Les Cahiers du GERAD ISSN: 0711–2440

Performance of the mathematical programming approach in credit scoring

Y. Mektoubi, B. Benyacoub,
M. Ouzineb, I. El Hallaoui

G–2019–06

January 2019

La publication de ces rapports de recherche est rendue possible grâce au soutien de HEC Montréal, Polytechnique Montréal, Université McGill, Université du Québec à Montréal, ainsi que du Fonds de recherche du Québec – Nature et technologies.

Dépôt légal – Bibliothèque et Archives nationales du Québec, 2019
– Bibliothèque et Archives Canada, 2019

The publication of these research reports is made possible thanks to the support of HEC Montréal, Polytechnique Montréal, McGill University, Université du Québec à Montréal, as well as the Fonds de recherche du Québec – Nature et technologies.

Legal deposit – Bibliothèque et Archives nationales du Québec, 2019
– Library and Archives Canada, 2019

GERAD HEC Montréal
3000, chemin de la Côte-Sainte-Catherine
Montréal (Québec) Canada H3T 2A7

Tél. : 514 340-6053
Téléc. : 514 340-5665
info@gerad.ca
www.gerad.ca

Performance of the mathematical programming approach in credit scoring

Yassine Mektoubi [b]
Badreddine Benyacoub [c]
Mohamed Ouzineb [a,b]
Issmail El Hallaoui [a,d]

[a] *GERAD, Montréal (Québec), Canada, H3T 2A7*

[b] *SI2M, National Institute of Statistics and Applied Economics, Rabat, Morocco, B.P. 6217*

[c] *Laboratory of Mathematics, Computing and Applications, Faculty of Sciences, University Mohammed-V, Rabat, Morocco, B.P, 8007*

[d] *Department of Mathematics and Industrial Engineering, Polytechnique Montréal (Québec) Canada, H3C 3A7*

y.mektoubi@gmail.com@gmail.com
benyacoubb@gmail.com
mohamed.ouzineb@cirrelt.net
issmail.el-hallaoui@polymtl.ca

January 2019
Les Cahiers du GERAD
G–2019–06

Abstract: Operations Research (OR) has a very important role to play in credit scoring for building models that can help the lending organization to make a good decision about accepting or rejecting a new applicant (borrower). Many standard statistical and machine learning techniques are used in the literature, for example, Linear Discriminant Analysis, Linear Regression, Hidden Markov Model, Support Vector Machine and Artificial Neural Network. In this paper, we propose a new approach for credit scoring. The idea is to combine statistics and OR modeling to solve this problem. An originality of our approach is that it proceeds by three steps. First, we use the cross validation method for separating the testing part from the training part. In the second step, a mathematical programing approach is used in discriminant analysis. In the last step, we use resampling techniques (like Jackknife and Bootstrap procedures) for estimating the parameters for the discriminant mathematical programming model. The performance of the proposed approach is validated on two Australian and German public credit datasets.

Keywords: Classification, discriminant analysis, statistics methods, MP methods, Linear programming

1 Introduction

After the recent world financial crisis, more attention was given from banks and financial institutions to credit risk, since it can cause great cost losses to owners, managers, workers, lenders, clients, community and government. Therefore, it is very important to predict bankruptcy and decide whether to grant credit to new applicants or not. One of the primary tools used by banks is credit scoring or scorecards, it's used to estimate how likely an applicant is in default. Originally, it was conducted using subjective judgments of human expert, but it's impossible nowadays because of the large number of applicants, the vast amount of information, the great commercial competition, the slowness of the process and the frequent errors. Credit scoring is a model developed to determine if loan customers belong to a good applicant group or a bad applicant group. Therefore, it's a classification problem where a new applicant must be categorized into one of the predefined classes based on a number of attributes that describe the economic and the socio-demographic situation of the applicant. In the other hand, the Basel II Accord requires from capital institutions to estimate the probability of default, the exposure at default and the loss given default, which encourages banks to enhance their statistical and analytical tools. Therefore, different approaches have been used to improve credit risk management.

