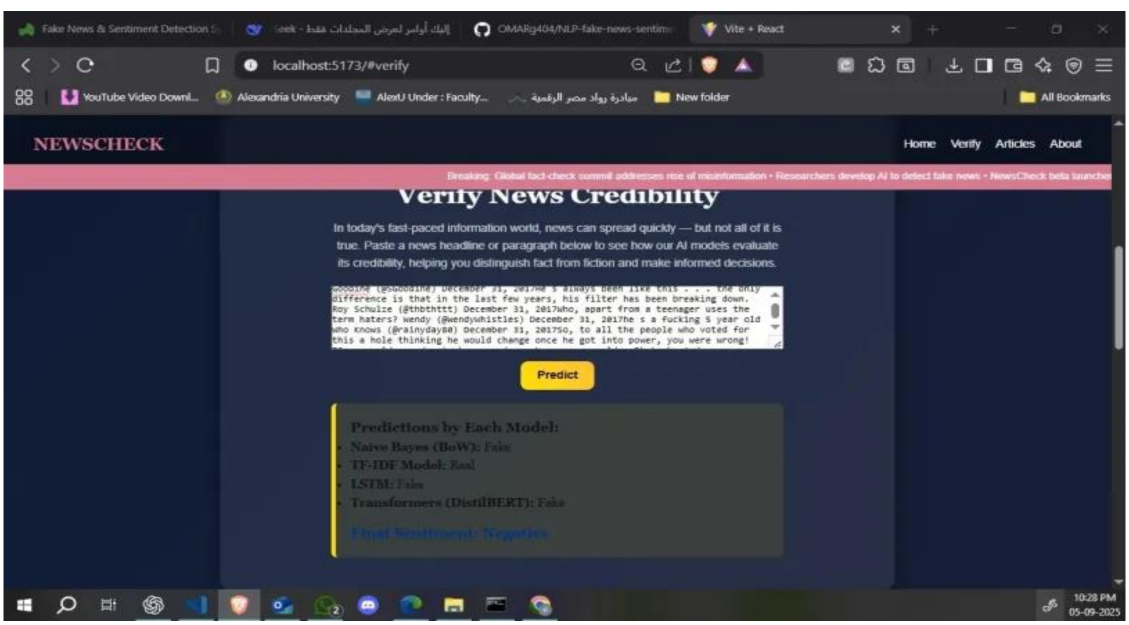




Output:

- Fake / Real prediction
- Sentiment: Positive / Neutral / Negative





5. Results

5.1 Classical Models

- Naïve Bayes (Bag-of-Words) achieved accuracy of 98.86%
- Linear SVM (TF-IDF) achieved accuracy of 99.61%

=== Training Logistic Regression = ==

Accuracy: 0.98864142538975

	precision	recall	F1- score	support
0	0.99	0.99	0.99	4696
1	0.99	0.99	0.99	4284
accuracy			0.99	8980
macro avg	0.99	0.99	0.99	8980
weighted avg	0.99	0.99	0.99	8980

=== Training Linear SVM ===

Accuracy: 0.9961024498886414

	precision	recall	F1- score	support
0	0.99	0.99	0.99	4696
1	0.99	0.99	0.99	4284
accuracy			0.99	8980
macro avg	0.99	0.99	0.99	8980
weighted avg	0.99	0.99	0.99	8980

5.2 Sequential Models

- LSTM (GloVe Embeddings) achieved accuracy of 58.0%

Epoch 1/5 | Val Accuracy: 0.9587 | Val Loss: 0.1607
Epoch 2/5 | Val Accuracy: 0.9399 | Val Loss: 0.2362
Epoch 3/5 | Val Accuracy: 0.9548 | Val Loss: 0.1595
Epoch 4/5 | Val Accuracy: 0.9605 | Val Loss: 0.1549
Epoch 5/5 | Val Accuracy: 0.5979 | Val Loss: 0.6167 <<< COLLAPSE

=== Test Set Classification Report ===

	precision	recall	F1- score	support
0	0.57	0.50	0.53	4283
1	0.59	0.66	0.62	4696
accuracy			0.58	8980
macro avg	0.58	0.58	0.58	8980
weighted avg	0.58	0.58	0.58	8980

