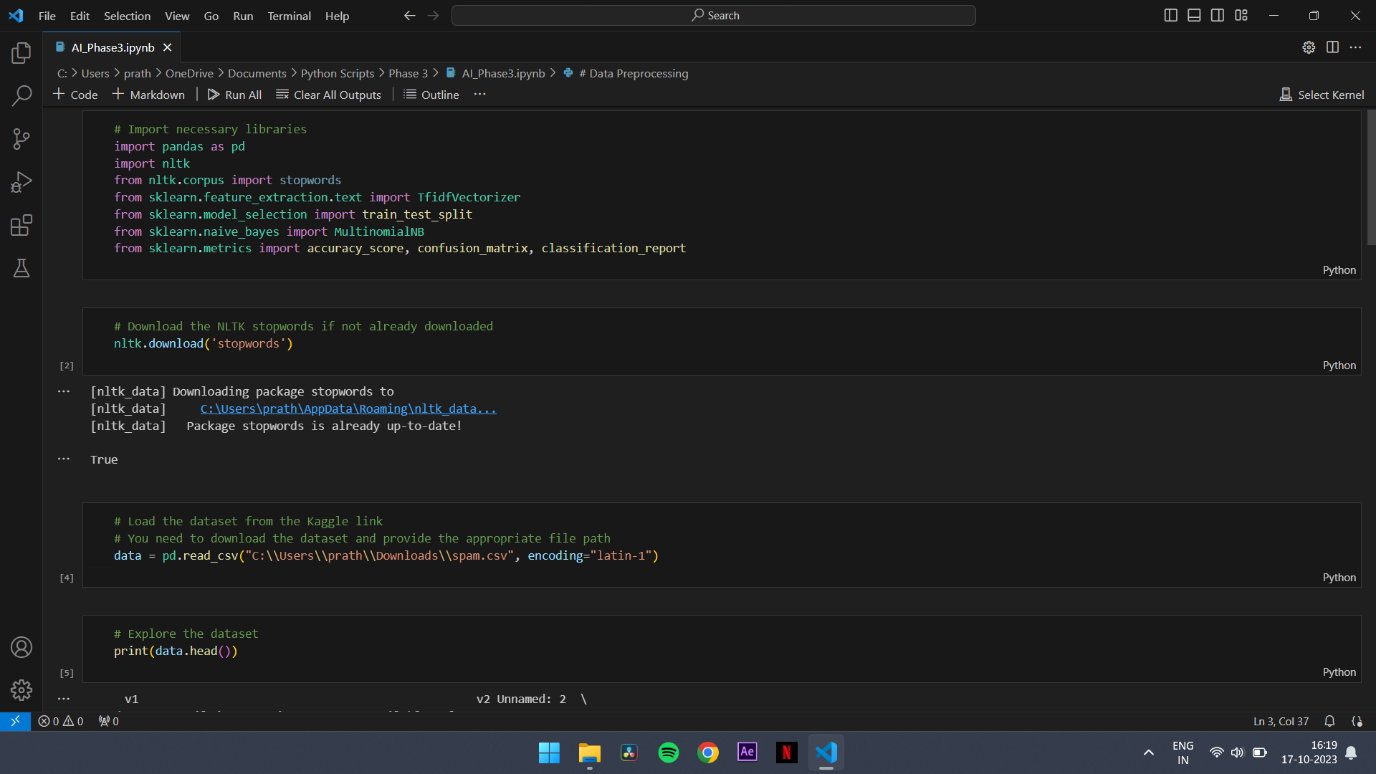
**# Spam Classification using Multinomial Naive Bayes**

This Python script performs spam classification using the Multinomial Naive Bayes algorithm. It utilizes the NLTK library for text preprocessing and the scikit-learn library for machine learning.

