Model #101: Credit Card Default Model
Performance Validation Results

George Brown

1. The Production Model

For the model development guide, we looked at 30,000 credit card observations from Taiwanese customers. The data was collected over a six-month period in 2005. The goal was to use predictor variables and featured engineering variables to predict our target variable which was will the people default on their credit card payments.

We looked at four different models including a random forest model, a gradient boosting model, a logistic regression model, and a naïve bayes model. We looked at all four methods using a training dataset and then from there tested the method using a testing dataset. All four models produced good results with at least 70% accuracy for the training and the test models. For this model production we are going to look at the logistic regression model.

For the logistic regression model, we started out with nine variables that the random forest and gradient boosting models had determined were important. The variables that were used are listed below. For the logistic regression model, we used a stepwise function to predict the default variable. The stepwise model either adds or removes predictor variables one at a time depending on the variable's significance. It either adds the most significant variable or removes the most insignificant variable.

Table 1: Variables for Logistic Regression Model

Variables used for Logistic Regression Model
1. Max Delinquency
2. Average Utility
3. Utility Growth Rate per 6 Months
4. Average Pay Ratio
5. Average Payment Amount
6. Max Bill Amount
7. Bill Sum
8. Max Payment Amount
9. Balance Growth Over 6 Months

The following table shows the result of our logistic regression model. Of the nine variables, five of them were determined important enough to keep when predicting default. The variables were average payment amount, average utility, max payment amount, max delinquency, and bill sum.

Table 2: Logistic Regression Model Summary

	Estimate	Std. Error	z Value	P Value	
(Intercept)	-1.81E+00	3.96E-02	-45.551	2.00E-16	***
Avg_Pay_Amt	-9.52E-05	1.34E-05	7.093	1.31E-12	***
Avg_Util	2.81E-01	7.11E-02	3.95	7..81E-05	***
Max_pay_Amt	1.32E-05	2.72E-06	4.871	1.11E-06	***
Max_DLQ	7.13E-01	2.02E-02	35.232	2.00E-16	***
Bill_Sum	2.94E-07	8.94E-08	3.288	1.01E-03	**
		Signif. Codes		**= .01	***=.001

2. Model Development Performance

The logistic regression model using the step wise function produced the following results.

Table 3: Logistic Regression Model Training and Test Metrics

Model #3A: Logistic Regression-Training												
Actual Class	Predicted Class 0	Predicted Class 1	Totals		Actual Class	Predicted Class 0	Predicted Class 1	TP	0.62	TP+TN	1.41	AUC	0.70
0	9,279	2,478	11,757		0	0.79	0.21	TN	0.79	Precision	0.46	Sensitivity	0.62
1	1,308	2,115	3,423		1	0.38	0.62	Type I Error	0.21	Recall	0.62	Specificity	0.79
								Type II Error	0.38	F1	0.68		

Table 5: Confusion matrix and classification metrics for Model #3A : Logistic Regression-Training

Model #3B: Logistic Regression-Testing												
Actual Class	Predicted Class 0	Predicted Class 1	Totals		Actual Class	Predicted Class 0	Predicted Class 1	TP	0.63	TP+TN	1.41	AUC	0.71
0	4,508	1,258	5,766		0	0.78	0.22	TN	0.78	Precision	0.44	Sensitivity	0.63
1	578	979	1,557		1	0.37	0.63	Type I Error	0.22	Recall	0.63	Specificity	0.78
								Type II Error	0.37	F1	0.68		

Table 6: Confusion matrix and classification metrics for Model #3B : Logistic Regression-Test

