Model #101: Credit Card Default Model

Performance Monitoring Guide

Mimi Trinh

Section 1: Production Model

Using the results of the first two models of random forest and gradient boosting, we identify a pool of interesting predictors to use in the logistic regression model. Specifically, we remove the four insignificant demographic variables MARRIAGE, AGE, SEX, EDUCATION and leave the seven payment-related predictors remain in the predictor pool to develop a logistic regression model. Then among these seven variables, we use the stepwise automatic variable selection method to arrive at the optimal logistic regression model.

Figure 1 – Logistic Regression Model Summary Result

```
Call:
glm(formula = DEFAULT ~ LIMIT_BAL + AVG_BILL_AMT + AVG_PAY_AMT +
    AVG_UTIL + MAX_BILL_AMT + MAX_PAY_AMT + MAX_DLQ, family = binomial(),
    data = model_train)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.4782  -0.7867  -0.6500  -0.2307   4.9009

Coefficients:
               Estimate Std. Error z value Pr(>|z|)
(Intercept)  -8.342e-01  5.612e-02 -14.865  < 2e-16 ***
LIMIT_BAL    -1.960e-06  2.622e-07  -7.473 7.82e-14 ***
AVG_BILL_AMT  9.209e-06  1.826e-06   5.043 4.59e-07 ***
AVG_PAY_AMT  -1.787e-04  1.638e-05 -10.905  < 2e-16 ***
AVG_UTIL      2.838e-01  8.895e-02   3.190  0.00142 **
MAX_BILL_AMT -8.600e-06  1.719e-06  -5.003 5.65e-07 ***
MAX_PAY_AMT   2.970e-05  3.125e-06   9.504  < 2e-16 ***
MAX_DLQ      -4.876e-06  2.102e-06  -2.319  0.02037 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 16205  on 15179  degrees of freedom
Residual deviance: 15464  on 15172  degrees of freedom
AIC: 15480

Number of Fisher Scoring iterations: 6
```

The result above shows that all seven predictors are statistically significant at 95% confidence level, each with p-value less than 0.05 alpha. Thus, the stepwise automatic variable selection algorithm indicates that all variables in the model are significant. Among these seven predictors, LIMIT_BAL, AVG_PAY_AMT, MAX_BILL_AMT, MAX_DLQ have negative coefficients, which mean that they have a negative correlation with the dependent variable. In other words, the lower the limit balance and the lower the average payment amount and the lower of the maximum bill and the lower the maximum delinquency value, the higher the probability of default on payment. The other three predictors AVG_BILL_AMT, AVG_UTIL, MAX_PAY_AMT have positive coefficient, which mean that these variables have a positive correlation with the response variable. In other words, the higher the average billing amount and the higher the utilization rate and the higher the maximum payment amount, the higher the chance of default on payment.

Section 2: Model Development Performance

Figure 2 – Model Performance on Train Dataset

	No Default Actual	Default Actual	Sum
No Default Predicted	7443	1332	8775
Default Predicted	4314	2091	6405
Sum	11757	3423	15180

The classification table above is used to calculate the following performance metrics for the train dataset.

- TPR = 2091 / 3423 = 61.09%
- FPR = 4314 / 11,757 = 36.69%
- Accuracy = (7443 + 2091) / 15,180 = 62.81%

