## **Title : Email Spam Detection using Logistic Regression**

**Goal/Objective:** Classifying the incoming emails as spam or not spam.

**Task Completion Steps:**

**Problem Definition**

Classify the incoming emails as spam or not spam.

**Data Collection**

Use a public dataset containing labeled email messages.

**Data Preprocessing**

Clean the data: removing the noise, lowercasing, strip HTML, etc.

**Feature Extraction**

Converting text into numerical features using TF-IDF.

**Model Selection and Training**

I used the Logistic Regression to train the model.

**Train-Test Split**

Dividing the data into training and testing sets.

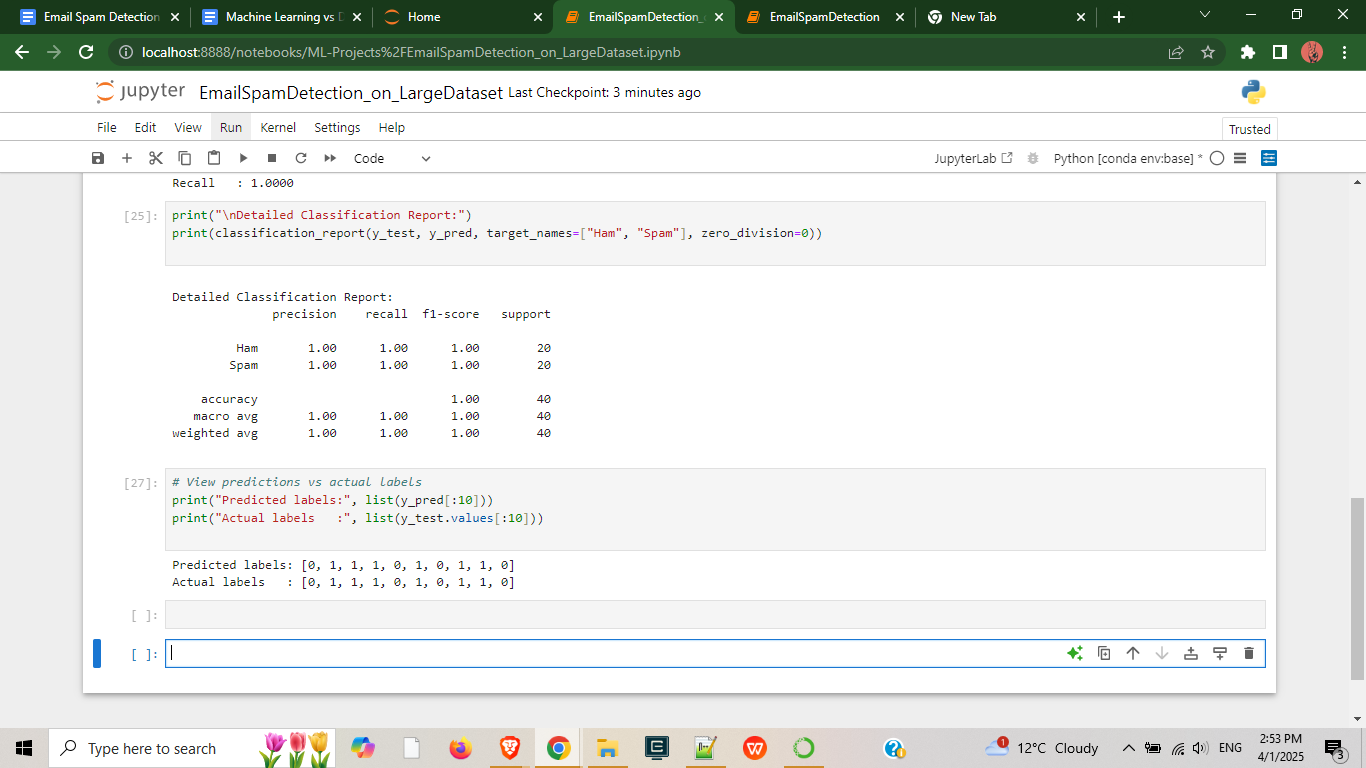
**Model Evaluation**

I used the metrics like accuracy, precision, recall to evaluate the performance.

