FRA ALTERNATE PROJECT

DEEPAK SINGH MANOLA

GREAT LEARNING

15TH JAN 2024

Please find below graded individual assignment for FRA (Extended Project)

You are requested to create an Indian credit risk (default) model, using the data provided in the spreadsheet.

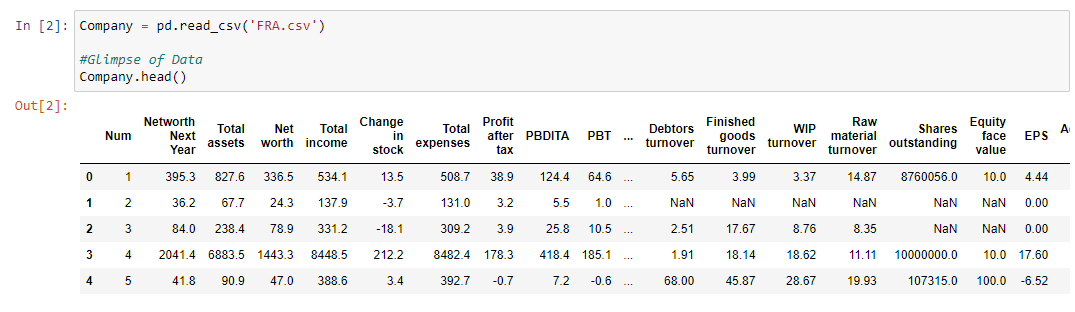
Hints:

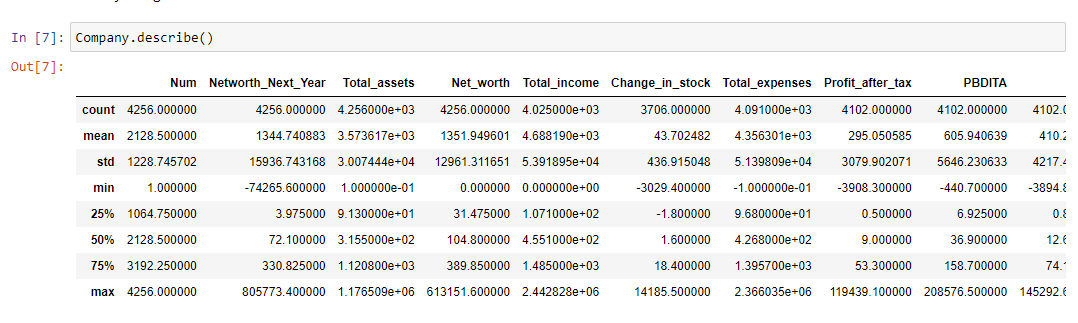
Dependent variable - We need to create a default variable which should take the value of 1 when net worth next year is negative & 0 when net worth next year is positive.

Validation Dataset - We need to build the model on train dataset and check the model performance measures on validation dataset

| FRA Rubric(Extended Project) | | |
| --- | --- | --- |
| **Criteria** | **Ratings** | **Pts** |
| This criterion is linked to a Learning Outcome Outlier Treatment |  | 8.0 pts |
| This criterion is linked to a Learning Outcome Missing Value Treatment |  | 8.0 pts |
| This criterion is linked to a Learning Outcome Check for multicollinearity |  | 8.0 pts |
| This criterion is linked to a Learning Outcome Univariate & bivariate analysis |  | 10.0 pts |
| This criterion is linked to a Learning Outcome Split the data into train and test (70:30) |  | 5.0 pts |
| This criterion is linked to a Learning Outcome Build Logistic Regression Model (using stats model library) on the most important variables on Train Dataset and choose the optimum cutoff. Also, showcase your model-building approach |  | 15.0 pts |
| This criterion is linked to a Learning Outcome Validate the Model on Test Dataset and state the performance matrices. Also, state interpretation from the model |  | 10.0 pts |
| This criterion is linked to a Learning Outcome Build Random Forest Model on Train Data and validated on the Test Data |  | 10.0 pts |
| This criterion is linked to a Learning Outcome Compare the performance of Logistics Regression and Random Forest |  | 7.0 pts |
| This criterion is linked to a Learning Outcome Quality of Business Report |  | 9.0 pts |
| Total Points: 90.0 | | |

