CredX Operational Acquisition Risk Analytics

PGDDS-Jul-2017 Batch - Capstone Projects: BFSI

Sabyasachi Sahoo -	DDA1720194
Pradosh Kumar Jena -	DDA1720238
Rahul Mukerjee -	DDA1720171
Chandrapal Panwar -	DDA1720177

Background & Objective

Background

- CredX is a leading credit card provider that gets thousands of credit card applicants every year. But in the past few years, it has experienced an increase in credit loss.
- The CEO believes that the best strategy to mitigate credit risk is to acquire "the right customers".

Objective

- The objective is to help CredX identify the right customers using predictive models. We need to determine the factors affecting credit risk and create strategies to mitigate the acquisition risk and assess the financial benefit of your project.
- We need to assess and explain the potential financial benefit of our project to the management of the bank and identify the metrics we are trying to optimize, explain how the analysis and the model works, and share the results of the model.
- Build an application scorecard and identify the cut-off score below which you would not grant credit cards to applicants.

Problem Solving Methodology – Analysis Flow

Data Preparation

1. Merge Demographic and Credit Bureau Data on Applicant ID
2. Remove Duplicate Rows
3. Explore data by Univariate and Multivariate Analysis
4. Check NAs and NANs
5. Outlier Treatment

IV Analysis and WOE

1. Compute **Information** Value
2. Populate features with their WOE values
3. Create data frames with and without WOE Values

Model Building

1. Split data into train and test
2. Use SMOTE for balancing data
3. Build Models using Logistic Regression, Decision Tree, Random Forest, SVM

Model Evaluation

16. Check Model on Test Data
17. Evaluate Model Metrics

16. Build Application Score Card from Logistic Regression Model
17. Build Financial Strategies using the chosen model

Data Understanding-Demographic and Credit Bureau Data

Demographic Data

- This information provided by the applicants at the time of credit card application
- It contains customer-level information on age, gender, income, marital status, etc.

- Age
- Gender
- Income
- Marital Status
- Education
- Profession
- Number of Dependents

Credit Bureau Data

- It provides an entire ABC of past transaction history (for example, previous loan status ,credit card history, default status) of prospects
- The credit bureau data has information on the Payment behavior(number of times payment delays),outstanding balance, credit utilization.

- Outstanding Balance
- Past Due 30,60,90 DPD
- Average Credit Card Utilization
- Total Number of Trades
- Presence of Home Loan
- Number of Inquiries

Assumptions & Constraints

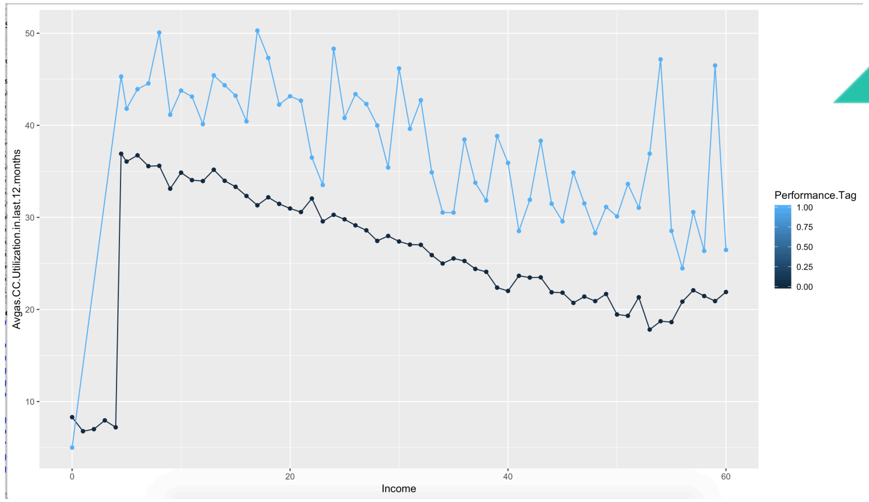
Assumptions

i.The missing data/outliers/invalid data treatment has been done either by replacing median values/limit values.
ii.Rejected application data is only used for scorecard verification. NOT for EDA/ modelling.
iii. It is assumed that all of the outstanding balance for the defaulted users would be lost and attributed as credit loss.

Constraints

i. The data is highly imbalanced. Only 4.2% of total data is about the defaulters.
We would use various sampling technique i.e. SMOTE while generating the models.

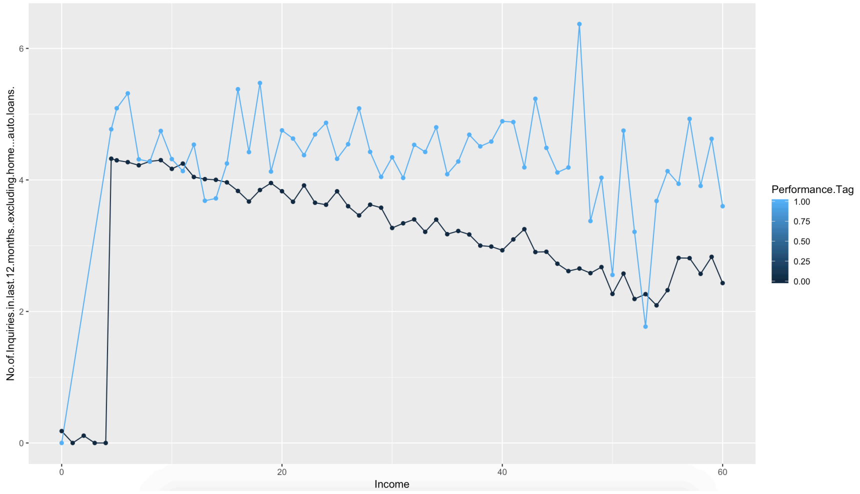
Insights derived by EDA

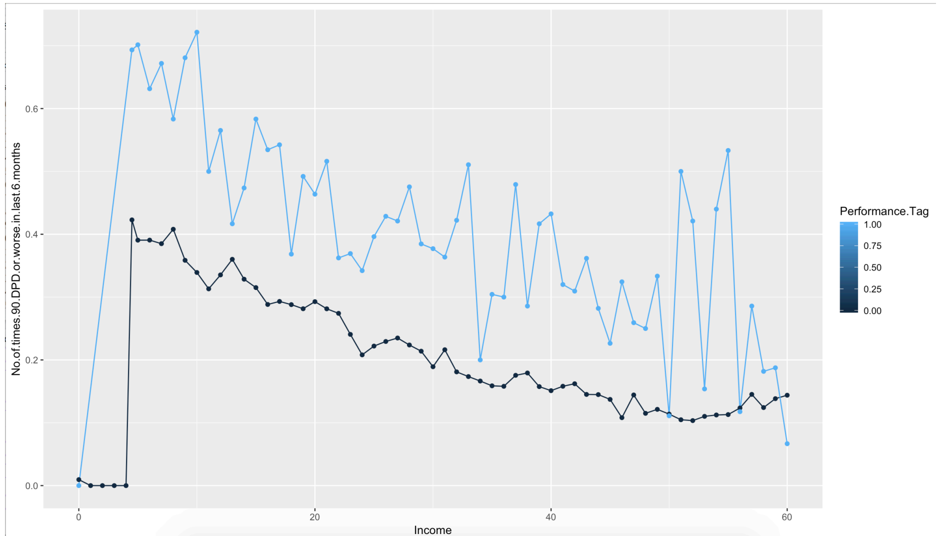


Income increasing

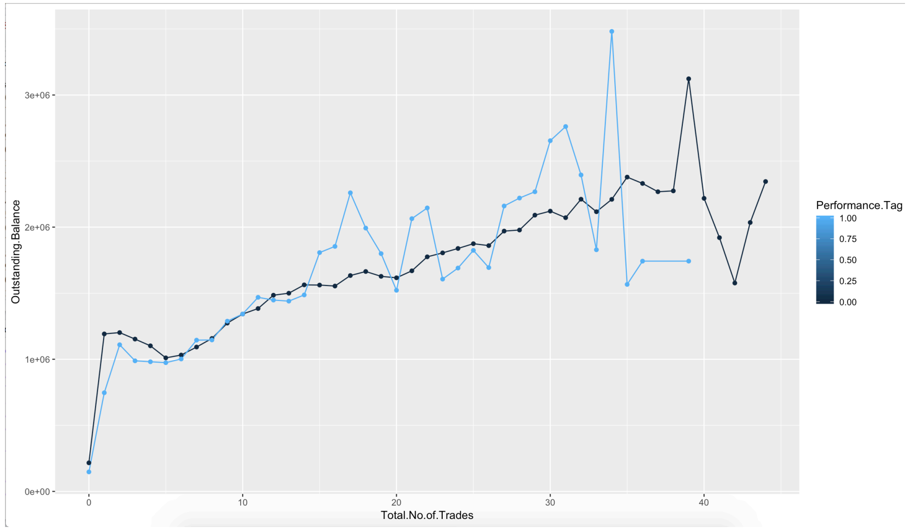
Avg-cc-utilisation decreasing.
No.of loan inquiries decreasing.
No of 90days delays decreasing.

Lower chance of default



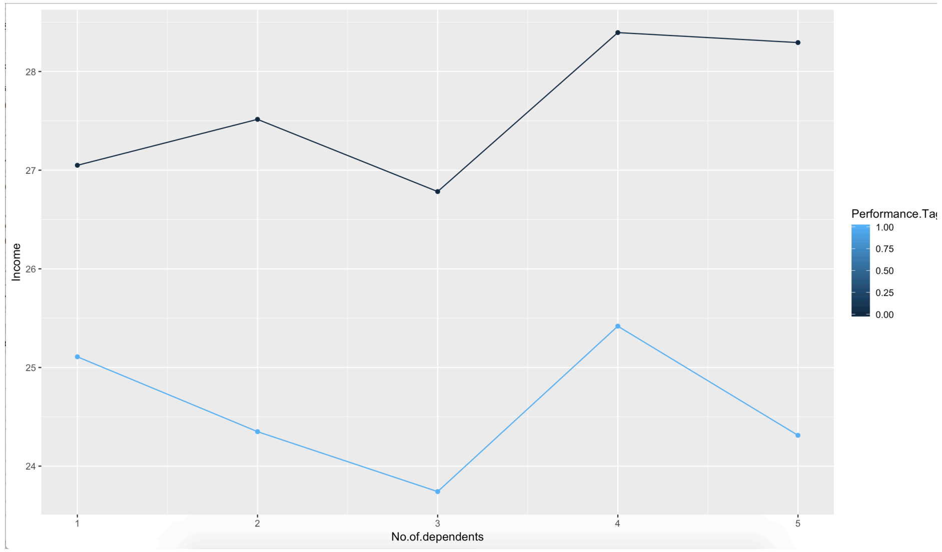


Insights derived by EDA



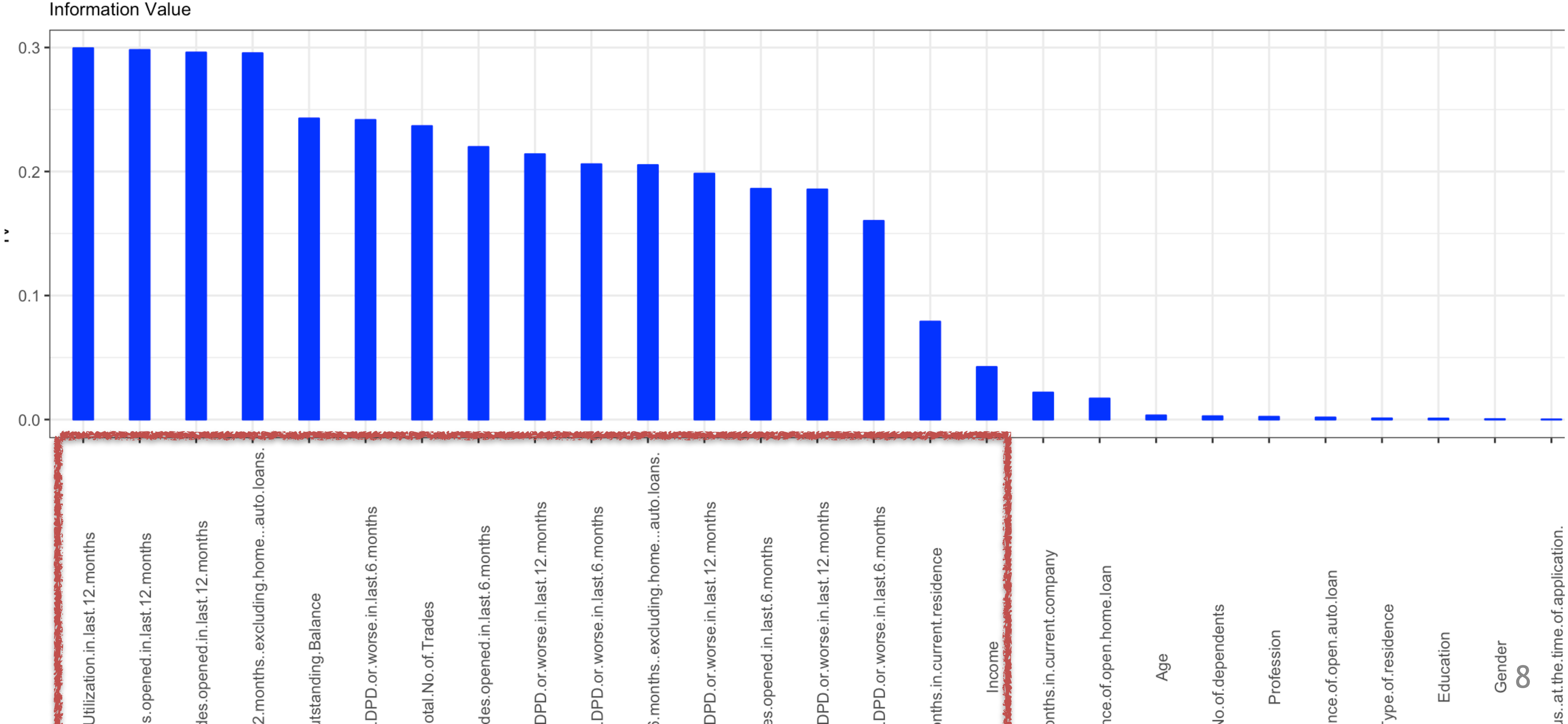
Higher outstanding balance
Higher total no of trades

Higher chance of default

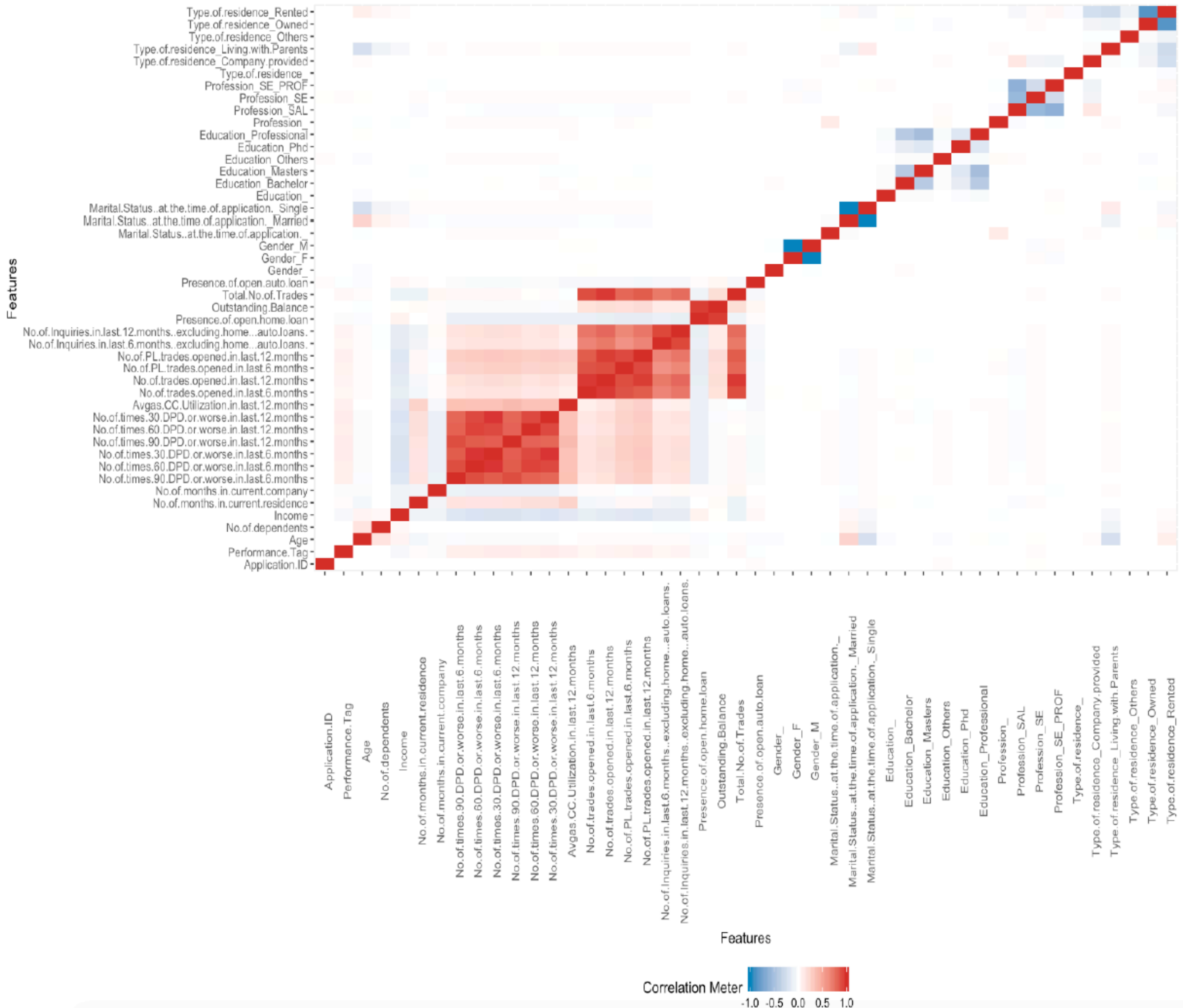


Lower the income per No. of dependants

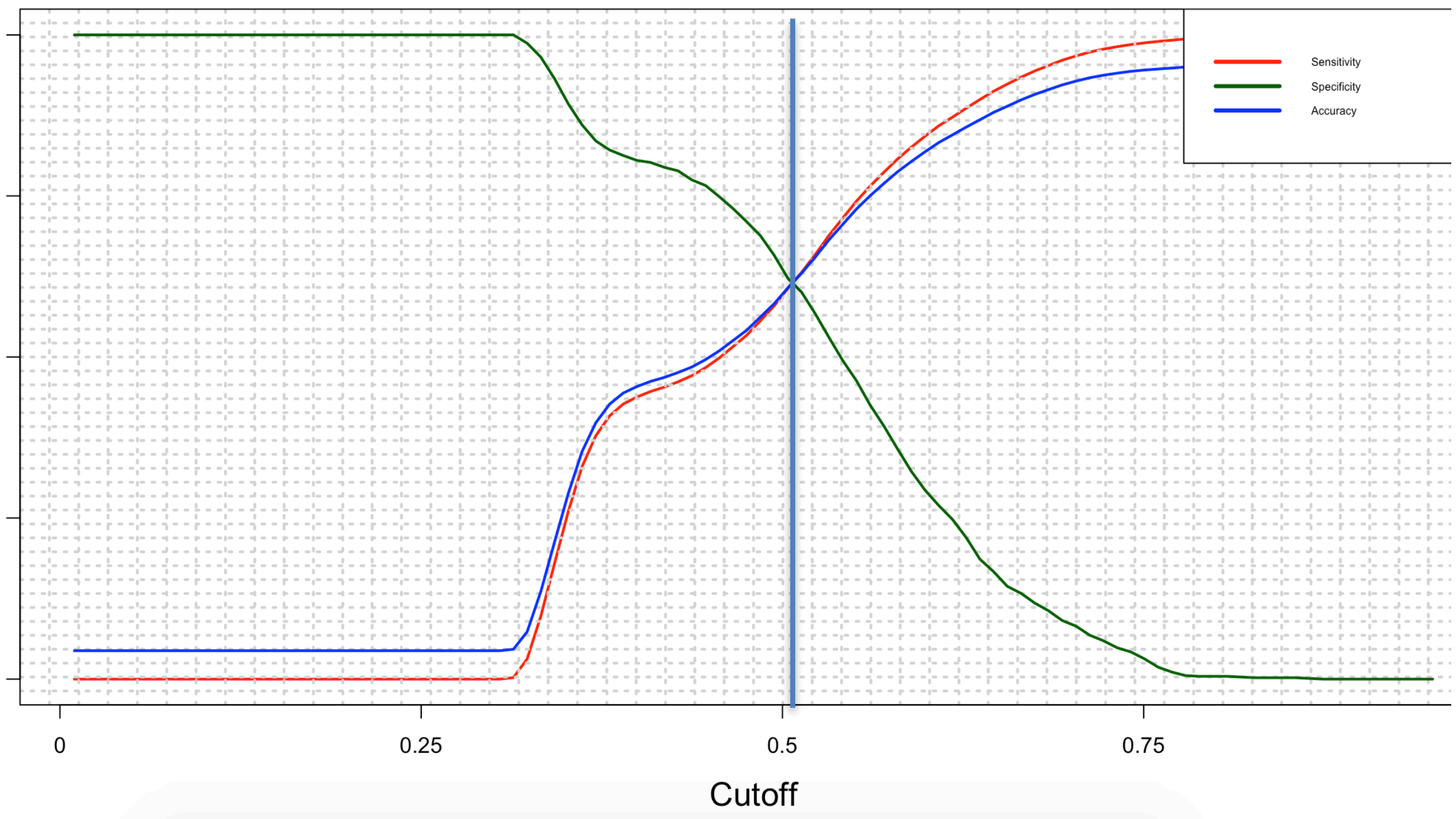
Findings from WOE/IV Analysis



Findings from Correlational analysis



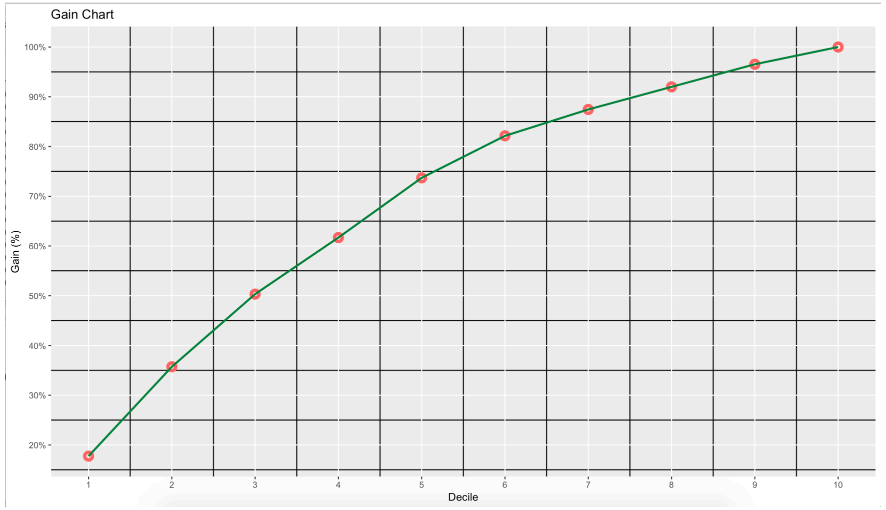
Logistic Model and Results

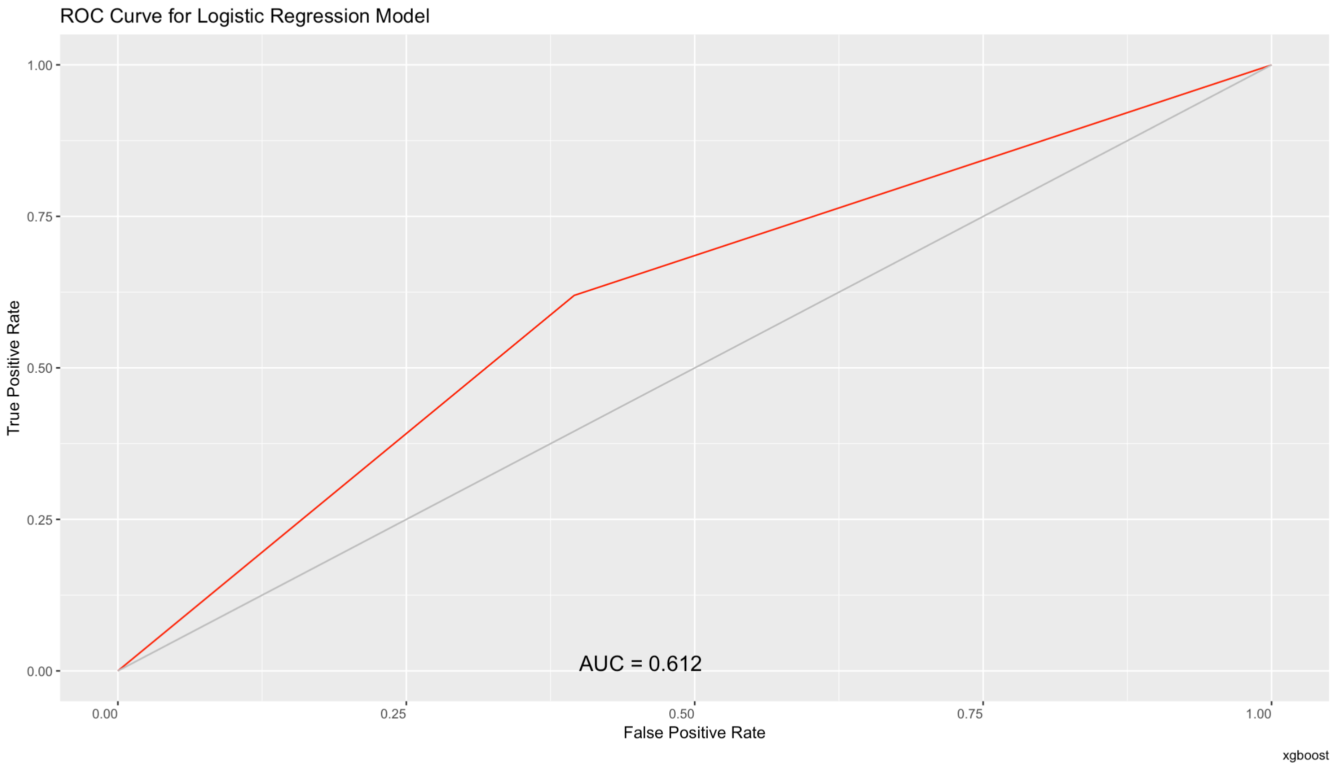


Logistic Regression Model Evaluation	
Evaluation Metric	**Value**
Optimal cutoff probability	0.502
Accuracy	60.52%
specificity	61.96%
sensitivity	60.45%
KS stat	22.43%
Area under curve	61.20%

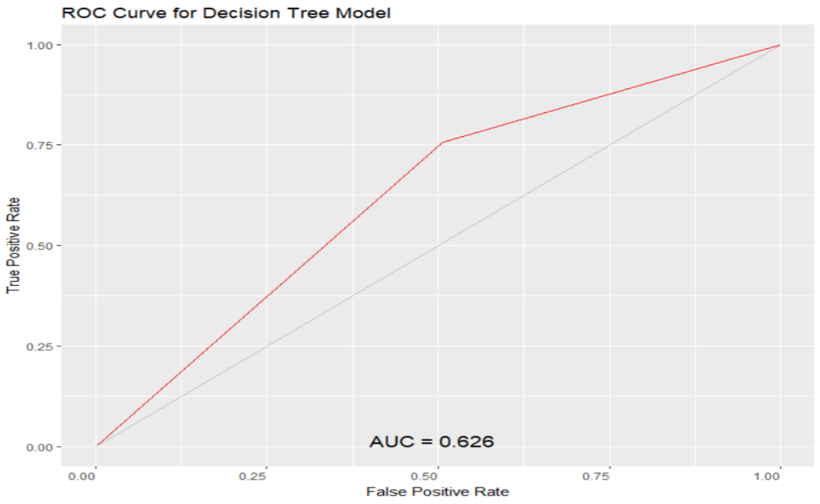
Logistic Model and Results-II

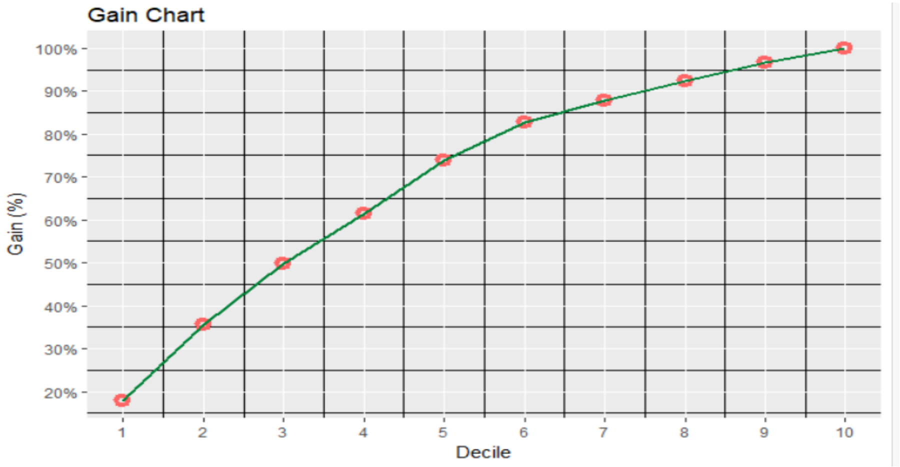
bucket	total	totalresp	Cumresp	Gain	Cumlift
1	2096	164	164	17.74891775	1.774891775
2	2096	166	330	35.71428571	1.785714286
3	2096	135	465	50.32467532	1.677489177
4	2096	105	570	61.68831169	1.542207792
5	2096	111	681	73.7012987	1.474025974
6	2096	78	759	82.14285714	1.369047619
7	2096	49	808	87.44588745	1.249226964
8	2096	42	850	91.99134199	1.149891775
9	2096	42	892	96.53679654	1.072631073
10	2095	32	924	100	1

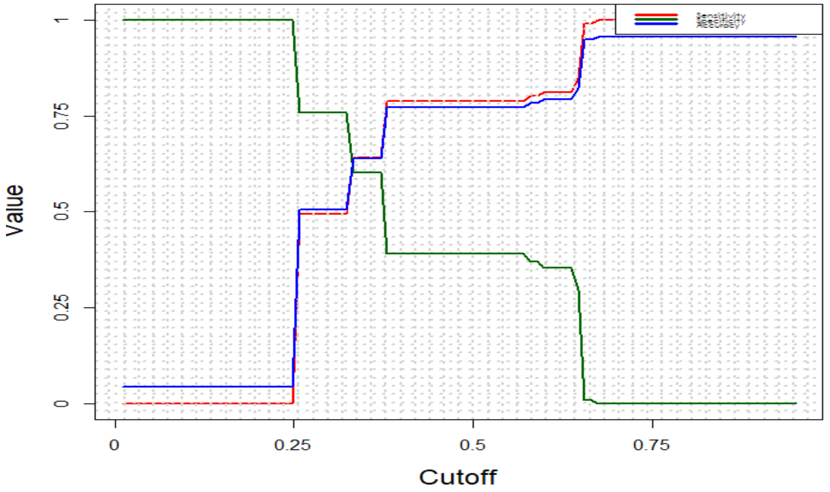




Decision Tree Model and Results

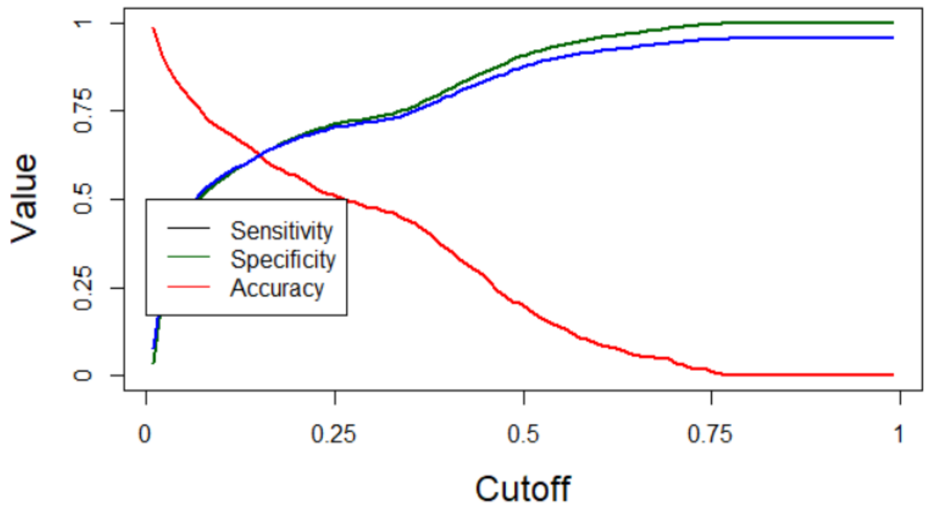


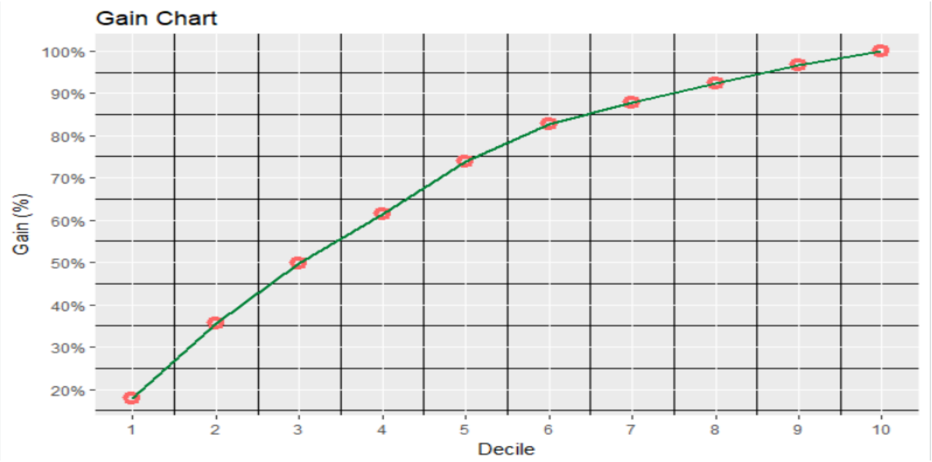


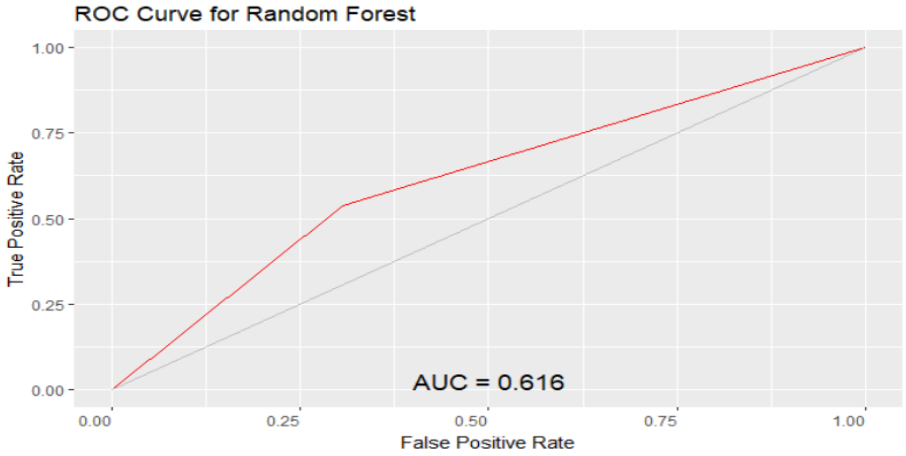



```
          Reference
Prediction     0     1
         0  9889   223
         1 10149   697
```

Random Forest Model and Results



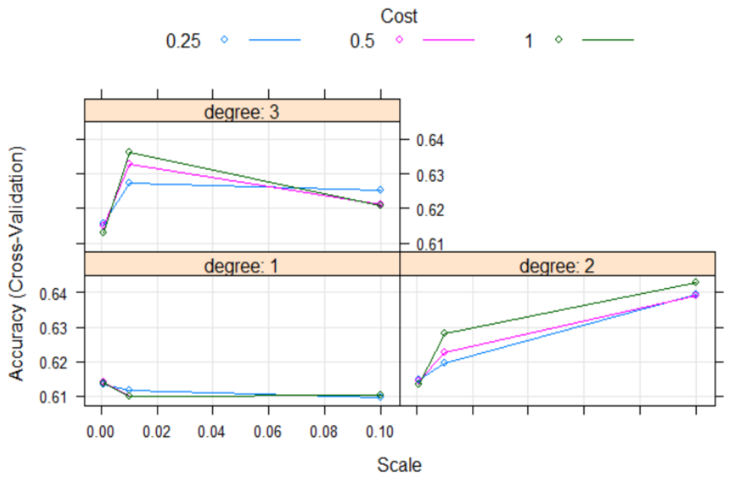




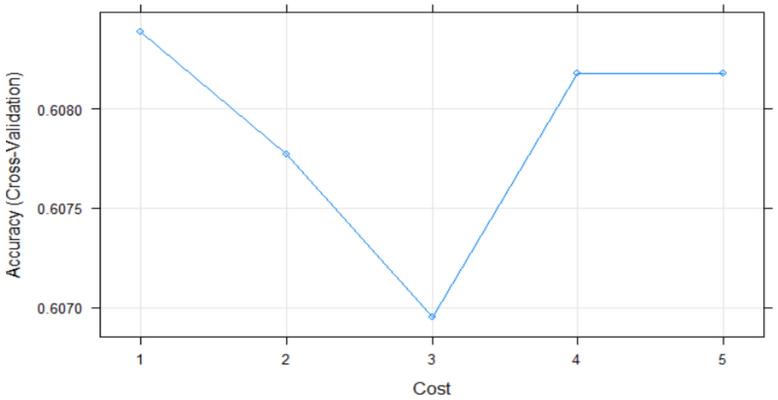
	Reference	
Prediction	no	yes
no	13879	424
yes	6159	496

SVM Model and Results

Polynomial Kernel

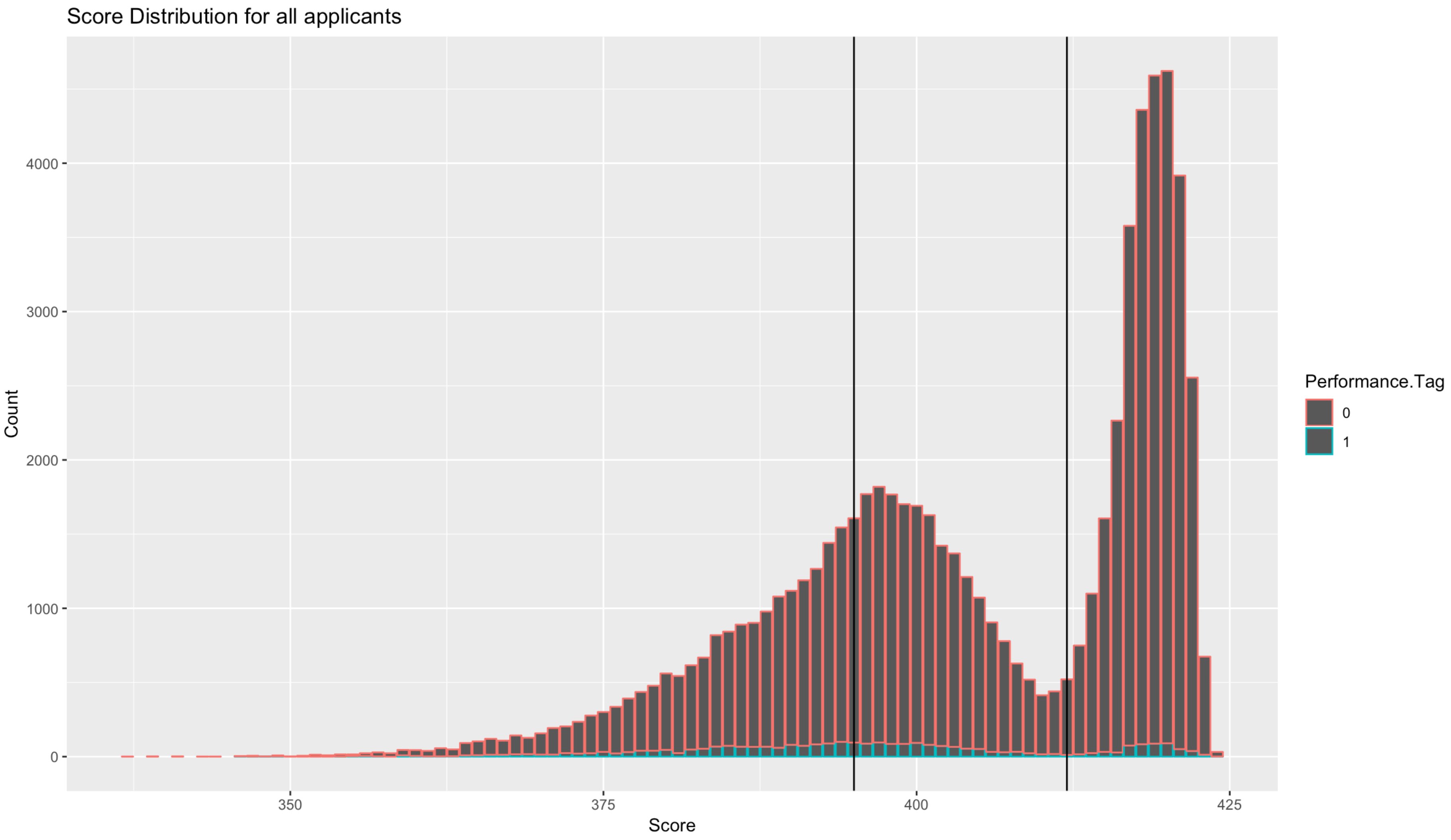


Linear Kernel



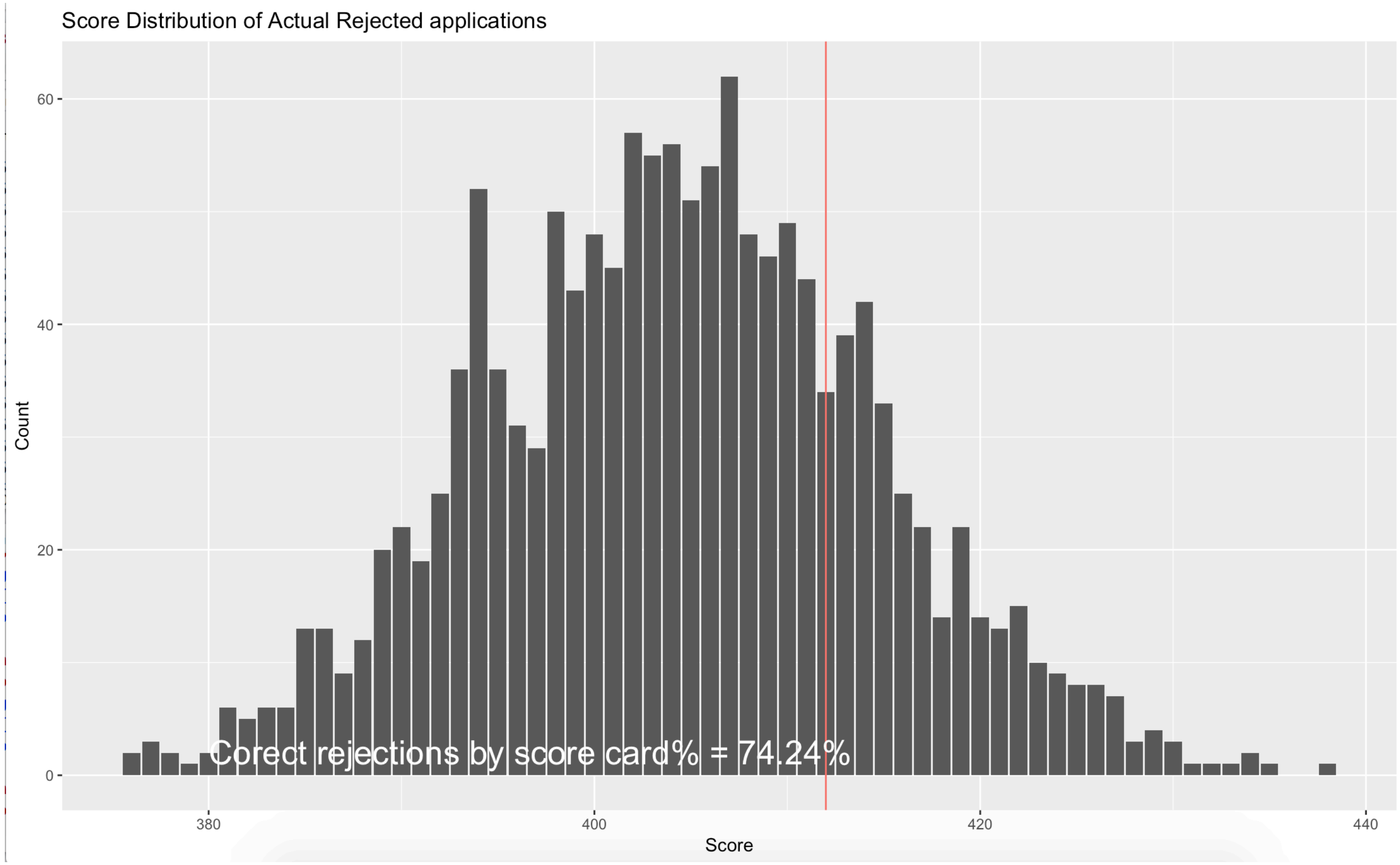
- We are doing this svm modelling on a small dataset due to resource crunch.
- The sensitivity , accuracy and specificity looks consistent in case of linear model
- Accuracy for polynomial kernel and radial kernel are too high but their specificity values are abysmal.

Application Scorecard



Percentile	Score
0	337
20	391
40	400
60	415
80	419
100	424

Application Scorecard-II



<u>Applying score card on Rejected Population :</u>

74% of actual rejected applicants detected

Financial Advantage

Objectives :

1. From P&L Perspective, the objective is to minimize "Net Credit Loss".

2. Scorecard is used for determining desired trade-off between risk level and approval rate.

Comparing results:

Scenario	Cut-off Score	Approval rate	Net Credit Loss
Without Scorecard - Current	NA	98%	3711178006.00
With Scorecard - conservative Cut off	412	43%	532840686.00
With Scorecard - Balanced Cut off	395	71%	1755465850.00

Conclusion

Final inferences made from the analysis are as below -

- With suggested optimal cut off score of 395, **avg. 71% of applicants would be approved**. Hence 29% applicants would be rejected.

- Assumption regarding credit loss- Outstanding balance of a defaulter is considered as credit loss for the specific user.

- **Potential credit loss avoided by applying the model/scorecard implementation - 52%.**