**## Prerequisites**

Before running the code, make sure you have the following libraries installed:

- pandas

- nltk

- scikit-learn

You can install these libraries using `pip`:

```bash

pip install pandas nltk scikit-learn

```

Additionally, you'll need to download the NLTK stopwords dataset. You can do this by running the following code:

```python

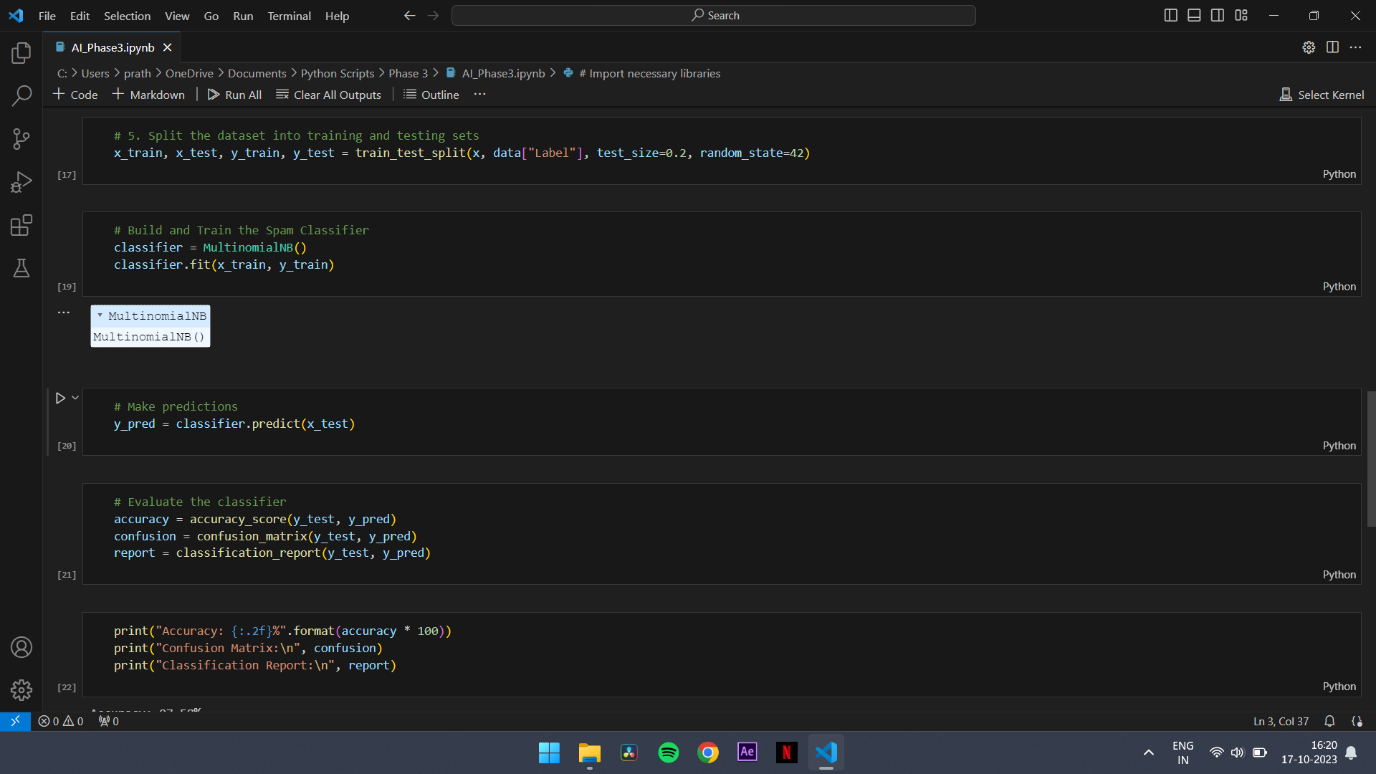
nltk.download('stopwords')

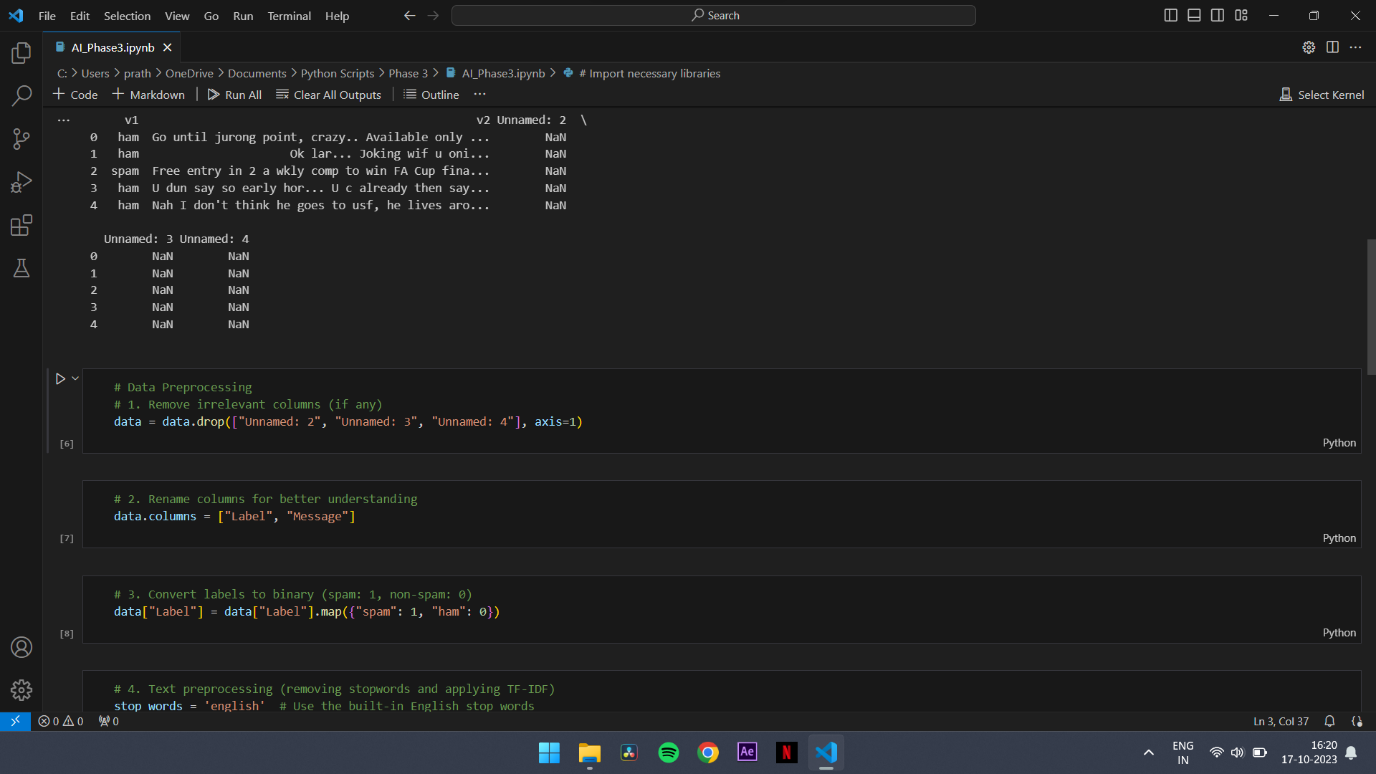
```

