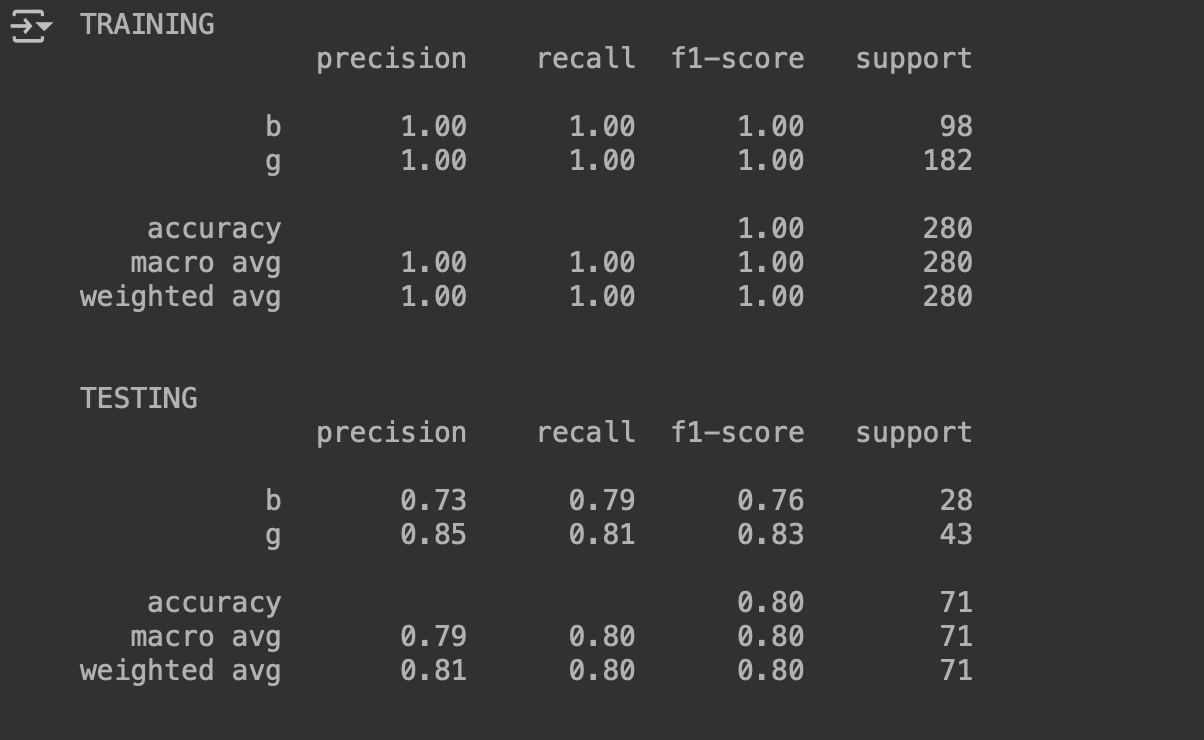
**MACHINE LEARNING FROM DATA**

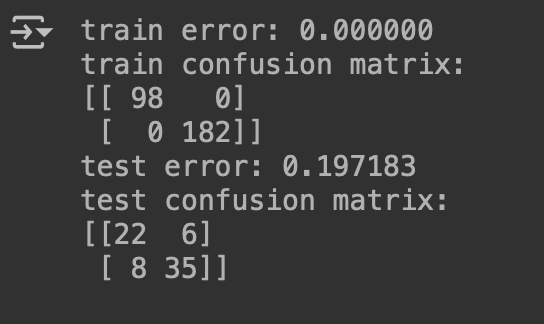
**Report: Lab Session 8 – Decision Trees and Random Forests**

**Questions**

Default parameters:

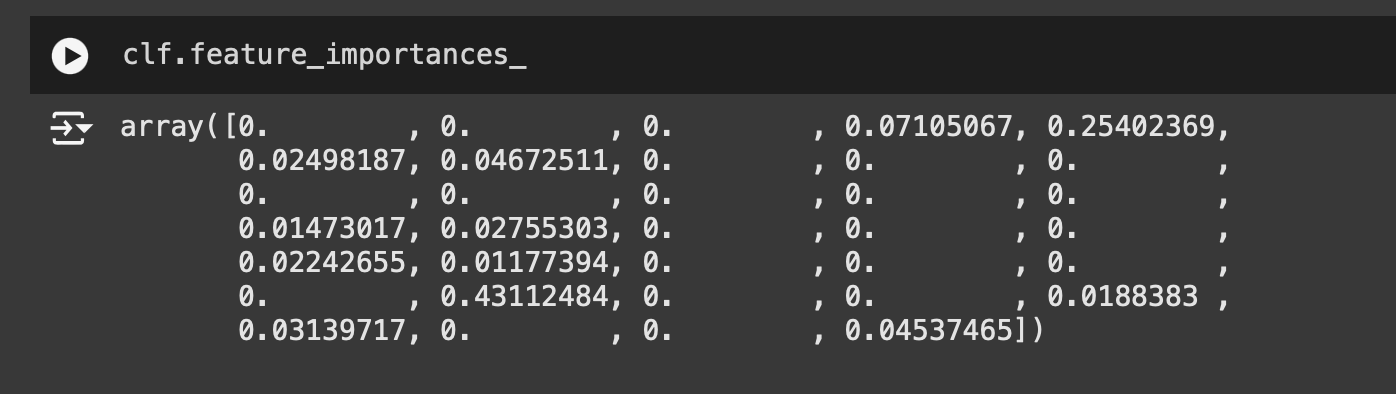
Q1: For the tree trained with the default parameters, copy the training, and test classification errors and the confusion matrices.





Feature importances:

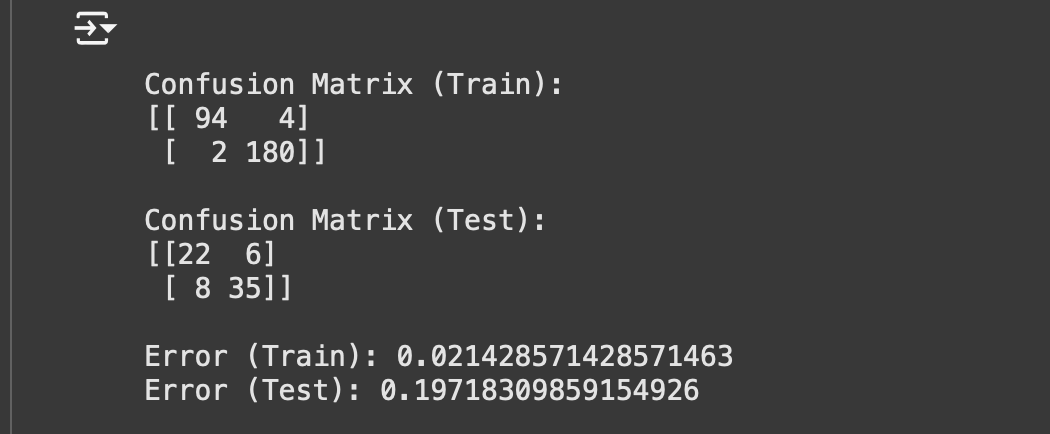
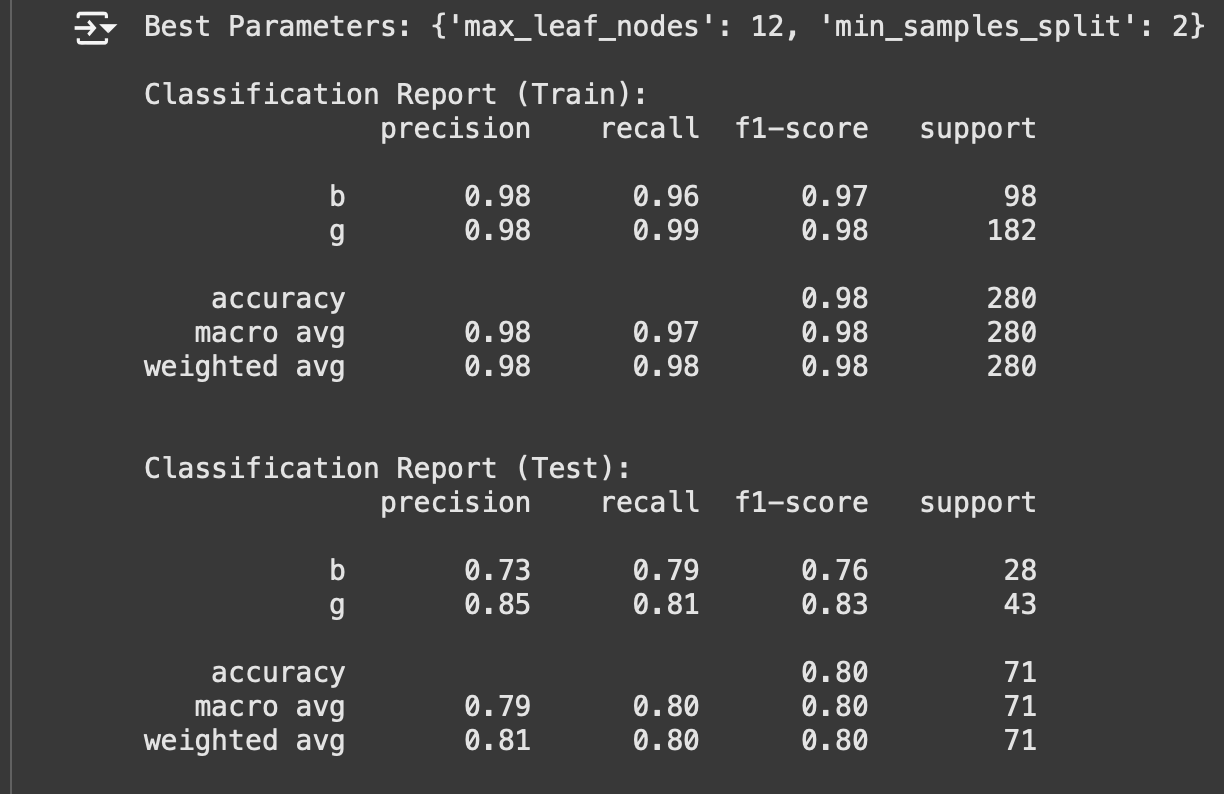
Q2: Visualize the tree and analyze the questions at each node. Compute the feature-importances. Which are the most relevant features for the classification task?



