

Predictive Modeling Approaches for Payroll Issuers

H.A. Pérez¹, J.A. Marmolejo², J.Velasco³, and J.G. Fuentes⁴
{hugo.perez@anahuac.mx¹, jose.marmolejo@anahuac.mx², jvelasco@cimat.mx³}

Center of Top Management in Engineering and Technology (CADIT), Anahuac University, Av. Universidad Anáhuac 46, Col. Lomas Anáhuac, Huixquilucan, 52786, State of Mexico, Mexico¹
Faculty of Engineering, Anáhuac University, Av. Universidad Anáhuac 46, Col. Lomas Anáhuac, Huixquilucan, 52786, State of Mexico, Mexico²
CONACYT Research Fellow -Center For Research in Mathematics, F. Bartolomé de las Casas 314, 20259, Aguascalientes, Mexico³

Abstract. Nowadays, in most banks, vast amounts of data are available in order to make business decisions and enhance the institution's know-how. The present study refers to transactional data systems used by companies that manage payroll outsourced services. We propose two practical approaches for analyzing this type information. One approach consists of testing traditional techniques for predictive modeling and, the other of building a credit score card using a credit scoring methodology. Several experiments were executed using specialized software in order to obtain the best credit score model for payroll issuers. Experimental results show that for most cases, decisions tree models are better than both logistic regression models and ensemble models. In one approach, we also show how the Quantile Grouping Method gives the lowest missclassification rate.

Keywords: Credit Scoring, Logistic Regression, Decision Tree, Artificial Neural Networks (ANNs), Ensemble Models.

1 Introduction

Typically, in financial institutions exists an area that functions evaluating customers for credit loans. Evaluation is a difficult activity especially since it is a core process for the banks. Credit scoring models have still a primary way to differentiate good from bad applicants, and the accuracy of credit scoring is critical to financial institution's profitability [1, 2]. In order to assure reliability regarding this issue, many data mining techniques have been implemented [3] for different credit score models [4]. Also, these models estimate the probability of default on individual loans or pools of transactions. A default is considered to have occurred with regard to a particular obligor (a customer), for example, when the bank considers that the obligor is unlikely to pay its credit obligations to the banking group in full, without recourse by the bank to legal actions such as realizing security [5]. Usually, many researchers interested in credit risk models are focused in studying data based on transactions of consumers or persons [6, 7]. However, in this work our customers are payroll issuers and that means that the company controls the dispersion of money to the final issuer. To date, our research group has not found a case as in this study. We built credit models following less conventional ways that allow to evaluate a number of techniques common in practice. We propose two approaches focused in a data transactional model from a Mexican bank. These approaches address two possible uses, i.e. prediction or classification. One consists of applying three

traditional techniques in data mining using all variables available. The second follows the methodology proposed in [8] which considers a characteristic analysis in order to select only significant variables that contribute to the predictive power of the model. In each approach, a comparison based on misclassification rate, Gini Index, Receiver Operating Characteristic (ROC) Index and Kolmogorov-Smirnov (KS) statistics was performed.

The paper is organized as follows. First, we present a brief review about credit scoring models and their applications. Then, we describe the case study and the database. Both approaches and their corresponding results are presented in the following sections. Finally, conclusions are presented and discussed.

2 Credit Scoring Models

Credit Scoring Models have been widely used as techniques for financial institutions regarding loan applications. An introduction to this topic is presented in [9]. In fact, some reviews about these models and their applications are shown in [10] and [11]. In [12] the authors give an ample description about the applications, and propose a method to construct these models using statistical techniques and Artificial Intelligence (AI) techniques. The authors compare techniques such as decision trees, logistic regression and neural networks models to construct a two-step method. In [13], they tested a decision tree and ANNs with a multilayer perceptron algorithm in order to analyze the capacity of credit unions members to settle their commitments. In most cases, the interest was centered in improving the performance of credit scoring models [14].

Traditional statistical techniques were used to determine the risk regarding the credit of customers including logistic regression, discriminant analysis, neural networks and decision trees. Some other techniques also popular were used, as shown in [15], where the use of new algorithms are proposed as in the case of support vector machines and least squares support vector machines.

