**Bank Credit Card Customers Churn Prediction Model Using Supervised Classification Model**

1. **Introduction**

There’s a higher attrition rate among credit card customers in the banking industry. Acquiring new customers are more costly compared to retaining existing customers. As such, it is in the best interest of any Bank to keep a close watch on its customers to monitor any signs of potential churning down the line. This project is for assessing and predicting credit card customer attrition in a retail bank. A set of predictor variables was identified and added to data set obtained from a website and used to train a set of predictive models. The goal of this project is to provide a new approach to identify potential churners so marketing retention strategies be developed accordingly.

**Objective**

Aim to accomplish the following from this study:

1. Identify and visualize which factors contribute to customer churn:
2. Build a prediction model that will perform the following:
   * Classify if a customer is going to churn or not
   * Preferably and based on model performance, choose a model that will attach a probability to the churn to make it easier for customer service to target low hanging fruits in their efforts to prevent churn
3. **Data**

There are 10,000 rows and 20 attributes

**Data set review & preparation**

Reviewed this further to identify what attributes will be necessary and what data manipulation needs to be carried out before exploratory analysis and prediction modelling

No Missing values in the data set

CreditScore 0

Age 0

Tenure 0

Balance 0

Limit 0

FEE\_LEVIED 0

CLD 0

utilization 0

Revolver 0

NumOfProducts 0

spenddecrease 0

paymentratio 0

HasCrCard 0

EstimatedSalary 0

Exited 0

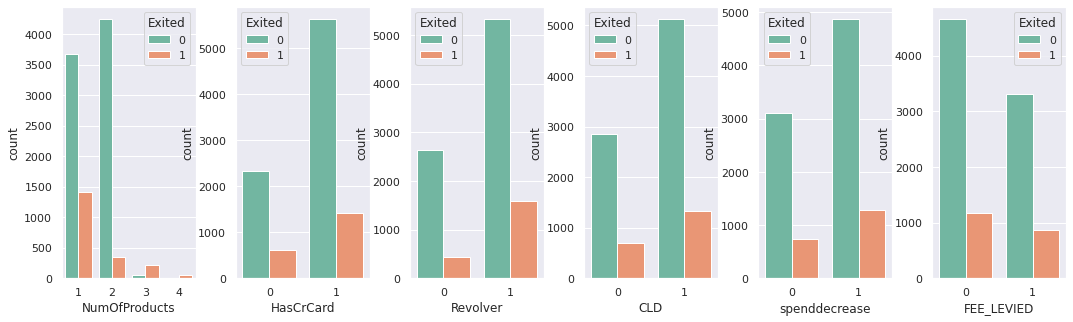
Dropped below unnecessary fields from the data set

"RowNumber", "CustomerId", "Surname","Geography","Gender"

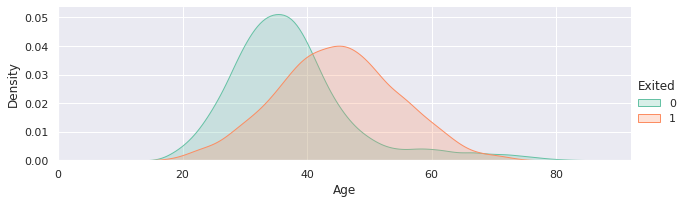
In this data set there are 5 categorical variables and other are continuous.

**Exploratory Data Analysis**

Objective of EDA is to get an understanding as to how the given attributes relate to the 'Exit' status.



Customers with higher number of products has a higher chances to churn



40-60 Age group has a higher chance to churn

1. **Methodology (Solution Approach, Tools)**

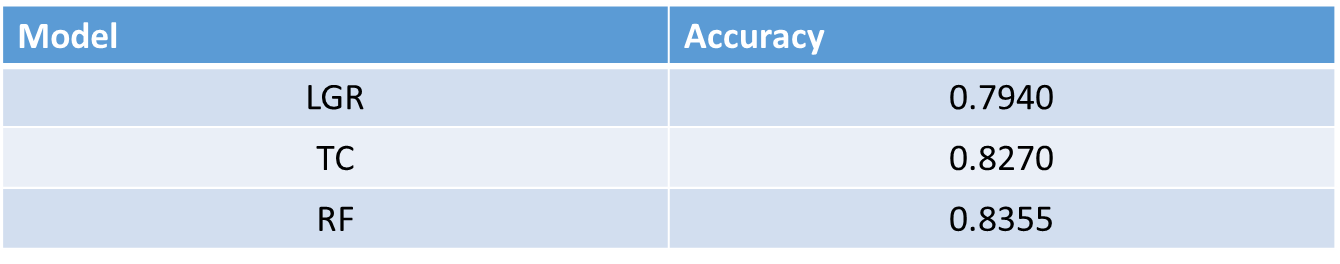
40-60 Age group has a higher chance to churn

Supervised learning Algorithm used for model development. It was used to learn the mapping function between Independent and Dependent Variables. Goal is to learn the underlying pattern in Training data set and validating its accuracy on Test Data.

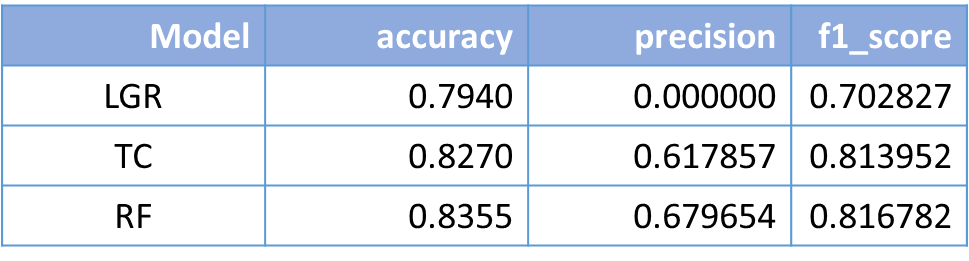
Logistic Regression Model, Tree Classifier and Random Forest Classifiers were used

1. **Results**

Accuracy of Models



Comparing Models



1. **Conclusion**

Based on model performance Random Forest Classifier is the best model for prediction

1. **Discussion**

From the above results, main aim is to predict the customers that will possibly churn so they can be put in some sort of scheme to prevent churn hence the recall measures on the 1's is of more importance than the overall accuracy score of the model.

From the review of the fitted models above, the best model that gives a decent balance of the accuracy and precision is the random forest where according to the fit on the training set, with a precision score on 1's of 0.83, out of all customers that the model thinks will churn, 83% do actually churn.

However, in as much as the model has a high accuracy, it still misses some of those who end up churning. This could be improved by retraining the model with more data over time while in the meantime working with the model to save customers that would have churned