1. EDA

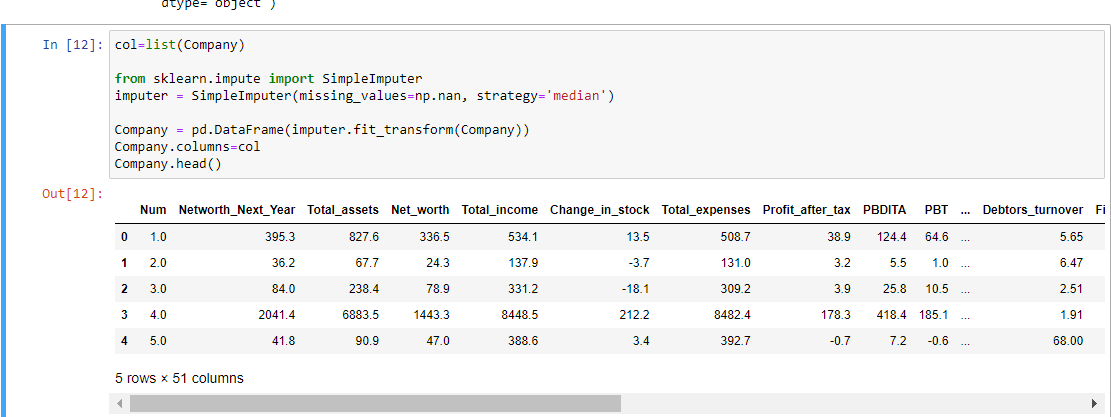




1. This criterion is linked to a Learning Outcome Outlier Treatment

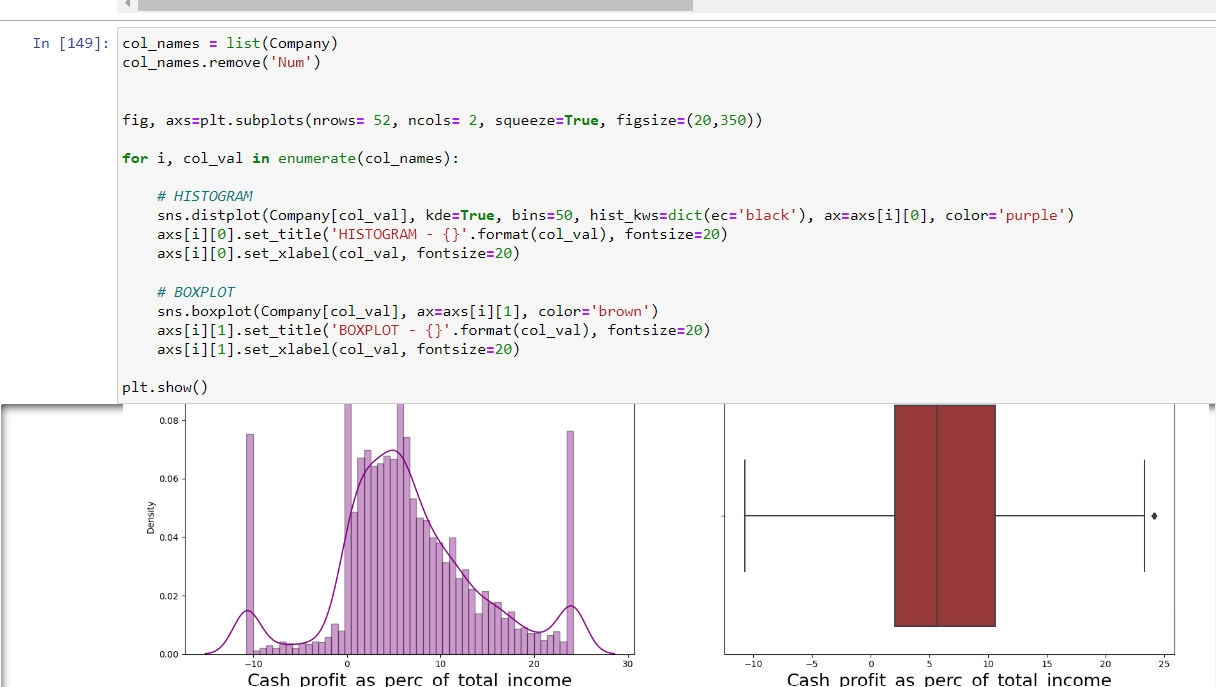


1. This criterion is linked to a Learning Outcome Missing Value Treatment

