Spam Classification Project Documentation

# 📌 Project Overview

This project builds a spam classifier using machine learning techniques. The goal is to train a model that can distinguish between spam and ham (non-spam) messages based on SMS text data.

# 📂 Dataset: spam.csv

Source: SMS Spam Collection Dataset  
Columns:  
- v1: Label (ham or spam)  
- v2: SMS message text  
  
Preprocessing Applied:  
- Renamed columns to 'label' and 'message'  
- Converted label to numerical (ham = 0, spam = 1)  
- Text cleaning: lowercasing, punctuation removal, stopword filtering, stemming

# 📒 Notebook: spam-classify1.ipynb

Steps Covered:  
1. Data Loading & Cleaning  
2. Label Encoding  
3. Text Preprocessing  
4. Feature Extraction  
5. Model Training  
6. Model Evaluation

# ⚙️ How to Run

1. Install Required Packages:  
 pip install pandas numpy sklearn nltk matplotlib seaborn  
  
2. Download NLTK Resources (inside notebook):  
 import nltk  
 nltk.download('stopwords')  
  
3. Run the Notebook:  
 Open spam-classify1.ipynb in Jupyter Notebook or VS Code and run all cells.

# 🧩 Dependencies

- Python ≥ 3.7  
- pandas  
- numpy  
- scikit-learn  
- nltk  
- matplotlib  
- seaborn

# 📊 Model Output

- Accuracy Score: e.g., 97%+ depending on the model  
- Confusion Matrix: Shows true positives/negatives  
- Classification Report: Precision, Recall, F1-score for both spam and ham