**Instructions to Run Code from DOS prompt:**

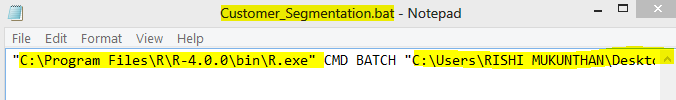
**List of Coding Files that are submitted:**

**R:**

1. Santander\_EDA.R
2. Santander\_Prediction\_Development.R
3. Santander\_Predict\_Final\_Deploy\_NB.R
4. Santander\_Predict\_Final\_Deploy\_XGB.R
5. Santander\_Predict\_Final\_Deploy\_NB.bat
6. Santander\_Predict\_Final\_Deploy\_XGB.bat

For running Final models (NB and XGB) R scripts files from the windows command prompt, a batch file (.bat) need to be prepared. I have attached the .bat files for the 2 R scripts.

You can just run the .bat file which will open the command prompt and start executing the R script. Before running the .bat file, edit the .bat file path with the corresponding R installation path and path of R script to be run.

* Example: My R installation/R.exe file path is **"C:\Program Files\R\R-4.0.0\bin\R.exe"**
* R script path: **"** **C:\Users\RISHI MUKUNTHAN\Desktop\Data Science\Projects\Edwisor\_Santander\_Customer\_Transaction\Submission\Deployment R Scripts\Santander\_Predict\_Final\_Deploy\_NB.R"**

**What each R script does?**

**Santander\_Predict\_Final\_Deploy\_NB.R** – Takes in train.csv and trains a Naïve Bayes model and makes predictions on the test.csv and outputs a prediction file.

**Input**: train.csv, test.csv

**Output:** NB\_Predictions\_R.csv

**Santander\_Predict\_Final\_Deploy\_XGB.R** – Takes in train.csv and trains a XGBoost model and makes predictions on the test.csv and outputs a prediction file.

**Input**: train.csv, test.csv

**Output:** XGB\_Predictions\_R.csv

**Python:**

1. **Santander\_EDA.ipynb**
2. **Santander\_Prediction\_Development.ipynb**
3. **Santander\_Prediction\_Final\_Deploy\_NaiveBayes.ipynb**
4. **Santander\_Prediction\_Final\_Deploy\_XGB.ipynb**
5. **Santander\_Predict\_Final\_Deploy\_NaiveBayes.py**
6. **Santander\_Predict\_Final\_Deploy\_XGB.py**
7. **Santander\_Prediction\_ANN.ipynb**

**File 1.** Jupyter notebook file where I have performed EDA.

**File 2.** Jupyter notebook file where data preprocessing and model development and experimenting with different model and evaluating models has been done.

**File 3&4.** Jupyter notebooks for chosen models.

**File 5.** Is the .py script which is to be executed from the windows command prompt by running with the below command. Open command prompt from the path where you have downloaded the code files in your system and run the python command to run the .py scripts.

**Input:** train.csv, test.csv

**Output:** Test\_Prediction\_NB.csv

**File 6.** Is the .py script which is to be executed from the windows command prompt by running with the below command. Open command prompt from the path where you have downloaded the code files in your system and run the python command to run the .py scripts.

**Input:** train.csv, test.csv

**Output:** Test\_Prediction\_XGB.csv

**Python Libraries and their version used in the project:**

joblib 0.13.2

matplotlib 3.3.0

numpy 1.19.0

pandas 1.0.4

scikit-learn 0.23.1

seaborn 0.9.0

xgboost 1.1.1

imblearn 0.0

tensorflow 2.0.0

**Libraries loaded in R:**

"ggplot2", "corrgram", "DMwR", "caret", "randomForest", "unbalanced", "C50", "dummies", "e1071", "Information", "MASS", "rpart", "gbm", "ROSE", 'sampling', 'DataCombine', 'inTrees’, ‘tidyverse’, ‘funModeling’, ‘Hmisc’,’ factoextra’,’ dplyr’,’ caTools’,’ cvAUC’,’ pROC’, ‘xgboost’