Fake News Detection using Machine Learning

■ Project Overview

This project aims to classify news articles as Fake News or True News using machine learning techniques.

The dataset consists of two CSV files:

• Fake.csv → contains fake news articles

• True.csv → contains genuine news articles

We preprocess the dataset, extract features using TF-IDF vectorization, and train multiple machine learning classifiers to evaluate accuracy.

■■ Workflow

1. Data Preparation

• Merged Fake.csv and True.csv datasets.

• Labeled:

• - Fake news → class = 0

• - True news → class = 1

• Dropped irrelevant columns like title, subject, date.

• Shuffled the dataset for unbiased training.

• Cleaned text data by removing:

• - URLs, punctuation, numbers, special characters

• - Extra spaces and line breaks

2. Feature Extraction

• Used TF-IDF Vectorizer (TfidfVectorizer) to convert text into numerical features.

• Training and testing data split: 75% train / 25% test.

3. Machine Learning Models

We applied and compared four classifiers:

• Logistic Regression (LR)

• Decision Tree (DT)

• Gradient Boosting (GB)

• Random Forest (RF)

4. Model Evaluation

• Accuracy measured using accuracy_score.

• Performance compared with classification_report (Precision, Recall, F1-score).

5. Manual Testing Function

A custom function manual_testing(news) was built:

• Takes user input (any news text).

• Cleans and vectorizes the text.

• Predicts whether it is Fake News or True News using all four models.

■ Results

• Logistic Regression: High accuracy, good for linear patterns.

• Decision Tree: Performed well but prone to overfitting.

• Gradient Boosting: Balanced accuracy with strong performance.

• Random Forest: Best overall accuracy with strong generalization.

■ How to Run

1. Install dependencies:

pip install pandas numpy scikit-learn matplotlib seaborn

2. Place Fake.csv and True.csv in the project folder.

3. Run the Python script (fake_news_detection.py).

4. To test manually:

news = input("Enter the news text: ")

manual_testing(news)

■ Project Files

• Fake.csv → Fake news dataset

• True.csv → True news dataset

• fake_news_detection.py → Main project script

• README.pdf → Project documentation

■■■ Author

ANSHIKA KASHYAP