A Classifier Evaluation for Payments' Default Predictions in a Brazilian Retail Company

96

Strauss Carvalho Cunha, Emanuel Mineda Carneiro, Lineu Fernando Stege Mialaret, Luiz Alberto Vieira Dias, and Adilson Marques da Cunha

Abstract

This article presents an investigation about the performance of classification algorithms used for predicting payments' default. Classifiers used for modelling the data set include: Logistic Regression; Naive-Bayes; Decision Trees; Support Vector Machine; *k*-Nearest Neighbors; Random Forests; and Artificial Neural Networks. These classifiers were applied to both balanced and original data using the Weka data mining tool. Results from experiments revealed that Logistics Regression and Naive Bayes classifiers had the best performance for the chosen data set.

Keywords

Data mining · Classifier algorithms · Area under curve · Logistic regression

96.1 Introduction

In recent years, there have been increases in credit granting by financial institutions and retail companies. This phenomenon is caused by favorable factors like better market conditions, greater consumption inherent to the growth of economies and increases in population incomes [1]. Regarding predictions of defaulting payments, the use of classification methods have been necessary and mandatory, implemented by the so-called credit scoring systems [2]. Classification methods have been widely implemented, by using algorithms, statistical techniques, and Machine Learning (ML), among other emerging areas of Artificial Intelligence (AI) [3].

In this research, the case study is a Brazilian retailer company with hundreds of stores spread throughout the country providing customers with credit cards. Companies aim to develop systems that allows to identify credit defaulting customers, among other needed functionalities. Computer systems must make use of data mining algorithms for classifying credit card defaulting customers.

The main goal of this paper is to assess classification algorithms for credit analysis in a Brazilian retail company. It is organized as follows: Sect. 96.2 deals with several classification algorithms; Sect. 96.3 describes the data mining experimentation algorithms; and finally, Sect. 96.4 presents some conclusions, recommendations and suggestions for future work.

96.2 Classification Algorithms

Choosing the most suitable classifier for defaulting customers is based on financial data and considered an important decision. Performance's comparison among several classifiers is one way to take this decision.

Louzada et al. [4] and Aniceto [5] have identified in their surveys the following most used classifiers: (1) Artificial Neural Networks (ANN); (2) Support Vector Machine (SVM); (3) Linear Regression; (4) Decision Trees (DT); (5) Fuzzy Logic; (6) Genetic Algorithms (GA); (7) Discriminant Analysis; (8) Bayesian Networks (Naive-Bayes—NB); and (9) Ensemble Methods. Others authors have identified a series of other classifiers used in the credit scoring: (1) Expert Systems; (2) *k*-Nearest Neighbors (KNN); (3) Clustering; (4)

S. C. Cunha
Brazilian Federal Service of Data Processing - SERPRO, RJ, Brazil
e-mail: strauss.carvalho@serpro.gov.br

E. M. Carneiro · L. A. V. Dias · A. M. da Cunha
Computer Science Department, Brazilian Aeronautics Institute of Technology - ITA, Sao Jose dos Campos, SP, Brazil
e-mail: mineda@ita.br; vdias@ita.br; cunha@ita.br

L. F. S. Mialaret (✉)
Federal Institute of Education, Science and Technology of Sao Paulo - IFSP, Jacarei, SP, Brazil
e-mail: lmialaret@ifsp.edu.br

© Springer International Publishing AG, part of Springer Nature 2018
S. Latifi (ed.), *Information Technology – New Generations*, Advances in Intelligent Systems and Computing 738,
https://doi.org/10.1007/978-3-319-77028-4_96

Case-Based Reasoning; (5) Random Forests (RF); among other types of different classifiers that have been investigated and used in the credit scoring applications [6–10]. In this research, the choice of classifiers was based on their popularity for credit scoring.

96.3 Classification Algorithms Experimentation

For this investigation, a sample of the data set was used, duly validated by the company. The initial sample of the data set used had 6158 records, of which 4461 records were related to the non-defaulting customers and 1696 records were related to the defaulting customers containing eight attributes, including the attribute of the target class.

A series of experiments were carried out, to verify which classifiers best fit the context of granting credit. For experiments, the Weka Experiment Environment was selected from the Weka tool, allowing to perform comparisons between classifiers, duly evaluated by statistical testing.

In this assessment, six sample of data sets were used with the following characteristics:

1. Sample A1, without applying the gain ratio metric, by using the natural imbalance;
2. Sample A2, without applying the gain ratio metric, by using oversampling (the SMOTE technique);
3. Sample A3, without applying the gain ratio metric, by using undersampling;
4. Sample A4, with the application of the gain ratio metric, by selecting the most significant attributes (above the 0.01 threshold);
5. Sample A5, with the application of the gain ratio metric, by using oversampling (the SMOTE technique); and
6. Sample A6, with application of the metric gain ratio, by using undersampling.

In order to evaluate the data sets described above, the following seven classifiers were used: (1) LR; (2) N; (3) DT, also named Algorithm J48 in Weka; (4) SVM; (5) kNN, also named IBK algorithm in Weka; (6) RF; and (7) ANN. Table 96.1 shows assessment results.

