Northwestern MSDS-498

Model #101: Credit Card Default Model

Performance Validation Results

Andrew Stevens

September 1, 2022

1. The Production Model

To develop a logistic regression model, the feature importance from two of the other models developed for this study, Random Forest and XGboost, were calculated and the top three were taken from each. The resulting explanatory variables are: pay_max, payment_avg, payment_max, ratio_avg, balance_growth_6mo. Logistic regression is one of the simplest data science approaches to fitting a model – using an algebraic expression with the number of variables chosen as a hyperparameter. The coefficient for each variable is provided in table 1. A review of the p-values tells that all the input variables are likely to be relevant, as the p-values are very low.

Table 1: Logistic Regression Model Summary

Variable Name	coef	std err	z	P-values
payment_avg	.55	3.45E-06	0.715	1.66834366e-03
ratio_avg	-1.20 e-04	0.003	-0.62	0.00000000e+00
balance_growth_6mo	1.59 e-05	0.088	-3.351	0.00000000e+00
payment_max	-.11	0.018	-5.356	0.00000000e+00
pay_max	-4.88 e-04	0.021	34.966	2.44573664e-04

2. Model Development Performance

The logistic regression model has proven to generalize well, as it performs similarly on the test data set in comparison to the training set. The overall performance, however, may not be sufficient to ensure business profitability, dependent on the cost function.

a. Training Data Model Performance

It is expected that the model will perform best on the training data set, as this was where the optimization was performed for selecting the coefficients. The Kolmogorov-Smirnov (KS) test statistic measures “goodness of fit” through the degree of separation between the positive and negative distributions (Darling, 1957). A perfect score would be 100, with complete separation; a zero would be a completely useless model providing no separation. For the variable selection-based logistic regression model, the KS is 63.7% at decile 11. This indicates poor performance. This is confirmed with the ROC-AUC curve, displaying an AUC of only thirty-eight percent.

Table 2: Logistic Regression Model Kolmogorov-Smirnov chart: Training Data

Decile	min_prob	max_prob	events	nonevents	event_rate	nonevent_rate	Cum eventrate	Cum noneventrate	Abs(KS)
1	0.847361	1.000000	621	567	5.80%	4.34%	5.80%	4.34%	1.5
2	0.789447	0.847201	873	317	8.15%	2.43%	13.94%	6.77%	7.2
3	0.751708	0.789324	867	322	8.09%	2.47%	22.03%	9.23%	12.8
4	0.745929	0.751694	1148	40	10.71%	0.31%	32.75%	9.54%	23.2
5	0.739833	0.745928	1126	63	10.51%	0.48%	43.26%	10.02%	33.2
6	0.731642	0.739803	1088	101	10.15%	0.77%	53.41%	10.80%	42.6
7	0.714909	0.731622	1024	162	9.56%	1.24%	62.97%	12.04%	50.9
8	0.691021	0.714889	859	328	8.02%	2.51%	70.98%	14.55%	56.4
9	0.657596	0.690936	662	531	6.18%	4.07%	77.16%	18.61%	58.5

10	0.630473	0.657556	349	626	3.26%	4.79%	80.42%	23.41%	57.0
11	0.618600	0.630473	1027	374	9.58%	2.86%	90.00%	26.27%	63.7
12	0.594564	0.618599	304	881	2.84%	6.75%	92.84%	33.02%	59.8
13	0.571472	0.594539	164	1030	1.53%	7.89%	94.37%	40.90%	53.5
14	0.553281	0.571454	112	1076	1.05%	8.24%	95.42%	49.14%	46.3
15	0.537969	0.553279	51	1138	0.48%	8.71%	95.89%	57.86%	38.0
16	0.526030	0.537948	44	1145	0.41%	8.77%	96.30%	66.62%	29.7
17	0.517093	0.526028	14	1172	0.13%	8.97%	96.43%	75.60%	20.8
18	0.510239	0.517059	39	1152	0.36%	8.82%	96.80%	84.42%	12.4
19	0.504224	0.510236	132	1046	1.23%	8.01%	98.03%	92.43%	5.6
20	0.500010	0.504222	211	989	1.97%	7.57%	100.00%	100.00%	0.0

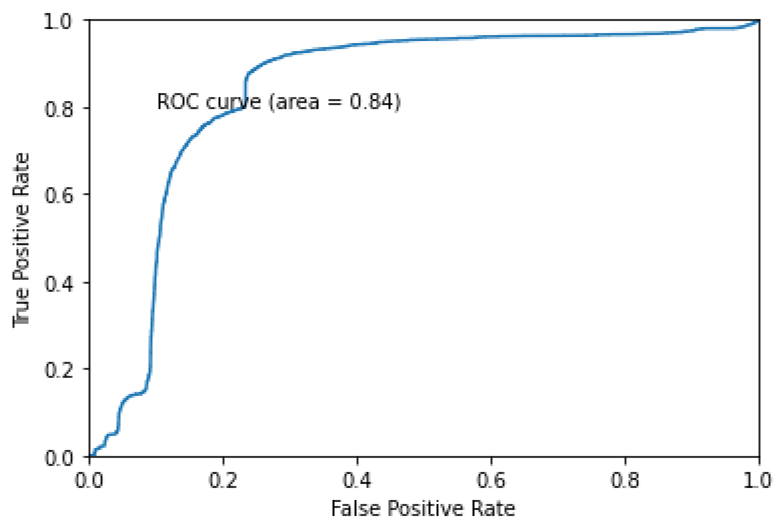

Figure 1: ROC: Training Performance

b. Test Data Model Performance

When the model was executed using the test data set, the KS is 57.4% at decile 10 – slightly improved from the training set, though the inclusion of the ROC-AUC plot reveals that overall performance is in fact slightly worse.

