International Journal of Business, Human and Social Sciences
ISSN: 2517-9411
Vol:14, No:7, 2020

Machine Learning Techniques in Bank Credit Analysis

Fernanda M. Assef, Maria Teresinha A. Steiner

Abstract**—**The aim of this paper is to compare and discuss better classifier algorithm options for credit risk assessment by applying different Machine Learning techniques. Using records from a Brazilian financial institution, this study uses a database of 5,432 companies that are clients of the bank, where 2,600 clients are classified as non-defaulters, 1,551 are classified as defaulters and 1,281 are temporarily defaulters, meaning that the clients are overdue on their payments for up 180 days. For each case, a total of 15 attributes was considered for a one-against-all assessment using four different techniques: Artificial Neural Networks Multilayer Perceptron (ANN-MLP), Artificial Neural Networks Radial Basis Functions (ANN-RBF), Logistic Regression (LR) and finally Support Vector Machines (SVM). For each method, different parameters were analyzed in order to obtain different results when the best of each technique was compared. Initially the data were coded in thermometer code (numerical attributes) or dummy coding (for nominal attributes). The methods were then evaluated for each parameter and the best result of each technique was compared in terms of accuracy, false positives, false negatives, true positives and true negatives. This comparison showed that the best method, in terms of accuracy, was ANN-RBF (79.20% for non-defaulter classification, 97.74% for defaulters and 75.37% for the temporarily defaulter classification). However, the best accuracy does not always represent the best technique. For instance, on the classification of temporarily defaulters, this technique, in terms of false positives, was surpassed by SVM, which had the lowest rate (0.07%) of false positive classifications. All these intrinsic details are discussed considering the results found, and an overview of what was presented is shown in the conclusion of this study.

Keywords**—**Artificial Neural Networks, ANNs, classifier algorithms, credit risk assessment, logistic regression, machine learning, support vector machines.

I. Introduction

THE term "credit", in the context of banking, is a type of transaction or contract through which a financial institution makes a certain amount of money available to someone (individual or company) with the promise of returning the money later in one or more installments. This is a complex decision, as this operation involves risk and can lead to serious consequences [1].

Credit Risk Assessment is a classic decision-making problem, in which the merit of a given individual is granted credit (or not) based on estimates of his potential of paying back the loan along with the interest rates determined by the bank [2]. With the fast development of financial products and services, a bank's credit departments always need to collect large amounts of data, which risk analysts use to build appropriate credit scoring models to evaluate an applicant's credit risk accurately [3].

The aim of this work is to propose efficient and effective methodologies to determine whether to grant credit to companies that are clients of a Brazilian financial institution. To achieve this goal, different evaluation technologies are explored in relation to specify parameters that will be discussed in Section III. The methods presented are different classifiers of Machine Learning (ML), namely: Artificial Neural Networks (ANN) both Multilayer Perceptron (MLP; ANN-MLP) and Radial Basis Functions (RBF; ANN-RBF); LR; and SVM.

After obtaining the results, this study seeks to discuss a comparison between the methods used regarding three different points: the accuracy of the tested methods, the rate of false positives found in the predictions and the rate of classification errors. This discussion is intended to suggest the best methodology for decision making, considering different aspects of each model.

The contribution of this research lies in different aspects of the field of study in decision-making about the problem of credit risk. The first aspect focuses on the analysis of a class of client, not only defaulters and non-defaulters as normally considered, but a classification for clients who are temporarily defaulters. This class represents clients who delay the payment of their debts with the financial institution. The study is also intended to contribute to the discussion on the influence of other analyses of the results of ML techniques at the time of decision making and what their implications may be.

In addition to this introduction, in Section II, studies related to the theme that sought applications and/or used similar techniques to those employed in this study are presented. The methodology of the work is addressed in Section III, from the data collection and its coding to the parameters tested in each of the four methods used (ANN-MLP; ANN-RBF; LR and SVM). The results and a discussion of them can be found in Section IV. Finally, the conclusion and suggestions for future studies are presented in Section V.

II. Related Work

Credit risk assessment is an important and challenging data-mining problem in the domain of financial analysis. Since its introduction in 1950, it has been widely used and, more recently, used for granting loans, especially for credit cards and loan agreements [4]. This problem basically consists of assessing the risk associated with a loan from a financial organization to a certain company or individual.

