

Credit Scoring

Mario Melchiori

6/5/2019

Objetives

The bank for which we do this work is active in the loans with repayment through payroll deductions a employees of the state government. Those employees have legal job stability, so is very difficult that an employee is fired. Given both, job stability and the sure payment, the default probability is very low. But as we will see in last paragraph of *Debtor classification* section below, it is posible that as only one-level discrepancy is allowed in relation to the information submitted by financial institutions to the Credit Information Data Base, if there is a greater discrepancy between the rating of the bank and the lower classification awarded by at least two other banks, and total loans from such banks account for 40% or more of the total informed, the our bank will be required to reclassify the debtor to at least the level immediately above from registering the highest level of indebtedness with the comparison institutions. In practical terms, the bank will must provisioning even though the debtor is normal on its books

The objetive of this project is pick the variables that predict if a debtor whose classification differs from ours, but is not yet required to be reclassified for not fulfilling some of the three conditions set forth in the regulations:

- a discrepancy of two or more notches between the rating of the bank and the lower classification awarded
- by at least two other banks, and
- the total loans from such banks account for 40% or more of the total informed,

complies them, within the year in which the data was selected.

The bank also requires build a score system and pick a cutoff that reach a reclassified rate lesser than 0.045, but allowing high approval rate.

Regulations

Debtor classification, provisioning and collateral guarantees

Debtor classification

Regulations establish guidelines for classifying debtors from the point of view of their credit quality and compliance with their commitments, according to the evaluation performed for that purpose by the financial institution.

- The guidelines vary depending on whether commercial consumer or housing loans are involved. Commercial loans of up to 40% of the determined amount (*) may be considered, for classification purposes, as consumption loans and treated as such at the bank's discretion.
- Debtors and all their loans are included in one of six categories or situations of decreasing credit quality:

Commercial Loans	Consumer or Housing Loans	Arrears
1. Normal ¹		up to 31 days
2. Special follow-up ²	2. Low risk	up to 90 days
3. Substandard	3. Middle risk	up to 180 days
4. High Insolvency Risk ³	4. High risk	up to 1 year
5. Unrecoverable		more than a year
6. Unrecoverable based on technical criteria ⁴		

- The basic criterion for evaluation is the repayment capacity in relation to the debt or the commitments that are the object of the financial institution guarantee. For the commercial portfolio, evaluation is made on the basis of repayment capacity and debtor cash flows, and in the case of consumer and housing loans it is based on debt payment compliance and the legal status of the debtor. Indicators used in the case of commercial loans include liquidity, financing structure, compliance with payment of obligations, quality of management and administration, IT systems, prospects for the customer's business sector, its position within the sector, its legal standing and the existence of refinancing or debt discounts. The evaluation criteria for the consumer and housing portfolios is exclusively objective - the degree of compliance with obligations and the legal situation of the debtor.
- When loans are fully collateralized by preferred class A collateral, evaluation of repayment capacity is not required.
- Minimum classification frequency: As a general rule, this should be annual. Nevertheless, classification should take place:
 - during the course of each quarter for customers whose debts are equivalent to 5% or more of the financial institution's RC;
 - during the course of each half-year term in the case of customers whose debts range from 1% (or the equivalent to the determined amount ⁵, whichever is lower) to less than 5% of the financial institution's RC.

In addition, the bank should review a debtor's situation when any of the following circumstances take place:

- a. when a debtor has debts equivalent to at least 10% of the total notified to the Credit Information Debtor Base in another financial institution, and that institution lowers the customer's rating on the mentioned database;
- b. when there are changes to any of the objective classification criteria (arrears or legal situation);
- c. when there is more than a one-level discrepancy between the classification assigned by the financial institution and at least two other institutions, and certain requirements have been met.

Re-appraisal must be carried out promptly in the case of customers with debts totaling 1% or more of the financial institution's RPC or the equivalent to the determined amount, whichever is lower.

Only one-level discrepancy is allowed in relation to the information submitted by financial institutions to the Credit Information Data Base. If there is a greater discrepancy between the rating of the bank and the lower classification awarded by at least two other banks, and total loans from such banks

account for 40% or more of the total informed, the bank will be required to reclassify the debtor to at least the level immediately above that registering the highest level of indebtedness with the comparison institutions.