Once the modeling technique is defined, it is possible to determine the advantages and disadvantages of its approach. Mester [16] proposed that building scoring models is faster, cheaper and more objective than *Expert* credit analysis. To assign a qualification dramatically reduces the time required to approve a credit line.

Another, more importante point, to be discussed about theses works, is that it is crucial to address the gap between academic and practical applications of credit scoring, because most researchers lack real-world data sets. In that respect, we consider the main contribution of the present the use of a real case from a Mexican bank regarding payroll issuers. We think that currently credit scorecard practices are still popular in banking institutions.

When both approaches were run using the software SAS Enterprise Miner 6.2, only a few options were modified in order to be practical for decisions makers, and in several options we used those provided by default by the software. [17] is an example of the use of this software regarding competitive and practical aspects for the benefit of professionals interested in these topics.

3 Case Study

In the Mexican banking system, payroll loans are one of the most profitable business for credit issuers. This particular loan is given to people who receive their salary payment on a payroll account. The customer signs a contract giving permission to the bank to retain the corresponding amount (monthly, weekly, semimonthly, etc.) that pays for his/her credit. For this phase, the customer's credit worthiness has been evaluated as well as his credit payment capacity (avoiding over-indebtedness). However, a potential risk can occur and may have a significant impact on the delinquency rates: churn of payroll issuers. If an issuer stops paying the payroll, every credit anchored to that issuer will default regardless of the customers credit quality.

The low level of financial education (good practices) in Mexico makes it difficult for the bank to collect the money lent, even when the responsibility of paying it back is entirely of the customer's. If delinquency goes up it will derive on several negative consequences, for example:

- increase of regulatory credit reserves,
- increase of collections costs,
- deterioration of relationship with customers.

The aim of this paper is to show to the bank under study, two approaches that address and solve this specific problem. In fact, there are many reasons for a payroll issuer (company) to end its commercial relationship with a bank, some of them are:

- bad service (regarding operational issues) of payroll payments,
- the company's bankruptcy,
- asset migration to another institution.

On the one hand, the company's relationship with the bank is managed by the wholesale banking. On the other, payroll loans are intended for individuals (retail banking). These two viewpoints create a difference in strategic purposes. In our previous work [18] we used the Credit Scoring Methodology to determine a stochastic model using issuer's transactional variables that allowed us to infer whether an issuer will churn in the following months, and we suggested at that point, a set of operative strategies in order to avoid churn risk and at the same time, to strengthen the bank's relationships with its customers and the payroll issuers. The data used in the present study was obtained from a real transactional data system base from a Mexican bank that it is available only for academic purposes and not for public use.

4 Previous Work

In our previous work, we proposed a credit scoring model for payroll issuers from a Mexican Bank. This work was a successful test of our concept, as we obtained results that were validated in both the practical and the theoretical aspects. Nevertheless, our model was only tested using one technique constructed by a code programming that followed a methodology proposed by Sidiqui [19]. However, in order to improve these results, specifically in the modeling phase, in this paper we propose two approaches: for the first approach, the main objective is to predict the likelihood of churn of the company. For the second approach, which is similar to the one shown in our recent investigation [18] the aim intended is to provide experimentation using SAS Enterprise Miner 6.2 in order to find the best model to establish a credit score.

4.1 Data description

The database considers 20,000 transactions collected during October 2014, and they are organized in 19 tables, each one corresponding to a monthly summary for every issuer, focusing on three variables:

- payroll amount (in US dollars),
- number of issued payroll payments,
- number of distinct payroll payment dates.

These data was collected over six months of observations as well as six months of actual performance. The target variable, in order to meet business requirements, was defined as follows:

- Bad issuer: one who within a month makes absolutely no payment for any of its customers, twice or more often in the performance period.
- Good issuer: one who makes a monthly payment for all of its customers, at least five times over a total of six, in the performance period.

In our past research, we provide the description of the process of the construction of the data set, considering 52 variables and the observation of eight different time periods, which overlapped each monthly with its respective anchor [18]. All variables were continuous, except the target variable, which was binary. Furthermore, no missing values were detected and variables did not fall within a normal distribution according to a KS Test. Variable Reduction was applied in order to improve the performance prediction using the technique called *Variable clustering* as suggested in [20]. Finally, we selected 14 variables which represent each cluster, that explains 90% of the variance. These variables that were the final data set are listed below:

Table 1. List of variables in data set.

