# Project 4: Spam Filtering

1. **What is a spam filtering algorithm?**

   Designed to automatically identify and remove unwanted email messages, known as spam, from a user's inbox. The algorithm typically uses various techniques to analyze incoming emails and determine whether they are ham or spam.

2. **Performance of the model in terms of different evaluation metrics such as confusion matrix, precision, recall, and F1-score.**

   - Confusion matrix:

     P: Actual Spam

     N: Actual Non-Spam(ham)

     TP: correctly classified as spam.

     FP: incorrectly classified as spam. (non-spam that was classified as spam)

     TN: correctly classified as non-spam. (ham)

     FN: incorrectly classified as non-spam (spam that was classified as non-spam)

```
[[1450    3]
 [  27  192]]
              precision    recall  f1-score   support

         ham       0.98      1.00      0.99      1453
        spam       0.98      0.88      0.93       219

    accuracy                           0.98      1672
   macro avg       0.98      0.94      0.96      1672
weighted avg       0.98      0.98      0.98      1672
```

3. **Explain how you can improve the performance of the algorithm.**

   **Tuning this hyperparameter can improve the model's accuracy by reducing overfitting and increasing its ability to generalize to new data.**