The features with a higher value are the most important features, the ones with value = 0, are not contributing to the classification task.

Regularization:

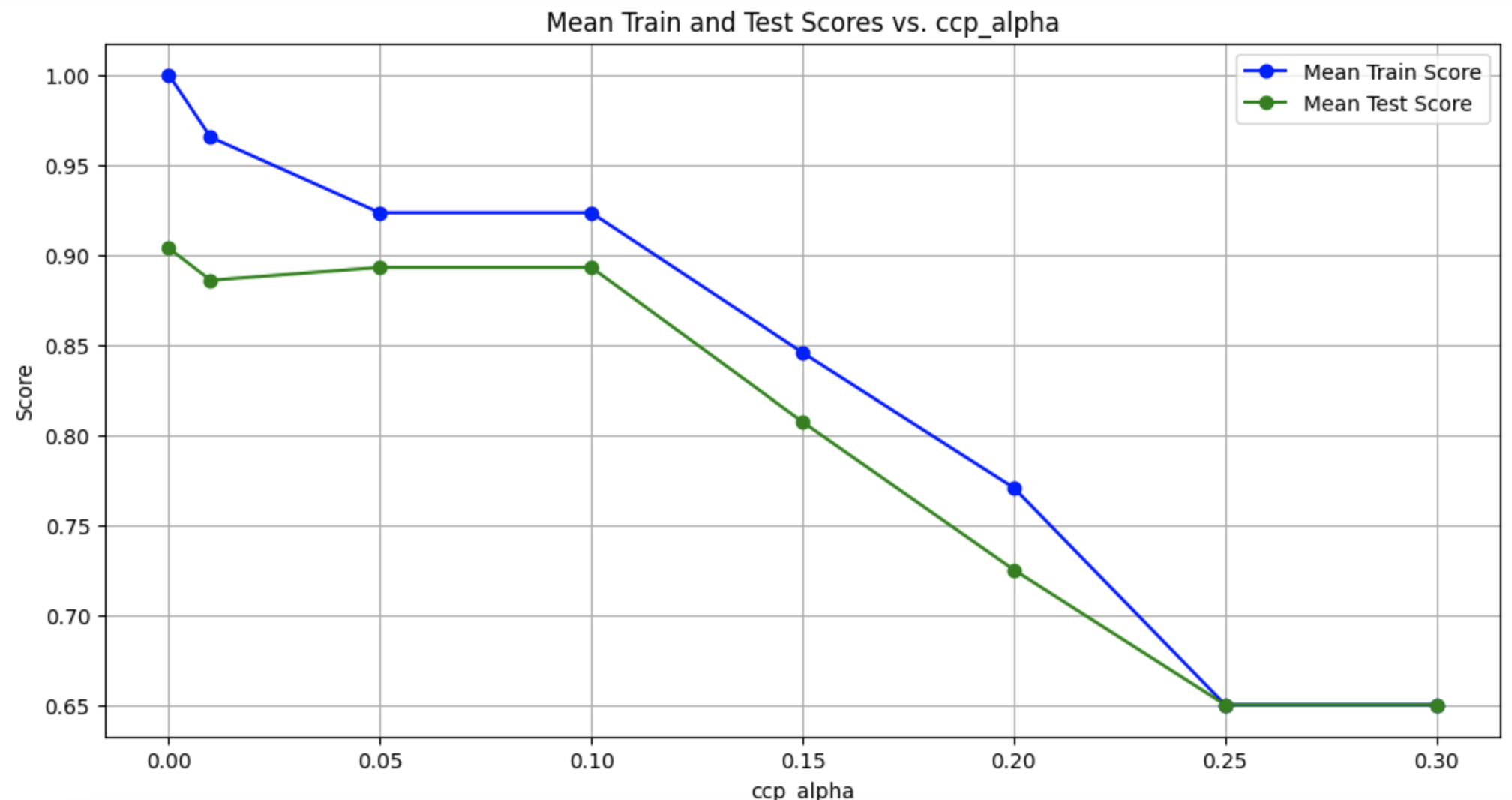