Variable name	Variable description
V_AVG_F_4	Distinct monthly average payroll payment dates in the last 4 months
V_AVG_F_3	Number of decreases in monthly distinct payroll payment dates in the last 3 months
V_DEC_F_3	Number of decreases in monthly distinct payroll payment dates in the last 5 months
V_DEC_N_3	Number of decreases in monthly issued payroll payment in the last 3 months
V_DEC_N_5	Number of decreases in monthly issued payroll payment in the last 5 months
V_INC_F_3	Number of increases in monthly distinct payroll payment dates in the last 3 months
V_INC_F_5	Number of increases in monthly distinct payroll payment dates in the last 5 months
V_INC_M_3	Number of increases in monthly payroll payment amount in the last 3 months
V_INC_M_4	Number of increases in monthly payroll payment amount in the last 4 months
V_INC_M_6	Number of increases in monthly payroll payment amount in the last 6 months
V_INC_N_n	Number of increases in monthly issued payroll payment in the last 3 months
V_INC_N_n	Number of increases in monthly issued payroll payment in the last 5 months
V_MED_N_4	Median of monthly issued payroll payment in the last 4 months
V_RACHA_M_5	Maximum consecutive payroll payment in the last 5 months

For this study, we used SAS Enterprise Miner 6.2 under the assumption that it is a common software used for data modeling in banks. The sample data was partitioned 70:30 that is 70% for training and 30% for validation, for both approaches presented in this work.

5 Application of approaches

5.1 Approach I. Predictive Modeling

In this section, three classical predictive techniques were modeled: Decision Trees, Logistic Regression and ANNs, plus an ensemble model that was constructed using the two best models. The selection criteria for these techniques was from out point of view, indeed adequate for business perspectives.

We constructed 18 Decision trees changing the parameters: 1) Splitting Criteria, 2) Max Branch and 3) Max Depth. In Table 2 we can observe the resulting configuration.

Table 2. Experiments for Decision Trees Models.

Splitting Criteria	(Max Branch, Max Depth)
ProbChisq	(2,6); (3,6); (4,6); (5,6);(6,6); (12,12)
Entropy	(2,6); (3,6); (4,6); (5,6);(6,6); (12,12)
Gini	(2,6); (3,6); (4,6); (5,6);(6,6); (12,12)

When the Logistic Regression Model was used, we produced the logit transformation for response variable which is a known link function for many practitioners. Futhermore, the selection methods were: backward, forward and stepwise, which correspond to logistic regression models.

Only for analytical reasons and not for a practical interpretation, an ANNs technique was implemented. The models have the architecture of a Multilayer Perceptron with 3, 6, 9, 12 and 15 hidden units with default training technique that depends on the number of weights that are applied during the execution of the calculation. Also, the model selection criteria was the one that maximizes the profit or minimizes the loss for the cases that are in the validation data set. Thus, five ANNs models were tested.

For ensemble models we use two functions that combine the models. In this case there were *Tree* and *Tree16* which will be then explained, and correspond to Average, Maximum and Voting. Specifically with the *voting* function, it was necessary to chose two options for posterior probabilities: Average and Proportion. Thus, four ensemble models were tested. Finally, we built a total of 30 models of prediction.

5.1 Approach II. Credit Scoring Models

This approach aims to provide an outline of experimentation in order to find the best model to build a scorecard. The methodology appears in [19]. To this end, the Credit Scoring option from SAS Enterprise Miner is used. In this software, the use of nodes to achieve the desired statistical modeling is common; so once the sample is separated randomly into a training sample and a sample validation, the interactive grouping node is added. This node is used to eliminate weak features, or those that do not follow business logic, and to group those characteristics in order to produce a model in a scorecard format.

Since only continuous variables are studied, the grouping interactive node consists of a discretization process. This analysis will be done including the following steps: First, a binning method in a number of quantiles which divides the data into quantiles or buckets, where the values of the interval variables are placed into a specified number of evenly spaced bins. In our case, we use a value of 50 bins. Then, the pre-binning interval variable level vary along with all other variables, which then run through a grouping algorithm that is based on a decision tree model. After that, the scorecard is produced. Finally, the Weight of Evidence for each grouping is used in the logistic regression, as a prediction model. For this investigation the Backward, Forward and Stepwise Algorithms were used.

