UNIVERSIDADE FEDERAL DO RIO DE JANEIRO
INSTITUTO COPPEAD DE ADMINISTRAÇÃO

DIEGO PAGANOTI FONSECA

A TWO-STAGE FUZZY NEURAL APPROACH FOR CREDIT RISK ASSESSMENT

Rio de Janeiro

2019

DIEGO PAGANOTI FONSECA

A TWO-STAGE FUZZY NEURAL APPROACH FOR CREDIT RISK ASSESSMENT

Master's dissertation presented to the Instituto COPPEAD de Administração, Universidade Federal do Rio de Janeiro, as part of the mandatory requirements in order to obtain the degree of Master in Business Administration (M.Sc.).

SUPERVISOR: Peter Fernandes Wanke, D.Sc.

Rio de Janeiro

2019

CIP - Catalogação na Publicação

F676t

Fonseca, Diego Paganoti
A TWO-STAGE FUZZY NEURAL APPROACH FOR CREDIT RISK ASSESSMENT / Diego Paganoti Fonseca. -- Rio de Janeiro, 2019.
77 f.

Orientador: Peter Fernandes Wanke.
Dissertação (mestrado) - Universidade Federal do Rio de Janeiro, Instituto COPPEAD de Administração, Programa de Pós-Graduação em Administração, 2019.

1. Modelos Preditivos. 2. Credit Scoring. 3. Sistema de Inferência Fuzzy. 4. Redes Neurais. 5. Administração - Teses. I. Fernandes Wanke, Peter, orient. II. Título.

Elaborado pelo Sistema de Geração Automática da UFRJ com os dados fornecidos pelo(a) autor(a), sob a responsabilidade de Miguel Romeu Amorim Neto - CRB-7/6283.

DIEGO PAGANOTI FONSECA

A TWO-STAGE FUZZY NEURAL APPROACH FOR CREDIT RISK ASSESSMENT

Master's dissertation presented to the Instituto COPPEAD de Administração, Universidade Federal do Rio de Janeiro, as part of the mandatory requirements in order to obtain the degree of Master in Business Administration (M.Sc.).

Approved by:

Peter Fernandes Wanke

Peter Fernandes Wanke, D.Sc
(COPPEAD/UFRJ)

Otávio Henrique dos Santos Figueiredo

Otávio Henrique dos Santos Figueiredo, D.Sc
(COPPEAD/UFRJ)

Virgílio José Martins Ferreira Filho

Virgílio José Martins Ferreira Filho, D.Sc
(COPPE/UFRJ)

ACKNOWLEDGEMENTS

First, this work is dedicated to my father, who passed away during this research and, unfortunately, was not able to see its conclusion. In any case, he definitely is a fundamental part of, not only it, but all my life.

In this regard, I am fully grateful to my mom, who vigorously handled all the emotions during this difficult period, giving me all the necessary support and love to keep going towards my objective. I thank my brother as well, due to his sense of unity and his partnership in such hard moments.

I am also deeply thankful to my supervisor Peter Fernandes Wanke, who patiently guided me through this path, giving all the relevant advices and directions to conclude it.

Special thanks also to CAPES, who granted me with a very helpful financial support. Without this benefit, I wouldn’t be able to finish my research.

I would like to thank our amazing institution, Instituto COPPEAD de Administração and all collaborators. Simply put, their work changes lives. And finally, I am grateful to have the opportunity to be part of the 2017 cohort, full of brilliant people that contributed significantly to my development.

ABSTRACT

FONSECA, Diego Paganoti. Title: A TWO-STAGE FUZZY NEURAL APPROACH FOR CREDIT RISK ASSESSMENT. 2019. 77p. Dissertation project (Master in Business Administration) - Instituto COPPEAD de Administração, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2019

This study intends to assess the clients' credit risk of a Brazilian private credit card provider through a two-stage process, involving soft computing techniques and the usage of different credit score ratings. Then, a discussion about performance is conducted, not only in general terms, but also demographically segregated. One of the approaches comprises of the sole usage of a soft computing algorithm – artificial neural networks (ANN) – for client classification into solvent or non-solvent situation, having market available credit score rating as input; while the second approach is a proposed two-stage process, that is the usage of a fuzzy inferenced input to a similar ANN. This process includes also, in a previous step, an ANN regression task, using a credit score rating already available as response in order to conduct the fuzzy reasoning step. The main takeaway is that, although not presenting the best results comparing to market options, it is possible to create a competitive credit score rating with past information using a fuzzy inference system.

Keywords: Predictive Modeling. Credit Scoring. Fuzzy Inference System. Neural Networks. Classification.

LIST OF FIGURES

FIGURE 1: A NEURAL NETWORK WITH D INPUT NEURONS, M OUTPUT NEURONS AND ONE HIDDEN LAYER WITH N NEURONS ____________ 8
FIGURE 2: TERMS OF THE LINGUISTIC VARIABLE AGE ______________ 12
FIGURE 3: THE BASIC COMPONENTS OF A FUZZY SYSTEM __________ 13
FIGURE 4: MODELING STRUCTURE________________________________ 19
FIGURE 5: DISTRIBUTION OF SCORES THROUGH CLIENT'S SOLVENCY 20
FIGURE 6: CORRELATION MATRIX AMONG VARIABLES______________ 23
FIGURE 7: HISTOGRAMS OF NUMERICAL VARIABLES _______________ 23
FIGURE 8: LEFT: EXAMPLE OF TWO VARIABLES BEFORE SPATIAL SIGN TRANSFORMATION. RIGHT: AFTER TRANSFORMATION __________ 24
FIGURE 9: NETWORK STRUCTURE FOR REGRESSION STEP _________ 24
FIGURE 10: NEURAL NETWORK REGRESSIONS ERRORS AND R^2 ____ 25
FIGURE 11: BEST MODEL FITTED WITHIN THE 100 BOOTSTRAPPED RESAMPLES____ 25
FIGURE 12: VARIABLE IMPORTANCE, ORDERED BY MEDIAN _________ 27
FIGURE 13: SCHEMATIC VIEW OF FUZZY SCORE CREATION _________ 27
FIGURE 14: NUMERICAL VARIABLES AND DEFAULT SITUATION_______ 28
FIGURE 15: RULE BLOCK 1 MEMBERSHIP FUNCTIONS (DEMOGRAPHIC), WHERE SCORES.DEMO IS THE INTERMEDIARY SCORE GENERATED ____ 32
FIGURE 16: RULE BLOCK 2 MEMBERSHIP FUNCTIONS (FINANCE), WHERE SCORES.FIN IS THE INTERMEDIARY SCORE GENERATED 33
FIGURE 17: RULE BLOCK 3 MEMBERSHIP FUNCTIONS (ASSETS), WHERE SCORES.ASSETS IS THE INTERMEDIARY SCORE GENERATED___ 33
FIGURE 18: RULE BLOCK 4 MEMBERSHIP FUNCTIONS (FINAL), WHERE SCORES.FINAL IS THE FINAL FUZZY SCORE GENERATED ________ 34
FIGURE 19: DISTRIBUTION AND DEFAULT SITUATION FOR FUZZY SCORE ____ 34
FIGURE 20: CLASSIFICATION MODEL STRUCTURE__________________ 36
FIGURE 21: SCHEMATIC REPRESENTATION OF HOW MODELS WERE TRAINED AND TESTED _______________________________________ 36
FIGURE 22: GENERAL RESULTS FOR MLPS (NAMED WITH THE SCORES USED AS PREDICTOR) AND LOGISTIC MODELS (LOGIT) __________ 40
FIGURE 23: ROC CURVES FOR BEST MODEL (IN TERMS OF AUC) FOR EACH METHOD ____ 41
FIGURE 24: COHEN'S D ESTIMATE FOR THE SCORES (FROM SEPARATION OF DEFAULTED AND SOLVENT CLIENTS) __________ 43
FIGURE 25: RESULTS FOR TOPSIS PERFORMANCE SEGREGATED BY AGE____ 44
FIGURE 26: RESULTS FOR TOPSIS PERFORMANCE SEGREGATED BY INCOME ____ 44

FIGURE 27: RESULTS FOR TOPSIS PERFORMANCE SEGREGATED BY GENDER 45
FIGURE 28: RESULTS FOR TOPSIS PERFORMANCE SEGREGATED BY REGION 45
FIGURE 29: RESULTS FOR TOPSIS VARIABILITY SEGREGATED BY AGE 46
FIGURE 30: RESULTS FOR TOPSIS VARIABILITY SEGREGATED BY INCOME 46
FIGURE 31: RESULTS FOR TOPSIS VARIABILITY SEGREGATED BY GENDER 46
FIGURE 32: RESULTS FOR TOPSIS VARIABILITY SEGREGATED BY REGION 47

LIST OF TABLES

TABLE 1. METHODS FOR CREDIT RISK EVALUATION ____ 5
TABLE 2: DESCRIPTIVE STATISTICS FOR NUMERICAL VARIABLES ____ 17
TABLE 3: DISTRIBUTION OF CATEGORICAL VARIABLES ____ 17
TABLE 4: D ESTIMATES FOR EACH SCORE RATING ____ 20
TABLE 5: PROPORTIONS OF BANKRUPTED OVER SOLVENT CLIENTS 29
TABLE 6: AVERAGE IMPORTANCE OF THE RULE BLOCKS ____ 30
TABLE 7: AVERAGE COMPUTING TIMES FOR EACH MODEL (MINUTES AND SECONDS) ____ 40
TABLE 8: STANDARD DEVIATIONS OF MODELS' RESULTS ____ 41
TABLE 9: TOPSIS RESULTS BASED ON MEDIAN OF RESULTS AND STANDARD DEVIATIONS ____ 42
TABLE 10: BLOCK 1 DEMOGRAPHIC RULES ____ 51
TABLE 11: RULE BLOCK 2 FINANCE ____ 55
TABLE 12: RULE BLOCK 3 ASSETS ____ 56
TABLE 13: RULE BLOCK FINAL ____ 57
TABLE 14: TOPSIS PERFORMANCE SEGREGATED PER REGION ____ 59
TABLE 15: TOPSIS PERFORMANCE SEGREGATED BY GENDER ____ 59
TABLE 16: TOPSIS PERFORMANCE SEGREGATED BY INCOME ____ 59
TABLE 17: TOPSIS PERFORMANCE SEGREGATED BY AGE ____ 59
TABLE 18: TOPSIS VARIABILITY SEGREGATED BY REGION ____ 60
TABLE 19: TOPSIS VARIABILITY SEGREGATED BY GENDER ____ 60
TABLE 20: TOPSIS VARIABILITY SEGREGATED BY INCOME ____ 60
TABLE 21: TOPSIS VARIABILITY SEGREGATED BY AGE ____ 60

LIST OF ABBREVIATIONS

ANN – Artificial Neural Networks

AUC – Area Under Curve

FED – Federal Reserve System

FICO – Fair Isaac Corporation

FRBS – Fuzzy Rule Based Systems

LN – Lexis Nexis

MCDM – Multi-Criteria Decision Making

MLP – Multilayer Perceptron

NYSE – New York Stock Exchange

ROC – Receiver Operating Characteristic

ROC – Receiver Operating Characteristic

TOPSIS – Technique for Order Preference by Similarity to Ideal Solution

TS – Takagi Sugeno

TU – Trans Union

SUMMARY

1. INTRODUCTION **1**
1.1 Research motivation **1**
1.2 Study Structure **3**
2 LITERATURE REVIEW **4**
2.1 Credit Risk Assessment **4**
2.2 Soft Computing Models **7**
2.2.1 Artificial Neural Networks 7
2.2.2 Fuzzy Logic 11
2.2.3 Combining ANN and FRBS 15
3 METHODOLOGY **16**
3.1 The Dataset **16**
3.2 Modeling Structure **17**
3.3 Phase 1: Preprocessing and Regression (Fuzzy Stage) **19**
3.3.1 Benchmark Score 19
3.3.2 Preprocessing 20
3.3.3 Regression 22
3.4 Phase 2: Fuzzy Score Creation (Fuzzy Stage) **26**
3.4.1 Data Analysis 28
3.4.2 Rules and Membership Functions 30
3.4.3 Fuzzy Inference System 31
3.5 Phase 3: Classification **34**
3.5.1 Model Pre-processing and Tuning (Neural Stage) 34
3.5.2 Treating Class Imbalance 35
3.5.3 Logistic Regression as a Benchmark 36
4 RESULTS AND DISCUSSION **37**
4.1 Metrics Used **37**
4.1.1 Sensitivity and Specificity 37
4.1.2 Geometric Means and Area Under Curve of Receiver Operating Characteristic 38
4.1.3 TOPSIS 38
4.2 Models Evaluation and Discussion **39**

4.2.1 General Results (all dataset) 39
4.2.2 Segregated by Selected Demographic Variables 44
5 CONCLUSION 47
5.1 Limitations **49**
5.2 Recommendations **50**
APPENDIX A – RULE BLOCKS FOR PHASE 2 51
APPENDIX B – RESULTS TABLES 59
BIBLIOGRAPHY 61

1 INTRODUCTION

1.1 Research motivation

It is undoubtedly true that uncertainty plays a big role when it comes to how business is conducted in organizations. While it is feasible to consider its beneficial role to some sectors, such as consultancy and advisory (legal, accounting, etc.), it is a common sense that uncertainty is a source of big concern, and hence, a negative aspect, for most industries. Everyday businesses must deal with doubts about future events concerning weather, politics, environment, law, economics, foreign trading and affairs, technology and marketing trends, among many other aspects.

Traditional approaches to deal with uncertainty are discussed by Courtney et al. (1997). The standard practice is to lay out a vision of future events precise enough, that they can be captured, for example, in a discounted cash flow analysis. However, such kind of attitude tend to lead executives to assess the future in a binary way – either certain or completely unpredictable, which may result in underestimation of uncertainty. But there is the other extreme: to assume that the world is entirely unpredictable. This approach is also dangerous and misleading, since it could make managers to base strategic decisions or long-term beliefs primarily on gut instinct or subjective judgements. Infamous examples include when the Prussian emperor Wilheim II stated, in the end of 19th century, that automobiles were a transitory phenomenon (CHESSHYRE, 2009) or when former Microsoft CEO Steve Ballmer stated, in 2007 that “there’s no chance that the iPhone is going to get any significant market share” (LIEBERMAN, 2007).

Decisions are intrinsically made by considering (or not) the likelihood about future events and uncertainty about them. And one prudent and very common way of doing so, apart from using intuition, is to analyze past information. Considering the recent developments in telecommunications and media, concomitant with the advent of the Internet, information of all kind has become more readily available to organizations and people across the world. As discussed by Kuhn and Johnson (2013), the human brain can consciously and subconsciously assemble a considerably big amount of data, however, it is impossible for it to process the even greater amount of relevant information, which is easily and consistently obtained in shorter periods of time today.

Nowadays, it is reasonable to consider that any people who intends to, either learn or gather information, with the aim to solve any given problem or to make a critical decision will, at least once, use a search engine, such as Google. This tool, as many others with the same objective, uses techniques that gather current information available, looks for patterns relevant to our queries and returns answers. A vast array of fields of science and technology, from chemistry to finance, developed tools and ways to make use of these kind of engines, which are part of a computer science field of study called “machine learning”, a subset of artificial intelligence field (GANGULY, 2018). It comprises techniques that give computers the ability to progressively improve performance on a specific task with data, without being explicitly programmed (SAMUEL, 1959). Such improvement is achieved by learning with an available set of data and, while its performance gets improved, more accurate predictions are generated.

A commonly used term to describe recent developed techniques to deal with uncertainty is “predictive modelling”. Kuhn and Johnson’s (2013) definition for it is as follows: “the process of developing a mathematical tool or model that generates an accurate prediction”. And to make predictions using information and tools related to machine learning is not only a growing trend, but also a prudent way of dealing with uncertainty about future events.

However, there are some criticism on traditional predictive modelling. Maybe due to the “flash crash” that occurred with the NYSE stocks in May 6th 2010 (PHILLIPS, 2010) and also the subprime crisis that took place in 2008, Rodriguez (2011) wrote an article in which he stated that “the models (…) fail to incorporate the nearly endless human capacity to behave irrationally, which cannot be neatly reduced to an algorithmic equation”.

In fact, and probably without any intention to do so, such article resonated what was the stimulus for the initial development of the theory of fuzzy sets, back in 1965. According to Bojadziev & Bojadziev (2007), such theory was motivated by the perception that traditional techniques of system analysis were not effective. Its development was a point of departure for the creation of sufficient knowledge to acquire the capability to model the imprecision and uncertainty inherent to the real world and to capture the problem-solving heuristics that human beings use every day.

On such grounds, this study aims at exploring two soft computing techniques for predictive modelling in the context of credit risk assessment. The first is all about the usage of machine learning and is called Artificial Neural Network (ANN). The other’s main feature is to address vagueness, called Fuzzy Logic Rule Based System (FRBS).

In that sense, a two-stage process comprising both fields of study (resulting in a “fuzzy-neural” approach) is explored.

Predicting bankruptcy of companies or creditworthiness of consumers is an imprecise and ambiguous practice (KOROL, 2012) and, in this manner, this work is applied both in the context of credit scoring and default classification, from the viewpoint of assessing credit consumer bankruptcy threat. The dataset to be used is composed by nearly 9,500 clients of a Brazilian credit card operator, containing 19 variables, one of them informing if the client had previously failed to pay. Moreover, 5 of such variables are score ratings of private consumer credit scoring agencies. The remaining variables comprises demographics, financial and other sort of information from each of the clients. In part, this work extends the methodology proposed by Tomasz Korol (2012), whereby a fuzzy classification system is developed for credit card clients based on a pack of metrics such as demographics, finance and security conditions. Here, though, a fuzzy score is generated and applied as input to train an ANN aiming at classifying clients under groups of potentially non-default and default clients. Essentially, the final objective is to explore differences in performance between (i) an ANN model containing a fuzzy generated input and (ii) ANN models, each of them containing (or not), a variable that is a market score rating, under different demographic circumstances and from the viewpoint of uncertainty and vagueness concepts. Finally, a comparison with a widely used and successful technique for credit risk assessment, the logistic regression, is also made.

1.2 Study Structure

This work comprises five chapters, where the first contextualize the areas of study, introduces some aspects of the work, present its objectives and its relevance. Then, the second chapter develops the literature review, in which the following topics will be presented: a discussion related to previous works on traditional ways of credit assessment and conceptual explanation of ANN and fuzzy logic. The third chapter will present modeling facts and the methodology of the study, going from the basic aspects of the data, then to the preprocessing techniques, then to the proposition of a new credit score rating (based upon fuzzy logic concepts) and finally to how the classification is conducted. The fourth chapter explores the performance of the approach suggested (with a fuzzy input), compared to the other methods for credit assessment, including brief explanation of the metrics used for comparison and methodology proposed for such evaluation. Finally, the fifth chapter brings the study's conclusion, its limitations and suggestions for future research.

2 LITERATURE REVIEW

2.1 Credit Risk Assessment

The predictive power of creditworthiness, either of a company or a retail client, is of fundamental relevance for any kind of financial institution and their core businesses. Not only bad decisions can lead to losses due to client bankruptcies, but also good opportunities can be put aside. In the 'loss side', an infamous example of the impact of bad credit risk decisions on the economy is the financial meltdown that occurred in 2008-2009 – the worst since 1929-, ultimately a consequence of the US subprime crisis (TORBAT, 2008). Banks and mortgage lenders suffered significant losses due to house finance payment defaults, a phenomenon that led to a stressed financial system, contributing largely to such a large economic downturn (CAMPBELL; COCCO, 2015). LAHSASNA (2009) remembers that one of the main reasons for such was the increasing offer of higher risk loans to higher risk borrowers and that credit rating agencies should have helped on assessing the risk involved in such riskier transactions.

Nevertheless, credit risk evaluation poses a classic problem in terms of decision-making regarding uncertainty, because an individual's worth is usually based on estimates of potential future incomes. Moreover, this information may lack accuracy, since it is possible that an individual possesses multiple (and perhaps undeclared) income streams (IGNATIUS et al., 2016). Above that, it should be considered that credit screening and its conclusion also depends on the lending institution's flexibility in interpreting customer's and market's information toward future forecasted trends, which can result in different ways to measure the credit worthiness of a customer. Ignatius et al. (2016) exemplifies this latter issue mentioning the adoption of the FICO scoring (a credit scoring model developed by Fair Isaac and Company) by the United States Federal Reserve as an effort to promote uniformity in credit evaluation.

From all that, it is clear to realize the magnitude of the importance of a sound and trustworthy credit risk decision support tool. Analysts are no more faced with a dilemma on whether to predict or guess the creditworthiness of entities (either consumers or businesses). Rather, they now need to evaluate which forecasting method is better at minimizing errors of prediction.

Credit scoring is usually the most popular technique to evaluate creditworthiness of applicants based on their characteristics (LAHSASNA, 2009). And conventional probabilistic-based models such as Discriminant and Logistic Regression have often been applied to build such credit scoring models (IGNATIUS, 2016). However, in addition to these, several other methods arose over the recent years. Generally, forecasting models are grouped into three groups: statistical, theoretical and soft computing methods (table 1). Upon such taxonomy, according to an extensive literature review made by Aziz & Dar (2006), 64% of case studies used statistical models, 25% used soft computing techniques and 11% other types of models.

Table 1. Methods for Credit Risk Evaluation

Statistical Methods	Soft Computing Methods	Theoretical Models
Discriminant Analysis	Artificial Neural Networks	Hazard Model
Logistic Regression	Fuzzy Logic	Other Credit Risk Models
Probit	Genetic Algorithms	
Decision Tree	Support Vector Machines	
Scoring		

Source: KOROL, 2012

Besides the above, Grace and Williams (2016) mention a "traditional and more heuristic method" than all others, by which the granting of credit to borrowers is based upon judgmental concept using experience of credit officers. This approach naturally carries setbacks like high cost of training finance professionals, inappropriate decisions, longer periods of time required to evaluate different cases and the chances of guiding different decisions (by different officers or institutions) for the same case (HANDZIC and AURUM 2001).

Statistical models for credit risk assessment, though being the most used in case studies (logistic regression specifically as the main benchmark method (LOUZADA et al., 2016)), usually require variables to follow assumptions such as (i) normal distribution, (ii) inter independence, (iii) high discriminative ability of separating solvent from insolvent entities, (iv) absence of missing information and (v) clear and exclusive definition of membership into either a group or another (KOROL, 2012). Discriminant analysis and logistic regression, for instance, assume multivariate normality and homoscedasticity, which are features not often present in real world of credit institutions data (GIANG, 2005; HUANG et al., 2004). Although some of these assumptions can be considered as setbacks, statistical models, in general, address latent problems of the aforementioned “traditional and heuristic method”. For example, one of the greatest advantages of credit scoring models is the attempt to correct the intrinsic bias inherent to such simpler methods by means of uniformizing how a candidate is perceived. However, it is still possible that they’ll misclassify applicants to a credit line and that they won’t easily accommodate new contexts (GRACE and WILLIAMS, 2016). In addition, in a comparison study made by Louzada et al. (2016), the Logistic model performed reasonably well, being among the top five credit risk classifiers out of 9 different models.

Soft computing models’ main feature, in another vein, is that they can handle imprecisely defined problems, incomplete data, uncertainty and approximation, features present in the issue of both retail and business bankruptcy prediction. They are also suitable for use in dynamic and continuously evolving contexts and environmental changes, whereby they gather knowledge and formulate rules of inference (in case of fuzzy logic, for instance), with the objective of making predictions and classifications about situations based on previously observed data. While statistical modeling assumes precision, reliability and accuracy of variables, fuzzy based models, for example, work upon the thesis that precision and certainty do carry a cost and decision making should exploit tolerance for these characteristics whenever possible (KOROL, 2012; ALIEV et al., 2012).