Provisioning

- Loan provisioning must be performed on the basis of the classification assigned to the debtor. No provision is required for loans for up to 30 days granted to other financial institutions (if not past due), for loans granted to the public non-financial sector, or unused balances of current account overdraft agreements.
- The following minimum provisioning levels are to be applied on total debt:

Debtor Category Concept		With preferred collateral	Without preferred collateral
1.	Normal	1%	1%
2.	a) Under Observation and low risk	3%	5%
	b) Under negotiation or with refinancing agreements	6%	12%
3.	Substandard and Middle risk	12%	25%
4.	Doubtful and High risk	25%	50%
5.	Unrecoverable	50%	100%
6.	Unrecoverable based on technical criteria	100%	100%

- The SEFyC may require additional provisioning if it determines that the current level is inadequate.
- Accrual of interest on customer debts classified as “under negotiation or with refinancing agreements” when arrears of more than 90 days in the payment of obligations are recorded, and those in the “substandard” or “medium risk”, “high risk”, and “unrecoverable” categories must be provided for at 100% as from the moment they are classified in any of those categories. The financial institution may opt to interrupt interest accrual.
- Customer debt classified as “unrecoverable” and fully provided for must be written off as from the seventh month subsequent to that in which such actions were taken.
- Inclusion of debtors in the “unrecoverable based on technical criteria” category results in the obligation to provision loans at 100%, including renewals, stays, forbearance -express or tacit - granted after such classification, once 90 or 180 days have elapsed as from the date on which the first of such financing measures were taken.
- Provision for the normal portfolio is of a global nature, while in the case of the other categories the allocation of provisions for each debtor is made on an individual basis.
- New loans in pesos hedged against commodity price risk to which the debtor may be exposed are subject to the minimum provisioning established for the category immediately preceding that of the debtor.

Data

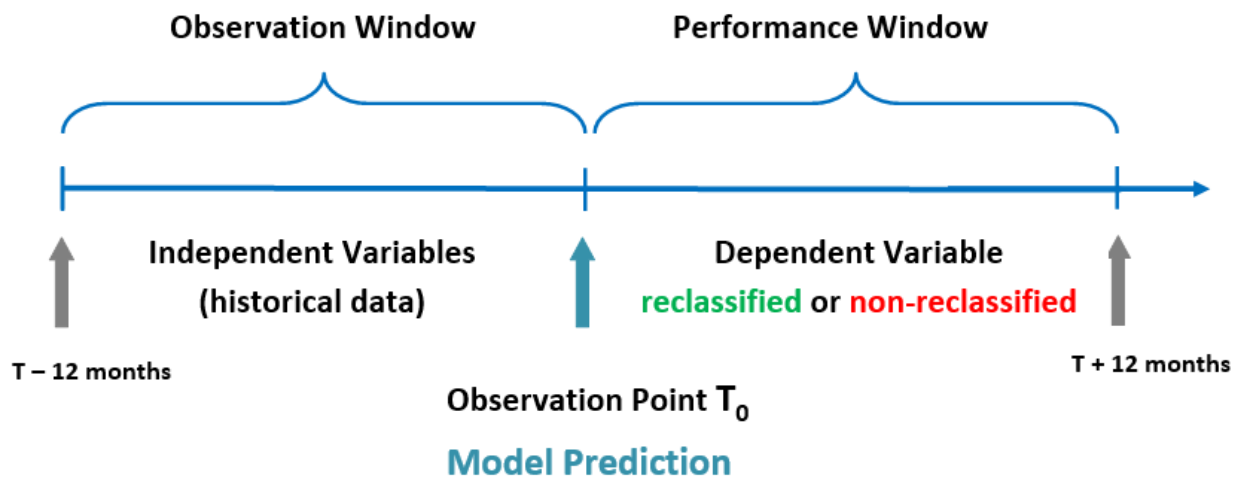


Figure 1. Utilising Historical Data to Predict Future Outcomes

The data were provided by the bank. The data were obtained from the Credit Information Data Base, which includes all debtors with the Argentine financial system. A filtering process was carried out in order to work with the target market, that is, human persons employed by the state government where the bank operates. With that target the bank only chose, in T_0 (Figure 1), those debtors whose classification differs from theirs, but is not yet required to be reclassified for not fulfilling some of the three conditions set forth in the regulations:

- a discrepancy of two or more notches between the rating of the bank and the lower classification awarded
- by at least two other banks, and
- the total loans from such banks account for 40% or more of the total informed. In short, data do not contains debtors that paying in time and form theirs debts in T_0 .

