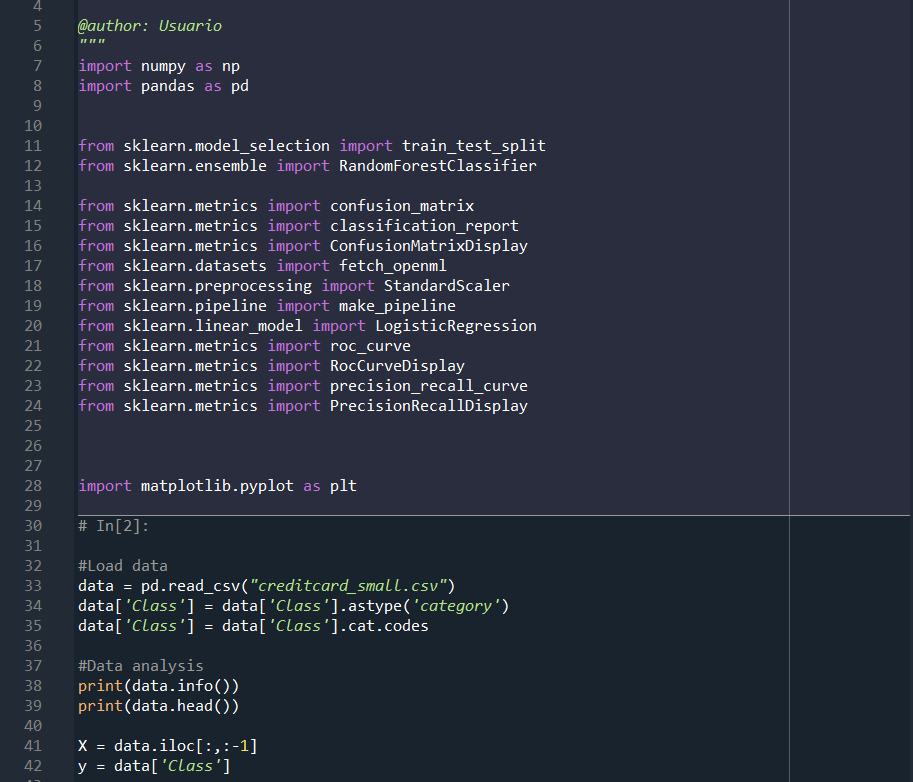
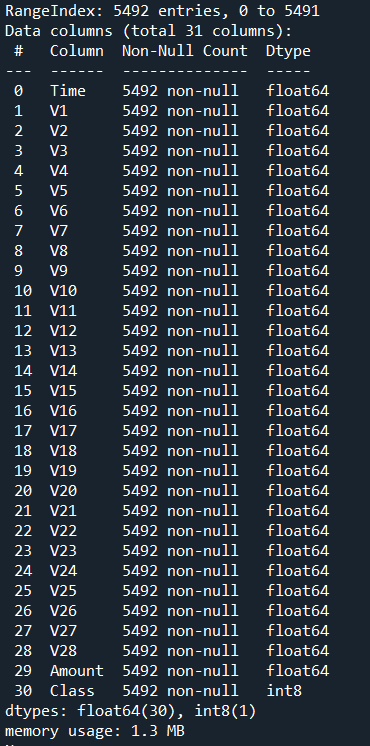
***Week 4: Metrics, Dataset Creditcard\_Small***

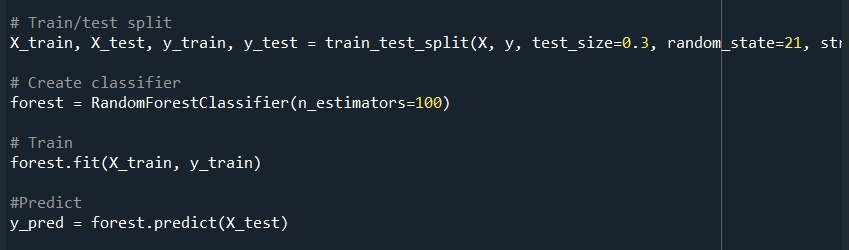
First, we will import the data set required for this exercise.