Finally, theoretical models focus on the use of qualitative information for prediction of bankruptcy. These methods rely on the causes of the collapse, rather than the symptoms per se (such as general characteristics of the observed data), which is the point of focus for soft computing and statistical methods (KOROL, 2012). It is done using statistical techniques for drawing quantitative conclusions over theoretical arguments. Proportional hazard modeling, for example, is a class of survival model, by which the time that passes before some event occurs is related to a variable (in credit analysis, the event could be a client default). Survival analysis is an alternative to logistic regression, and it is argued that one of its advantages is that the time to default can also be modeled, not just whether an applicant will default or not (DIRICK et al., 2017).

2.2 Soft Computing Models

For general prediction, soft computing modelling moderately stands out among the other methods briefly presented before, notably from 1980s on, when a rapid growth in its popularity arose for all kind of uses, mainly in technology, such as data compression for HDTV, audio recording, speech recognition, image understanding and related fields. (ZADEH, 1993; ALIEV et al., 2012). Back in the beginning of 1990s, it was already argued that soft computing should be viewed as the foundation of artificial intelligence, rather than hard computing (ZADEH, 1993). Such growth occurred mainly due to a feature it possesses, which is the ability to approximate a solution to a precisely (or, more often, imprecisely) formulated problem, a crucial difference from hard computing. Zadeh (1993) exemplifies it by mentioning the problem of parking a car. Although it has a very simple solution because the final position isn't specified exactly (and, hence, would be an approximated solution), it would take hours or maybe days of maneuvering to achieve fractions of a millimeter of distance and seconds of an arc angle in case such a level of precision was specified. In essence, soft computing is a consortium of artificial intelligence paradigms, such as Fuzzy Logic, Neurocomputing, Evolutionary Computing, Probabilistic Computing and Chaotic Computing, all of which enable one to solve important real-world problems that wouldn't be solved by other existing technologies (ALIEV et al., 2012).

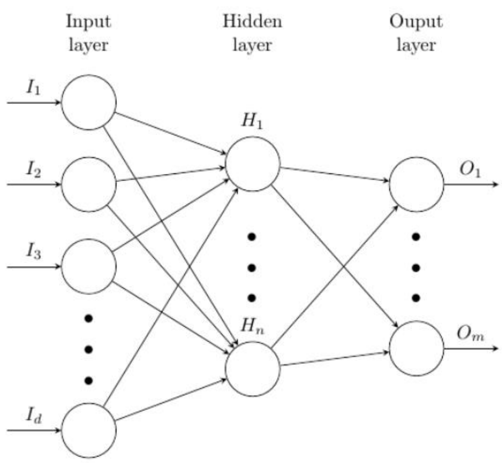
Fuzzy Logic and ANN are techniques that have been recently applied as emerging technologies to both industrial and non-industrial areas, like credit risk evaluation. While the first is concerned with imprecision and approximate reasoning, ANN's capacity is maximized in learning procedures (ALIEV et al., 2012). Both will be the focus of this work, due to recent increase in adoption and their successful applications to credit risk assessment in loan and credit institutions (BAHRAMMIRZAEE, 2010; GRACE and WILLIAMS, 2016).

2.2.1 Artificial Neural Networks

Zadeh, the proposer of fuzzy mathematics, once noted that the perfect model of soft computing is the human brain. And Artificial Neural Networks are a suitable candidate for such task, being viable computational models for a wide variety of problems, like pattern classification, speech synthesis, speech recognition, curve fitting, curve approximation etc. (ALIEV et al., 2012). Such way of modelling comprises a class of nonlinear regression or classification techniques inspired by theories about how the human brain works (KUHN and JOHNSON, 2013). Essentially, it contains many basic computing devices (neurons) that are connected to each other in a complex communication network, through which the model is able to carry out highly complex computations and adapt itself by way of changing its structures and parameters based on external signals or information (SHALEV-SCHWARTZ, 2014; LAHSASNA, 2009). These computing devices are arranged in layers and such information flow in the modeling is determined by a weighted connections system and, as such, they have the ability of learning from examples. In such scheme, two ways of learning exists: supervised and unsupervised. In the first, a set of examples comprising inputs and outputs is given, while the latter learns only with the input data, by which data is divided into groups based in common patterns (LAHSASNA, 2009).

Its basic structure is composed by neurons organized throughout the network as Input Layer -consisting of all input variables-, a single or multiple Hidden Layers (if multiple, the network is named Multilayer Perceptron or MLP) and an Output Layer, consisting of one or multiple response variables (Figure 1 illustrates an ANN with a single hidden layer).

Figure 1: A Neural Network with d input neurons, m output neurons and one hidden layer with n neurons



Source: BRUMMELHUIS and LUO, 2017

Broadly speaking, layers are usually fully connected by synapses, to which a weight w_i is attached indicating the effect of the corresponding neuron, and all data pass through it to "transmit" a signal to the output. In other words, each single neuron transforms an input vector $x = (x_1, \ldots ; x_n)$ into a single output $o\ (x)$ by getting a linear combination $\sum_i^n w_i\, x_i$ of the inputs, to which a constant bias w_0 is added, and, finally, a transformation f is applied. Then, to increase the modelling flexibility, hidden layers can be included. A general representation of a single hidden layer network with J hidden neurons is as follows:

$$o(x) = f.\left(w_0 + \sum_{j=1}^{J} w_j \, .\, f\left(w_{0j} + \sum_{i=1}^{n} w_{ij}x_i\right)\right) = f.\left(w_0 + \sum_{j=1}^{J} w_j \, .\, f\big(w_{0j} + \mathbf{w}_\mathrm{j}^\mathrm{T}\mathbf{x}\big)\right),$$

in which f usually takes the form of a smooth differentiable, bounded, nondecreasing and nonlinear function, such as the logistic one, defined by:

$$f(u) = \frac{1}{1 + e^{-au}}$$

In this function, a is an adaptable parameter. Other choices of f are hyperbolic tangent or linear functions and they are usually denoted as the activation functions. In the original idea, f was a threshold function applied to the simplest form of an ANN, called perceptron, consisting of an input layer with n inputs and one output layer with one neuron, which would transmit a signal if the combination of inputs $w_{0j} + \mathbf{w}_\mathrm{j}^\mathrm{T}\mathbf{x}$ was sufficiently strong to "beat" the threshold (GÜNTHER and FRITSCH, 2010; KUHN and JOHNSON, 2013; BRUMMELHUIS and LUO, 2017).

In a broader case, such as a multiple layered network, all hidden and output neurons calculate an output $f\Big(g\big(z_1, \ldots ; z_p\big)\Big) = f(g(\boldsymbol{z}))$ from the outputs of all preceding neurons $z_1, \ldots, z_p$, where f denotes the activation function (that could take the forms mentioned before) and g the integration function (GÜNTHER and FRITSCH, 2010; ZELL et al., 1995). In other words, an MLP scheme works the following way: as inputs in the input layer takes the variables of vector z, its outputs, after duly processed by activation and integration functions, serve as inputs for the neurons of the first Hidden Layer, whose outputs accordingly serve as inputs for the next Hidden Layer, and so on (a term commonly used to describe this process is "feedforward"). In the final Layer, for regression purposes, a linear combination connects hidden units of the final Hidden Layer to the output layer, while for classification purposes, it is used a sigmoidal function (e.g. logistic). In the latter case, however, a transformation is needed since the results are not "probability-like" and, hence, do not add up to 1 among different classes. In this manner, the so-called Softmax transformation is made (KUHN and JOHNSON, 2013; BRUMMELHUIS and LUO, 2017):

$$\pi_k = \frac{e^{\mathbf{w}_\mathrm{k}^\mathrm{T} o + \mathbf{w}_{\mathrm{k}_0}}}{\sum_{l=1}^{k} e^{\mathbf{w}_\mathrm{l}^\mathrm{T} o + \mathbf{w}_{\mathrm{l}_0}}}$$

where π_k, which is a function of input vector x and all initial, intermediary and final network weights, is interpreted as the model prediction, in terms of probability, for a given input to be in the kth class. This work makes use of both regression and classification methods but emphasizes the latter in order to predict possibilities of credit default among consumers.

Having this setting, the main purpose of an ANN is to “learn” the weights w_i of each neuron via an optimization procedure, or, the so-called learning functions. In a supervised learning algorithm, an actual output is compared to a predicted output, which is given by previous neurons and their individual parameters (or weights). As such, the network iteratively calculates all weights to the point in which the error (or the difference among predicted output and observed output) is minimum. A commonly used learning method for that purpose is the backpropagation algorithm, which modifies the weights to find a local minimum of the error function. A modified, but more efficient, version of it is the resilient backpropagation, which (i) is more suitable for supervised learning in MLP structures (referred also as “deep learning” due to its multiple layer structure (ADDO et al., 2018)), given its higher complexity, and (ii) its faster training rate (GÜNTHER and FRITSCH, 2010; ZELL et al., 1995; McCAFFREY, 2015). For the regression task, this work uses the standard backpropagation algorithm in a dual-layered structure, while for classification tasks, resilient backpropagation in an MLP is used.

Neural networks were extensively studied in the 1980s and early 1990s, but with mixed empirical success. With the advent of more advanced and efficient algorithms and computational power throughout 1990s and 2000s, the effectiveness of the usage of this methodology expanded considerably (SHALEV-SCHWARTZ, 2014). For credit risk evaluation purposes, ANN are considered good alternative to more traditional statistical models since there are studies concluding that its performance is superior in term of classification accuracy and capacity for faster decisions (LAHSASNA, 2009; GRACE and WILLIAMS, 2016). According to Grace and Williams (2016), the first model for credit risk evaluation was proposed by Odom and Sharda (1990), who obtained better results for neural networks compared with multivariate discriminant analysis. Other positive results were obtained by Jensen (1992), Desai et al. (1996), Eletter and Yaseen (2010), Ghatge et al. (2013) and Brummelhuis and Luo (2017), all of whom verified a superior performance of neural networks over traditional statistical modeling techniques for credit risk evaluation. On another vein, Vellido (1999) indicated that more than 75% of neural networks studies in a diverse range of businesses (from 1992 to 1998) were based on feedforward MLP trained by backpropagation. The same study reported that, among the papers researched, the two most recurrent contributions of its use were (i) “ANN yields better results than other techniques” and (ii) “ANN are shown to offer new insights into the application”. For credit risk scoring purposes, it has been the most researched technique, when used alone, over the years, considering 187 publications from 1992 to 2015 (LOUZADA et al., 2016). All in all, considering the amount of time dedicated to study this soft computing technique, it can be argued that, in general, neural networks potentially pose as a very successful method in learning and estimating default tendency of a borrower. However, this is true as long as the data is carefully analyzed, preprocessed and properly trained (GRACE and WILLIAMS, 2016).

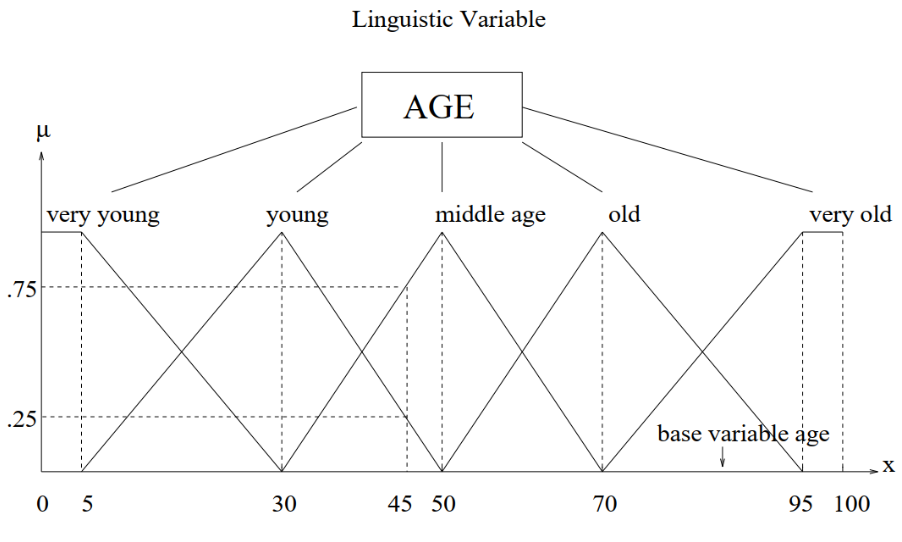
2.2.2 Fuzzy Logic

Fuzzy mathematics gained a broader reach when started being discussed by Lotfi Zadeh (1965), who was a Computing Science professor at the University of California and introduced the concept of fuzzy set theory. He argued that there was a need for differentiation of randomness and fuzziness, since, in his view at the time, the most usual way of dealing with imprecision was to employ the concepts and techniques of probability theory. By doing so, it was being accepted that imprecision could be equated with randomness, while the major source of imprecision in many decision processes could be, in fact, fuzziness (ZADEH and BELLMAN, 1970).

But a similar discussion started way earlier. Since when the principle that every proposition is either true or false has been declared, doubts were raised concerning its accuracy. Bojadziev & Bojadziev (2007) exemplifies it by noting that future events cannot be treated as facts or, in other words, as true or false. As such, this classical two-valued, or binary, logic isn’t enough to describe the future. In that sense, Lukasiewicz, in 1920, proposed a three-valued logic rationale to deal with that kind of unknown events, which possibly inspired the proposition of fuzzy set theory (BOJADZIEV & BOJADZIEV, 2007).

In essence, while classic logic follows a bivariate two-valued approach for a given statement (either being true of false or, in other words, 0 or 1, which are referred to as “crisp values”, in fuzzy terms), fuzzy logic accepts the propositions that consider values inside a 0 to 1 interval, creating room for “half true” and “half false” propositions, which are defined as membership functions. As a result, each object has a degree of belonging to a particular set, where 1 and 0 indicates full-membership and non-membership, respectively. This capability stems from the inability of classical logic to capture the vague language, common-sense reasoning and heuristic way of thinking that people use to solve problems every day. Fuzzy set theory is based on this concept, assuming that there are imprecise boundaries between two sets, in which the transition from non-membership to membership in a subset of a reference set is gradual, instead of abrupt, as it is in classical logic. Thus, it can be seen as a broad conceptual framework that actually encompasses the classical logic. This feature creates room for the usage of linguistic variables, as does human common reasoning, to handle imprecise information as a matter of degree of membership among different sets. For example, a fuzzy set can be represented by terms such as “very young”, “young”, “middle aged”, “old” and “very old”, as can be seen in Figure 2, where μ corresponds to the membership degree of each of the subsets (BOJADZIEV & BOJADZIEV, 2007; LAHSASNA, 2009).

Figure 2: Terms of the linguistic variable Age



Source: BOJADZIEV & BOJADZIEV, 2007

As to the way the fuzzy reasoning is created, it is relevant to note that, since fuzzy values are a matter of degree, this theory makes use of logical operators such as “minimum”, “maximum” and “complementary” instead of “and”, “or” and “not” as in Boolean Logic. The reasoning is ruled by the If-Then Rules, which are conditional statements that comprises the fuzzy logic reasoning, containing antecedent values (elements before “then”) and consequent values (elements after “then”). All such information is processed through a fuzzy logic reasoning, in which the set of If-Then rules are used to derive conclusions using linguistic variables (BOJADZIEV and BOJADZIEV, 2007). Figure 3 shows how an inference process occurs, where crisp input can be understood as a numerical value. In such scheme, two basic types of inference engines (to perform the If-Then fuzzy operations and reach its conclusions) can be used: (i) Mamdani and (ii) Takagi Sugeno (TS). The former gives a single fuzzy set as output and is generally described as (RIZA, 2015):

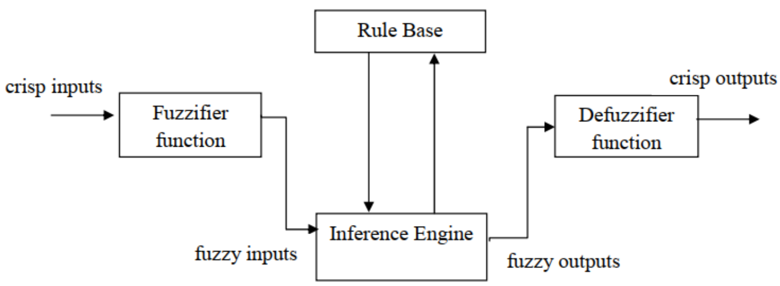
$$If\ x_n\ is\ A_n\ and\ x_{n+1}\ is\ A_{n+1}\ and\ \ldots\ then\ y\ is\ B,$$

while the latter, on the other hand, gives as a result a crisp function, such as:

$$If\ x_n\ is\ A_n\ and\ x_{n+1}\ is\ A_{n+1}\ and\ \ldots\ then\ y\ is\ y = f(x_n, x_{n+1}, \ldots)\ .$$

Both are extensively used and Mamdani is a class of modelling that allows the usage of natural language to describe input and output rules, being widely used for data analysis and data mining problems. In that sense, LAHSASNA (2009) remarks that Hoffmann (2002) compared the accuracy of two hybrid modelling techniques for credit scoring purposes - a genetic and a neuro fuzzy rule-based system approaches-, whereas the first used TS and the second used Mamdani engine. Its conclusion was that, despite presenting good accuracy for credit scoring purposes, one of the drawbacks of the model using TS engine was the lack of comprehensiveness compared to the model that used the Mamdani engine. So, in general, the Mamdani engine is more intuitive and well-suited to human input (LAHSASNA, 2009; KUMARASWAMY, 2018). In any case, the present work makes use of the Mamdani inference engine.

Figure 3: The basic components of a fuzzy system



Source: LAHSASNA, 2009

In soft computing, fuzzy logic plays an essential role, since it serves as a way to compute information with words. To design a system processor for handling knowledge represented in a linguistic or uncertain numerical form, a fuzzy system is needed. In general, it has been applied successfully to a broad array of fields, such as industrial, robotics, complex decision making and diagnosis, data compression and many other fields (ALIEV, 2012).

Fuzzy logic concepts can also be applied to finance and its diverse array of fields. In a credit risk analysis, it is fair to consider that business and consumer bankruptcy constitute an imprecise and ambiguous type of prediction. Failure processes can be affected by many internal and external factors that cannot be precisely defined and the mere allegation that a company or an individual is at risk of default (and even the statement “100% bankrupted”) must be considered imprecise. And, although many phenomena in finance and economics are fuzzy, they are still treated as if they were crisp (KOROL, 2012). This remark is in accordance with the fact that, as mentioned before, most of the techniques used for credit analysis in businesses are statistical methods, which generally are based in probability theory (such as Logistic Regression (KUHN and JOHNSON, 2013)). As such, they rely mainly on the assumption of randomness, which isn’t necessarily the major component of imprecision, according to Zadeh and Bellman (1970).

In that manner, although constituting a mere 2.5% of previous researches using solely fuzzy logic for credit scoring (LOUZADA et al., 2016) there are several researches that verified the usefulness of fuzzy logic to credit risk analysis. Louzada et al. (2016) founds points out that fuzzy performed generally better than other 7 methods, mainly when in the presence of imbalanced number of bad players (defaulters). ABDULRAHMAN et al. (2014) found, when developing a credit score model for a microfinance institution in Ghana, that a fuzzy modeling was more effective in evaluating credit applicants due to the involvement of human judgement. Grace and Williams (2016) remarks that Chen and Chiou (1999) proposed a fuzzy credit rating for addressing commercial credit applicants in Taiwan, a model that outperformed the then used methodology, which was a credit rating table. Daliyev (2015) developed a fuzzy expert system to address credit risk of applicants from Kazakhstan’s largest city (Almaty), getting results as good as 84% of accuracy, 10% higher than a regression comparable method. Finally, Tomas Korol (2012) proposed a fuzzy score model based on 500 polish consumers containing 10 demographical and financial variables which he divided in 3 fuzzy if-then rule blocks: (i) demographics (age, educational level, etc.), (ii) finance (monthly income, type of employment etc.), (iii) financial security (value of assets). He then proposed a fourth rule block, containing the forecasted outputs of all three previous rule blocks and, based on these evaluated inputs, the model forecasts the final credit scoring output, which was used straight to classify clients either as solvent or potentially default. Using two test samples (a balanced and an unbalanced), it was possible to achieve 88.75% and 91% of overall effectiveness. As a crucial step, the present work makes use of the method proposed by Korol to propose a fuzzy score, with the addition of a further analysis using a classification ANN having such score as input.

2.2.3 Combining ANN and FRBS

Recent research revealed drawbacks related to the use of standalone artificial intelligence techniques for handling real world problems. Limitations occur mainly due to the mass and vagueness of datasets, complexity of the real-world problems and uncertainty or unclear precision of available information. In that manner, there is a trend, based on a review covering the period 2005-2015, pointing new researches to the direction of hybrid modeling instead of traditional standalone soft computing methods, by way of integrating two or more different techniques (RAJAB and SHARMA, 2018).

In fact, Louzada et al. (2016) shows that, out of 187 papers analyzed, roughly 50% aimed the proposition of new methods of credit scoring, of which almost 20% accounted for hybrid models. Moreover, they showed that, hybridization modeling went from being used in 14% of cases to 18% of cases when comparing publications before 2006 with after 2012, respectively, which makes it, when considering all time spans (from 1992 to 2015), the second most adopted technique of a research, behind Neural Networks.

There are different ways of combining soft computing modelling techniques and the integration of Neural Networks with Fuzzy Logic is the most investigated one. Such hybridization is also amongst the most dominant for addressing various difficult research problems in business (LAHSASNA, 2009; RAJAB and SHARMA, 2018).

Two remarkable setbacks of the standalone usage of these two models explain the rising popularity of hybridization: while the lack of interpretability inherent to ANNs, due to its complexity, makes it frequently being called as a “black-box” technique, fuzzy systems generally carry a poor learning capability (ALIEV, 2012). The combination of the two techniques tries to solve such drawbacks and, in that sense, two common types of hybridization exists: Neural-Fuzzy and Fuzzy-Neural. Both of them combine the learning and connectionist skill of a neural network model with the interpretability capacity of a fuzzy system (ALIEV, 2012; LAHSASNA, 2009).

In general, this combination generates single algorithms that captures the singular positive feature of both methods, and several researches on credit risk analysis were developed. Yao et al. (2009) designed a six-layer network for commercial banks credit assessment (having one of such layers representing the fuzzy rule step) and such hybrid model proved to be better than a pure neural network in terms of error prediction. Constantinescu et al. (2010) proposed a neuro-fuzzy algorithm, with ease of use for credit scoring purposes and good performance on simulated data. Amorim et al. (2007) tested different neuro-fuzzy techniques against the traditional MLP technique in a large-scale credit risk assessment of a Brazilian financial institution, concluding that hybrid models, though presenting same accuracy, are an interesting alternative to traditional neural network modeling due to its interpretability and knowledge discovery.

The present work, however, makes use of both methodologies (ANN and fuzzy logic) independently and in a two-stage approach, by which: (i) a fuzzy system is developed manually (where the rules are based on (a) pure data characteristics and (b) information collected from an ANN regression made in a previous step) and (ii) a neural classification step is conducted.

3 METHODOLOGY

3.1 The Dataset

The dataset contains information from the client base of a Brazilian credit card operator, as of September 2015. There is a total of 9,458 observations with 19 variables, of which 11 comprise numerical information and 8 are categorical. Among these variables, 6 carry crucial information for credit evaluation purposes: 5 of them are score ratings from private consumer credit reporting agencies and one indicates whether the client have defaulted or not. At most, 237 observations contain missing information about 2 variables, at most. Tables 2 and 3 show general descriptive information.

