# 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, classification\_report, confusion\_matrix

import nltk

from nltk.corpus import stopwords

from nltk.tokenize import word\_tokenize

import string

# Download NLTK resources

nltk.download('punkt')

nltk.download('stopwords')

# Load the dataset

url = "/content/spam.csv"

df = pd.read\_csv(url)

# Preprocess the data

df['label'] = df['label'].map({'ham': 0, 'spam': 1})  # Convert labels to binary values

X = df['message']

y = df['label']

# Split the data 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 into numerical data (Bag of Words model)

vectorizer = CountVectorizer()

X\_train\_vec = vectorizer.fit\_transform(X\_train)

X\_test\_vec = vectorizer.transform(X\_test)

# Build and train the model

model = MultinomialNB()

model.fit(X\_train\_vec, y\_train)

# Make predictions on the test set

y\_pred = model.predict(X\_test\_vec)

# Evaluate the model

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

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

print("\nClassification Report:\n", classification\_report(y\_test, y\_pred))

# Example of predicting a new message

new\_message = ["Free entry in 2 a weekly comp to win FA Cup final tkts 21st May 2005. Text FA to 87121 to receive entry question(std txt rate)"]

new\_message\_vec = vectorizer.transform(new\_message)

prediction = model.predict(new\_message\_vec)

print("\nPrediction for new message:", "Spam" if prediction[0] == 1 else "Ham")