# Fake News Detection - Project Report

## 1. Introduction

The aim of this project was to develop a machine learning model capable of identifying fake news articles using text classification techniques. The system was trained on a labeled dataset containing both fake and real news articles. The final model was deployed as a simple application that allows users to input news content and receive a classification result.

## 2. Approach

a. Data Collection and Labeling

We used the publicly available Fake and Real News Dataset from Kaggle, which includes two CSV files:

- Fake.csv – Containing fabricated news articles

- True.csv – Containing authentic, verified news articles

Each article was labeled as:

- 0 for fake

- 1 for real

b. Preprocessing and Cleaning

Preprocessing included removal of punctuation, conversion to lowercase, tokenization, stopword removal, and lemmatization.

c. Feature Extraction

We used TF-IDF Vectorization with max\_features=5000 to convert text into numerical vectors.

d. Model Selection and Training

We trained a Multinomial Naive Bayes model using an 80/20 train-test split. This model was chosen for its efficiency in text classification.

## 3. Challenges Faced

- Data Imbalance: Slight variations in dataset sizes caused minor class imbalance.

- Text Noise: Headlines or repeated phrases introduced bias.

- Model Limitations: Naive Bayes does not understand word order or context.

- Deployment Compatibility: Required careful alignment of preprocessing and vectorization.

## 4. Model Performance

The final model achieved the following on the test data:

Accuracy: ~93%

Precision (Real): 0.92

Recall (Real): 0.94

F1-Score (Real): 0.93

This performance indicates good reliability for detecting fake news.

## 5. Improvements and Future Work

- Use of Deep Learning: Incorporating LSTM or BERT can improve context understanding.

- Multilingual Support: Extend detection to multiple languages.

- Real-Time Scraping: Enable real-time classification of news from the web.

- Explainable AI: Add explanations for each classification decision.

## Conclusion

The fake news detection system demonstrates that even simple models like Naive Bayes, when supported with good preprocessing and features, can tackle real-world NLP problems effectively. The system serves as a reliable baseline for more advanced fake news detection systems.