Table 2: Descriptive statistics for numerical variables

Variable	Observations	Min.	1st Quartile	Median	Mean	3rd Quartile	Max.	GMD
Age	9,458	12	29	38	38.81	48	77	14.14
Monthly Income	9,458	0	1,700	3,400	4,237.00	5,700	22,200	3,659.00
Monthly Rent	9,458	0	400	1,000	1,233.00	1,800	7,000	1,152.00
Credit Line Requested	9,458	1,500	2,000	5,000	8,723.00	15,000	40,500	8,519.00
Credit Line Approved	9,458	1,000	1,200	4,200	7,893.00	14,000	38,300	8,441.00
Monthly Spend	**9,221**	0	287	740	1,013.00	1,361	23,074	1,014.00
Score Serasa	9,458	0	27,25	157	247.97	450	800	270.00
Score FICO Money	9,458	100	310	548	518.00	729	850	281.30
Score TransUnion	9,458	200	264	575	513.60	717	800	246.20
Score LexisNexis	9,458	200	315	500	467.70	607	690	181.10
Score Unit4	9,458	-4	296	390	395.30	486	800	176.30

Source: From the Author

Table 3: Distribution of categorical variables

Variable	Missing	Classes			
Lives in Capital	0	*Yes*	3,228	*No*	6,230
Region	0	*South*	1,342	*North*	493
		Center West	1,053	*Northeast*	1,587
		Southeast	4,983		
Residence Rent or Own	0	*Yes*	6,066	*No*	3,392
Residence Duration	0	*6 months or less*	1,045	*7-12 months*	1,099
		1-2 years	1,765	*3+ years*	5,549
Bank Account Duration	**14**	*6 months or less*	1,854	*7-12 months*	774
		1-2 years	1,917	*3+ years*	4,899
Facebook Gender	0	*Male*	4,824	*Female*	4,634
Facebook Profile Duration	0	*3 months or less*	148	*4-12 months*	408
	0	*1 year or more*	6,657	*3+ years*	2,245
Defaulted	0	*Yes*	2,047	*No*	7,411

Source: From the Author

3.2 Modeling Structure

The modeling structure comprises 2 stages, subdivided in 3 phases: The Fuzzy Stage, comprising (i) Preprocessing and Regression and (ii) Fuzzy Score Creation for a new proposed score and the Neural Stage, comprising the (iii) Classification step (Figure 4). All computations are made using R software, version x64 3.5.2, through different packages in an Intel Core i34010U CPU @ 1.7GHz 4 GB RAM machine.

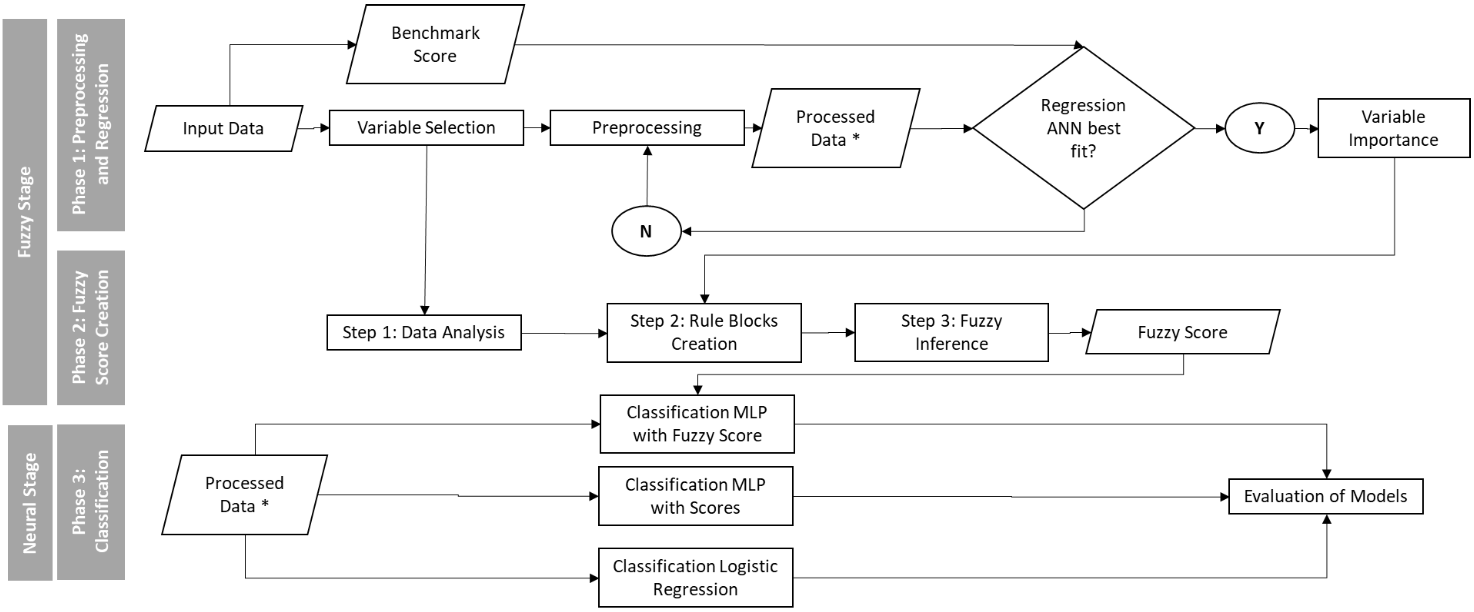
In the Preprocessing and Regression phase, among the 5 credit score ratings available for each observation, one is chosen to serve as training output for a dual layer neural network regression step, in which 100 models will be created, through the bootstrap sampling technique. Although considered a “black-box” model, as mentioned before, a recently developed tool to identify the relative importance of each input variable of the network is used here. In essence, the objective of this phase is to identify which variables would best serve to create, in this dataset context, the credit score rating selected as benchmark. This information is useful for the proposition of a new score in the second phase.

The Fuzzy Score Creation phase comprises a proposition of a new credit score rating based upon the idea of vagueness as one of the sources of imprecision, as discussed before, and is inspired by Korol's methodology for credit scoring (2012). For that purpose, a deep understanding of how the data is distributed among insolvent and solvent clients is made. Moreover, the variables are divided in 3 different blocks (demographics, finance and assets related), to which several If-Then rules are created. Each of such 3 blocks generates one crisp output for all 9,458 observations, which could be understood as intermediary credit score ratings. Next, these 3 intermediary scores serve as variables for a new fuzzy system, whose If-then rules are created with the knowledge acquired in phase one, regarding the importance of variables. With that, a Fuzzy Score is created.

In the final phase, 7 groups of neural networks for classification purposes are created, each of them trained through 100 bootstrapping resamples, to be tested in the "out-of-bag" set of samples (KUHN and JOHNSON, 2013). All of them will have as training output the variable "Defaulted" and as input, the selected variables in phase 1, preprocessed as such, plus one credit score rating (one modeling will be performed without any credit score rating). For one of the groups, the Fuzzy Score created in phase 2 is used, and, for 5 of each of the 6 remaining, the already available score ratings are used. The remaining one will be trained with no credit rating score. Plus, one group of 100 logistic regressions, with no score rating as input, is also made to serve as a statistical model benchmark for comparison with the neural network models.

Comparisons in terms of general performance and variability are made in the end, with the use of a multi criteria decision method (MCDM) to integrate the following results: (i) True Positive Rate (also known as Sensitivitiy), (ii) True Negative Rate (also known as Specificity), (iii) G-Means and (iv) Area Under Curve of Receiver Operating Characteristic (AUC of ROC). The results are also compared in terms of socio-demographic segmentation (Region, Age and Income) in the same terms.

Figure 4: Modeling Structure



Source: From the Author

3.3 Phase 1: Preprocessing and Regression (Fuzzy Stage)

3.3.1 Benchmark Score

In this step, each score rating already available is plotted in boxplots with regards to clients' default situation (Figure 5).

From the rationale that clients with lower chances of default tend to be better scored, it is possible to conclude that the credit scoring agencies that better evaluated this dataset are FICO, TransUnion and LexisNexis. By computing the significance of the difference among score ratings of defaulters and non-defaulters for each of the scores, a t-test resulted in significant differences among all 5 cases (very low p-value).

However, according to Coe (2002), a t-test isn't the proper tool for a considerably big dataset, as it is the case here and, hence, the Cohen's d Effect Size is used. This measurement, simply put, is a way of quantifying the size of the difference between two groups or, in other words, its magnitude (COE, 2002). Its calculation is as follows (COHEN, 1988):

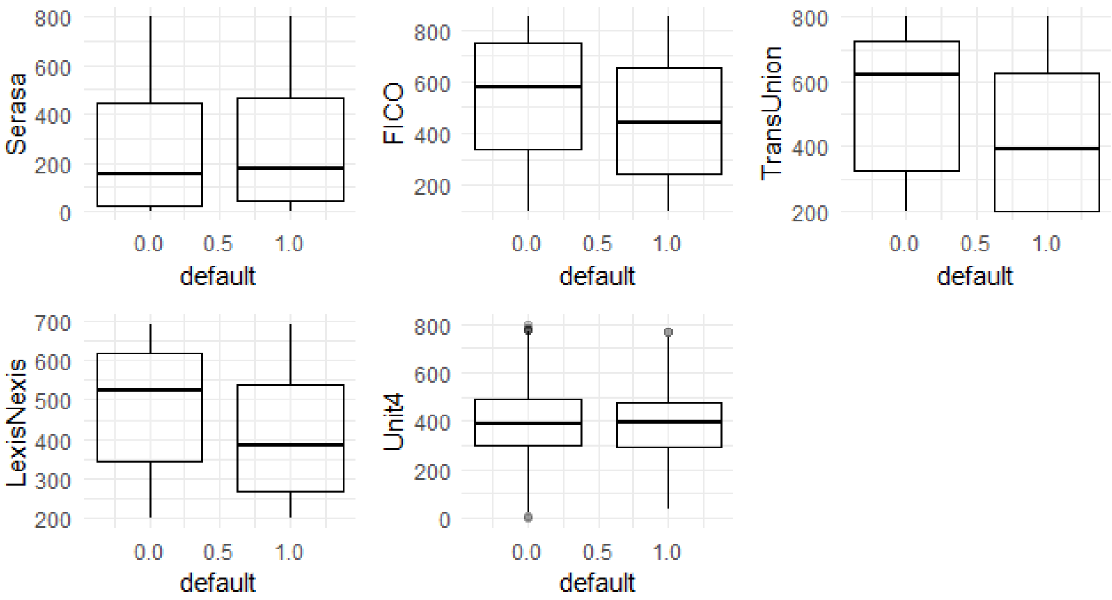
$$d = \frac{\mu_{default} - \mu_{nondefault}}{s_{pooled}},$$

where s is the pooled standard deviation:

$$s_{pooled} = \sqrt{\frac{(n_{default}-1)s^2_{default} + (n_{nondefault}-1)s^2_{nondefault}}{n_{default} + n_{nondefault}}},$$

and in which s^2 is the variance. The thresholds indicating the effect size magnitude are: |d|<0.2 as "negligible"; |d|<0.5 as "small"; |d|<0.8 as "medium" and, otherwise, "large" (COHEN, 1992). Table 4 lists the effect sizes for the differences among defaulter's and non-defaulted' ratings for each of the 5 scores. Hence, TransUnion is chosen for the Regression step, in Section 3.3.3. For this step, all calculations were done using the package effsize (TORCHIANO and TORCHIANO; 2018)

Figure 5: Distribution of scores through client's solvency



Source: From the Author

Table 4: d estimates for each score rating

Rating Method	*d* estimate	Magnitude
Score TransUnion	0.53262	medium
Score LexisNexis	0.50466	medium
Score FICO Money	0.36024	small
Score Unit4	0.07308	small
Score Serasa	-0.07548	negligible

Source: From the Author

3.3.2 Preprocessing

There are several reasons for the failure of predictive models. This statement is stronger when a highly dimensional and big dataset is considered, as is the case here. Four common reasons are: (i) inadequate pre-processing, (ii) inadequate model validation, (iii) unjustified extrapolation (using the model to an unusual data space) and (iv) over-fitting the model to the existing data. A model building process must consider all these aspects, specially nowadays, when higher data access and higher variety of tools to analyze it is easily available, making naïve model applications not related to the desired research objectives more recurrent. Clearly, the availability of large quantities of records is not a protection against an uninformed use of the data, and so, the credibility of model building had recently reduced (KUHN and JOHNSON, 2013).

The present work minds all such aspects (except for extrapolation, since the test set is obtained from the same dataset), however, it emphasizes the pre-processing step due to its thorough process, as it demands knowledge about the dataset analyzed and time to test different techniques. In any case, such stage is crucial, since failing to appropriately adjust the variables prior to modeling may produce models that have less-than-optimal predictive performance. It usually involves addition, deletion or transformation of training set data.

For neural networks modelling, in general, it is a binding condition that the practitioner (KUHN & JOHNSON, 2013; MONDAL, 2016):

i. adds, whenever the case, variables to the dataset through transformation of all dummies into numerical information,
ii. either (a) centers and scales and/or (b) normalizes all variables, a step that aims at improving numerical stability of some calculations, and
iii. removes (i) near-zero variance (due to its low information value) and (ii) highly correlated variables (since redundant predictors add more complexity and adds computational cost to the model building and can generate unstable models).

All such computations can be made via the caret R package (KUHN, 2015).

Following such rationale, firstly, the categorical variables are transformed into dummies. It is worth to note that this process creates “one minus” variables than the number of categories to avoid high correlation (AMUNATEGUI, 2014). Then, near-zero variance variables are removed. In this process, the dummy variable “Facebook Profile Duration_4-12 months” is removed, since it represents only 408 observations in a nearly 9.5 thousand space.

Next, one of each pair of highly correlated variables is removed, using as threshold 0.75 of correlation (Figure 6). In this manner, removed variables are “Credit Line Requested” (highly correlated with “Credit Line Approved”) and dummy variable “Facebook Profile Duration_3+years” (highly correlated with “Facebook Profile Duration_1+year”). Figure 6 shows a matrix correlation.

Other measurements are needed depending on the data intrinsic characteristics. Figure 7 shows how the variables (including the output for the Regression step - the

TransUnion score) are distributed. It is possible to realize that the data contains either or both skewness and outliers (Figure 8, left chart plots two variables that apparently have outliers). As for skewness, Korol (2012) argues that neural networks are not subject to drawbacks concerning the assumption of normal distribution for inputs like financial ratios. This vision is corroborated by L'Souza and Gupta (2006), who argued that MLPs don't make any assumptions regarding probability density functions, a statement that proved right during the model tuning in the Regression step, since traditional remedies for skewness (such as log transformation) did not improve model performance for the dataset considered in the present work. As for outliers, it has been proved they have the potential to deeply affect model performance (KHAMIS, 2005). And a data transformation that can handle this issue is the spatial sign (SEERNELS et al., 2006), which basically makes all variables of each observation to "reside" into the same space through the following calculation (KUHN and JOHNSON, 2013):

$$x_{ij} = \frac{x_{ij}}{\sum_{j=1}^{P} x_{ij}^2} ,$$

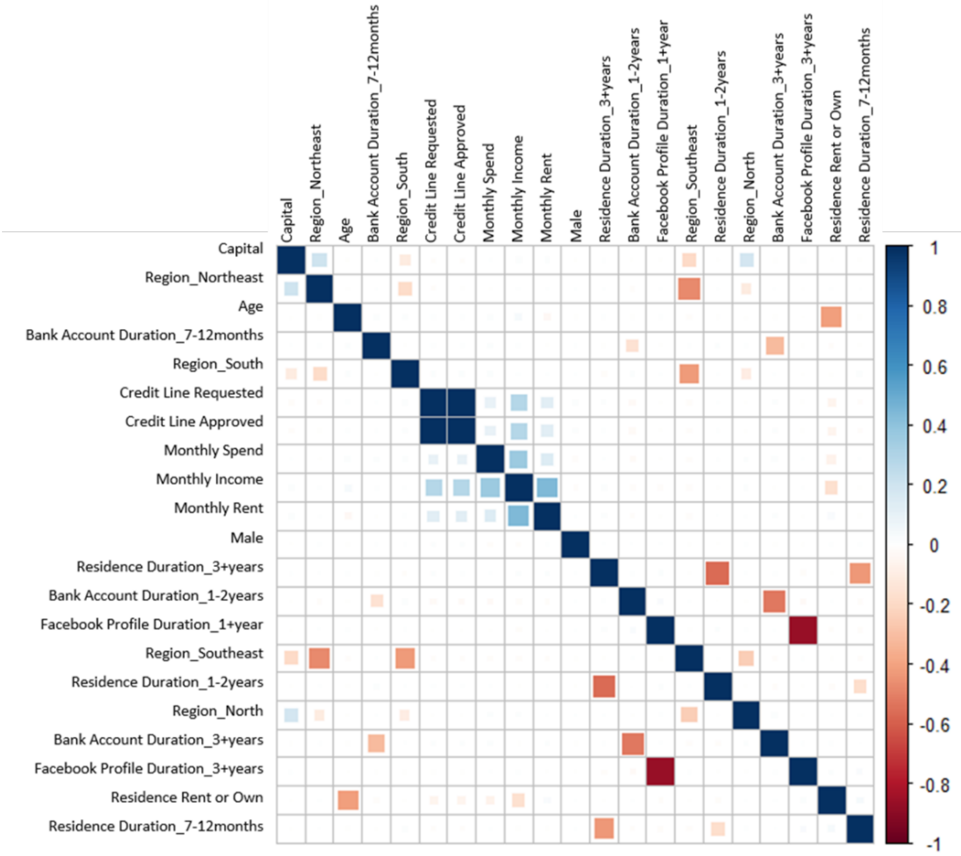
where x is the ith observation of the jth variable. The right chart of Figure 8 clearly demonstrates that almost all outliers are within the data mainstream space after transformation.

Then, all missing information from variables "Monthly spend" and "Bank Account Duration" are imputed through their respective medians, a fast and simple method, but less accurate than other methods (KUHN, 2015), although such inaccuracy would be harmless due to the size of the data in comparison with the amount of missing data. Finally, in addition to centering and scaling (inherent to spatial sign transformation) the data was also rescaled to range [0,1] due to better modeling results during the model tuning.

3.3.3 Regression

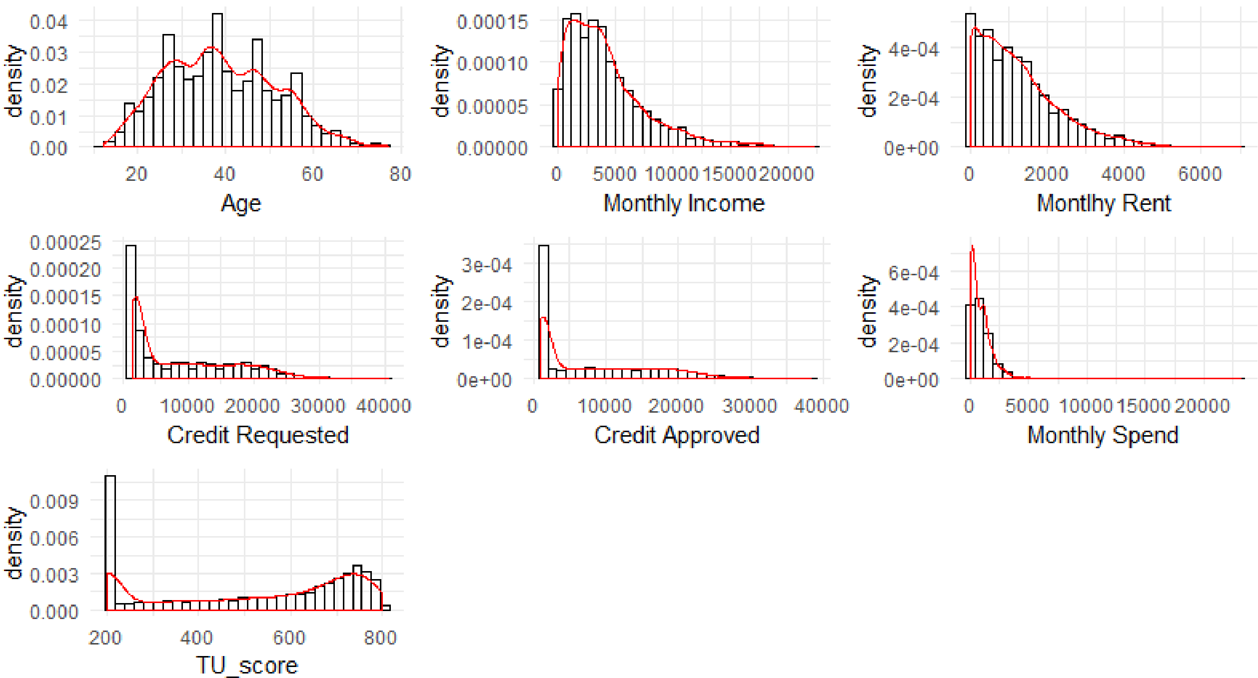
As mentioned earlier, the regression is aimed at constructing a model that, in light of the variables of the dataset, helps to understand which of them hypothetically are most related to the chosen credit score rating TransUnion. The R packages used for this step are RSNNS (BERGMEIR, 2012) and NeuralNetTools (BECK, 2015).

Figure 6: Correlation Matrix among variables



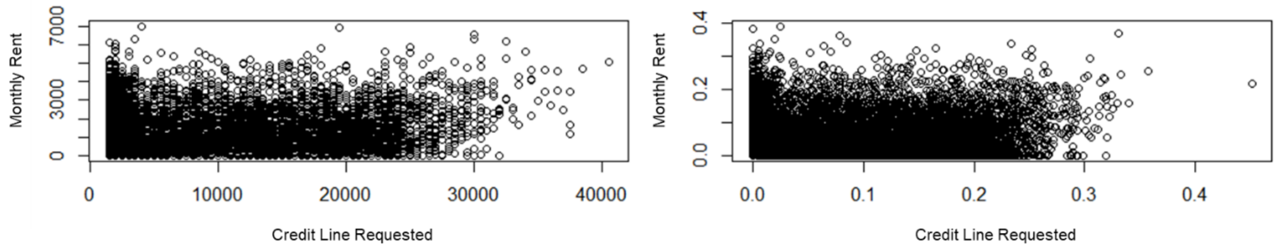
Source: From the Author

Figure 7: Histograms of numerical variables



Source: From the Author

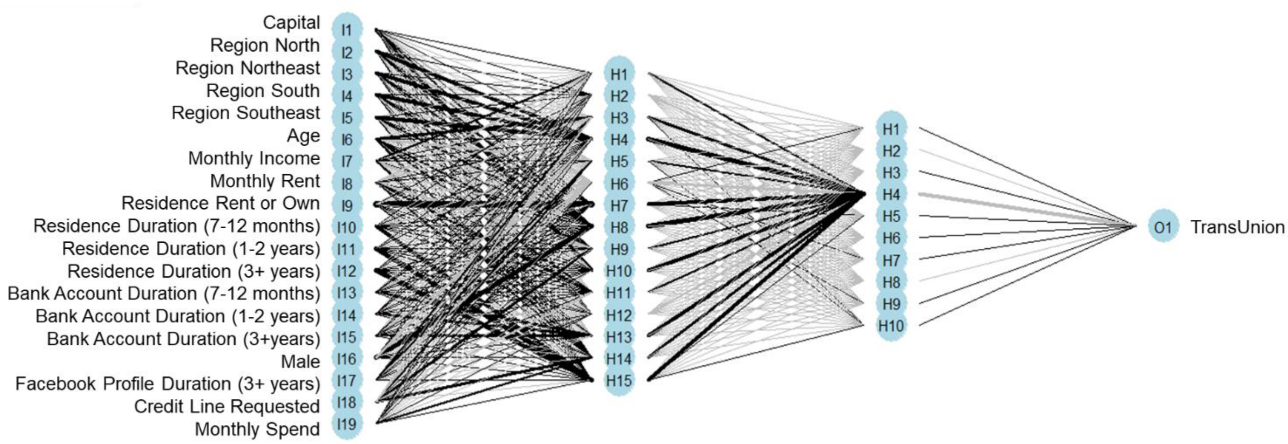
Figure 8: Left: Example of two variables before spatial sign transformation. Right: After transformation



Source: From the Author

After an exhaustive fine-tuning process, where lots of different settings of neural networks were tested, a final and best performing one was selected: a feedforward MLP with 2 hidden layers, the first with 15 neurons and the second with 10 hidden neurons. The activation and learning functions are Randomized Weights and Backpropagation, respectively (Figure 9, where the bold lines represent the strongest “connections”).

