Supporting Documentation

Student ID: 20439708

Instructions for Running the Churn Model

To run the churn model, the following steps should be followed:

- Open the Databricks notebook named "Churn_Prediction_20439708.html" and run sll the cells sequentially till Section 6 -"Unbiased Accuracy of Model".
- Then run the next cell code of Section 7 - "Model Implementation" to save the file of trained model with the name "Model_final.pkl".
- Open Databricks notebook named "Predict_20439708.html" in Databricks and load the saved "Model_final.pkl" file in the Databricks environment.
- Load or connect to the SQL database, which has tables related to the transactional data by entering the url of the database in the "filepath" variable under first section called "Loading Database" section or load the SQL database in the environment. Ensure the database has same schema as shown in Appendix of the submitted report, to avoid any error in the execution of code.
- Run each cell sequentially under "Predict.html" Databricks notebook.
- The final output of the prediction will contain a CSV file containing the list of potential churners as predicted by the model.

Data Cleaning

No extra cleaning was required.

List of Figures

Figure Number	Figure Caption	Method used
1	Approach for creating Lagged Features	Powerpoint- shapes
2	Variable Importance based on filter based method	Python - seaborn library using bar plot function (Section: 5b in databricks code)
3	Variable Importance based on model based method	Python - seaborn library using bar plot function (Section: 5c in databricks code)
4	PDP for feature "no purchase till date"	Python - pdpbox library using pdp_plot function (Section 8 in databricks code)
A.1	SQL Database Schema showing relationship between SQL tables	Online tool- Lucid Chart
B.1	Pen profile descriptive statistics	Python - groupby and describe function (Section 9 in databricks code)