Project 6: Fake News Article Classification

Objective:

The objective of this project is to classify news articles as either Fake or Real using machine learning and natural language processing techniques.

Tools and Technologies Used:

- Python (Google Colab)
- Libraries: pandas, numpy, matplotlib, seaborn, nltk, scikit-learn, joblib

Dataset Description:

The dataset used consists of two files: 'Fake.csv' and 'True.csv'. Each file contains news articles labeled as either fake or real. The combined dataset includes text-based news articles with a label column (0 for Fake, 1 for Real). The dataset was sourced from Kaggle.

Data Preprocessing:

- Combined both datasets and assigned labels (0 for fake, 1 for real)
- Cleaned the text data by converting to lowercase, removing punctuation, and removing stopwords using NLTK
- Applied stemming to reduce words to their root forms
- Used TF-IDF Vectorizer to convert text into numerical features

Model Building:

Two models were trained:

1. Logistic Regression
2. Multinomial Naive Bayes

The dataset was split into training and testing sets (80:20). Both models were trained and evaluated to determine which performed better.

Project 6: Fake News Article Classification

Evaluation:

The models were evaluated using accuracy, precision, recall, F1-score, and confusion matrix. Naive Bayes showed slightly better performance for this dataset.

Conclusion:

The project demonstrates how machine learning and NLP techniques can be used effectively for binary text classification tasks like fake news detection. The trained model and TF-IDF vectorizer were saved using joblib for future use or deployment.

Author Name: [Your Name]

Date: 26-07-2025