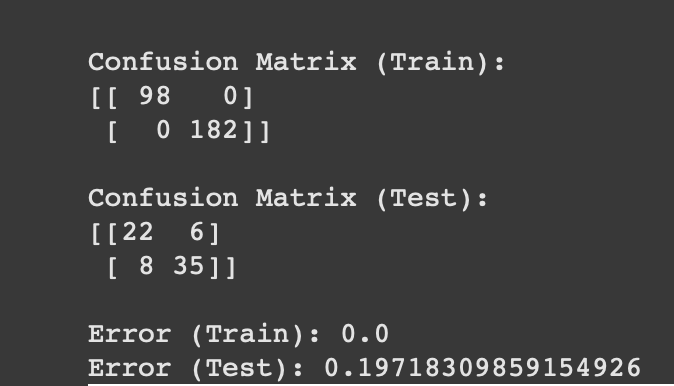
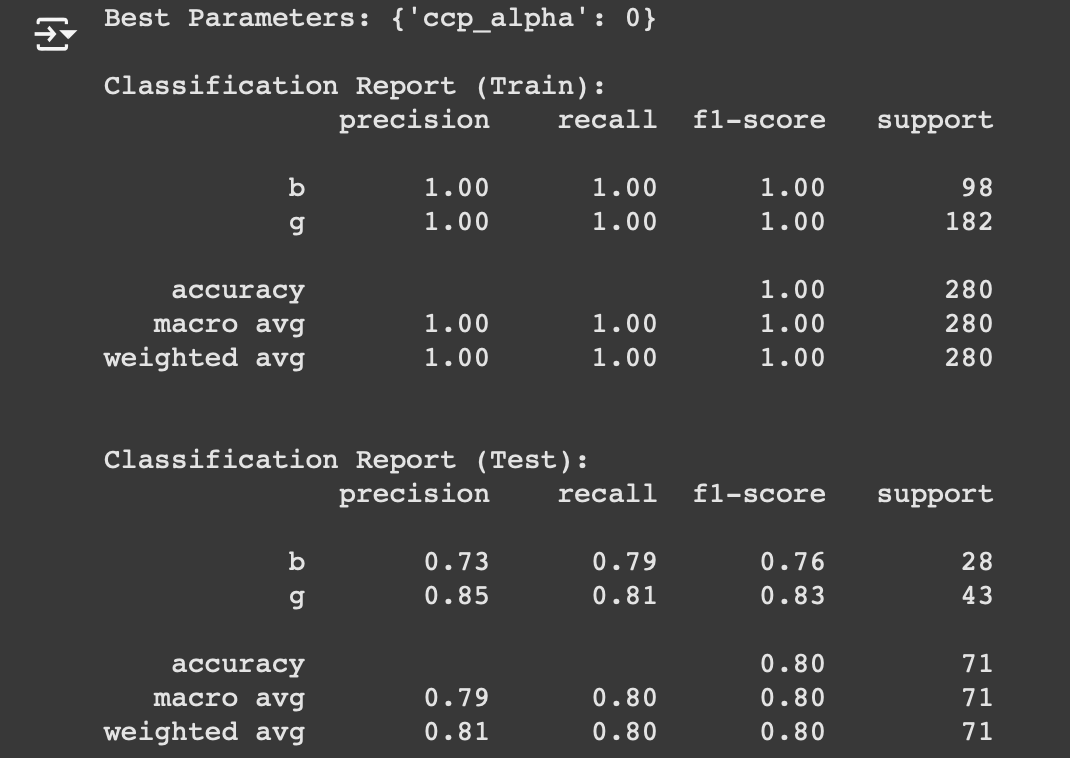
Q3: Copy and analyze the results (errors, confusion matrices and plots) and compare with the results obtained with the default parameters.



