Model #101: Credit Card Default Model

Performance Monitoring Plan

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 [1], 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' coefficients[2] 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.

[1] To see the feature importance plots, refer to Appendix A.

[2] 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[3] 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	TN	0.75	Precision	0.43	Sensitivity	0.67
0	8787	2,970	11757		0	0.75	0.25	Type I Error	0.25	Recall	0.67	Specificity	0.75
1	1,144	2,279	3423	41.1%	1	0.33	0.67	Type II Error	0.33	F1	0.53	Accuracy	0.7290

[3]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 (Y=1)	NonTarget (Y=0)	Target Density	NonTarget Density	Target CDF	NonTarget CDF	KS 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 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 0	Predicted Class 1	Totals	**KS**	Actual Class	Predicted Class 0	Predicted Class 1	TP	0.68	TP+TN	1.43	**AUC**	0.7633
0	4,815	1,581	6,396		0	0.75	0.25	Type I Error	0.25	Recall	0.68	Specificity	0.75
1	499	1,058	1,557	40.9%	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 (Y=1)	NonTarget (Y=0)	Target Density	NonTarget Density	Target CDF	NonTarget CDF	KS 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.

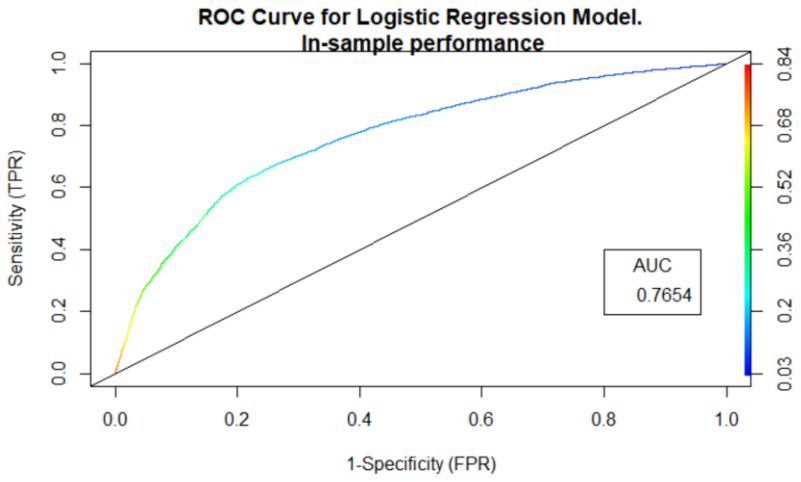


Figure 2.

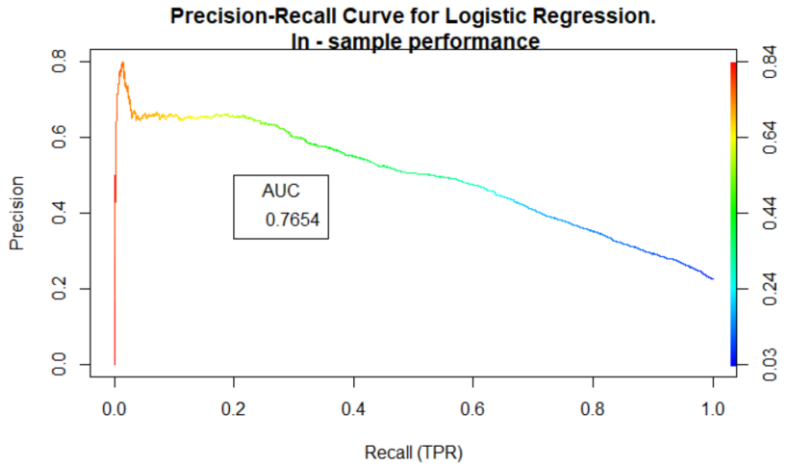


Figure 3.

