

Credit Risk Analysis Applying Logistic Regression, Neural Networks and Genetic Algorithms Models

Eric Bacconi Gonçalves¹, Maria Aparecida Gouvêa²

¹Department of Marketing, São Paulo State University (USP), Brazil

²Department of Business Administration, São Paulo State University (USP), Brazil

Received: 14 Aug 2021,

Received in revised form: 15 Sep 2021,

Accepted: 22 Sep 2021,

Available online: 30 Sep 2021

©2021 The Author(s). Published by AI
Publication. This is an open access article
under the CC BY license

(<https://creativecommons.org/licenses/by/4.0/>).

Keywords—*credit risk, credit scoring models, genetic algorithms, logistic regression, neural networks.*

Abstract—*Most large Brazilian institutions working with credit concession use credit models to evaluate the risk of consumer loans. Any improvement in the techniques that may bring about greater precision of a prediction model will provide financial returns to the institution. The first phase of this study introduces concepts of credit and risk. Subsequently, with a sample set of applicants from a large Brazilian financial institution, three credit scoring models are built applying these distinct techniques: Logistic Regression, Neural Networks and Genetic Algorithms. Finally, the quality and performance of these models are evaluated and compared to identify the best. Results obtained by the logistic regression and neural network models are good and very similar, although the first is slightly better. Results obtained with the genetic algorithm model are also good, but somewhat inferior. This study shows the procedures to be adopted by a financial institution to identify the best credit model to evaluate the risk of consumer loans. Use of the best fitted model will favor the definition of an adequate business strategy thereby increasing profits.*

I. INTRODUCTION

With the currency stability achieved by the Economical Plano Real in 1994, financial loans became a good business for the banks that no longer made such large profits from currency devaluation (Bresser-Pereira & Nakano, 2002). To replace this profitability, the need to increase investment alternatives was felt at the end of the inflation period. Thereafter institutions have endeavored to expand their credit portfolios. However, loans could not be offered at random to all the applicant clients, therefore ways to evaluate the candidates were required.

Some years ago, when applying for a loan, the client filled in a proposal for evaluation by one or more analysts (Abdou & Pointon, 2011). They then issued an opinion regarding the request. Although effective, the process was slow because it did not accommodate the analysis of many requests. As such, the model for the analysis of the concession of credit was initially introduced

in financial institutions aiming to speed up evaluation of proposals.

Models of analysis for extension of credit known as models of credit scoring are based on historical information from the databank on existing clients, in order to assess whether the prospective client will have a greater chance of being a good or bad payer. The models of credit scoring are added to the institution's systems permitting on-line credit evaluation.

1.1 Objectives of the Study

Based on the data of a sample, the intention is to:

- Develop three credit scoring models by using three statistical/computational techniques: Logistic Regression, Neural Networks, Genetic Algorithms
- Compare the models developed in terms of the quality of fitness and prediction indicators;
- Propose a model for the classification of clients

II. THEORETICAL BASIS

In this section, the theoretical concepts that will support the theme of this work will be presented.

2.1 Consumer Credit

The expression consumer credit may be understood as a form of trade where a person obtains money, goods or services and vouches to pay for this in the future, adding a premium (interest) to the original value (Crook et al., 2007).

Currently, consumer credit is a large industry operating worldwide. Major retailers spur their sales by supplying credit. Automobile companies, banks and other segments utilize consumer credit lines as an additional alternative to make profit. On the other hand, consumer credit injects resources into the economy, permitting production and economic expansion of a country, thereby bringing development to the nation (Lewis, 1992).

However to make credit widely available does not mean to distribute credit at random to all those requesting it; there is a factor associated to consumer credit which is crucial in the decision of making credit available or not: the risk.

2.2 Credit Risk

On the financial market, credit risk is the oldest form of risk (Caouette et al., 2008). It is the upshot of a financial transaction, contracted between the supplier of funds (giver of credit) and the user (taker of credit). Prior to any sophistication resulting from financial engineering, the mere act of lending a sum to someone entails the probability of it not being repaid, the uncertainty regarding return. This is, in essence, the credit risk which may be defined as the risk of a counterpart, in an agreement of credit concession, not to meet his/her obligation.

According to Caouette et al. (2008 p.1), "if credit may be defined as the expectation of receiving a sum of money in a given period, credit risk is a chance that this expectation is not fulfilled".

The activity of credit concession is a basic function of banks, therefore credit risk takes on a relevant role in the composition of an institution's risks and may be found in the operations where there is a transfer of money to the clients as well as in those where there is only a possibility of usage, the pre-conceded limits. Primary types of a bank credit operation are: loans, financing, discount of payables, advancement to depositors, advancement of exchange, leasing operations, surety bonds and warranties etc.

In these operations risk may take on different forms; to be conceptually familiar with them helps to orient management and mitigation.

In the universe of consumer credit, pledge of future payment involves the idea of risk. As the future cannot be fully predicted, all consumer credit involves risk, because assurance of payment does not exist (Lewis, 1992). Analysis of credit is charged with the task of estimating the risk involved in the concession or not of credit.

The maximum risk that the institution may accept relies on the policy adopted by the company. Risk presented by the applicant is of major significance for the process of credit concession, and various queries must be considered in its evaluation.

2.3 Evaluation of the Credit Risk

Evaluation of risk is the main issue for concession of credit. If the risk is poorly evaluated the company will certainly lose money, be it because of acceptance of clients who will generate losses to the business or because of the refusal of good clients who would generate profits for the business. Companies who have a better evaluation than their competitors in the concession of credit have an advantage over the others as they are less vulnerable to the consequence of the wrong decisions when providing credit.