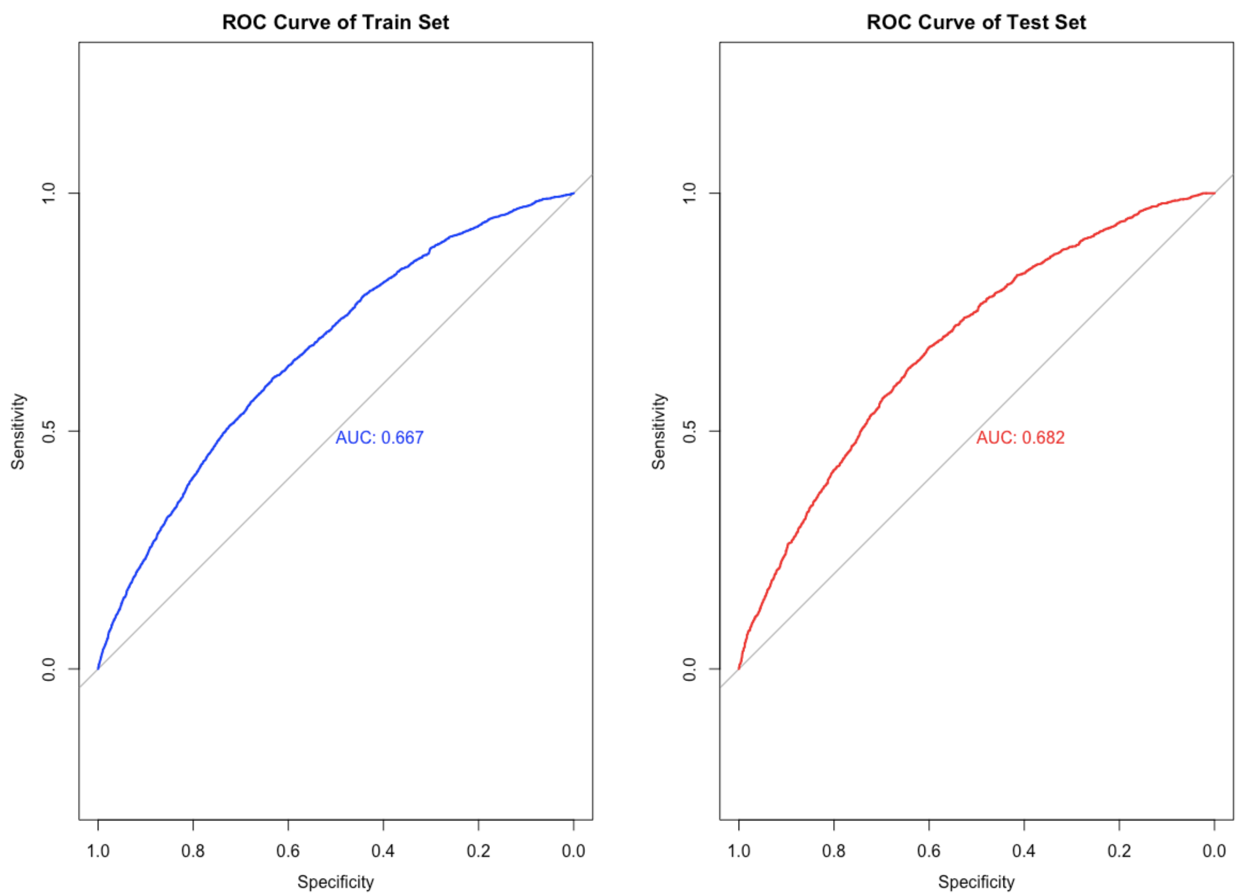
Figure 3 – Model Performance on Test Dataset

	No Default Actual	Default Actual	Sum
No Default Predicted	3470	506	3976
Default Predicted	2296	1051	3347
Sum	5766	1557	7323

The classification matrix above is used to calculate the following performance metrics for the test dataset.

- TPR = 1051 / 1557 = 67.5%
- FPR = 2296 / 5766 = 39.82%
- Accuracy = (3470 + 1051) / 7323 = 61.74%

Figure 4 – ROC Curve and AUC on both Train and Test Dataset



The two ROC curves and AUC above for train and test sets are very similar to one another. Thus, we can conclude that there's no overfitting issue in the logistic regression model. In general, the higher the AUC, the better the model. The AUC value ranges from 0.5 to indicate no diagnostic ability to 1.0 to indicate perfect diagnostic ability. In this scenario, the AUC on the train set is 0.667 whereas the AUC on the test is 0.682, which mean that the logistic regression model doesn’t successfully capture the response variable. Thus, though this is a good model to start, there still much work needed to be done to improve the model. Moreover, since the AUC numbers for both train and test sets are similar to one another, we can conclude that the model doesn’t have an overfitting issue.