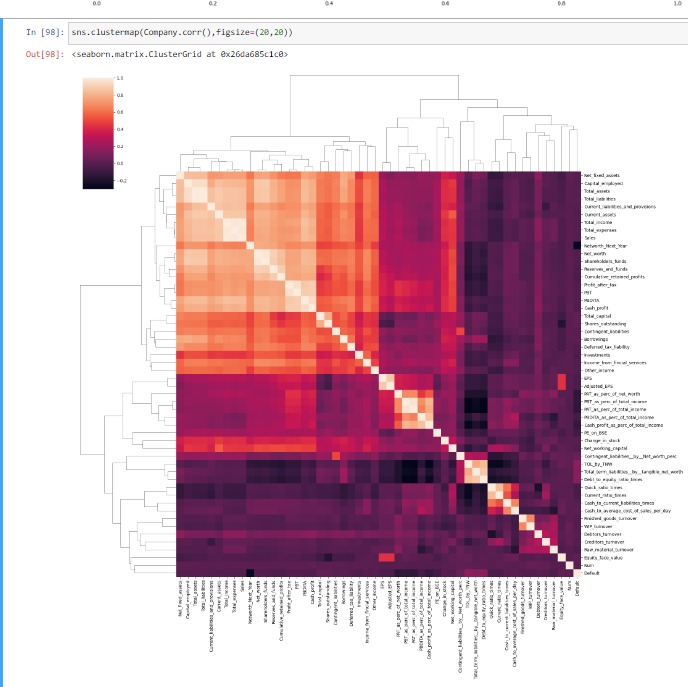


1. This criterion is linked to a Learning Outcome Univariate & bivariate analysis

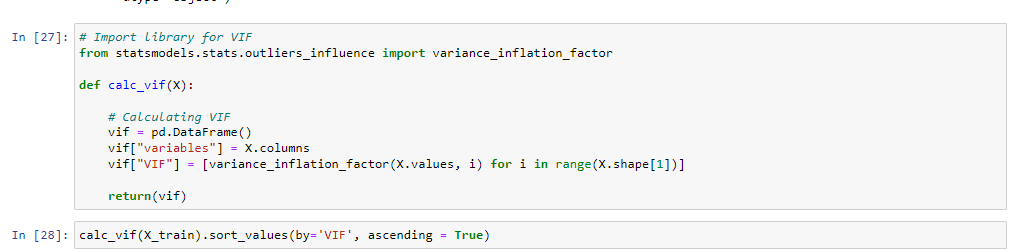
Check for 149 (Univariate Analysis)