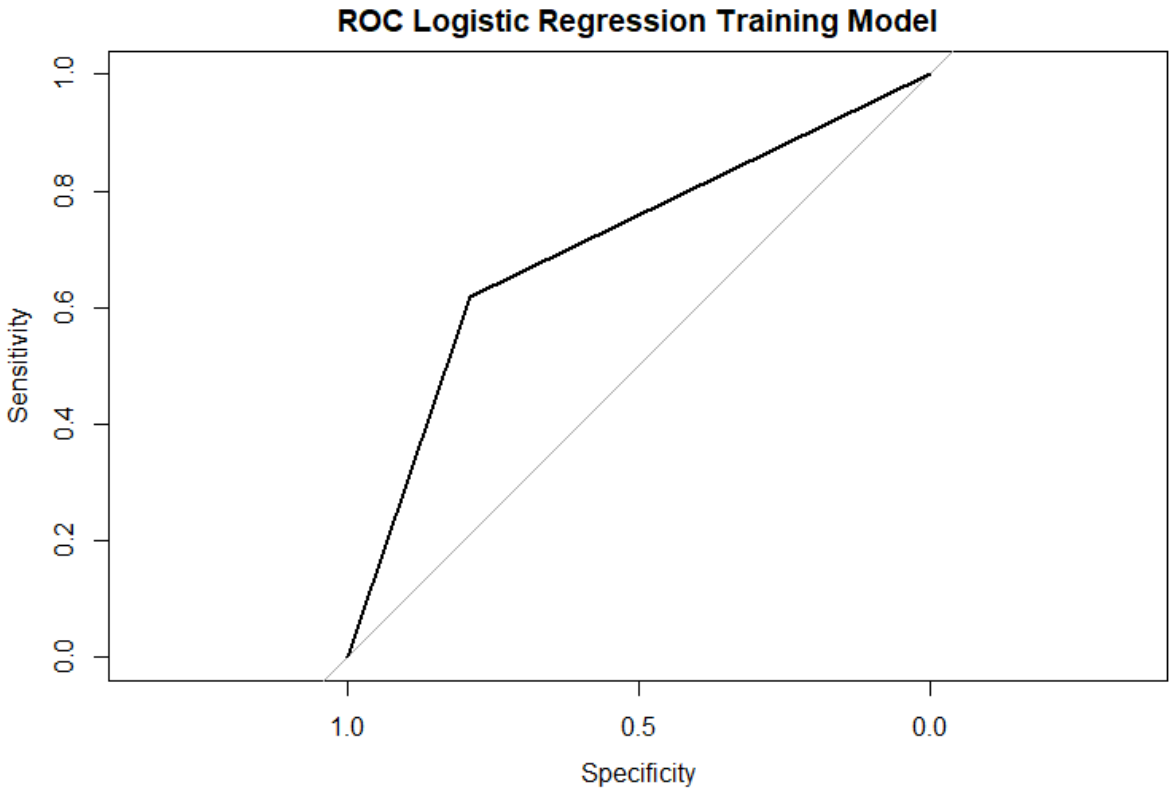
This model was a good model with an accuracy of 75% for both the training and the test model. These metrics for the tables above were all very important however, we want to look at the AUC statistics and the ROC curve for both the training and the test model.

The ROC curve is calculated by plotting the sensitivity and specificity. The ROC is the curve and the AUC tells us how well the model is at predicting results. It will tell us how well it is at predicting 0 if it a 0 and 1 if it is a 1. The higher the AUC the better the model is. Both of our AUC scores came out with just about the same result. The training model had a score of .70 and the testing model had a score of .71. As with most of the other metrics both the training and the test model had about the same results.

