| **Criteria** | **Ratings** | **Points** |
| --- | --- | --- |
| **Statistical Summary and Null value check** | Areas of Focus: Q1.a. Providing Inferences from the step "check datatype, statistical summary, shape, null values or incorrect imputation" could have been attempted with supporting details. (It's only an Observation. No marks are being deducted.) Areas performed well: Q1. Checking and treatment of incorrect imputed data with required data imputation techniques. e.g. Treating -ive values in Experience column is handled without dropping. | 5/5 |
| **EDA** | Areas of focus: Q2. EDA: Highlighting the presence of highly correlated variables as a part of Bi-variate analysis is missing either using correlation matrix and pair plot e.g: experience and age are highly correlated. Q2. Listing out the required findings from Uni-variate and Bi-variate analysis "share your findings" as a part of below question "EDA: Study the data distribution in each attribute and target variable, share your findings" are either very limited or missing. Q2. (Optional) Uni-Variate Analysis: Explaining/Studying the target column distribution as a part of Uni-variate analysis could have been more meaningful or relevant. e.g. Apart from printing/highlighting the value counts of class labels, highlighting the presence of class imbalance and its impact on overall model performance could have been attempted. | 19.75/20 |
| **Data Split** | Areas performed well: Q3. Splitting the data with Random state seeding for controlling the randomness for reproducibility of the results by explicitly seeding the "random\_state" pseudo random number generator. | 5/5 |
| **Modelling** | Areas of focus: Q4. The reasoning behind dropping the following list of features before model building is not clear from your code and due to this step, your overall model performance has been impacted at Line 29. "'Family','Education','Online','CCAvg','Experience','Income','Mortgage','Personal Loan','Securities Account'" | 14.75/15 |
| **Model Parameters** | Areas performed well: Q5. The approach being followed by including the test dataset to showcase all the rows where the predicted class is not equal to the observed class is attempted well with supporting output results. | 10/10 |
| **Business Insights** | Areas of focus: Q6. Overall your answer to the reported question (Q: Give Business understanding of your model?) is correct but can be improved by illustrating following approaches as well. (It's only an Observation. No marks are being deducted after considering your overall approach.) a. It is expected to conclude the metric of choice as 'Recall'(Reducing the False Negative rate) or 'Precision' (Reducing the False Positive rate) for the current business problem along with providing the reason on why it has to be a choice of best evaluation metric in the current business setting. b. Highlighting the importance of FP and FN in the context of business value proposition (i.e. which one is more important to business and why ) and also interpreting the values of FP and FN scores while concluding optimal the model performance. e.g. Following kind of interpretation is being expected. "If false positives are high then the bank will predict more interest income than they will actually get. This will be bad for financial planning, as the bank may budget incorrectly and over spend. If false negatives are high then the bank will predict lower income due interest than they will actually generate. Although this is not ideal for accurate financial planning, the consequences will be that they have more "money in the bank" at the end of the budgeted period. So false negatives are not too bad. True negatives add to the accuracy of the prediction and thus better financial planning. Now consider two of the performance metrics of the model. Recall = TP /(TP + FN) Precision = TP / (TP + FP) If False Negatives are high, then the recall value will be low. The bank makes more money than they expected. If False Positives are high, then the precision will be low. The bank makes less money than expected. Therefore, the Precision is more important than the Recall in this case. The accuracy and the precision of our model are quite high, the recall is not high. Since recall is not that important, so we can conclude that the current model is quite good." | 5/5 |
|  | Points | 59.5/60 |