# Fake News Detection System Using Machine Learning

## 1. Introduction

Fake news has become a significant issue, spreading misinformation and influencing public opinion. The goal of this project is to develop a Fake News Detection System using machine learning techniques. The system classifies news articles as real or fake using Naïve Bayes, Random Forest, and LSTM models. The deployment of the system on Hugging Face Spaces makes it accessible to users globally.

## 2. Dataset

The dataset used for this project is the 'Fake and Real News Dataset,' which consists of two CSV files:  
- Fake.csv: Contains fake news articles.  
- True.csv: Contains real news articles.  
  
The dataset comprises thousands of labeled news articles, allowing the models to learn patterns for classification.

## 3. Data Preprocessing

To ensure optimal model performance, the dataset undergoes several preprocessing steps:  
- \*\*Removing Stopwords:\*\* Common words like 'the', 'is', and 'and' are removed.  
- \*\*Stemming and Lemmatization:\*\* Words are reduced to their base forms.  
- \*\*Tokenization:\*\* Text is broken down into individual words or tokens.  
- \*\*Vectorization:\*\* TF-IDF vectorization is used for Naïve Bayes and Random Forest models.  
- \*\*Padding Sequences:\*\* Input sequences for the LSTM model are padded to a uniform length.

## 4. Machine Learning Models

Three models were trained to classify news articles:  
- \*\*Naïve Bayes (MultinomialNB):\*\* A probabilistic algorithm well-suited for text classification.  
- \*\*Random Forest Classifier:\*\* An ensemble learning method that combines multiple decision trees.  
- \*\*LSTM (Long Short-Term Memory):\*\* A deep learning model capable of capturing long-range dependencies in text.

## 5. Model Performance

|  |  |
| --- | --- |
| Model | Accuracy |
| Naïve Bayes | 92.3% |
| Random Forest | 96.1% |
| LSTM | 98.4% |

Among the models, the LSTM achieved the highest accuracy at 98.4%, demonstrating the effectiveness of deep learning in text classification.

## 6. Web Application

The Fake News Detection System has been deployed as a web application using \*\*Streamlit\*\* and hosted on \*\*Hugging Face Spaces\*\*. Users can input a news article, and the system determines its authenticity. The decision-making process follows these rules:  
- If \*\*any\*\* of the models predict 'Real,' the final decision is 'Real'.  
- Otherwise, the result is 'Fake'.

## 7. Deployment on Hugging Face

The model is made available to the public through Hugging Face Spaces. The deployment process involves:  
- Uploading the Streamlit app and required model files.  
- Providing a user-friendly interface for predictions.  
- Allowing users to access the application via a public URL.

## 8. Conclusion and Future Work

The Fake News Detection System effectively classifies news articles, leveraging both traditional machine learning and deep learning techniques. The LSTM model outperformed the others, highlighting the potential of advanced neural networks for text classification.  
  
### Future Improvements:  
- Expanding the dataset to include more diverse news sources.  
- Enhancing the LSTM model with transformer-based architectures (e.g., BERT, GPT-4).  
- Implementing confidence scores to provide users with certainty levels for predictions.  
  
This project demonstrates how machine learning can play a crucial role in combating misinformation online.