**READ ME ( PROJECT OVERVIEW ) FILE**

**📹 YouTube Spam Detection**

****

Welcome to the **YouTube Spam Detection** project! This repository hosts a machine learning model designed to detect and filter out spam comments on YouTube, making it easier for creators and viewers to enjoy a spam-free experience.

**🚀 Project Overview**

Spam comments on YouTube can range from irrelevant links to promotional messages, phishing attempts, and scams, creating a frustrating experience for genuine users. This project tackles this challenge by building a model that can distinguish between spam and non-spam comments with high accuracy, helping to keep YouTube comment sections clean and relevant.

**🔍 Key Features**

* **Efficient Spam Detection:** Detect spam comments with precision to enhance user experience on YouTube.
* **Machine Learning-Based:** Leverages popular ML algorithms and natural language processing (NLP) to understand comment patterns.
* **Real-World Dataset:** Trained on a diverse dataset of YouTube comments to improve detection accuracy.
* **Easy Integration:** Designed to be integrated with YouTube or other platforms for automated spam detection.

**📂 Project Structure**

bash

Copy code

├── data/ # Contains the dataset for training and testing (e.g., dat.csv)

├── notebooks/ # Jupiter notebooks for data exploration and model training

├── models/ # Trained machine learning models

├── src/ # Source code for data preprocessing, training, and evaluation

│ ├── data\_preprocessing.py

│ ├── model\_training.py

│ └── evaluation.py

├── README.md # Project overview and instructions

└── requirements.txt # List of dependencies

**📊 Dataset**

The model is trained on a dataset of YouTube comments, including both spam and non-spam instances. **Note:** Due to privacy and compliance policies, only sample or anonymized data may be shared in this repository.

**🛠️ Installation**

1. **Clone the repository**:

bash

Copy code

git clone https://github.com/yourusername/youtube-spam-detection.git

cd youtube-spam-detection

1. **Install dependencies**:

bash

Copy code

pip install -r requirements.txt

1. **Download or place the dataset** in the data/ directory.

**🚅 Quick Start**

1. **Data Preprocessing**: Run the data preprocessing script to clean and prepare the comments.

bash

Copy code

python src/data\_preprocessing.py

1. **Train the Model**: Train the spam detection model using the training data.

bash

Copy code

python src/model\_training.py

1. **Evaluate the Model**: Run the evaluation script to test the model's performance.

bash

Copy code

python src/evaluation.py



**📈 Model Performance**

After training, the model achieves a **high accuracy** in distinguishing between spam and non-spam comments. The evaluation metrics such as **accuracy**, **precision**, and **recall** are available in the project documentation.

**🎓 Tech Stack**

* **Python**
* **scikit-learn**
* **Natural Language Processing (NLP) libraries**: NLTK, SpaCy
* **Pandas & NumPy** for data manipulation

**🤝 Contributions**

Contributions are welcome! If you have suggestions, issues, or would like to contribute, please fork the repository and create a pull request.

**📜 License**

This project is licensed under the MIT License. See the LICENSE file for details.