**Assignment\_2: Ensemble Methods for Classification**

**Instructions:**

Prepare a Python notebook to build, train and evaluate models on the given datasets. Please read the instructions carefully.

**Business Case:** Build a classifier model to predict the credit card defaulter for given dataset

**Task: Credit Card Defaulter Dataset**

Please click the icon below to view/download the dataset



1. Import Libraries/Dataset (1 mark)

* Import the required libraries and the dataset

1. Data Visualisation and Augmentation (0.5\*6 = 3 marks)

* Plot at least two EDA graphs (use matplotlib/seaborn/any other library)
* Prepare data to be able to build a classification model
* Bring the train and test data in the required format
* Perform missing values check
* Perform scaling of data
* Print the shapes of train and test data

1. Random Forest Classifier Model Building (0.5\*3 = 1.5 marks)

* Build a random forest classification model
* Train the model on the train dataset
* Print the model summary

1. XGBoost Classifier Model Building (0.5\*3 = 1.5 marks)

* Build a XGBoost classification Model
* Train the model on the Train Data Set
* Print the model summary

1. Model Evaluation (1 + 1 = 2 marks)

* Check the random forest classifier model’s performance by printing accuracy, confusion matrix, F1 score and the AUC-ROC curve
* Check the XGBoost classification model performance by printing accuracy, confusion matrix, F1 score and the AUC-ROC curve

1. Compare the performance of random forest classification and XGBoost classification models for given dataset (1 mark)