Credit Risk Scorecard Monitoring

Published on May 17, 2017

Shailendra S
Hiring Machine Learning Rockstars :)

Introduction

Nowadays, Retail Banks are more focused on finding or discriminating the right clients and the wrong ones (Defaulters). From a Credit Risk perspective a Good Client will be a customer/applicant who has least chances to do default (a row risk client) i.e. the applicant has low chances to perform default in his obligations. This detection process of identifying or separating a Good & bad applicant/client is where Credit Risk Scorecard comes into play. It is an automated application which help banks to consistently assess each client in a shorter period of time ie. To detect his chances of delinquency (Probability of Default). With integration of loan approval I.T applications it help banks to speed up the loan application process, reducing the man hours resulting in increasing in the productivity with complete transparency.

Credit Risk Scorecard is basically a group of features, which is statistically determined to be predictive in distinguishing Good and Bad applicants.

Problem

Financial institutions have started paying lot of attention in tracking the performance of existing Scorecard Models. This tracking of PD models helps them in understanding the population shift in their data and knowing the change in delinquency pattern of users which will help them in improving their bad definition and validating clients.

In this article we take up a problem to propose and implement a simple scorecard tracking methodology that can be used by Banks to track their Credit risk Models. In this we will be focusing on how to measure the performance of existing PD model by tracking its performance over the Development and Current data. This will give us information's on how well the Scorecard is performing, whether it needs any further tuning or not, the current Bad definition still holds good for the current data or not. Helping them in visualizing insights and distribution on the current population shift and determining existing delinquency pattern of customers which will further help financial constitutions to derive a new Bad definition depending upon the current delinquency patterns of customers.

However for different models we will have different variables/characteristics but more or less the tracking process will be the same.

Investigation

gather the data with all the required fields in a specific database format. I a
generally include the "Good" "Bad" definitions, establishing the Developm
windows, defining the data exclusions. The following are the general Data
generally collected from applications from past two to five years depending
requirement:

- Unique Identification number/Loan ID
- Loan Application Date
- Accept/Reject Indicator
- Loan Disbursement Date
- Product code
- Emi Frequency
- Emi Date
- EMI payment date
- Current account Status

Information Gathering :

Scorecards are developed using an assumption that performance shown by clients in future will reflect the performance of the past users i.e assuming the economic conditions as constant we judge customers on the basis of past customer's delinquency pattern.

In order to perform this tracking analysis we need to gather data for accounts opened during a specific time period and then monitor their performance for another specific period of time and see whether there is any change in any delinquency pattern in these two data. We need to create two data sets from the user database ie. Development data, Recent Data. Based on these two data sets we will be performing tracking analysis and will see whether current Bad definition still holds good for Recent Data. Monitoring the scorecard is divided into Front-End and Back-End analysis :

1. Front-End – Measures the degree of variance in score distribution between the Development and Current populations. It includes:

a. Population Stability Index (PSI) :Measures the change in score distribution.

b. Characteristic Stability Index :Measures the change at characteristic le compares this performance to that of Development population.

a. Vintage Analysis – Analyse portfolio performance on a vintage basis.

b. Gini & KS comparisons – Quantifies the strength of the scorecard.

Front End Analysis

Population Stability Index : Population stability measures changes in the between the development population and the recent population. This chang distribution is measured by the population stability index. It an important m shift in population for credit scorecards.

A good place to start this comparison is by checking how two populations a across the risk bands created through the scorecard. The following is a repr of the shift in population among Development and Current data. Here Curr the population distribution for the Current data and 'Development Apps%' distribution for the Development data in each score bands. This change in s is measured by the Population Stability Index, which for this analysis is 0.0 little or no change in the score distributions.

Population Stability Report								
Development Window: May 2013 - April 2014								
Current Window: August 2015 - July 2016								
Score	Development apps #	Current apps #	Development	Current	Change (decimal) (5)-(4)	Ratio (5)/(4)	WoE ln(7)	Contribution to Index (8)x(6)
1	2	3	4	5	6	7	8	9
0-261	3,738	3,023	10%	8%	-2%	0.76	-0.27	0.01
262-273	3,491	3,761	10%	10%	0%	1.01	0.01	0
274-283	3,787	4,001	10%	10%	0%	0.99	-0.01	0
284-291	3,493	4,907	10%	13%	3%	1.32	0.28	0.01
292-298	3,004	3,438	8%	9%	1%	1.08	0.07	0
299-305	3,378	4,006	9%	10%	1%	1.12	0.11	0
306-312	3,329	3,868	9%	10%	1%	1.09	0.09	0
313-330	6,345	6,505	17%	17%	-1%	0.96	-0.04	0
331-341	3,005	2,496	8%	6%	-2%	0.78	-0.25	0
342+	2,867	2,723	9%	7%	-1%	0.89	-0.11	0
Total	**36,437**	**38,728**	**100%**	**100%**				**0.02**

The last column in the above table is what we care for. Index for each score band which is calculated by taking the difference between Recent and development applications (in %) and multiplying it with the log of ratio between current and development.