Evaluation of risk of a potential client can be carried out in two ways:

1. By judgment, a more subjective way involving a more qualitative analysis;
2. By classifying the taker by means of evaluation models, involving a more quantitative analysis.

Currently, almost all large sized companies working with concession of credit use a combination of both.

The models called credit scoring are used for the evaluation of risk of credit by classification of the applicant. They permit measurement of the credit applicant's risk, to support the decision taking (concession or not of credit).

2.4 Credit Scoring Models

The pioneer of credit models was Henry Wells, executive of the Spiegel Inc. who developed a credit scoring model during the Second World War (Lewis, 1992).

Wells needed tools that would allow inexperienced analysts to perform credit evaluation, because many of its qualified employees had been recruited for the War.

During the fifties the scoring models were disseminated in the American banking industry. The first models were based upon pre-established weights for certain given characteristics, summing the points to reach a classification score.

More extensive use of the models in the sixties transformed business in the American market (Thomas, 2000). Not only companies in the financial area, but also the large retailers began to use credit scoring models to carry out credit sales to their consumers. Retailers such as Wards, Bloomingdale's and J.C. Penney were some of the pioneers in this segment.

In Brazil the background is shorter. Financial institutions started to make an intensive use of credit scoring models only in the mid-nineties.

There are some steps to be followed to construct a credit scoring model; such as:

1. Survey of a historical background of the clients

The basic supposition to construct a model of credit evaluation is that the clients have the same behavior pattern over time; therefore models are constructed based upon past information. The availability and quality of the data bank are fundamental for the success of the model (Jain et al., 2020)

2. Classification of clients according to their behavior pattern and definition of the dependent variable

In addition to good and bad clients there are also the excluded clients, those who have peculiar characteristics and should not be considered (for instance, workers in the institution) and the indeterminate clients, those on the threshold of being good or bad, still without a clear position about them. In practice, institutions consider only the good and bad clients to build the model because it is much easier to work with binary response models. This tendency to work only with good and bad clients is also noticed in academic works (Amaral & Iquiapaza, 2020; Gonçalves et al., 2013; Locatelli et al., 2015; Ríha, 2016).

3. Selection of a random sample representative of the historical background

It is important that the samples of good and bad clients have the same size so as to avoid any possible bias due to size difference. There is no fixed number for the sample; however Lewis (1992) suggests a sample of 1,500 good clients and 1,500 bad clients to achieve robust results. Habitually three samples are used, one for building of the model, another for the validation of the model and a third to test the model.

4. Descriptive analysis and preparation of data

This consists of analyzing, according to statistic criteria, each variable that will be utilized in the model.

5. Choice and application of techniques to be used in the construction of the model

Logistic Regression, Neural Networks and Genetic Algorithms will be used in this work. Hand & Henley

(1997) further stress Discriminant Analysis, Linear Regression and Decision Trees as methods that can be used in practice. There is no method that is clearly better than the others, everything depends upon how the elected technique fits the data.

6. Definition of the comparison criteria of the models

Measurement for the comparison of the models will be defined here, normally by the rate of hits and the Kolmogorov-Smirnov (KS) statistics.

7. Selection and implementation of the best model

The best model is chosen using the previously defined criteria. As such, the implementation of the model must be programmed. The institution must adjust its systems to receive the final algorithm and program its utilization in coordination with the other areas involved.

III. METHODOLOGICAL PROCEDURES

3.1 Description of the Study

A financial institution wishes to grant loans to its clients and therefore it requires a tool to assess the level of risk associated to each loan to support the decision making process. To set up this project, information on the history of the clients that contracted personal credit was made available.

The product under study is personal credit. Individual credit is a rapid and practical consumer credit operation. The purpose of the loan does not need to be stated, and the loan will be extended according to the applicant's credit scoring.

Another characteristic of the product in question is the lack of requirement of goods as a guarantee of payment. The modality with pre-fixed interest rates with the loan terms ranging from 1 to 12 months was focused for this study.

3.2 The Data

To carry out this study a random selection was made in a universe of clients of the bank, 10,000 credit contracts, considered as good and 10,000 considered as bad. All these contracts had already matured, that is to say the sample was collected after the due date of the last installment of all contracts. This is an historical database with monthly information on the utilization of the product. Based upon this structure, the progress of the contract could be accompanied and particularized when the client did not pay one or more installments.

In the work, the sample is divided into three sub-samples coming from the same universe of interest: one

for construction of the model, 8,000 data (4,000 good and 4,000 bad), the second for validation of the constructed model, 6,000 data (3,000 good and 3,000 bad) and the third also with 6,000 (with the same equal division) to test the model obtained.

3.3 The Variables

The available explanatory variables have characteristics that can be divided into two groups: Reference File Variables, and Variables of Utilization and Restriction. Reference File Variables are related to the client and the Utilization and Restriction Variables regard the restriction of credit and notes about the client's other credit operations existing in the market.

The Reference File Variables as well as those of Utilization and Restriction are collected when the client contracts the product.

3.4 Definition of the Dependent Variable

This definition of the Dependent Variable, also called Performance Definition, is directly related to the institution's credit policy. For the product under study, clients delinquent for 60 or more days were considered Bad (default) and clients with a maximum delinquency of 20 days were considered Good.