According to [3], there is a wide range of methodologies for

Fernanda Assef is with the Pontifícia Universidade Católica do Paraná, Brazil (e-mail: fermassef@gmail.com).

solving credit risk classification problems. The credit assessment methods can be classified into six distinct categories, namely: Theoretical Frameworks [5], [6]; Statistical Methods [7], [8]; Decision Theory [9]-[11]; ANNs [12]-[14]; SVMs [15]-[17]; and, finally, Data Envelopment Analysis [18], [19].

The field of ML techniques has seen a significant growth in credit risk assessment research in recent years. Loterman et al. [20] compared 24 techniques for classifying bank loans, including heuristics, ANN, SVM and statistical models. These techniques were evaluated in six real databases, with two techniques (ANN and B-OLS - Beta Ordinary Least Squares) having the best performance in one of the databases in terms of Mean Absolute Error (MAE).

ANN were also found among the selected papers either to focus on enhancing credit scoring models in different aspects by testing different parameters [21], or combined with other techniques such as SVM, kNN (K-Nearest Neighbors) and Decision Trees (DT) [22].

Other models used to classify clients were proposed based on Bayesian latent variables. The proposed methods were compared with techniques such as Linear Discriminant Analysis (LDA), ANN, SVM, Classification and Regression Trees (CART) and Multivariate Adaptive Regression Splines (MARS) [23], [24]. In a similar way to the present study, a methodology was proposed that divided clients into "good payers", "cannot pay" and "will not pay". In [25], a clustering algorithm, LR and Multinomial LR were used, along with ANN in order to conduct this bank loan analysis.

It should be highlighted that, analogous to [25], a third classification was introduced. Thus, seeking the best option to aid decision making, different parameters of ML algorithms were tested, as in [21]. With the chosen parameters, four different methods were tested and compared, as in [20], [23] and [24], the methods being ANN-MLP, ANN-RBF, LR and SVM.

III. Methodology

In this section, the techniques used to solve the Credit Risk Problem of the institution are explained and exemplified. Fig. 1 shows in detail the steps used in the development of this work.

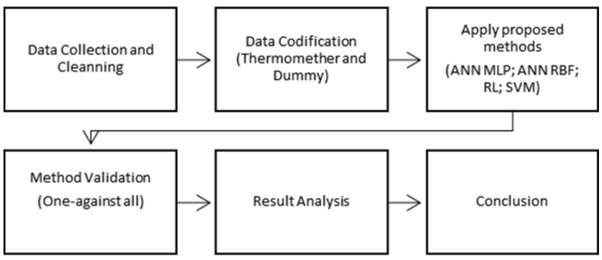

Fig. 1 Research process

The data used in this work were obtained from a large Brazilian financial institution regarding the granting of credit to companies located in Paraná State, Brazil. Historical records from January 1996 to June 2017 were used, with over 39,000 credit operations involving 5,432 companies. The data on the operations during the period in question were classified regarding their payment situations as follows:

- Class I – Non-defaulters: clients whose repayments are made within the deadline;
- Class II – Temporarily Defaulters: clients who are up to 180 days in arrears;
- Class III – Defaulters: clients whose repayments are over 180 days in arrears.

In this way, 2,600 clients were classified as non-defaulters, 1,281 as temporarily defaulters and 1,551 as defaulters. With the help of employees of this institution, 15 attributes of each client were defined that were considered in the application of the methods: Legal Status of the Company; Annual Gross Revenue; Number of Employees; Company Activity; Year of Incorporation; Date of Company's Opening; Market Segment; Loan Risk; Scale of Establishment of Credit Limit; Established Limit; Application; Overdue Payments; Total Debt; Indebtedness to the National Financial System; and Restrictions on Loans. Initially, these attributes were transformed into binary values as they were found in different forms (ordinal or nominal) and scales. Thermometer codification was applied to ordinal attributes (example in Table I), and dummy codification [26] was applied to nominal attributes (Table II).