Table 3: Logistic Regression Kolmogorov-Smirnov chart: Test Data

Decile	min_prob	max_prob	events	nonevents	event_rate	nonevent_rate	Cum eventrate	Cum noneventrate	Abs(KS)
1	0.844020	1.000000	132	232	5.31%	4.85%	5.31%	4.85%	0.5
2	0.761118	0.843945	172	191	6.92%	4.00%	12.22%	8.85%	3.4
3	0.748044	0.761095	316	47	12.71%	0.98%	24.93%	9.83%	15.1
4	0.740543	0.748020	336	27	13.51%	0.56%	38.44%	10.40%	28.0
5	0.728590	0.740524	312	52	12.55%	1.09%	50.99%	11.49%	39.5
6	0.708947	0.728516	267	96	10.74%	2.01%	61.72%	13.50%	48.2
7	0.680528	0.708897	210	153	8.44%	3.20%	70.16%	16.70%	53.5
8	0.652125	0.680519	116	247	4.66%	5.17%	74.83%	21.87%	53.0
9	0.630474	0.652074	63	225	2.53%	4.71%	77.36%	26.57%	50.8
10	0.615234	0.630473	258	181	10.37%	3.79%	87.74%	30.36%	57.4
11	0.595886	0.615219	75	288	3.02%	6.03%	90.75%	36.39%	54.4
12	0.575906	0.595753	56	307	2.25%	6.42%	93.00%	42.81%	50.2
13	0.558044	0.575895	25	339	1.01%	7.09%	94.01%	49.91%	44.1
14	0.544576	0.558022	17	346	0.68%	7.24%	94.69%	57.15%	37.5
15	0.532980	0.544536	7	356	0.28%	7.45%	94.97%	64.60%	30.4
16	0.523318	0.532968	6	357	0.24%	7.47%	95.22%	72.07%	23.1

17	0.515433	0.523315	9	355	0.36%	7.43%	95.58%	79.49%	16.1
18	0.509821	0.515431	7	356	0.28%	7.45%	95.86%	86.94%	8.9
19	0.504226	0.509784	26	337	1.05%	7.05%	96.90%	93.99%	2.9
20	0.500070	0.504222	77	287	3.10%	6.01%	100.00%	100.00%	0.0

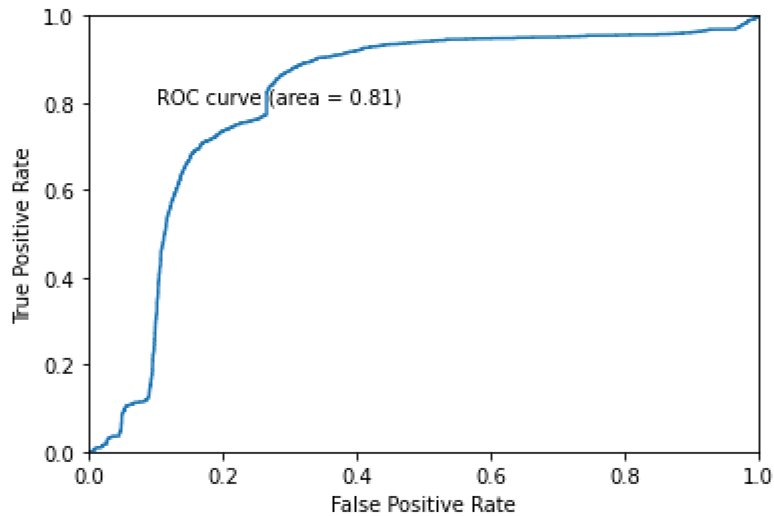

Figure 2: ROC: Test Performance

3. Performance Monitoring Plan

In order to ensure continued positive results, the model performance must be monitored through test statistics and thresholds must be established for when action is required. Specific performance values are not as relevant for a business as the impacts they produce. If a model accuracy reduction of five-percent results in twenty-percent increase in costs, 50-percent decrease in profits, five-percent is much worse in reality than it initially seems. A more relevant performance plan may require a more complex model. Classification of defaulting customers binarily may be insufficient. The model may also need to monitor and predict severity of that default, with high loss (either graduated or threshold of dollar amount) being recognized as another risk factor. To begin, a simple plan can be devised.

