Fake News Detection Project

Introduction:

This project aims to classify news articles as Fake or Real using Natural Language Processing (NLP) techniques. It leverages datasets containing real and fake news articles, and employs machine learning algorithms for classification.

Abstract:

The system uses text preprocessing (tokenization, stop-word removal, stemming) and TF-IDF vectorization to convert textual data into numerical form. Logistic Regression and Naive Bayes classifiers are trained to identify fake news patterns. A Streamlit web app provides an interactive interface for predictions.

Tools Used:

 - Python 3

- Libraries: Pandas, NumPy, Scikit-learn, NLTK, Joblib

- Visualization: Matplotlib, Seaborn

 - Web App: Streamlit

Steps Involved:

1. Combine and preprocess Fake.csv and True.csv datasets.

2. Clean the text using regex, stopword removal, and stemming.

 3. Convert text to numerical features using TF-IDF.

4. Train Logistic Regression and Naive Bayes models.

5. Evaluate models using accuracy, F1-score, and confusion matrix.

6. Save trained models and vectorizers using Joblib.

7. Build a Streamlit app for real-time predictions.

Conclusion:

The Logistic Regression model achieved an accuracy of ~98%, outperforming Naive Bayes (~93%). This demonstrates the effectiveness of feature engineering and text classification techniques in detecting fake news.