Figure 9: Network structure for Regression step

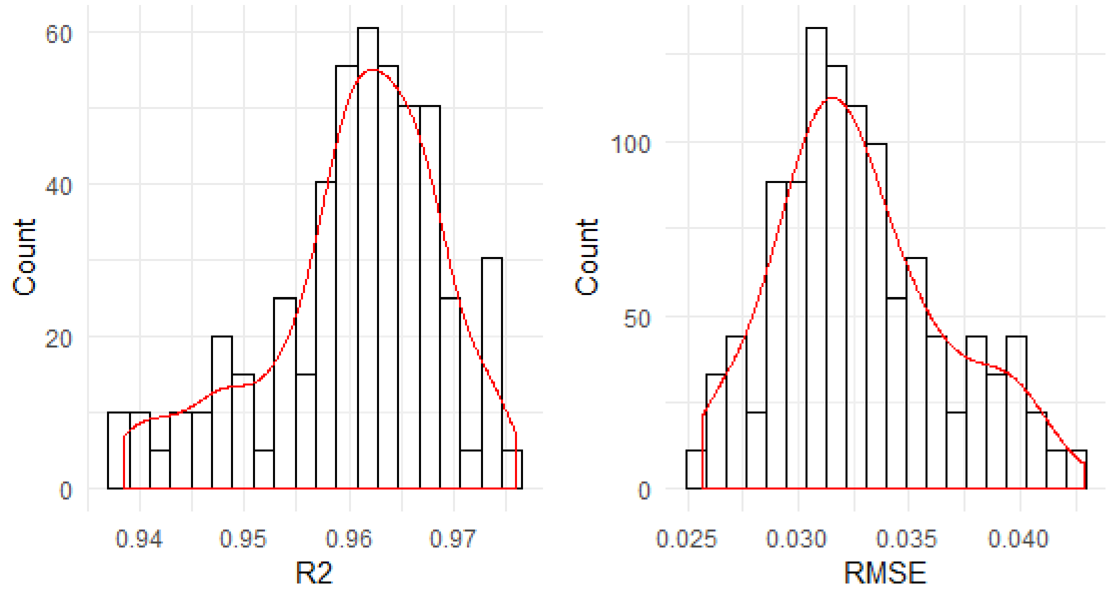


Source: From the Author

The chosen resampling method, both for the Regression and the Classification steps, is bootstrap, which implies in a random selection of samples, with replacement, for a subset to be trained that has the same size as the original dataset. Then, for testing, all samples not selected for the training set are selected (called “out-of-bag” samples, KUHN and JOHNSON, 2013). The reason for choosing bootstrap method is because its error rates, in general, tend to have less variance than the most utilized technique for such purpose, which is the k-fold cross validation (EFFRON and TIBSHIRANI, 1995). For the present work, since several models will be compared, error variance in the test set could be a big problem. Kuhn and Johnson (2013) also mention that this technique has a considerable amount of bias, but for it isn’t necessarily a problem here because every model is trained and tested under the same circumstances and with the same modeling, differing only in the inputs (the scores). For this resampling step, the package rsample was used (KUHN, 2019)

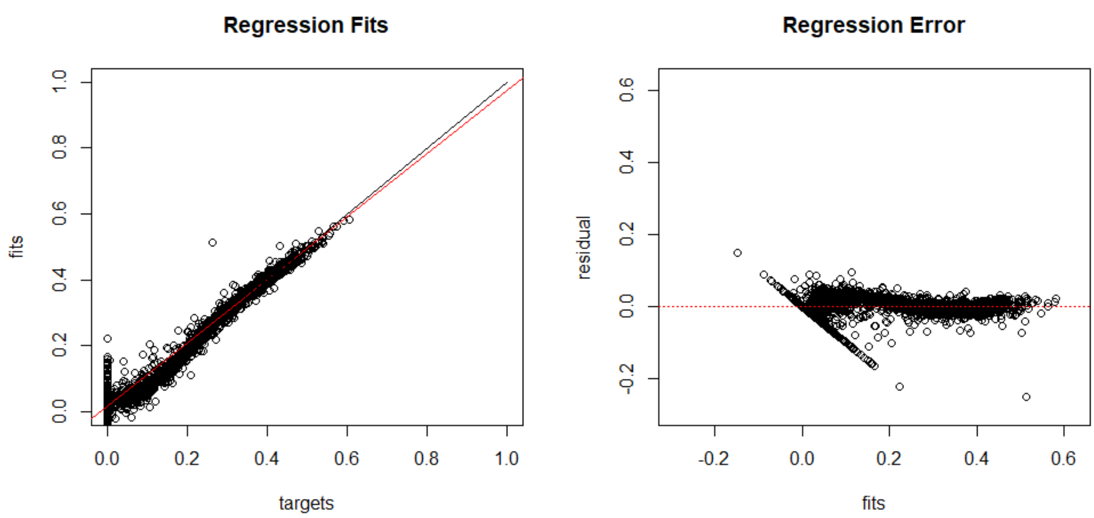
Figure 10 and 11 present the results for the Regression step, both in terms of R^2 and RMSE (Root Mean Standard Error), showing a high degree of fitness. But despite it seems that there is overfitting in this step, it does not necessarily represent a problem because the intention in this phase is to understand the relationship among variables and the score TransUnion.

Figure 10: Neural Network Regressions errors and R²



Source: From the Author

Figure 11: Best model fitted within the 100 bootstrapped resamples



Source: From the Author

Even though a neural network model lacks interpretability due to its high complexity and flexibility (ALIEV, 2012), it is argued that the weights that connect variables are analogous to parameter coefficients in a standard regression model and, hence, can describe relationships between variables. According to Beck (2018), essentially, these weights dictate the signal and intensity of the information flow throughout the network. Garson's (1991) method decomposes all weights connections, whereby input connections are multiplied by the output connections (resulting in $w_{i,j}$, the weight for variable i in the hidden layer j) and the importance for each variable is obtained as shown in (1). Olden (2002) proposed a different method, since it doesn't consider absolute values, but instead, raw values, as shown in (2):

$$Importance_i = \sum_j \frac{|w_{i,j}|}{|\sum_j w_{i,j}|} \text{ (1); } Importance_i = \sum_j w_{i,j} \text{ (2)}$$

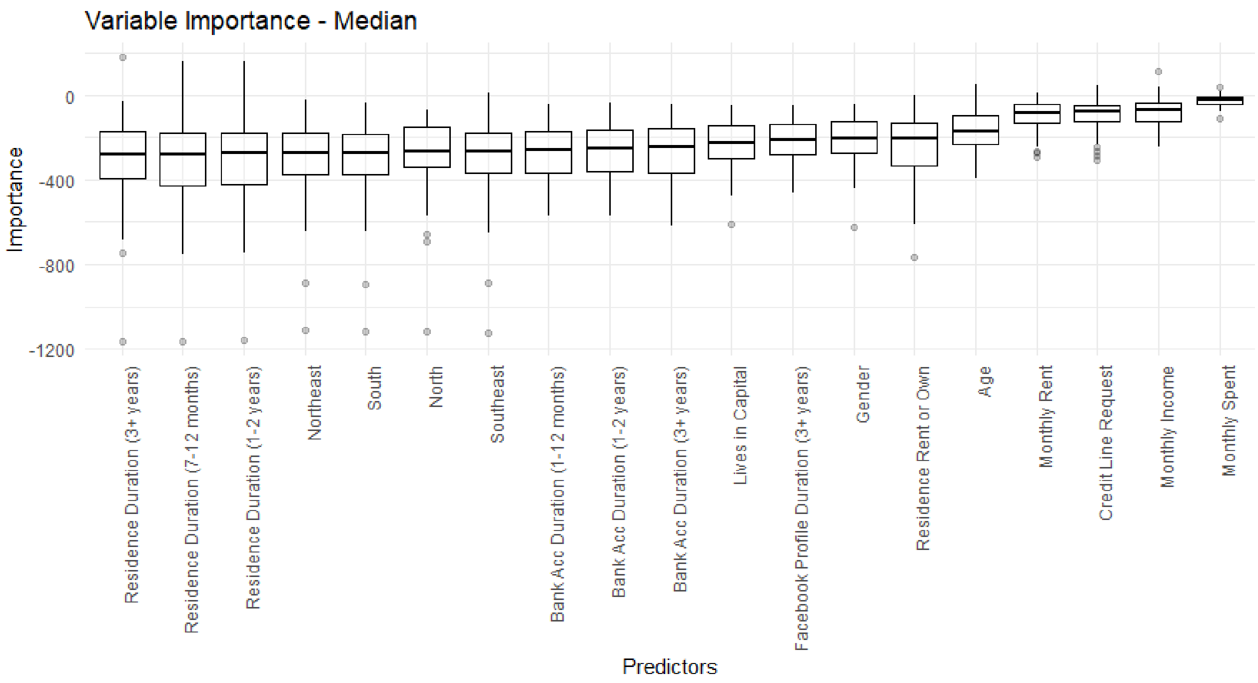
In a comparison between the latter and other 9 methods (including Garson's), it was concluded that Connection Weights provides the best methodology for accurately quantifying variable importance (OLDEN, 2004).

The importance of each variable was computed from each bootstrapped model and such information was ordered through median (Figure 12). In general, categorical variables were more important than numerical ones and a possible explanation for that is the binary characteristic of the dummy variables, which possibly contributes more intensely to the information flow throughout the network weights. The information gathered here is ordered by its absolute median and is used in Phase 2 for creation of the final rule block.

3.4 Phase 2: Fuzzy Score Creation (Fuzzy Stage)

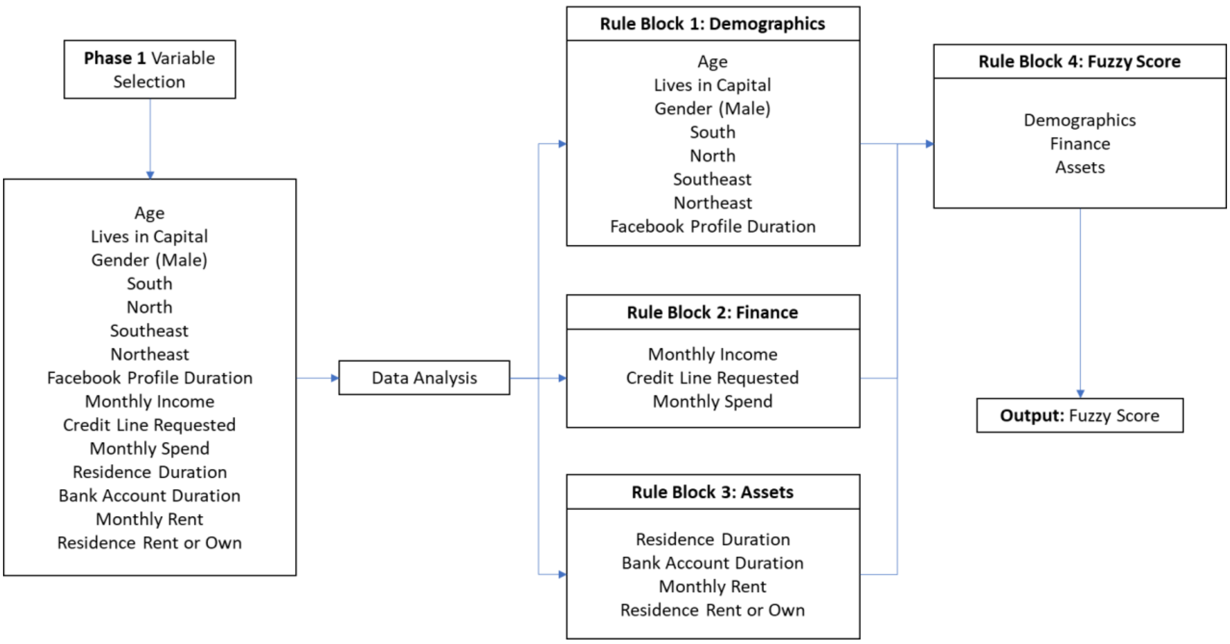
In this phase, data will be first analyzed in light of the possibility of default occurrence and, with such information, 3 groups of fuzzy rules are created in order to incorporate vagueness of usual language under the different variables (an illustrative example could be: “if an individual is older, he has a slightly higher score”). A final set of rules will be created, with the 3 blocks of crisp values obtained beforehand and whose rules are created considering the importance information obtained in phase 2 (Figure 13). The R package used for this phase is frbs, abbreviation for Fuzzy Rule Based Systems (RIZA et al, 2015). It also contains several algorithms for regression and classification based on a diverse range of techniques, including neural networks (neuro fuzzy modeling). These could be helpful with the creation of the rules by, for example, retrieving them after a regression having the TransUnion score as a variable of response. However, due to high error rates, this step wasn't conducted and, instead, rules were created manually.

Figure 12: Variable importance, ordered by median



Source: From the Author

Figure 13: Schematic view of Fuzzy Score creation

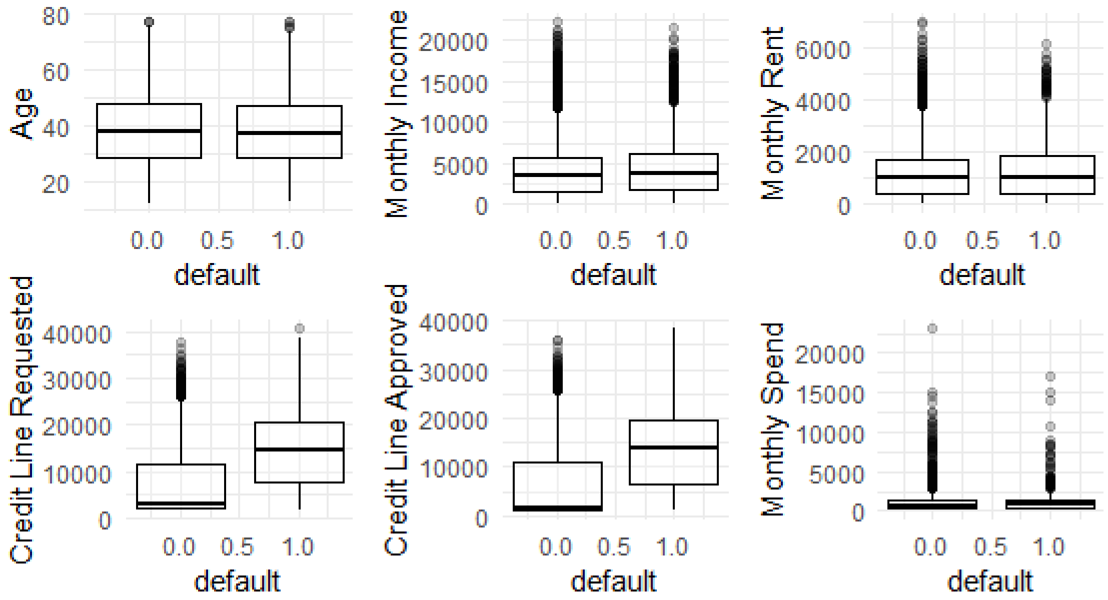


Source: From the Author

3.4.1 Data Analysis

This step served as a knowledge building process for the creation of the rules (Figure 14 and Table 7). Figure 14 shows boxplots for the numeric variables. The reasoning here is fairly simple: a higher median for defaulted clients will imply in a lower score (e.g.: the median of defaulted clients is higher than solvent ones in the variable “Credit Line Requested”, hence, a higher request for credit implies in a lower score). As for the Table 7, which shows the rationale for the categorical variables, the proportion among defaulted clients over solvent clients is analyzed and a reasoning is made (eg.: the proportion of defaulted male individuals is higher than within female individuals, thus, to be a woman implies in a higher score rating).

Figure 14: Numerical variables and default situation



Source: From the Author

The following reasonings were made for the proposition of the Demographic set of rule blocks:

- Higher Age is related to slightly lower chances of failure, then generates slightly higher scores,
- To live in capitals or not doesn’t affect chances of failure,
- To be male is related to higher chances of failure, then generates lower scores,
- To live in the Northeast Region doesn’t affect chances of failure,
- To live in the North Region is related to slightly higher chances of failure, then generates slightly lower scores,

- To live in the South Region is related to slightly higher chances of failure, then generates slightly lower scores,
- To live in the Southeast Region is related to lower chances of failure, then generates higher scores,
- To live in the Center-west Region is related to slightly lower chances of failure, then generates slightly higher scores,
- Higher Facebook Profile Durations are related to slightly higher chances of failure, then generates slightly lower scores.

Table 5: Proportions of bankrupted over solvent clients

Variable	Condition	Proportion Bankrupt/Solvent
Lives in Capital	Yes	0.27639
	No	0.27612
Region	South	0.31698
	Centerwest	0.27482
	North	0.31467
	Southeast	0.25548
	Northeast	0.29869
Gender	Male	0.29538
	Female	0.25685
Facebook Profile Duration	3 months or less	0.17460
	4-12 months	0.28707
	1 year or more	0.26728
	3+ years	0.30904
Residence Duration	6 months or less	0.73588
	7-12 months	0.61618
	1-2 years	0.64492
	3+ years	0.09751
Bank Acc Duration	6 months or less	0.32999
	7-12 months	0.24437
	1-2 years	0.25294
	3+ years	0.27115
Residence Rent or Own	Yes	0.24982
	No	0.29146

Source: From the Author

For the proposition of the Financial block, the reasoning for the rules were as follows:

- Higher Monthly Incomes are related to slightly higher chances of failure, then generates slightly lower scores,
- Higher Credit Line Requests are related to higher chances of failure, then generates lower scores,

- Higher Monthly Spend is related to slightly higher chances of failure, then generates slightly lower scores.

Finally, the rules for the Assets block were based on the following rationale:

- Higher Residence Duration is related to lower chances of failure, then generates higher scores,
- Higher Bank Account Duration is related to slightly lower chances of failure, then generates slightly higher scores,
- Higher Monthly Rent is related to slightly higher chances of failure, then generates slightly lower scores and
- To have a rented residence is related to higher chances of failure, then generates a lower score.

The Rule Block 4 reasoning was based on the general importance of the variables included in Blocks one, two and three, which was gathered through the average of all variables and compared among themselves (Table 6).

Table 6: Average Importance of the Rule Blocks

Rule Block	Relative Importance	Impact on Score
Assets	44.32%	High
Demographics	44.75%	High
Finance	10.94%	Low

Source: From the Author

3.4.2 Rules and Membership Functions

The rules were created considering the reasoning presented in the previous section. In that sense, 180 rules are created for the Rule Block 1, 27 for Rule Block 2 and 54 for Rule Block 3 (Annex 1). The outputs for all Rule blocks present 4 linguistic terms, which are “low”, “slightly low”, “slightly high” and “high”, ranging from 200 to 800. This setting is motivated by the distribution of the TransUnion score values (as shown in Figure 7, last histogram), barely showing “medium” values and ranging from 200 to 800. For Rule Block 1, variables Age and Facebook Profile Duration have each three linguistic terms as inputs (“low”, “medium” and “high”) and the remaining were binary variables with linguistic terms “yes” and “no”. As for Rule Block 2, all three variables present three linguistic variables (“low”, “medium” and “high”). Rule Block 3 variables presents the same three linguistic variables, except for Residence Rent or Own, which presents “yes” and “no”. Finally, Rule Block 4 had the same 3 linguistic terms as inputs for the three variables.

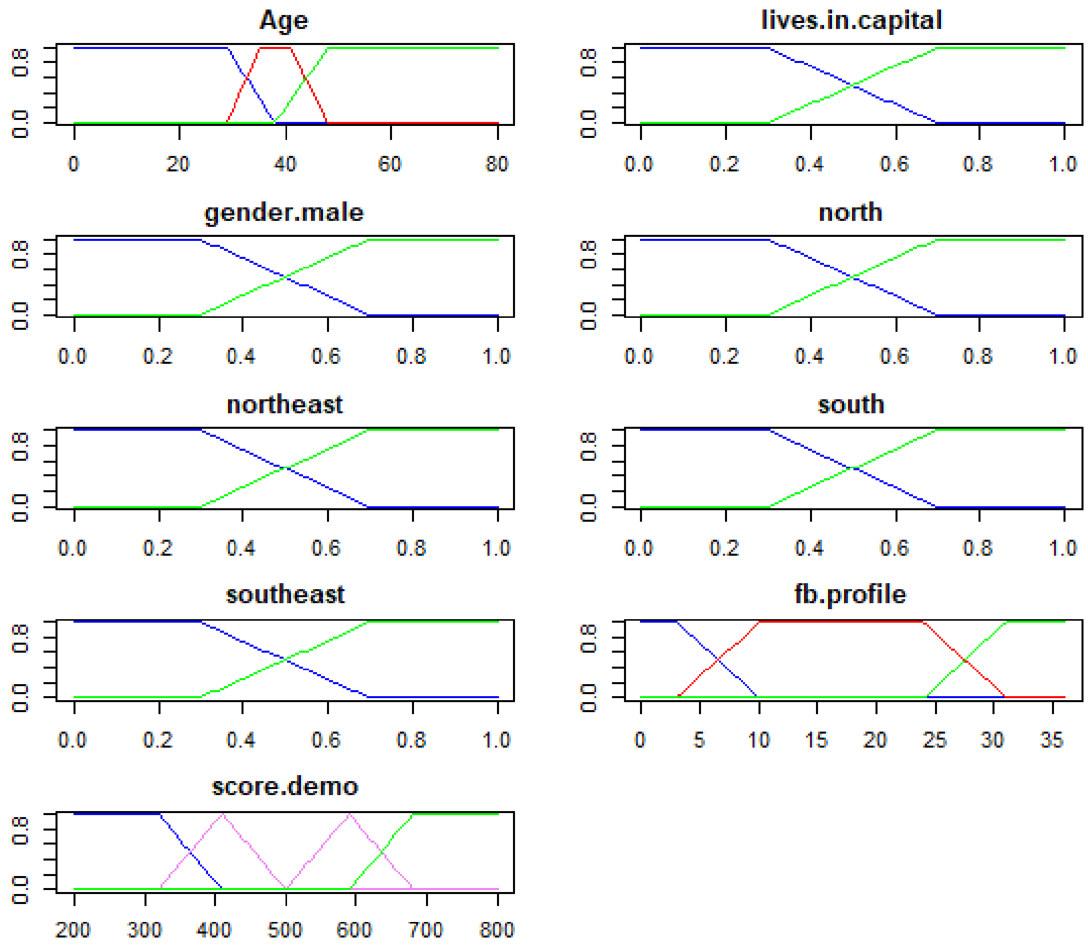
The membership functions of all inputs present trapezoidal shapes and for numerical variables, their format followed percentiles in the x axis (where the slope changes) such as 0%, 25%, 50% for the “low” set; 25%, 40%, 60% and 75% for the “medium” set and 50%, 75% and 100% for the “high” set. For the binary variables, “yes” and “no” inputs formed trapezoidal shapes, but either the membership result is 100% or 0% in these cases. For the categorical dummy variables, all input formats followed trapezoidal shapes and the intervals in the x axis for variables like “Residence Duration” and “Bank Account Duration” were as follows: 0, 6 and 10 months for the “low” set; 6, 10, 18 and 22 months for “medium” set and 18, 22 and 36 months for “high” set. For the variable “Facebook Profile Duration”, the trapezoidal shape followed values in the x axis such as: 0, 3 and 10 months for the “low” set; 3, 10, 24 and 31 months for “medium” set and 24, 31 and 36 for the “high” set. As for the outputs, all of them present trapezoidal and triangular shapes and, in percentiles of the range 200 to 800, they are: 0%, 20% and 35% for “low” set, 20%, 35%, 50% for the “slightly low” set, 35%, 50%, 65% for the “medium” set, 50%, 65% and 80% for the “slightly high” set and 65%, 80% and 100% for the “high” set. Figures 15-18 show the membership plots for the input variables and the outputs of each Rule Block. The motivation for the number of linguistic terms and formats of membership functions of all inputs were based upon the case study 17 (“A Client Financial Risk Tolerance Model”) of the Bojadziev & Bojadziev (2007) book, except for the medium sets, which were triangular instead of trapezoidal. The setting of the values in the x axis was based upon a model tuning of each Rule Block, by checking the best models in terms of how separated defaulted and solvent clients were (i.e. solvent clients with higher values as outputs and the opposite for the defaulted clients).

