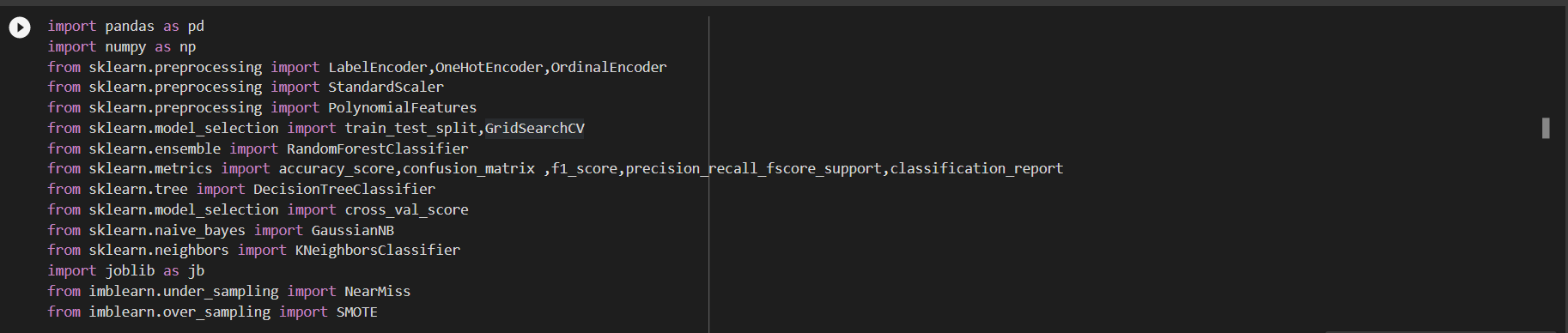
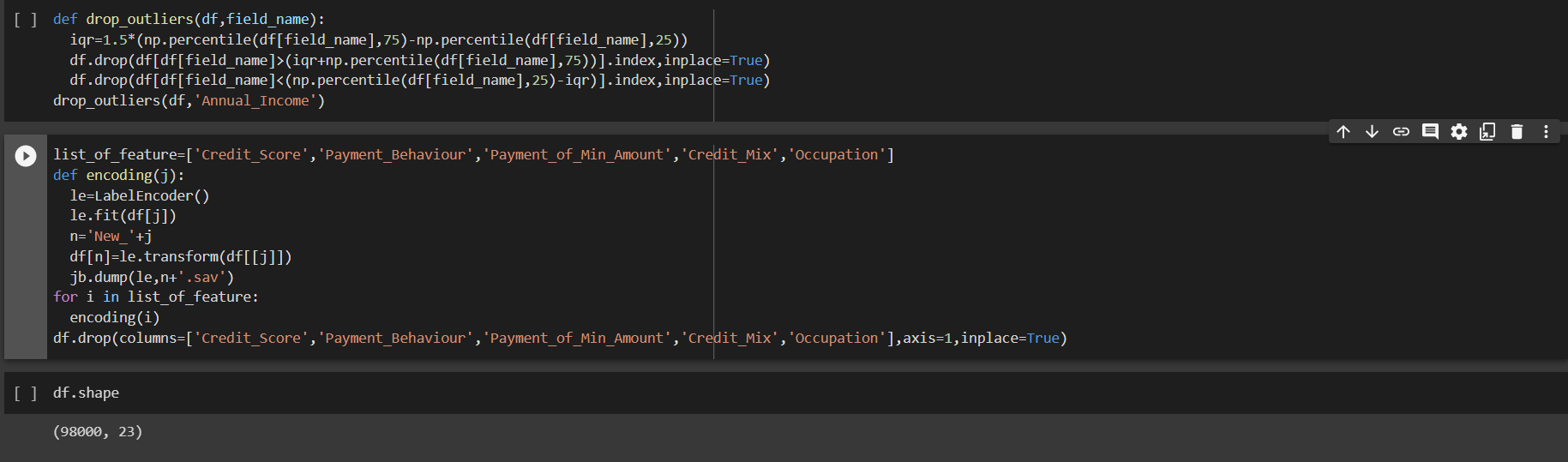
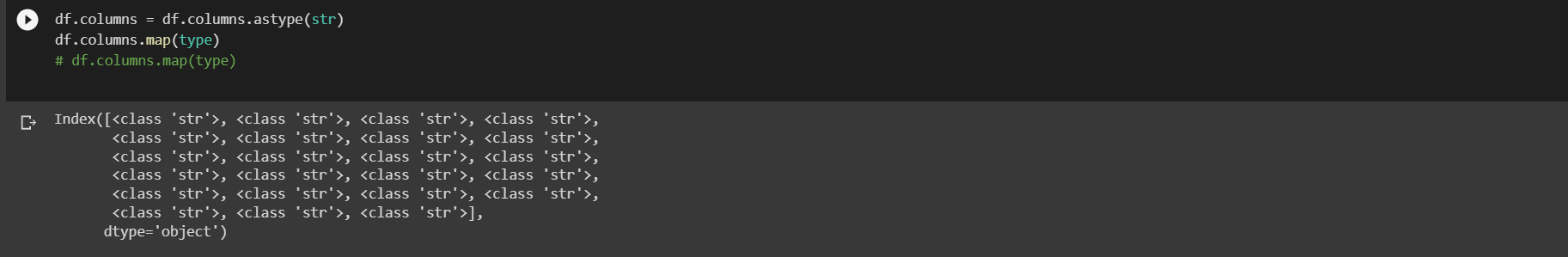
|  |  |  |  |
| --- | --- | --- | --- |
|  |  | |  |
|  |  |  |  |
|  | Documentation of Credit score project  Group 6 | |  |
|  |  |  |  |
|  | | | |
|  |  |  |  |
|  |  |  |  |

