Fake News Detection using Machine Learning

# 1. Abstract

This project aims to detect fake news articles using machine learning algorithms.   
With the rise of misinformation, a reliable model that can automatically detect fake news is highly valuable.   
The dataset used includes both fake and real news, and the model is trained using logistic regression.

# 2. Introduction

Fake news detection is the process of classifying news content as real or fake.   
Machine learning models can assist in this task by learning patterns in text data that distinguish fake news from real news.

# 3. Dataset Information

The dataset was taken from Kaggle and contains two CSV files: Fake.csv and True.csv.  
Each file contains the text of the news article and other metadata. Labels are assigned manually (0 for fake, 1 for real).

# 4. Tools and Libraries

- Python  
- Pandas  
- NumPy  
- Scikit-learn  
- Matplotlib  
- Seaborn  
- Streamlit (for deployment)

# 5. Data Preprocessing

The text data was cleaned and transformed using TF-IDF Vectorizer to convert textual data into numerical features.  
Stopwords were removed and the data was split into training and test sets.

# 6. Machine Learning Model

A Logistic Regression model was trained on the TF-IDF features. The model achieved an accuracy of over 98% on the test data.  
It was able to effectively distinguish between fake and real news articles.

# 7. Evaluation

The model was evaluated using accuracy, confusion matrix, and classification report.   
A confusion matrix was plotted to visualize the performance.

# 8. Conclusion

The project successfully demonstrates that machine learning techniques can be effectively used for fake news detection.  
The high accuracy indicates the potential of these models in real-world applications.

# 9. References

Dataset: https://www.kaggle.com/datasets/clmentbisaillon/fake-and-real-news-dataset