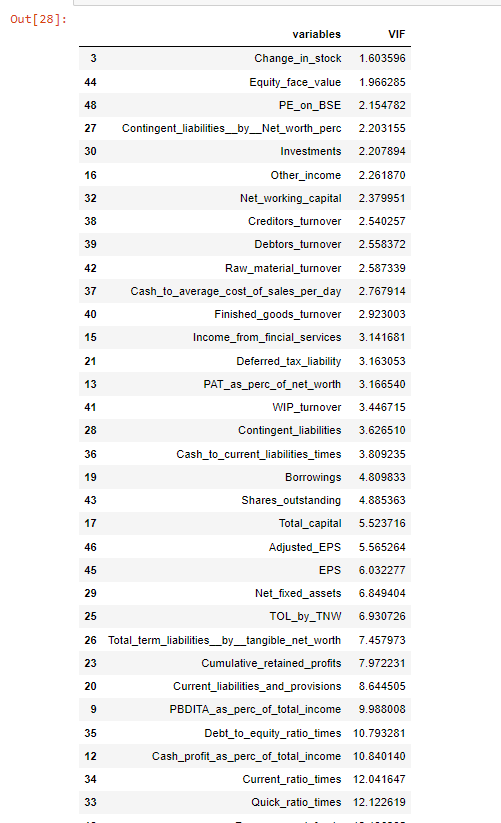


Bivariate Analysis

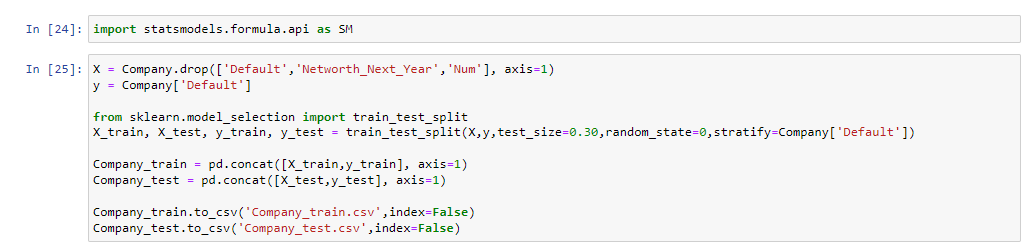


1. This criterion is linked to a Learning Outcome Check for multicollinearity

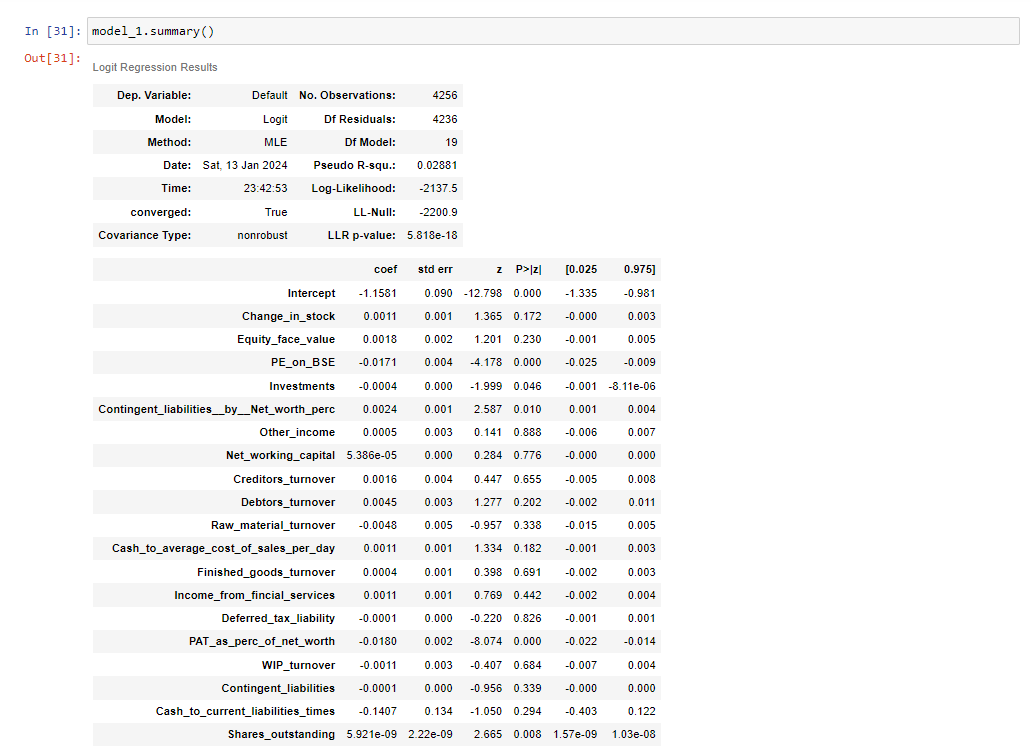


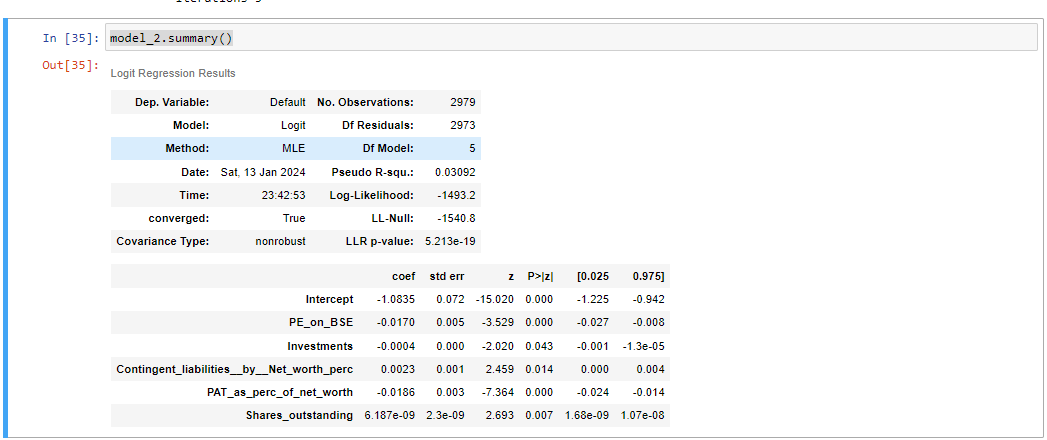


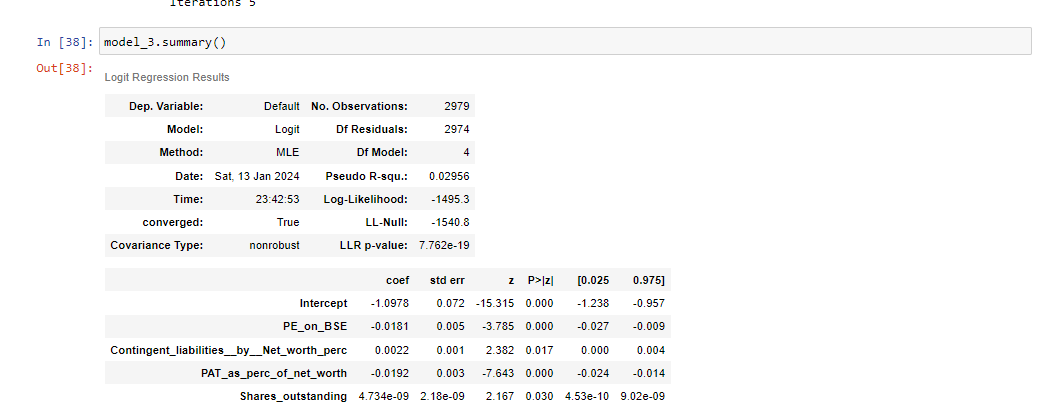
1. This criterion is linked to a Learning Outcome Split the data into train and test (70:30)



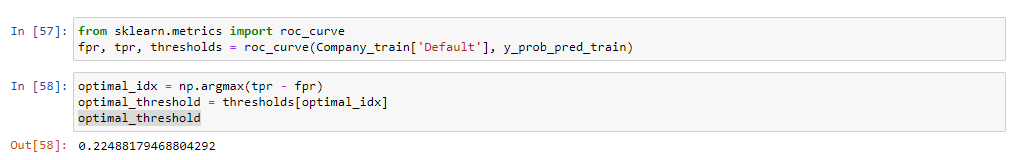
1. This criterion is linked to a Learning Outcome Build Logistic Regression Model (using stats model library) on the most important variables on Train Dataset and choose the optimum cutoff. Also, showcase your model-building approach