* In this documentation we will show briefly what we had worked and some explanation of our steps
* in the next steps we explain the modeling notebooks:

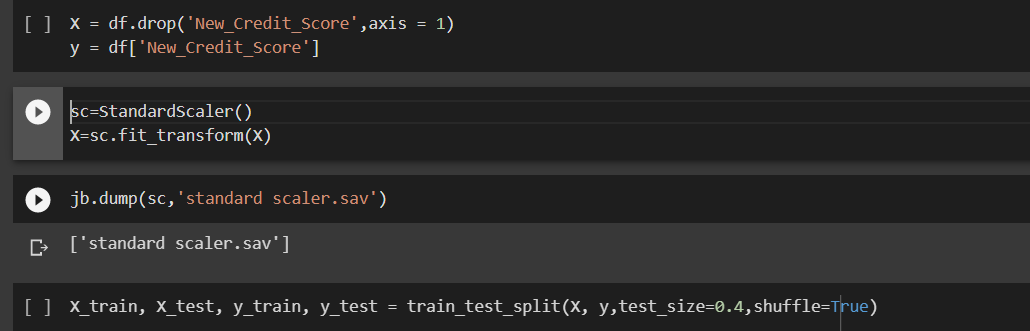
1. Firstly, we import the important libraries مas (numpy, pandas, etc.) that used to deal with the dataset and during the work we import another feature to model and measurements the accuracy and others.



