# Import necessary libraries

import pandas as pd

from sklearn.model\_selection import train\_test\_split

from sklearn.feature\_extraction.text import TfidfVectorizer

from sklearn.naive\_bayes import MultinomialNB

from sklearn.metrics import accuracy\_score, confusion\_matrix, classification\_report

# Load the dataset

# Replace 'your\_dataset.csv' with your actual dataset file

dataset = pd.read\_csv('your\_dataset.csv')

# Assuming your dataset has 'text' column for SMS messages and 'label' column for spam or legitimate

X = dataset['text']

y = dataset['label']

# Split the dataset into training and testing sets

X\_train, X\_test, y\_train, y\_test = train\_test\_split(X, y, test\_size=0.2, random\_state=42)

# Convert text data to TF-IDF features

tfidf\_vectorizer = TfidfVectorizer(max\_features=5000)

tfidf\_train = tfidf\_vectorizer.fit\_transform(X\_train)

tfidf\_test = tfidf\_vectorizer.transform(X\_test)

# Train a Naive Bayes classifier

naive\_bayes\_classifier = MultinomialNB()

naive\_bayes\_classifier.fit(tfidf\_train, y\_train)

# Make predictions on the test set

predictions = naive\_bayes\_classifier.predict(tfidf\_test)

# Evaluate the model

accuracy = accuracy\_score(y\_test, predictions)

conf\_matrix = confusion\_matrix(y\_test, predictions)

classification\_rep = classification\_report(y\_test, predictions)

print(f'Accuracy: {accuracy:.2f}')

print('\nConfusion Matrix:')

print(conf\_matrix)

print('\nClassification Report:')

print(classification\_rep)