3.4.3 Fuzzy Inference System

As mentioned before, all outputs were obtained through the Mamdani inference engine. The defuzzification, a process in which linguistic outputs are transformed into crisp values, happened through weighted averages. In that sense, Rule Block 1, 2 and 3 generated, each, a crisp output for each of the observations. Then, these outputs were used as inputs for Rule Block 4, resulting in the final score (Fuzzy Score). Figure 19 shows how it is distributed and how defaulted and solvent clients are distributed.

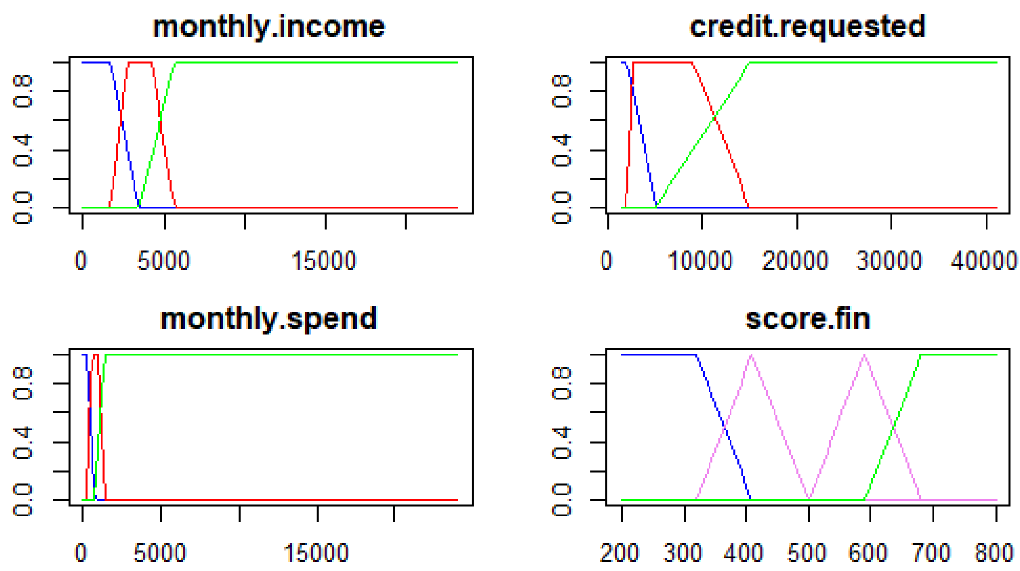
The proposed method seems to separate defaulted and solvent clients relatively well, having in mind its Cohen's d factor, which is 0.73.

Figure 15: Rule Block 1 Membership Functions (Demographic), where scores.demo is the intermediary score generated



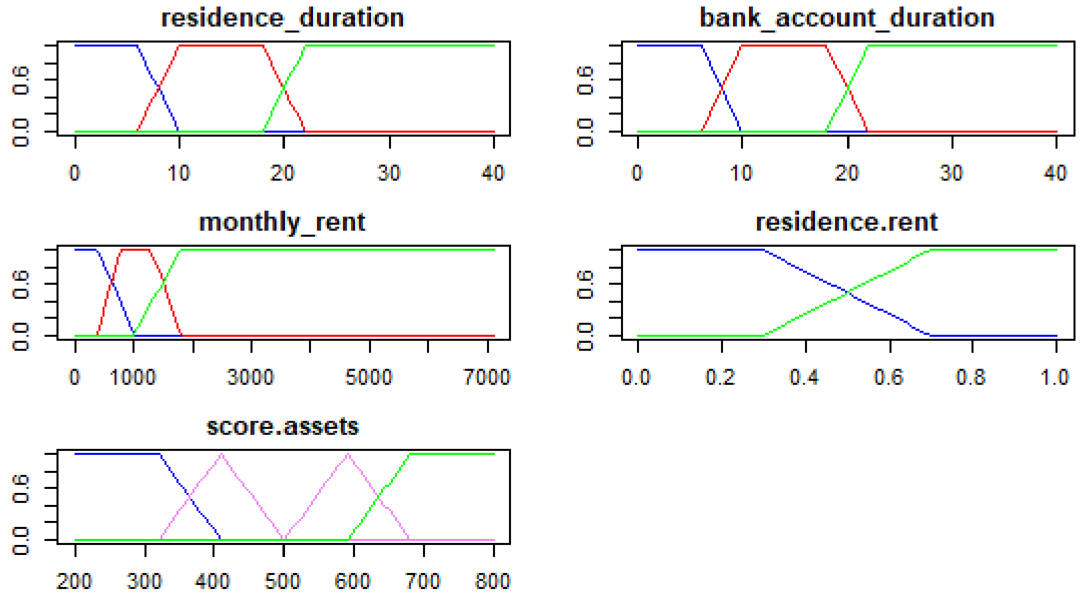
Source: From the Author

Figure 16: Rule Block 2 Membership Functions (Finance), where scores.fin is the intermediary score generated



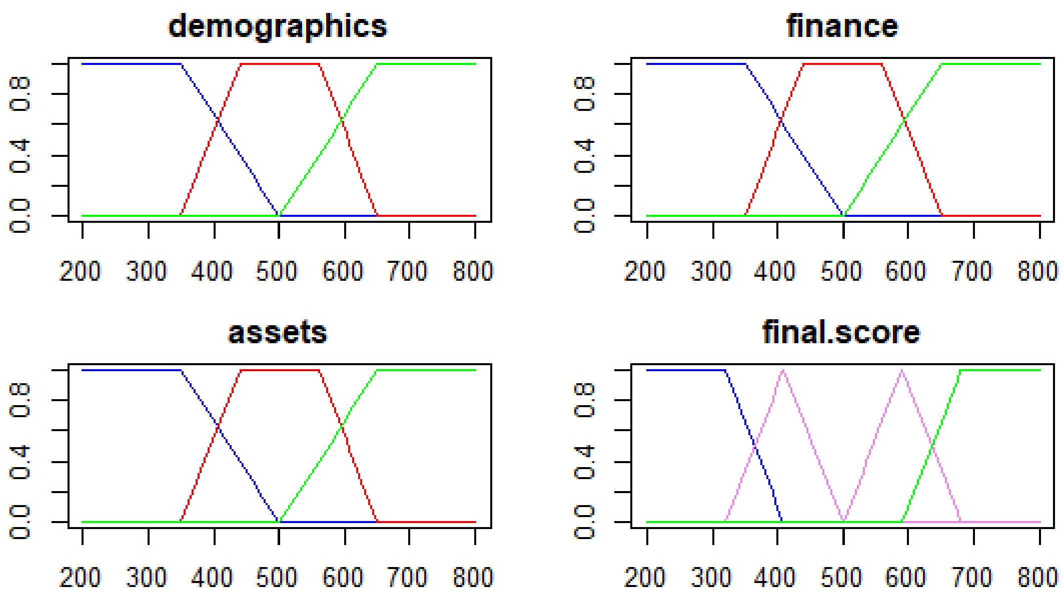
Source: From the Author

Figure 17: Rule Block 3 Membership Functions (Assets), where scores.assets is the intermediary score generated



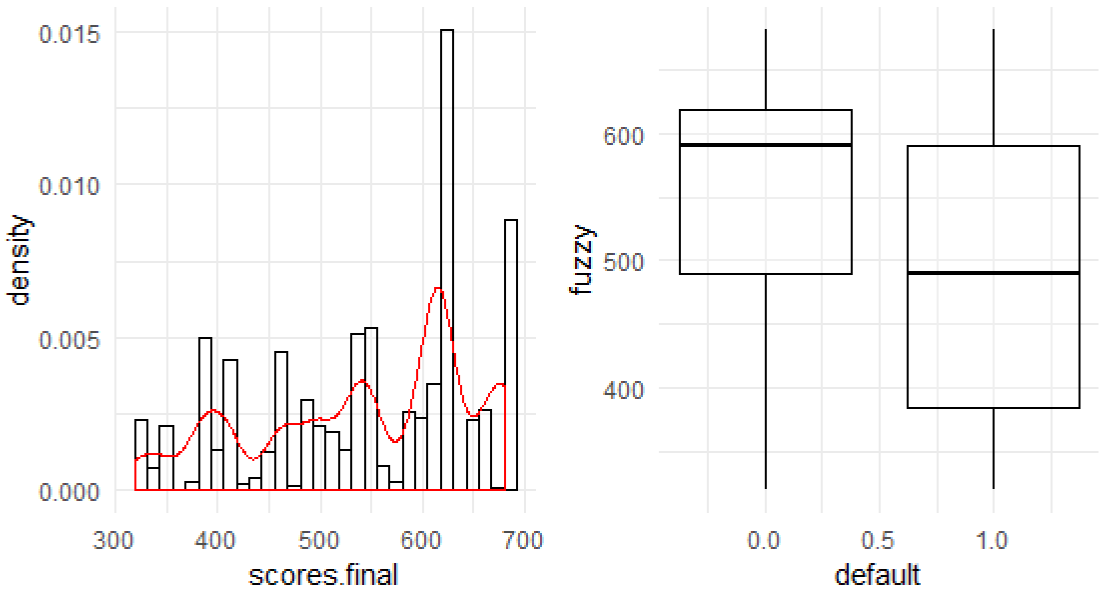
Source: From the Author

Figure 18: Rule Block 4 Membership Functions (Final), where final.score is the final Fuzzy score generated



Source: From the Author

Figure 19: Distribution and default situation for Fuzzy Score



Source: From the Author

3.5 Phase 3: Classification (Neural Stage)

3.5.1 Model Pre-processing and Tuning

In this final phase, each of the six score ratings (the five original ones and the Fuzzy Score) was added to all selected variables in the Fuzzy Stage in order to train and test an MLP network, having as output the default situation of each observation.

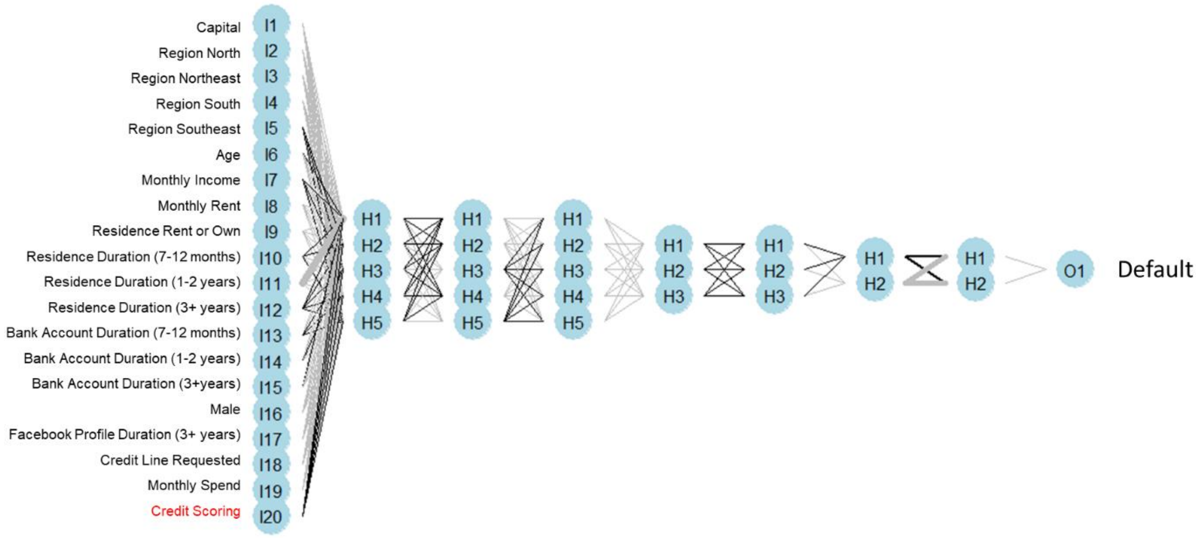
A set of models with no scores in its set of variable predictors was also created. The R packages used were RSNNS (BERGMEIR, 2012) for model creation, caret for pre-processing (KUHN, 2015) and rsample for resampling (KUHN, 2019).

All models had inputs preprocessed the same way as in Phase 1 (center, scale, normalization and spatial sign transformation, plus median impute for missing values). After tuning, where several settings were tested, one best performing one was selected: MLPs with 7 hidden layers using (i) random weights initialization and (ii) resilient backpropagation functions (Figure 21).

3.5.2 Treating Class Imbalance

In addition, it was used a package called DMwR (TORGO, 2015), to treat the imbalance problem. This step was deemed necessary because the data is relatively imbalanced, where 21.6% of samples are defaulted clients (as previously shown in Table 2). According to Kuhn & Johnson (2013), when modeling with discrete classes, as is the case here, the relative frequency of classes can have a significative impact on the effectiveness of the model, since the classification rules, in the end, get overwhelmed by the prevalent class and the rare examples are somewhat ignored, according to Lunardon & Menardi (2014). They also point out that, there are several remedies for treating imbalance, however, the most common are altering the class distribution to obtain a more balanced sample. The algorithm used here is called SMOTE, for Synthetic Minority Over-Sampling Technique, and resolves class imbalance by synthetizing new instances belonging to the minority classes (using the k-nearest neighbors' methodology) and/or by under-sampling the majority class, a method that has been proved better than simple oversampling with replacement or the standalone usage of under-sampling (CHAWLA et al., 2002). Kuhn and Johnson (2013) mention that, although models should be trained with a balanced dataset, it is important to test them with a test set consistent with the state of nature of the data and, hence, imbalance must be present so that honest estimates of future performance can be computed. Lunardon & Menardi (2014) simulated a model building and testing using a balanced training set and an imbalanced test set (5% of minority class presence), achieving good results. Similarly, in the present work, the SMOTE algorithm is applied upon the 100 training samples, created through stratified bootstrap sampling, meaning that the models are trained on a balanced dataset and tested within the out-of-bag samples in the presence of imbalance (Figure 20).

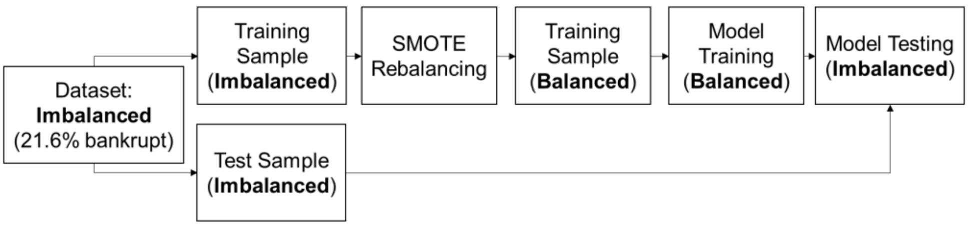
Figure 20: Classification model structure



Source: From the Author

With all that setting, 7 MLP models were ran 100 times in bootstrapped balanced resamples, one of them being generated with no scores.

Figure 21: Schematic representation of how models were trained and tested



Source: From the Author

3.5.3 Logistic Regression as a Benchmark

Finally, an additional set of 100 models was created, comprising logistic regression modeling, that followed the same pre-processing steps and variables for the MLPs generated. This is a very popular way of modeling, due to its simplicity and ability to make inferential statements about model's variables (KUHN & JOHNSON, 2013). Broadly speaking, this is a statistical learning technique that models the probability of an occurrence to belong to a particular category. It can be achieved through a function called logistic (S-shaped, values ranging between 0 and 1), whose parameters (or, the variable predictors) are estimated through a concept called Maximum Likelihood, which is a regression method where such parameters are calculated so the probability of occurrence of an event corresponds as closely as possible to observed event (JAMES et al., 2013).

All in all, 800 models were created: 700 being neural networks in 7 different settings and 100 logistic regressions (as explained above).

4 RESULTS AND DISCUSSION

4.1 Metrics Used

4.1.1 Sensitivity and Specificity

Common measurements of classification models' performance are Sensitivity and Specificity. Sensitivity measures the ratio of true positives, which, in this case, is the ratio of rightly predicting clients that failed. Specificity, on the other hand, measures the ratio of correctly classifying the true negative ones, or the solvent clients in the context of credit assessment. Both are important, however, under a credit card company perspective, managers would particularly wish to avoid incorrectly classifying an individual who will default, whereas incorrectly classifying an individual who will not default, though still to be avoided, would be less problematic (JAMES et al., 2013). This is because a defaulted client carries principal, interests and fees as costs, while a solvent "non-client" (in this condition as a consequence of being predicted as defaulted, hence, not being granted with a credit) only carries the cost of opportunity implied in the interests and fees. Even so, a 100% accuracy in terms of sensitivity, for example, could be easily attained by simply classifying all instances as defaulted clients, which would then result in a 0% accuracy in terms of specificity. In a hypothetical extreme situation, if a credit card company follows strictly a model like this, it would likely avoid having clients at all and the business would probably cease to exist. Hence, both metrics should be looked at.

For this calculation, a threshold probability (i.e. the limit value to classify an instance to one category or another) is needed and, depending on the model, the best performance would not necessarily be achieved with 50% as a cutoff (KUHN & JOHNSON, 2013). In that sense, the threshold for calculating them will be the one that provides the highest Youden's J index (Youden, 1950) for each of the models' results, which is simply the sum of true positive and negative rates minus one. This index is, according to Kuhn and Johnson (2013), a simple and an appropriate method for combining sensitivity and specificity into a single value. This measure is taken in order to put all models under the same conditions (their best), while setting a 50% threshold, for example, for all models would provide better results to those that classify with more accuracy in such cut-off.

4.1.2 Geometric Means and Area Under Curve of Receiver Operating Characteristic

As mentioned before, it is important to look at both sensitivity and specificity and one way of doing so is through the G-means, or geometric means (KUBAT & MATWIN, 1997), a very often used method by researches for evaluating classifiers on imbalanced datasets (NGUYEN et al., 2009). In this work, this step is done via computation of the square roots of the product among sensitivity and specificity, calculated as explained before.

Another commonly used and effective tool to measure model performance is ROC (for Receiver Operating Characteristics) and its AUC (Area Under Curve), which are independent of probability cutoffs and can produce meaningful contrasts between models. The ROC quickly shows, in a chart, different measurements of sensitivity (whose values are shown in the y axis) and specificity (values are shown in the x axis) for different levels of probability thresholds for classification in either one or another category. In that sense, the better the model, the more shifted the curve is towards the upper left corner of the plot and the higher is its AUC (KUHN and JOHNSON, 2013). This metric is also computed for the present work.

4.1.3 TOPSIS

All in all, 4 metrics are computed, along with training times for each model. Though, according to García et al. (2016), the usage of single performance evaluation measures may lead to unreliable conclusions and, sometimes, contradictory results regarding the best modeling process. Through the comparison of different algorithms over real-life bankruptcy and credit risk datasets, they showed that two Multiple Criteria Decision-Making (MCDM) techniques, – based upon G-means, AUC, Specificity and Sensitivity (among others) –, yielded a more reliable analysis to assess model performance. In summary, MCDM comprises several analytical tools to judge the pros and cons of a finite set of alternatives based on a finite set of criteria or attributes (YOON and HWANG, 1995).

In this regard, aiming at having a single and reliable result that reflects all the metrics above, a TOPSIS (for Technique for Order Preference by Similarity to Ideal Solution) result is also provided. The principle behind TOPSIS is to find the best alternative that simultaneously have the lowest distance to the positive ideal solution and maximum distance to the worst solution. The positive ideal alternative has the best levels for all criteria considered, while negative ideal has the worst levels (GARCÍA et al., 2016; WANKE et al., 2018).

The calculations for this step are made through the topsis R package and its setting are as follows: a number of alternatives, which are the 8 prediction models – 7 MLPs and 1 logistic regression - and a number of decision criteria, which are the performance measures (4 in total). The inputs are the medians, the criteria's weights are set to be equal and all criteria provides positive effects (i.e. the higher their value, the better). It is fair to note that, as previously discussed, although specificity is more critical than specificity under a credit card company perspective, criteria's weights were set to be equal because outperforming models in sensitivity have the potential to underperform in terms of specificity, and vice versa (KUNH & JOHNSON, 2013). The same is done for ROC and Geometric Means, since it is not obvious which one is a better measure of performance.

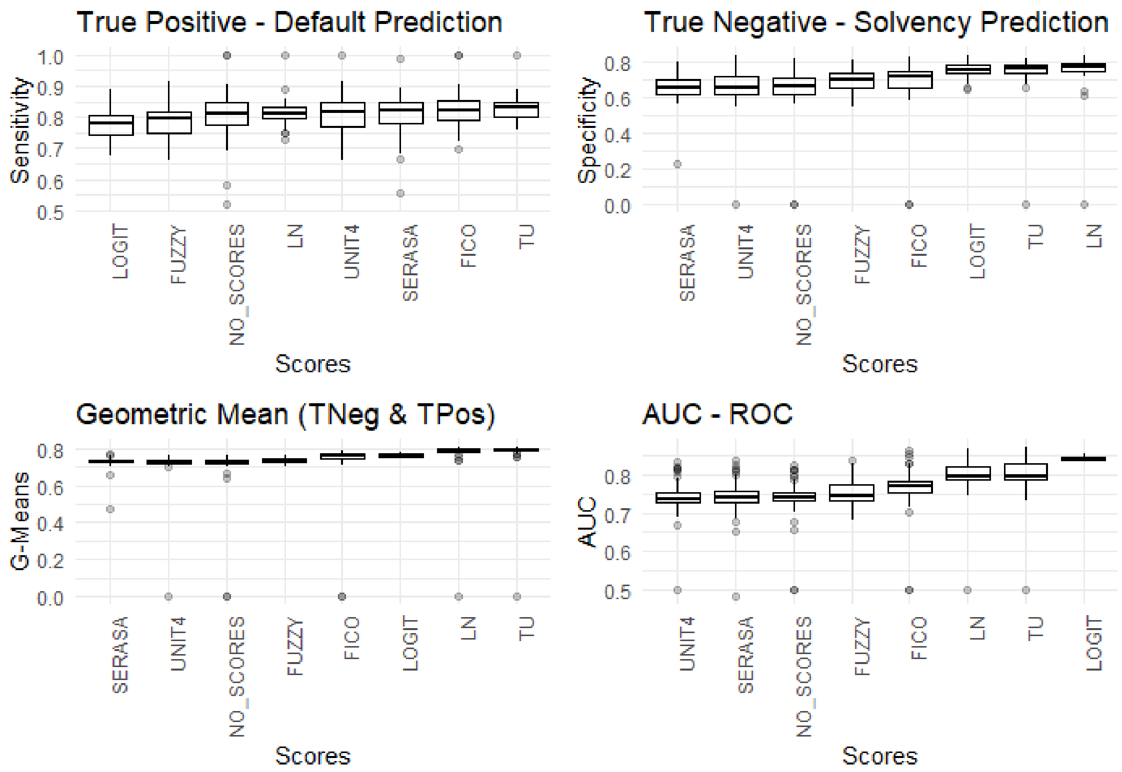
Finally, it's fair to understand how consistent the results are, among the different test sets. For that, the standard deviation of the models will also be computed into a TOPSIS ranking, in which targets criteria's weights are the same and all will provide positive effects (in the sense that models with results presenting high variance generate higher TOPSIS scores). This discussion is important in predictive modeling due to the trade-off existent between (i) the bias of a model (or how far the functional format of a model is to the true relationship between predictors and response) and (ii) its variance. In general, more complex methods (i.e. those that greatly adapts to a training data) can present high variance, due to overfitting, but low bias. The opposite is true when an underfitted method is unable to model the true relation between predictors and response variable (KUHN & JONHSON, 2013; JAMES et al., 2013).