TABLE I
Thermometer Codification for Ordinal Attributes

Annual Gross Revenue (AGR) x 1000	Inputs		
AGR ≤ 1000	0	0	0
1000 < AGR ≤ 2000	0	0	1
2000 ≤ AGR ≤ 3000	0	1	1
AGR ≥ 3000	1	1	1

TABLE II
Dummy Codification for Nominal Attributes

Market segment	Inputs	
Micro enterprise	0	0
Small Business	0	1
Average or Large Company	1	0

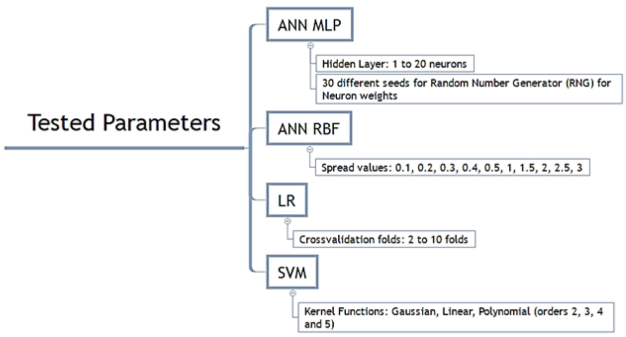

Fig. 2 Tested parameters for each ML method

After the coding of the models (ANN-MLP; ANN-RBF; LR and SVM), the parameters for each of the methods evaluated were chosen for consideration in their tests. The parameters that were chosen are shown in Fig. 2. The methodology chosen for the evaluation of each method was that of "one-against-all", where one of the categories "faces" the others. The results are then analyzed in terms of percentages of being a non-default client in relation to the other classes analyzed (defaulters and temporarily defaulters), for instance. All the tests conducted for each of the methods were performed using a Samsung computer, with Windows 10, Intel CORE i7 8th Gen., using MATLAB R2018b software.

IV. Results

Applying the methodology outlined, we begin this section by presenting the individual results of each method (ANN-MLP; ANN-RBF; LR and SVM), dividing them into sections that represent what was obtained in the classification of non-defaulters, defaulters and those who are temporarily defaulters. Initially, the results of the methods will be presented individually, and the methods will then be compared for each type of classification. It should be highlighted that each method was assessed as follows: for ANN-MLP, 70% of the instances were used for training and 30% for the test; for ANN-RBF, 70% for training and 30% for the test; for LR, cross validation with different numbers of folds and, finally, for SVM cross-validation with 10 folds was used.

A. Non-Defaulter Classification

The tests were executed following the previous mentioned parameters and, in Table III, the best results are presented for each tested method. In observing Table III, where TP stands for True Positive, FP is False Positive, FN is False Negative, and TN is True Negative, we can believe that, if the decision-maker wishes to enjoy greater reliability when assessing whether the client will be a non-defaulter, he can use ANN-RBF with a spread of 0.3. Under the conditions presented above, his decision will be 79.2% accurate and he will have a lower rate of false positives (where the method believes that the client would be a non-defaulter, but he would not pay or would pay in arrears).

TABLE III
COMPARISON BETWEEN NON-DEFAULTERS' BEST RESULTS

	Parameters	Accuracy	TP	FP	FN	TN
RBF	Spread 0.3	79.20%	32.11%	3.18%	17.64%	47.07%
MLP	20 neurons; Seed 16	72.80%	33.43%	12.76%	14.43%	39.38%
LR	Cross-validation with 4 folds	70.50%	36.25%	17.86%	11.62%	34.28%
	Cross-validation with 7 folds	70.50%	36.25%	17.84%	11.62%	34.30%
	Cross-validation with 8 folds	70.50%	34.33%	11.51%	17.80%	36.36%
SVM	Linear Kernel Function	70.40%	40.93%	20.75%	8.82%	29.50%

B. Defaulter Classification

Following an evaluation of Table IV, we perceived that even with the high accuracy rates of the methods, the use of only this version of one-against-all for the classification of defaulters would not be reliable enough to aid decision making. This is due to a greater chance of identification of the other categories (true negatives) than true positives, i.e., actual defaulters.

TABLE IV
COMPARISON BETWEEN DEFAULTERS' BEST RESULTS

	Parameters	Accuracy	TP	FP	FN	TN
RBF	Spread 0.1	97.74%	20.89%	1.13%	1.13%	76.85%
MLP	16 neurons; Seed 29	90.80%	18.40%	4.00%	5.20%	72.40%
SVM	Linear Kernel Function	88.57%	13.06%	2.47%	8.96%	75.51%
LR	Cross-validation with 10 folds	88.00%	15.10%	3.55%	8.49%	72.86%

C. Temporarily Defaulter Classification

In Table III we perceived that although ANN-RBF with a spread of 0.3 had better accuracy in general (75.37%), the rates of true positives were low. An analysis of the rest of the information shows that this also occurred in the assessment of defaulters. It appears that the number of defaulters and temporarily defaulters is not high enough for the rate of true positives to be greater. Thus, the algorithm seems to be more reliable when classifying "other categories" for the "one-against all", which can be verified by the rate of true negatives in both Tables IV and V.