The first approach used is based on statistical classification methods. Hand et al. (1997) presented a review of statistical classification approaches used in credit scoring and assessed each method, such as discriminate analysis, linear regression, logistic regression and decision trees. Lessmann et al. (2015) realizeed a benchmark that includes 41 classifiers using 8 datasets, measured their performances with 6 indicators that include novel classifiesr such as Percentage Correctly Classified, AUC, Partial GINI Index, the H-measure, the Brier Score and Kolmogrov-Smirnov Statistic and examined the correspondence between empirical results obtained using different accuracy indicators.

Tsai et al. (2014) developed a novel hybrid financial distress model based on combining the clustering technique and classifier ensembles. They used two clustering (Self Organized Maps and K-Means) techniques and three classification techniques (logistic regression, multilayer-perception, neural network). Joaquín et al. (2014) improved the work of Nami et al. (2009) in which they created many ensemble classifiers based on complex ones. But the authors show that using simple classifiers outperforms Nami and Lumini results. Sanchez et al. (2012) introduced composite ensembles that use different strategies for diversity induction so that the diversity inducted by one method can be improved by the diversity produced by the other method. They found that the classifier ensemble performed better than any single classifier. The authors used two resampling-based ensemble (bagging and adaboost) and two attribute-based algorithms (random subspace and rotation forest). Alaraj et al. (2016a) developed a credit scoring model based on a combination of heterogeneous ensemble of classifiers and combined their rankings using a new combination rule called the consensus approach, through which the classifiers work as a team to reach an agreement on all the data points' final outputs. Alaraj et al. (2016b) focused on improving predictive performance of hybrid ensemble credit scoring model through the combination of two data pre-processing methods in feature selection and data filtering. Hongshan Xiao et al. (2016) proposed an ensemble classification approach based on supervised clustering for credit scoring. Supervised clustering is employed to partition the data samples of each class into a number of clusters. Most researchers utilize random sampling to generate training subsets for constructing the base classifiers. Therefore, their diversity is not guaranteed, which may lead to a degradation of the overall classification performance. Clusters from different classes are then pairwise combined to form a number of training subsets. In each training subset, a specific base classifier is constructed.

Benyacoub and El Bernoussi (2014) modeled the probability to be a good or a bad borrower using HMM and compared it with other techniques such as LDA and LR. A Hidden Markov Model corresponds to the modeling of two stochastic processes, a hidden process perfectly modeled by a Markov chain and an observed process dependent on the hidden states. The authors tested the classifier on two datasets of banks, and used random selection for resampling data. The authors compared the results with nine other classifiers using different performance indicators such as AUC, error type I and II, accuracy and cost function. Sebastien Aupetit in his thesis focused on three directions namely, the improvement of the learning of the pattern, the creation of a new model and the understanding by the visualizations of dissimilarity. For the first direction, he used three meta-heuristic bio-mimetic-based populations that are applied to spaces of solutions of different natures. The author has adopted HMM with symbol substitution since they have more powerful expression power without increasing the complexity of the algorithms. The author used several meta-heuristics to learn the HMM model, such as Random heuristics, simulated annealing method, taboo search method, genetic algorithm, ant colony, and particle swarm optimization.

The mathematical programming approach was first applied in classification by Magasarian (1965). Freed and Glover (1981a) presented an evaluation of the LP approach in discriminant analysis. They provided a variety of test conditions involving both normal and non-normal populations. Bajgier and Hill (1982) showed the potential of LP approach in classification where the normality of the population is upheld. Erenguc et al (1990) presented a survey of different linear programming models and there experimental results. These formulations include the objective of minimizing the maximum deviation, minimizing the sum of exterior deviations and the objective combining minimizing a measure of exterior deviations and maximizing a measure of interior deviation. Other variants of these formulations include weighted objectives or constraints to overcome unbound solutions. Glover (1990) presented and improved version of the LP model by showing how to increase the scope and flexibility of these models and he also used successive goal method by establishing a series of conditional objectives to obtain refined solutions. Ziari et al (1997) used re-sampling techniques to identify statistical parameter estimates for the mathematical programming model used in classification problems. They used Jackknife, deleted-d Bootstrap and Bootstrap procedures to prioritize and rank the variables and to identify statistically significant variables. Freed and Glover (1986) presented different potential difficulties that may arise with linear programming approach used in discriminant analysis. Theses difficulties may limit the effectiveness of the model. The authors provide some solutions to overcome these difficulties to improve the stability and the accuracy of the model and provide a better normalization.