The LR classifier was selected as the base classifier for selection because it is traditionally used in credit granting scenarios [11]. For the sample A1, which did not use any attribute selection metrics and was imbalanced, the LR classifier obtained the best performance with the AUC-ROC = 0.89. For the A2 sample, which was available without using any attribute selection technique, but was balanced using the SMOTE technique, the best performing classifier was the RF classifier with the AUC-ROC = 0.91. For the sample A3, where no variable selection metric was applied and it was balanced using the undersampling technique, the classifier that presented the best result was again the LR classifier, along with the NB classifier, with a value of the AUC-ROC = 0.89.

Table 96.1 Classifier evaluation results using AUC-ROC metric

	AUC-ROC values for classifiers						
Sample	LR	NB	DT	SVM	kNN	RF	ANN
A1	0.89	0.89	0.85	0.76	0.85	0.87	0.87
A2	0.9	0.90	0.89	0.82	0.9	0.91	0.9
A3	0.89	0.89	0.87	0.78	0.88	0.88	0.88
A4	0.89	0.89	0.85	0.76	0.88	0.88	0.88
A5	0.89	0.89	0.88	0.78	0.89	0.89	0.89
A6	0.89	0.89	0.87	0.78	0.88	0.88	0.88

In samples A4, A5, and A6, the gain ratio metric for attribute selection was applied, with attributes discarded with a value lower than 0.01. In the case of the imbalanced A4 sample, the LR classifier, together with the NB classifier, with a value of the AUC-ROC = 0.89, presented the best result in terms of classification. For the A5 sample, which was balanced using the SMOTE technique, the best classifier was also the LR classifier, along with the NB, kNN, and RF classifiers, with a value of the AUC-ROC = 0.89. Finally, for the A6 sample, balanced with the undersampling technique, the best result was the LR classifier, along with the NB classifier, presenting the AUC-ROC = 0.89.

It can be observed that some classifiers obtain an improvement of performance when applying attribute's selection techniques and imbalance's reduction. One example is the *k*NN classifier, which improves its performance by up to 6%. In other cases, such as the LR classifier, there is no significant improvement in performance, which confirms the resilience of this algorithm to data imbalance [9].

96.4 Conclusion

The main goal of this investigation was to assess classification algorithms for credit analysis in a Brazilian retail company. The following classifiers were used for modelling the data set including: Logistic Regression (LR); Naive-Bayes (NB); Decision Trees (DT); Support Vector Machine (SVM); *k*-Nearest Neighbors (k-NN); Random Forests (RF); and Artificial Neural Networks (ANN).

Experiments have shown that Logistic Regression (LR) and Naive-Bayes (NB) classifiers have performed better in comparison with other classifiers for a specific data set. It was also possible to verify that class imbalance and attribute selection have affected classification performance for certain classifiers.

As future work, one interesting issue would be to assess the classifiers using others attribute selection metrics. Other line of future research would be the evaluation of classifier algorithms with others performance measures.

Acknowledgements The authors would like to thank: (1) the Brazilian Aeronautics Institute of Technology (ITA); (2) the Casimiro Montenegro Filho Foundation (FCMF); the Software Engineering Research Group (GPES) members; and the 2RP Net Enterprise for their infrastructure, data set, assistance, advice, and financial support for this work.

References

1. V. García, A.I. Marqués, J.S. Sánchez, An insight into the experimental design for credit risk and corporate bankruptcy prediction systems. J. Intell. Inf. Syst. **44**(1), 159–189 (2015)
2. J. Abellán, J.G. Castellano, A comparative study on base classifiers in ensemble methods for credit scoring. Expert. Syst. Appl. **73**, 1–10 (2017)
3. K. Kennedy, Credit scoring using machine learning, Doctoral thesis, Dublin Institute of Technology, 2013
4. F. Louzada, A. Ara, G. Fernandes, Classification methods applied to credit scoring: a systematic review and overall comparison. Surv. Oper. Res. Manag. Sci. **21**(2), 117–134 (2016)
5. M.C. Aniceto, Estudo comparativo entre técnicas de aprendizado de máquina para estimação de risco de crédito, Dissertação (Mestrado em Administração), Universidade de Brasília, Brasília, Brazil, 2016
6. A. Hooman, G. Marthandan, W.F.W. Yusoff, M. Omid, S. Karamizadeh, Statistical and data mining methods in credit scoring. J. Dev. Areas **50**(5), 371–381 (2016)
7. A. Riasi, D. Wang. Comparing the performance of different data mining techniques in evaluating loan applications. Int. Bus. Res. **9**(7), 164–187 (2016).
8. S. Lessmann, B. Baesens, H.-V. Seow, L.C. Thomas, Benchmarking state-of-the-art classification algorithms for credit scoring: an update of research. Eur. J. Oper. Res. **247**(1), 124–136 (2015)
9. I. Brown, C. Mues, An experimental comparison of classification algorithms for imbalanced credit scoring data sets. Expert. Syst. Appl. **39**, 3446–3453 (2012)
10. W. Lin, Y. Hu, C. Tsai, Machine learning in financial crisis prediction: a survey. IEEE Trans. Syst. Man Cybern. C Appl. Rev. **42**(4), 421–436 (2012)
11. H.A. Abdou, J. Pointon, Credit scoring, statistical techniques and evaluation criteria: a review of the literature. Intell. Syst. Accounts Finance Manag. **18**, 59–88 (2011)