Project 6: Credit Card Users Churn Prediction

**Submission type**: File Upload **Due Date**: Jul 23, 8:30 PM **Total Score**: 60

**Background & Context**

The Thera bank recently saw a steep decline in the number of users of their credit card, credit cards are a good source of income for banks because of different kinds of fees charged by the banks like annual fees, balance transfer fees, and cash advance fees, late payment fees, foreign transaction fees, and others. Some fees are charged on every user irrespective of usage, while others are charged under specified circumstances.

Customers’ leaving credit cards services would lead bank to loss, so the bank wants to analyze the data of customers’ and identify the customers who will leave their credit card services and reason for same – so that bank could improve upon those areas

You as a Data scientist at Thera bank need to come up with a classification model that will help bank improve their services so that customers do not renounce their credit cards

**Objective**

1. Explore and visualize the dataset.
2. Build a classification model to predict if the customer is going to churn or not
3. Optimize the model using appropriate techniques
4. Generate a set of insights and recommendations that will help the bank

**Data Dictionary:**

* CLIENTNUM: Client number. Unique identifier for the customer holding the account
* Attrition\_Flag: Internal event (customer activity) variable - if the account is closed then 1 else 0
* Customer\_Age: Age in Years
* Gender: Gender of the account holder
* Dependent\_count: Number of dependents
* Education\_Level: Educational Qualification of the account holder
* Marital\_Status: Marital Status of the account holder
* Income\_Category: Annual Income Category of the account holder
* Card\_Category: Type of Card
* Months\_on\_book: Period of relationship with the bank
* Total\_Relationship\_Count: Total no. of products held by the customer
* Months\_Inactive\_12\_mon: No. of months inactive in the last 12 months
* Contacts\_Count\_12\_mon: No. of Contacts in the last 12 months
* Credit\_Limit: Credit Limit on the Credit Card
* Total\_Revolving\_Bal: Total Revolving Balance on the Credit Card
* Avg\_Open\_To\_Buy: Open to Buy Credit Line (Average of last 12 months)
* Total\_Amt\_Chng\_Q4\_Q1: Change in Transaction Amount (Q4 over Q1)
* Total\_Trans\_Amt: Total Transaction Amount (Last 12 months)
* Total\_Trans\_Ct: Total Transaction Count (Last 12 months)
* Total\_Ct\_Chng\_Q4\_Q1: Change in Transaction Count (Q4 over Q1)
* Avg\_Utilization\_Ratio: Average Card Utilization Ratio

**Best Practices for Notebook :**

* The notebook should be well-documented, with inline comments explaining the functionality of code and markdown cells containing observations and insights.
* The notebook should be run from start to finish in a sequential manner before submission.
* It is preferable to remove all warnings and errors before submission.
* The notebook should be submitted as an HTML file (.html)

**Best Practices for Presentation :**

Like in real-world projects, the ultimate destination of any project or work is generally an executive or decision-making meeting, where you are supposed to present your solution to the business problem, based on the project/work you have done. The purpose of this presentation is to simulate that kind of experience, and to draw the attention of your audience (a business leader like CMO, COO, CFO or CEO) to the key points of your project, which are:

* Business overview of the problem and solution approach
* Key findings and insights which can drive business decisions
* Model overview and performance summary
* Business recommendations

Please keep the following points in mind while making the presentation:

* Focus on explaining the takeaways in an easy-to-understand manner.
* Inclusion of the potential benefits of implementing the solution will give you the edge.
* Copying and pasting from the notebook is not a good idea, and it is better to avoid showing codes unless they are the focal point of your presentation.
* Please submit the presentation in PDF format only.

**Submission Guidelines :**

1. There are two parts to the submission:
   1. A well commented Jupyter notebook [format - .html]
   2. A presentation as you would present to the top management/business leaders [format - .pdf ] (you have to export/save the .pptx file as .pdf)
2. Any assignment found copied/ plagiarized with other groups will not be graded and awarded zero marks
3. Please ensure timely submission as any submission post-deadline will not be accepted for evaluation
4. Submission will not be evaluated if,
   1. it is submitted post-deadline, or,
   2. more than 2 files are submitted

**Scoring guide (Rubric) - CreditCard Users Churn Prediction**

| **Criteria** | **Points** |
| --- | --- |
| **Perform an Exploratory Data Analysis on the data**  - Univariate analysis - Bivariate analysis - Use appropriate visualizations to identify the patterns and insights - Any other exploratory deep dive | 4 |
| **Illustrate the insights based on EDA**  Key meaningful observations on the relationship between variables | 4 |
| **Data Pre-processing**  Prepare the data for analysis - Missing value Treatment, Outlier Detection(treat, if needed- why or why not ), Feature Engineering, Prepare data for modeling | 4 |
| **Model building - Logistic Regression**  - Make a logistic regression model - Improve model performance by up and downsampling the data - Regularize above models, if required | 5 |
| **Model building - Bagging and Boosting**  - Build Decision tree, random forest, bagging classifier models - Build Xgboost, AdaBoost, and gradient boosting models | 8 |
| **Hyperparameter tuning using grid search**  - Tune the best 3 models using grid search and provide the reason behind choosing those models - Use pipelines in hyperparameter tuning | 8 |
| **Hyperparameter tuning using random search**  - Tune the best 3 models using random search and provide the reason behind choosing those models - Use pipelines in hyperparameter tuning | 8 |
| **Model Performances**  - Compare the model performance of all the models - Comment on the time taken by the grid and randomized search in optimization | 4 |
| **Actionable Insights & Recommendations**  - Business recommendations and insights | 3 |
| **Report - Overall quality**  - Structure and flow - Crispness - Visual appeal - All key insights and recommendations covered | 8 |
| **Notebook - Overall quality**  - Structure and flow - Well commented code | 4 |
| Points | 60 |