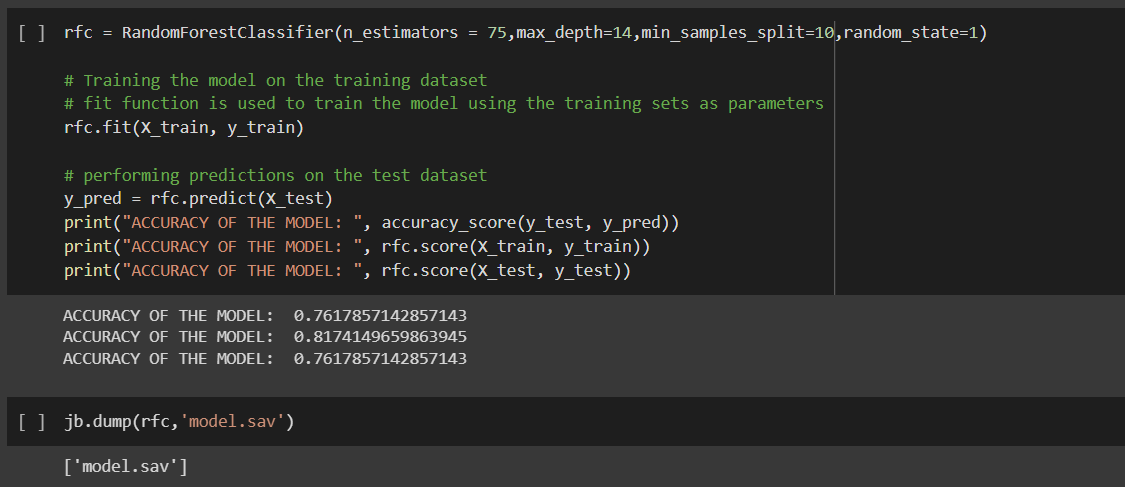
1. Secondly, we import the dataset and checked if there is any missing value and checked the percentage of each element in target value to check if there is imbalance data.
2. Dropped the feature that’s doesn’t give us an information and been unique value in dataset and may make any misleading in our model.
3. we extract the outliers of (annual\_Income) and handling it by dropped it to imporve the fitting of model and give more accuracy.
4. After that the suitable criteria of encoding the categorical feature.
5. And I have used label encoder by function to introduce the feature name and output the new featuer encoded.
6. Then I convert the name of columns has encoded from integer to string.



1. After the preprocessing of dataset I split it to (train / test) then Standardization of a dataset to improve the performance and reduce the time estimated for training.

