Retail Credit Scoring

Commercial Banking, Corp.

RFP #: RC - S1.H1

February 1, 2019

**Orange Team #1**

Morgan Groves, Bill Jenista, Mia Wu, Pierce Secola & Dave Hiltbrand

# Executive Summary

This report is a response to Commercial Banking, Corp’s (hereafter the Company) request for Proposal RC – S1.H1 for analytics services to create a scorecard and mapping system to support its banking services. Using the information in the RFP and the data provided, the analytics team developed a scorecard that suggests a single cutoff score of 500 and both decreases the event rate to 2.4% and increases the acceptance rate to 78%. Furthermore, the team recommends the Company make future decisions based on the results from the scorecard.

# 

# Background

Lending institutions mitigate risks by offering resources based on information provided by the customer or gathered by the firm. Our analytics team develops credit scorecards to improve clients’ returns on lending risks. We build a mathematical model that takes information about potential customers such as applicant’s income and oldest line of credit. Using advanced techniques we create new variables that contain the greatest information value in predicting the probability that the applicant will default on the loan. Our goal is to use data provided by a firm using default and acceptance rates for loans they have already issued. We also infer the possibility that an applicant who was originally denied a loan would have defaulted to build a more robust dataset. Using this methodology we successfully ingest a customer profile and return a score which can be used to help banks and lenders determine if the risk presented by customer is outweighed by the return on interest.

# Analysis

After thorough research, our team decided to build the scorecard and determine the credit score cut off point for the department of Revolving Line of Credit in SAS enterprise miner. The team was given two datasets to create the scorecard. One which included information about 3000 previous customers who were accepted for credit cards. There were 25 original input variables and a target variable which indicates if a customer has defaulted an account before. The other dataset contained all the rejected applicants. To account for the different ratios between good and bad loans as well as the size of the accepted and rejected datasets, a weight variable was assigned in the dataset. In order to validate the performance of the scorecard, the accepted loans dataset was split into 70% training data and 30% validation data. Then, we implemented a binning technique to group variables. This technique also discovered the seven most important variables for the model based on their weights of evidence. Our data was now prepared to build a scorecard.

Using guidelines in the proposal to build the initial scorecard, we assigned a score of 500 to applicants with odds-ratio 20/1 and let the doubling the odds be associated with a change of 50 points. In addition, we accounted for the existing acceptance rate of 75%, current event rate of 3.23%, expected revenue of accepted good customers, and the expected cost of accepted bad customers into this scorecard. The area under the curve of this initial scorecard model is 0.724, which indicates our scorecard accurately predicted 72% of the defaults in the validation data. Based on the trade-off chart, the optimal cut off point for a credit score is 516 for the maximization of profit. To make sure that our scorecard has complied with the FDIC, the rejected dataset was utilized as our reject inference and the fuzzy inference was chosen by our team to remove the bias resulting from exclusion of rejects.

After implementing the reject inference technique our final scorecard was built using both the accepted and rejected datasets. We again partitioned the data with a 70% training and 30% validation split. The same seven input variables were included in the model and the area under the curve statistic is 0.731 which indicates similar prediction performance to the original scorecard. The optimal cut-off point for credit score has decreased to 500. This can be gleaned in the trade off plots from Figure 1 below. At credit score of 500, the cumulative event rate is 0.83% lower than the previous one and the approval rate is increased by 3% of the previous rate.