In this work, we only propose two approaches in order to find the best model for study case, so we do not show the scorecard points. In Table 3 we show the complete flow diagram for developing the credit scoring model; a total of 30 models are calculated. The diagram of modeling process is shown in **Figure 1**.

Table 3. Complete Flow Diagram for Credit Scoring.

Binning Method	Grouping Method	Splitting Worth Criteria	Selection Method
Quantile	Optimal	Entropy	Backward, Forward, Stepwise
	Quantile	Chi-Squared	Backward, Forward, Stepwise
	Monotonic Even Rate		Backward, Forward, Stepwise
Bucket	Optimal	Entropy	Backward, Forward, Stepwise
	Quantile	Chi-Squared	Backward, Forward, Stepwise
	Monotonic Even Rate		Backward, Forward, Stepwise
	Optimal		Backward, Forward, Stepwise

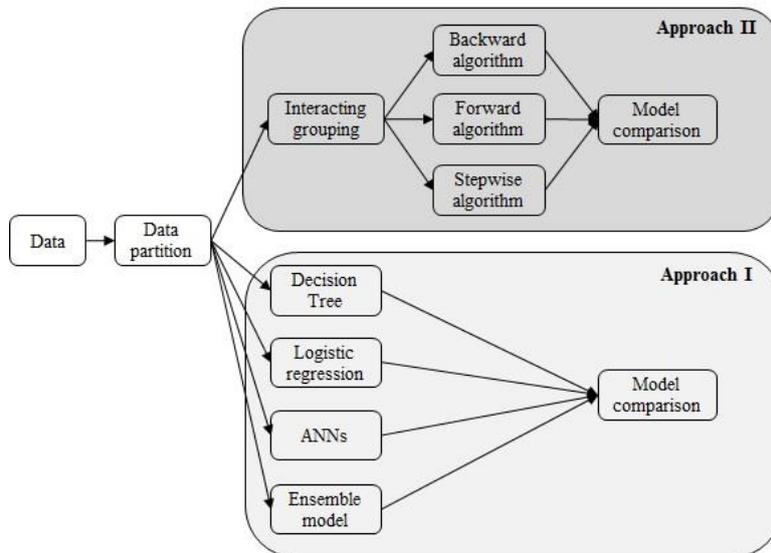


Fig. 1. Modeling Diagram for two approaches.

6 Results

6.1 Approach I

Table 4 we show that the best model obtained is a decision tree *Tree16* with a Gini Index as splitting rule, with a maximum of two leaves and a maximum depth of 6, based on misclassification rate equal to 8.472% for the training sample and 8.899% for the validation sample.

Similarly, the ROC Index, as shown in **Figure 2**, an area under the curve (AUC) of 0.698 and 0.683 for the training sample and validation, respectively. Indeed, the other Fit Statistics show that the Gini Index and the KS statistic are also competitive in terms of their predictive power. These results show an adequate performance of decision trees models, even in an ensemble model (*Ensembl7*) that was added to the two best decision tree models -*Tree* and *Tree16*- but these two do not improve the results of the best. In fact, the models *Tree* and *Tree7* also had the maximum number of leaves but were only constructed using Entropy and ProbChisq as splitting criteria, respectively.

Table 4. Results Fit Statistics for Approach I.

