**Spam Message Detection Web App**

A simple machine learning-based \*\*Spam Detection\*\* web application built with \*\*Flask\*\*. The app uses a \*\*Naïve Bayes classifier\*\* to detect whether a given text message is \*\*Spam\*\* or \*\*Not Spam\*\*.

🚀 Features

- Classifies SMS messages as \*\*Spam\*\* or \*\*Not Spam\*\*

- Uses \*\*TF-IDF Vectorization\*\* for text preprocessing

- Trained with the \*\*Multinomial Naive Bayes\*\* model

- Interactive and responsive UI using HTML + Flask

- Web-based form for live testing

📁 Project Structure

spam-detection/  
│  
├── spam.csv # Dataset (Ham/Spam messages)  
├── spam\_detection.py # Main Flask application and ML logic  
├── templates/  
│ └── index.html # Frontend HTML template  
└── README.md # Project documentation (this file)

🧠 Technologies Used

- Python 3.10+

- Flask

- Pandas

- Scikit-learn

- Regular Expressions

- HTML/CSS (via Flask templating)

📦 Installation & Setup

**1. Clone the repository or download the files**

```bash

git clone https://github.com/your-username/spam-detection.git

cd spam-detection

**2. Install the required Python libraries**

pip install -r requirements.txt

If you don't have a requirements.txt, install manually:

pip install flask pandas scikit-learn

**3. Run the Flask App**

python spam\_detection.py

**4. Open in Browser**

Go to [http://127.0.0.1:5000](http://127.0.0.1:5000/)

**📊 Dataset Information**

* Source: [spam.csv](https://www.kaggle.com/datasets/uciml/sms-spam-collection-dataset)
* Columns:
  + label: ham (not spam) or spam
  + message: the raw SMS text
* Preprocessing includes:
  + Lowercasing
  + Removing numbers and punctuation
  + TF-IDF vectorization

**📈 Model Info**

* **Algorithm**: Multinomial Naïve Bayes
* **Text Vectorization**: TF-IDF (top 5000 features)
* **Accuracy**: ~97% on test set

**🧠 How Prediction Works**

1. User submits a message via the form.
2. The message is cleaned and converted to a vector using TF-IDF.
3. The trained Naïve Bayes model predicts the label.
4. Result is displayed as “Spam” or “Not Spam”.

**✅ Future Improvements**

* Add model training from UI
* Support for email spam classification
* Export prediction logs
* Host using Heroku / Render / SAP BTP

**👩‍💻 Author**

**Swetha Sridharan**  
📍 Student, Sri Ramakrishna Institute of Technology  
💻 Interested in ML, Web Development, and SAP

**📜 License**

This project is open source and available under the [MIT License](https://chatgpt.com/c/LICENSE).