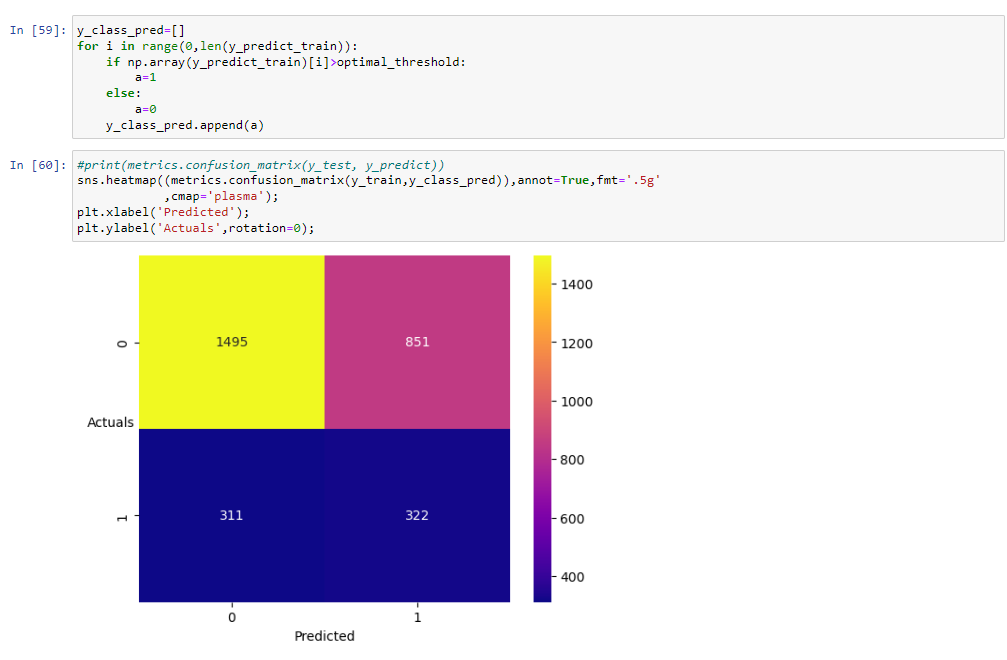




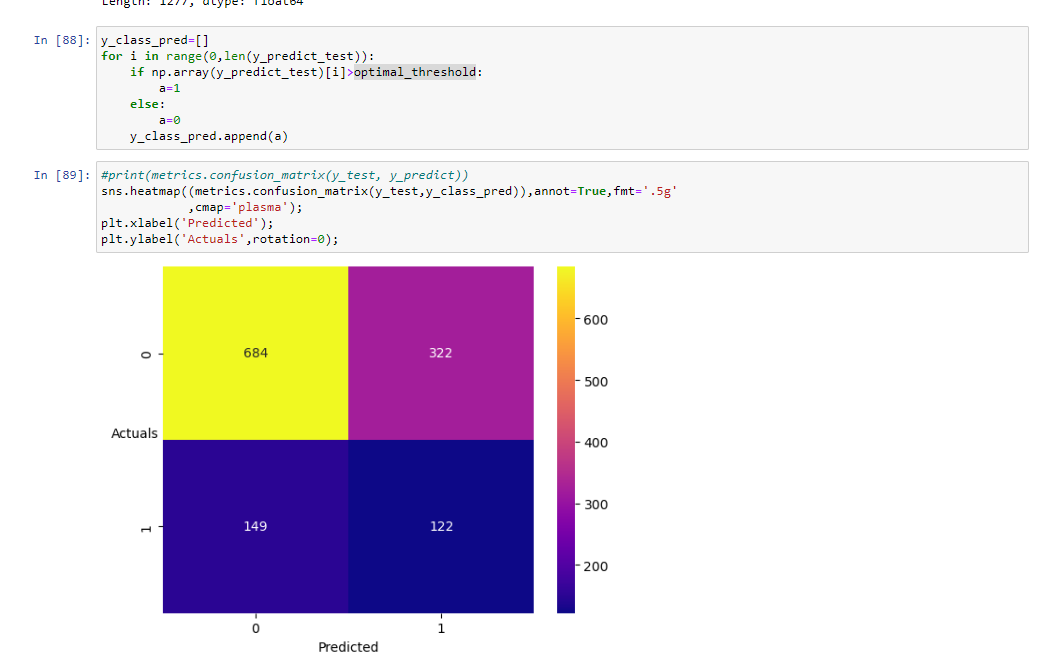
Optimum Threshold



Train Dataset –

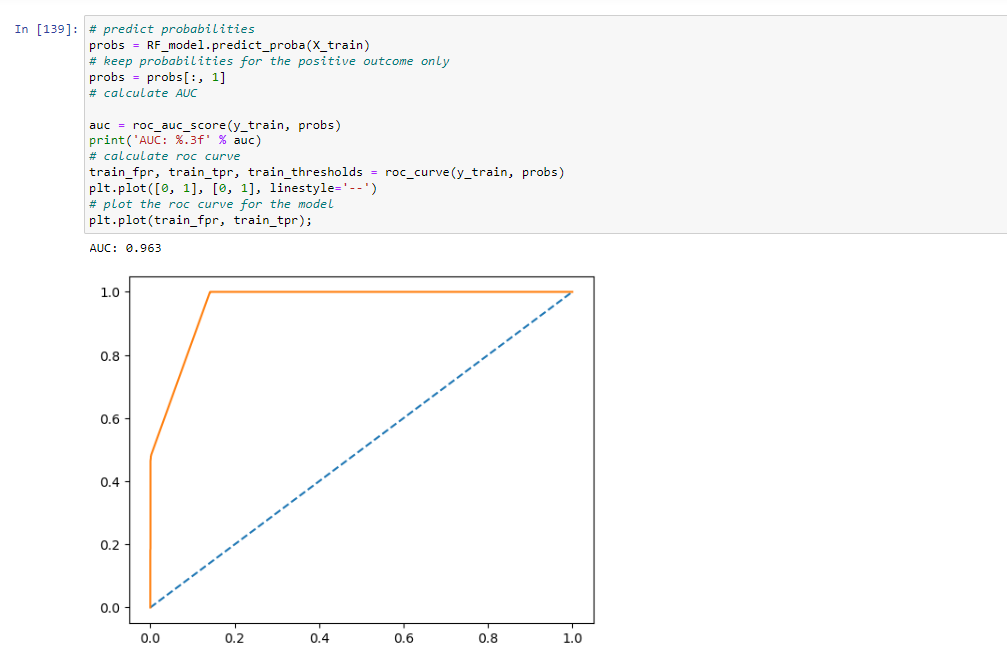


Test Dataset –



1. This criterion is linked to a Learning Outcome Build Random Forest Model on Train Data and validated on the Test Data

Train -



Test –

