Model #101: Credit Card Default Model Performance Monitoring Plan

Tannia Dubon
MSDS 498

1. The Production Model

The logistic regression model selected for production was developed using the glm() function and it was trained on a subset of variables. The variables themselves were selected using both the output from a Random Forest model that was trained on all the variables in the modeling suite and a logistic regression variable selection (regsubsets) method.

The logistic regression model was selected from the suite of models developed given that it is the most common method applied to scorecards (Thomas, 2009, p. 79) and its practical implementation and revalidation in a production environment with large data sets (Bhatti, 2018).

The use of the variable PAY_1, the only non-engineered variable included, will need to be addressed with the respective business unit given that the data that was received contained entries that were not listed in the data dictionary, and the use of the codes was inconsistent. It will be helpful to define the business rules for how the whole set of variables (PAY_1 – PAY_6) are used, to ensure that the variable imputations or adjustments are applied appropriately.

The precise code used to run this model was:

```
log_reg <- glm(DEFAULT ~ PAY_1 + Max_Util + Max_Pmt_Amt + Max_DLQ +
 Bal_Growth_6mo_21390_428792 + Avg_Pmt_Amt_2833_12092 +AGE_25_32,
 data=train_df, family=binomial)
```

The table of coefficients in Image 1 presents the significance levels of each of the variables and their estimated coefficient values.

```
MODEL INFO:
Observations: 15180
Dependent Variable: DEFAULT
Type: Generalized linear model
  Family: binomial
  Link function: logit

MODEL FIT:
χ²(14) = 2693.97, p = 0.00
Pseudo-R² (Cragg-Uhler) = 0.25
Pseudo-R² (McFadden) = 0.17
AIC = 13541.41, BIC = 13655.83

Standard errors: MLE
```

	Est.	S.E.	z val.	p
(Intercept)	-1.48	0.04	-34.29	0.00
PAY_11	0.38	0.07	5.55	0.00
PAY_12	1.76	0.08	21.91	0.00
PAY_13	1.68	0.21	8.04	0.00
PAY_14	1.61	0.41	3.94	0.00
PAY_15	-0.32	0.57	-0.56	0.57
PAY_16	-0.13	0.78	-0.17	0.86
PAY_17	11.76	132.53	0.09	0.93
PAY_18	-0.47	0.88	-0.54	0.59
Max_Util	0.21	0.05	4.04	0.00
Max_Pmt_Amt	-0.00	0.00	-5.34	0.00
Max_DLQ	0.30	0.02	15.74	0.00
Bal_Growth_6mo_21390_4287921	-0.26	0.13	-1.97	0.05
Avg_Pmt_Amt_2833_120921	-0.39	0.05	-7.86	0.00
AGE_25_321	-0.16	0.05	-3.38	0.00

Image 1: Table of Coefficients for Logistic Regression Model

All of the variables, with the exception of dummy variables PAY_15-18, have significance levels that would lead us to reject the null hypothesis, based on an alpha level of .05. The dummy variable PAY_17 has the largest coefficient. The value of the coefficient represents the average increase in DEFAULT for each unit of change when PAY_1 = 7, while holding all other predictor variables constant. This is a potential point of concern and requires close attention as it may result in a less robust model during re-validation. Table 1 below shows just how small the occurrence of PAY_1 = 5-8 really is.

Table 1: Pay_1 Values by Default										
		-1	1	2	3	4	5	6	7	8
DEFAULT	0	19975	2436	823	78	24	13	5	2	8
	1	3207	1252	1844	244	52	13	6	7	11

2. Model Development Performance

This model performed relatively well on the training data. The Sensitivity indicates that the true positives for DEFAULT were correctly predicted 70% of the time. The Specificity reflects the rate at which the true negatives were predicted, that is 83% of the time.

Model Logistic Regression train data													
Actual	Predicted Class		Totals		Actual	Predicted Class		TP	0.70	TP+TN	1.53	AUC	0.64
Class	0	1			Class	0	1	TN	0.83	Precision	0.33	Sensitivity	0.70
0	11,284	2,300	13,584		0	0.83	0.17	Type I Error	0.17	Recall	0.70	Specificity	0.83
1	473	1,123	1,596		1	0.30	0.70	Type II Error	0.30	F1	0.75		

Image 2: Confusion Matrix for Training Data

The area under the curve (AUC) displays the tradeoffs between the Sensitivity and the Type I Errors for the model at different levels. The optimal area under the curve is 64%, a point on the curve at which Sensitivity equals .70 and the Type I Error equals .17.