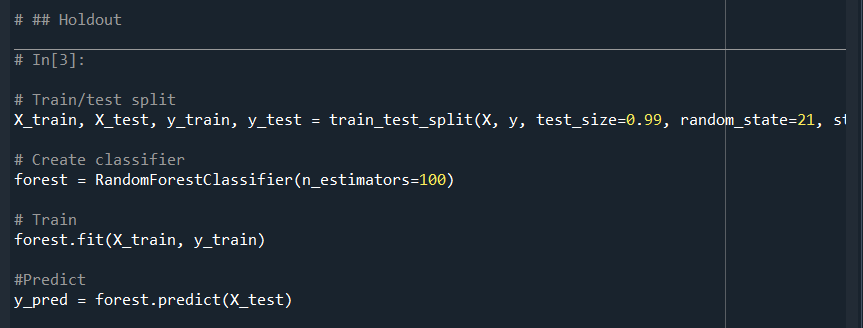


After importing the data, we can run the code to observe what type of information is stored in the dataset Creditcard\_small:



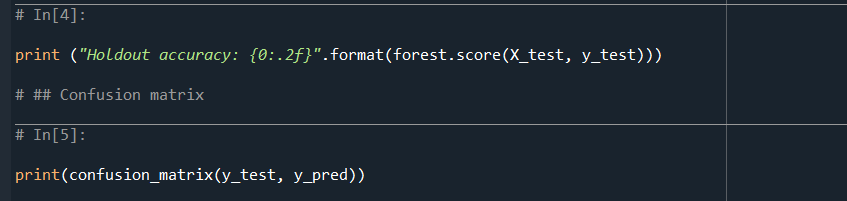
We split, create the classifier, train and predict.



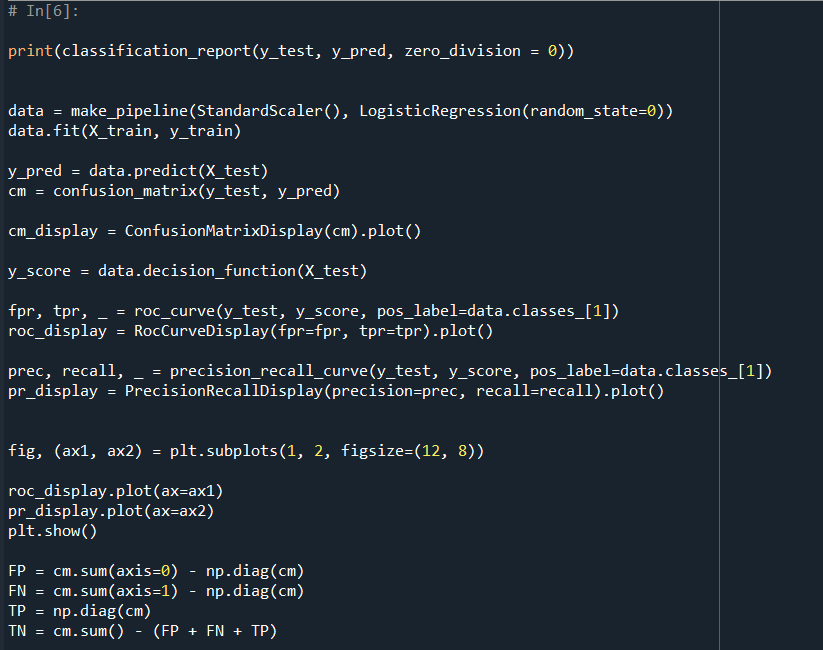


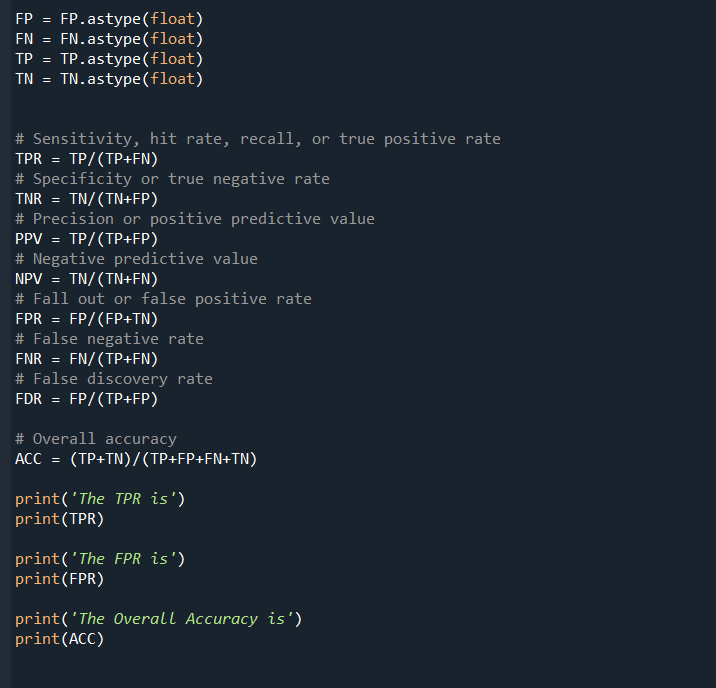
With the test size set to 0.3 the number of results obtained was far from the number of samples we have so we modified the set size just to see if there were any improvements.