PSI is calculated by :PSI=((Recent% - Dev%) * (ln (Current%Dev%))

below:

PSI Value	Inference	Actic
Less than 0.1	Insignificant change	No action required
0.1 – 0.25	Some minor change	Check other scorecard m
Greater than 0.25	Major shift in population	Change occurred, perform analysis

In this case we have observed a slight movement from the lower and highe toward the middle score bands, however, the shift is immaterial. As we are PSI value according to the benchmark so we won't need a further Characte case. So, the Population Stability Index is one of the metrics to keep a chec conditions – however, the idea is clear that one has to capture robust metric look on the ever changing economic winds to prevent a crash landing.

Characteristic Analysis:

Characteristic Analysis measures the change between development and the population at a characteristic level. The change in distributions are added for each characteristic to estimate the impact on the total score. It answers which variable is causing a shift in population distribution. It compares the distribution of an independent variable in the current data set to a development data set. It detects shifts in the distributions of input variables that are submitted for scoring over time.

It helps us to determine which variable has the most influence in causing the model score shift.In above case as the change at an overall score level was insignificant, the change at a characteristic level will follow the same trend. If a scorecard constitutes 9 characteristics and a change in any of these 9 characteristic of more than 5 points is deemed significant. In below example we will see how to perform a Characteristic Analysis for one variable (Salary) we can use the same methodology on the remaining characteristics to detect the influencing variable

Attribute	Dev apps #	Current apps #	Dev %	Current %	(decimal) (5)-(4)	Ratio (5)/(4)	WoE ln(7)	n to Index (8)x(6)	at
1	2	3	4	5	6	7	8	9	
<3L	5,298	4,564	15 %	12%	-3%	0.81	-0.21	0.01	
>=3L-<6L	5,308	4,853	15 %	13%	-2%	0.86	-0.15	0	
>=6L-<10L	3,410	3,287	9%	8%	-1%	0.91	-0.1	0	
>=10L-<20L	3,665	4,298	10 %	11%	1%	1.1	0.1	0	
>=20L	18,756	21,726	51 %	56%	5%	1.09	0.09	0	
Total	36,437	38,728	100 %	100%				0.01	

Finally the summation of this score is used to detect whether there is any sh variable or not. Salary has a score difference of 0.54, indicating little or no Generally a characteristic with more than score difference of 5 is considere significant.

Final Decision report

In the end a Final Decision report is prepared by analyzing accept rate for each score band. This report is generated for both Development & Recent populations and the trend in acceptance rate is compared using this final report. The Final Decision report analyses accept rate by score band. The expected trend is accept rate increasing as score increases, thereby minimizing the impact of riskier accounts on the portfolio performance.

Back End Analysis

Vintage Analysis: Vintage Analysis in credit Risk models helps you to understand the Maturity of a portfolio and to establish the independent variable. The independent variable in credit risk modelling usually depends of the maturity and the default point. For example, one of the standards in Basel II is to modeling the probability that a client hit the 90 day past due during the next 12 months. The examination of loans by the period in which they were originated is knows a Vintage Analysis. In addition of monitoring your portfolio's performance we are also analyzing how loans of different age are performing i.e how they performs over their life span. It tells when the delinquency rate gets constant over a tenure so that we can set a trademark or we can fix a period throughout which we gonna monitor each applicant. This vintage curve is monitored for both the Development and Current data across all the 30,60,90 DPD and the change in curve is monitored across both the data. It helps in knowing whether the current Bad definition holds for the current data or not. It checks for the shift in delinquency pattern of applicants over the loan tenure.

Roll Rate Analysis : Comparing worst delinquencies in a specified previous "n" months with respect to the next "n" months and then calculating the number of accounts(in %) who maintained their delinquency or got better/worst.It is a method of analyzing

move into backward delinquent bucket or move into forward delinquency
worse with time. In simpler words its purpose is to evaluate the probability
whether he will move into forward bucket ie getting worse or moving into
getting better with time. It gives an idea on no return point ie. it helps to ide
no return. Typically customers who reach 90+ DPD bucket have the least c
cured or to roll back thus confirming the bad definition.