The remainder of this paper is organized as follows. Section II summarizes a description of Credit Scoring Problem. In Section III, we provide a description of resampling techniques. Section IV shows the application on Australia and German credit datasets. Finally, concluding remarks are made in Section V.

2 Credit scoring problem

2.1 Problem description and assumptions

The majority of lending organizations must assess the risk level of granting a credit to a new applicant. To do so, they use credit scoring models that are developed from training sets consisting of people in their records who were given loans in the past.

Given a sample of n previous borrowers or applicants, n_G of them are good and n_B are bad. Each applicant is characterized by p variables $X = (X_1, X_2, ..., X_p)$ (age, sex, salary, ...). Denote by A the set of all possible combinations of values of the variables X. The first work to do by the lending organization is to divide A into two sets : A_G represents the answers given by the good clients and A_B represents the bad ones. The response of each applicant i denoted by $(x_{i1}, x_{i2}, ..., x_{ip})$. The second work is to choose, for a given cutoff value c, weights or scores $(w_1, w_2, ..., w_p)$ so that $\sum_{j=1}^{p} w_j x_{ij} \geq c$ if the client i in the sample is a good one and $\sum_{j=1}^{p} w_j x_{ij} \leq c$ if the client i in the sample is bad.

2.2 Mathematical formulation of the problem

Generally, it is difficult to get a perfect separation of the good from the bad ones. One can allow possible errors by introducing non negative variables a_i. Therefore, $\sum_{j=1}^{p} w_j x_{ij} \geq c - a_i$ if applicant i in the sample is a good one, and, $\sum_{j=1}^{p} w_j x_{ij} \leq c + a_i$ if applicant i is a bad one. The objective is to find weights $(w_1, w_2, ..., w_p)$ that minimize the sum of the absolute values of these deviations (MSD) and the problem can be formulated as the following linear program:

$$\text{Min} \sum_{i=1}^{n} a_i \tag{1}$$

subject to :

$$\sum_{j=1}^{p} w_j x_{ij} \geq c - a_i \qquad \forall i \in G_1 \tag{2}$$

$$\sum_{j=1}^{p} w_j x_{ij} \leq c + a_i \qquad \forall i \in G_2 \tag{3}$$

$$\sum_{j=1}^{p} \left(n_B \sum_{\substack{i=1 \\ i \in G_1}}^{n_G} x_{ij} - n_G \sum_{\substack{i=1 \\ i \in G_2}}^{n_B} x_{ij} \right) w_j = 1 \tag{4}$$

$$a_i \geq 0 \quad \forall i, \text{ and } w_j \geq 0 \quad \forall j. \tag{5}$$

Where:

c :	The hyperplane cutoff between the 1st and the 2nd group
a_i :	The absolute value deviation
w_i :	The weights such as $\sum_{j=1}^{p} w_j x_{ij}$ is the equation of the hyperplane separating the two groups
x_{ij} :	The value of feature j of client i
n :	The total number of population in the two groups
n_G :	The number of population of the 1st group (good clients)
n_B :	The number of population of the 2nd group (bad clients)
G_1 :	The set of good clients
G_2 :	The set of bad clients.

The constraint (4) is the normalization constraint introduced by Glover in 1990 to deal with the problem relating trivial solutions and the problem of fixing the cutoff.

For a problem of classification with two distinguished group, the main objective of the classifier is to determine a weighting vector $w = (w_1, w_2, ..., w_p)$ and a scalar c so that it assigns, as correctly as possible, the observations (i=1,2,. . . ,n) from the group G_1 to other group G_2 based on the linear discriminant function which can be expressed as $Z_i = w_1X_1 + w_2X_2 + \ldots + w_pX_p$ using the separating value c.