This selection was made in T_0 (Figure 1). The variables, depending to type (stock in T_0 or flow through $T-12$ to T_0), were compute from $T-12$ to T_0 . On $T+12$ was compute if the debtors reclassified or non-reclassified.

The debtors reclassified represent 16.42% of total. For more information about data to try cook book (<http://rpubs.com/mrmelchi/496179>).

The model

Logistic regression model is one of the most commonly used statistical technique for solving binary classification problem. It is an acceptable technique in almost all the domains. Two concepts - weight of evidence (WOE) and information value (IV) evolved from the same logistic regression technique. These two terms have been in existence in credit scoring world for more than 4-5 decades. They have been used as a benchmark to screen variables in the credit risk modeling projects such as probability of default. They help to explore data and screen variables. It is also used in marketing analytics project such as customer attrition model, campaign response model etc.

In this project we using this tools by changing the variable to predict, in this case is other than “default” like in most of credit risk model, instead, a new concept labeled “reclassified” is employed.

The model is completed through the following steps:

- Automated optimal Weight-of-Evidence (WOE) and Information Value (IV) binning is executed.
- The variates with Information Value (IV) bins with greater power are selected.
- Transformation all features within the data set so that the new data contains only weight-of-evidence valued transformations.
- Models using logistic regression is created.
- Credit scorecard is created.
- Cutoff that maximize the sensibility, that is the ability of our model to correctly identify those debtors reclassified (true positive rate) but allowing a high the approval rate, is computed.
- Good of fit statistic and Statistical measures of discriminatory power are calculated.

Logistic regression model is heavily researched. We do not include information about it, but you can try the following link (<https://rafalab.github.io/dsbook/examples-of-algorithms.html#logistic-regression>) for more information. Weight of evidence (WOE) and Information value (IV) are concepts less knowed, so below we develop a brief explanation about them.

Weight of Evidence (WOE)

Since this is a concept from credit scoring world, it is generally described as a measure of the separation of good and bad customers. “Bad Customers” refers to the customers who defaulted on a loan, in our case, reclassified a debt, and “Good Customers” refers to the customers who paid back loan, in our case, they did not reclassified.

WOE Calculation

Distribution of yes - % of reclassified Customers

Distribution of no - % of non-reclassified Customers

WOE is calculated by taking the natural logarithm (log to base e) of division of % reclassified Customers and % of non-reclassified Customers.

$$WOE = \log \left(\frac{\%non - reclassified}{\%reclassified} \right)$$

Steps of Calculating WOE:

- For a continuous variable, split data into n parts (n depends on the distribution).
- Calculate the number of reclassified and non-reclassified in each group (bin)
- Calculate the % of reclassified and % of non-reclassified in each group.
- Calculate WOE by taking natural log of division of % of non-reclassified and % of reclassified

Note : For a categorical variable, you do not need to split the data (Ignore Step 1 and follow the remaining steps)

Usage of WOE

Weight of Evidence (WOE) helps to transform a continuous independent variable into a set of groups or bins based on similarity of dependent variable distribution in our case: reclassified and non-reclassified.

- *For continuous independent variables*: Create bins (categories / groups) for a continuous independent variable and then combine categories with similar WOE values and replace categories with WOE values. Use WOE values rather than input values in your model.
- *For categorical independent variables*: Combine categories with similar WOE and then create new categories of an independent variable with continuous WOE values. In other words, use WOE values rather than raw categories in your model. The transformed variable will be a continuous variable with

WOE values. It is same as any continuous variable.

Why combine categories with similar WOE?

It is because the categories with similar WOE have almost same proportion of reclassified and non-reclassified. In other words, the behavior of both the categories is same.

Rules related to WOE

- Each category (bin) should have at least 5% of the observations. We relax this rule in this project.
- Each category (bin) should be non-zero for both non-reclassified and reclassified.
- The WOE should be distinct for each category. Similar groups should be aggregated.
- The WOE should be monotonic, i.e. either growing or decreasing with the groupings.
- Missing values are binned separately.