		Development Window: May 2013 - April 2014				
		Current Month Bucket				
	Delinquency	Current	0-29 DPD	30-59 DPD	60-89 DPD	90+ I
	Current	89%	11%	0%	0%	
	0-29 DPD	78%	18%	4%	0%	
	30-59 DPD	11%	43%	38%	6%	
	60-89 DPD	9%	25%	16%	11%	
Previous Month	90+ DPD	3%	2%	1%	1%	

For example in this table we are comparing delinquency buckets of present month with respect to previous month. In Row 1 11% of the applicants who were in current (zero delinquency) bucket went to 0-29 DPD. In Row 2 78% applicants went from 0-29 DPD to Current bracket (roll backward), 4% went to 30-59 DPD (roll forward) ie. went into higher delinquency bucket. As we can see in the last column of the table there is a “Backward roll rate” column so what does it infer??

In Row 3 we can see that 54% of the applicants who were in 30-59 DPD last month rolled backward to previous DPD bucked ie. improvement in their delinquency. While in Row 5 only 7% of the applicants went to previous buckets. So our definition of bad ie. 90 DPD makes more sense. Conversely 0-29 or 30+ DPD will be less significant as 78% and 54% of the customers respectively have the chances to roll back to previous buckets thus confirming out Bad Definition of 90+ DPD as correct.

We should be Comparing this analysis for both the Development and Recent Population data and see if there’s any significant change in BAD definition or not ie. the point of no return has changed or not .

		Current Month Bucket				
	Delinquency	Current	0-29 DPD	30-59 DPD	60-89 DPD	90+ D
	Current	89%	11%	0%	0%	
	0-29 DPD	72%	22%	6%	0%	
	30-59 DPD	9%	39%	44%	6%	
	60-89 DPD	9%	21%	18%	9%	
Previous Month	90+ DPD	1%	2%	1%	1%	

- The backward roll rate for 60-89 DPD for both the development (50%) (48%) populations are largely in line, however, there has been a deteriorati backward roll rate for the 90+ DPD from development (7%) to the recent ((If the backward roll rate for the 90+ DPD segment improved significantly population, the bad definition could potentially be inappropriate)

- As the backward roll-rate has slightly deteriorated for the earlier delinc and more for the 90+ DPD range, the chosen bad definition of 90+ DPD ov window holds for both the development and recent populations

Gini and KS Statistics: It is a scorecard performance statistics which measure the overall strength of the scorecard in separating Good and Bad accounts. This was conducted for both the Development and Recent populations.

KS Statistics

This technique is mostly used to validate a PD model ie. its ability to discriminate between Good and Bad customers giving us the predictive power of the existing model. It is a point estimate and tells the score-band where the difference between cumulative good and cumulative bad customers is maximum.

Ks =Cumulative % Event-Cumulative % Non Event

Steps :

1. Arrange the scorebands (deciles) in increasing order.

2. Find the difference between Cumulative % Good and Cumulative % Good for each decile.

3. The maximum difference is the Ks value.

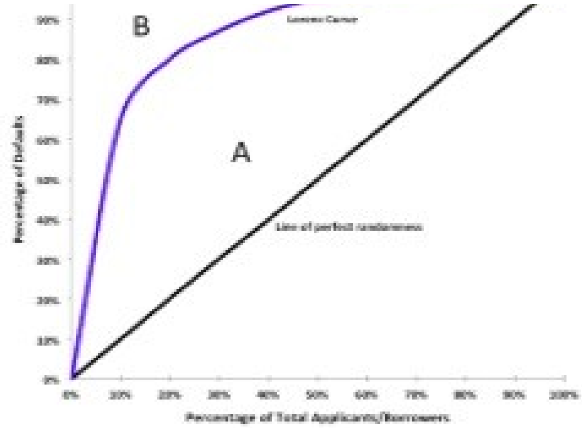
Ideally KS value should be in first of the 3 Deciles and score lying between 40-70%.

Ginni Coefficient

It a common measurement technique which helps us to measure the effecti scorecard in discriminating Goods/Bads. It is mostly used for assessing pre a credit risk model. It measures the degree to which the model has better di power than the model with random scores.

The GINI co-efficient for measuring credit models also has values between higher value means that a particular credit model can better discriminate an risky borrowers. A value of 1 means that the model predicts perfectly, and which borrowers will repay and which borrowers will default. A value of 0 model is completely random, or in other words, it is the statistical equivale resulting in a 50/50 probability of repayment or default for each applicant.