In practical credit scoring, it consists to estimate the parameters w and a decision rule cutoff value c minimizing the number of misclassifications for the dataset. Generally, the parameter vector w and the value c are combined in such a discriminant function to determines the classification model or the classifier.

In practice again, the dataset is divided into two subsets: one for training used to build the model and the other one for testing used to evaluate the performance of the model. In this order, different performance measures are used to quantify the performance of the model. Firstly, average accuracy is used to compare models. Secondly, the performance of these models is measured in terms of Type I error or Type II error and AUC (defined in the following subsection).

2.3 Accuracy types of error and AUC

The most direct criterion that can measure quantitatively the performance of a classifier is the accuracy. It can be calculated, as following :

$$Accuracy = \frac{The\ number\ correctly\ classified\ cases}{The\ total\ number\ of\ cases}$$

For credit scoring problem, two types of error rates should be considered [27]. The first one is type I error that corresponds to the rate of the number of 'bad' customers being categorized as 'good' among the 'good' customers which is noted by FP (False Positive), here the 'bad' borrower is accepted as creditworthy. The following equations show the process to calculate the Type I error rate :

$$Type\ I\ error = \frac{FP}{FP + TN}$$

where TN (True Negative) is the number of 'bad' customers being categorized as as 'bad' (correctly classified).

The second one is Type II error corresponding to the rate of 'good' customers being categorized as 'bad' among the 'bad' customers which noted by FN (False Negative), in this case a 'good' borrower is rejected.

$$Type\ II\ error = \frac{FN}{FN + TP}$$

where TP (True Positive) is the number of 'good' customers being categorized as 'good' (correctly classified).

Two different techniques of resampling were used to re-estimate these performance measures (or parameters). The two methods *Jackknife* and *Bootstrap*, presented in the following section, are commonly used to estimate parameters for performance metrics.

3 Resampling techniques

The Bootstrap and Jackknife resampling procedures have been used to estimate the statistical error. The two methods are nonparametric techniques that do not request statistical assumptions for the estimation. Different samples selected from the original data are used to examine and evaluate the variability of the estimator. Both methods use a subset from the original dataset and each subset generates different pseudo values.

3.1 Jackknife procedure

The Jackknife was first proposed by Quenouille (1949) as a procedure for correcting bias. Later, it was refined and given its current name by Turkey (1956), he used the procedure to estimate standard error and construct confidence limits for a large class of estimators. The use of Jackknife is practical when it is difficult or impossible to calculate estimators' standard error. Keith Knight (2000) noted that the Jackknife estimate of the standard error is equivalent to the *delta method* for large samples.

The Jackknife estimator of a parameter is found by systematically leaving out each observation from a dataset and calculating the estimate and then finding the average of these calculations. Given a sample of size n, the Jackknife estimate is found by aggregating the estimates of each $n-1$ sized sub-sample. Let the sample be divided into t $(i = 1, 2, \ldots, t)$ subsets which are equal size each d, $n = td$. The main idea behind the Jackknife procedure lies in systematically recomputing the estimate parameter leaving out one observation, known as the *deleted-one* Jackknife procedure $(t = n, d = 1)$, or more observations at a time from the complete sample, known as the *deleted-d* Jackknife procedure ($d > 1$).

Let $W^{'}$ be an estimator of the parameter W obtained using LP by considering the whole sample of observations where size of the sample is equal to n and $W_i^{'}$ is its estimator corresponding to the case of leaving out the i^{th} subset.