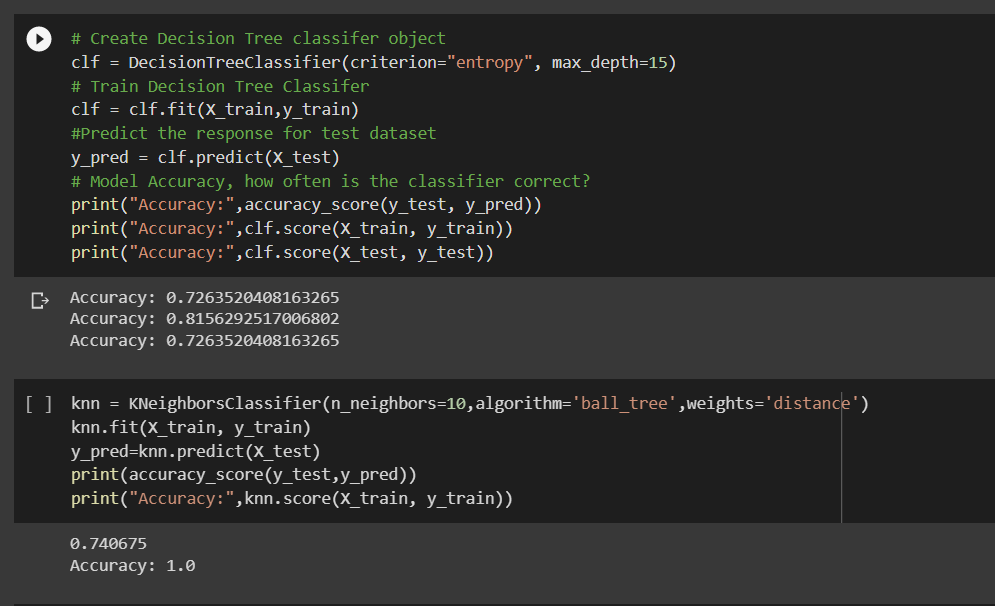


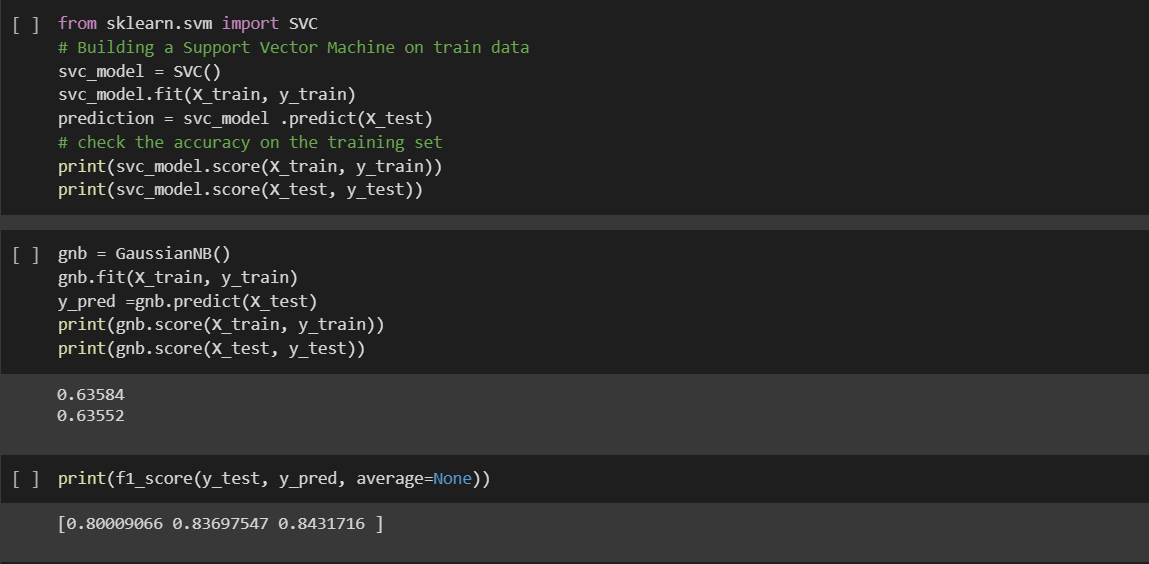
1. Selecting the best model (Random forest classifier ) that give me the best accuracy and precision and f1-score.

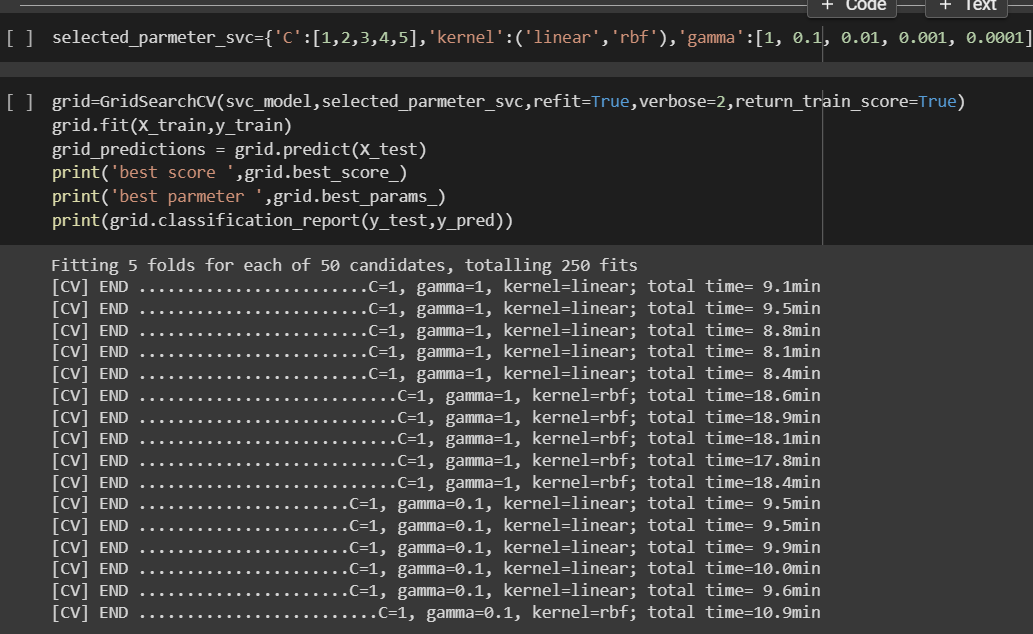


* Accuracy of training data : 76%
* Accuracy of test data : 81%
* We try a lot of class like decsion tree classifier and svc ,GaussianNB , knn