TABLE V
COMPARISON BETWEEN TEMPORARILY DEFAULTERS' BEST RESULTS

	Parameters	Accuracy	TP	FP	FN	TN
RBF	Spread 0.3	75.37%	7.69%	4.09%	20.54%	67.68%
MLP	11 neurons; Seed 27	73.00%	9.90%	8.40%	18.60%	63.10%
	13 neurons; Seed 26	73.00%	8.60%	7.00%	20.00%	64.40%
SVM	Linear Kernel Function	71.84%	0.14%	0.07%	28.09%	71.70%
LR	Cross-validation with 2 folds	71.20%	2.61%	2.91%	25.94%	68.54%
	Cross-validation with 8 folds	71.20%	2.10%	2.30%	26.45%	69.15%
	Cross-validation with 9 folds	71.20%	2.06%	2.30%	26.49%	69.15%

V. Conclusions

The objective of the present research was to define three models for the classification of "defaulting" or "non-defaulting" or "temporarily defaulting" customers of legal entities that use bank credit services. For this purpose, the historical records provided by a financial institution were used.

MLP ANNs were applied, using the *Levenberg-Marquardt* learning algorithm and with a topology with only one hidden layer in which the number of neurons varied from 1 to 20, using the one-against-all model. The same test was performed for RBF-type ANNs, with the largest number of neurons tested being represented by the number of instances to be trained by the network. The third method was the RL using different values for cross validation folds. The last method was SVM, tested for different kernel functions.

Considering the results obtained, some important points deserve to be highlighted. After the application of the proposed methodologies and the selection of the best results within each ML method, we perceived that for the three classifications tested (non-default, default and temporarily default), the performance of the ANN-RBF algorithm was superior in terms of accuracy. Only the size of its spread varied in accordance with the class that was under analysis at the time.

With the results achieved, it is suggested that the decision-maker should test the three classifications with the best methods presented in Tables III-V to be equipped with as much information as needed when it comes to granting credit. As the main intention of a financial institution when using these classifier algorithms (ML) is to identify as assertively as possible those clients who are prone to being "defaulters", it is preferable to use a technique with greater precision. Thus, based on what has been shown, using only ANN-RBF, the analysis would mean a more accurate classification. Furthermore, if a decision is made to minimize false positives (the main aspect of this work), the use of ANN-RBF is suggested.

Suggestions for future works include: tests with hybrid techniques with a view to improving performance; equilibrium of classes, since the non-defaulter class had more instances in relation to the others; use of evaluation metrics other than the ones presented here, such as AUC (Area Under the Curve) and F-Score; and the inclusion of insights of the decision maker in the form of ranking clients based on his experience of working with this kind of information on a daily basis.

Acknowledgment

The first author would like to thank CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brazil) for her masters and doctorate financial incentives. The second author would like to acknowledge CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico – Brazil) for promoting her research through financial incentive.

References

[1] Fenerich, A. T., Steiner, M. T. A., Steiner Neto, P. J., Tochetto, E. A., Tsutsumi, D. P., Assef, F. M., dos Santos, B. S. 2018. Use of Machine Learning techniques in bank risk analysis. Proceedings of International Conference on Computers and Industrial Engineering (CIE48), Auckland, New Zealand, 2-5 December.

[2] Fahner, G. 2012. Estimating causal effects of credit decisions, International Journal of Forecasting, 28(1), pp. 248–260.

[3] Zhang Z., Gao G., Shi Y. 2014. Credit risk evaluation using multi-criteria optimization classifier with kernel, fuzzification and penalty factors. European Journal of Operational Research. 237(1), pp. 335-348.

[4] Luo, S., Kong, X., Nie, T. 2016. Spline Based Survival Model for Credit Risk Modelling, European Journal of Operational Research, 253(3), pp. 869 – 879.

[5] Bolton, C. 2009. Logistic Regression and its Application in Credit Scoring, Universidade de Pretoria, 238p.

[6] Redondo, J. M., Ortin, F. 2017. A Saas Framework for Credit Risk Analysis Services. IEEE América Latina Transactions, 15(3) pp. 474-481.