Let $W_i^{'} = (w_1^i, w_2^i, ..., w_p^i)$ be the linear discriminant function coefficient at step i, i.e., after deleting of the i^{th} d observation set from the full sample. The Jackknife delete-d ($d = 1$ or $d > 1$) procedure used to provide estimates of LP coefficient is as follows:

1. Given an n sized sample from the population.
2. Divide the dataset into t independent subsets of size d.
3. Alternately omit the k^{th} d observations set and estimate the linear discriminant function coefficient $W_i^{'}$, $i = 1, \cdot, t$. Thus, we obtain t deleted-d Jackknife parameters $W_1^{'}, W_2^{'}, \ldots, W_t^{'}$.
4. Calculate the Jackknife parameter corresponding to the linear discriminant function coefficient estimate $W_j^{'*} = (w_1^*, w_2^*, \cdot, w_p^*)$ where each coefficient is given by Efron (1982) as follows :

$$w_s^* = \sum_i^t \frac{w_s^i}{t}, s = 1, 2, \ldots, p \tag{6}$$

and the standard error $\sigma = (\sigma_1, \sigma_2, \ldots, \sigma_p)$ where the standard error coefficient σ_s is given as follows :

$$\sigma_s = [(1 - \frac{1}{t})\sum_{i}^{t}(w_s^i - w_s^*)^2]^{\frac{1}{2}}, s = 1, 2, \ldots, p \tag{7}$$

3.2 Bootstrap procedure

The Bootstrap method was firstly developed and published by Bradley Elfon in "Bootstrap methods: another look at the Jackknife" (1979). A Bayesian extension was introduced by Rubin D. in 1981.

The method specifies that B samples will be generated randomly from the original dataset with replacement, with each sample set being of identical size as the original set. Let $W = (w_1, w_2, \ldots, w_p)$ be the parameter of the discriminant function and $W_i^{'} = (w_1^i, w_2^i, \ldots, w_p^i)$ be its estimator for the ith random sample. The Bootstrap procedure corresponding to LP is as follows:

1. for $i = 1$ to B do:
   step 1 : Select randomly n observations from the original sample with replacement.
   step 2 : Train the LP parameter $W_i^{'} = (w_1^i, w_2^i, \ldots, w_p^i)$ using the n selected observations.
2. Calculate the mean of the estimated parameters $W_1^{'}, W_2^{'}, \ldots, W_B^{'}$ yielding the Bootstrap parameter $W_b^{'*} = (w_1^*, w_2^*, \ldots, w_p^*)$ such as

$$w_s^* = \frac{1}{B}\sum_{i=1}^{B} w_s^i, \quad s = 1, 2, \ldots, p.$$

3. Calculate the standard error $\sigma = (\sigma_1, \sigma_2, \ldots, \sigma_p)$ of the parameter $W^{'}$ given by the following :

$$\sigma_s = [(\frac{1}{B-1})\sum_{i}^{B}(w_s^i - w_s^*)^2]^{\frac{1}{2}}, \quad s = 1, 2, \ldots, p.$$

4 Application and numerical results

In this paper, the classification model is applied in credit scoring. For this purpose, we will use two databases, the Australian and the German database. The first is The Australian Credit Approval, it contains 690 instances with 14 attributes, 6 of them are numerical and 8 are categorical. The second Database is The German Credit Data, it contains 1000 clients with 20 attributes, 7 of them are numerical and 13 are categorical. The two databases have two classes, the good client or credit that is not delinquent and the bad client or credit that is delinquent. The Australian database contains 44.5% of good clients and 55.5% of bad clients; the German Database contains 70% good clients and 30% bad clients.

The database is divided in 2 sets, the learning database which is used to construct the model and to estimate the parameters and the testing database which is used to evaluate the classification power and validate of the model. In order to get the best performance of a classification technique, the train data set and test data set should be taken separately from the available population. In addition, the cases in the two sample sets should be independent, to make sure that there is no relationship among them.

The most popular method for performance assessment of a classification technique is the k-fold cross-validation method. This method has been used to maximizes the use of the data and approximately the same class proportions as the original data set in all obtained subsets. The procedure of k-fold cross-validation is as follow :

1. The data set is split into k mutually folds of nearly equal size.
2. Choose the first subset for testing set and the $k-1$ remainder for training set.
3. Build the model on the training set.
4. Evaluate the model on the testing set by calculating the accuracy error type I rate and error type II rate.

