

# Fake News Detection Using Machine Learning

## Introduction

In the digital age, the spread of fake news has become a pressing issue. Misinformation can have harmful consequences on society, especially when it influences public opinion on critical topics such as health, politics, and national security. This project aims to combat fake news by using Natural Language Processing (NLP) and Machine Learning (ML) to build a web-based Fake News Detection System that can classify articles as real or fake based on their content.

## Abstract

This project implements a binary text classification system using a supervised machine learning model to detect fake news. We trained a Logistic Regression classifier on a labeled dataset containing both fake and real news articles. Text data was vectorized using the TF-IDF method to convert words into numerical features. The final model is integrated with a user-friendly web interface using Streamlit, allowing users to input news content and receive real-time classification feedback.

## Tools Used

- Python 3.10
- scikit-learn – for ML model and vectorization
- Pandas – for dataset handling
- Streamlit – to build the web app
- Pickle – to save and load the trained model
- Jupyter Notebook / VS Code – for development

## Steps Involved in Building the Project

1. Dataset Collection
  - Used a publicly available dataset of fake and real news from Kaggle.
  - The dataset includes news headlines, article texts, and labels (FAKE or REAL).
2. Data Preprocessing
  - Cleaned text data and encoded labels as binary (0 for FAKE, 1 for REAL).
  - Split dataset into training and test sets.
3. Feature Extraction
  - Applied TF-IDF Vectorization to convert text into numerical form.
4. Model Training
  - Trained a Logistic Regression model using scikit-learn.
  - Evaluated model accuracy on test data.
5. Model Deployment
  - Saved the trained model and TF-IDF vectorizer using Pickle.
  - Built a user interface using Streamlit where users can paste a news article and get an instant result: FAKE or REAL.

## 6. Testing

- Deployed the app locally and on Streamlit Cloud for public use.

## Conclusion

This project demonstrates how machine learning can be effectively used to tackle real-world problems like misinformation. The resulting Fake News Detection System is accurate, lightweight, and easy to use. It offers users a quick way to verify the authenticity of online news content, helping to promote more informed and responsible consumption of information.