Table 4: RAG Thresholds

Status	KS threshold	AUC threshold
Green	>55%	>80%
Amber	50-55%	75-80%
Red	<50%	<75%

These thresholds are set assuming that current performance is sufficient – though that is not likely the case with the low initial performance values. It can also be helpful to monitor multiple measures if one is insufficient for recognizing all areas of potential loss. While the model continues to score in the green domain no action is necessary, and the model is considered healthy. Performance falling into the amber region would require more active monitoring and preparation for response within three months with re-validation. When KS or AUC falls below the red threshold, immediate redevelopment is required. The thresholds will also be periodically re-evaluated annually to determine if business impacts have changed and require corresponding adjustment of the monitoring plan.

4. Performance Monitoring Results

When the model was executed using the validation data set, the KS is 57.4% at decile 10 – identical to the test data performance. This aligns with logic since the test and validation sets were not used in training the logistic regression model. The KS and AUC are both within Green status, so no action is necessary. Validation set will be retested at the next interval.

Table 5: Logistic Regression Kolmogorov-Smirnov chart: Validation Data

Decile	min_prob	max_prob	events	nonevents	event_rate	nonevent_rate	Cum eventrate	Cum noneventrate	Abs(KS)
1	0.844855	1.000000	129	244	5.21%	4.90%	5.21%	4.90%	0.3
2	0.763020	0.844816	158	215	6.39%	4.31%	11.60%	9.21%	2.4
3	0.748299	0.762923	332	41	13.42%	0.82%	25.02%	10.03%	15.0
4	0.739267	0.748296	347	26	14.03%	0.52%	39.05%	10.56%	28.5
5	0.726594	0.739214	310	62	12.53%	1.24%	51.58%	11.80%	39.8
6	0.705583	0.726572	262	111	10.59%	2.23%	62.17%	14.03%	48.1
7	0.676429	0.705561	197	176	7.96%	3.53%	70.13%	17.56%	52.6
8	0.643865	0.676346	123	250	4.97%	5.02%	75.10%	22.58%	52.5
9	0.628869	0.643736	199	174	8.04%	3.49%	83.14%	26.07%	57.1
10	0.610942	0.628862	128	244	5.17%	4.90%	88.32%	30.97%	57.4
11	0.590840	0.610860	72	301	2.91%	6.04%	91.23%	37.01%	54.2
12	0.572703	0.590802	38	335	1.54%	6.72%	92.76%	43.73%	49.0
13	0.556909	0.572681	29	344	1.17%	6.90%	93.94%	50.63%	43.3
14	0.542815	0.556873	20	353	0.81%	7.08%	94.75%	57.72%	37.0
15	0.532046	0.542763	10	362	0.40%	7.26%	95.15%	64.98%	30.2

16	0.522812	0.532034	7	366	0.28%	7.34%	95.43%	72.33%	23.1
17	0.515352	0.522782	6	364	0.24%	7.30%	95.68%	79.63%	16.0
18	0.509529	0.515346	6	370	0.24%	7.43%	95.92%	87.06%	8.9
19	0.504222	0.509512	38	324	1.54%	6.50%	97.45%	93.56%	3.9
20	0.500009	0.504222	63	321	2.55%	6.44%	100.00%	100.00%	0.0

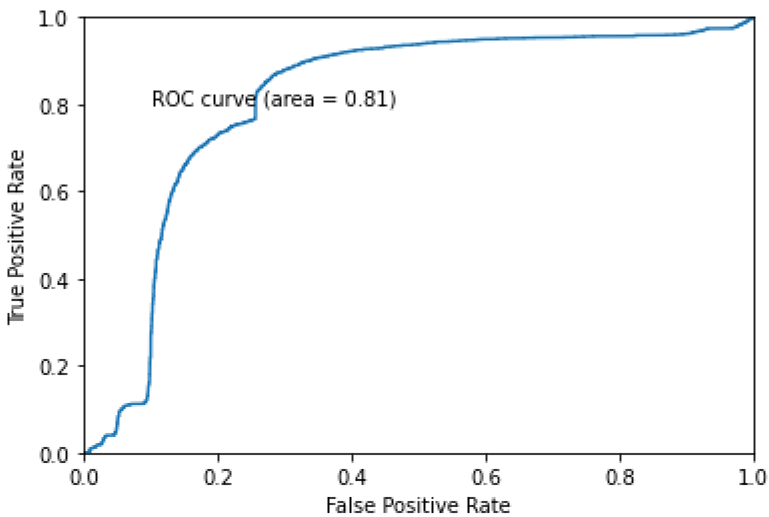

Figure 3: ROC: Validation Performance

5. Bibliography

1. Model Evaluation. https://www.saedsayad.com/model_evaluation_c.htm. Accessed 20 Aug. 2022.

2. Darling, D. A. "The Kolmogorov-Smirnov, Cramer-von Mises Tests." The Annals of Mathematical Statistics, vol. 28, no. 4, Dec. 1957, pp. 823–38. DOI.org (Crossref), https://doi.org/10.1214/aoms/1177706788.