5. Alternately choose the following subset for testing set and the $k-1$ remainder for training set.
6. The structure of the model is then trained k times each time using $k-1$ subsets (training set) for training and the performance of the model is evaluated $k-1$ on the remaining subset (testing set).
7. The predictive power of classifier is obtained by averaging the k validation fold estimates found during the k runs of the cross validation process.

The common values for k are sometimes 5 and 10. The Cross validation method is used in this work to assess the performance of classification techniques and we choose 10 as value for k for our experiments evaluation method.

Figure 1 shows the proposed process of application LP in credit scoring problem. The whole building process of model can be presented into three phases. First, the cross validation procedure is applied to split the dataset into two different subsets. Secondly, the LP model is trained by the training set. The unknown LP parameters are estimated on using a resampling procedure (Jackknife or Bootstrap). Then the linear discriminant function is employed to classify the observations given in testing set and finally evaluate the performance of the model.

The linear model is coded in R and solved using the Rglpk library. The evaluate the models, we are going to use many the accuracy, Error type 1 and error type 2.

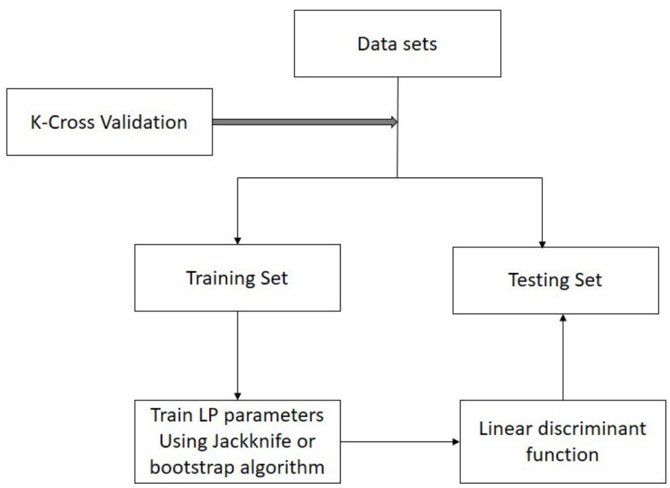

Figure 1: The flow chart for creating a credit scoring using LP model and Jackknife/Bootstrap procedures.

4.1 Results

At first, we solved the linear program in (1–5) with c constant equal to 0.5 and with various cross validation and Bootstrap parameters. The result of the linear program applied to the Australian database using Bootstrap resampling technique is shown in Table 1 and the one using Jackknife resampling technique is shown in Table 2. Table 1 shows that the best result are obtained with a cross validation of size 10 and a Bootstrap of size 100. This model offers an accuracy of $91,30\%$.

We applied the same linear model with various values of constant c and various resampling techniques settings to the German database. We obtained the results in Table 3 using Bootstrap resampling technique, and the one using Jackknife resampling technique is shown in Table 4.

Table 3 shows that the best results are obtained with a cross validation of size 10 and a Bootstrap of size 20. This model reaches an accuracy of 76%. It appears also that the model performance drops dramatically with a positive constant c.

Table 1: Classification results with LP and Bootstrap resampling technique for Australian Database datasets and $c = 0$.

Cross Validation size	Resampling parameter	Accuracy	Type I error	Type II error
00.00	20.00	00.78	0.09	0.06
10.00	20.00	00.48	0.03	0.13
10.00	50.00	00.88	0.04	0.13
10.00	100.00	00.91	0.01	0.07
05.00	100.00	00.65	0.02	0.32
10.00	150.00	00.83	0.04	0.13
10.00	200.00	00.81	0.09	0.10

Table 2: Classification results with LP and Jackknife Resampling technique for the Australian Database datasets.

Value of c	Cross Validation size	Accuracy	TypeI error	Type II error
0.50	10.00	00.83	0.10	0.07
-0.50	10.00	00.58	0.00	0.42
1.00	10.00	00.59	0.06	0.35
-1.00	10.00	00.49	0.00	0.50
-0.50	05.00	00.53	0.00	0.46
0.50	05.00	00.57	0.00	0.42

Table 3: Classification results with LP and Bootstrap Resampling technique for German Database datasets.