4.2 Models Evaluation and Discussion

4.2.1 General Results (all dataset)

As expected, an evaluation of the results considering the measurements separately can be inconclusive or misleading (Figure 22).

Figure 22: General Results for MLPs (named with the scores used as predictor) and Logistic Models (Logit)



Source: From the Author

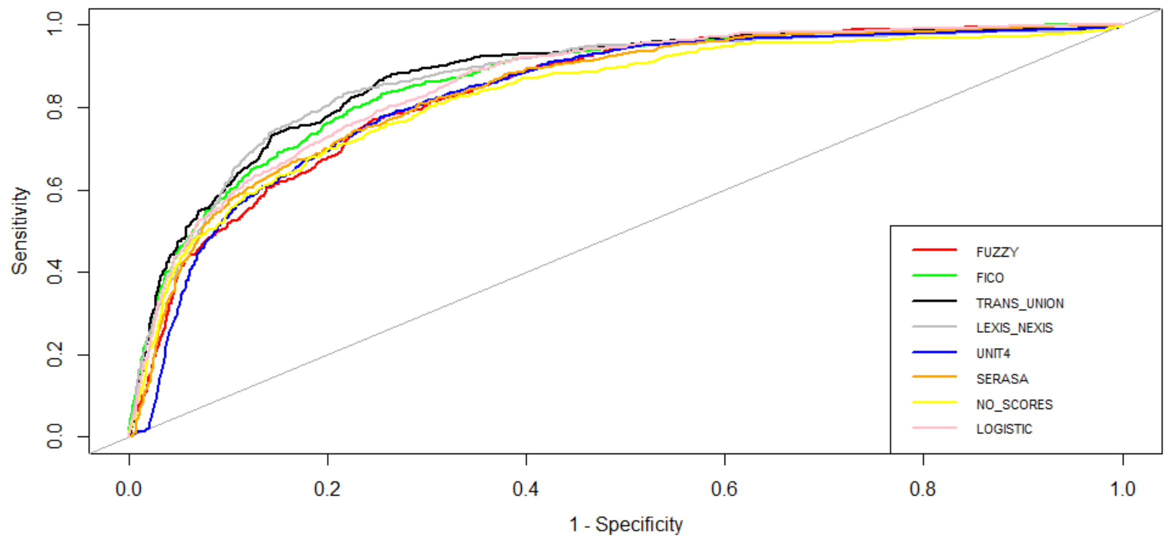
Table 7: Average Computing times for each model (minutes and seconds)

Model	Logistic Regression	TransUnion	Fuzzy	Lexis Nexis	Serasa	No Scores
Time	10.0 S	2 M 3.9 S	2M 12.9 S	2M 25.2 S	2M 43.0 S	2M 43.5 S

Source: From the Author

While an MLP containing the Trans Union score as a predictor is the best in terms of G-means and Sensitivity metrics, Lexis Nexis MLP and Logistic model without scores perform better in Specificity and AUC metrics, respectively. Trans Union’s and Lexis Nexis’ good performances can be seen in Figure 23, which shows the best ROC for each model. Their curves slightly detach from the other models towards the upper left corner of the plot, a result consistent with their performance in Sensitivity and Specificity metrics. As for the proposed Fuzzy Score, it shows MLPs with intermediary to low performances in relation to the others, being the 7th in terms of Sensitivity (better than the Logistic model), 5th in the Sensitivity, AUC and G-means metrics. In terms of computing time, Logistic Regression clearly stands out as the quickest one.

Figure 23: ROC curves for best model (in terms of AUC) for each method



Source: From the Author

As for how consistent the results are, Table 7 shows the standard deviation of the models. If a comparison is made, Fuzzy MPL and Logistic model results seems to vary less than the others in most metrics.

Table 8: Standard Deviations of models' results

Models	Sensitivity	Specificity	AUC	G-means
FICO	5.0%	11.4%	4.8%	10.8%
FUZZY	5.5%	6.0%	3.2%	1.4%
LEXIS_NEXIS	3.4%	8.4%	4.1%	8.0%
LOGIT	4.0%	3.7%	0.6%	0.7%
NO_SCORES	6.6%	11.0%	4.3%	10.4%
SERASA	5.9%	7.2%	4.0%	3.0%
TRANS_UNION	3.5%	8.3%	4.2%	8.0%
UNIT4	6.3%	9.5%	3.6%	7.4%

Source: From the Author

It is relatively easy to realize which models perform better and worst, either in terms of performance or variability, but it's hard to order them in a proper rank to make a relative comparison. For this reason, two TOPSIS are created, as can be seen in Table 8.

In terms of performance, MLPs trained with the scores Trans Union and Lexis Nexis clearly performed better, which demonstrates their effectiveness, at least in a similar setting, in predictive modeling. Moreover, the good performance of the Logistic method, a simpler model which used the same variable inputs (with no scores) and the same preprocessing steps as all MLPs, shows that a complex modeling technique like neural networks do not necessarily generates significantly better performances. The MLP containing the proposed Fuzzy score as input had, as expected from Figure 22, intermediary results, but still better than MLPs containing other two scores: UNIT4 and Serasa. Finally, it is notorious the increment in model performance in MPLs having a score as input for the training data. Except for the UNIT4, the scores used in this work proved to be more informative than noisy or irrelevant for predictive modeling purposes.

Table 9: TOPSIS results based on median of results and standard deviations

Performance		Variability	
Model	**TOPSIS**	**Model**	**TOPSIS**
TRANS_UNION	0.77	NO_SCORES	0.92
LEXIS_NEXIS	0.75	FICO	0.87
LOGIT	0.73	UNIT4	0.71
FICO	0.49	TRANS_UNION	0.64
FUZZY	0.27	LEXIS_NEXIS	0.64
SERASA	0.19	SERASA	0.48
UNIT4	0.17	FUZZY	0.36
NO_SCORES	0.17	LOGIT	0.05

Source: From the Author

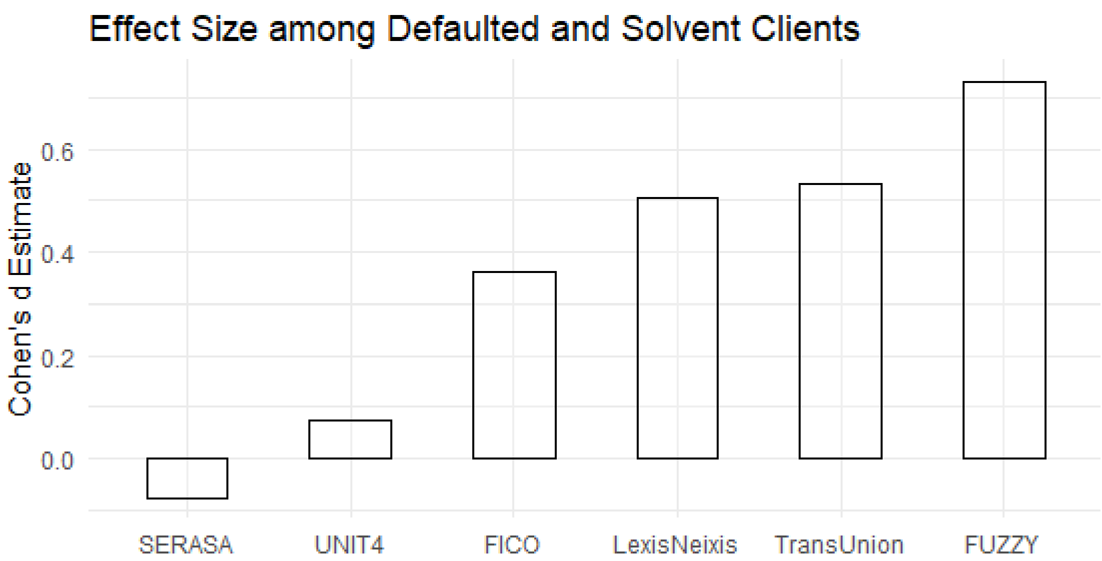
The above results followed somewhat the Cohen's d estimate showed as to the scores' informative power to the training set (Figure 24). In the lower batch, UNIT4 and Serasa didn't seem to separate well defaulted and solvent clients. In an intermediary level, there is the FICO score. TransUnion and Lexis Nexis showed a relatively high Cohen's d estimate, which is reflected in the MLPs having them as input. The Fuzzy score is an exception, maybe due to a more even distribution when compared, for example, with the Trans Union's one (Figures 19 and 7).

As for the variability, the Logistic model showed the lowest value, implying that it can be considered more stable than the others for the current setting, along with the MPL having a Fuzzy score input. Trans Union and Lexis Nexis' MLPs showed relative higher variability and an MPL without scores or with FICO had the worst results in this regard.

From all the above, some remarks can be made are. First, Logistic model is a simple method that presented relatively satisfying results, while being the most stable compared to the neural networks. Possibly, it results from the fact that, in general, neural networks have the tendency to overfit the relationship between predictors and response due to its large number of regression coefficients, which are, in this case, the neurons (KUHN & JOHNSON, 2013). While the Logistic model presented 19 coefficients, the MPLs had 44 to 45 coefficients. And, as mentioned before, there is a trade-off between bias and variance, whereas, in this context, while Trans Union and Lexis Nexis as inputs for MLPs presented good performance results (they fitted reasonably well, thus, lower bias), they showed relatively higher instability (high variance).

Second, in general, the scores were informative inputs for the model building, since, not only they were responsible for increments in the result, but most of them helped the models to become more stable. Considering this, the UNIT4 as input for modeling an MLP can nearly be considered as an irrelevant variable (due to its poor informative power for prediction purposes and to the fact that it barely improved stability over a model with no scores), however not the same can be said about Serasa, since, although it showed a low predictive power as an input, at least it helped the MLP to achieve a lower variability.

Figure 24: Cohen's d Estimate for the Scores (from separation of defaulted and solvent clients)



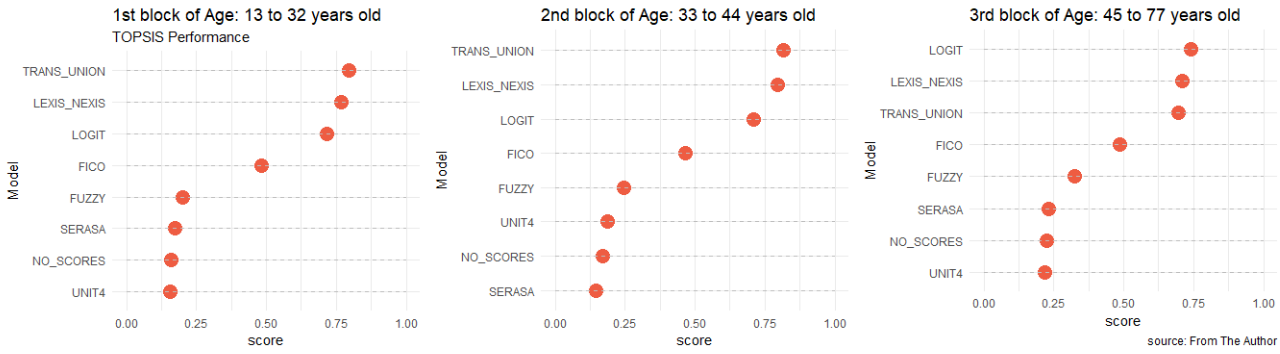
Source: From the Author

Finally, the proposed Fuzzy score proved also to be an effective input for the MLP. It improved the MLP performance (not only over the model without scores, but also when compared to other two scores) while also bringing more stability. Thus, it acted in both positive ways with regards to the bias-variance trade-off: it helped the model to achieve a lower bias, considering that has one more variable (an informative score) and, at the same time, it trained a model that proved to be more consistent than the one without scores. Hence, the Fuzzy score can be considered a relevant and informative input variable for the modeling set discussed here.

4.2.2 Results Segregated by Selected Demographic Variables

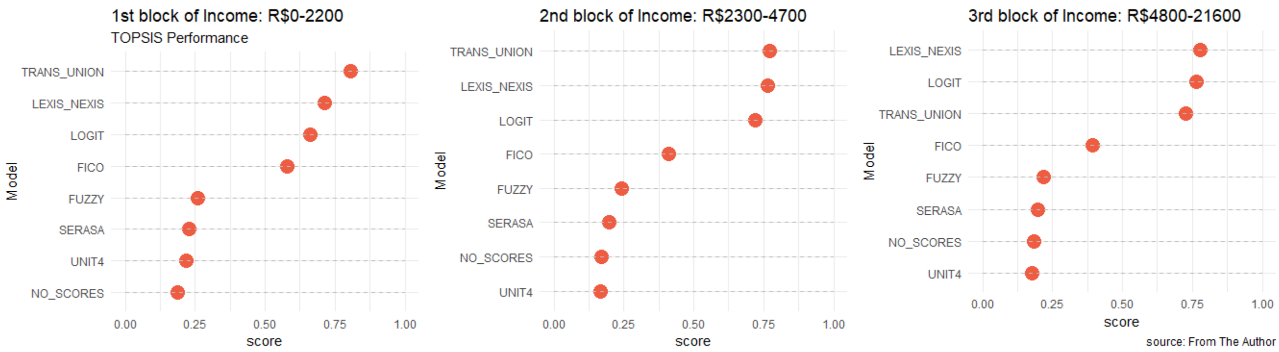
A further breakdown of the analysis is made, where the results are divided among different demographic segments in terms of Age, Income, Gender and Region. For the numeric variables Age and Income, the samples were segregated in three parts: the first block contains samples up until the 33rd quantile, the second block ranges from the 33rd up until the 66th quantile and the third block ranges from the 66th block until the highest value. Figures 25 to 28 and Appendix B show the final TOPSIS results for performance.

Figure 25: Results for TOPSIS Performance Segregated by Age



Source: From the Author

Figure 26: Results for TOPSIS Performance Segregated by Income

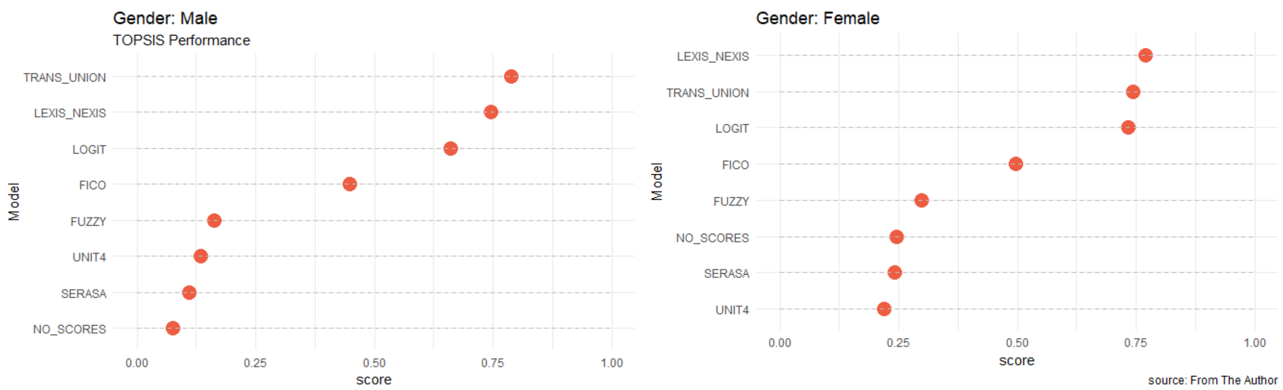


Source: From the Author

In general, the segregated data had similar results to the complete set of data. Except for minor differences, all subsets presented three blocks of results, where MPLs with Trans Union and Lexis Nexis had the best results, along with the Logistic model.

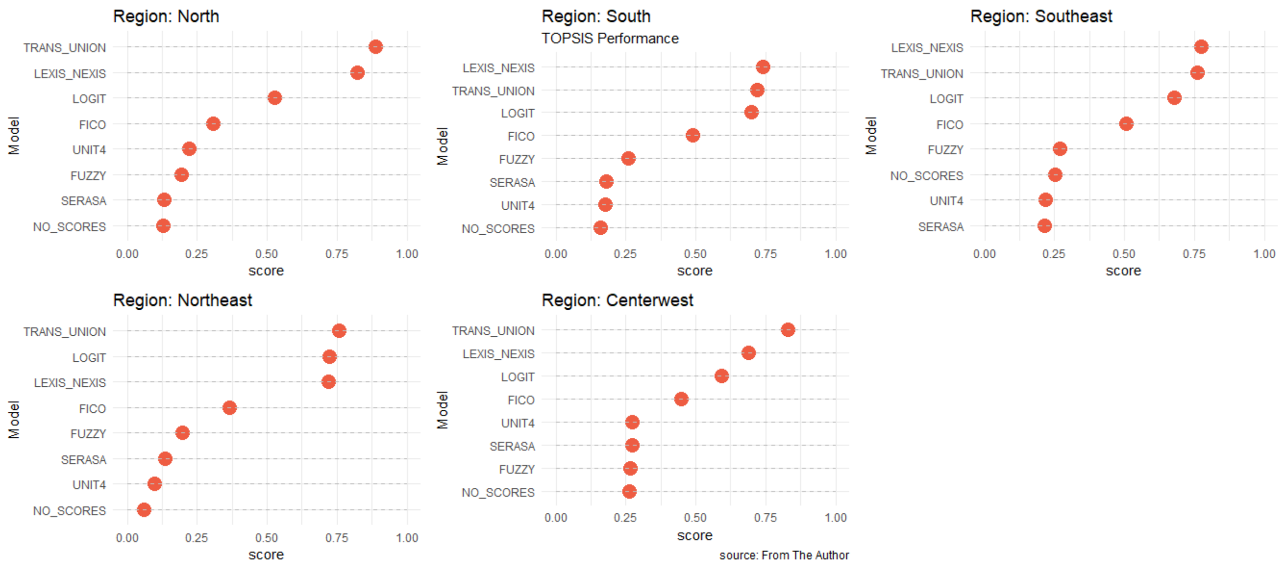
MPLs with FICO score generally presented intermediary results in the different blocks. In a third group, MLPs with Fuzzy, UNIT4 and Serasa scores performed worse than the others, along with the MLP with no scores.

Figure 27: Results for TOPSIS Performance Segregated by Gender



Source: From the Author

Figure 28: Results for TOPSIS Performance Segregated by Region

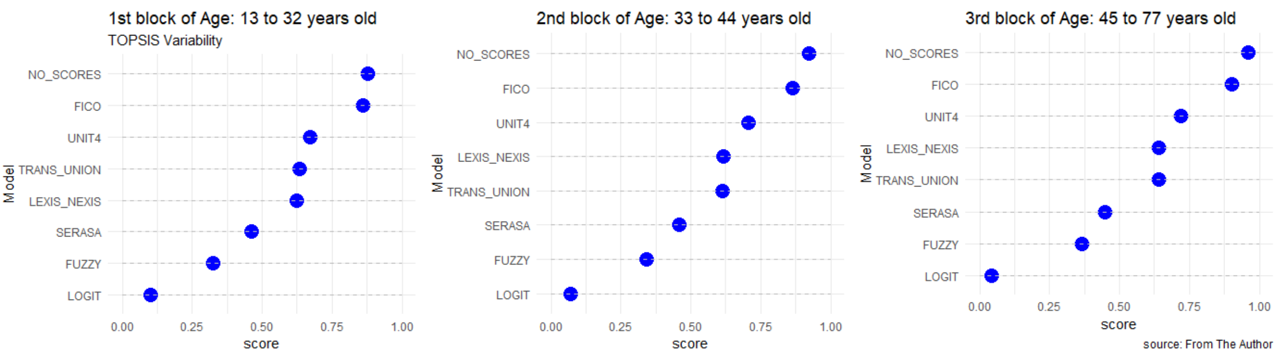


Source: From the Author

In a standalone analysis, the MPL with a Fuzzy score as input performed better in the third block of Age and within the female clients, while presenting slightly better results in the lower income category and in the Southeast region.

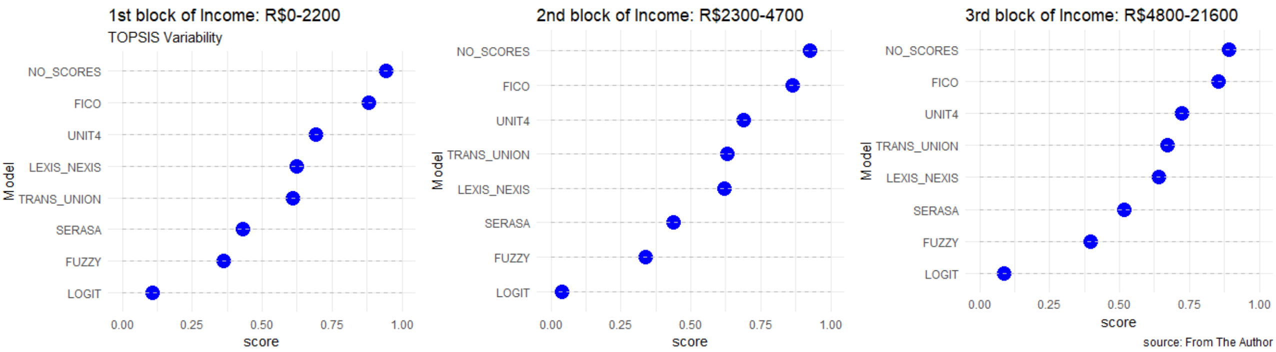
The same analysis is made for the variability of the models (Figures 29 to 32). Again, the results were similar to the analysis in the complete set of data, where Logistic method and MLP with Fuzzy score as input consistently showed lower variability than the other models. On another vein, an MLP without scores and with the FICO score showed the highest variability relatively to the others.