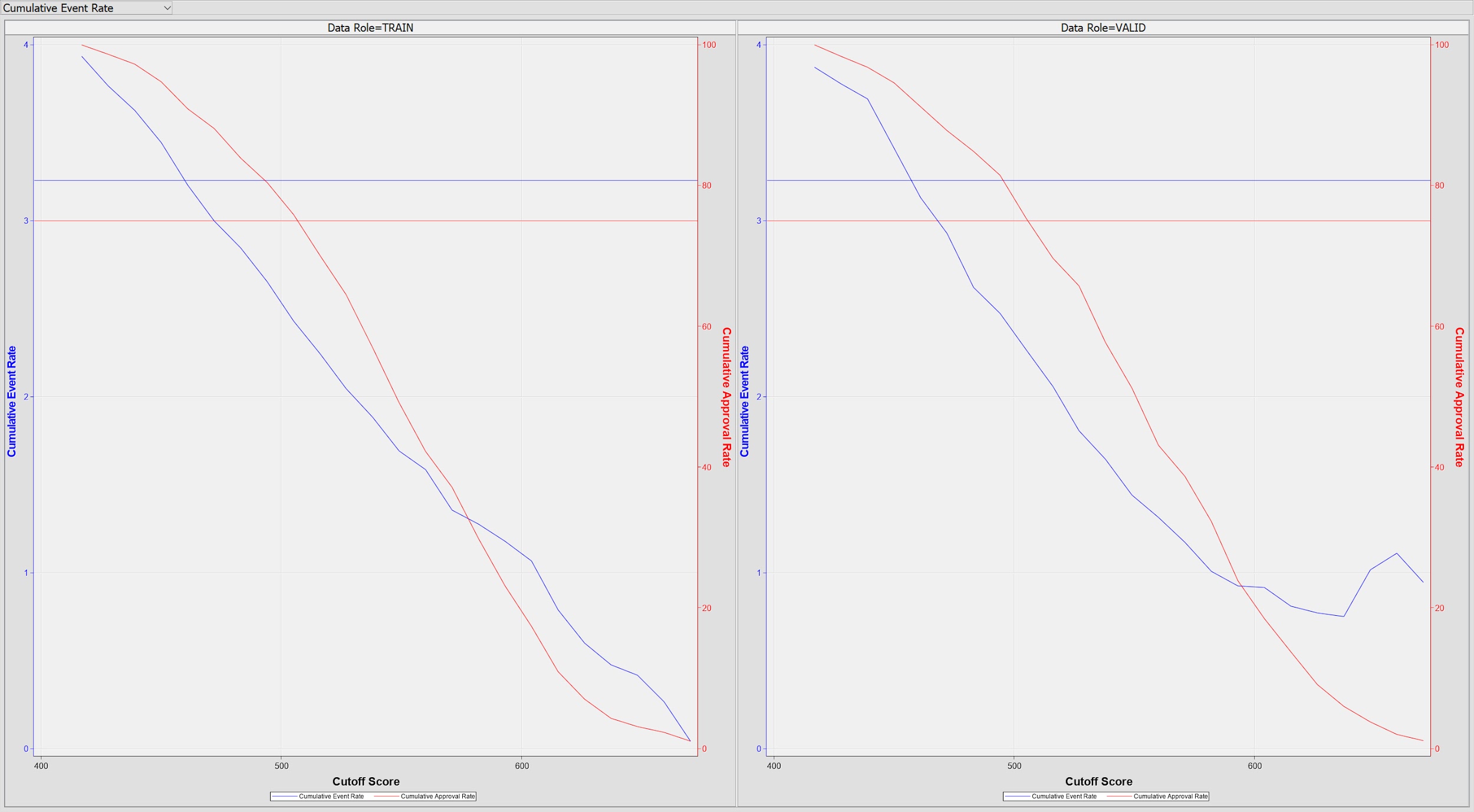
****

Figure 1 - Trade off plots for training data on the left and validation data on the right showing acceptance and event rates.

# Conclusion

While the preexisting acceptance and event rate currently deployed by the Company is satisfactory, the analytics team is proud to present a new scorecard that the Company can utilize immediately. Through the implementation of the scorecard, the Company will have higher acceptance rates paired with lower default at one clearly defined cutoff score. At the end of the day our methodology will allow the bank to increase their customer base while safely mitigating risks.

# Appendix

A.1 - Scorecard output from SAS E-Miner