Clients designated as undetermined represent a group whose credit behavior is not sufficiently clear to assign them as good or bad customers. In practice, clients who are not clearly defined as good or bad are analyzed separately by the credit analyst, based upon qualitative analysis.

3.5 Logistic Regression

In the models of logistic regression, the dependent variable is, in general a binary variable (nominal or ordinal) and the independent variables may be categorical (as long as dichotomized after transformation) or continuous (Almeida et al., 2020).

The model of Logistic Regression is a particular case of the Generalized Linear Models (Lopes et al., 2017). The function which characterizes the model is given by (Ye & Bellotti, 2019):

$$\ln\left(\frac{p(X)}{1-p(X)}\right) = \beta'X = Z$$

$\beta' = (\beta_0, \beta_1, \beta_2, \dots, \beta_n)$: vector of the parameters associated to the variables

$p(X) = E(Y=1|X)$: probability of the individual has been classified as good, given the vector X.

This probability is expressed by (Gonçalves et al., 2013):

$$p(X) = E(Y) = \frac{e^{\beta'X}}{1 + e^{\beta'X}} = \frac{e^Z}{1 + e^Z}$$

Initially, in this work all variables will be included for the construction of the model; however in the final logistic model, only some of the variables will be selected. The choice of the variables will be done by means of the method forward stepwise, which is the most widely used in models of logistic regression.

Fensterstock (2005) points out the following advantages in using logistic regression for the construction of models:

- The generated model takes into account the correlation between variables, identifying relationships that would not be visible and eliminating redundant variables;
- It takes into account the variables individually and simultaneously;
- The user may check the sources of error and optimize the model.

In the same text, the author further identifies some disadvantages of this technique:

- In many cases preparation of the variables takes a long time;
- In the case of many variables the analyst must perform a pre-selection of the more important, based upon separate analyses;
- Some of the resulting models are difficult to implement.

3.6 Artificial Neural Networks

Artificial Neural Networks are computational techniques that present a mathematical model based upon the neural structure of intelligent organisms and who acquire knowledge through experience.

It was only in the eighties that, because of the greater computational power, neural networks were widely studied and applied. Rojas (1996) underlines the development of the backpropagation algorithm as the turning point for the popularity of neural networks.

An artificial neural network model processes certain characteristics and produces replies like those of the human brain. Artificial neural networks are developed using mathematical models in which the following suppositions are made (Rojas, 1996):

1. Processing of information takes place within the so-called neurons;
2. Stimuli are transmitted by the neurons through connections;

3. Each connection is associated to a weight which, in a standard neural network, multiplies itself upon receiving a stimulus;

4. Each neuron contributes for the activation function (in general not linear) to determine the output stimulus (response of the network).

The pioneer model by McCulloch and Pitts (McCulloch & Pitts, 1943) for one processing unit (neuron) can be summarized in:

- Signals are presented upon input;
- Each signal is multiplied by a weight that indicates its influence on the output of the unit;
- The weighted sum of the signals which produces a level of activity is made;
- If this level exceeds a limit, the unit produces an output.

There are input signals X_1, X_2, \dots, X_p and corresponding weights W_1, W_2, \dots, W_p and the limit being k .

In this model the level of activity is given by:

$$a = \sum_{i=1}^p W_i X_i$$

And the output is given by:

$$y = 1, \text{ if } a \geq k$$

$$y = 0, \text{ if } a < k$$

Three characteristics must be taken into account in the definition of a model of neural networks: the form of the network called architecture, the method for determination of the weights, called learning algorithm; and the activation function.

Architecture relates to the format of the network. Every network is divided in layers, usually classified into three groups (Akkoç, 2012):

- Input Layer where the patterns are presented to the network;
- Intermediate or Hidden layers in which the major part of processing takes place, by means of the weighted connections, they may be viewed as extractors of characteristics;
- Output Layer, in which the end result is concluded and presented.

There are basically three main types of architecture: feedforward networks with a single layer; feedforward networks with multiple layers and recurring networks.

1. Feedforward networks with a single layer are the simpler network, in which there is only one input layer and one output layer. Some networks utilizing this architecture are: the Hebb Network, perceptron, ADALINE, among others.

2. Multilayered feedforward networks are those having one or more intermediate layers. The multilayer perceptron networks (MLP), MADALINE and of a radial base function are some of the networks utilizing this architecture.

3. Recurrent networks: in this type of network, the output layer has at least one connection that feeds back the network. The networks called BAM (Bidirectional Associative Memory) and ART1 and ART2 (Adaptive Resonance Theory) are recurring networks.

The most important quality of neural networks is the capacity to “learn” according to the environment and thereby improve their performance (Deiu-merci & Mayou, 2018).

There are essentially three types of learning:

1. Supervised Learning: in this type of learning the expected reply is indicated to the network. This is the case of this work, where a priori it is already known whether the client is good or bad.

2. Non-supervised Learning: in this type of learning the network must only rely on the received stimuli; the network must learn to cluster the stimuli;

3. Reinforcement Learning: in this type of learning, behavior of the network is assessed by an external reviewer.

Berry & Linoff (2004) point out the following positive points in the utilization of neural networks:

- They are versatile: neural networks may be used for the solution of different types of problems such as: prediction, clustering or identification of patterns;
- They are able to identify non-linear relationships between variables;
- They are widely utilized, can be found in various software.

