Explanation of Spam Detection Code

This script is used to train and save a spam detection model using a real-world SMS dataset.

1. IMPORTS

- pickle: To save the trained model and vectorizer.

- pandas: For data manipulation.

- sklearn: For machine learning functionalities like vectorization, model training, and data splitting.

2. LOAD DATASET

- Downloads a dataset from a GitHub URL.

- It's a TSV (tab-separated) file with labels (ham/spam) and SMS messages.

3. ENCODE LABELS

- Converts text labels to binary: 'ham' to 0, 'spam' to 1.

4. TRAIN/TEST SPLIT

- Splits the data into 80% training and 20% testing.

5. TEXT VECTORIZATION

- Uses TfidfVectorizer to convert SMS messages into numerical vectors based on term importance.

6. MODEL TRAINING

- Uses Multinomial Naive Bayes classifier to train on the TF-IDF vectors.

7. SAVE MODEL AND VECTORIZER

- Saves the trained model and vectorizer as .pkl files using pickle.

8. CONFIRMATION

- Prints a message indicating successful training and saving.

This code prepares a foundational machine learning model for spam detection which can be integrated into applications like email or SMS filters.