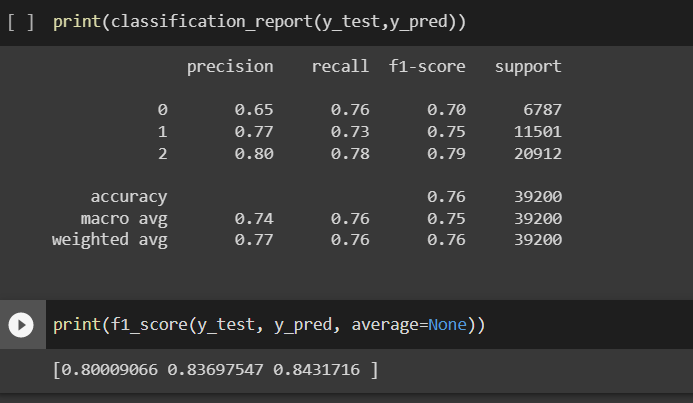
And we use grid search cv to retrive best parameter and best score.



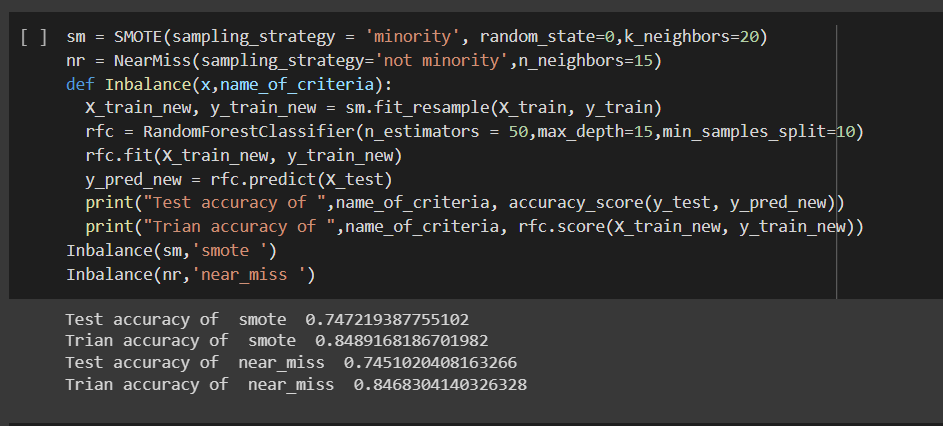




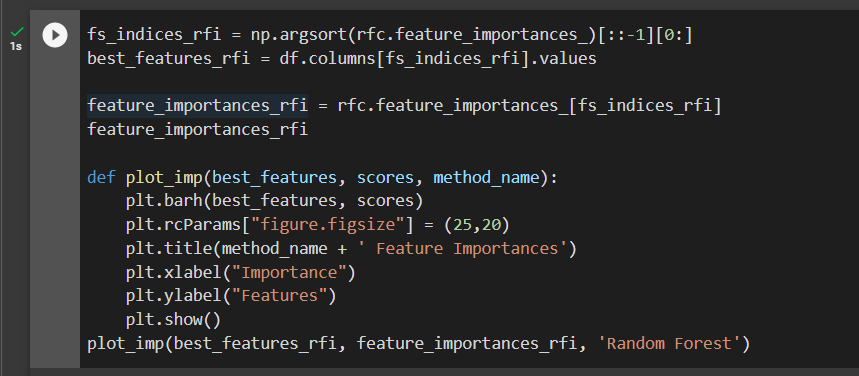
* that’s the important measure we need to judge on our module :

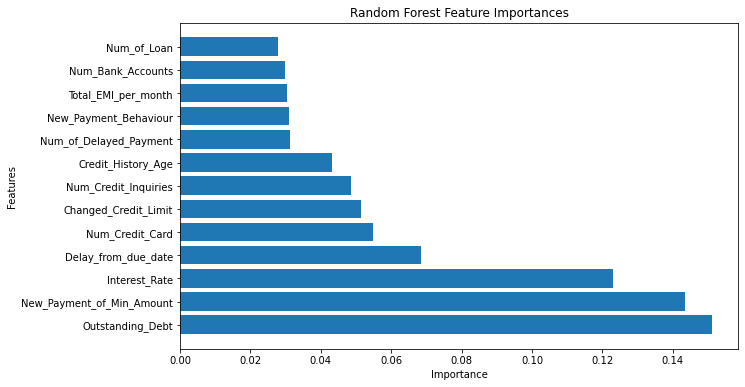


10.Finally we used SMOTE and MISSNEAR to remove the overfitting but it was not efficient result.