Model	Train				Validate			
	Missclassification Rate	ROC Index	Gini Index	KS Statistics	Missclassification rate	ROC Index	Gini Index	KS Statistics
<i>Tree16</i>	0.08472	0.698	0.397	0.375	0.08899	0.683	0.367	0.352
Ensembl7	0.08472	0.698	0.397	0.375	0.08899	0.683	0.367	0.352
Ensembl6	0.08472	0.698	0.397	0.375	0.08899	0.683	0.367	0.352
Ensembl8	0.08472	0.657	0.313	0.313	0.08899	0.641	0.283	0.283
Ensembl5	0.08536	0.698	0.397	0.375	0.08932	0.683	0.367	0.352
<i>Tree</i>	0.08543	0.698	0.396	0.375	0.08932	0.683	0.366	0.352
<i>Tree7</i>	0.08536	0.716	0.432	0.393	0.08949	0.703	0.407	0.374
Tree10	0.08393	0.763	0.526	0.421	0.08965	0.73	0.459	0.370
Tree11	0.08372	0.746	0.492	0.410	0.08965	0.706	0.412	0.356
Tree12	0.08365	0.743	0.487	0.407	0.08999	0.708	0.416	0.358
Tree17	0.08265	0.746	0.493	0.410	0.09032	0.735	0.471	0.375
Tree8	0.08508	0.671	0.341	0.326	0.09032	0.669	0.338	0.330
Tree18	0.08308	0.749	0.497	0.408	0.09065	0.733	0.466	0.383
Tree9	0.08429	0.735	0.470	0.399	0.09082	0.714	0.429	0.369
Tree13	0.08401	0.751	0.501	0.410	0.09082	0.734	0.468	0.378
Tree14	0.08386	0.725	0.451	0.385	0.09098	0.686	0.373	0.327
Tree15	0.08465	0.73	0.461	0.386	0.09098	0.702	0.404	0.341
Neural3	0.08815	0.751	0.503	0.385	0.09098	0.756	0.513	0.380
Neural	0.08793	0.745	0.490	0.381	0.09215	0.743	0.485	0.370
Neural5	0.08893	0.753	0.506	0.389	0.09248	0.753	0.505	0.378
Tree2	0.08579	0.719	0.437	0.377	0.09265	0.706	0.413	0.352
Neural4	0.08836	0.743	0.486	0.378	0.09282	0.75	0.500	0.382
Tree3	0.08622	0.669	0.337	0.325	0.09281	0.671	0.341	0.332
Tree4	0.08622	0.669	0.337	0.325	0.09282	0.671	0.341	0.332
Tree5	0.08622	0.669	0.337	0.325	0.09282	0.671	0.341	0.332
Tree6	0.08622	0.669	0.337	0.325	0.09282	0.671	0.341	0.332
Neural2	0.08701	0.768	0.536	0.405	0.09382	0.755	0.510	0.393
Reg2	0.09422	0.731	0.463	0.370	0.09448	0.743	0.487	0.366
Reg3	0.09422	0.731	0.463	0.370	0.09448	0.743	0.487	0.366

Reg 0.09444 0.731 0.462 0.369 0.09465 0.742 0.484 0.362

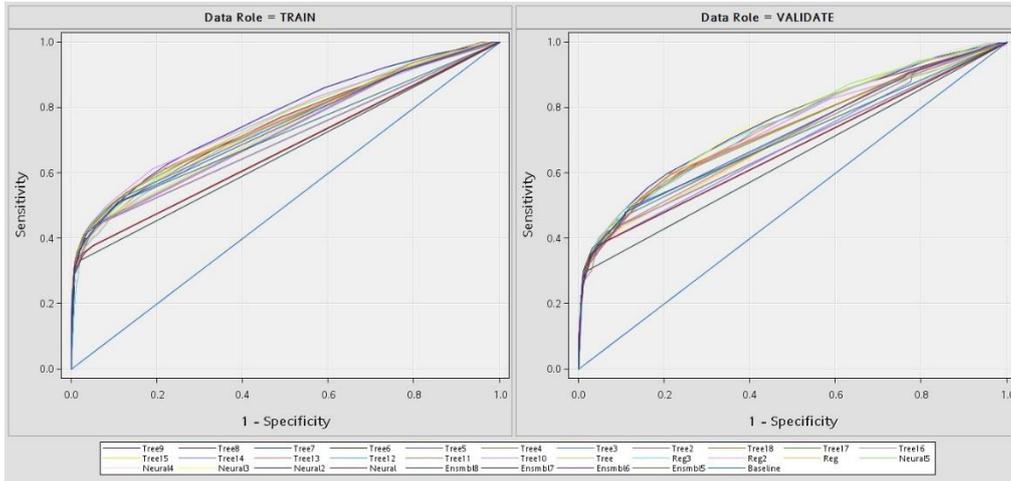


Fig. 2. ROC Chart for training and validation sample for Approach I.

6.2 Approach II

In this approach, the idea was to test the performance of logistic regression models implemented in software as a core for building a scorecard that helps the interpretation of the decision maker.