Number of Bins (Groups)

In general, 10 or 20 bins are taken. Ideally, each bin should contain at least 5% cases. The number of bins determines the amount of smoothing - the fewer bins, the more smoothing. If someone asks you ' "why not to form 1000 bins?" ' The answer is the fewer bins capture important patterns in the data, while leaving out noise. Bins with less than 5% cases might not be a true picture of the data distribution and might lead to model instability.

Handle Zero Event/ Non-Event

If a particular bin contains no event or non-event, you can use the formula below to ignore missing WOE. We are adding 0.5 to the number of reclassified and non-reclassified in a group.

$$\text{AdjustedWOE} = \log \left(\frac{\frac{\text{Number of non-reclassified in a group} + 0.5}{\text{Number of non-reclassified}}}{\frac{\text{Number of reclassified in a group} + 0.5}{\text{Number of reclassified}}} \right)$$

How to check correct binning with WOE

- 1. The WOE should be monotonic i.e. either growing or decreasing with the bins. You can plot WOE values and check linearity on the graph.
- 2. Perform the WOE transformation after binning. Next, we run logistic regression with 1 independent variable having WOE values. If the slope is not 1 or the intercept is not $\ln(\% \text{ of non-reclassified} / \% \text{ of reclassified})$ then the binning algorithm is not good.

Benefits of WOE

- 1. It can treat outliers. Suppose you have a continuous variable such as annual salary and extreme values are more than 500 million dollars. These values would be grouped to a class of (let's say 250-500 million dollars). Later, instead of using the raw values, we would be using WOE scores of each classes.
- 2. It can handle missing values as missing values can be binned separately.
- 3. Since WOE Transformation handles categorical variable so there is no need for dummy variables.

Information Value (IV)

Information Value Formula

Information value is one of the most useful technique to select important variables in a predictive model. It helps to rank variables on the basis of their importance. The IV is calculated using the following formula :

$$IV = \sum (\% \text{ of non reclassified} - \% \text{ of reclassified}) * WOE$$

Information Value Variable Predictiveness

According to Siddiqi ⁶, by convention the values of the IV statistic in credit scoring can be interpreted as follows.

IV	Predictiveness
Less than 0.02	Not useful for prediction
0.02 to 0.1	Weak predictive Power
0.1 to 0.3	Medium predictive Power
0.3 to 0.5	Strong predictive Power
>0.5	Suspicious Predictive Power

Luckily R has several packages to compute WOE and Information Value (IV). In this project we use the R package `smbinning` (<https://cran.r-project.org/web/packages/smbinning/index.html> (<https://cran.r-project.org/web/packages/smbinning/index.html>)).

This package uses *Supervised Discretization* that divides a continuous feature into groups (bins) mapped to a target variable. The central idea is to find those cutpoints that maximize the difference between the groups. In the past, analysts used to iteratively move from Fine Binning ⁷ to Coarse Binning ⁸, a very time consuming process of finding manually and visually the right cutpoints (if ever). Nowadays with algorithms like ChiMerge or Recursive Partitioning, two out of several techniques available, analysts can quickly find the optimal cutpoints in seconds and evaluate the relationship with the target variable using metrics such as Weight of Evidence and Information Value.