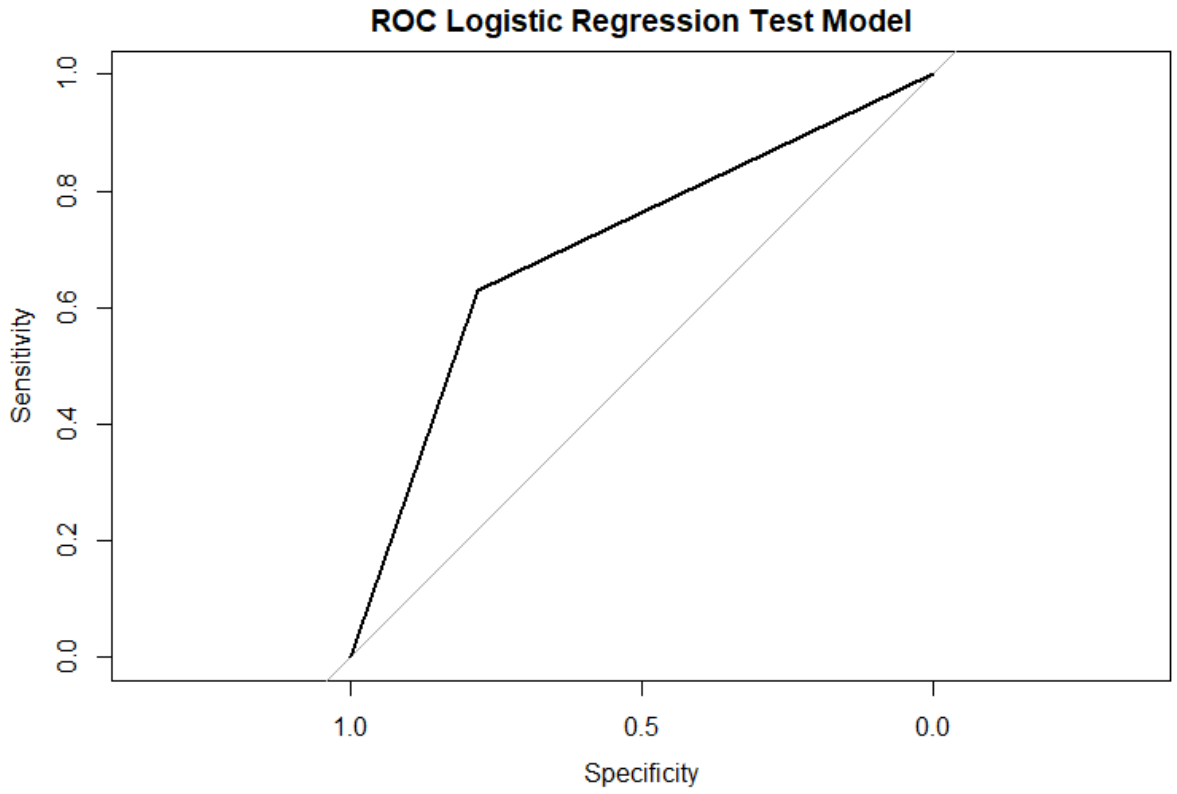
Table 4: AUC Scores for Logistic Model

Model	Logistic Training	Logistic Test
AUC Score	.70	.71

Graph 5: ROC Curve Logistic Training Model



Graph 6: ROC Curve Logistic Test Model



The next step is to look at the KS statistic of the Logistic Models. The KS statistic is also a non-parametric test that compares the distributions of two datasets. It measures the separation between the positive and negative distribution. The best KS score would be a 100 and that is if the model is perfect at separating the positive and negative results. It would be a 0 if it cannot separate any of the results. The higher the KS score the better.

We will now produce a lift chart for the KS statistic for both the training and the test datasets. We will use twenty groups to produce these lift charts. Below is the chart for the training set for the KS statistic and the test set. It is obvious that there are more target values hit for an actual default in the early deciles versus the later deciles. The largest numbers for Y=1 are in the first few deciles for both datasets. The KS statistic is calculated by looking at the target CDF and subtracting the non-target CDF. In this case, decile six has the largest KS statistic at 40.8% for

the training set and 40.9% for the test set. This shows that this is the largest gap between the two classes which are predicted default and non-default. The larger the gap, the better the indicator that model is working. When the KS statistic gets lower then we may need to adjust the model. Even though the training set had about twice as many observations as the test set, the numbers were very similar in terms of the KS statistic. For both models' deciles 5-8 produced the highest KS statistics.

