Classification –
Credit Card Dataset

Dolores Umbridge

"Oh, you won't need ink"

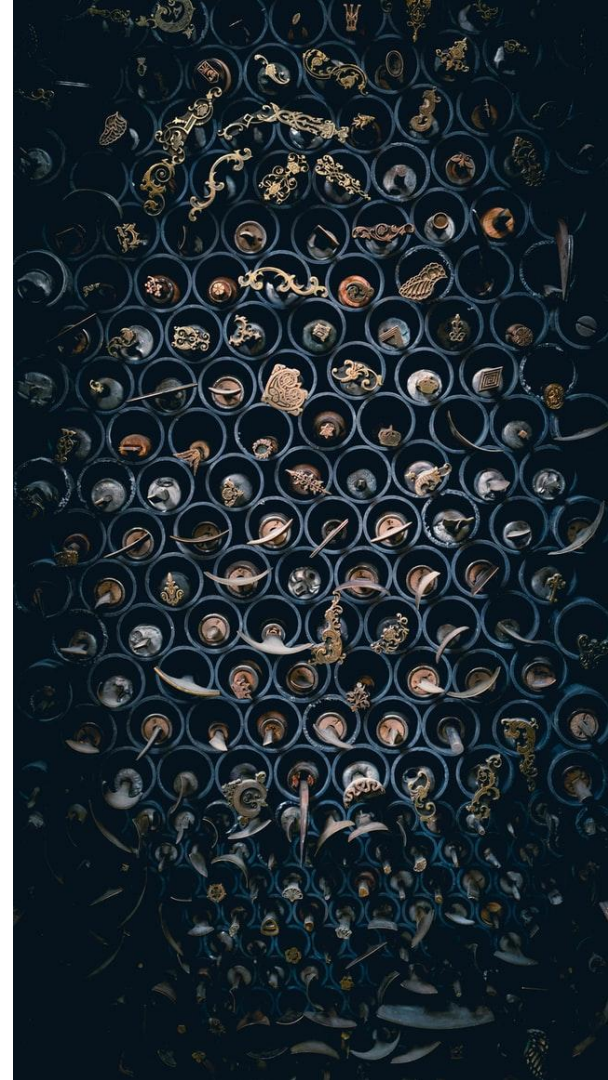# The Team

Karina

Davis

Prebitha

Supervised by

Kostadin

The Credit Card Dataset

Worked on the dataset using MySQl workbench, Tableau and python using Jupyter Notebook

From our findings its safe to say -
1. Dataset Highly Imbalanced
2. Biased to Offer Accepted = 'NO'

How did we approach this?

STEP 1
We used the **Logistic Regression Model**

STEP 2
A few numerical columns we turned into categorical

STEP 3
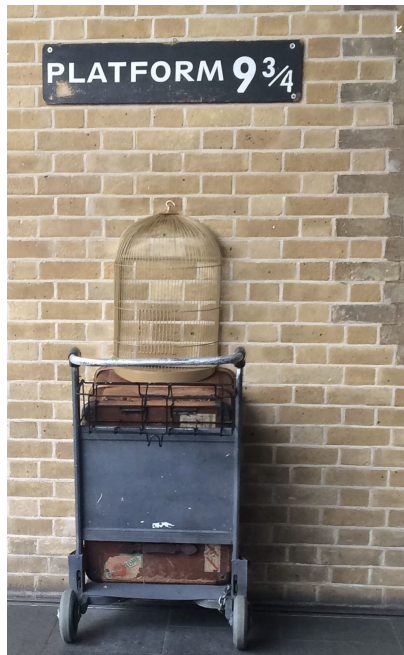Treated the imbalance with **SMOTE**

