

Fake News & Hate Speech Detector

This project uses Natural Language Processing (NLP) and Machine Learning to classify text as: - **Real vs Fake News - Hate Speech vs Neutral Speech**

The system supports data preprocessing, training, evaluation, and a Streamlit web interface for interactive testing.

Features

- Text preprocessing (cleaning, tokenization, stopword removal)
- TF-IDF vectorization
- Logistic Regression / SVM classification
- Dual-class detection (fake news + hate speech)
- Streamlit interface for user input

Project Structure

```
fake-news-hate-speech-detector/  
|-- src/  
|   |-- main.py          # Streamlit UI  
|   |-- train.py         # Model training script  
|   |-- preprocess.py    # Text cleaning utilities  
|   |-- model.py         # Model load/save utilities  
|  
|-- data/  
|   |-- fake_news.csv  
|   |-- hate_speech.csv  
|  
|-- models/  
|   |-- fake_news_model.joblib  
|   |-- hate_speech_model.joblib  
|  
|-- requirements.txt  
|-- README.md
```

Installation

```
git clone <your_repo_link>  
cd fake-news-hate-speech-detector  
pip install -r requirements.txt
```

Training

```
python src/train.py
```

Running the App

```
streamlit run src/main.py
```

Resume Example Entry

- Built and deployed a dual-task NLP classifier to detect misinformation and hate speech using TF-IDF and SVM models, including an interactive Streamlit web interface.

Next Enhancements

- Swap TF-IDF with BERT embeddings
- Add explainability (SHAP / LIME)
- Add multilingual support