The best model, based on the misclassification rate, was the model *Scorecard7*, see Table 5 ahead. This model is characterized by the use of the Quantile Binning Method and the Quantile as Grouping Method, and using the algorithm Backward. This model had 9.365% of misclassified rate for training data, and 9.498% for validation data.

Table 5. Results Fit Statistics for Approach II.

Model	Train				Validate			
	Missclassification Rate	ROC Index	Gini Index	KS Statistics	Missclassification rate	ROC Index	Gini Index	KS Statistics
<i>Scorecard7</i>	0.09365	0.751	0.502	0.393	0.09498	0.759	0.519	0.396
Scorecard28	0.09036	0.730	0.460	0.367	0.09498	0.732	0.464	0.353
Scorecard13	0.09244	0.758	0.515	0.413	0.09515	0.762	0.523	0.403
Scorecard15	0.09244	0.758	0.515	0.413	0.09515	0.762	0.523	0.403
Scorecard	0.09165	0.758	0.517	0.411	0.09515	0.762	0.524	0.413
Scorecard3	0.09165	0.758	0.517	0.411	0.09515	0.762	0.524	0.413
Scorecard14	0.09258	0.758	0.516	0.413	0.09515	0.760	0.520	0.395
Scorecard8	0.09351	0.751	0.501	0.385	0.09515	0.759	0.518	0.390
Scorecard29	0.09044	0.729	0.459	0.367	0.09532	0.733	0.466	0.354
Scorecard30	0.09044	0.729	0.459	0.367	0.09532	0.733	0.466	0.354
Scorecard10	0.09294	0.755	0.510	0.411	0.09548	0.761	0.522	0.405
Scorecard11	0.09294	0.755	0.510	0.411	0.09548	0.761	0.522	0.405
Scorecard12	0.09294	0.755	0.510	0.411	0.09548	0.761	0.522	0.405
Scorecard4	0.09236	0.758	0.516	0.418	0.09548	0.762	0.523	0.404

Scorecard6	0.09236	0.758	0.516	0.418	0.09548	0.762	0.523	0.404
Scorecard17	0.09136	0.730	0.460	0.378	0.09548	0.731	0.463	0.368
Scorecard18	0.09136	0.730	0.460	0.378	0.09548	0.731	0.463	0.368
Scorecard19	0.09136	0.730	0.460	0.378	0.09548	0.731	0.463	0.368
Scorecard5	0.09265	0.758	0.517	0.416	0.09582	0.760	0.520	0.401
Scorecard2	0.09194	0.759	0.517	0.409	0.09598	0.761	0.521	0.407
Scorecard9	0.09422	0.752	0.503	0.389	0.09598	0.760	0.520	0.390
Scorecard16	0.09136	0.730	0.461	0.375	0.09615	0.732	0.463	0.363
Scorecard23	0.09136	0.730	0.461	0.375	0.09615	0.732	0.463	0.363
Scorecard24	0.09136	0.730	0.461	0.375	0.09615	0.732	0.463	0.363
Scorecard25	0.09136	0.730	0.461	0.373	0.09615	0.731	0.462	0.362
Scorecard26	0.09136	0.730	0.461	0.373	0.09615	0.731	0.462	0.362
Scorecard27	0.09136	0.730	0.461	0.373	0.09615	0.731	0.462	0.362
Scorecard20	0.10665	0.691	0.383	0.292	0.10898	0.682	0.363	0.262
Scorecard21	0.10665	0.691	0.383	0.292	0.10898	0.682	0.363	0.262
Scorecard22	0.10665	0.691	0.383	0.292	0.10898	0.682	0.363	0.262

In **Figure 3**, the ROC index is observed with an area under the curve of 0.759 for data validation. These results were slightly better than those obtained in our research, which were 10% of missclassification rate and 0.750 for ROC Index. **Figure 4** shows that there is a discrimination or an adequate separation of Cumulative Distributions, between good and bad issuers.