STEP 4
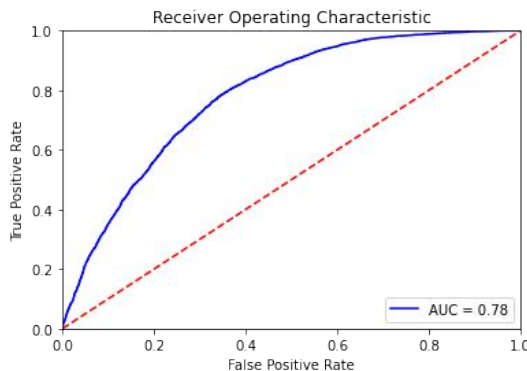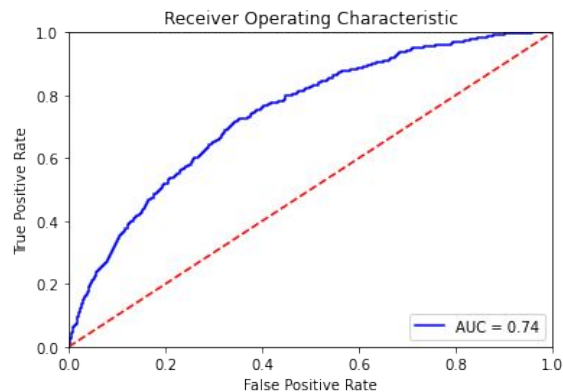We accessed the Model based on the **Area Under the Curve**

STEP 5
A few **METHODS**

STEP 6
Validated our model

How did we treat the imbalance?

Accuracy = 94.19%
AUC = 0.74

Accuracy = 69%
AUC = 0.78

Conclusion

With SMOTE and IQRs the Model got better at prediction but our accuracy dropped. With Upsampling using 'DAVIS Method' was overfitting but we had the highest accuracy of 77%

Method A

Mailer Column

Q1 Balance
+ Average Balance

Average Balance
+ Overdraft Protection

Conclusion

Our accuracy got a teeny bit better with dropping Average Balance + Overdraft Protection

Method B

Binned Household Size
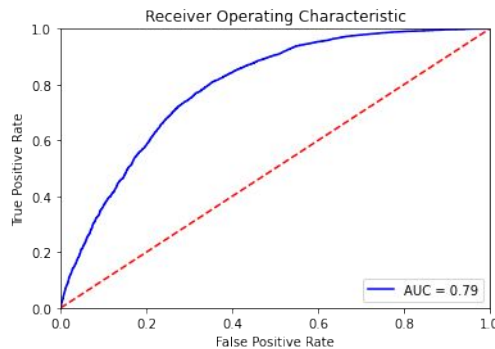+ dropped Overdraft Protection
+ dropped Average Balance

Accuracy
72.32%
AUC
0.79

Conclusion

The Accuracy & AUC increases with binning with q-cut

Binned Bank Accounts Opened
+ dropped Overdraft Protection
+ dropped Average Balance

