Project Milestones – Supervised Learning

Criteria	Pts	Week	Topic
Part 1. 1 a - Read all the 3 CSV files as DataFrame and store them into 3 separate variables.	1.0 pts	Common part	Basic Data prep operations
Part 1.1 b - Print Shape and columns of all the 3 DataFrames	1.0 pts	Common part	Basic Data prep operations
Part 1.1 c - Compare Column names of all the 3 DataFrames and clearly write observations	1.0 pts	Common part	Basic Data prep operations
Part 1.1 d - Print DataTypes of all the 3 DataFrames.	1.0 pts	Common part	Basic Data prep operations
Part 1.1 e - Observe and share variation in 'Class' feature of all the 3 DaraFrames	1.0 pts	Common part	Basic Data prep operations
Part 1.2 a - Unify all the variations in 'Class' feature for all the 3 DataFrames	1.0 pts	Common part	Basic Data prep operations
Part 1.2 b - Combine all the 3 DataFrames to form a single DataFrame	1.0 pts	Common part	Basic Data prep operations
Part 1.2 c - Print 5 random samples of this DataFrame	1.0 pts	Common part	Basic Data prep operations
Part 1.2 d - Print Feature-wise percentage of Null values	1.0 pts	Common part	Basic Data prep operations
Part 1.2 e - Check 5-point summary of the new DataFrame	1.0 pts	Common part	Statistical Analysis
Part 1.3 a - Visualize a heatmap to understand correlation between all features	2.0 pts	Week 1	Statistical Analysis
Part 1.3 b - Share insights on correlation	2.0 pts	Week 1	Statistical Analysis
Part 1.3 c - Visualize a pairplot with 3 classes distinguished by colors and share insights	2.0 pts	Week 2	Visualization
Part 1.3 d - Visualize a jointplot for 'P_incidence' and 'S_slope' and share insights	2.0 pts	Week 1	Visualization
Part 1.3 e - Visualize a boxplot to check distribution of the features and share insights.	2.0 pts	Week 1	Statistics & Visualiztion
Part 1.4 a - Split data into X and Y.	1.0 pts	Week 2	Preparing data for bulilding model
Part 1.4 b - Split data into train and test with 80:20 proportion	1.0 pts	Week 2	Preparing data for bulilding model
Part 1.4 c - Train a Supervised Learning Classification base model using KNN classifier	2.0 pts	Week 3	KNN
Part 1.4 d - Print all the possible performance metrics for both train and test data	2.0 pts	Week 3	Performance Metrics
Part 1.5 a - Experiment with various parameters to improve performance of the base model	2.0 pts	Week 3	Parameter Tuning - Research required
Part 1.5 b - Clearly showcase improvement in performance achieved	1.0 pts	Week 3	Performance Metrics
Part 1.5 c - Clearly state which parameters contributed most to improve model performance.	1.0 pts	Week 3	Research required

Project Milestones – Supervised Learning

Criteria	Pts	Week	Topic
Part 2.1 a - Read both the Datasets 'Data1' and 'Data 2' as DataFrame and store them into two separate variables	1.0 pts	Common part	Data Preparation
Part 2.1 b - Print shape and Column Names and DataTypes of both the Dataframes.	1.0 pts	Common part	Data Preparation
Part 2.1 c - Merge both the Dataframes on 'ID' feature to form a single DataFrame	2.0 pts	Common part	Data Preparation
Part 2.1 d - Change Datatype of below features to 'Object'	1.0 pts	Common part	Data Preparation
Part 2.2 a - Visualize distribution of Target variable 'LoanOnCard' and clearly share insights	2.0 pts	Common part	Data Preparation
Part 2.2 b - Check the percentage of missing values and impute if required	1.0 pts	Common part	Data Preparation
Part 2.2 c - Check for unexpected values in each categorical variable and impute with best suitable value.	2.0 pts	Common part	Data Preparation
Part 2.3 a- Split data into X and Y	1.0 pts	Week 1	Data Preprocessing
Part 2.3 b - Split data into train and test. Keep 25% data reserved for testing	1.0 pts	Week 1	Data Preprocessing
Part 2.3 c - Train a Supervised Learning Classification base model - Logistic Regression	2.0 pts	Week 2	Logistic Regression
Part 2.3 d - Print evaluation metrics for the model and clearly share insights. [	1.0 pts	Week 2	Performance Metrics
Part 2.3 e - Balance the data using the right balancing technique.	2.0 pts	Week 3	Research required
Part 2.3 f- Again train the same previous model on balanced data	1.0 pts	Week 3	Logistic Regression
Part 2.3 g- Print evaluation metrics and clearly share differences observed	2.0 pts	Week 3	Performance Metrics
Part 2.4 a- Train a base model each for SVM, KNN	4.0 pts	Week 4	KNN & SVM
Part 2.4 b - Tune parameters for each of the models wherever required and finalize a mode	3.0 pts	Week 4	Research required
Part 2.4 c- Print evaluation metrics for final mode	1.0 pts	Week 4	Performance Metrics
Part 2.4 d- Share improvement achieved from base model to final model	2.0 pts	Week 4	Research required