

**Faculty of Engineering & Technology**

**Electrical & Computer Engineering Department**

**ARTIFICIAL INTELLIGENCE- ENCS3340**

**Project # 2 Report**

**Prepared by:**

Musab Masalmah: 1200078

Wasim Atta: 1200920

**Instructors:** Dr. Yazan Abu Farha

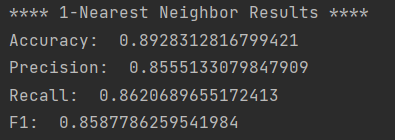
**Sections:** 2

**Date:** 11/7/2023

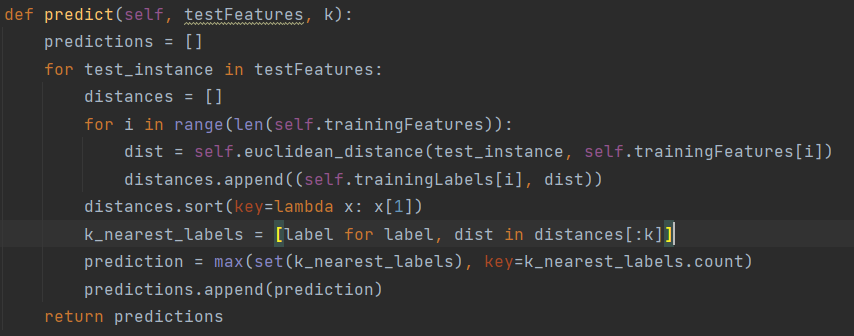
Results and Analysis:

1- k-NN Classifier Performance:

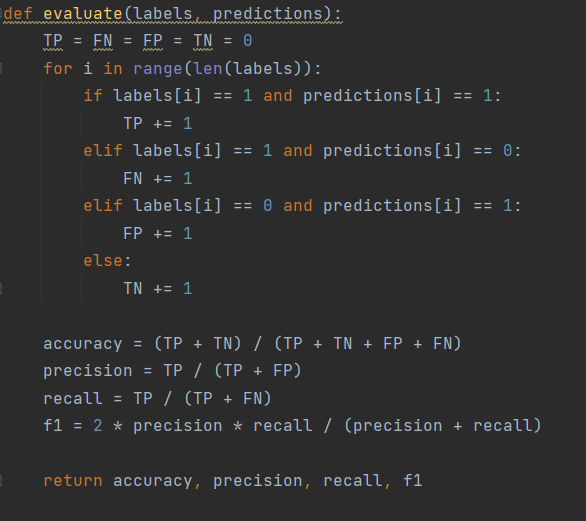
The k-NN classifier achieved the following performance metrics on the test set:



These results indicate that the classifier performs well in distinguishing between spam and non-spam emails, with a high proportion of correctly classified instances. Further analysis and experimentation can provide a more comprehensive evaluation of its performance.



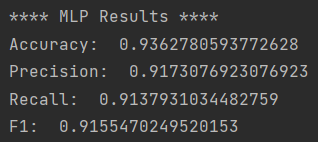
This method calculates the distances between each test instance and all training instances using Euclidean distance. It sorts the distances and selects the **k** nearest neighbors. The method determines the majority label among the **k** nearest neighbors and assigns it as the predicted label for the test instance. Finally, it returns the list of predicted labels for all test instances.



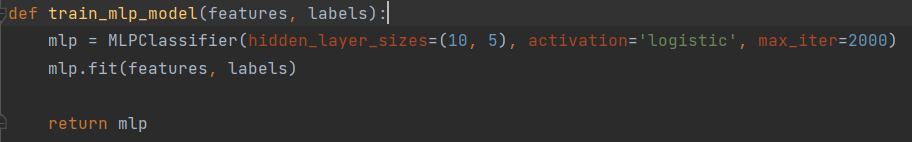
And this simple method we use it to calculate the accuracy, precision, recall and F1 , and we get the equations from the course slides.

2- MLP Performance:

The MLP classifier achieved the following performance metrics on the test set:



The MLP classifier achieved an accuracy of 0.9363, indicating a high proportion of correctly classified instances. The precision of 0.9173 highlights the classifier's ability to accurately identify spam emails. With a recall of 0.9138, the MLP classifier effectively captures a large portion of actual spam instances. The F1-score of 0.9155 represents a balanced measure of precision and recall, indicating overall strong performance in spam classification.



This function trains an MLP classifier with two hidden layers, the first layer containing 10 neurons and the second layer containing 5 neurons. It uses the logistic activation function and a maximum of 2000 iterations for training. The trained MLP classifier is returned as the output of the function.

-The Predict and The Evaluate function we use in the MLP is the same one as the above in the K-NN.

The Confusion Matrix:

1- k-NN Classifier :

This is the confusion matrix for k-NN Classifier:

|  | **Predicted Spam** | **Predicted Non-Spam** |
| --- | --- | --- |
| Actual Spam | 3965 | 551 |
| Actual Non-Spam | 624 | 461 |

2- MLP Classifier :

This is the confusion matrix for MLP Classifier:

|  | **Predicted Spam** | **Predicted Non-Spam** |
| --- | --- | --- |
| Actual Spam | 1257 | 87 |
| Actual Non-Spam | 84 | 1173 |

Suggested ways to improve the models:

1-Improve the performance of k-NN:

Experiment with different values of k to find the optimal number of neighbors for the given task, Implement feature scaling or normalization to ensure that all features have equal importance in distance calculations.

2-Improve the performance of MLP:

Increase the complexity of the model by adding more hidden layers or increasing the number of neurons in each layer, Implement regularization techniques, such as L1 or L2 regularization, to prevent over fitting and improve generalization.

Conclusion:

The project involved testing the performance of the k-NN and MLP classifiers on a dataset of email examples, Overall, both classifiers exhibited good performance in distinguishing between spam and non-spam emails. The k-NN classifier showcased simplicity and competitive results, while the MLP classifier demonstrated stronger performance with its ability to capture complex patterns.