Value of c	Cross Validation size	Resampling parameter	Accuracy	TypeI error	Type II error
0.50	00.00	20.00	0.409	0.10	0.167
0.50	10.00	20.00	00.30	0.02	0.68
-0.50	10.00	20.00	00.76	0.24	0.00
-0.50	10.00	50.00	00.75	0.25	0.00
0.50	10.00	50.00	00.29	0.01	0.70
-0.20	10.00	20.00	00.70	0.30	0.00
-0.80	05.00	20.00	00.65	0.35	0.00
-0.50	10.00	100.00	00.66	0.34	0.00

Table 4: Classification results with LP and Jackknife resampling technique for German Database datasets.

Value of c	Cross Validation size	Accuracy	TypeI error	Type II error
-0.50	10.00	00.72	0.28	0.00
0.50	10.00	00.47	0.28	0.00
1.00	10.00	00.50	0.14	0.36
-1.00	10.00	00.67	0.33	0.00
-0.50	05.00	00.68	0.32	0.00
0.50	05.00	00.61	0.14	0.25

Given the results above, linear programming model gives a very good score of classification and outperforms in both Australian and German databases. The model is the best one when applied to the Australian dataset. It also appears that the Bootstrap technique gives better results than the Jackknife technique.

5 Conclusion

Credit scoring is an important subject that held the attention of bank and financial institution. Many methods were used to improve the accuracy of the models. We presented a survey of different approaches used in credit scoring. We also introduced two different resampling techniques namely Jackknife and Bootstrap and applied them to a linear programming model to improve its accuracy.

From the results found above, we can say that, first, the linear programming approach gives the best classification approach. Second, the Bootstrap resampling technique performs better that the Jackknife approach both in accuracy score. Third, the linear programming model is better in term of computational requirements and we can easily modify the objective function to meet the preferences of the decision maker. In the other hand, the linear model performance drops when applied to other database, so a certain steps of cleaning, preprocessing and formatting data is required. Finally, the database used in this study is small compared to today real bank databases. So it is important to test the performance of these models with large databases and different kinds of data (structured and unstructured data).

References

[1] Badreddine, B., Souad, B., Abdelhak, Z., and E. Ismail. Classification with Hidden Markov Model, Applied Mathematical Sciences, 2014. 8(50), 2483–2496.

[2] Abdou, H.A. and Pointon, J. (2011). Credit scoring, statistical techniques and evaluation criteria: a review of the literature. Intelligent Systems in Accounting Finance and Management, 18, 59–88. doi:10.1002/isaf.325

[3] Joaquán, A. and C.J. Mantas. Improving experimental studies about ensembles of classifiers for bankruptcy prediction and credit scoring. Expert Systems with Applications, 2014. 41, 3825–3830.

[4] Nanni, L. and A. Lumini. An Expermental Comparison of Ensemble of Classifiers of Bankruptcy Prediction and Credit Scoring, Expert Systems with Applications, 2009. 36(2), 3028–3033.

[5] Alaraj, M., Abbod, M. and Z. Hunaiti (2014). Evaluating consumer loans using neural networks ensembles, International Conference on Machine Learning, Electrical and Mechanical Engineering (ICMLEME'2014), Dubai, pp. 55–61.

[6] Alaraj, M., Abbod, M. and B. Al-Hnaity (2015). Evaluation of consumer credit in Jordanian banks: a credit-scoring approach, Proceedings of the 2015 17th UKSIM-AMSS International Conference on Modelling and Simulation, IEEE Computer Society, Cambridge, pp. 127–130.

[7] Hongshan, X., Zhi, X. and W. Yu. Ensemble classification based on supervised clustering for credit scoring. Applied Soft Computing, 2016. 43, 73–86.

[8] Marqués, A.I., García, V. and Sánchez, J.S. Two-level classifier ensembles for credit risk assessment Expert Systems with Applications. Expert Systems with Applications: An International Journal, 39(2), 10916–10922.

[9] Benyacoub, B. Classification with Hidden Markov Model. (2014).