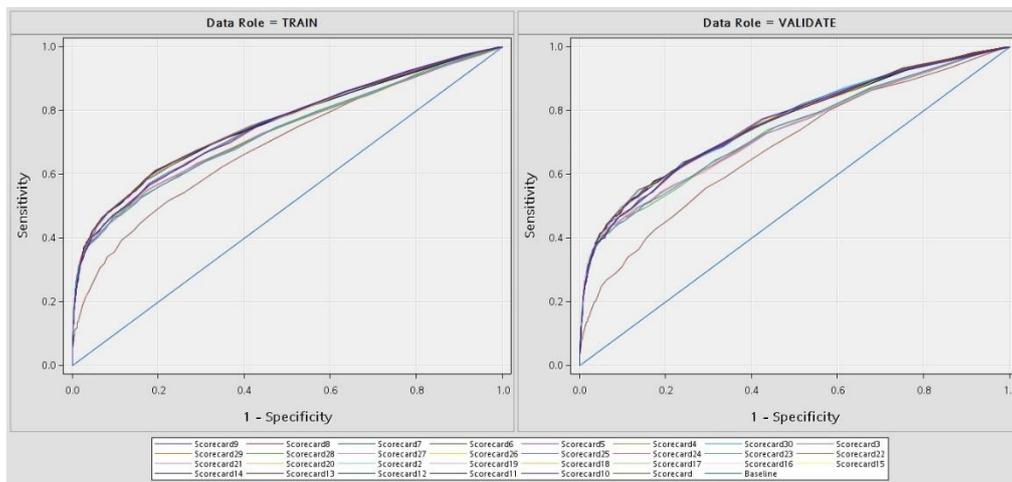


Fig. 3. ROC Chart for training and validation sample for Approach II.

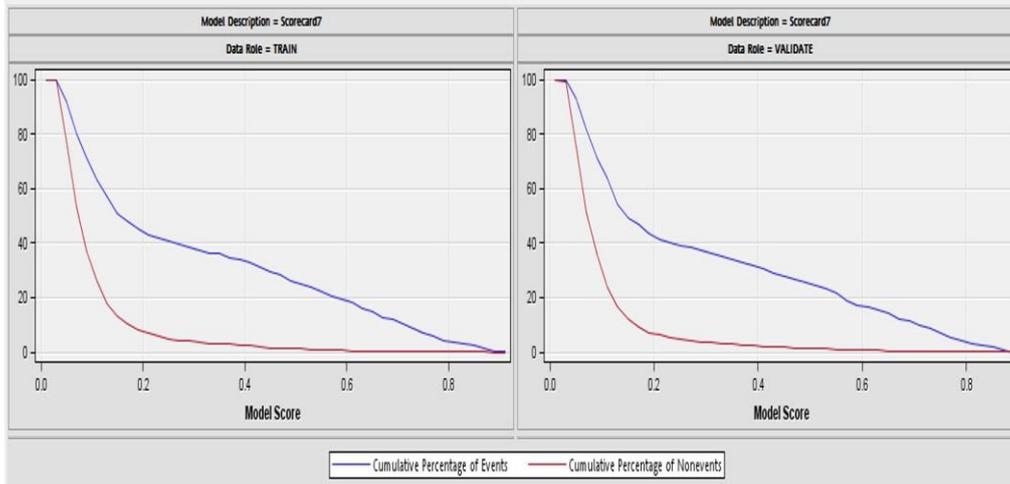


Fig. 4. Cumulative Distribution Functions: training and validation sample for Approach II.

7 Conclusions

This paper presents two approaches for predictive models of payroll issuers in a Mexican bank. In recent years, the statistical and AI techniques for credit scoring models was still useful, and in our case, the study results were consistent and performance successfully. Additionally, the software used is an adequate tool to run this modeling exercise, which means to provide a practical decision-making framework for the industry. In general, the results show that it is possible to construct several credit scores for transactional models for payroll issuers. Also, a scorecard system is still useful as well as an easy tool for business applications. In both approaches the results are promising, and can improve depending on the business objectives. In that sense, we show different models for prediction purposes and the credit score models were a tradeoff between description and prediction approaches. Therefore, in approach I it is necessary to know how to generate value from prediction models without the explanation of the phenomenon. In the approach II, the value of KS Statistics (± 0.4) was adequate because it showed separation between the distribution of good and bad issuers, see **Figure 4**. Future work should replicate these approaches using another time frame of observation in order to validate the timeline further. We think it is possible to expect that future implementation of the best models obtained in our study will show an adequate stability, and discriminatory power. Also, we would like to mention that the definition of target variable is still controversial but the model will be, once again, capable to provide solutions for these business decision-making problems.