Table 7: Lift Chart KS Statistic Training Set

Logistic Model KS Statistic Training Set

Decile	Obs	Target (Y=1)	NonTarget (Y=0)	Target Density	NonTarget Density	Target CDF	NonTarget CDF	KS Stat
1	759	460	299	13.4%	2.5%	13.4%	2.5%	10.9%
2	759	361	398	10.5%	3.4%	24.0%	5.9%	18.1%
3	759	376	383	11.0%	3.3%	35.0%	9.2%	25.8%
4	759	371	388	10.8%	3.3%	45.8%	12.5%	33.3%
5	759	279	480	8.2%	4.1%	54.0%	16.6%	37.4%
6	759	261	498	7.6%	4.2%	61.6%	20.8%	40.8%
7	759	125	634	3.7%	5.4%	65.2%	26.2%	39.0%
8	759	148	611	4.3%	5.2%	69.6%	31.4%	38.2%
9	759	106	653	3.1%	5.6%	72.7%	36.9%	35.7%
10	759	102	657	3.0%	5.6%	75.6%	42.5%	33.1%
11	759	78	681	2.3%	5.8%	77.9%	48.3%	29.6%
12	759	139	620	4.1%	5.3%	82.0%	53.6%	28.4%
13	759	103	656	3.0%	5.6%	85.0%	59.2%	25.8%
14	759	115	644	3.4%	5.5%	88.3%	64.7%	23.7%
15	759	90	669	2.6%	5.7%	91.0%	70.3%	20.6%
16	759	77	682	2.2%	5.8%	93.2%	76.2%	17.1%
17	759	61	698	1.8%	5.9%	95.0%	82.1%	12.9%
18	759	69	690	2.0%	5.9%	97.0%	88.0%	9.1%
19	759	61	698	1.8%	5.9%	98.8%	93.9%	4.9%
20	759	41	718	1.2%	6.1%	100.0%	100.0%	0.0%
Totals	**15,180**	**3423**	**11757**	**100.0%**	**100.0%**			

Table 8: Lift Chart KS Statistic Test Set:

Logistic Model KS Statistic Test Set

Decile	Obs	Target (Y=1)	NonTarget (Y=0)	Target Density	NonTarget Density	Target CDF	NonTarget CDF	KS Stat
1	367	210	157	13.5%	2.7%	13.5%	2.7%	10.8%
2	366	175	191	11.2%	3.3%	24.7%	6.0%	18.7%
3	366	173	193	11.1%	3.3%	35.8%	9.4%	26.5%
4	366	161	205	10.3%	3.6%	46.2%	12.9%	33.2%
5	366	140	226	9.0%	3.9%	55.2%	16.9%	38.3%
6	366	110	256	7.1%	4.4%	62.2%	21.3%	40.9%
7	366	74	292	4.8%	5.1%	67.0%	26.4%	40.6%
8	366	49	317	3.1%	5.5%	70.1%	31.9%	38.3%
9	366	51	315	3.3%	5.5%	73.4%	37.3%	36.1%
10	366	41	325	2.6%	5.6%	76.0%	43.0%	33.1%
11	367	38	329	2.4%	5.7%	78.5%	48.7%	29.8%
12	366	63	303	4.0%	5.3%	82.5%	53.9%	28.6%
13	366	47	319	3.0%	5.5%	85.5%	59.5%	26.1%
14	366	52	314	3.3%	5.4%	88.9%	64.9%	24.0%
15	366	38	328	2.4%	5.7%	91.3%	70.6%	20.7%
16	366	33	333	2.1%	5.8%	93.4%	76.4%	17.1%
17	366	31	335	2.0%	5.8%	95.4%	82.2%	13.3%
18	366	30	336	1.9%	5.8%	97.4%	88.0%	9.4%
19	366	23	343	1.5%	5.9%	98.8%	93.9%	4.9%
20	367	18	349	1.2%	6.1%	100.0%	100.0%	0.0%
Totals	**7,323**	**1557**	**5766**	**100.0%**	**100.0%**			

3. Performance Monitoring Plan

Part of the modeling performance is putting together a monitoring plan. In this section we will create a table outlining the metric threshold for the KS statistic. The threshold will be scored based off the KS statistic. The red means the model needs redevelopment. Amber means the model needs to be re-validated in three months and green means the model is performing as expected. The model will be revalidated at the standard interval of six months. Below is a chart of our three thresholds.