As for the disadvantages the authors state:

- Results cannot be explained: no explicit rules are produced, analysis is performed inside the network and only the result is supplied by the “black box”;
- The network can converge towards a lesser solution: there are no warranties that the network will find the best possible solution; it may converge to a local maximum.

3.7 Genetic Algorithms

The idea of genetic algorithms resembles the evolution of the species proposed by Darwin: the algorithms will evolve with the passing of generations and the candidates for the solution of the problem one wants to solve “stay alive” and reproduce (Silva et al., 2019).

The algorithm is comprised of a population which is represented by chromosomes that are merely the various possible solutions for the proposed problem. Solutions that are selected to shape new solutions (starting from a cross-over) are selected according to the fitness of the parent chromosomes. Thus, the more fit the chromosome is, the higher the possibility of reproducing itself. This process is repeated until the rule of halt is satisfied, that is to say to find a solution very near to that hoped for.

Every genetic algorithm goes through the following stages:

Start: initially a population is generated formed by a random set of individuals (chromosomes) that may be viewed as possible solutions for the problem.

Fitness: a function of fitness is defined to evaluate the “quality” of each one of the chromosomes.

Selection: according to the results of the fitness function, a percentage of the best fit is maintained while the others are rejected (Darwinism).

Cross-over: two parents are chosen and based upon them an offspring is generated, based on a specific cross-over criterion. The same criterion is used with another chromosome and the material of both chromosomes is exchanged. If there is no cross-over, the offspring is an exact copy of the parents.

Mutation is an alteration in one of the genes of the chromosome. The purpose of mutation is to avoid that the population converges to a local maximum. Thus, should this convergence take place, mutation ensures that the population will jump over the minimum local point, endeavoring to reach other maximum points.

Verification of the halt criterion: once a new generation is created, the criterion of halt is verified and should this criterion not have been met, one returns to the stage of the fitness function.

The following positive points in the utilization of genetic algorithms must be highlighted:

- Contrariwise to neural networks they produce explicable results (Berry & Linoff, 2004)
- Their use is easy (Berry & Linoff, 2004)
- They may work with a large set of data and variables (Fensterstock, 2005)

Some of the disadvantages pointed out in literature are:

- They continue to be seldom used for problems of assessment of risk credit (Fensterstock, 2005)
- Require a major computational effort (Berry & Linoff, 2004)
- Are available in only a few softwares (Berry & Linoff, 2004)

Criteria for Performance Evaluation

To evaluate performance of the model two samples were selected, one for validation and the other for test. Both were of the same size (3,000 clients considered good and 3,000 considered bad, for each one). In addition to the samples, other criteria are used, which are presented in this section.

3.8 Score of Hits

The score of hits is measured by dividing the total of clients correctly classified, by the number of clients included in the model.

Similarly, the score of hits of the good and bad clients can be quantified.

In some situations it is much more important to identify a good client than a bad client (or vice versa); in such cases, often a more fitting weight is given to the score of hits and a weighted mean of the score of hits is calculated.

In this work, as there is not a priori information on what would be more attractive for the financial institution (identification of the good or bad clients), the product between the score of hits of good and bad clients (I_h) will be used as an indicator of hits to evaluate the quality of the model. This indicator will privilege the models with high scores of hits for both types of clients. The greater the indicator is the better will be the model.

3.9 The Kolmogorov-Smirnov Test

The Kolmogorov-Smirnov (KS) is the other criterion often used in practice and used in this work (Fonseca et al., 2019; Lin, 2013; Machado, 2015).

The KS test is a non-parametric technique to determine whether two samples were collected from the same population (or from populations with similar distributions) (Jaklič et al., 2018). This test is based on the accumulated distribution of the scores of clients considered good and bad.

To check whether the samples have the same distribution, there are tables to be consulted according to the significance level and size of the sample (Siegel & Castellan Jr, 2006). In this work, as the samples are large, tendency is that all models reject the hypothesis of equal distributions. The best model will be that with the highest

value in the test, because this result indicates a larger spread between the good and bad.

IV. RESULTS

This section will cover the methods to treat variables, the application of the three techniques under study and the results obtained by each one of them, comparing their performance. For descriptive analysis, categorization of data and application of logistic regression the SPSS for Windows v.21.0 software was used, the software SAS Enterprise Miner. 14.1 was used for the selection of the samples and application to the neural network; for the genetic algorithm a program developed in Visual Basic by the authors was utilized.

4.1 Treatment of the Variables

Initially, the quantitative variables were categorized.

The deciles (values below which 10%, 20% etc. of the cases fall) of these variables were initially identified for categorization of the continuous variables. Starting from the deciles, the next step was to analyze them according to the dependent variable. The distribution of good and bad clients was calculated by deciles and then the ratio between good and bad was calculated, the so called relative risk (RR).

Groups presenting a similar relative risk (RR) were re-grouped to reduce the number of categories by variable.

The relative risks were also calculated for the qualitative variables to reduce the number of categories, whenever possible. According to (Gouvêa et al., 2012) there are two reasons to make a new categorization of the qualitative variables. The first is to avoid categories with a very small number of observations, which may lead to less robust estimates of the parameters associated to them. The second is the elimination of the model parameters, if two categories present a close risk, it is reasonable to group them in one single class.