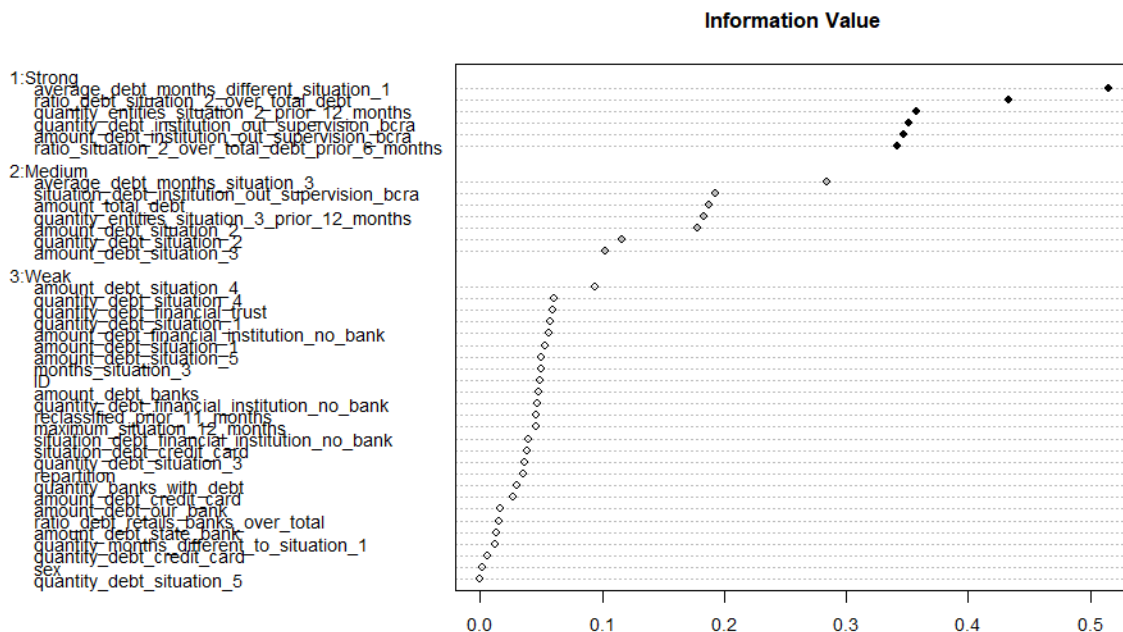


Figure 2. Variable Predictiveness Power

Below, we show the IV of the variable with the greatest predictive power 0.5147. You can see that the *BadRate* column is monotonic.

##	Cutpoint	CntRec	CntGood	CntBad	CntCumRec	CntCumGood	CntCumBad	PctRec
## 1	<= 3.9167	1236	53	1183	1236	53	1183	0.1361
## 2	<= 11.5	2279	212	2067	3515	265	3250	0.2510
## 3	<= 17.2223	1270	158	1112	4785	423	4362	0.1399
## 4	<= 24.1112	1122	188	934	5907	611	5296	0.1236
## 5	<= 36	1166	243	923	7073	854	6219	0.1284
## 6	<= 52.6667	898	239	659	7971	1093	6878	0.0989
## 7	> 52.6667	1108	398	710	9079	1491	7588	0.1220
## 8	Missing	0	0	0	9079	1491	7588	0.0000
## 9	Total	9079	1491	7588	NA	NA	NA	1.0000

##	GoodRate	BadRate	Odds	LnOdds	WoE	IV
## 1	0.0429	0.9571	0.0448	-3.1055	-1.4784	0.1779
## 2	0.0930	0.9070	0.1026	-2.2773	-0.6501	0.0847
## 3	0.1244	0.8756	0.1421	-1.9513	-0.3242	0.0132
## 4	0.1676	0.8324	0.2013	-1.6030	0.0241	0.0001
## 5	0.2084	0.7916	0.2633	-1.3346	0.2926	0.0121
## 6	0.2661	0.7339	0.3627	-1.0143	0.6129	0.0450
## 7	0.3592	0.6408	0.5606	-0.5788	1.0483	0.1817
## 8	NaN	NaN	NaN	NaN	NaN	NaN
## 9	0.1642	0.8358	0.1965	-1.6271	0.0000	0.5147

NOTE: Information value is not an optimal feature (variable) selection method when you are building a classification model other than binary logistic regression (random forest or SVM) as conditional log odds (which we predict in a logistic regression model) is highly related to the calculation of weight of evidence. In other words, it's designed mainly for binary logistic regression model. Also think this way: Random forest can detect non-linear relationship very well so selecting variables via Information Value and using them in random forest model might not produce the most accurate and robust predictive model.

Results

On test set, with a cutoff of 569 points the model reach a 0.3381 approval rate and 0.0408 reclassified rate. Statistical measures of discriminatory power would be considered to be “good” at separating reclassified from non-reclassified debtors. ROC curve is 0.7873, KS statistic is 0.4486 and Gini statistic is 0.5745.