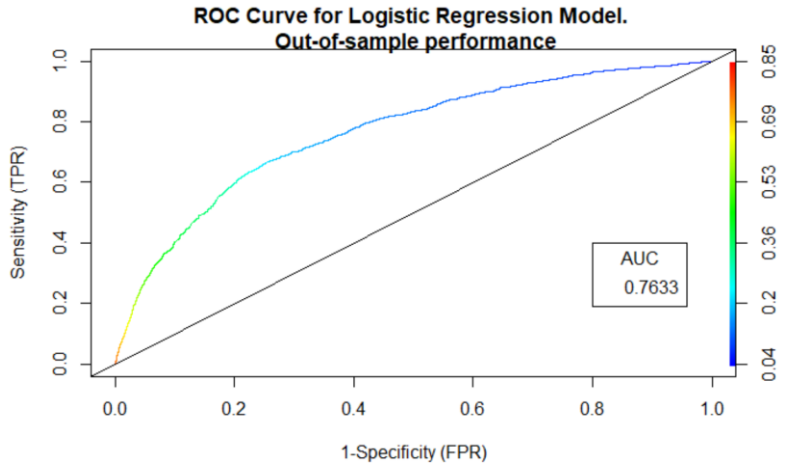
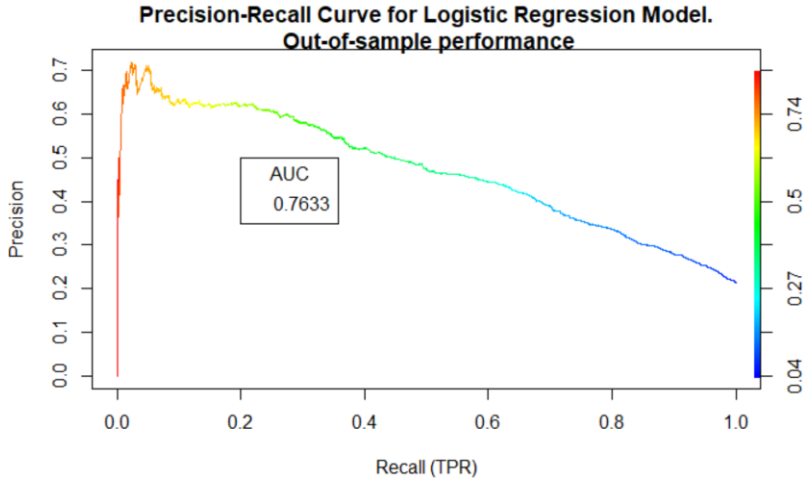


Figure 4.



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.).

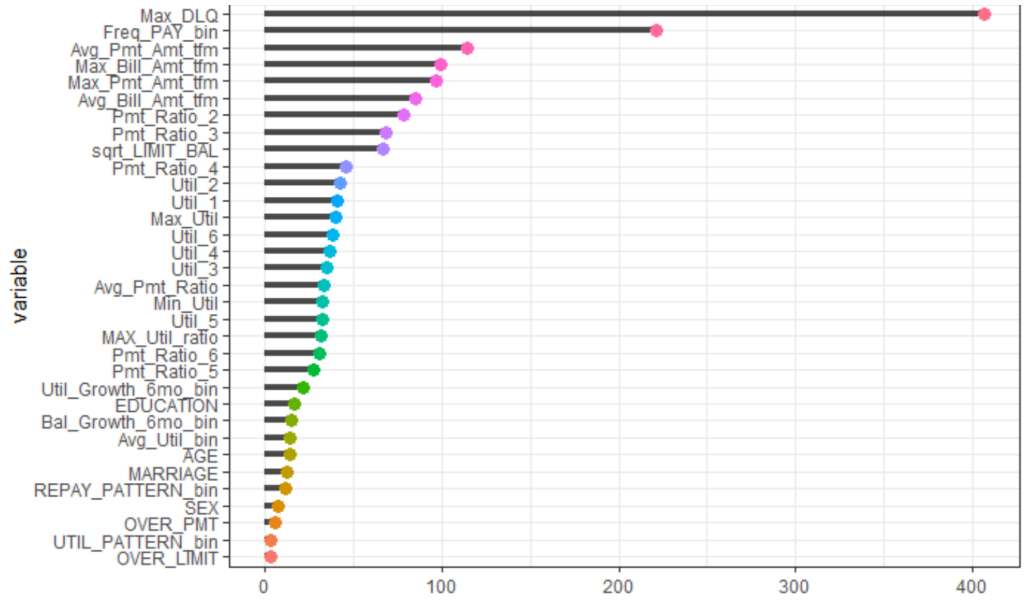
Table 4. Recommended Actions based on the KS value.