Besides clustering of categories, RR helps to understand whether this category is more connected to good or to bad clients. This method of clustering categories is explained by Hand & Henley (1997)

When working with the variables made available, heed was given to the following:

- The variables gender, first acquisition and type of credit were not re-coded as they are already binary variables;
- The variable profession was clustered according to the similarity of the nature of jobs;

- The variables commercial telephone and home telephone were recoded in the binary form as ownership or not;

- The variables commercial ZIP Code and home ZIP Code were initially clustered according to the first three digits, next the relative risk of each layer was calculated and later a re-clustering was made according to the similar relative risk, the same procedure adopted by Hand & Henley (1997);

- The variable salary of the spouse was discarded from the analysis because much data was missing;

- Two new variables were created, percentage of the amount loaned on the salary and percentage of the amount of the installment on the salary. Both are quantitative variables, which were categorized in the same way as the remainder.

4.2 Logistic Regression

For the estimation of the model of logistic regression, a sample of 8,000 cases equally divided in the categories of good or bad was utilized.

Initially, it is interesting to evaluate the logistic relationship between each independent variable and the dependent variable TYPE.

Since one of the objectives of this analysis was to identify which variables are more efficient for the characterization of the two types of bank clients, a stepwise procedure was utilized. The elected method of selection was forward stepwise.

With categorical variables, evaluation of the effect of one particular category must be done in comparison with a reference category. The coefficient for the reference category is 0.

Variables with a logistic coefficient estimated negative indicate that the focused category, with regard to the reference, is associated to a decrease of the odds and therefore a decrease in the probability of having a good client.

There are two statistical tests to evaluate the significance of the final model: the chi-square test of the change in the value of $-2LL$ (-2 times the log of the likelihood) and the Hosmer and Lemeshow test.

Table 1 presents the initial value of $-2LL$, considering only the model's constant, its end value, the improvement and the descriptive level to measure its significance.

Table1: Chi-Square test

-2LL	Chi-Square (improvement)	Degrees of freedom	Significance
11090.355			
9264.686	1825.669	28	0.000

The model of 28 variables disclosed that the reduction of the -2LL measure was statistically significant.

The Hosmer and Lemeshow test considers the statistical hypothesis that the predicted classifications in groups are equal to those observed. Therefore, this is a test of the fitness of the model to the data.

The chi-square statistic presented the outcome 3.4307, with eight degrees of freedom and descriptive level equal to 0.9045. This outcome leads to the non rejection of the null hypothesis of the test, endorsing the model's adherence to the data.

4.3 Neural Network

In this work, a supervised learning network will be used, as it is known a priori whether the clients in question are good or bad. According to Potts (1998: 44), the most used structure of neural network for this type of problem is the multilayer perceptron (MLP) which is a network with a feedforward architecture with multiple layers. Consulted literature (Akkoç, 2012; Deiu-merci & Mayou, 2018; Olson et al., 2012; Ríha, 2016) supports this statement. The network MLP will also be adopted in this work.

The MLP networks can be trained using the following algorithms: Conjugate Descending Gradient, Levenberg-Marquardt, Back propagation, Quick propagation or Delta-bar-Delta. The most common (Rojas, 1996) is the Back propagation algorithm which will be detailed later on.

The implemented model has an input layer of neurons, a single neuron output layer, which corresponds to the outcome whether a client is good or bad in the classification of the network. It also has an intermediate layer with three neurons, since it was the network which presented the best outcomes, in the query of the higher percentage of hits as well as in the query of reduction of the mean error. Networks which had one, two or four neurons were also tested in this work.

Each neuron of the hidden layer is a processing element that receives n inputs weighted by weights W_i . The weighted sum of inputs is transformed by means of a nonlinear activation function $f(\cdot)$.

The activation function used in this study will be the logistic function $\frac{1}{1 + e^{(-g)}}$, where

$$g = \sum_{i=1}^p W_i X_i$$

is the weighted sum of the neuron inputs.

Training of the networks consists in finding the set of W_i weights that minimizes one function of error. In this work for the training will be used the Back propagation algorithm. In this algorithm the network operates in a two step sequence. First a pattern is presented to the input layer of the network. The resulting activity flows through the network, layer by layer until the reply is produced by the output layer. In the second step the output achieved is compared to the desired output for this particular pattern. If not correct, the error is estimated. The error is propagated starting from the output layer to the input layer, and the weights of the connections of the units of the inner layers are being modified, while the error is backpropagated. This procedure is repeated in the successive iterations until the halt criterion is reached.

In this model the halt criterion adopted was the mean error of the set of validation data. This error is calculated by means of the module of the difference between the value the network has located and the expected one. Its mean for the 8,000 cases (training sample) or the 6,000 cases (validation sample) is estimated. Processing detected that the stability of the model took place after the 94th iteration. In the validation sample the error was somewhat larger (0.62 x 0.58), which is common considering that the model is fitted based upon the first sample.

Initially, the bad classification is of 50%, because the allocation of an individual as a good or bad client is random; with the increase of the iterations, the better result of 30.6% of error is reached for the training sample and of 32.3% for the validation sample.

Some of the statistics of the adopted network are in table 2.

Table2: Neural network statistics

Obtained statistics	Test	Validation
Misclassification of cases	0.306	0.323
Mean error	0.576	0.619
Mean square error	0.197	0.211
Degrees of freedom of the model	220	
Degrees of freedom of the error	7780	
Total degrees of freedom	8000	

Besides the misclassification and the mean error, the square error and the degrees of freedom are also presented. The average square error is calculated by the average of the squares of the differences between that observed and that obtained from the network.

The number of degrees of freedom of the model is related to the number of estimated weights, to the connection of each of the attributes to the neurons of the intermediate layer and to the binding of the intermediate layer with the output.