Accuracy
71.93%
AUC
0.79
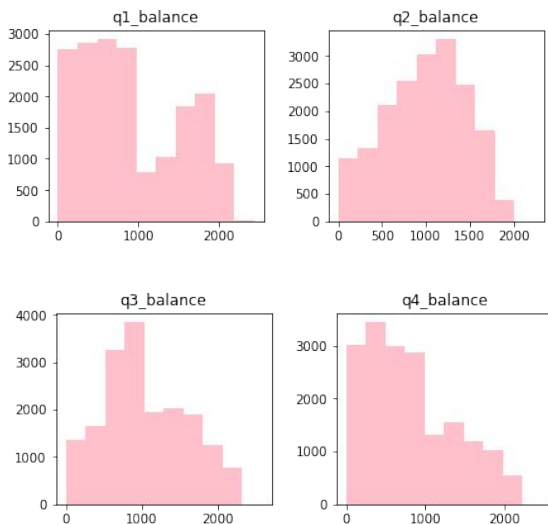
Method C

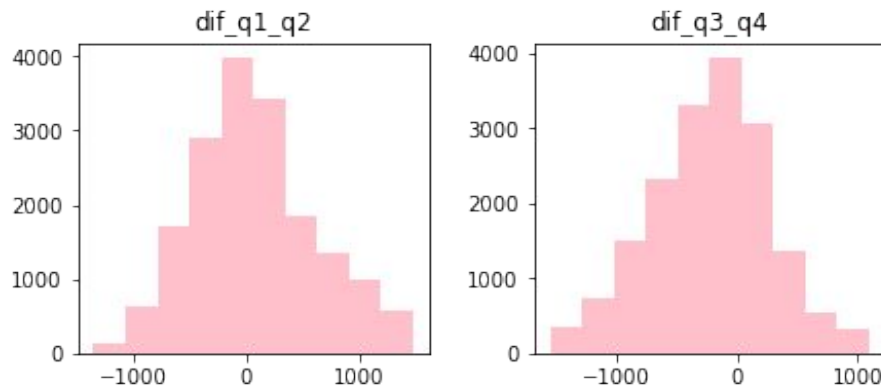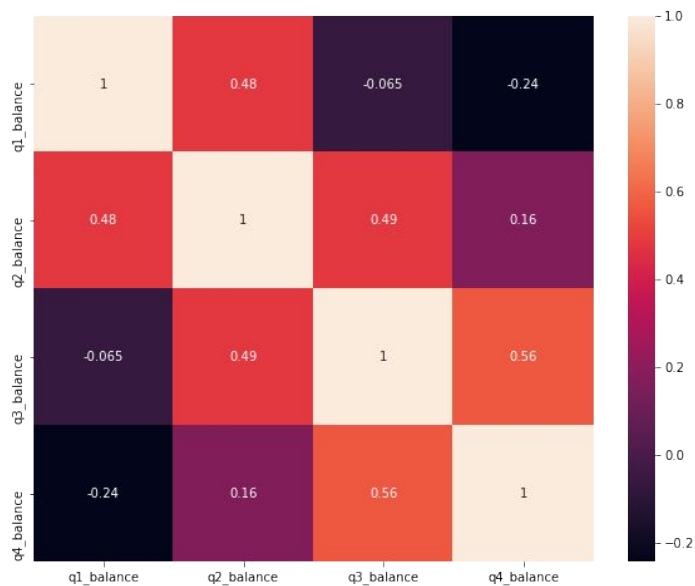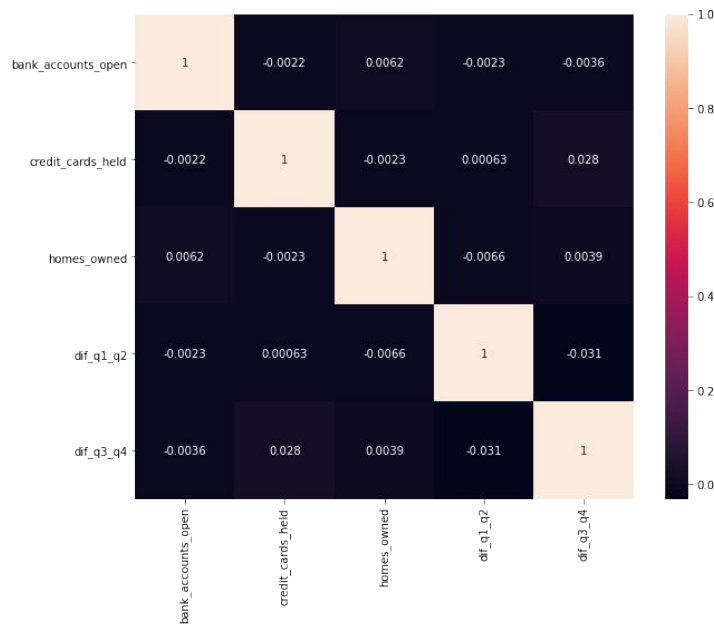Breaking down the quarterly balances into half yearly balances.

After

Method C

Before

After

Validate and Summarize

SMOTE + IQRs

AUC
0.78

Dropping
Columns

Increase in
accuracy by
decimal points

Accuracy
72.32%
AUC
0.79

Binning - q-cuts

Wrangling the
quarterly
balances to half
yearly balances

Columns with better
Normal Distribution
No Collinearity

Accuracy
72.01%
AUC
0.79

Roadblock

A lack of information of what the client wants

What will we add more?

- Calculate the confusion matrix precision to chose the better hypothesis

- Maybe use one more model

Dolores Umbridge

"Oh, you won't need ink"

THANK YOU