Then we will calculate the Holdout accuracy and create the confusion matrix:

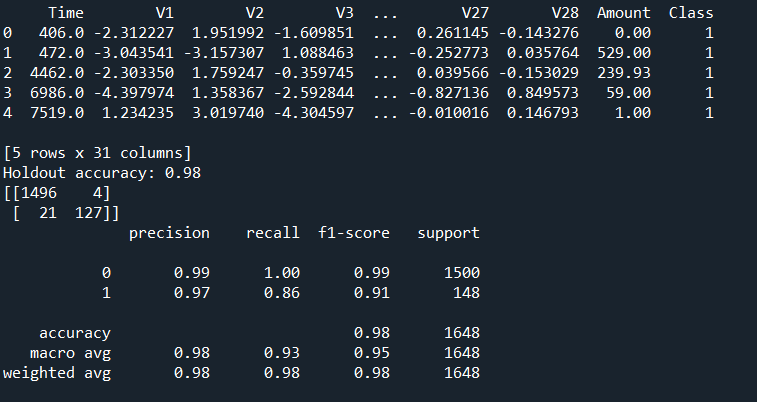


And finally, we plot the confusion matrix, the ROC curve and the TPR, FPR and total accuracy.

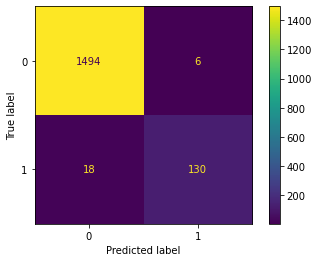




After running the previous code, we obtain the following plots and information printed in the console:

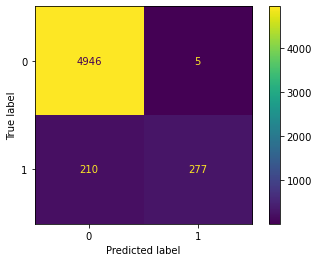


The confusion matrix: (Test size = 0.3)

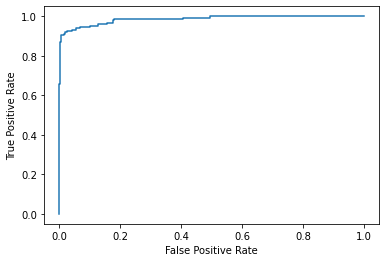


This matrix is quite easy to understand. On the left we have 1494 values that were meant to be 0 and that the model correctly predicted. Beneath we see 18 values that were meant to be 1 and that the model also predicted. The ones on the right, the purple ones, mean that there were six False Positives, which means six ones that were classified as zeros and 130 False Positives, zeros that were classified as ones.

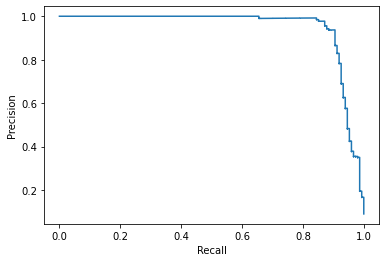
The confusion matrix: (Test size = 0.99)



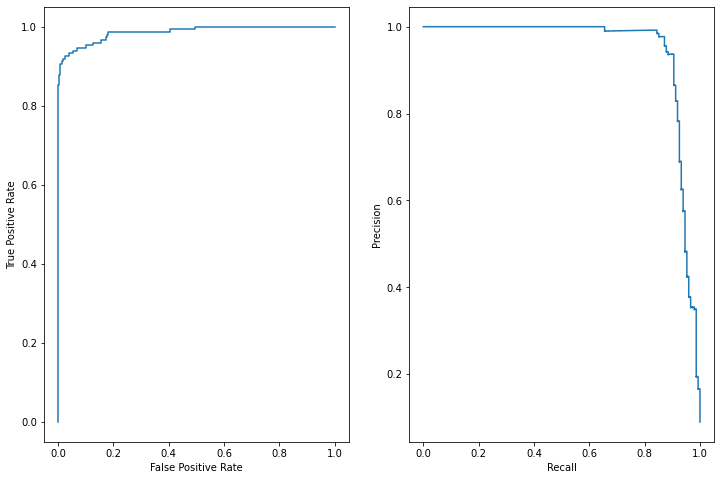
ROC curve: (Test size = 0.3)



Precision and accuracy (Test size = 0.3)



And the previous two plots in the same one:





And with set size set to 0.99:

