



## **Model #101: Credit Card Default Model**

### **Performance Validation Results**

## 1. The Production Model

The Logistic regression classifier is built using the variables previously identified as important by Random Forest algorithm according to various performance measures, F1, Gini Index and TPR <sup>1</sup>, and Recursive Partitioning and Regression (Rpart) and Boruta algorithms. The scoring method has been used to eliminate the useless features as follows, one point was granted to a variable if it is identified as a strong predictor by any classifier, -1 if the feature is rejected by any algorithm, and -0.5 if it is condemned as a tentative predictor (see Appendix B.). The points have been added up and the predictors with a score less than -1 are identified as weak predictors. As a result SEX, EDUCATION, OVER\_LIMIT, and UTIL\_PATTERN\_bin have been excluded from the further analysis. Automated stepwise variable selection method, that remove and enter predictors based on AIC values in a stepwise manner until there is no variable left to remove (enter) any, has been applied to the remaining predictors to obtain the final Logistic regression model. Based on the p-value for MARRIAGE coefficient it has been concluded that the variable is not statistically significant at an alpha level (statistical significance level) of 0.05, and thus has been removed from the model which did not compromise the classifier's predictive performance (see Appendix C.). The final model's coefficients<sup>2</sup> are presented in the Table 1. along with the estimated variance inflation factors (VIF). VIFs for Util\_1 and Util\_2 are larger than 10 suggesting that the variables are correlated but since it does not affect the predictive capabilities of the classifier, both features have been retained in the model.

---

<sup>1</sup> To see the feature importance plots, refer to Appendix A.

<sup>2</sup> Some independent variables have been replaced with their WOE scores that allowed to establish monotonic relationship to the dependent variable. Since the transformation is based on logarithmic value of distributions which is aligned with the logistic regression output function, WOE transformation is believed to improve the prediction results. The coefficients could not be interpreted per se but should be multiplied with WOE to obtain a score.

**Table 1: The Coefficient table. VIF values included**

|                   | Est.   | 2.5%   | 97.5%   | z val. | P      | VIF   |
|-------------------|--------|--------|---------|--------|--------|-------|
| (Intercept)       | -0.24  | -0.45  | -0.02   | -2.17  | 0.03   |       |
| Util_1            | 0.50   | 0.08   | 0.91    | 2.34   | 0.02   | 15.26 |
| Util_2            | 0.45   | 0.05   | 0.84    | 2.21   | 0.03   | 13.79 |
| Util_5            | -0.39  | -0.67  | -0.10   | -2.67  | 0.01   | 5.76  |
| MAX_Util_ratio    | -0.41  | -0.59  | -0.23   | -4.45  | 0.0000 | 2.73  |
| Avg_Pmt_Amt_tfm   | -0.10  | -0.13  | -0.07   | -6.18  | 0      | 3.08  |
| Max_Bill_Amt_tfm  | -0.01  | -0.02  | -0.001  | -2.24  | 0.02   | 7.31  |
| WOE_Freq_PAY      | -0.005 | -0.01  | -0.004  | -14.21 | 0      | 1.30  |
| WOE_Avg_Util      | -0.003 | -0.01  | -0.0001 | -2.02  | 0.04   | 4.17  |
| WOE_LIMIT_BAL     | -0.002 | -0.003 | -0.0003 | -2.37  | 0.02   | 2.18  |
| WOE_MAX_DLQ       | -0.01  | -0.01  | -0.01   | -27.90 | 0      | 1.41  |
| WOE_OVER_PMT      | -0.003 | -0.005 | -0.001  | -3.40  | 0.001  | 1.41  |
| WOE_REPAY_PATTERN | -0.004 | -0.01  | -0.002  | -3.28  | 0.001  | 1.33  |
| Pmt_Ratio_6       | 0.22   | 0.08   | 0.37    | 2.97   | 0.003  | 2.53  |

## 2. Model Development Performance

The Logistic Regression classifier produced similar results out-of-sample and in-sample (and for cross validation) which suggests that the model is stable (not overfitted to the training data) and should perform equally good on the unseen cases. The ability of the logistic model to discriminate between the reliable and risky customers have been measured using Kolmogorov–Smirnov test (KS) and Area under the ROC Curve (AUC). The model displays good discriminatory power on the train data, KS = 41.1% and AUC = 0.7653802 (see Table 1.1 and 1.2). The test results do not differ much from the results obtained on the training data set with KS equal to 40.9% and AUC of 0.7632956 (see Table 3.1 and 3.2).

The cutoff value is tuned to maximize F1 score<sup>3</sup> and realize the tradeoff between increasing the expected profit and minimizing the expected loss due to lending to risk default customers. The model is accurate given that the response variable is unbalanced, and threshold has been tuned to increase Recall. The accuracy obtained on the train data set is 73% (see Table 1.) and test accuracy is 74% (see Table 3.).

The resulted Recall values from the training and test data sets are 67% and 68% respectively. The model can detect 68% of delinquent accounts for the new data. Precision for the training data set is 43% and test data – 40%. The model can capture 40% of default customers for the new data. Bearing in mind that in credit scoring context the cost associated with False Negative is higher than Type I error, the obtained values for Precision and Recall are satisfactory. Overall, the quality of the model is good with tolerable Type I and Type II errors of 0.25 and 0.33 respectively produced on the training cases, and 0.25 and 0.32 on the test data set. F1 values of 0.53 and 0.52 for the seen and unseen observations indicate the model is good at realizing the established trade-off objective.

**Table 1.1.**

| Logistic Regression Model: Threshold tuned - 0.21. In-sample. |                 |       |        |       |              |                 |      |               |      |        |      |             |        |
|---|-----------------|-------|--------|-------|--------------|-----------------|------|---------------|------|--------|------|-------------|--------|
| Actual Class  | Predicted Class |       | Totals | KS    | Actual Class | Predicted Class |      | TP            | 0.67 | TP+TN  | 1.41 | AUC         | 0.7654 |
|   | 0               | 1     |        |       |              | 0               | 1    |               |      |        |      |             |        |
| 0   | 8787            | 2,970 | 11757  | 41.1% | 0            | 0.75            | 0.25 | Type I Error  | 0.25 | Recall | 0.67 | Specificity | 0.75   |
| 1   | 1,144           | 2,279 | 3423   |       | 1            | 0.33            | 0.67 | Type II Error | 0.33 | F1     | 0.53 | Accuracy    | 0.7290 |

<sup>3</sup>F1 score is the weighted average of Precision and Recall. Therefore, this score takes both false positives and false negatives into account. F1 is usually more useful than accuracy, especially in case of an uneven class distribution because it is sensitive to imbalances in data

**Table 1.2. Computed Lift chart and KS statistic for the train data set using half-deciles.**

| Decile | Obs    | Target<br>(Y=1) | NonTarget<br>(Y=0) | Target<br>Density | NonTarget<br>Density | Target<br>CDF | NonTarget<br>CDF | KS<br>Stat |
|--------|--------|-----------------|--------------------|-------------------|----------------------|---------------|------------------|------------|
| 1      | 759    | 496             | 263                | 14.5%             | 2.2%                 | 14.5%         | 2.2%             | 12.3%      |
| 2      | 759    | 451             | 308                | 13.2%             | 2.6%                 | 27.7%         | 4.9%             | 22.8%      |
| 3      | 759    | 336             | 423                | 9.8%              | 3.6%                 | 37.5%         | 8.5%             | 29.0%      |
| 4      | 759    | 293             | 466                | 8.6%              | 4.0%                 | 46.0%         | 12.4%            | 33.6%      |
| 5      | 759    | 302             | 457                | 8.8%              | 3.9%                 | 54.9%         | 16.3%            | 38.6%      |
| 6      | 759    | 237             | 522                | 6.9%              | 4.4%                 | 61.8%         | 20.7%            | 41.0%      |
| 7      | 759    | 173             | 586                | 5.1%              | 5.0%                 | 66.8%         | 25.7%            | 41.1%      |
| 8      | 759    | 142             | 617                | 4.1%              | 5.2%                 | 71.0%         | 31.0%            | 40.0%      |
| 9      | 759    | 146             | 613                | 4.3%              | 5.2%                 | 75.3%         | 36.2%            | 39.1%      |
| 10     | 759    | 130             | 629                | 3.8%              | 5.4%                 | 79.1%         | 41.5%            | 37.5%      |
| 11     | 759    | 106             | 653                | 3.1%              | 5.6%                 | 82.2%         | 47.1%            | 35.1%      |
| 12     | 759    | 95              | 664                | 2.8%              | 5.6%                 | 84.9%         | 52.7%            | 32.2%      |
| 13     | 759    | 94              | 665                | 2.7%              | 5.7%                 | 87.7%         | 58.4%            | 29.3%      |
| 14     | 759    | 89              | 670                | 2.6%              | 5.7%                 | 90.3%         | 64.1%            | 26.2%      |
| 15     | 759    | 86              | 673                | 2.5%              | 5.7%                 | 92.8%         | 69.8%            | 23.0%      |
| 16     | 759    | 69              | 690                | 2.0%              | 5.9%                 | 94.8%         | 75.7%            | 19.1%      |
| 17     | 759    | 54              | 705                | 1.6%              | 6.0%                 | 96.4%         | 81.7%            | 14.7%      |
| 18     | 759    | 48              | 711                | 1.4%              | 6.0%                 | 97.8%         | 87.7%            | 10.0%      |
| 19     | 759    | 38              | 721                | 1.1%              | 6.1%                 | 98.9%         | 93.9%            | 5.0%       |
| 20     | 759    | 38              | 721                | 1.1%              | 6.1%                 | 100.0%        | 100.0%           | 0.0%       |
| Totals | 15,180 | 3423            | 11,757             | 100.0%            | 100.0%               |               |                  |            |

**Table 2.**

| Logistic regression Champion Model: 10 - fold cross - validation<br>results. Stratification applied |      |           |      |             |      |
|---|------|-----------|------|-------------|------|
| TP  | 0.27 | TP+TN     | 1.23 | AUC         | 0.77 |
| TN  | 0.95 | Recall    | 0.27 | Sensitivity | 0.28 |
| Type I Error  | 0.05 | Precision | 0.62 | Specificity | 0.95 |
| Type II Error   | 0.73 | F1        | 0.52 | Accuracy    | 0.80 |

**Table 3.1.**

| Logistic Regression Model: Threshold tuned - 0.22. Out-of-sample. |                 |       |        |       |              |                 |      |               |      |           |      |             |        |
|---|-----------------|-------|--------|-------|--------------|-----------------|------|---------------|------|-----------|------|-------------|--------|
| Actual Class  | Predicted Class |       | Totals | KS    | Actual Class | Predicted Class |      | TP            | 0.68 | TP+TN     | 1.43 | AUC         | 0.7633 |
|   | 0               | 1     |        |       |              | 0               | 1    | TN            | 0.75 | Precision | 0.40 | Sensitivity | 0.68   |
| 0   | 4,815           | 1,581 | 6,396  | 40.9% | 0            | 0.75            | 0.25 | Type I Error  | 0.25 | Recall    | 0.68 | Specificity | 0.75   |
| 1   | 499             | 1,058 | 1,557  |       | 1            | 0.32            | 0.68 | Type II Error | 0.32 | F1        | 0.50 | Accuracy    | 0.7385 |

**Table 3.2. Computed Lift chart and KS statistic for the test data set using half-deciles.**

| Decile | Obs   | Target<br>(Y=1) | NonTarget<br>(Y=0) | Target<br>Density | NonTarget<br>Density | Target<br>CDF | NonTarget<br>CDF | KS<br>Stat |
|--------|-------|-----------------|--------------------|-------------------|----------------------|---------------|------------------|------------|
| 1      | 367   | 228             | 139                | 14.6%             | 2.4%                 | 14.6%         | 2.4%             | 12.2%      |
| 2      | 366   | 207             | 159                | 13.3%             | 2.8%                 | 27.9%         | 5.2%             | 22.8%      |
| 3      | 366   | 147             | 219                | 9.4%              | 3.8%                 | 37.4%         | 9.0%             | 28.4%      |
| 4      | 326   | 125             | 201                | 8.0%              | 3.5%                 | 45.4%         | 12.5%            | 33.0%      |
| 5      | 386   | 131             | 255                | 8.4%              | 4.4%                 | 53.8%         | 16.9%            | 36.9%      |
| 6      | 386   | 128             | 258                | 8.2%              | 4.5%                 | 62.0%         | 21.3%            | 40.7%      |
| 7      | 366   | 80              | 286                | 5.1%              | 5.0%                 | 67.2%         | 26.3%            | 40.9%      |
| 8      | 366   | 63              | 303                | 4.0%              | 5.3%                 | 71.2%         | 31.6%            | 39.7%      |
| 9      | 366   | 60              | 306                | 3.9%              | 5.3%                 | 75.1%         | 36.9%            | 38.2%      |
| 10     | 366   | 71              | 295                | 4.6%              | 5.1%                 | 79.6%         | 42.0%            | 37.7%      |
| 11     | 367   | 41              | 326                | 2.6%              | 5.7%                 | 82.3%         | 47.6%            | 34.6%      |
| 12     | 366   | 42              | 324                | 2.7%              | 5.6%                 | 85.0%         | 53.3%            | 31.7%      |
| 13     | 366   | 51              | 315                | 3.3%              | 5.5%                 | 88.2%         | 58.7%            | 29.5%      |
| 14     | 366   | 42              | 324                | 2.7%              | 5.6%                 | 90.9%         | 64.3%            | 26.6%      |
| 15     | 366   | 29              | 337                | 1.9%              | 5.8%                 | 92.8%         | 70.2%            | 22.6%      |
| 16     | 366   | 35              | 331                | 2.2%              | 5.7%                 | 95.1%         | 75.9%            | 19.1%      |
| 17     | 366   | 27              | 339                | 1.7%              | 5.9%                 | 96.8%         | 81.8%            | 15.0%      |
| 18     | 366   | 16              | 350                | 1.0%              | 6.1%                 | 97.8%         | 87.9%            | 9.9%       |
| 19     | 366   | 16              | 350                | 1.0%              | 6.1%                 | 98.8%         | 93.9%            | 4.9%       |
| 20     | 367   | 18              | 349                | 1.2%              | 6.1%                 | 100.0%        | 100.0%           | 0.0%       |
| Totals | 7,323 | 1557            | 5,766              | 100.0%            | 100.0%               |               |                  |            |

The threshold could be further adjusted based on a company’s growth goals and risk appetite. The model’s in -sample and out-of-sample performances in terms of Precision and Recall and FPR and TPR for various cutoff values are illustrated with Precision – Recall (see Figures 2. and 4) and ROC curves (see Figures 1 and 3.).

**Figure 1.**

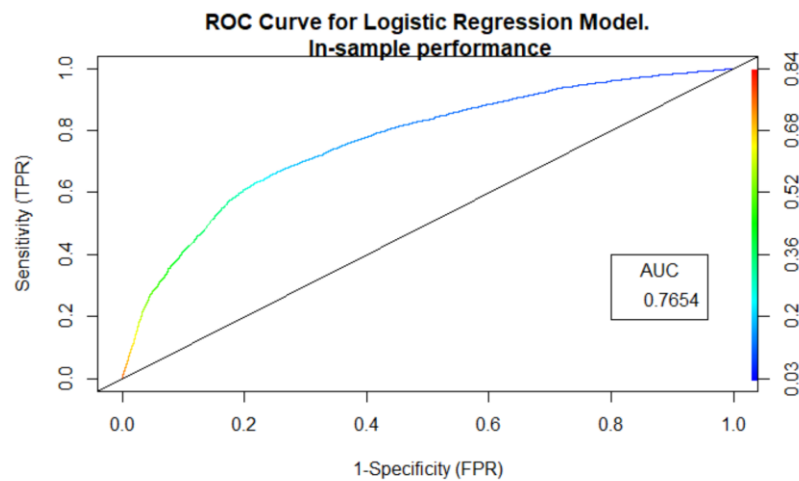


Figure 2.

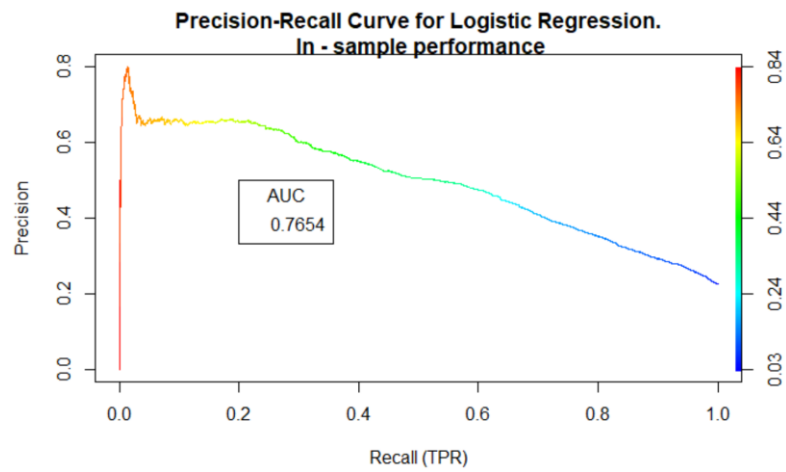


Figure 3.

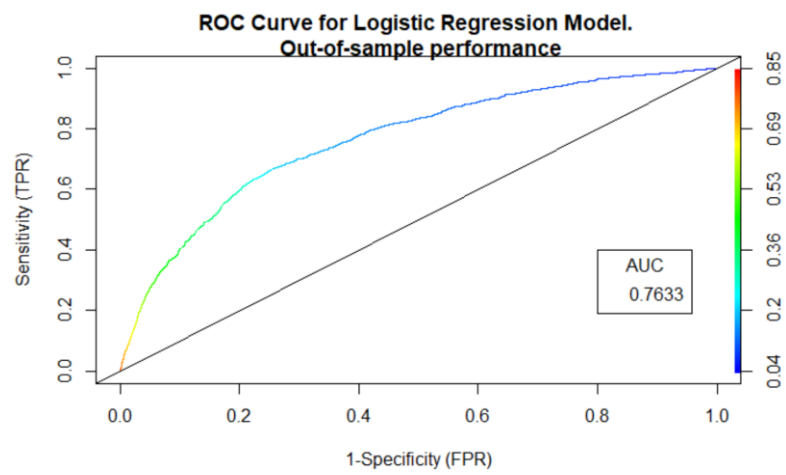
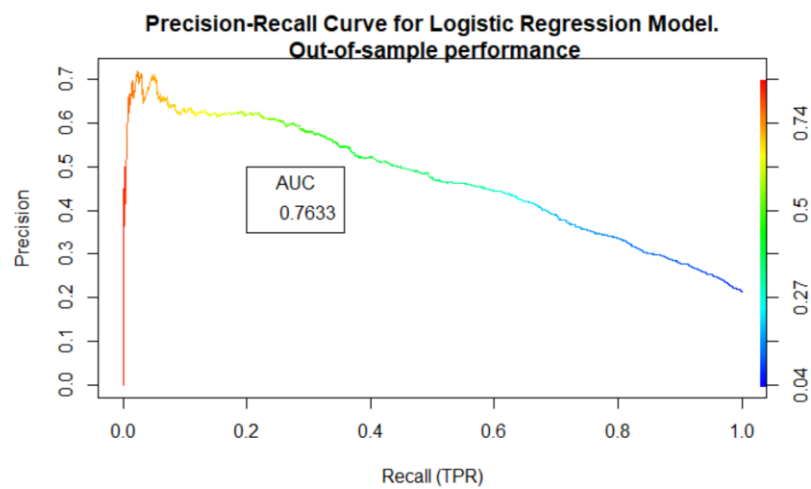


Figure 4.



### 3. Performance Monitoring Plan

Since the changes in external conditions might lead to deterioration in the model's performance the following monitoring plan has been developed to supervise and maintain accuracy and efficiency of the model (see Table 4.).

**Table 4. Recommended Actions based on the KS value.**

| KS (%)  | Status | Action  |
|---|--------|---|
| >30   | Green  | NO ACTION REQUIRED<br>(Model is performing as expected) |
| 20 - 30   | Amber  | Model needs to be re-validated<br>in three months       |
| < 20  | Red    | Model needs redevelopment                               |
| <i>Model will be re-validated at the standard interval of six months.</i> |        |   |

### 4. Performance Monitoring Results

Within the model monitoring framework, the model's performance has been evaluated on the validation data set<sup>4</sup>. The obtained KS statistic of 43.9% indicates the GREEN model's performance status (see Table. 4.) and suggests that the model is performing as expected. Thus, no action is required at this moment, and the model will be re-validated in six months per established model reevaluation schedule.

---

<sup>4</sup> The validation data set consists of 1747 observations. A-priori probabilities of the response: 0 - 0.7791116; 1 - 0.2208884. The summary statistics of the data set is presented in Appendix D.



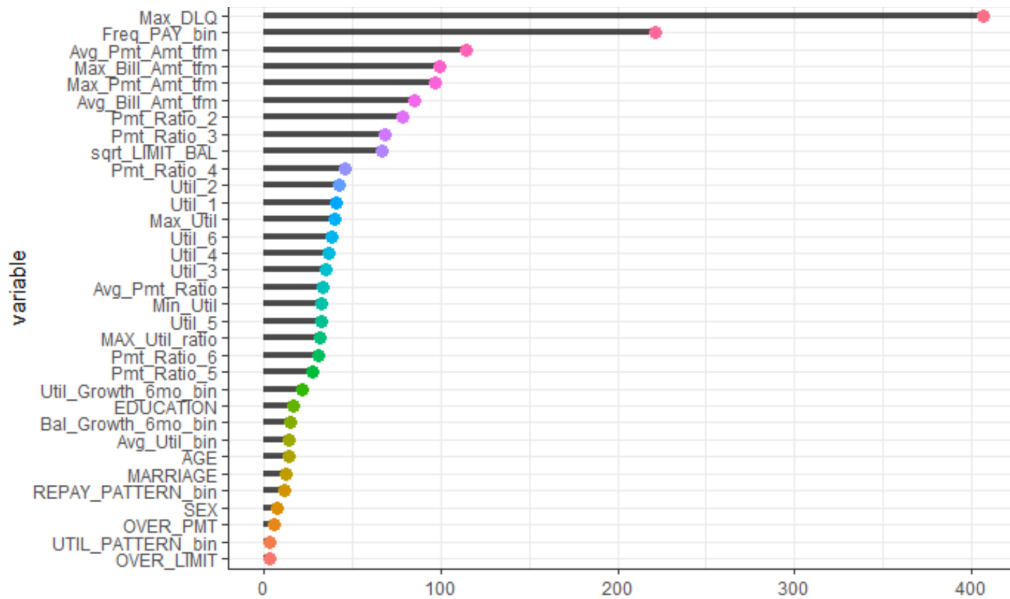
**Table 5. Computed Lift chart and KS statistic for the validation data set using half-deciles.**

| Decile        | Obs          | Target<br>(Y=1) | NonTarget<br>(Y=0) | Target<br>Density | NonTarget<br>Density | Target<br>CDF | NonTarget<br>CDF | KS<br>Stat |
|---------------|--------------|-----------------|--------------------|-------------------|----------------------|---------------|------------------|------------|
| 1             | 375          | 247             | 128                | 14.9%             | 2.2%                 | 14.9%         | 2.2%             | 12.7%      |
| 2             | 375          | 220             | 155                | 13.3%             | 2.7%                 | 28.2%         | 4.8%             | 23.4%      |
| 3             | 375          | 167             | 208                | 10.1%             | 3.6%                 | 38.3%         | 8.4%             | 29.9%      |
| 4             | 375          | 174             | 201                | 10.5%             | 3.4%                 | 48.8%         | 11.8%            | 36.9%      |
| 5             | 374          | 145             | 229                | 8.8%              | 3.9%                 | 57.5%         | 15.8%            | 41.8%      |
| 6             | 375          | 110             | 265                | 6.6%              | 4.5%                 | 64.2%         | 20.3%            | 43.9%      |
| 7             | 375          | 70              | 305                | 4.2%              | 5.2%                 | 68.4%         | 25.5%            | 42.9%      |
| 8             | 375          | 69              | 306                | 4.2%              | 5.2%                 | 72.6%         | 30.8%            | 41.8%      |
| 9             | 375          | 62              | 313                | 3.7%              | 5.4%                 | 76.3%         | 36.1%            | 40.2%      |
| 10            | 374          | 48              | 326                | 2.9%              | 5.6%                 | 79.2%         | 41.7%            | 37.5%      |
| 11            | 375          | 47              | 328                | 2.8%              | 5.6%                 | 82.1%         | 47.3%            | 34.7%      |
| 12            | 375          | 36              | 339                | 2.2%              | 5.8%                 | 84.2%         | 53.1%            | 31.1%      |
| 13            | 375          | 34              | 341                | 2.1%              | 5.8%                 | 86.3%         | 59.0%            | 27.3%      |
| 14            | 375          | 50              | 325                | 3.0%              | 5.6%                 | 89.3%         | 64.5%            | 24.8%      |
| 15            | 374          | 30              | 344                | 1.8%              | 5.9%                 | 91.1%         | 70.4%            | 20.7%      |
| 16            | 375          | 32              | 343                | 1.9%              | 5.9%                 | 93.1%         | 76.3%            | 16.8%      |
| 17            | 375          | 44              | 331                | 2.7%              | 5.7%                 | 95.7%         | 82.0%            | 13.8%      |
| 18            | 375          | 33              | 342                | 2.0%              | 5.9%                 | 97.7%         | 87.8%            | 9.9%       |
| 19            | 375          | 23              | 352                | 1.4%              | 6.0%                 | 99.1%         | 93.8%            | 5.3%       |
| 20            | 375          | 15              | 360                | 0.9%              | 6.2%                 | 100.0%        | 100.0%           | 0.0%       |
| <b>Totals</b> | <b>7,497</b> | <b>1656</b>     | <b>5,841</b>       | <b>100.0%</b>     | <b>100.0%</b>        |               |                  |            |

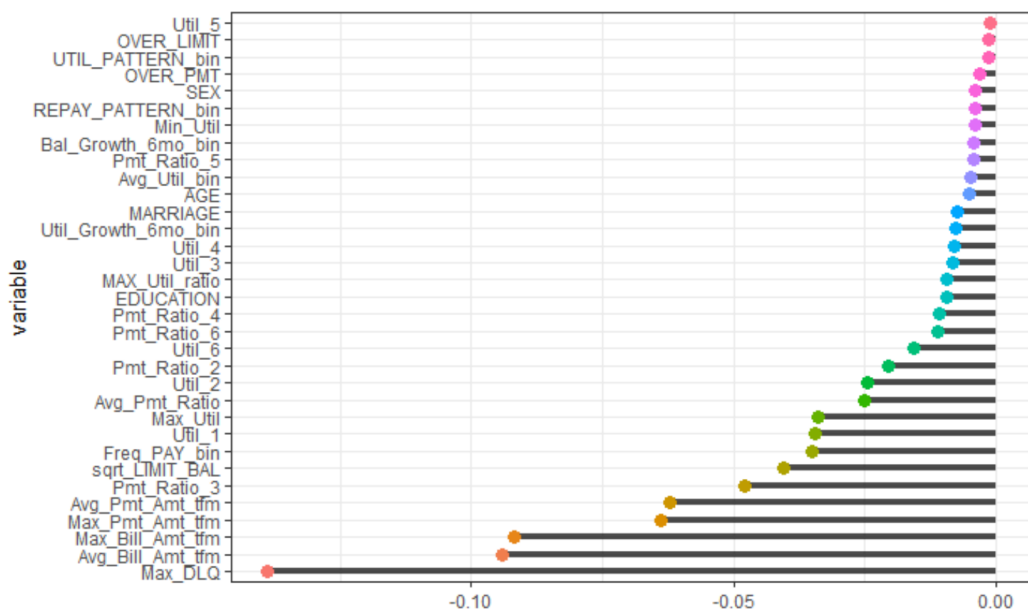
## Appendix A.

The Appendix includes the variable importance plots measured by Random Forest that show the importance of each feature according to given performance criteria F1, TPR and a mean decrease in Gini.

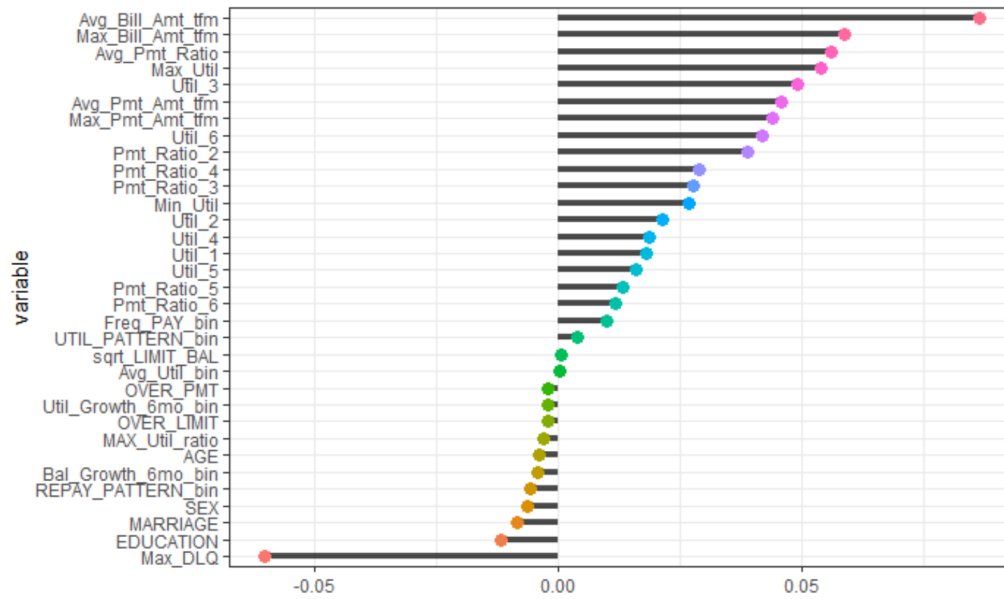
### Importance ranking of variables (Gini Index).



### Variable importance plot (F1 measure).



Variable importance plot (TPR).



## Appendix B.

The Appendix presents the table of aggregated feature importance scores according to various feature ranking techniques. The features with high positive and negative scores are highlighted which are indicators of variables' very strong and weak predictive power.

The scoring system: one point was granted to a variable if it is identified as a strong predictor by any classifier, -1 if the feature is rejected by any algorithm, and -0.5 if it is condemned as a tentative predictor

|                  | Rpart     | Random<br>Forest<br>F1 | Random<br>Forest<br>Gini Idx | Random<br>Forest<br>TPR | Boruto    | Negative<br>score | Positive<br>score | Total<br>Score |
|------------------|-----------|------------------------|------------------------------|-------------------------|-----------|-------------------|-------------------|----------------|
| sqrt_LIMIT_BAL   | Confirmed | Confirmed              | Confirmed                    | Rejected                | Confirmed | 1                 | 1                 | 0              |
| SEX              | Rejected  | Tentative              | Tentative                    | Confirmed               | Rejected  | 3                 | 0                 | -3             |
| EDUCATION        | Tentative | Rejected               | Confirmed                    | Confirmed               | Rejected  | 2                 | 0                 | -2             |
| MARRIAGE         | Tentative | Confirmed              | Confirmed                    | Confirmed               | Tentative | 1                 | 0                 | -1             |
| AGE              | Rejected  | Confirmed              | Confirmed                    | Confirmed               | Confirmed | 1                 | 0                 | -1             |
| Avg_Bill_Amt_tfm | Confirmed | Confirmed              | Confirmed                    | Confirmed               | Confirmed | 0                 | 3                 | 3              |
| Avg_Pmt_Amt_tfm  | Confirmed | Confirmed              | Confirmed                    | Confirmed               | Confirmed | 0                 | 3                 | 3              |
| Pmt_Ratio_2      | Confirmed | Confirmed              | Confirmed                    | Confirmed               | Confirmed | 0                 | 1                 | 1              |
| Pmt_Ratio_3      | Confirmed | Confirmed              | Confirmed                    | Confirmed               | Confirmed | 0                 | 2                 | 2              |
| Pmt_Ratio_4      | Confirmed | Confirmed              | Confirmed                    | Confirmed               | Confirmed | 0                 | 1                 | 1              |
| Pmt_Ratio_5      | Confirmed | Confirmed              | Confirmed                    | Confirmed               | Confirmed | 0                 | 1                 | 1              |
| Pmt_Ratio_6      | Confirmed | Confirmed              | Confirmed                    | Confirmed               | Confirmed | 0                 | 1                 | 1              |
| Avg_Pmt_Ratio    | Confirmed | Confirmed              | Confirmed                    | Confirmed               | Confirmed | 0                 | 1                 | 1              |
| Util_1           | Confirmed | Confirmed              | Confirmed                    | Confirmed               | Confirmed | 0                 | 1                 | 1              |
| Util_2           | Confirmed | Confirmed              | Confirmed                    | Confirmed               | Confirmed | 0                 | 0                 | 0              |
| Util_3           | Confirmed | Confirmed              | Confirmed                    | Confirmed               | Confirmed | 0                 | 1                 | 1              |
| Util_4           | Confirmed | Confirmed              | Confirmed                    | Confirmed               | Confirmed | 0                 | 0                 | 0              |
| Util_5           | Confirmed | Rejected               | Confirmed                    | Confirmed               | Confirmed | 1                 | 0                 | -1             |
| Util_6           | Confirmed | Confirmed              | Confirmed                    | Confirmed               | Confirmed | 0                 | 0                 | 0              |

|                     |           |           |           |           |           |     |   |      |
|---------------------|-----------|-----------|-----------|-----------|-----------|-----|---|------|
| Avg_Util_bin        | Confirmed | Confirmed | Confirmed | Tentative | Confirmed | 0.5 | 0 | -0.5 |
| OVER_LIMIT          | Tentative | Rejected  | Rejected  | Tentative | Confirmed | 3   | 0 | -3   |
| OVER_PMT            | Tentative | Confirmed | Tentative | Tentative | Confirmed | 1.5 | 0 | -1.5 |
| Bal_Growth_6mo_bin  | Confirmed | Confirmed | Confirmed | Confirmed | Confirmed | 0   | 0 | 0    |
| Util_Growth_6mo_bin | Tentative | Confirmed | Confirmed | Tentative | Confirmed | 1   | 0 | -1   |
| Max_Util            | Confirmed | Confirmed | Confirmed | Confirmed | Confirmed | 0   | 1 | 1    |
| Min_Util            | Confirmed | Confirmed | Confirmed | Confirmed | Confirmed | 0   | 0 | 0    |
| MAX_Util_ratio      | Confirmed | Confirmed | Confirmed | Confirmed | Confirmed | 0   | 0 | 0    |
| Max_Bill_Amt_tfm    | Confirmed | Confirmed | Confirmed | Confirmed | Confirmed | 0   | 3 | 3    |
| Max_Pmt_Amt_tfm     | Confirmed | Confirmed | Confirmed | Confirmed | Confirmed | 0   | 3 | 3    |
| Max_DLQ             | Confirmed | Confirmed | Confirmed | Confirmed | Confirmed | 0   | 4 | 4    |
| Freq_PAY_bin        | Confirmed | Confirmed | Confirmed | Confirmed | Confirmed | 0   | 3 | 3    |
| REPAY_PATTERN_bin   | Rejected  | Confirmed | Confirmed | Confirmed | Confirmed | 1   | 0 | -1   |
| UTIL_PATTERN_bin    | Tentative | Rejected  | Rejected  | Rejected  | Confirmed | 3.5 | 0 | -3.5 |

## Appendix C.

The coefficient table before MARRIAGE variable is removed from the final model.

**Table : The Coefficient table. VIF values included**

|                   | Est.   | 2.5%   | 97.5%  | z val. | p      | VIF   |
|-------------------|--------|--------|--------|--------|--------|-------|
| (Intercept)       | -0.20  | -0.63  | 0.23   | -0.93  | 0.35   |       |
| MARRIAGE1         | 0.06   | -0.32  | 0.44   | 0.31   | 0.75   | 1.03  |
| MARRIAGE2         | -0.12  | -0.50  | 0.26   | -0.64  | 0.52   | 1.03  |
| Util_1            | 0.50   | 0.08   | 0.91   | 2.34   | 0.02   | 15.27 |
| Util_2            | 0.45   | 0.05   | 0.85   | 2.22   | 0.03   | 13.80 |
| Util_5            | -0.38  | -0.66  | -0.10  | -2.63  | 0.01   | 5.76  |
| MAX_Util_ratio    | -0.42  | -0.60  | -0.24  | -4.52  | 0.0000 | 2.74  |
| Avg_Pmt_Amt_tfm   | -0.10  | -0.13  | -0.07  | -6.14  | 0      | 3.08  |
| Max_Bill_Amt_tfm  | -0.01  | -0.02  | -0.001 | -2.25  | 0.02   | 7.31  |
| WOE_Freq_PAY      | -0.005 | -0.01  | -0.004 | -14.13 | 0      | 1.30  |
| WOE_Avg_Util      | -0.002 | -0.005 | 0.0000 | -1.96  | 0.05   | 4.16  |
| WOE_LIMIT_BAL     | -0.002 | -0.004 | -0.001 | -2.69  | 0.01   | 2.20  |
| WOE_MAX_DLQ       | -0.01  | -0.01  | -0.01  | -27.82 | 0      | 1.41  |
| WOE_OVER_PMT      | -0.003 | -0.005 | -0.001 | -3.41  | 0.001  | 1.41  |
| WOE_REPAY_PATTERN | -0.004 | -0.01  | -0.002 | -3.19  | 0.001  | 1.33  |
| Pmt_Ratio_6       | 0.22   | 0.07   | 0.37   | 2.90   | 0.004  | 2.53  |

# Appendix D.

**Table : Summary Statistics for Validate data**

| Statistic           | N     | Mean       | St. Dev.   | Min         | Pctl(25) | Median    | Pctl(75)  | Max        |
|---------------------|-------|------------|------------|-------------|----------|-----------|-----------|------------|
| LIMIT_BAL           | 7,497 | 167,760.10 | 129,544.10 | 10,000      | 50,000   | 140,000   | 240,000   | 760,000    |
| Avg_Bill_Amt        | 7,497 | 45,526.43  | 63,458.11  | -8,867.33   | 4,598.50 | 21,167.83 | 58,398.33 | 503,890.00 |
| Avg_Pmt_Amt         | 7,497 | 5,273.71   | 9,415.89   | 0.00        | 1,125.00 | 2,416.67  | 5,620.33  | 194,217.80 |
| Pmt_Ratio_2         | 7,497 | 0.40       | 0.44       | 0           | 0.04     | 0.1       | 1         | 1          |
| Pmt_Ratio_3         | 7,497 | 0.41       | 0.44       | 0.00        | 0.05     | 0.11      | 1.00      | 1.00       |
| Pmt_Ratio_4         | 7,497 | 0.40       | 0.45       | 0.00        | 0.04     | 0.09      | 1.00      | 1.00       |
| Pmt_Ratio_5         | 7,497 | 0.40       | 0.45       | 0.00        | 0.04     | 0.08      | 1.00      | 1.00       |
| Pmt_Ratio_6         | 7,497 | 0.43       | 0.46       | 0           | 0.04     | 0.1       | 1         | 1          |
| Avg_Pmt_Ratio       | 7,497 | 0.41       | 0.40       | 0.00        | 0.05     | 0.19      | 0.86      | 1.00       |
| Util_1              | 7,497 | 0.42       | 0.39       | 0.00        | 0.02     | 0.32      | 0.82      | 1.00       |
| Util_2              | 7,497 | 0.40       | 0.38       | 0.00        | 0.02     | 0.30      | 0.80      | 1.00       |
| Util_3              | 7,497 | 0.39       | 0.38       | 0.00        | 0.02     | 0.28      | 0.76      | 1.00       |
| Util_4              | 7,497 | 0.36       | 0.36       | 0.00        | 0.01     | 0.26      | 0.68      | 1.00       |
| Util_5              | 7,497 | 0.34       | 0.34       | 0.00        | 0.01     | 0.22      | 0.61      | 1.00       |
| Util_6              | 7,497 | 0.32       | 0.34       | 0.00        | 0.01     | 0.19      | 0.59      | 1.00       |
| Avg_Util            | 7,497 | 0.37       | 0.34       | 0.00        | 0.03     | 0.29      | 0.68      | 1.00       |
| Bal_Growth_6mo      | 7,497 | 1,901.23   | 12,454.79  | -108,098.00 | -0.16    | 0.04      | 1.28      | 252,810.00 |
| Util_Growth_6mo     | 7,497 | 0.09       | 0.27       | -1.00       | -0.03    | 0.01      | 0.17      | 1.00       |
| Max_Util            | 7,497 | 0.48       | 0.39       | 0.00        | 0.07     | 0.44      | 0.92      | 1.00       |
| Min_Util            | 7,497 | 0.27       | 0.31       | 0.00        | 0.00     | 0.09      | 0.51      | 1.00       |
| MAX_Util_ratio      | 7,497 | 0.70       | 0.38       | 0.00        | 0.41     | 0.92      | 1.00      | 1.00       |
| Max_Bill_Amt        | 7,497 | 60,921.92  | 77,975.96  | -4,894      | 9,920    | 31,205    | 80,117    | 855,086    |
| Max_Pmt_Amt         | 7,497 | 15,850.14  | 37,705.18  | 0           | 2,184    | 5,000     | 12,000    | 1,024,516  |
| Freq_PAY            | 7,497 | -0.18      | 0.89       | -1          | -1       | 0         | 0         | 7          |
| REPAY_PATTERN       | 7,497 | 1.23       | 1.38       | 0           | 0        | 1         | 3         | 4          |
| UTIL_PATTERN        | 7,497 | 0.72       | 1.10       | 0           | 0        | 0         | 1         | 3          |
| sqrt_LIMIT_BAL      | 7,497 | 377.55     | 158.80     | 100.00      | 223.61   | 374.17    | 489.90    | 871.78     |
| Avg_Bill_Amt_tfm    | 7,497 | 30.82      | 12.91      | 0.00        | 22.01    | 31.65     | 39.90     | 57.39      |
| Avg_Pmt_Amt_tfm     | 7,497 | 7.52       | 2.06       | 0.00        | 7.03     | 7.79      | 8.63      | 12.18      |
| Max_Bill_Amt_tfm    | 7,497 | 34.16      | 12.60      | 0.00        | 26.50    | 34.61     | 42.83     | 60.10      |
| Max_Pmt_Amt_tfm     | 7,497 | 8.32       | 2.25       | 0.00        | 7.69     | 8.52      | 9.39      | 13.84      |
| WOE_AGE             | 7,497 | 0.51       | 13.20      | -24.67      | -9.95    | 10.78     | 10.78     | 10.78      |
| WOE_Freq_PAY        | 7,497 | 6.21       | 57.75      | -183.14     | 23.82    | 23.82     | 23.82     | 23.82      |
| WOE_Avg_Util        | 7,497 | 3.02       | 34.96      | -54.49      | -19.38   | -13.40    | 40.19     | 40.19      |
| WOE_Bal_Growth_6mo  | 7,497 | 2.63       | 34.15      | -42.87      | -42.87   | 27.60     | 27.60     | 29.88      |
| WOE_LIMIT_BAL       | 7,497 | 4.07       | 39.31      | -67.68      | -13.11   | -13.11    | 44.34     | 44.34      |
| WOE_MAX_DLQ         | 7,497 | 17.68      | 79.26      | -111.04     | -111.04  | 66.48     | 66.48     | 66.48      |
| WOE_OVER_LIMIT      | 7,497 | 0.94       | 16.22      | -41.59      | 7.12     | 7.12      | 7.12      | 7.12       |
| WOE_OVER_PMT        | 7,497 | 2.61       | 31.02      | -13.73      | -13.73   | -13.73    | -13.73    | 61.48      |
| WOE_REPAY_PATTERN   | 7,497 | 1.09       | 19.97      | -19.17      | -19.17   | -14.01    | 21.32     | 21.32      |
| WOE_Util_Growth_6mo | 7,497 | 2.02       | 29.05      | -43.09      | -7.45    | 7.59      | 37.48     | 37.48      |
| WOE_UTIL_PATTERN    | 7,497 | 0.95       | 18.63      | -13.50      | -13.50   | -13.50    | 24.97     | 24.97      |