[10] Benyacoub, B., Bernoussi, S.E., and A. Zoglat (2014). Building classification models for customer credit scoring. 2014 International Conference on Logistics Operations Management, pp. 107–111.

[11] Benyacoub, B., Moudden, I.E., Bernoussi, S.E., Zoglat, A., and M. Ouzineb (2015). Initial Model Selection for the Baum-Welch Algorithm Applied to Credit Scoring. In: Le Thi H., Pham Dinh T., Nguyen N. (eds) Modelling, Computation and Optimization in Information Systems and Management Sciences. Advances in Intelligent Systems and Computing, vol 360. Springer, Cham.

[12] Tsai, C.F. Combining cluster analysis with classifier ensembles to predict financial distress. Information Fusion, 16, 46–58.

[13] Efron, B. and R. Tibshirani. Bootstrap Methods for Standard Errors, Confidence Intervals, and Other Measures of Statistical Accuracy. Statistical Science 1 (1986). 1, 54–75. doi:10.1214/ss/1177013815. https://projecteuclid.org/euclid.ss/1177013815

[14] Erenguc, S.S. and G.J. Koehler (1990). Survey of mathematical programming models and experimental results for linear discriminant analysis. Managerial and Decision Economics. 11, 215–225. doi:10.1002/mde.4090110403

[15] Ferrier, G.D. and J.G. Hirschberg. Bootstrapping Confidence Intervals for Linear Programming Efficiency Scores: With an Illustration Using Italian Banking Data, Journal of Productivity Analysis (1997) 8(1), 19–33.

[16] Freed, N. and F. Glover (1986). Evaluating alternative linear programming models to solve the two–group discriminant problem. Decision Sciences, 17, 151–162. doi:10.1111/j.1540-5915.1986.tb00218.x

[17] Freed, N. and F. Glover (1986). Notes and Communications – Resolving certain difficulties and improving the classification power of LP discriminant analysis formulations. Decision Sciences, 17, 589–595. doi:10.1111/j.1540-5915.1986.tb00244.x

[18] Hand, D.J. and W.E. Henley (1997). Statistical Classification Methods in Consumer Credit Scoring: a Review. Journal of the Royal Statistical Society: Series A (Statistics in Society), 160, 523–541. doi:10.1111/j.1467-985X.1997.00078.x

[19] Houshmand A.Z., David J.L. and P.N. Ellinger (1997). Development of Statistical Discriminant Mathematical Programming Model Via Resampling Estimation Techniques, American Journal of Agricultural Economics, Agricultural and Applied Economics Association, 79(4), 1352–1362.

[20] Min, J.H. and Y.-C. Lee. A practical approach to credit scoring Expert Systems with Applications, 35(4), 1762–1770

[21] Jayagopal B. (2004) Applying Data Mining Techniques to Credit Scoring, Amadeus Software Limited.

[22] Lessmann, S., Baesens, B., Seow, H.V. and L.C. Thomas (2015). Benchmarking state-of-the-art classification algorithms for credit scoring: An update of research. European Journal of Operational Research. doi:10.1016/j.ejor.2015.05.030

[23] Thomas, L.C. A survey of credit and behavioural scoring: forecasting financial risk of lending to consumers International Journal of Forecasting, 2000, 16(2), 149–172.

[24] Thomas, L.C., Edelman, D.B. and J.N. Crook. Credit Scoring and its Application, chapter 5, 64–68

[25] Aupetit, S. Contributions aux Modèles de Markov Cachés : métaheuristiques d'apprentissage, nouveaux modèles et visualisation de dissimilarité. pp. 7–50

[26] Eddy S.R. Hidden markov models. Current opinion in structural biology, 1996.

[27] Houshmand A.Z., Leatham, D.J. and C.G. Turvey (1994). Application Of Mathematical Programming Techniques In Credit Scoring Of Agricultural Loans, Proceedings: 1994 Regional Committee NC–207, October 1994, Washington, DC 131475, Regional Research Committee NC–1014: Agricultural and Rural Finance Markets in Transition.