4.4 Genetic Algorithms

The genetic algorithm was used to find a discriminate equation permitting to score clients, and later, separate the good from the bad according to the score achieved. The equation scores the clients and those with a higher score are considered good, while the bad are those with a lower score. This route was adopted by Metawa et al., (2017) and Picinini et al. (2003).

The implemented algorithm was similar to that presented in Picinini et al. (2003). Each one of the 71 categories of variables was given an initial random weight. To these seventy one coefficients, one more was introduced, an additive constant incorporated to the linear equation. The value of the client score is given by:

$$S_j = \sum_{i=1}^{72} w_i (p_{ij}), \text{ where}$$

S_j = Score obtained by client j

w_i = Weight relating to the category i

p_{ij} = binary indicator equal to 1, if the client j has the category i and 0, conversely.

The following rule was used to define if the client is good or bad:

If $S_j \geq 0$, the client is considered good

If $S_j < 0$, the client is considered bad

As such, the problem the algorithm has to solve is to find the vector $W = [w_1, w_2, \dots, w_{72}]$ resulting in a classification criterion with a good rate of hits in predicting the performance of payment of credit.

Following the stages of a genetic algorithm, one has:

Start: a population of 200 individuals was generated with each chromosome holding 72 genes. The initial weight w_i of each gene was randomly generated in the interval $[-1, 1]$ (Picinini et al., 2003)

Fitness Function: each client was associated to the estimate of a score and classified as good or bad. By comparing with the information already known a priori on the nature of the client, the precision of each chromosome can be calculated. The indicator of hits (Ih), will be the fitness function, that is to say, the greater the indicator the better will be the chromosome.

Selection: In this work an elitism of 10% was used for each new generation, the twenty best chromosomes are maintained while the other hundred and eighty are formed by cross over and mutation.

Cross-over: to chose the parents for cross-over the method known as roulette wheel was used for selection among these twenty chromosomes that were maintained (Oreski et al., 2012). In this method, each individual is given one probability of being drawn according to its value of the fitness function.

For the process of exchange of genetic material a method known as uniform cross-over was used (Galvan, 2016). In this type of cross-over each gene of the offspring chromosome is randomly chosen among the genes of one of the parents, while the second offspring receives the complementary genes of the second father.

Mutation: in the mutation process, each gene of the chromosome is independently evaluated. Each gene of each chromosome has a 0.5% probability of undergoing mutation. Whenever a gene is chosen for mutation, the genetic alteration is performed, adding a small scalar value k in this gene. In the described experiment a value ranging from -0.05 and + 0.05 was randomly drawn.

Verification of the halt criterion: a maximum number of generations equal to 600 was defined as the halt criterion. After six hundred iterations, the fit chromosome will be the solution.

Results of the algorithm that had the highest Indicator of hits are presented here.

After execution of the algorithm, variables with a very small weight were discarded. In the work by Picinini et al. (2003) the authors consider that the variables with a weight lower than 0.15 or higher than -0.15 would be discarded because they did not have a significant weight for the model. In this work, after performing a sensitivity analysis, it was decided that the variables with a weight higher than 0.10 or lower than - 0.10 would be considered significant for the model. This rule was not applied for the constant, which was proven important for the model even with a value below cutoff.

4.5 Evaluation of the Models' Performance

After obtaining the models the three samples were scored and the Ih and KS were calculated for each of the

models. Table 3 shows the results of classification reached by the three models.

Table 3: Classification results

	Training			Validation			Test		
Logistic Regression	Bad	Good	% Correct	Bad	Good	% Correct	Bad	Good	% Correct
Bad	2833	1167	70.8	2111	889	70.4	2159	841	72.0
Good	1294	2706	67.7	1078	1922	64.1	1059	1941	64.7
Total	4127	3873	69.2	3189	2811	67.2	3218	2782	68.3
Neural Networks									
Bad	2979	1021	74.5	2236	764	74.5	2255	745	75.2
Good	1430	2570	64.3	1177	1823	60.8	1193	1807	60.2
Total	4409	3591	69.4	3413	2587	67.7	3448	2552	67.7
Genetic Algorithms									
Bad	2692	1308	67.3	1946	1054	64.9	2063	937	68.8
Good	1284	2716	67.9	1043	1957	65.2	1073	1927	64.2
Total	3976	4024	67.6	2989	3011	65.1	3136	2864	66.5

All presented good classification results, because, according to Picinini et al. (2003): “credit scoring models with hit rates above 65% are considered good by specialists”.

The hit percentages were very similar in the models of logistic regression and neural network and were somewhat lesser for the model of genetic algorithms. Another interesting result is that, except for genetic algorithms, the models presented the greatest rate of hits for bad clients, with a higher than 70% rate for bad clients in the three samples of the logistic and neural network models.

Table 4 presents results of the criteria Ih and KS which were chosen to compare the models.

Table 4: Comparison indexes

Ih	Training	Validation	Test
Logistic regression	47.9	45.1	46.6
Neural network	47.9	45.3	45.3
Genetic algorithm	45.7	42.3	44.2
KS	Training	Validation	Test
Logistic regression	38	35	37
Neural network	39	35	35
Genetic algorithm	34	30	32

KS values in all models can be considered good. Again, Picinini et al. (2003) explain: “The Kolmogorov-Smirnov test (KS) is used in the financial market as one of the efficiency indicators of the credit scoring models. A model which presents a KS value equal or higher than 30 is considered good by the market”. Here again, the logistic regression and neural network models exhibit very close results, superior to those achieved by the genetic algorithm.

