Documentation of Project

Project name:

Fake news detector

Name:

Dua Zahra

Roll no:

BSAI_036

SUBMITTED BY:

SIR RASIK ALI

DEATAIL:

This project is a Fake News Detector built using Python and machine learning. It classifies news articles as either Fake News or True News using a Logistic Regression model trained on a dataset of labeled news articles. The project demonstrates text preprocessing, feature extraction using TF-IDF, and model training for accurate classification.

Features

Detects whether a news article is fake or true.

Uses TF-IDF Vectorization to transform text into numerical data.

Implements a Logistic Regression classifier for text classification.

Provides high accuracy with robust evaluation metrics.

Technologies Used

Python

Machine Learning (Logistic Regression)

Natural Language Processing (TF-IDF)

Libraries: pandas, sickie-learn, pickle, streamlet

Dataset

The project uses two datasets:

Fake News Dataset (Fake.csv)

Contains news articles labeled as fake (0).

True News Dataset (True.csv)

Contains news articles labeled as true (1).

Sample Dataset Format:

text label

"Climate change is a hoax." 0

"NASA confirms the Earth is round." 1

Installation

Prerequisites

Python 3.7 or later

Required Libraries: Install via requirements.txt

Steps

Clone the repository:

bash

Copy code

git clone https://github.com/your-username/Fake-News-Detector.git

Navigate to the project directory:

bash

Copy code

cd Fake-News-Detector

Install dependencies:

bash

Copy code

pip install -r requirements.txt

Usage

1. Train the Model

Run the training script to train the Logistic Regression model and save the model along with the TF-IDF vectorizer:

bash

Copy code

python scripts/fake_news_detector.py

2. Run the Application

Start the Streamlit app to interact with the Fake News Detector:

bash

Copy code

streamlit run app.py

3. Input News Text

Paste a news article into the app's text box.

Click "Analyze" to see whether the article is classified as Fake News or True News.

Project Structure

graph

Copy code

Fake-News-Detector/

```
├── app.py                # Streamlit application
├── data/
│   ├── Fake.csv          # Fake news dataset
│   ├── True.csv          # True news dataset
├── models/
│   ├── model.pkl         # Trained Logistic Regression model
│   ├── tfidf_vectorizer.pkl  # TF-IDF vectorizer
├── scripts/
│   └── fake_news_detector.py # Training script
├── requirements.txt      # Python dependencies
└── README.md             # Project documentation
```

Results

Model Performance

Metric Fake News (0) True News (1) Overall

Confusion Matrix

The model shows high accuracy and minimal misclassification between classes.

Future Enhancements

Add support for URL-based news verification.

Integrate more advanced models like Random Forest, XGBoost, or BERT.

Deploy the application online using Heroku, AWS, or Google Cloud.