Figure 29: Results for TOPSIS Variability Segregated by Age



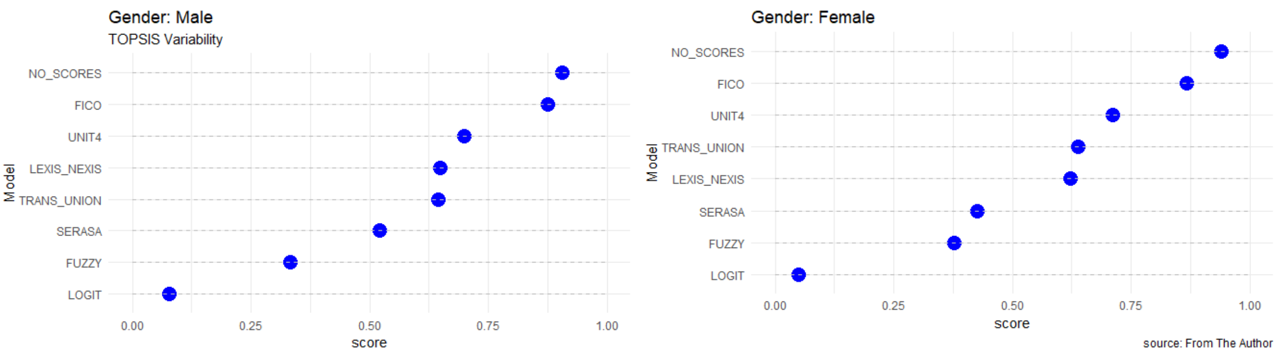
Source: From the Author

Figure 30: Results for TOPSIS Variability Segregated by Income



Source: From the Author

Figure 31: Results for TOPSIS Variability Segregated by Gender

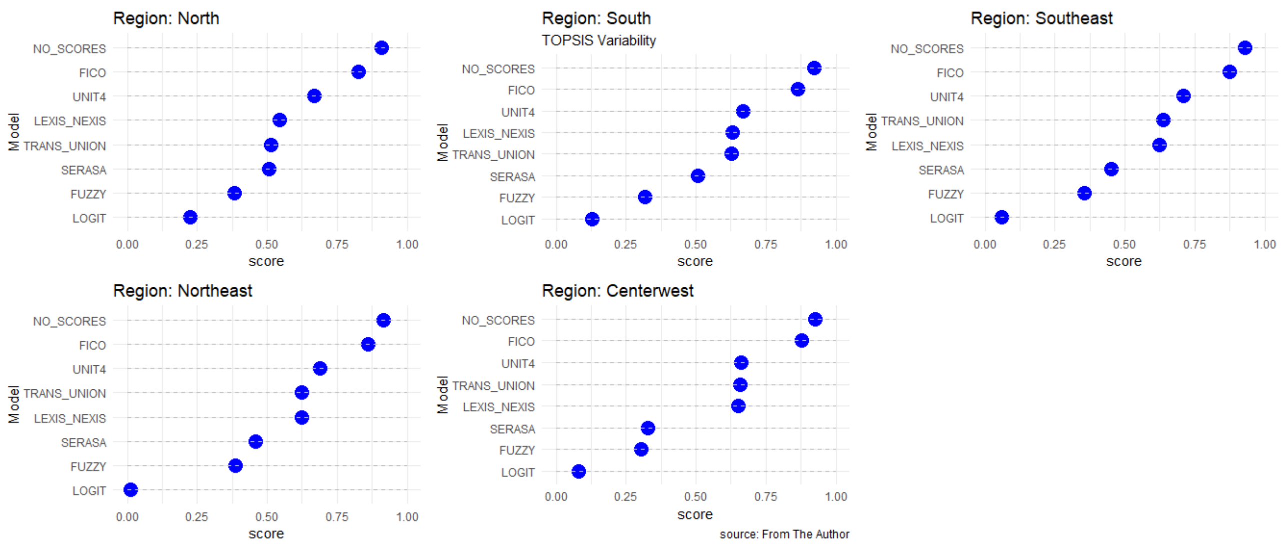


Source: From the Author

The MLPs with Fuzzy score seemed to vary less in the following group of subsets: first block of age (slightly lower variability than the other blocks), intermediary income (slightly lower variability than the other blocks), male gender and Center-west region.

Notably, in some subsets where Fuzzy MLPs seemed to perform better, they also presented higher variability, when comparing within each subset. For example, while showing lower relative performance in the Male subset, these MLPs showed lower relative variability for the same group of people. The same is evident when looking at their results in the younger people subset: relatively lower performance, but lower variability as well, comparatively. This fact seems to be in conjunction with the bias-variance trade-off, as discussed before.

Figure 32: Results for TOPSIS Variability Segregated by Region



Source: From the Author

A final remark that worth to be made is that the MLPs containing fuzzy score, while addressing the issue of imprecision through fuzziness (since they generally performed better than some MLPs with and without scores), weren't sufficiently as good as the model purely based on randomness and probability theory, which is the case of the statistical method Logistic Regression. In line with Zadeh and Bellman's (1970) discussion, who said that imprecision is not only generated by randomness, but also by fuzziness, it could be the case that there is space for exploring probability-based models combined with a fuzzy rule-based system, thus, addressing both issues to diminish imprecision.

5 CONCLUSION

The present work intended to present a method, adapted from Korol (2012), for dealing with uncertainty on a practical application for predictive modelling purposes in credit risk assessment. The results obtained allowed a discussion about uncertainty in terms of predictability power of the models used and their consistency over different samples.

More specifically, the objective was to analyze the predictive capability of models with different score ratings as inputs, one of them being generated through a fuzzy rule-based system. The predictive models used were Multi-Layer Perceptrons (MLPs), a type of neural network frequently called deep learning, aiming at classifying credit card holders into solvent or insolvent. Thus, the process comprised two stages: fuzzy (score creation) and neural (classification). The data included several types of information from clients of a Brazilian credit card provider, including 5 different scores from Score Rating Agencies. But before all that, it was necessary to perform a thorough process of data understanding and pre-processing, in terms of selecting the more suitable variables and the proper data transformation for model fitting, as indicated by Kuhn and Johnson (2013).

In the Fuzzy stage, a score was created based upon a methodology proposed by Korol (2012). In the present research, though, (i) the knowledge for the fuzzy rules was acquired from the data itself (not from an expert), (ii) the output was an actual score rating (Korol's output was the classification output: either solvent or not) and (iii) the score had some knowledge gathered from a benchmark score rating, selected within the data. In fact, this benchmark score rating was used as a response variable for a neural network regression fit, whose parameters importance was used as guidance in one of the steps necessary for the creation of the proposed fuzzy score.

The neural stage consisted of a comparison between the fuzzy proposed score and the other 5 score ratings, all of them acting as inputs for a classification MLP as well, where the predictions were compared (an MLP with no scores and a pure logistic regression were also added to the analysis).

The results indicated good predictability on the models trained with Trans Union and Lexis Nexis score ratings. However, they showed higher variability, while a pure logistic method performed as good as them, but with the lowest variability among all. The first remark is that a much simpler model, like Logistic Regression, is able to perform as well or sometimes better than a highly complex one, like a neural network. As for the proposed fuzzy score, it generated models with intermediary predictability power, but with high consistency (low variability) among different samples. Not only it was better than an MLP with no scores, but it's also beaten two other score ratings (Serasa and UNIT4) in terms of performance. In general, similar results were obtained in a segregated analysis (divided by segments of Age, Region, Gender and Income variables), where, it was possible to realize that, in some segments where the fuzzy method had relatively better performance, it had also higher variability.

In general, it can be said that the proposed method was competitive to the pure application of some market available score ratings as inputs to an MLP since it was able to reduce uncertainty by, at the same, improving predictability and reducing variability when compared to a model with no scores. Hence the combination of a fuzzy with a neural methodology, though not ideal, was satisfactory, which can be a signal that the vagueness inherent to the information in a clients' database was, at least, somewhat treated. It could also serve as a good straightforward indicative of financial health of clients, since it was apparently able to separate more clearly, in terms of financial rating, insolvent from solvent clients than the other scores (due to its higher Cohen's d effect size). Moreover, most of the time, its creation involved pure and simple analysis of the present data (except for the use of a benchmark score), being a viable and cheaper solution than the score rating provided by rating agencies. Finally, since it is about fuzzy logic, its building can also involve the input of specialist practitioners and human being reasoning or heuristics, when applicable.

5.1 **Limitations**

The research is limited in different aspects. While it explores a difficult-to-implement process, its data was also limited to a specific timing and context. Additionally, it didn't prove to be the best among comparable methods.

More specifically, the methodology proposed was an arduous one, since it involved various steps, such as preprocessing, selection of variables, model tuning and the manual creation of fuzzy rules. It was also a time-consuming process, due to the learning time that the Multilayer Perceptrons took to fit the models (2 minutes each in average, compared with 10 seconds for the simpler and, in general, more accurate, Logistic model). With regards to the rules creation, the lack of standardization in such manual process have the potential to generate a huge variety of results, which could undermine its position as an attracting alternative to other traditional methods, under a managerial perspective.

With regards to the database, the methodology was tested upon a specific setting, comprising specific variables that could not be available in other contexts or companies. Conclusions on other settings could be different. The data also lacked other time spans to serve as validation and testing sets, that could make the analysis more robust.

Finally, the score ratings that were used as comparison standpoints towards the proposed fuzzy score are generated through undisclosed methodologies. In that sense, any conclusion as to why they performed better could be limited, which makes it difficult for further improvements in the research in the same directions of, for example, the best performance comparable, which was the Trans Union.

5.2 **Recommendations**

Several ways of exploring the methodology can be further made, mainly with regards to the data itself and to the modeling process.

In terms of the data, future research could try to explore the methodology within other databases involving credit risk assessment, such as the public available Statlog German Dataset (HOFMANN, 1994). In addition, other businesses that imply contexts of uncertainty, imprecision, vagueness and, hence, the need for predictive modeling could serve as environment for testing the methodology. Finally, combinations of the models could be tried aiming at obtaining better performances compared to their individual results.

As for the modeling process, it is recommended that hybrid models with an inbuilt combination of fuzzy and neural network models, aiming at the automatic creation of fuzzy rules for example, are exhaustively tuned to be used in the fuzzy stage. This step would eliminate the lack of standardization inherent to the manual creation of fuzzy rules and would make the modelling process more efficient and replicable. Another possibility is to explore the Takagi-Sugeno inference engine in place of the Mamdani, since it is proved that it can potentially provide more satisfactory results in terms of performance, although usually lacking interpretability. Finally, as previously stated, the same methodology could be explored using statistical learning, aiming at exploring randomness as a source of imprecision with more emphasis. For that, two ways could be possible: (i) using a statistical learning method in the classification phase (logistic regression instead of an ANN, for example) or (ii) using a logistic regression to capture the relationship between variables in the fuzzy rules creation.

APPENDIX A – RULE BLOCKS FOR PHASE 2

Table 10: Block 1 Demographic Rules

input	*input*	*input*	*input*	*input*	*input*	*input*	*input*	*input*	*output*
Age	**Capital**	**Male**	**North**	**Northeast**	**South**	**Southeast**	**CenterWest**	**FB Duration**	**Demographics**
high	yes	yes	yes	no	no	no	no	high	low
high	no	yes	yes	no	no	no	no	high	low
high	yes	no	yes	no	no	no	no	high	slightly low
high	no	no	yes	no	no	no	no	high	slightly low
medium	yes	yes	yes	no	no	no	no	high	low
medium	no	yes	yes	no	no	no	no	high	low
medium	yes	no	yes	no	no	no	no	high	slightly low
medium	no	no	yes	no	no	no	no	high	slightly low
low	yes	yes	yes	no	no	no	no	high	low
low	yes	no	yes	no	no	no	no	high	low
low	no	yes	yes	no	no	no	no	high	low
low	no	no	yes	no	no	no	no	high	low
high	yes	yes	no	yes	no	no	no	high	slightly low
high	no	yes	no	yes	no	no	no	high	slightly low
high	yes	no	no	yes	no	no	no	high	slightly high
high	no	no	no	yes	no	no	no	high	slightly high
medium	yes	yes	no	yes	no	no	no	high	low
medium	no	yes	no	yes	no	no	no	high	low
medium	yes	no	no	yes	no	no	no	high	slightly low
medium	no	no	no	yes	no	no	no	high	slightly low
low	yes	yes	no	yes	no	no	no	high	low
low	yes	no	no	yes	no	no	no	high	slightly low
low	no	yes	no	yes	no	no	no	high	low
low	no	no	no	yes	no	no	no	high	slightly low
high	yes	yes	no	no	yes	no	no	high	low
high	no	yes	no	no	yes	no	no	high	low
high	yes	no	no	no	yes	no	no	high	slightly low
high	no	no	no	no	yes	no	no	high	slightly low
medium	yes	yes	no	no	yes	no	no	high	low
medium	no	yes	no	no	yes	no	no	high	low
medium	yes	no	no	no	yes	no	no	high	slightly low
medium	no	no	no	no	yes	no	no	high	slightly low
low	yes	yes	no	no	yes	no	no	high	low
low	yes	no	no	no	yes	no	no	high	low
low	no	yes	no	no	yes	no	no	high	low
low	no	no	no	no	yes	no	no	high	low
high	yes	yes	no	no	no	yes	no	high	slightly high
high	no	yes	no	no	no	yes	no	high	slightly high

high	yes	no	no	no	no	yes	no	high	high
high	no	no	no	no	no	yes	no	high	high
medium	yes	yes	no	no	no	yes	no	high	slightly low
medium	no	yes	no	no	no	yes	no	high	slightly low
medium	yes	no	no	no	no	yes	no	high	high
medium	no	no	no	no	no	yes	no	high	high
low	yes	yes	no	no	no	yes	no	high	slightly low
low	yes	no	no	no	no	yes	no	high	slightly high
low	no	yes	no	no	no	yes	no	high	slightly low
low	no	no	no	no	no	yes	no	high	slightly high
high	yes	yes	no	no	no	no	yes	high	slightly low
high	no	yes	no	no	no	no	yes	high	slightly low
high	yes	no	no	no	no	no	yes	high	slightly high
high	no	no	no	no	no	no	yes	high	slightly high
medium	yes	yes	no	no	no	no	yes	high	slightly low
medium	no	yes	no	no	no	no	yes	high	slightly low
medium	yes	no	no	no	no	no	yes	high	slightly high
medium	no	no	no	no	no	no	yes	high	slightly high
low	yes	yes	no	no	no	no	yes	high	low
low	yes	no	no	no	no	no	yes	high	slightly high
low	no	yes	no	no	no	no	yes	high	low
low	no	no	no	no	no	no	yes	high	slightly high
high	yes	yes	yes	no	no	no	no	medium	low
high	no	yes	yes	no	no	no	no	medium	low
high	yes	no	yes	no	no	no	no	medium	slightly high
high	no	no	yes	no	no	no	no	medium	slightly high
medium	yes	yes	yes	no	no	no	no	medium	low
medium	no	yes	yes	no	no	no	no	medium	low
medium	yes	no	yes	no	no	no	no	medium	slightly low
medium	no	no	yes	no	no	no	no	medium	slightly low
low	yes	yes	yes	no	no	no	no	medium	low
low	yes	no	yes	no	no	no	no	medium	slightly low
low	no	yes	yes	no	no	no	no	medium	low
low	no	no	yes	no	no	no	no	medium	slightly low
high	yes	yes	no	yes	no	no	no	medium	slightly low
high	no	yes	no	yes	no	no	no	medium	slightly low
high	yes	no	no	yes	no	no	no	medium	slightly high
high	no	no	no	yes	no	no	no	medium	slightly high
medium	yes	yes	no	yes	no	no	no	medium	slightly low
medium	no	yes	no	yes	no	no	no	medium	slightly low
medium	yes	no	no	yes	no	no	no	medium	slightly high
medium	no	no	no	yes	no	no	no	medium	slightly high
low	yes	yes	no	yes	no	no	no	medium	low

low	yes	no	no	yes	no	no	no	medium	slightly low
low	no	yes	no	yes	no	no	no	medium	low
low	no	no	no	yes	no	no	no	medium	slightly low
high	yes	yes	no	no	yes	no	no	medium	low
high	no	yes	no	no	yes	no	no	medium	low
high	yes	no	no	no	yes	no	no	medium	slightly high
high	no	no	no	no	yes	no	no	medium	slightly high
medium	yes	yes	no	no	yes	no	no	medium	low
medium	no	yes	no	no	yes	no	no	medium	low
medium	yes	no	no	no	yes	no	no	medium	slightly low
medium	no	no	no	no	yes	no	no	medium	slightly low
low	yes	yes	no	no	yes	no	no	medium	low
low	yes	no	no	no	yes	no	no	medium	slightly low
low	no	yes	no	no	yes	no	no	medium	low
low	no	no	no	no	yes	no	no	medium	slightly low
high	yes	yes	no	no	no	yes	no	medium	slightly high
high	no	yes	no	no	no	yes	no	medium	slightly high
high	yes	no	no	no	no	yes	no	medium	high
high	no	no	no	no	no	yes	no	medium	high
medium	yes	yes	no	no	no	yes	no	medium	slightly high
medium	no	yes	no	no	no	yes	no	medium	slightly high
medium	yes	no	no	no	no	yes	no	medium	high
medium	no	no	no	no	no	yes	no	medium	high
low	yes	yes	no	no	no	yes	no	medium	slightly low
low	yes	no	no	no	no	yes	no	medium	high
low	no	yes	no	no	no	yes	no	medium	slightly low
low	no	no	no	no	no	yes	no	medium	high
high	yes	yes	no	no	no	no	yes	medium	slightly high
high	no	yes	no	no	no	no	yes	medium	slightly high
high	yes	no	no	no	no	no	yes	medium	high
high	no	no	no	no	no	no	yes	medium	high
medium	yes	yes	no	no	no	no	yes	medium	slightly low
medium	no	yes	no	no	no	no	yes	medium	slightly low
medium	yes	no	no	no	no	no	yes	medium	slightly high
medium	no	no	no	no	no	no	yes	medium	slightly high
low	yes	yes	no	no	no	no	yes	medium	slightly low
low	yes	no	no	no	no	no	yes	medium	slightly high
low	no	yes	no	no	no	no	yes	medium	slightly low
low	no	no	no	no	no	no	yes	medium	slightly high
high	yes	yes	yes	no	no	no	no	low	slightly low
high	no	yes	yes	no	no	no	no	low	slightly low
high	yes	no	yes	no	no	no	no	low	slightly high
high	no	no	yes	no	no	no	no	low	slightly high

medium	yes	yes	yes	no	no	no	no	low	low
medium	no	yes	yes	no	no	no	no	low	low
medium	yes	no	yes	no	no	no	no	low	slightly high
medium	no	no	yes	no	no	no	no	low	slightly high
low	yes	yes	yes	no	no	no	no	low	low
low	yes	no	yes	no	no	no	no	low	slightly low
low	no	yes	yes	no	no	no	no	low	low
low	no	no	yes	no	no	no	no	low	slightly low
high	yes	yes	no	yes	no	no	no	low	slightly low
high	no	yes	no	yes	no	no	no	low	slightly low
high	yes	no	no	yes	no	no	no	low	high
high	no	no	no	yes	no	no	no	low	high
medium	yes	yes	no	yes	no	no	no	low	slightly low
medium	no	yes	no	yes	no	no	no	low	slightly low
medium	yes	no	no	yes	no	no	no	low	slightly high
medium	no	no	no	yes	no	no	no	low	slightly high
low	yes	yes	no	yes	no	no	no	low	slightly low
low	yes	no	no	yes	no	no	no	low	slightly high
low	no	yes	no	yes	no	no	no	low	slightly low
low	no	no	no	yes	no	no	no	low	slightly high
high	yes	yes	no	no	yes	no	no	low	slightly low
high	no	yes	no	no	yes	no	no	low	slightly low
high	yes	no	no	no	yes	no	no	low	slightly high
high	no	no	no	no	yes	no	no	low	slightly high
medium	yes	yes	no	no	yes	no	no	low	low
medium	no	yes	no	no	yes	no	no	low	low
medium	yes	no	no	no	yes	no	no	low	slightly high
medium	no	no	no	no	yes	no	no	low	slightly high
low	yes	yes	no	no	yes	no	no	low	low
low	yes	no	no	no	yes	no	no	low	slightly low
low	no	yes	no	no	yes	no	no	low	low
low	no	no	no	no	yes	no	no	low	slightly low
high	yes	yes	no	no	no	yes	no	low	high
high	no	yes	no	no	no	yes	no	low	high
high	yes	no	no	no	no	yes	no	low	high
high	no	no	no	no	no	yes	no	low	high
medium	yes	yes	no	no	no	yes	no	low	slightly high
medium	no	yes	no	no	no	yes	no	low	slightly high
medium	yes	no	no	no	no	yes	no	low	high
medium	no	no	no	no	no	yes	no	low	high
low	yes	yes	no	no	no	yes	no	low	slightly high
low	yes	no	no	no	no	yes	no	low	high
low	no	yes	no	no	no	yes	no	low	slightly high

low	no	no	no	no	no	yes	no	low	high
high	yes	yes	no	no	no	no	yes	low	slightly high
high	no	yes	no	no	no	no	yes	low	slightly high
high	yes	no	no	no	no	no	yes	low	high
high	no	no	no	no	no	no	yes	low	high
medium	yes	yes	no	no	no	no	yes	low	slightly high
medium	no	yes	no	no	no	no	yes	low	slightly high
medium	yes	no	no	no	no	no	yes	low	high
medium	no	no	no	no	no	no	yes	low	high
low	yes	yes	no	no	no	no	yes	low	slightly low
low	yes	no	no	no	no	no	yes	low	slightly high
low	no	yes	no	no	no	no	yes	low	slightly low
low	no	no	no	no	no	no	yes	low	slightly high

Table 11: Rule Block 2 Finance

input	*input*	*input*	*output*
Monthly Income	**Credit Line Request**	**Monthly Spend**	**Finance**
high	high	high	low
high	high	medium	low
high	medium	high	slightly low
high	medium	medium	slightly low
high	medium	low	slightly high
high	low	medium	slightly high
high	low	low	slightly high
high	high	low	slightly low
high	low	high	slightly high
medium	high	high	low
medium	high	medium	slightly low
medium	high	low	slightly low
medium	medium	high	slightly low
medium	medium	medium	slightly high
medium	medium	low	slightly high
medium	low	high	slightly high
medium	low	medium	slightly high
medium	low	low	high
low	high	high	slightly low
low	high	medium	slightly low
low	high	low	slightly high
low	medium	high	slightly high
low	medium	medium	slightly high
low	medium	low	slightly high
low	low	high	slightly high

low	low	medium	high
low	low	low	high

Table 12: Rule Block 3 Assets

input	*input*	*input*	*input*	*output*
Residence Duration	**Bank Acc Duration**	**Monthly Rent**	**Residence Rent**	**Assets**
high	high	high	yes	slightly high
high	high	medium	yes	slightly high
high	high	low	yes	high
high	medium	high	yes	slightly low
high	medium	medium	yes	slightly high
high	medium	low	yes	slightly high
high	low	high	yes	slightly low
high	low	medium	yes	slightly low
high	low	low	yes	slightly high
medium	high	high	yes	slightly low
medium	high	medium	yes	slightly low
medium	high	low	yes	slightly high
medium	medium	high	yes	low
medium	medium	medium	yes	slightly low
medium	medium	low	yes	slightly low
medium	low	high	yes	low
medium	low	medium	yes	low
medium	low	low	yes	slightly low
low	high	high	yes	low
low	high	medium	yes	low
low	high	low	yes	slightly low
low	medium	high	yes	low
low	medium	medium	yes	low
low	medium	low	yes	low
low	low	high	yes	low
low	low	medium	yes	low
low	low	low	yes	low
high	high	high	no	high
high	high	medium	no	high
high	high	low	no	high
high	medium	high	no	high
high	medium	medium	no	high
high	medium	low	no	high
high	low	high	no	slightly high
high	low	medium	no	high
high	low	low	no	high

medium	high	high	no	slightly high
medium	high	medium	no	high
medium	high	low	no	high
medium	medium	high	no	slightly high
medium	medium	medium	no	slightly high
medium	medium	low	no	high
medium	low	high	no	slightly low
medium	low	medium	no	slightly high
medium	low	low	no	slightly high
low	high	high	no	slightly low
low	high	medium	no	slightly high
low	high	low	no	slightly high
low	medium	high	no	slightly low
low	medium	medium	no	slightly low
low	medium	low	no	slightly high
low	low	high	no	low
low	low	medium	no	slightly low
low	low	low	no	slightly high