References

- [1] Wang, G., Hao, J., Ma, J., Jiang, H.: A comparative assessment of ensemble learning for credit scoring. *Expert Systems with Applications* 38, pp. 223-230 (2011)
- [2] Thapliyal, M.P.: Data mining: a tool for banking industry. *International Journal of Emerging Research in Management and Technology* 4(4), pp. 84-88 (2015)
- [3] Keramati, A., Yousefi, N.: A proposed classification of data mining techniques in credit scoring. *Proceedings of the 2011 International Conference on Industrial Engineering and Operations Management Kuala Lumpur, Malaysia, January 22-24, (2011)*
- [4] Ince, H., Aktan, B.: A comparison of data mining techniques for credit scoring in banking: a managerial perspective. *Journal of Business Economics and Management* 10(3), pp. 233-240 (2009)
- [5] Brown, J.I.L.: *Developing Credit Risk Models using SAS Enterprise Miner and SAS/STAT: Theory and Applications*. SAS Institute Inc. (2014)
- [6] Khandani, A.E., Kim, A.J. Lo, A.W.: Consumer credit-risk models via machine learning algorithms. *Journal of Banking and Finance* 34(11), pp. 2767-2787 (2010)
- [7] Samreen, A., Batul-Zaidi, F.: Design and development of credit scoring model for the commercial banks of Pakistan: Forecasting Creditworthiness of Individual Borrowers. *International Journal of Business and Social Science* 3(17), pp. 155-166 (2012)
- [8] SAS Institute Inc.: *Building Credit Scorecards using Credit Scoring for SAS Enterprise Miner*. A SAS Best Practices Paper, pp. 1-23 (2009)
- [9] Thomas, L.C., Edelman, D.B., Crook, J.N.: *Credit Scoring and Its Applications*. Society for Industrial and Applied Mathematics, Philadelphia, PA (2002)
- [10] Sadatrasoul, S.M., Gholamian, M.R., Siami, M., Hajimohammadi, Z.: Credit scoring in banks and financial institutions via data techniques: a literature review. *Journal of AI and Data Mining* 1, pp. 119-129 (2013)
- [11] Abdou, A.H., Pointon, J.: Credit scoring, statistical techniques and evaluation criteria: a review of the literature. *Intelligent Systems in Accounting Finance and Management* 18, pp. 59-88 (2011)
- [12] Chye, K.H., Chi, T.W., Peng, G. C.: A two-step method to construct credit scoring models with data mining techniques. *International Journal of Business and Information* 1(1), pp. 96-118 (2006)
- [13] De Moraes Sousa, M., Santana Figueiredo, R.: Credit analysis using data mining: application in the case of a credit union. *JISTEM - Journal of Information Systems and Technology Management* 11(2), pp. 379-396 (2014)
- [14] Zhou, L., Lai, K.K., Yu, L.: Least squares support vector machines ensemble models for credit scoring. *Expert Systems with Applications* 37, pp. 127-133 (2010)
- [15] Baesens, A.: Benchmarking State-of-the-art classification algorithms for credit scoring. *Journal of the Operations Research* 54, pp. 627-635 (2003)
- [16] Mester, L.J.: *What is the Point of Credit Scoring?*. Business Review, Federal Reserve Bank (1997)
- [17] Yap, B.W., Ong, S.H., Husain, N.H.M.: Using data mining to improve assessment of credit worthiness via credit scoring models. *Expert Systems with Applications* 38, pp. 13274-13283 (2011)
- [18] Fuentes-Cabrera, J. , Pérez-Vicente, H.: Credit scoring model for payroll issuers: a real case. *Advances in Artificial Intelligence and Its Applications: 14th Mexican International Conference on Artificial Intelligence, MICAI 2015, Cuernavaca, Morelos, Mexico, October 25-31, 2015, Proceedings, Part II*. (2015)
- [19] Siddiqi, N.: *Credit Risk Scorecards: Developing and Implementing Intelligent Credit Scoring*. John Wiley and Sons, USA (2005)
- [20] SAS Institute Inc.: *SAS/STAT 9.3 User's Guide*. SAS Institute Inc, Cary NC (2011)