

## Final Internship Project

Start Date: 11<sup>th</sup> September 2023

Submission Date: 24<sup>th</sup> September 2023

Demo during class on: 25<sup>th</sup> September 2023, 7:30 PM

### A) Summary of the project

In this project, we will try to combine various nuances of all that we have learnt during the entire course. This will have elements of Data Analytics, Statistics and Machine Learning. All these components will have their due weightages and proportions.

This project is about the credit card portfolio of a major bank in Singapore.

### B) Technical deliverables

In this project, you are expected to complete the below objectives (as a minimum, but you can always enhance with additional efforts):

**Objective-1:** Import the “Raw\_Data\_Singapore Credit Dataset” file, and run it through all the possible set of supervised ML models that you can, at 3 different split ratios – 30%/25%/20%.

**Objective-2:** Based on a comparative analysis of the various models with respect to the evaluation parameters, provide a justification of which ML model we are finalizing as our Go-To model for this project. Highlight the evaluation parameter too. In short, create the final model by fitting on the test dataset and we will preserve the model as reference for future fits.

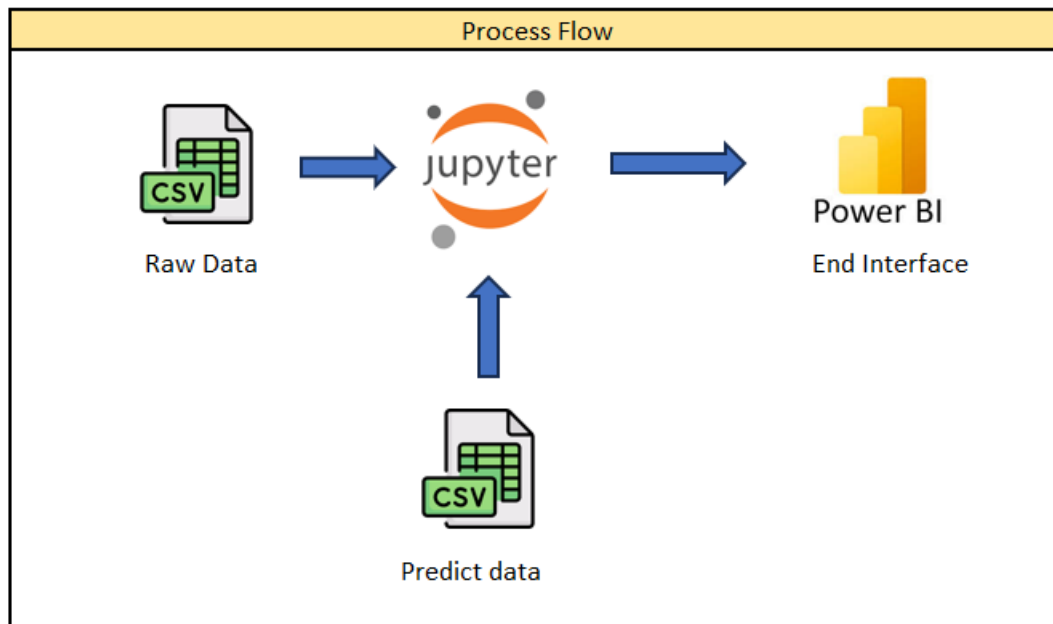
**Objective-3:** Now consider a production dataset “Predict\_Data\_Singapore Credit Dataset” file as an incoming new data, which has to undergo prediction using the finalized model in Objective-2. Please note, we do not have the target label in this dataset. The target label has to be generated via prediction for this dataset. Once the target label is appended to this dataset, we can call this as our Viz dataset.

**Objective-4:** Now, the output of Objective-3: Viz dataset (input file for the visualization exercise), needs to be imported into Power BI, and appropriate visualizations need to be created, which will serve as the final output for the clients.

The visualization must have the below pointers at the least:

- ✓ The name of the ML Model
- ✓ The evaluation Metrics
- ✓ The Test-Train split value
- ✓ Must have a visual to show proportion of Classes
- ✓ There should be visuals to show the distribution of the features, once any segment of the classes is selected. In other words, an answer as to which set of attributes leads to a particular class, can be visually answered.

## C) Visual Guide (Process Flow) and Hints of Sample Specimen



## A Sample Specimen