For the default parameters we can see the model is overfitting, now using gridSearch to find the best parameters, we can see that the error at the train set is bigger, but the error in the test set is the same as for the default parameters.

Tree pruning:

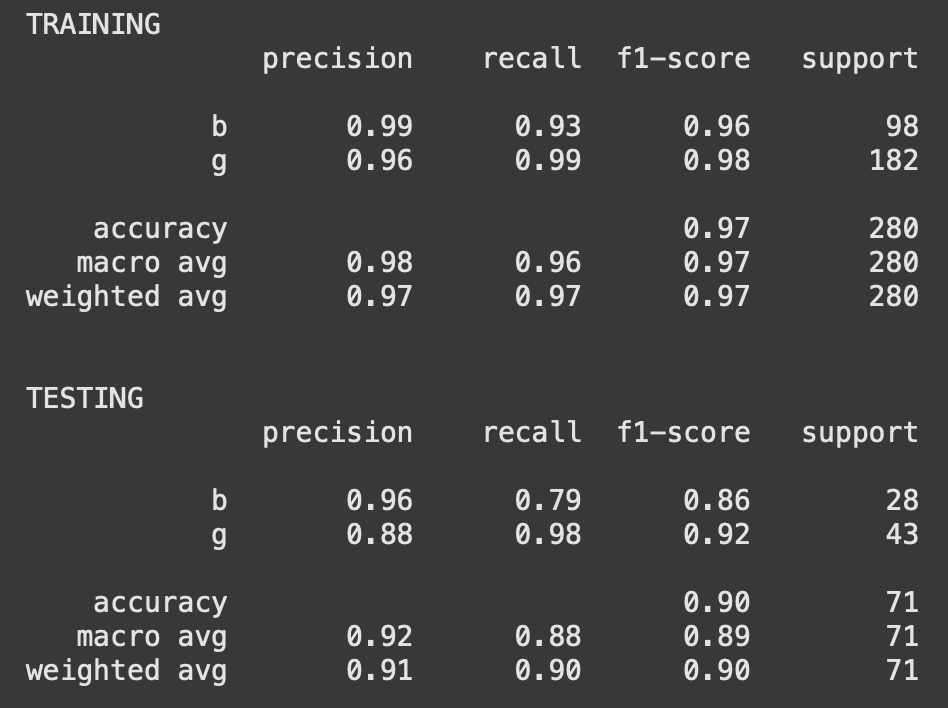
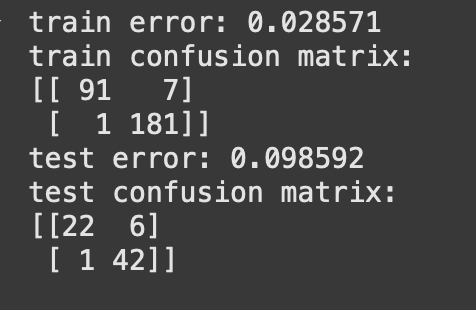
Q4: Copy and analyze the results (errors, confusion matrices and plots) and compare with the results obtained with the previous strategy (regularization).