Table 9: Model Threshold Classification

Red	Model needs redevelopment
Amber	Model needs to be re-validated in three months
Green	Model is performing as expected

For the next step we need to determine the metrics we are going to use for the KS statistic to determine the thresholds. We will evaluate the threshold based on the absolute KS statistic not the relative change from decile to decile. Each decile will have a different threshold as some KS statistics will be acceptable in one decile and they will not be in another decile. Below is the threshold status that we will determine how our model is performing. The green threshold will be scored off the test dataset. If the model performs up to the test set, then it will be considered adequate. If for some reason the model is terrible and does not even meet the red threshold then obviously, we will need to start over. In the next part we will run our model validation test to see if the validation model performs adequately. The hope is that it performs in the green threshold, but if we have a mixture of green and yellow results then it will be considered adequate. Unless there are all green results then we will still have to evaluate the model every few months.

Table 10: KS Statistic Threshold

Decile	KS Statistic Threshold		
1	3%	6%	10.80%
2	7%	10%	18.70%
3	12%	18%	26.50%
4	17%	25%	33.20%
5	20%	30%	38.30%
6	22%	32%	40.90%
7	22%	32%	40.60%
8	20%	30%	38.30%
9	19%	28%	36.10%
10	17%	25%	33.10%
11	13%	21%	29.80%
12	13%	20%	28.60%
13	12%	18%	26.10%
14	11%	16%	24.00%
15	7%	12%	20.70%
16	7%	10%	17.10%
17	4%	8%	13.30%
18	3%	6%	9.40%
19	1%	3%	4.90%
20	0%	0%	0.00%

4. **Performance Monitoring Results**

In this section we will produce a lift chart showing the KS statistics and see how our model performs with our validation dataset. We will use the chart to see what thresholds are met. Below are the results.

Table 11: Lift Chart KS Statistic Validation Set:

Logistic Model KS Statistic Validation Set

Decile	Obs	Target (Y=1)	NonTarget (Y=0)	Target Density	NonTarget Density	Target CDF	NonTarget CDF	KS Stat
1	375	225	150	13.6%	2.6%	13.6%	2.6%	11.0%
2	375	196	179	11.8%	3.1%	25.4%	5.6%	19.8%
3	375	211	164	12.7%	2.8%	38.2%	8.4%	29.7%
4	375	171	204	10.3%	3.5%	48.5%	11.9%	36.6%
5	374	135	239	8.2%	4.1%	56.6%	16.0%	40.6%
6	375	108	267	6.5%	4.6%	63.2%	20.6%	42.6%
7	375	61	314	3.7%	5.4%	66.8%	26.0%	40.9%
8	375	52	323	3.1%	5.5%	70.0%	31.5%	38.5%
9	375	53	322	3.2%	5.5%	73.2%	37.0%	36.2%
10	374	40	334	2.4%	5.7%	75.6%	42.7%	32.9%
11	375	35	340	2.1%	5.8%	77.7%	48.6%	29.2%
12	375	71	304	4.3%	5.2%	82.0%	53.8%	28.2%
13	375	55	320	3.3%	5.5%	85.3%	59.2%	26.1%
14	375	42	333	2.5%	5.7%	87.9%	64.9%	22.9%
15	374	36	338	2.2%	5.8%	90.0%	70.7%	19.3%
16	375	49	326	3.0%	5.6%	93.0%	76.3%	16.7%
17	375	32	343	1.9%	5.9%	94.9%	82.2%	12.7%
18	375	33	342	2.0%	5.9%	96.9%	88.0%	8.9%
19	375	29	346	1.8%	5.9%	98.7%	94.0%	4.7%
20	375	22	353	1.3%	6.0%	100.0%	100.0%	0.0%
Totals	**7,497**	**1656**	**5841**	**100.0%**	**100.0%**			

The model performed well. The model governance should know that all the deciles either performed well or need to be reevaluated in three months. Since some of the deciles performed not up to the threshold it would probably be a good idea to reevaluate the entire model in three

months. This model did have the highest overall KS statistic in decile 6 being at 42.6%. All in all, this model is performing very well and looks like it can be a very good performance model.