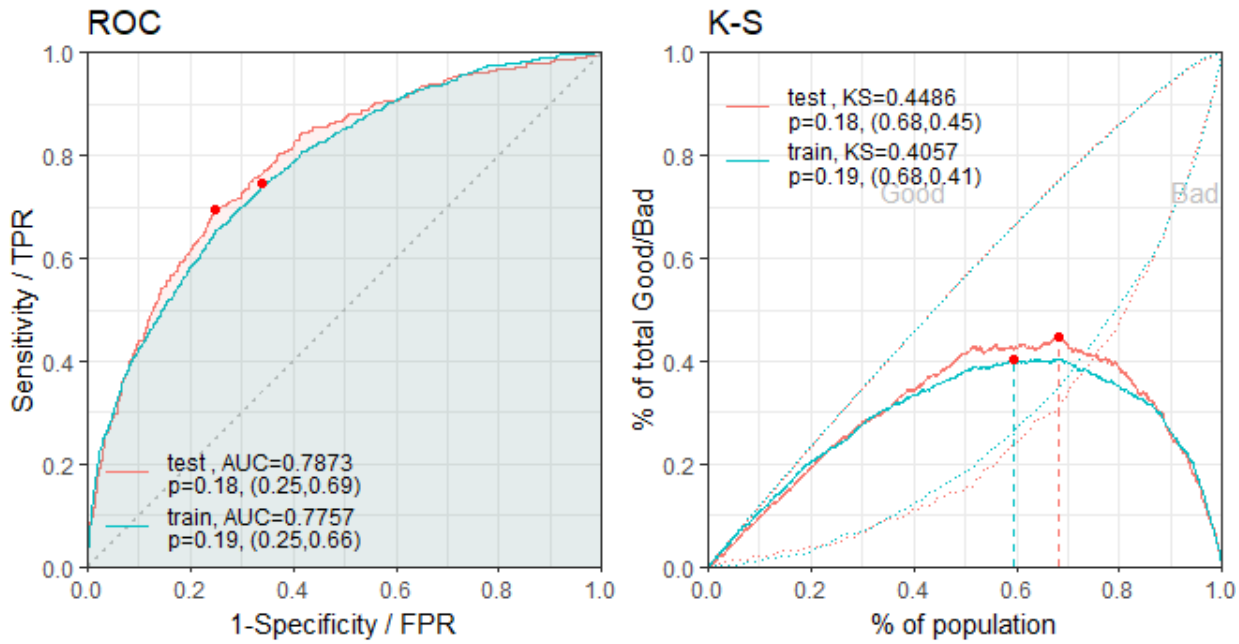


Figure 3. Statistical measures of discriminatory power

Also we conducted goodness-of-fit test, the result of Hosmer-Lemeshow Chi-square test and Spiegelhalter test suggest that according to the p-values, the estimated reclassified rate are quite close to the observed reclassified rate.

Finally, we compute Population Stability Index (PSI) to check how scoring on testing set is compared to the predicted probability from training data set. The rule of thumb for the PSI is as follows:

PSI	Predictiveness
Less than 0.1	Less than 0.1 inference insigni<U+FB01>cant change, no action required
0.1 to 0.25	inference some minor change, check other scorecard monitoring metrics
>0.25	inference major shift in population, need to delve deeper

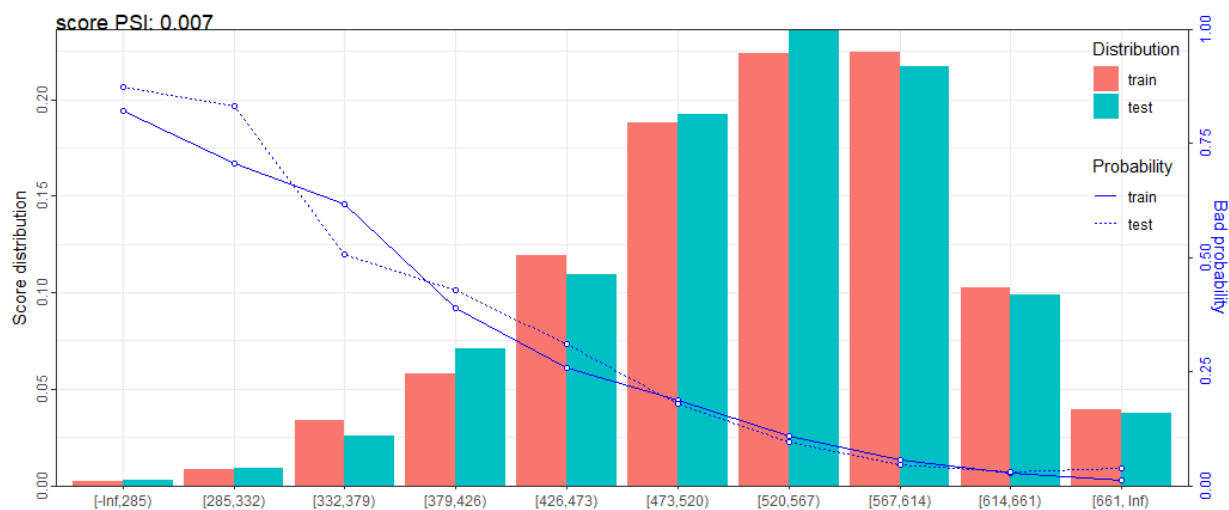


Figure 4. Population Stability Index (PSI)

