3. DEVELOP A PROGRAM FOR SPAM MAIL DETECTION

EX.N0 : 3	**DEVELOP A PROGRAM FOR SPAM MAIL DETECTION**
DATE : /0 /2025	

AIM:

To write a program to develop a program for Spam mail detection.

ALGORITHM:

Step 1: Start

Step 2: Import necessary NLP libraries.

Step 3: Load and preprocess the spam dataset.

Step 4: Convert text into numerical features using TF-IDF Vectorizer.

Step 5: Split the dataset into training and test sets.

Step 6: Train a classifier and Predict labels for test data.

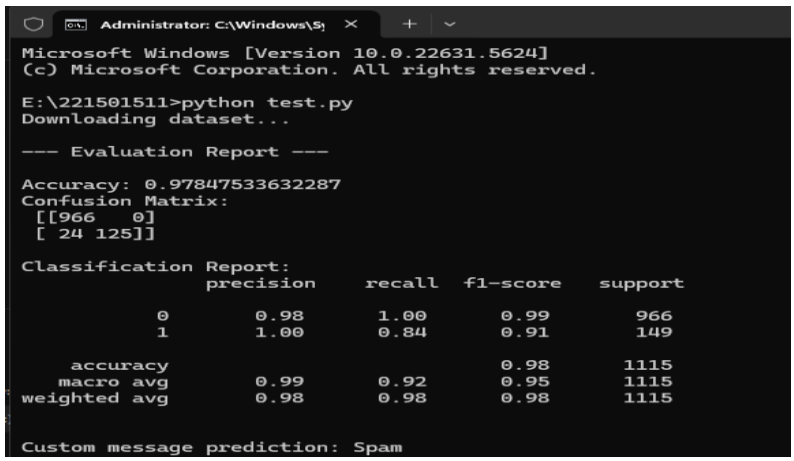
Step 7: Evaluate model performance using accuracy, confusion matrix.

PROGRAM:

```
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 urllib.request
import zipfile
import io
url = "https://archive.ics.uci.edu/ml/machine-learning-
databases/00228/smsspamcollection.zip"
response = urllib.request.urlopen(url)
zip_file = zipfile.ZipFile(io.BytesIO(response.read()))
```

```
zip_file.extractall("sms_data")
df = pd.read_csv("sms_data/SMSSpamCollection", sep='\t', header=None, names=['label', 'message'])
df['label'] = df['label'].map({'ham': 0, 'spam': 1})  # Convert label to binary
X_train, X_test, y_train, y_test = train_test_split(df['message'], df['label'], test_size=0.2, random_state=42)
vectorizer = TfidfVectorizer(stop_words='english')
X_train_tfidf = vectorizer.fit_transform(X_train)
X_test_tfidf = vectorizer.transform(X_test)
model = MultinomialNB()
model.fit(X_train_tfidf, y_train)
 y_pred = model.predict(X_test_tfidf)
print("\n--- Evaluation Report ---\n")
print("Accuracy:", accuracy_score(y_test, y_pred))
print("Confusion Matrix:\n", confusion_matrix(y_test, y_pred))
print("\nClassification Report:\n", classification_report(y_test, y_pred))
custom_message = ["Congratulations! You've won a free ticket to Bahamas. Text WIN to 12345."]
custom_vec = vectorizer.transform(custom_message)
result = model.predict(custom_vec)
print("\nCustom message prediction:", "Spam" if result[0] == 1 else "Ham")
```

OUTPUT:



RESULT:

Thus a program to develop a program for Spam mail detection has been executed successfully.