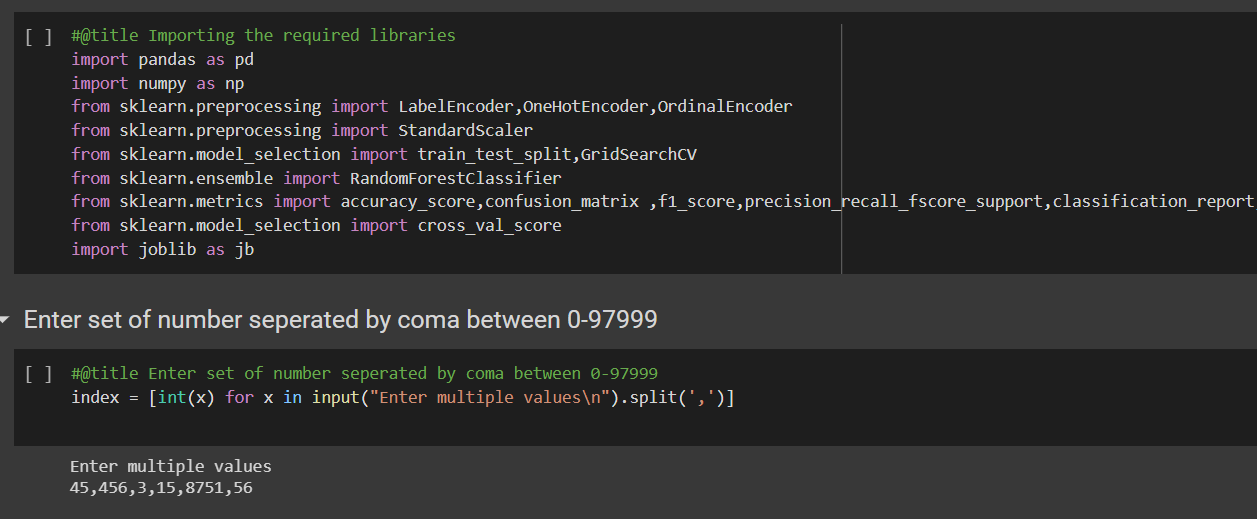
11.And here we extract the most important feature that effect on moduling and visualize it.





Here we will show the steps we following to deploy the model:

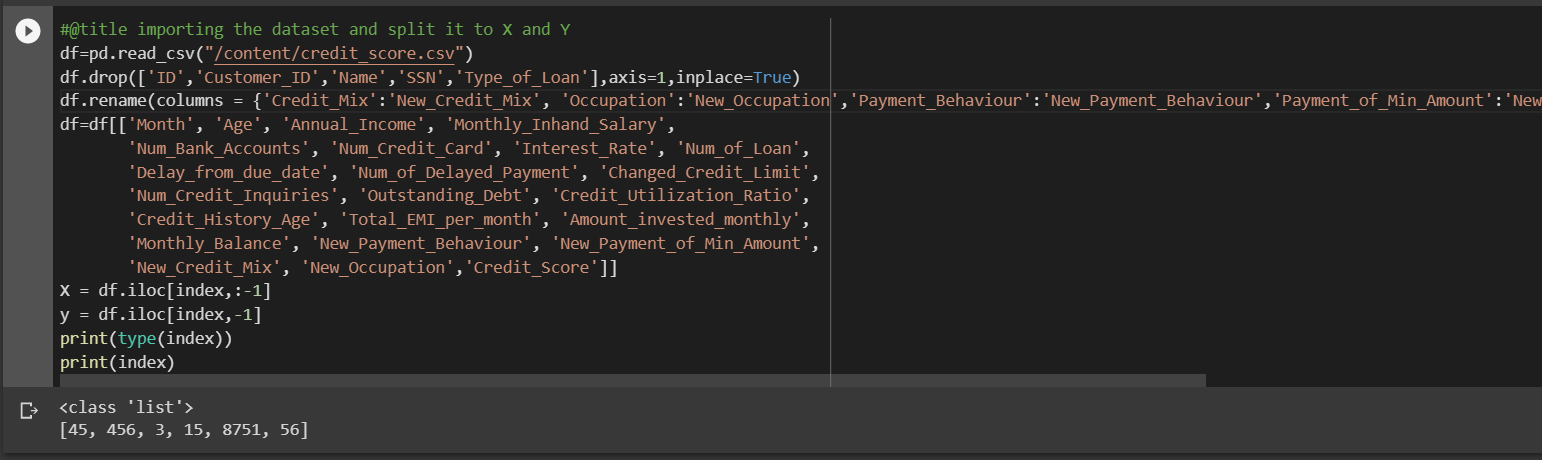
1. Firstly , we import the libraries we needed we created an input list to select row from the dataset .



1. Then we drop the feature we dropped while development

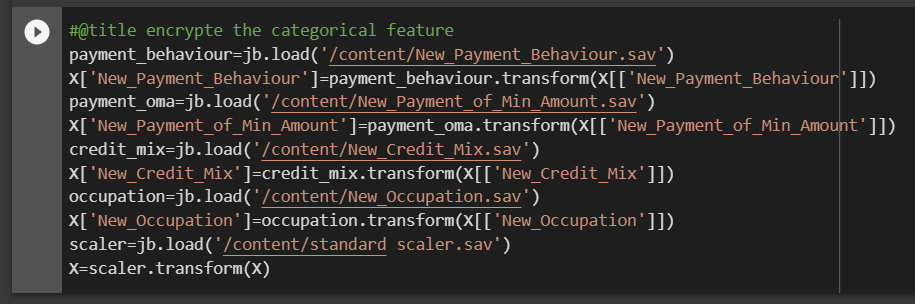
And rename of feature name that encoded in development (because the feature name introduced in moduling should

Be matching in predicted feature ) then split the dataset to X and Y



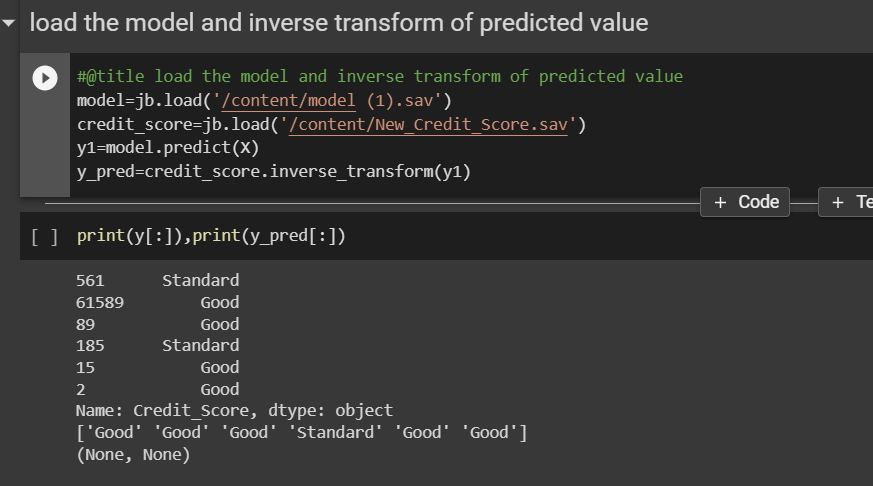
1. Here we load the file of encoding to transform the feature

And standard scaler



1. Here we load the model and make a predict for selected value and make an inverse transform to retrieve

The predicted value (Good\standard\poor) and compare with actual value



This is our work , we hope you like it