In choosing the model that best fits these data and analyzing according to the Ih and KS indicators, the model built by logistic regression was elected. Although results were very similar to those achieved by neural networks this model presented the best results in the test sample, suggesting that it is best fit for application in other databases. Nevertheless, it must be highlighted that the adoption of any one of the models would bring about good results for the financial institution.

V. CONCLUSION

The objective of this study was to develop credit scoring predictive models based upon data of a large financial institution by using Logistic Regression, Artificial Neural Networks and Genetic Algorithms.

When developing the credit scoring models some care must be taken to guarantee the quality of the model and its

later applicability. Precautions in the sampling, clear definition of criteria for the classification of good and bad clients and treatment of variables in the database prior to application of the techniques were the measures taken in this study, aiming to optimize results and minimize errors.

The three models presented suitable results for the database in question, which was supplied by a large retail bank operating in Brazil. The logistic regression model presented slightly better results to the model built by neural networks and both were better than the model based on genetic algorithms.

This study did not aim at a more detailed approach of the techniques focused. Neural networks and genetic algorithms presented an extensive range of structures and variations that may (and must) be better explored. Genetic algorithms, as they are a rather flexible method, not yet widely researched in problems of credit concession, may be used in diverse forms to optimize results.

In this type of problem, new techniques such as survival analysis should not be overlooked and merit attention in future studies.

REFERENCES

- [1] Abdou, H. A., & Pointon, J. (2011). CREDIT SCORING, STATISTICAL TECHNIQUES AND EVALUATION CRITERIA: A REVIEW OF THE LITERATURE. *Intelligent Systems in Accounting, Finance and Management*, 18(2-3), 59-88. <https://doi.org/10.1002/isaf.325>
- [2] Akkoç, S. (2012). An empirical comparison of conventional techniques, neural networks and the three stage hybrid Adaptive Neuro Fuzzy Inference System (ANFIS) model for credit scoring analysis: The case of Turkish credit card data. *European Journal of Operational Research*, 222(1), 168-178. <https://doi.org/10.1016/j.ejor.2012.04.009>
- [3] Almeida, F. P., Gouveia, R. G. L. de, Lima, M. K. G. de, Ribeiro, F. A. B. S., Mendonça, J. P., & Oliveira, J. do N. (2020). Co-occurrence of Economic, Political and Environmental Factors in the Perception of Social Groups in the Municipality of Uberlândia (Minas Gerais, Brazil) About Notified Cases of Dengue. *International Journal of Advanced Engineering Research and Science*, 7(4), 145-156. <https://doi.org/10.22161/ijaers.74.17>
- [4] Amaral, G. H. de O., & Iquiapaza, R. A. (2020). Determinantes de Inadimplência e de Recuperação de Crédito em um Banco de Desenvolvimento. *BASE – Revista de Administração e Contabilidade Da Unisinos*, 17(3), 483-519. <https://doi.org/10.4013/base.173.05>
- [5] Berry, M. J. a., & Linoff, G. S. (2004). Data mining techniques: for marketing, sales, and customer relationship management. In *Portal.Acm.Org*. https://books.google.com.br/books?hl=pt-BR&lr=&id=AyQfVTDJypUC&oi=fnd&pg=PR37&dq=berry+linoff&ots=KYIppqPXxG&sig=GfmmZURJGHyQzmCH0YUPuiAoXO0&redir_esc=y#v=onepage&q=berry+linoff&f=false
- [6] Bresser-Pereira, L. C., & Nakano, Y. (2002). Uma Estratégia de Desenvolvimento com Estabilidade. *Brazilian Journal of Political Economy*, 22(3), 533-563. <https://doi.org/10.1590/0101-31572002-1246>
- [7] Caouette, J. B., Narayanan, P., Nimmo, R., & Altman, E. I. (2008). *Managing Credit Risk: The Great Challenge for Global Financial Markets* (2nd ed.). John Wiley & Sons.
- [8] Crook, J. N., Edelman, D. B., & Thomas, L. C. (2007). Recent developments in consumer credit risk assessment. *European Journal of Operational Research*, 183(3), 1447-1465. <https://doi.org/10.1016/j.ejor.2006.09.100>
- [9] Deiu-merci, K. K., & Mayou, M. (2018). Network Data Security for the Detection System in the Internet of Things with Deep Learning Approach. *International Journal of Advanced Engineering Research and Science*, 5(6), 208-213. <https://doi.org/10.22161/ijaers.5.6.34>
- [10] Fensterstock, A. (2005). Credit scoring and the next step. *Business Credit*, 46-50.
- [11] Fonseca, S. E., Santos, A. de O., Pereira, M. V. L., & Camargos, M. A. de. (2019). Análise do Impacto de Variáveis Macroeconômicas no Desempenho Financeiro e Endividamento de Empresas Listadas na B3. *Revista Universo Contábil*, 14(4), 93-114. <https://doi.org/10.4270/ruc.2018429>
- [12] Galvan, P. (2016). Educational Evaluation and Prediction of School Performance through Data Mining and Genetic Algorithms. *International Journal of Advanced Engineering Research and Science*, 3(10), 215-220. <https://doi.org/10.22161/ijaers/3.10.34>
- [13] Gonçalves, E. B., Gouvêa, M. A., & Mantovani, D. M. N. (2013). Análise de risco de crédito com o uso de regressão logística. *Revista Contemporânea de Contabilidade*, 10(20), 139-160. <https://doi.org/10.5007/2175-8069.2013v10n20p139>
- [14] Gouvêa, M. A., Gonçalves, E. B., & Mantovani, D. M. N. (2012). Aplicação De Regressão Logística E Algoritmos Genéticos Na Análise De Risco De Crédito. *Revista Universo Contábil*, 84-102. <https://doi.org/10.4270/ruc.2012214>
- [15] Hand, D. J., & Henley, W. E. (1997). Statistical Classification Methods in Consumer Credit Scoring: a Review. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 160(3), 523-541. <https://doi.org/10.1111/J.1467-985X.1997.00078.X>
- [16] Jain, A., Patel, H., Nagalapatti, L., Gupta, N., Mehta, S., Guttula, S., Mujumdar, S., Afzal, S., Sharma Mittal, R., & Munigala, V. (2020). Overview and Importance of Data Quality for Machine Learning Tasks. *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 3561-3562. <https://doi.org/10.1145/3394486.3406477>
- [17] Jaklič, J., Grublješič, T., & Popovič, A. (2018). The role of compatibility in predicting business intelligence and analytics use intentions. *International Journal of Information Management*, 43(August), 305-318.