**## Dataset**

The script loads the spam dataset from a CSV file. The dataset consists of the following columns:

v1: Label (spam or ham)

v2: Message content 



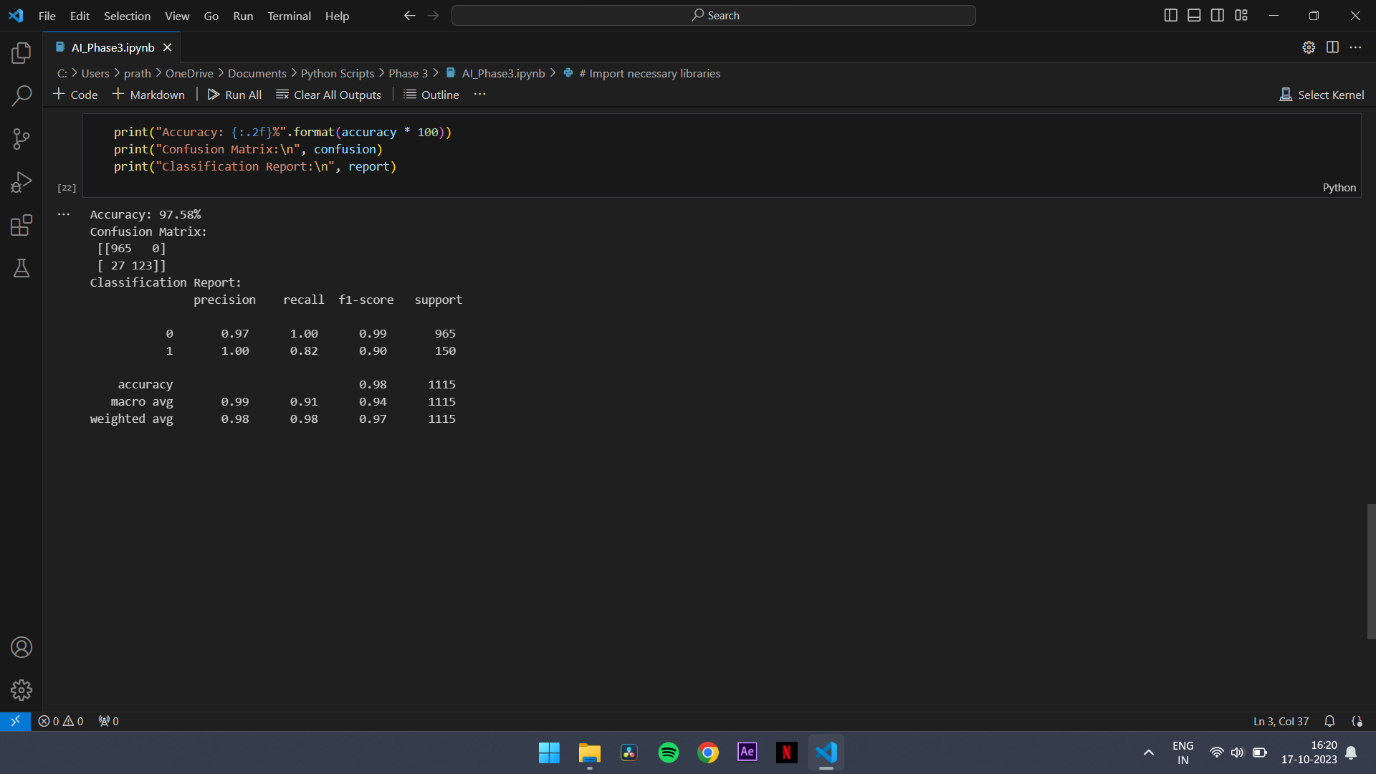
**## Data Preprocessing**

The following data preprocessing steps are performed:

- Irrelevant columns are removed: Unnamed: 2, Unnamed: 3, and Unnamed: 4.

- Column names are renamed for better understanding: Label, Message.

- Labels are converted to binary format (spam: 1, non-spam: 0).

- Text data is preprocessed by removing stopwords and applying TF-IDF vectorization.

**## Machine Learning Model**

The script uses a Multinomial Naive Bayes classifier to classify messages as spam or non-spam. The accuracy of the model is evaluated.

**## Results**

After training and testing the model, the following results are obtained:

Accuracy: 97.58%

Confusion Matrix:

```lua

[[965   0]

 [ 27 123]]

```

Classification Report:

```lua

Copy code

                   precision    recall  f1-score   support

           0       0.97      1.00      0.99       965

           1       1.00      0.82      0.90       150

  accuracy                           0.98      1115

 macro avg       0.99      0.91      0.94      1115

weighted avg     0.98      0.98      0.97      1115

```

**## Usage**

You can run the code by executing the provided Python script. Make sure to replace the dataset file path with the actual file path to your dataset.

```bash

python spam\_classification.py

```