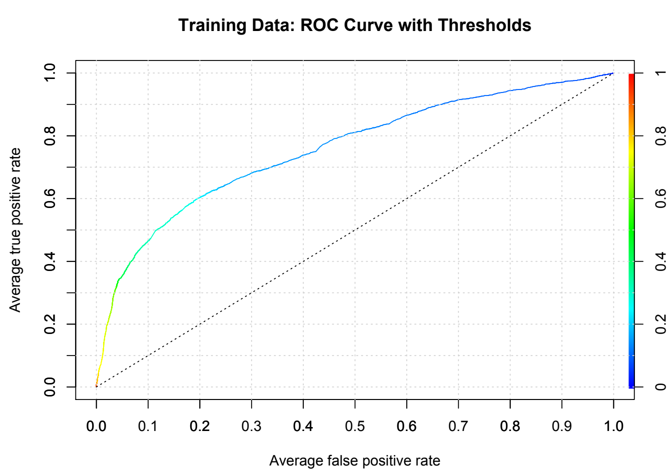

Image 2: ROC Curve for Training Data

Next, we consider the performance of the model on the testing data. The Sensitivity drops by 2%, so we can expect the true positives to be correctly predicted 68% of the time. The Specificity increases by 1%, so the true negatives are predicted at a slightly higher rate.

Model Logistic Regression test data													
Actual	Predicted Class		Totals		Actual	Predicted Class		TP	0.68	TP+TN	1.52	AUC	0.65
Class	0	1			Class	0	1	TN	0.84	Precision	0.34	Sensitivity	0.68
0	5,512	1,028	6,540		0	0.84	0.16	Type I Error	0.16	Recall	0.68	Specificity	0.84
1	254	529	783		1	0.32	0.68	Type II Error	0.32	F1	0.74		

Image 3: Confusion Matrix for Testing Data

We don't see a dramatic increase or decrease in the AUC of the testing data. It increases by 1%, to 65%, given the increase in the Specificity to .84. The plot of the ROC curve reflects this in Image 5.

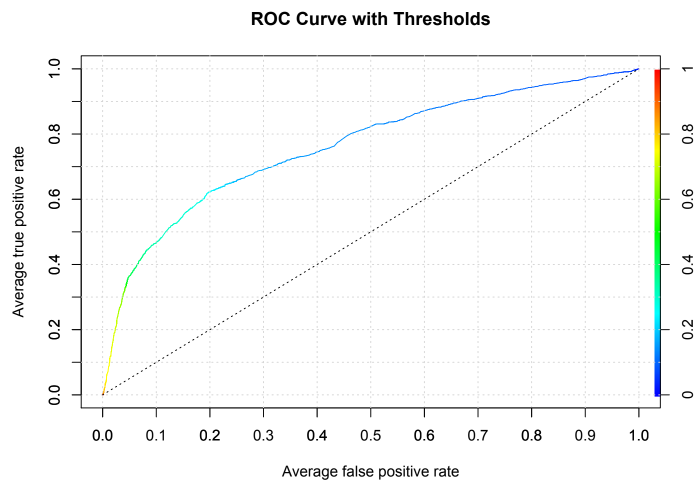

Image 3: ROC Curve for Testing Data

The Kolmogorov-Smirnov (KS) statistic scores the differences between the cumulative distribution functions using the densities of the binary response variable, by deciles or semi-deciles (Thomas, 2009, p. 111). The KS statistics for the training data are depicted in Table 2 below. The highest score of 40.2% represents the maximum difference between the cumulative distribution functions of the two groups and it takes place in the 6th semi-decile (the .70-.75 band). This score is expected to discriminate well (Thomas, 2009, p. 112), but it will need to be monitored as verification.

The 4th, 5th, 7th, and 8th semi-deciles were also close to the highest score, indicating that these are also areas of maximum distance and nearly have a good discrimination capacity.

Table 2. Kolmogorov-Smirnov Statistic Lift Chart for Training Data								
Semi-Decile	Obs	Target (Y=1)	NonTarget (Y=0)	Target Density	NonTarget Density	Target CDF	NonTarget CDF	KS Stat
1	759	563	196	16.4%	1.7%	16.4%	1.7%	14.8%
2	759	513	246	15.0%	2.1%	31.4%	3.8%	27.7%
3	759	338	421	9.9%	3.6%	41.3%	7.3%	34.0%
4	759	285	474	8.3%	4.0%	49.6%	11.4%	38.3%
5	759	210	549	6.1%	4.7%	55.8%	16.0%	39.7%
6	759	183	576	5.3%	4.9%	61.1%	20.9%	40.2%
7	759	144	615	4.2%	5.2%	65.3%	26.2%	39.2%
8	759	121	638	3.5%	5.4%	68.9%	31.6%	37.3%
9	759	107	652	3.1%	5.5%	72.0%	37.1%	34.8%
10	759	106	653	3.1%	5.6%	75.1%	42.7%	32.4%
11	759	166	593	4.8%	5.0%	79.9%	47.7%	32.2%
12	759	90	669	2.6%	5.7%	82.6%	53.4%	29.1%
13	759	110	649	3.2%	5.5%	85.8%	59.0%	26.8%
14	759	111	648	3.2%	5.5%	89.0%	64.5%	24.6%
15	759	87	672	2.5%	5.7%	91.6%	70.2%	21.4%
16	759	49	710	1.4%	6.0%	93.0%	76.2%	16.8%
17	759	61	698	1.8%	5.9%	94.8%	82.2%	12.6%
18	759	64	695	1.9%	5.9%	96.6%	88.1%	8.6%
19	759	46	713	1.3%	6.1%	98.0%	94.1%	3.9%
20	759	69	690	2.0%	5.9%	100.0%	100.0%	0.0%
Totals	15,180	3423	11,757	100.0%	100.0%			