Ginni is nothing but the ratio of the area under the curve (A) but above the randomness to the entire area above the line of perfect randomness (A+B). Lorentz curve the cumulative distribution with the line of perfect randomne below.

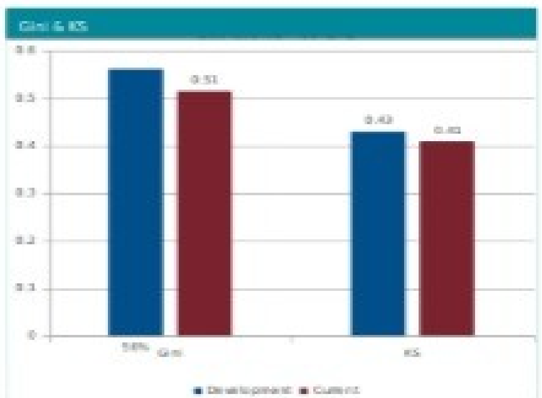


Comparing Ks and Ginni Stats for Development and Recent

Now we will be comparing these scorecard statistics on both th
and Recent data and see if there is any significant change or no

Development Group	Score_Band	% Good	% Bad	% Cum Good	% Cum bad		
0					0		
1	171-261	8.70%	42.40%	8.70%	42.40%	33.70%	1.90%
2	262-273	8.80%	15.10%	17.50%	57.50%	40.00%	4.40%
3	274-283	9.60%	12.80%	27.10%	70.30%	43.20%	6.10%
4	284-291	8.20%	6.80%	35.30%	77.10%	41.80%	6.50%
5	292-298	8.50%	6.20%	43.80%	83.30%	39.50%	6.80%
6	299-305	9.30%	4.20%	53.10%	87.50%	34.40%	8.10%
7	306-312	9.10%	4.10%	62.20%	91.60%	29.40%	8.30%
8	313-330	19.10%	5.70%	81.30%	97.30%	16.00%	17.50%
9	331-341	9.10%	2.10%	90.40%	99.40%	9.00%	9.10%
10	342-404	9.60%	0.60%	100.00%	100.00%	0.00%	9.40%
		100.00%	100.00%				56.00%

Group	Score_Band	% Good	% Bad	% Cum Good	% Cum bad	KS
0					0	
1	171-261	10.00%	43.00%	10.00%	43.00%	33.00
2	262-273	9.90%	16.40%	19.90%	59.30%	39.40
3	274-283	9.20%	10.70%	29.10%	70.10%	41.00
4	284-291	10.40%	6.80%	39.50%	76.90%	37.40
5	292-298	8.80%	4.80%	48.30%	81.70%	33.40
6	299-305	9.90%	5.50%	58.20%	87.10%	28.90
7	306-312	9.90%	5.00%	68.10%	92.10%	24.00
8	313-330	17.10%	5.10%	85.20%	97.20%	12.00
9	331-341	7.00%	2.00%	92.20%	99.20%	7.00
10	342-404	7.80%	0.70%	100.00%	100.00%	0.00
		100.00%	100.00%			



From the above table and plotted graph we can make following observation:

- The Gini Coefficient value has slightly reduced from 56% at time of development, to 51% for the recent population. A Gini of 51% is classified as very strong for an application scorecard and hence the predictive power of the scorecard will add value to the decision making process.

- The KS has reduced marginally from 43% to 41% which is still a good score for an application scorecard.

Conclusion

In this step-by-step guide to application scorecards tracking we have covered all the steps in tracking application scorecards using PSI, Characteristic Analysis, showing vintage analysis, rollrates, scorecard statistics etc. The following are the conclusions that we have drawn from our above problem case:

distributions in recent population as compare to development data.

- As the Population Stability Index does not indicated any significant ch distributions, so similarly, there are no characteristics indicating a shift in t This is an optional step it is needed only when we have a PSI>.25 showing distribution.

- Roll Rate analysis (Backward) showed us that our current default defi still holds good for the recent population.

- A marginal decrease in the Gini coefficient was observed, from 56% a 51% for the current period. A Gini of 51% is still considered strong for app scorecards as any value below 30 represents a weak model. The KS has red from 43% to 41%.

However we need to continue with the scorecard monitoring to ensure stab scorecard performance, specifically on population shift and certain characte monitoring going forward.

More from Shailendra S

Introduction to Reinforcement Learning
Shailendra S on LinkedIn

Moving towards world powered by Artificial Intelligence & Deep Learning.
Shailendra S on LinkedIn

Collecting Twitter Stream : Using Python & MongoDB
Shailendra S on LinkedIn

ic

See all 12 articles