5.3 Transformer Models

- DistilBERT (Contextual Embeddings) achieved accuracy of 94.4%

Fine-tuning Evaluation Metrics:
{'eval_loss': 0.1341, 'eval_accuracy': 0.96, 'epoch': 1.0}

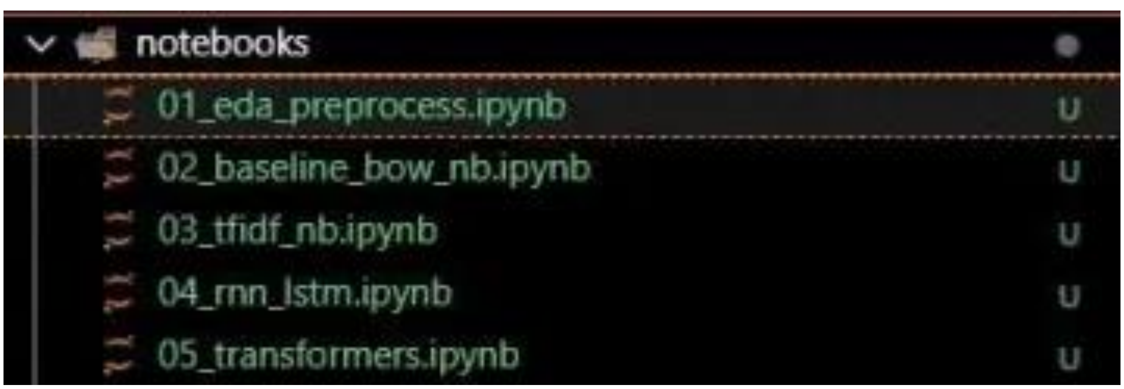
=== Test Set Classification Report ===
AUC Score: 0.9439

	precision	recall	F1- score	support
0	0.92	0.96	0.94	4284
1	0.96	0.92	0.94	4696
accuracy			0.94	8980
macro avg	0.94	0.94	0.94	8980
weighted avg	0.94	0.94	0.94	8980

5.4 Model Performance Comparison

The models were evaluated on a held-out test set. Results are summarized in the table below.

Model	Feature Extraction	Accuracy	Precision	Recall	F1-Score	AUC
Naïve Bayes	Bag-of-Words	0.9886	0.99	0.99	0.99	-
Logistic Regression	TF-IDF	0.9886	0.99	0.99	0.99	-
Linear SVM	TF-IDF	0.9961	1.00	1.00	1.00	1.00
LSTM	GloVe Embeddings	0.5800	0.58	0.58	0.58	-



6. Error Analysis

6.1 Common Misclassifications

- Sarcasm and humor often misclassified as real.
- Headlines with ambiguous context misclassified.
- The LSTM model's significant drop in performance is a classic case of overfitting. The model memorized the training data but failed to generalize to the validation and test sets.
- While the SVM score is perfect, this result is likely overly optimistic and may not generalize to real-world, out-of-sample data, suggesting potential data leakage or an overly simple test set.

6.2 Model Limitations

- Sequential Models (LSTM): Struggled with training stability and overfitting, despite their theoretical ability to handle long-range dependencies.
- Transformers (DistilBERT): Achieved robust and generalizable performance but required significant computational resources for fine-tuning, though inference is faster than BERT.
- Bias: The dataset is heavily focused on US political news (e.g., Trump, Clinton, Obama), so model performance may degrade on news from other domains or geographical origins.

7. Ethical Considerations

7.1 Dataset Bias

The dataset has a clear political bias, over-representing certain topics and outlets. A model trained on this data could inherit these biases and perform poorly on apolitical or international news.

7.2 Transparency and Interpretability

The "black box" nature of deep learning models is a concern. While we provided word-level insights (Figure 3), using explainability frameworks like SHAP or LIME in the future would be crucial for building trust.

7.3 Responsible AI Use

This tool is designed to support human fact-checkers, not replace them. There is a risk of misuse for censorship or automating the suppression of dissenting views. Ethical deployment requires transparency, ongoing bias audits, and clear communication of the model's limitations.

8. Conclusion

The project demonstrates a complete NLP pipeline for fake news and sentiment detection. from a simple yet powerful Linear SVM to a sophisticated Transformer-based model (DistilBERT) outperform classical and sequential models in accuracy and robustness. Deployment via Flask and React ensures accessibility. However, ethical considerations remain vital for responsible use