Predicting and Preventing Customer Churn

Saroj Raut

Bellevue University

**Executive Summary**

Customer Attrition has always been one of the key challenges impacting the bottom line in any industry and so is it for the Merchant acquiring industry. By being able to identify customers that are at risk, corrective actions can be taken to address the issue.

From the data exploration, status, and their relationship with the categorical variable was identified. About 20% of the customers have churned with most of the data from France. There is more female customer churn than that of male customers and most of the customers that churned are those with credit cards. The inactive member has a greater churn than the customer that are active. Neither the product nor the salary has a significant effect on the likelihood to churn.

Overall, predictions for customer churn were achieved with strong accuracy. From the review of the fitted models above, the best model that gives a decent balance of the recall and precision is the random forest where according to the fit on the training set, with a precision score on 1's of 0.88, out of all customers that the model thinks will churn, 88% do churn and with the recall score of 0.53 on the 1's, the model can highlight 53% of all those who churned. The precision of the model on previously unseen test data is slightly higher about predicting 1's i.e. those customers that churn. However, in as much as the model has high accuracy, it still misses about half of those who end up churning. This could be improved by providing retraining the model with more data over time while in the meantime working with the model to save the 41% that would have churned.

Businesses need to update and upgrade their business as per the changing need of the market. To gain a competitive edge in the market, it is necessary to focus on areas that are identified as a risk and take necessary corrective actions for it.

Predicting and Preventing Customer Churn

# Introduction

Customer attrition means that a customer stops his or her business relationship with a company. It is the loss of customers by a business. Businesses classify a customer as churned(inactive) after a certain period when the customer is not active or make any transaction from the company. By losing customers, the business loses both revenue and the marketing costs involved in replacing them with new customers.

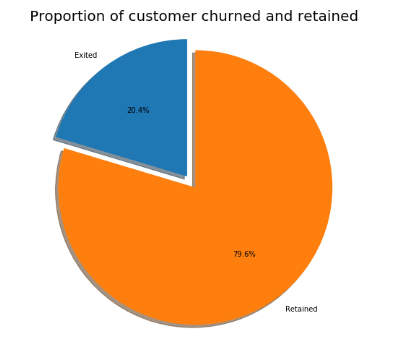
The purpose of this project is to develop an AI model for looking into customer activity and pattern to provide customers that are at risk of likely to leave soon. The model provides customers that are at risk or will likely leave soon by looking into customer activity and pattern. The model also provides LOB an ability to utilize model output data for identifying merchants at risk, finding microsegments of merchants at risk, and develop corrective action to lower

# Methods

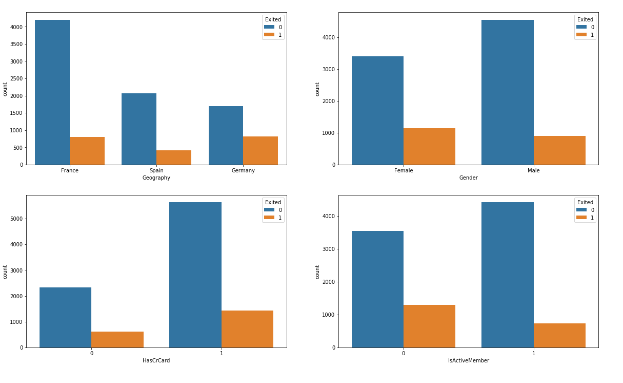
The data available is from the 2018 churn modeling Data obtained from Kaggle. There was data about customer CreditScore, Geography, Age, Balance, Tenure, etc. from 2018. Data exploration and data preparation are key to the success of building a good model. Data exploration was performed using Python, R, and data visualization to study the shape, summary, and data types for given Kaggle dataset. Exploration and analysis were very helpful in preparing the data for model training and feature generation.

Looking at the data was well worth it. CustomerID and surname were removed as they are specific to a customer. The surname would result in profiling, so we excluded that as well.

