

AIML Online

Project Milestone – Featurization, Model Selection and Tuning

The project is a comprehensive walk through of all the concepts you have learned so far. Please brush up all the concepts before attempting the project. Feature engineering, Model selection and Tuning are some of the concepts which you will get good exposure on throughout the project.

Criteria	Pts	Week	Topic
Q1 A - Import 'signal-data.csv' as DataFrame.	2.0 pts	Python	Pandas functions
Q1 B - Print 5 point summary and share at least 2 observations.	3.0 pts	Python	Pandas functions
Q2 A - Write a for loop which will remove all the features with 20%+ Null values and impute rest with mean of the feature.	5.0 pts	Data Preprocessing	For loop, Missing values treatment and imputing
Q2 B - Identify and drop the features which are having same value for all the rows.	3.0 pts	Week 1	Feature Engineering
Q2 C - Drop other features if required using relevant functional knowledge. Clearly justify the same.	2.0 pts	Week 1	Feature Engineering
Q2 D - Check for multi-collinearity in the data and take necessary action.	3.0 pts	Week 1	EDA, Feature Engineering
Q2 E - Make all relevant modifications on the data using both functional/logical reasoning/assumptions.	2.0 pts	Theoretical question	Observation from Q2 B, C, D and assumptions
Q3 A - Perform a detailed univariate Analysis with appropriate detailed comments after each analysis.	2.0 pts	EDA	Univariate Analysis
Q3 B - Perform bivariate and multivariate analysis with appropriate detailed comments after each analysis.	3.0 pts	EDA	Bivariate and Multivariate Analysis
Q4 A - Segregate predictors vs target attributes.	2.0 pts	Model Building	Separating dependent and Independent variables
Q4 B - Check for target balancing and fix it if found imbalanced.	3.0 pts	Week 1	Feature Engineering, Up sampling and Down Sampling
Q4 C - Perform train-test split and standardize the data or vice versa if required.	3.0 pts	Common part	Separating training data and test data
Q4 D - Check if the train and test data have similar statistical characteristics when compared with original data.	2.0 pts	Statistical Learning Revision	5 point summary of train, test and original data separately is expected here
Q5 A - Use any Supervised Learning technique to train a model.	2.0 pts	Supervised Learning Revision	Model building

Criteria	Pts	Week	Topic
Q5 B - Use cross validation techniques.	3.0 pts	Week 1 and Week 2	Cross validation Techniques, GridSearch and RandomSearchCV
Q5 C - Apply hyper-parameter tuning techniques to get the best accuracy.	3.0 pts	Week 2	Hyper parameters and tuning, model performance evaluation
Q5 D - Use any other technique/method which can enhance the model performance.	4.0 pts	Statistical Learning, Unsupervised Learning, EDA Revision	PCA, Standardization, Target balancing
Q5 E - Display and explain the classification report in detail.	3.0 pts	Supervised Learning Revision	Model performance evaluation, Classification matrix
Q5 F - Apply the above steps for all possible models that you have learnt so far.	5.0 pts	Supervised Learning and Ensemble Techniques Revision	Model building, hyper parameter tuning
Q6 A- Display and compare all the models designed with their train and test accuracies.	1.0 pts	Supervised Learning Revision	Model performance evaluation
Q6 B - Select the final best trained model along with your detailed comments for selecting this model.	1.0 pts	Common part	Overall observations
Q6 C - Pickle the selected model for future use.	2.0 pts	Common part	Pickling model for future use
Q6 D - Write your conclusion on the results.	1.0 pts	Common part	Detailed suggestions on the data points collected by the bank to do better Analysis in the future

Wishing the best for every step in this journey!

*****Happy Learning*****