Predicting Churn Credit Card Customers MVP

Nouf Alotaibi

Email: noufmitla@gmail.com

This project aims to build classification models to predict who the customers will get churn in Bank Unity. Therefore, the Bank can proactively go to the customer and provide them better services. I used a Credit Card Customers dataset from Kaggle with 0,127 customers records and 21 features. After doing an EDA, I build a Logistic Regression baseline model as shown below after splitting the data into training, validation, and testing with 75%, 15%, and 10%, respectively

Baseline Model

```
Train Scores:
- Accuracy score:  0.8878787878787879
- F-1 score:  0.5882922109337204
- Precision score:  0.7281437125748503
- Recall score:  0.4935064935064935
- F-beta of 2 score:  0.5275030366128752
```

```
Train Scores:
- Accuracy score:  0.8878787878787879
- F-1 score:  0.5882922109337204
- Precision score:  0.7281437125748503
- Recall score:  0.4935064935064935
- F-beta of 2 score:  0.5275030366128752
```

There is an imbalance in the data, so the accuracy score is not an appropriate measure to evaluate model performance. The number of observations is 10127, the number of Attired and existing customers is 1627 and 8500, respectively. So, a classifier that achieves an accuracy of 88% with an event rate of 16% is not accurate.

In our case, the cost of False Negatives is much higher than the cost of False Positives. The cost of a customer who is attired and predicted by the model as an existing customer is high, and the bank will lose them since they will not provide better services. Therefore, we want to decrease the FN's to increase recall.