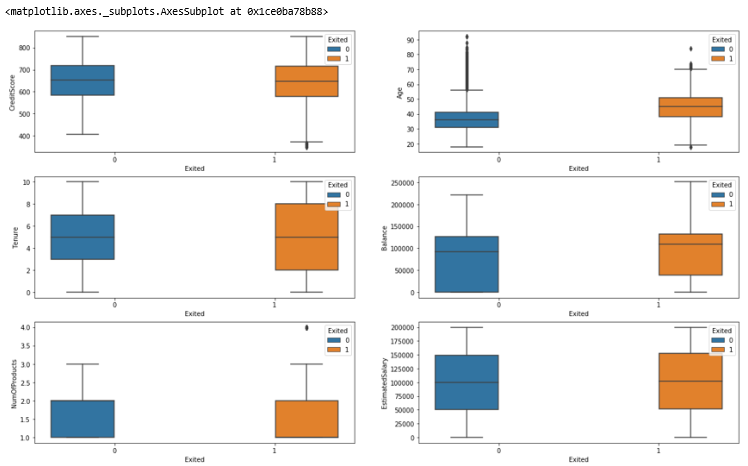
From the data exploration, about 20% of the customers have churned. So, the baseline model could be to predict that 20% of the customers will churn. Given 20% is a small number, we need to ensure that the chosen model does predict with great accuracy this 20% as it is of interest to the bank to identify and keep this bunch as opposed to accurately predicting the customers that are retained.



Most of the data is from persons from France. The proportion of female customers churning is also greater than that of male customers. Interestingly, most of the customers that churned are those with credit cards. Given that majority of the customers have credit cards could prove this to be just a coincidence. Unsurprisingly the inactive members have a greater churn. Worryingly is that the overall proportion of inactive members is quite high suggesting that the bank may need a program implemented to turn this group to active customers as this will have a positive impact on the customer churn.

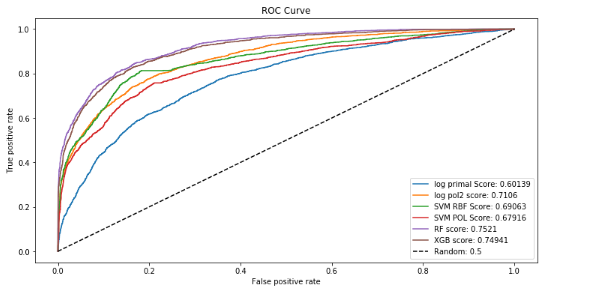


There is no significant difference in the credit score distribution between retained and churned customers. The older customers are churning at more than the younger ones alluding to a difference in service preference in the age categories. The bank may need to review its target market or review the strategy for retention between the different age groups. Neither the product nor the salary has a significant effect on the likelihood to churn.



# Results

From the review of the fitted models above, the best model that gives a decent balance of the recall and precision is the random forest where according to the fit on the training set, with a precision score on 1's of 0.88, out of all customers that the model thinks will churn, 88% do churn and with the recall score of 0.53 on the 1's, the model can highlight 53% of all those who churned.

The precision of the model on previously unseen test data is slightly higher regarding predicting 1's i.e. those customers that churn. However, in as much as the model has high accuracy, it still misses about half of those who end up churning. This could be improved by providing retraining the model with more data over time while in the meantime working with the model to save the 41% that would have churned :-)

# Discussion and Conclusion

In conclusion, it can be said that the Business needs to update and upgrade its business as per the changing need of the market. To gain a competitive edge in the market, it is necessary to focus on areas that are identified as a risk and take necessary corrective actions for it.

# Acknowledgments

We would like to acknowledge the in-depth analysis made available to the public by the Kaggle. We would also like to acknowledge Professor Fadi at Bellevue University for his feedback and thoughts throughout this term regarding this project. We would also like to thank

Our family for his patience of our many hours of research, and their willingness to let us bounce ideas off them. Finally, we would like to thank everyone from our class directly or indirectly in helping us complete this project.

# References

[1] <https://www.kaggle.com/kmalit/bank-customer-churn-prediction/notebook>

- <https://opendatascience.com/solving-merchant-attrition-using-machine-learning/>