# NEWS ARTICLE CLASSIFICATION (FAKE OR REAL)

## Introduction

With the rapid spread of digital content, fake news has become a major concern. This project aims to build a machine learning model that can classify news articles as Fake or Real using natural language processing (NLP).

## Abstract

The model was trained on a Kaggle dataset containing labeled news articles. Text data was preprocessed using NLTK, vectorized using TF-IDF, and classified using Logistic Regression. The final model achieved high accuracy and was deployed as a web application using Streamlit. Users can enter any news content and get instant classification results.

## Tools Used

Python, NLTK, Scikit-learn, Pandas, Streamlit, Joblib

## Steps Involved

1. Data Collection: Downloaded Fake.csv and True.csv from Kaggle and merged with labels.
2. Data Cleaning: Used NLTK to remove punctuation, lowercase text, and remove stopwords.
3. Text Vectorization: Applied TF-IDF to convert text into numerical vectors.
4. Model Training: Trained a Logistic Regression model with 80/20 train-test split.
5. Evaluation: Achieved strong performance with high accuracy and F1-score.
6. App Development: Built a simple interface using Streamlit for user input and prediction display.

## Conclusion

The fake news classifier successfully detects fake vs real news articles using basic NLP techniques. The web interface makes it accessible and useful. Future improvements could involve deep learning, real-time news scraping, and multilingual support.