PSI in our case is 0.007.

	train	test	overall
Data			
reclassified rate	0.1664	0.1590	0.1642
Model			
cutoff	0.0760	0.0760	
score cutoff - points -	569	569	
reclassified rate	0.0450	0.0408	
approval rate	0.3508	0.3380	
ROC	0.7757	0.7873	
KS	0.4057	0.4486	
Gini statistic	0.5513	0.5745	
psi			0.007
Hosmer - Lemeshow test p-value		0.2816	
Spiegelhalter test p-value		0.8815	

Figure 5. Results

Summary

A bank is active in the loans with repayment through payroll deductions a employees of the state government. Those employees have legal job stability, so is very difficult that an employee is fired. Given both, job stability and the sure payment, the default probability is very low, but argentine banking regulator requires that under certain circumstances a bank must provisioning (charge a loss) to debtor paying in a timely manner the debt contracted but delay pays of debt with other bank. This become very expensive for the bank, so, the bank want avoid give a loan to whose can fall in this situation.

The bank required us to design a model that predict if a debtor will be reclassified durant the next year.

The bank also requires build a score system and pick a cutoff that reach a reclassified rate lesser than 0.045, but allowing high approval rate.

The data were provided by the bank. The bank only chose those debtors whose classification differs from their, but is not yet required to be reclassified for not fulfilling some of the three conditions set forth in the regulations. In short, data do not contain debtors that paying in time and form their debts in all financial system Argentine neither debtors reclassified in T0 (Figure 1).

In this project we using logistic regression, Weight of evidence (WOE) and Information value (IV) but by changing the variable to predict, in this case is other than “default” like in most of credit risk model, instead, a new concept labeled “reclassified” is employed.

We find out a cutoff of 569 points, on test set the model reach a 0.3381 approval rate and 0.0408 reclassified rate. Statistical measures of discriminatory power would be considered to be “good” at separating reclassified from non-reclassified debtors. ROC curve is 0.7873, KS statistic is 0.4486 and Gini statistic is 0.5745.

The result of Hosmer-Lemeshow Chi-square test and Spiegelhalter test suggest that according to the p-values, the estimated reclassified rate from the model are quite close to the observed reclassified rate.

-
1. In the case of consumer or housing loans, current account overdrafts are considered to be performing until 61 days have elapsed from the date granted.↵
 2. Commercial loans in situation 2 are divided into: a) under observation, include those debtors up to 90 days past due in situations that if not controlled or corrected in a timely manner, could compromise their repayment capacity, and b) those under negotiation or with refinancing agreements, which include debtors that although unable to pay their obligations under the agreed conditions, have declared their intention of refinancing their debts no later than 60 days after becoming past due.↵
 3. This category includes debtors that have filed for creditor protection or an out-of-court preventive measure, or for which payment has been demanded in court. In the case of the consumer portfolio, debtors that have filed for creditor protection or are covered by out-of-court measures can record arrears of up to 540 days.↵
 4. This category includes debtors with arrears in excess of 180 days that are customers of banks that have been wound up or had their license revoked by the BCRA, residual entities of privatized banks, or trusts of which SEDESA is a beneficiary.↵
 5. The maximum value of total annual sales for the “micro” category belonging to the “commercial” sector pursuant to the regulations on “Determination of the condition of MiPyMEs” amounts to \$ 15,800,000.↵
 6. Credit Risk Scorecards: Developing and Implementing Intelligent Credit Scoring by Naeem Siddiqi, John Wiley and Sons, Sep 1, 2015↵
 7. Create 10/20 bins/groups for a continuous independent variable and then calculates WOE and IV of the variable.↵
 8. Combine adjacent categories with similar WOE scores.↵