Fake News Detection using Python & Machine Learning

---

1. Introduction

This project aims to detect whether a news article is Fake or Real using Machine Learning and Natural Language Processing techniques. The system uses TF-IDF vectorization for text feature extraction and Passive Aggressive Classifier for classification.

2. Objectives

- Detect fake news using Machine Learning.
- Apply Natural Language Processing techniques.
- Train and evaluate a classification model.
- Provide real-time prediction for user input.

3. Technologies Used

- Python
- Pandas
- NumPy
- Scikit-learn
- Joblib

4. Project Architecture

Step 1: Load Dataset (news.csv) Step 2: Preprocess Data (remove null values) Step 3: Convert text into TF-IDF features Step 4: Train Passive Aggressive Classifier Step 5: Evaluate model using accuracy and confusion matrix Step 6: Save model and vectorizer Step 7: Predict new news input

5. Model Details

The Passive Aggressive Classifier is well-suited for large-scale text classification problems. It updates the model only when misclassification occurs, making it efficient and accurate. TF-IDF (Term Frequency - Inverse Document Frequency) converts textual data into numerical format by measuring word importance in documents.

6. Project Structure

FakeNewsDetection/ ■ ■■■ data/ ■ ■■■ news.csv ■■■ src/ ■ ■■■ train_model.py ■ ■■■ predict.py ■■■ model/ ■■■ requirements.txt ■■■ README.md

7. Results

The model typically achieves high accuracy (above 90%) depending on dataset quality. It can classify news articles as FAKE or REAL effectively.

8. Conclusion

This project demonstrates how Machine Learning and NLP can help combat misinformation by automatically classifying news content. It is suitable for academic submission, portfolio showcase, and practical implementation.