Figure 5 – Lift Chart for Train Dataset

Train Set		Goods	Bads					
Decile	**Obs**	**Target (Y=1)**	**NonTarget (Y=0)**	**Target Density**	**NonTarget Density**	**Target CDF**	**NonTarget CDF**	**KS Stat**
1	759	759	0	11.9%	0.0%	11.9%	0.0%	11.9%
2	759	759	0	11.9%	0.0%	23.7%	0.0%	23.7%
3	759	759	0	11.9%	0.0%	35.6%	0.0%	35.6%
4	759	759	0	11.9%	0.0%	47.4%	0.0%	47.4%
5	759	759	0	11.9%	0.0%	59.3%	0.0%	59.3%
6	759	759	0	11.9%	0.0%	71.1%	0.0%	71.1%
7	759	759	0	11.9%	0.0%	83.0%	0.0%	83.0%
8	759	759	0	11.9%	0.0%	94.8%	0.0%	94.8%
9	759	333	426	5.2%	4.9%	100.0%	4.9%	95.1%
10	759	0	759	0.0%	8.6%	100.0%	13.5%	86.5%
11	759	0	759	0.0%	8.6%	100.0%	22.2%	77.8%
12	759	0	759	0.0%	8.6%	100.0%	30.8%	69.2%
13	759	0	759	0.0%	8.6%	100.0%	39.5%	60.5%
14	759	0	759	0.0%	8.6%	100.0%	48.1%	51.9%
15	759	0	759	0.0%	8.6%	100.0%	56.8%	43.2%
16	759	0	759	0.0%	8.6%	100.0%	65.4%	34.6%
17	759	0	759	0.0%	8.6%	100.0%	74.1%	25.9%
18	759	0	759	0.0%	8.6%	100.0%	82.7%	17.3%
19	759	0	759	0.0%	8.6%	100.0%	91.4%	8.6%
20	759	0	759	0.0%	8.6%	100.0%	100.0%	0.0%
Totals	**15,180**	**6405**	**8,775**	**100.0%**	**100.0%**			

Figure 6 – Lift Chart for Test Dataset

Test Set Goods Bads

Decile	Obs	Target (Y=1)	NonTarget (Y=0)	Target Density	NonTarget Density	Target CDF	NonTarget CDF	KS Stat
1	367	367	0	11.0%	0.0%	11.0%	0.0%	11.0%
2	366	366	0	10.9%	0.0%	21.9%	0.0%	21.9%
3	366	366	0	10.9%	0.0%	32.8%	0.0%	32.8%
4	366	366	0	10.9%	0.0%	43.8%	0.0%	43.8%
5	366	366	0	10.9%	0.0%	54.7%	0.0%	54.7%
6	366	366	0	10.9%	0.0%	65.6%	0.0%	65.6%
7	366	366	0	10.9%	0.0%	76.6%	0.0%	76.6%
8	366	366	0	10.9%	0.0%	87.5%	0.0%	87.5%
9	366	366	0	10.9%	0.0%	98.4%	0.0%	98.4%
10	366	52	314	1.6%	7.9%	100.0%	7.9%	92.1%
11	367	0	367	0.0%	9.2%	100.0%	17.1%	82.9%
12	366	0	366	0.0%	9.2%	100.0%	26.3%	73.7%
13	366	0	366	0.0%	9.2%	100.0%	35.5%	64.5%
14	366	0	366	0.0%	9.2%	100.0%	44.7%	55.3%
15	366	0	366	0.0%	9.2%	100.0%	53.9%	46.1%
16	366	0	366	0.0%	9.2%	100.0%	63.2%	36.8%
17	366	0	366	0.0%	9.2%	100.0%	72.4%	27.6%
18	366	0	366	0.0%	9.2%	100.0%	81.6%	18.4%
19	366	0	366	0.0%	9.2%	100.0%	90.8%	9.2%
20	367	0	367	0.0%	9.2%	100.0%	100.0%	0.0%
Totals	**7,323**	**3347**	**3,976**	**100.0%**	**100.0%**			

Using the semi-deciles or half-deciles with 20 groups, the two lift charts above are provided, one for train set and one for test set. From the lift charts above, the Kolmogorov-Smirnov (KS) statistics are calculated: 95.1 for train set and 98.4 for test set. Similar to the AUC numbers, the two KS statistics for train and test sets are very close to each other, which means that the predictive model has no overfitting issue.

Section 3: Performance Monitoring Plan

According to Lyn C. Thomas in the book "Consumer Credit Models – Pricing, Profit, and Portfolios," the rule of thumb is that KS statistics of 0.4 (or 40 as the metric used in this project) suggest good discrimination (p112). In this project, the KS statistics of 95.1 and 98.4 on train and test sets are very high above the threshold. To be more specific, this performance monitoring plan uses the table below outlining the metric threshold for the KS statistics that determine each RAG (red, amber, green) status. The KS statistics in this study uses an absolute change instead of a percentage change to define the performance status.

Figure 7 – KS Statistics Threshold for RAG Status

Performance Status	KS Statistics Threshold
Red	0 – 30
Amber	31 – 60
Green	61 – 100

In the RAG status, red category means that the model needs redevelopment. Amber means that the model needs to be re-validated in three months. Green means that the model is performing as expected. Therefore, using the table above, the KS statistics for the logistic regression predictive model falls in the green category, which means that the logistic regression model performs as well as expected. Thus, the model will be re-validated at the standard interval of six months.