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

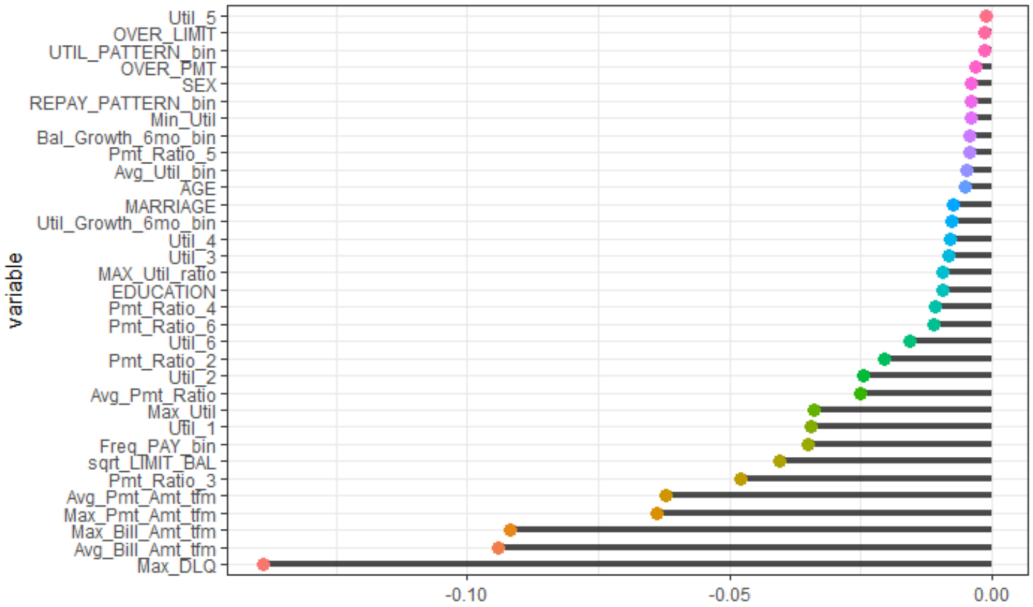
Appendix A.

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.

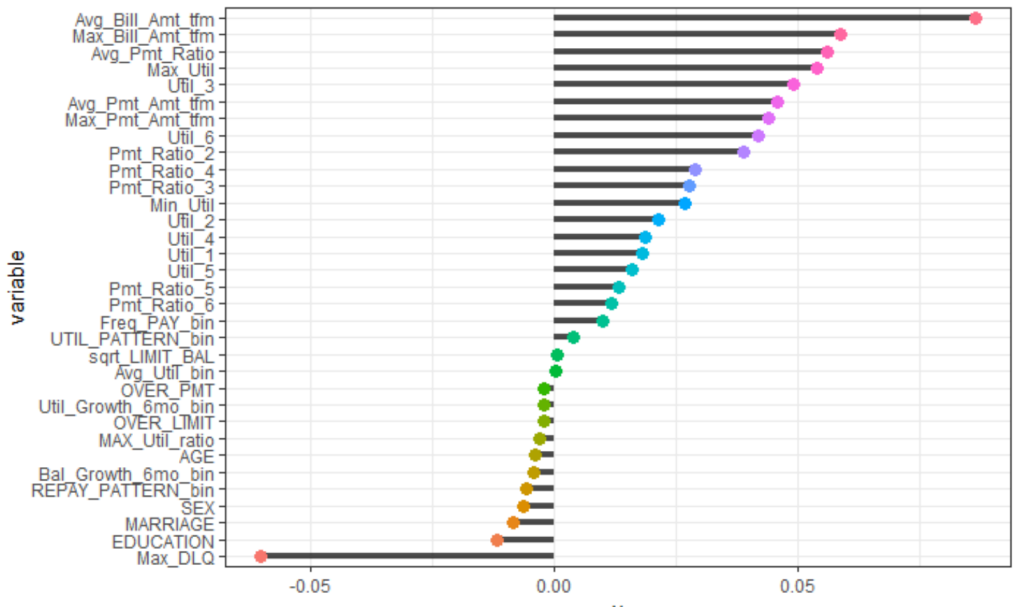
Importance ranking of variables (Gini Index).



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 Forest F1	Random Forest Gini Idx	Random Forest TPR	Boruto	Negative score	Positive score	Total 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