- <https://doi.org/10.1016/j.ijinfomgt.2018.08.017>
- [18] Lewis, E. M. (1992). *An Introduction to Credit Scoring*.
- [19] Lin, R. (2013). *The application and assessment of consumer credit scoring models in measuring consumer loan issuing risk of commercial banks in China*. May.
- [20] Locatelli, R. L., Afonso, T., Ramalho, W., & Silvério, R. A. de O. (2015). Determinantes da inadimplência no crédito habitacional direcionado a classe média emergente brasileira. *Revista de Finanças Aplicadas*, 1(1), 1–30. <http://repositorio.uninove.br/xmlui/handle/123456789/1112>
- [21] Lopes, M. G., Ciribeli, J. P., Massardi, W. D. O., & Mendes, W. D. A. (2017). Análise dos indicadores de inadimplência nas linhas de crédito para pessoa física: um estudo utilizando modelo de regressão logística. *Estudos Do CEPE*, 46, 75. <https://doi.org/10.17058/cepe.v0i46.11099>
- [22] Machado, A. R. (2015). *Collection Scoring via Regressão Logística e Modelo de Riscos Proporcionais de Cox*. Universidade de Brasília.
- [23] McCulloch, W. S., & Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. *The Bulletin of Mathematical Biophysics* 1943 5:4, 5(4), 115–133. <https://doi.org/10.1007/BF02478259>
- [24] Metawa, N., Hassan, M. K., & Elhoseny, M. (2017). Genetic algorithm based model for optimizing bank lending decisions. *Expert Systems with Applications*, 80, 75–82. <https://doi.org/10.1016/J.ESWA.2017.03.021>
- [25] Olson, D. L., Delen, D., & Meng, Y. (2012). Comparative analysis of data mining methods for bankruptcy prediction. *Decision Support Systems*, 52(2), 464–473. <https://doi.org/10.1016/j.dss.2011.10.007>
- [26] Oreski, S., Oreski, D., & Oreski, G. (2012). Hybrid system with genetic algorithm and artificial neural networks and its application to retail credit risk assessment. *Expert Systems with Applications*, 39(16), 12605–12617. <https://doi.org/10.1016/j.eswa.2012.05.023>
- [27] Picinini, R., Oliveira, G. M. B., & Monteiro, L. H. A. (2003). Detecção de problemas de crédito em empresas de pequeno porte usando redes neurais e algoritmo genético. *SIMPÓSIO BRASILEIRO DE AUTOMAÇÃO INTELIGENTE*.
- [28] Ríha, J. (2016). *Artificial Intelligence Approach to Credit Risk* [Charles University]. file:///E:/Downloads/DPTX_2013_2_11230_0_415651_0_151649.pdf
- [29] Rojas, R. (1996). Neural networks: a systematic introduction. In *Springer Science & Business Media*. https://books.google.com.br/books?hl=pt-BR&lr=&id=4rESBwAAQBAJ&oi=fnd&pg=PA3&ots=V Bf8cRZWqP&sig=wKOJYQs4mZa3iR1F56RB-rzB6zM&redir_esc=y#v=onepage&q&f=false
- [30] Siegel, S., & Castellan Jr, N. J. (2006). *Estatística não-Paramétrica Para Ciências do Comportamento* (2nd ed.). Bookman.
- [31] Silva, M. F. da, Silva, W. G. da, Carvalho, R. L. de, Silva, E. M. da, & Almeida, T. da S. (2019). Analysis of Genetic Algorithm for synthesis digital systems modeled in finite state machine. *International Journal of Advanced Engineering Research and Science*, 6(7), 218–222. <https://doi.org/10.22161/ijaers.6726>
- [32] Thomas, L. C. (2000). A survey of credit and behavioural scoring: forecasting financial risk of lending to consumers. *International Journal of Forecasting*, 16(2), 149–172. [https://doi.org/10.1016/S0169-2070\(00\)00034-0](https://doi.org/10.1016/S0169-2070(00)00034-0)
- [33] Ye, H., & Bellotti, A. (2019). Modelling recovery rates for non-performing loans. *Risks*, 7(1), 1–17. <https://doi.org/10.3390/risks7010019>