For this strategy, we can see that we have overfitting another time the train error is 0 again, and the test error is big.

Random Forests:

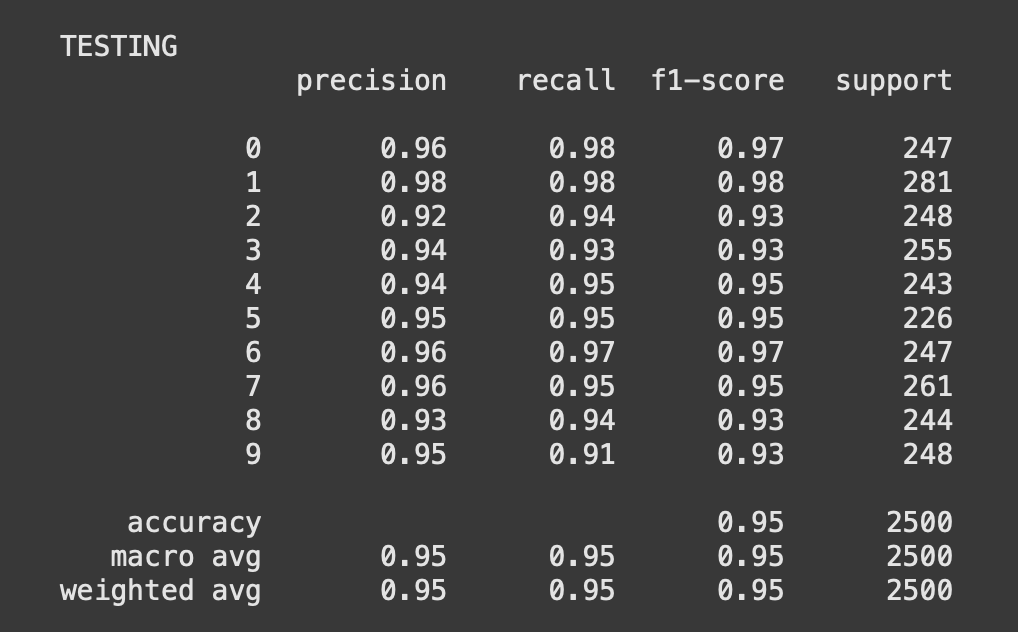
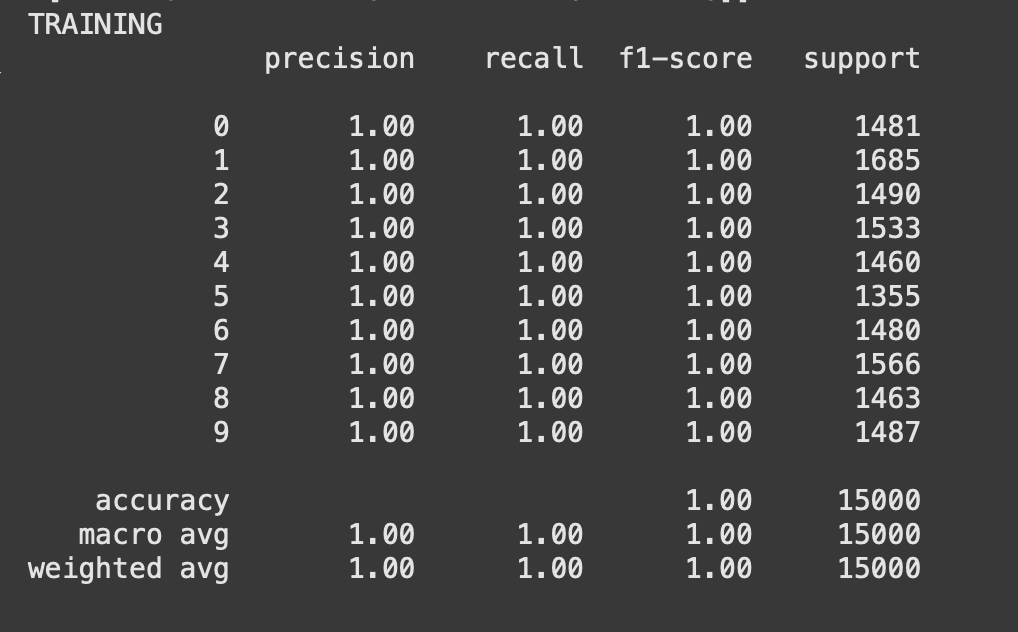
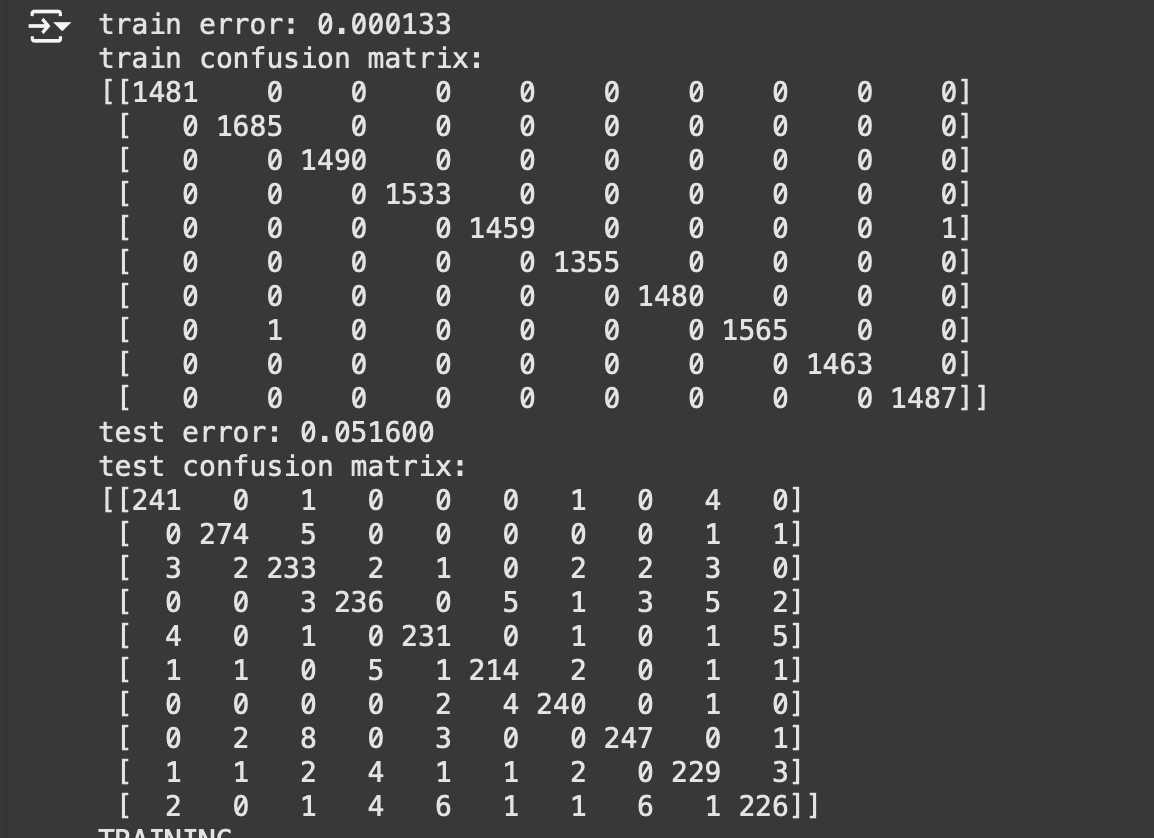
Q5: Copy and analyze the results (errors, confusion matrices and plots) and compare with the results obtained with decision trees.



For this strategy we can see is not overfitting anymore, and we find a very good error at test and train sets.

Random Forests on MNIST:

Q6: Copy and analyze the results (errors, confusion matrices and plots) and compare with the results obtained in Lab7 using Neural Networks.



Comparing with the model in the last lab, we can see that the error for the train and test sets are significantly better, being arround 18% of error for the strategy with NN and 5% now.

So we can say that this model generalizes better than the other one.