[7] Acevedo, Y. V. N., Quintero, J. F. L., Marín, C. E. M., González, C. 2016. Business Rules Model for the Automation in the Receipt of Credit Applications by Financial Institutions Based on ArchiMate. IEEE América Latina Transactions, 14(6), pp. 2801-2806.

[8] Diakoulaki, D., Mavrotas, G., Papayannakis, L. 1992. A multicriteria approach for evaluating the performance of industrial firms, Omega, vol. 20, no. 4, pp. 467 – 474, 1992.

[9] Messier, W. F., Hansen, J. V. 1985. Inducing Rules for expert system development an example using default and bankruptcy data, Management Science, 9, pp. 253 – 266.

[10] Bryant, K. 2001. ALEES: an agricultural loan evaluation expert system, Expert Systems with Applications, 21, pp. 75 – 85.

[11] Zhang, D., Zhou, X., Leung, S. C., Zheng, J. 2010. Vertical bagging decision trees model for credit scoring, Expert Systems with Applications, 37(12), pp. 7838 – 7843.

[12] Desai, V., Crook, J., Overstreet Jr, G. 1996. A comparison of neural networks and linear scoring models in credit union environment, European Journal of Operational Research, 95, pp. 24–37.

[13] West, D. 2000. Neural Network Credit Scoring Models, Computers & Operations Research,27 (11), pp. 1131–1152.

[14] Pavlenko, T., Chernyak, O. 2010. Credit risk modeling using Bayesian networks, International Journal of Intelligent Systems, 25(4) pp. 326 – 344.

[15] Gestel, T. V., Baesens, B., Garcia, J. Dijcke, P. V. 2003. A support vector machine approach to credit scoring, Journal of Banking & Finance, vol. 2, pp. 73 – 82.

[16] Belloti T., Crook, J. 2009. Support Vector Machines for credit scoring and discovery of significant features, Expert Systems with Applications, 36(2), pp. 3302 – 3308.

[17] Zhang, Z., Gao, G., Shi, Y. 2014. Credit risk evaluation using multi-criteria optimization classifier with kernel, fuzzyfication and penalty factors, European Journal of Operational Research, 237(1), pp. 335 – 348.

[18] Shanmugam, R., Johnson, C. 2007. At a crossroad of data envelopment and principal component analysis, Omega, vol. 35(4), pp. 351 – 364.

[19] Iazzolino, G., Bruni, M. E., Beraldi, P. 2013. Using DEA and financial ratings for credit evaluation: an empirical analysis, Applied Economics Letters, 20, (14), pp. 1310 – 1317.

[20] Loterman, G., Brown, I., Martens, D., Mues, C., Baesens, B. 2012. Benchmarking regression algorithms for loss given default modeling, International Journal of Forecasting, 28, pp. 161 – 170.

[21] Zhao, Z., Xu, S., Kang, B. H., Kabir, M. M. J., Liu, Y., Wasinger, R. 2015. Investigation and improvement of multi-layer per-ceptron neural networks for credit scoring. Expert Systems with Applications. 42 (7), pp. 3508-3516.

[22] García, V., Marqués, A. I., Sánchez, J. S. 2014. An insight into the experimental design for credit risk and corporate bankruptcy prediction systems. Journal of Intelligent Information Systems. 44 (1), pp. 159-189.

[23] Kao, L. -J., Chiu, C. -C., Chiu, F. -Y. 2012. A Bayesian latent variable model with classification and regression tree approach for behavior and credit scoring, Knowledge Based Systems,36, pp. 245 – 252.

[24] Abdou, H. A., Tsafack, M. D. D., Ntim, C. G., Baker, R. D. 2016. Predicting creditworthiness in retail banking with limited scoring data, Knowledge Based Systems, 103, pp. 89 - 103.

[25] Bravo, C., Thomas, L. C., Weber, R. 2015. Improving credit scoring by differentiating defaulter behavior, Journal of Operations Research Society, 66, pp.771 – 781.

[26] Steiner, M. T. A., Nievola, J. C., Soma, N. Y., Shimizu, T., & Steiner Neto, P. J. 2007. Extração de regras de classificação a partir de redes neurais para auxílio à tomada de decisão na concessão de crédito bancário. Pesquisa Operacional, 27(3), pp. 407-426.