The KS scores were also calculated for the testing data. The highest score was 41.9% in the 6th semi-decile; reflecting a 1.9% increase from the best score on the training data. Table 3 also shows that the 5th and 7th semi-deciles have good discrimination.

Table 3. Kolmogorov-Smirnov Statistic Lift Chart for Testing Data								
Semi-Decile	Obs	Target (Y=1)	NonTarget (Y=0)	Target Density	NonTarget Density	Target CDF	NonTarget CDF	KS Stat
1	367	257	110	16.5%	1.9%	16.5%	1.9%	14.6%
2	366	240	126	15.4%	2.2%	31.9%	4.1%	27.8%
3	366	169	197	10.9%	3.4%	42.8%	7.5%	35.3%
4	366	115	251	7.4%	4.4%	50.2%	11.9%	38.3%
5	366	109	257	7.0%	4.5%	57.2%	16.3%	40.8%
6	366	91	275	5.8%	4.8%	63.0%	21.1%	41.9%
7	366	56	310	3.6%	5.4%	66.6%	26.5%	40.1%
8	366	55	311	3.5%	5.4%	70.1%	31.9%	38.3%
9	366	47	319	3.0%	5.5%	73.2%	37.4%	35.8%
10	367	46	321	3.0%	5.6%	76.1%	43.0%	33.1%
11	366	78	288	5.0%	5.0%	81.1%	48.0%	33.2%
12	366	40	326	2.6%	5.7%	83.7%	53.6%	30.1%
13	366	46	320	3.0%	5.5%	86.6%	59.2%	27.5%
14	366	39	327	2.5%	5.7%	89.1%	64.8%	24.3%
15	366	31	335	2.0%	5.8%	91.1%	70.6%	20.5%
16	366	34	332	2.2%	5.8%	93.3%	76.4%	16.9%
17	366	28	338	1.8%	5.9%	95.1%	82.3%	12.9%
18	366	20	346	1.3%	6.0%	96.4%	88.3%	8.1%
19	366	30	336	1.9%	5.8%	98.3%	94.1%	4.2%
20	367	26	341	1.7%	5.9%	100.0%	100.0%	0.0%
Totals	7,323	1,557	5,766	100.0%	100.0%			

3. Performance Monitoring Plan

In this section, the performance benchmarks listed below are based on a K-S statistic of 41.9% from the testing data. These thresholds along with routine re-validation procedures are necessary given that the critical area identified by the K-S statistic would not be valid in a scenario where the parameters for the distributions change.

Table 4. Performance Monitoring Rules

Status	K-S Statistic Threshold	Definition
Red	27 – 34 %	The model needs redevelopment.
Amber	35 – 41 %	Model needs to be revalidated in 3 months
Green	42 %	Model is performing as expected. To be re-validated every 6 months

The corresponding confusion matrices benchmark the errors associated with the K-S statistic thresholds. These allow us to articulate what we expect to see in the different scenarios and to ensure alignment with risk thresholds.

Table 5. Confusion Matrices for Each Threshold Level

Confusion Matrix using K-S Scores for Testing Data with a 42% Cutoff

	Actual Goods	Actual Bads		Number so Predicted
Predicted Goods	890	941		1,831
Predicted Bads	667	4,825		5,492
Actual Numbers	1,557	5,766		7,323
Specificity	0.571612075			
Type I Error	0.428387925			
Sensitivity	0.836801942			
Type II Error	0.163198058			
Total Error	0.22			

Confusion Matrix using K-S Scores for Testing Data with a 35.8% Cutoff

	Actual Goods	Actual Bads		Number so Predicted
Predicted Goods	1,092	1,837		2,929
Predicted Bads	465	3,929		4,394
Actual Numbers	1,557	5,766		7,323
Specificity	0.701348748			
Type I Error	0.298651252			
Sensitivity	0.681408255			
Type II Error	0.318591745			
Total Error	0.31			

Table 5. Confusion Matrices for Each Threshold Level				
Confusion Matrix using K-S Scores for Testing Data with a 27.5% Cutoff				
	Actual Goods	Actual Bads		Number so Predicted
Predicted Goods	1,303	3,091		4,394
Predicted Bads	254	2,675		2,929
Actual Numbers	1,557	5,766		7,323
Specificity	0.836865768			
Type I Error	0.163134232			
Sensitivity	0.463926465			
Type II Error	0.536073535			
Total Error	0.46			