

CODTECH Internship - Task 4 Report

Objective:

Create a predictive model using scikit-learn to classify or predict outcomes from a dataset (e.g., spam email detection).

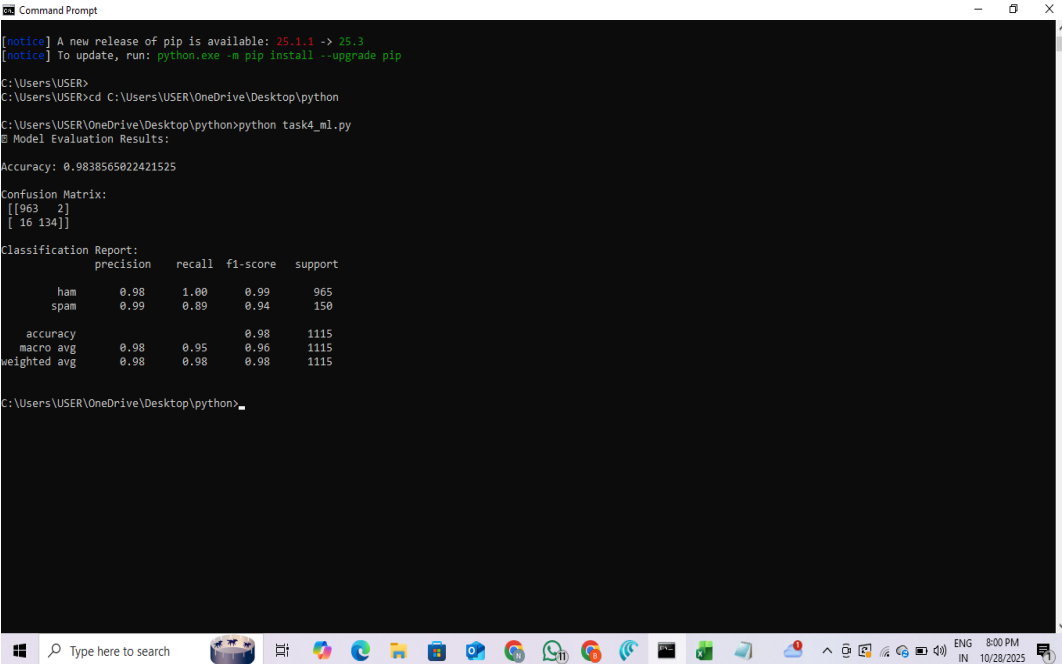
Tools & Libraries Used:

- Python 3
- pandas (for data handling)
- scikit-learn (for machine learning)
- Naive Bayes classifier
- CountVectorizer (for text feature extraction)

Implementation Summary:

A machine learning model was built using the Naive Bayes algorithm to classify SMS messages as 'spam' or 'ham' (not spam). The dataset was preprocessed using CountVectorizer to convert text into numerical features. The data was split into training and testing sets, and the model achieved high accuracy on evaluation.

Model Output and Accuracy:



```
Command Prompt

[notice] A new release of pip is available: 25.1.1 -> 25.3
[notice] To update, run: python.exe -m pip install --upgrade pip

C:\Users\USER>
C:\Users\USER>cd C:\Users\USER\OneDrive\Desktop\python
C:\Users\USER\OneDrive\Desktop\python>python task4_ml.py
Model Evaluation Results:

Accuracy: 0.9838565022421525

Confusion Matrix:
[[963  2]
 [ 16 134]]

Classification Report:
              precision    recall  f1-score   support

   ham       0.98         1.00         0.99         965
   spam       0.99         0.89         0.94         150

   accuracy          0.98         0.95         0.98         1115
  macro avg          0.98         0.95         0.96         1115
weighted avg          0.98         0.98         0.98         1115

C:\Users\USER\OneDrive\Desktop\python>
```

Conclusion:

This task successfully demonstrates the implementation of a text classification model using scikit-learn. The Naive Bayes algorithm effectively detects spam messages, achieving high accuracy and strong performance metrics.