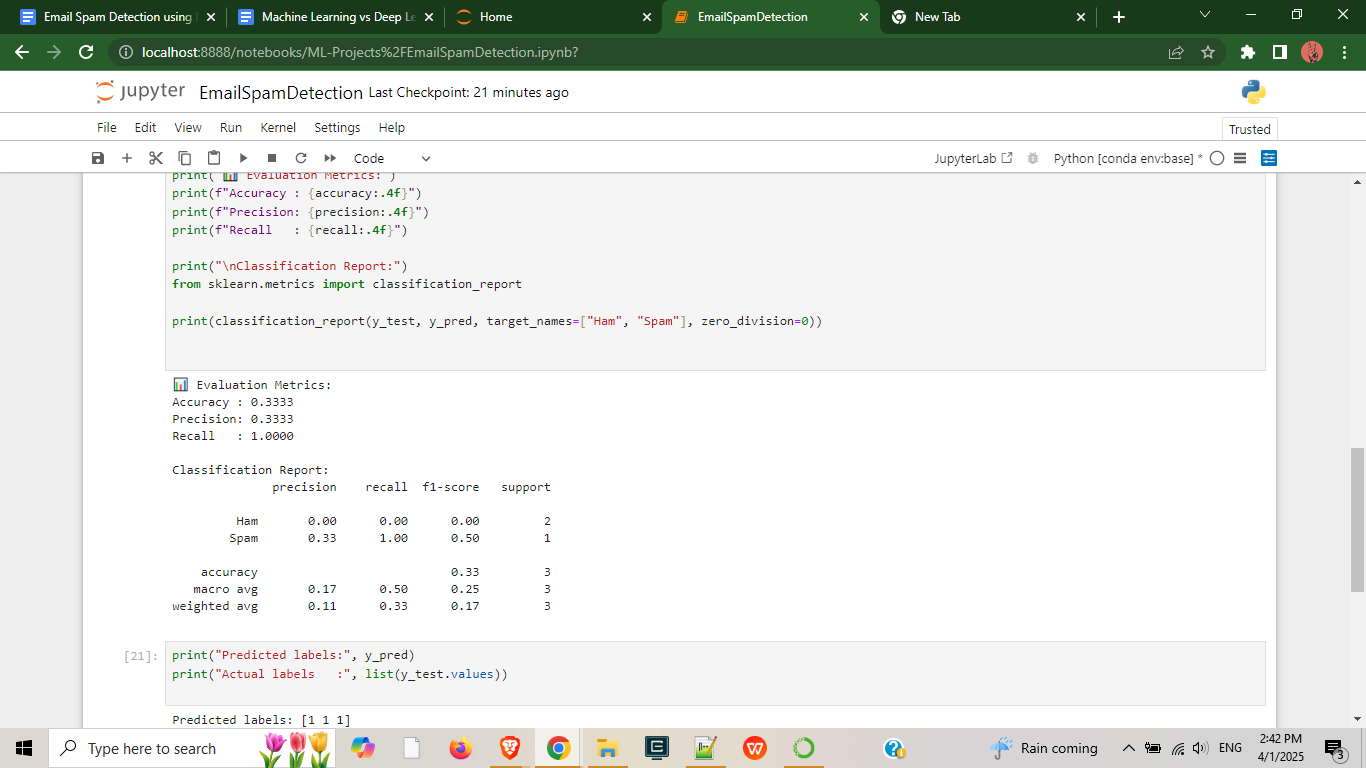
**Results Interpretation**

I ran the model on two datasets(Large dataset and small dataset)

Results when using the Large Dataset



Result when using the smaller dataset



**Tools/resources Used:**

For this project I have used mainly Jupyter notebook,publicly available datasets,Logistic Regression model

**Value Proposition (Small Dataset):** A lightweight spam detection model ideal for the quick demonstrations, showcasing how the machine learning can classify emails even with the minimal data.

**Value Proposition (Large Dataset):** A robust, high accuracy spam detection system trained on realistic and balanced data, offering scalable and reliable performance for the real world email filtering solutions.