Table 13: Rule Block Final

input	input	input	output
demographics	finance	assets	Final Score
high	high	high	high
high	high	medium	high
high	high	low	slightly high
high	medium	high	high
high	medium	medium	slightly high
high	medium	low	slightly high
high	low	high	high
high	low	medium	slightly high
high	low	low	slightly low
medium	high	high	high
medium	high	medium	slightly high
medium	high	low	slightly low
medium	medium	high	slightly high
medium	medium	medium	slightly high
medium	medium	low	slightly low
medium	low	high	slightly high
medium	low	medium	slightly low
medium	low	low	low
low	high	high	slightly high
low	high	medium	slightly low

low	high	low	low
low	medium	high	slightly low
low	medium	medium	slightly low
low	medium	low	low
low	low	high	slightly low
low	low	medium	low
low	low	low	low

APPENDIX B – RESULTS TABLES

Table 14: TOPSIS Performance segregated per Region

Southeast		North		South		Northeast		Center-West	
TOPSIS	**Model**	**TOPSIS**	**Model**	**TOPSIS**	**Model**	**TOPSIS**	**Model**	**TOPSIS**	**Model**
0.774	LEXIS_NEXIS	0.886	TRANS_UNION	0.740	LEXIS_NEXIS	0.757	TRANS_UNION	0.830	TRANS_UNION
0.760	TRANS_UNION	0.822	LEXIS_NEXIS	0.721	TRANS_UNION	0.722	LOGIT	0.689	LEXIS_NEXIS
0.679	LOGIT	0.526	LOGIT	0.698	LOGIT	0.719	LEXIS_NEXIS	0.591	LOGIT
0.506	FICO	0.309	FICO	0.489	FICO	0.365	FICO	0.449	FICO
0.268	FUZZY	0.220	UNIT4	0.258	FUZZY	0.199	FUZZY	0.274	UNIT4
0.253	NO_SCORES	0.193	FUZZY	0.179	SERASA	0.136	SERASA	0.274	SERASA
0.216	UNIT4	0.133	SERASA	0.176	UNIT4	0.097	UNIT4	0.264	FUZZY
0.213	SERASA	0.130	NO_SCORES	0.159	NO_SCORES	0.058	NO_SCORES	0.261	NO_SCORES

Source: From the Author

Table 15: TOPSIS Performance segregated by gender

Male		Female	
TOPSIS	**Model**	**TOPSIS**	**Model**
0.787	TRANS_UNION	0.769	LEXIS_NEXIS
0.745	LEXIS_NEXIS	0.745	TRANS_UNION
0.662	LOGIT	0.733	LOGIT
0.449	FICO	0.497	FICO
0.163	FUZZY	0.298	FUZZY
0.135	UNIT4	0.245	NO_SCORES
0.110	SERASA	0.241	SERASA
0.076	NO_SCORES	0.220	UNIT4

Source: Source from the Author

Table 16: TOPSIS Performance segregated by Income

Income Block 1		Income Block 2		Income Block 3	
TOPSIS	**Model**	**TOPSIS**	**Model**	**TOPSIS**	**Model**
0.807	TRANS_UNION	0.770	TRANS_UNION	0.779	LEXIS_NEXIS
0.711	LEXIS_NEXIS	0.764	LEXIS_NEXIS	0.765	LOGIT
0.662	LOGIT	0.721	LOGIT	0.726	TRANS_UNION
0.580	FICO	0.410	FICO	0.394	FICO
0.260	FUZZY	0.241	FUZZY	0.218	FUZZY
0.229	SERASA	0.198	SERASA	0.198	SERASA
0.219	UNIT4	0.171	NO_SCORES	0.185	NO_SCORES
0.189	NO_SCORES	0.168	UNIT4	0.177	UNIT4

Source: From the Author

Table 17: TOPSIS Performance segregated by Age

Age Block 1		Age Block 2		Age Block 3	
TOPSIS	**Model**	**TOPSIS**	**Model**	**TOPSIS**	**Model**
0.793	TRANS_UNION	0.81422	TRANS_UNION	0.740	LOGIT
0.767	LEXIS_NEXIS	0.793944	LEXIS_NEXIS	0.710	LEXIS_NEXIS
0.715	LOGIT	0.708076	LOGIT	0.695	TRANS_UNION
0.483	FICO	0.466557	FICO	0.487	FICO
0.199	FUZZY	0.244411	FUZZY	0.323	FUZZY
0.173	SERASA	0.188355	UNIT4	0.233	SERASA
0.159	NO_SCORES	0.168141	NO_SCORES	0.224	NO_SCORES
0.158	UNIT4	0.144055	SERASA	0.217	UNIT4

Source: From the Author

Table 18: TOPSIS Variability segregated by Region

Southeast		South		Northeast		North		Center-West	
TOPSIS	**Model**	**TOPSIS**	**Model**	**TOPSIS**	**Model**	**TOPSIS**	**Model**	**TOPSIS**	**Model**
0.930	NO_SCORES	0.921	NO_SCORES	0.916	NO_SCORES	0.910	NO_SCORES	0.927	NO_SCORES
0.874	FICO	0.865	FICO	0.860	FICO	0.827	FICO	0.879	FICO
0.709	UNIT4	0.668	UNIT4	0.687	UNIT4	0.669	UNIT4	0.662	UNIT4
0.638	TRANS_UNION	0.629	LEXIS_NEXIS	0.625	TRANS_UNION	0.543	LEXIS_NEXIS	0.657	TRANS_UNION
0.623	LEXIS_NEXIS	0.626	TRANS_UNION	0.623	LEXIS_NEXIS	0.513	TRANS_UNION	0.650	LEXIS_NEXIS
0.452	SERASA	0.505	SERASA	0.459	SERASA	0.506	SERASA	0.328	SERASA
0.355	FUZZY	0.319	FUZZY	0.387	FUZZY	0.384	FUZZY	0.305	FUZZY
0.058	LOGIT	0.129	LOGIT	0.012	LOGIT	0.225	LOGIT	0.079	LOGIT

Source: From the Author

Table 19: TOPSIS Variability segregated by Gender

Male		Female	
TOPSIS	**Model**	**TOPSIS**	**Model**
0.906	NO_SCORES	0.939	NO_SCORES
0.875	FICO	0.868	FICO
0.699	UNIT4	0.711	UNIT4
0.650	LEXIS_NEXIS	0.638	TRANS_UNION
0.644	TRANS_UNION	0.622	LEXIS_NEXIS
0.521	SERASA	0.426	SERASA
0.332	FUZZY	0.378	FUZZY
0.077	LOGIT	0.050	LOGIT

Source: From the Author

Table 20: TOPSIS Variability segregated by Income

Income Block 1		Income Block 2		Income Block 1	
TOPSIS	**Model**	**TOPSIS**	**Model**	**TOPSIS**	**Model**
0.943	NO_SCORES	0.925	NO_SCORES	0.893	NO_SCORES
0.881	FICO	0.865	FICO	0.852	FICO
0.692	UNIT4	0.689	UNIT4	0.724	UNIT4
0.624	LEXIS_NEXIS	0.629	TRANS_UNION	0.671	TRANS_UNION
0.611	TRANS_UNION	0.619	LEXIS_NEXIS	0.640	LEXIS_NEXIS
0.433	SERASA	0.439	SERASA	0.515	SERASA
0.362	FUZZY	0.337	FUZZY	0.395	FUZZY
0.109	LOGIT	0.039	LOGIT	0.088	LOGIT

Source: From the Author

Table 21: TOPSIS Variability segregated by Age

Age Block 1		Age Block 2		Age Block 3	
TOPSIS	**Model**	**TOPSIS**	**Model**	**TOPSIS**	**Model**
0.878	NO_SCORES	0.921	NO_SCORES	0.960	NO_SCORES
0.860	FICO	0.864	FICO	0.902	FICO
0.672	UNIT4	0.704	UNIT4	0.718	UNIT4
0.632	TRANS_UNION	0.618	LEXIS_NEXIS	0.642	LEXIS_NEXIS
0.624	LEXIS_NEXIS	0.612	TRANS_UNION	0.640	TRANS_UNION
0.461	SERASA	0.457	SERASA	0.448	SERASA
0.326	FUZZY	0.342	FUZZY	0.366	FUZZY
0.101	LOGIT	0.072	LOGIT	0.043	LOGIT

Source: From the Author

BIBLIOGRAPHY

ABDULRAHMAN, Umar Farouk Ibn; PANFORD, Joseph Kobina; HAYFRON-ACQUAH, James Ben. **Fuzzy logic approach to credit scoring for micro finance in ghana: a case study of KWIQPLUS money lending**. International Journal of Computer Applications, v. 94, n. 8, 2014.

ADDO, Peter; GUEGAN, Dominique; HASSANI, Bertrand. **Credit risk analysis using machine and deep learning models**. Risks, v. 6, n. 2, p. 38, 2018.

ADNAN AZIZ, M.; DAR, Humayon A. **Predicting corporate bankruptcy: where we stand?.** Corporate Governance: The international journal of business in society, v. 6, n. 1, p. 18-33, 2006.

ALIEV, Rafik Aziz; FAZLOLLAHI, Bijan; ALIEV, Rashad Rafik. **Soft computing and its applications in business and economics.** Springer, 2012.

AMUNATEGUI, Manuel. "**Brief Walkthrough Of The dummyVars Function From {caret}**" (2014). Catalonia. Spain. Available at: https://amunategui.github.io/dummyVar-Walkthrough/

ARTHUR, S. **Some Studies in Machine Learning Using the Game of Checkers**. IBM Journal of Research and Development. Vol. 3 (3): p.210–229, 1959.

BAHRAMMIRZAEE, Arash. **"A comparative survey of artificial intelligence applications in finance: artificial neural networks, expert system and hybrid intelligent systems."** Neural Computing and Applications 19.8 (2010): 1165-1195.

BECK, Marcus W. **NeuralNetTools: Visualization and Analysis Tools for Neural Networks**. Journal of statistical software, v. 85, n. 11, p. 1, 2018.

BELLMAN, Richard E.; ZADEH, Lotfi Asker. **Decision-making in a fuzzy environment**. Management science, v. 17, n. 4, p. B-141-B-164, 1970.

BERGMEIR, Christoph Norbert, and José Manuel Benítez Sánchez. **"Neural networks in R using the Stuttgart neural network simulator: RSNNS."** American Statistical Association, 2012.

BOJADZIEV, G.; BOJADZIEV, M., **Fuzzy Logic for Business, Finance, and Management**. Second Ed., Singapore: World Scientific Publishing Co. Ltd., 2007

BRUMMELHUIS, Raymond, and Zhongmin LUO. "**CDS Rate Construction Methods by Machine Learning Techniques.**" arXiv preprint arXiv:1705.06899 (2017).

CAMPBELL, John Y., and Joao F. COCCO. **"A model of mortgage default."** The Journal of Finance 70.4 (2015): 1495-1554.

CHAWLA, N. V., BOWYER, K. W., HALL, L. O., & Kegelmeyer, W. P. (2002). **SMOTE: synthetic minority over-sampling technique**. Journal of artificial intelligence research, 16, 321-357.

CHEN, Liang-Hsuan; CHIOU, Tai-Wei. **A fuzzy credit-rating approach for commercial loans: a Taiwan case.** Omega, v. 27, n. 4, p. 407-419, 1999.

CHESSHYRE, T. **“Stuttgart Mercedes-Benz museum”**. The Times, June 20th, 2009. Available at: < https://www.thetimes.co.uk/article/stuttgarts-mercedes-benz-museum-pj5vvf0ff8v>

COE, Robert. **"It's the effect size, stupid: What effect size is and why it is important."** (2002). Leeds, UK. 2002. Available at:< http://www.leeds.ac.uk/educol/documents/00002182.htm>

COHEN, Jacob. **"A power primer."** Psychological bulletin 112.1 (1992): 155.

COHEN, Jacob. **"Statistical power analysis for the behavioral sciences**”. 1988, Hillsdale, NJ: L. Lawrence Earlbaum Associates 2 (1988).

COURTNEY,H; KIRKLAND, J; VIGUERIE, P., **Strategy Under Uncertainty**. HBS Nov-Dec 1997. Boston, MA: HARVARD BUSINESS SCHOOL PUBLISHING, p.2, 1997.

DALIYEV, A. (2015). **Credit risk assessment by means of fuzzy logic prediction.** Journal of Multidisciplinary Engineering Science and Technology (JMEST), 2(5)

DE AMORIM, B. P.; VASCONCELOS, Germano C.; BRASIL, Lourdes Mattos. **Hybrid neural systems for large scale credit risk assessment applications.** Journal of Intelligent & Fuzzy Systems, v. 18, n. 5, p. 455-464, 2007.

DESAI, Vijay S., Jonathan N. CROOK, and George A. OVERSTREET JR. **"A comparison of neural networks and linear scoring models in the credit union environment."** European Journal of Operational Research 95.1 (1996): 24-37.

DIRICK, Lore, Gerda CLAESKENS, and Bart BAESENS. **"Time to default in credit scoring using survival analysis: a benchmark study."** Journal of the Operational Research Society 68.6 (2017): 652-665.

EFRON, Bradley; TIBSHIRANI, Robert J. **Cross-validation and the bootstrap: Estimating the error rate of a prediction rule**. Division of Biostatistics, Stanford University, 1995.

ELETTER, Shorouq Fathi, and Saad Ghaleb YASEEN. **"Applying neural networks for loan decisions in the Jordanian commercial banking system."** International Journal of Computer Science and Network Security 10.1 (2010): 209-214.

GANGULY, Dev Shankar. **Top Differences Between Artificial Intelligence, Machine Learning & Deep Learning**. URL (version: 2018). Available at: https://hackernoon.com/top-differences-between-artificial-intelligence-machine-learning-deep-learning-d39cb6f6feaa

GARCÍA, V., MARQUÉS, A. I., CLEOFAS-SÁNCHEZ, L., & SÁNCHEZ, J. S. (2016, April). **Model Selection for Financial Distress Prediction by Aggregating TOPSIS**

and PROMETHEE Rankings. In International Conference on Hybrid Artificial Intelligence Systems (pp. 524-535). Springer, Cham.

GARSON, G. David. **"Interpreting neural-network connection weights."** AI expert 6.4 (1991): 46-51.

GHATGE, A. R., and P. P. HALKARNIKAR. **"Ensemble neural network strategy for predicting credit default evaluation."** International Journal of Engineering and Innovative Technology (IJEIT) Volume 2 (2013): 223-225.

GIANG, H. (2005). **Using neural network in predicting corporate failure**. Journal of Social Sciences, 1(4), 199–202. doi:10.3844/jssp.2005.199.202

GOH, Anthony TC. **"Back-propagation neural networks for modeling complex systems."** Artificial Intelligence in Engineering 9.3 (1995): 143-151.

GRACE, Asogbon Mojisola, WILLIAMS, and Samuel Oluwarotimi. **"Comparative analysis of neural network and fuzzy logic techniques in credit risk evaluation."** International Journal of Intelligent Information Technologies (IJIIT) 12.1 (2016): 47-62.

GÜNTHER, Frauke, and FRITSCH, Stefan. **"neuralnet: Training of neural networks."** The R journal 2.1 (2010): 30-38.

HANDZIC, M., & AURUM, A. (2001, September 19-21). **Knowledge discovery: Some empirical evidence and directions for future research**. Proceedings of the 5th International Conference on Wirtschafts Informatics WI'01, Augsburg, Germany. doi:10.1007/978-3-642-57547-1_86

HOFFMANN, Frank et al. **Comparing a genetic fuzzy and a neurofuzzy classifier for credit scoring.** International Journal of Intelligent Systems, v. 17, n. 11, p. 1067-1083, 2002

HOFMANN, P. D. H. Statlog (German Credit Data) Data Set. **UCI Repository of Machine Learning Databases**, 1994.

HUANG, Z., CHEN, H., HSU, C.-J., CHEN, W.-H., & WU, S. (2004). **Credit rating analysis with support vector machines and neural networks: A market comparative study. Decision Support Systems**, 37(4), 543–558. doi:10.1016/S0167-9236(03)00086-1

IGNATIUS, J., HATAMI-MARBINI, A., RAHMAN, A. et al. "**A fuzzy decision support system for credit scoring"**. Neural Comput & Applic (2018) 29: 921. Springer London. May 2018.

JAMES, Gareth et al. **An introduction to statistical learning**. New York: springer, 2013.

JENSEN, Herbert L. **"Using neural networks for credit scoring."** Managerial finance 18.6 (1992): 15-26.

KHAMIS, Azme, et al. **"The effects of outliers data on neural network performance."** Journal of Applied Sciences 5.8 (2005): 1394-1398.

KOROL, T. **Fuzzy Logic in Financial Management, Fuzzy Logic - Emerging Technologies and Applications.** Shanghai. p.260-285. 2012

KUBAT, M., & MATWIN, S. (1997, July). **Addressing the curse of imbalanced training sets: one-sided selection**. In Icml (Vol. 97, pp. 179-186).

KUHN, M. "**Package 'rsample.**" General Resampling Infrastructure (2019).

KUHN, M., J. Wing, and S. Weston. "**Package 'caret'.**" Classification and regression training (2015).

KUHN, Max; JOHNSON, Kjell. **Applied predictive modeling.** New York: Springer, 2013.

KUMARASWAMY, Visalini, **What is the difference between mamdani and sugeno in fuzzy logic?**, URL (version: 2018-07-04). Available at: https://math.stackexchange.com/q/2840703

L'SOUZA, Rita; GUPTA, Manu. **Hand Gesture Recognition for Human Robot Interaction**. Massachusetts Institute of Technology. URL (version: 2006) Available at: http://courses.media.mit.edu/2006fall/mas622j/Projects/manu-rita-MAS_Proj/MLP.pdf

LAHSASNA, Adel. **Evaluation of credit risk using evolutionary-fuzzy logic scheme**. 2009. Doctoral Dissertation. Universiti Malaya.

LIEBERMAN, D. **"CEO Forum: Microsoft's Ballmer having a 'great time'"**. USA Today, April 29th, 2007. Available at: < https://usatoday30.usatoday.com/money/companies/management/2007-04-29-ballmer-ceo-forum-usat_N.htm >

LOUZADA, Francisco, Anderson ARA, and Guilherme B. FERNANDES. **"Classification methods applied to credit scoring: Systematic review and overall comparison."** Surveys in Operations Research and Management Science 21.2 (2016): 117-134.

LUNARDON, Nicola; MENARDI, Giovanna; TORELLI, Nicola. **ROSE: A Package for Binary Imbalanced Learning**. R Journal, v. 6, n. 1, 2014.

McCAFFREY, J. **"How to use resilient backpropagation to train neural networks",** California, US. 2015. Available at: https://visualstudiomagazine.com/articles/2015/03/01/resilient-back-propagation.aspx

MONDAL, Ariful. **"Classifications in R: Response Modeling/Credit Scoring/Credit Rating using Machine Learning Techniques"** (2016). Available at: https://rstudio-pubs-static.s3.amazonaws.com/225209_df0130c5a0614790b6365676b9372c07.html#22_univariate_and_bivariate_analysis

NGUYEN, Giang Hoang; BOUZERDOUM, Abdesselam; PHUNG, Son Lam. **Learning pattern classification tasks with imbalanced data sets.** In: Pattern recognition. InTech, 2009.

ODOM, M., & SHARDA, R. (1990). "**A neural network model for bankruptcy prediction.**" Proceedings of the International Joint Conference on Neural networks (pp. 163-168)

OLDEN, Julian D., and Donald A. Jackson. **"Illuminating the "black box": a randomization approach for understanding variable contributions in artificial neural networks."** Ecological modelling 154.1-2 (2002): 135-150.

OLDEN, Julian D., Michael K. Joy, and Russell G. Death. **"An accurate comparison of methods for quantifying variable importance in artificial neural networks using simulated data."** Ecological Modelling 178.3-4 (2004): 389-397.

PHILLIPS, M. **"Nasdaq: Here's Our Timeline of the Flash Crash"**. Wall Street Journal, May 11th, 2010. Available at: < https://blogs.wsj.com/marketbeat/2010/05/11/nasdaq-heres-our-timeline-of-the-flash-crash/>

RAJAB, Sharifa, and Vinod SHARMA. **"A review on the applications of neuro-fuzzy systems in business."** Artificial Intelligence Review 49.4 (2018): 481-510.

RIZA, Lala Septem et al. **frbs: Fuzzy rule-based systems for classification and regression in R.** 2015.

RODRIGUEZ, M, **The Failure of Predictive Modeling and Irrational Thinking.** Miami, US. 2011. Available at: < http://miamidreamteam.com/failure-predictive-modeling-irrational-thinking/>

SERNEELS, Sven, Evert De Nolf, and Pierre J. Van Espen. **"Spatial sign preprocessing: a simple way to impart moderate robustness to multivariate estimators."** Journal of Chemical Information and Modeling 46.3 (2006): 1402-1409.

SERNEELS, Sven; DE NOLF, Evert; VAN ESPEN, Pierre J. **Spatial sign preprocessing: a simple way to impart moderate robustness to multivariate estimators**. Journal of Chemical Information and Modeling, v. 46, n. 3, p. 1402-1409, 2006.

SHALEV-SHWARTZ, Shai, and Shai BEN-DAVID. **Understanding machine learning: From theory to algorithms**. Cambridge university press, 2014.

TORBAT, Akbar E. (2008). **Global Financial Meltdown and the Demise of Neoliberalism**. Global Research, Center for Research on Globalization, Available at: http://www.globalresearch.ca/index.php?context=va&aid=10549

TORCHIANO, Marco; TORCHIANO, Maintainer Marco. **Package 'effsize'**. 2018.

TORGO, Luis; TORGO, Maintainer Luis. **Package ‘DMwR’**. Comprehensive R Archive Network, 2015.

VELLIDO, A., LISBOA, P. J. G., and VAUGHAN, J. (1999). “**Neural networks in business: A survey of applications”** (1992–1998). Expert Systems with Applications, 17, 51–70.

WANKE, Peter; AZAD, Abul Kalam; EMROUZNEJAD, Ali. **Efficiency in BRICS banking under data vagueness: A two-stage fuzzy approach**. Global Finance Journal, v. 35, p. 58-71, 2018.

YAO, Ping; WU, Chong; YAO, Minghui. **Credit risk assessment model of commercial banks based on fuzzy neural network**. In: International Symposium on Neural Networks. Springer, Berlin, Heidelberg, 2009. p. 976-985.

YOON, K. Paul; HWANG, Ching-Lai. **Multiple attribute decision making: an introduction.** Sage publications, 1995.

YOUDEN, W. J. (1950). **Index for rating diagnostic tests**. Cancer, 3(1), 32-35.

ZADEH, Lofti A. **"Information and control."** Fuzzy sets 8.3 (1965): 338-353.

ZADEH, Lotfi A. **"Fuzzy logic, neural networks and soft computing."** Safety Evaluation Based on Identification Approaches Related to Time-Variant and Nonlinear Structures. Vieweg+ Teubner Verlag, 1993. 320-321.

ZELL, Andreas, et al. **“Stuttgart Neural Network Simulator, User Manual, Version 4.2."** University of Stuttgart, Institute for Parallel and Distributed HighPerformance Systems, Technical Report 6 (1995).