

Article

Variations of Particle Swarm Optimization for Obtaining Classification Rules Applied to Credit Risk in Financial Institutions of Ecuador

Patricia Jimbo Santana ¹, Laura Lanzarini ² and Aurelio F. Bariviera ^{3,*}

¹ Facultad de Ciencias Administrativas, Carrera de Contabilidad y Auditoría, Universidad Central del Ecuador, Quito 170129, Ecuador; prjimbo@uce.edu.ec

² Instituto de Investigación en Informática LIDI, Facultad de Informática, Universidad Nacional de la Plata, La Plata C1900, Buenos Aires, Argentina; laural@lidi.info.unlp.edu.ar

³ Department of Business, Universitat Rovira i Virgili, Avenida de la Universitat, 43204 Reus, Spain

* Correspondence: aurelio.fernandez@urv.cat; Tel.: +34-977-759-833

Received: 9 September 2019; Accepted: 24 December 2019; Published: 30 December 2019



Abstract: Knowledge generated using data mining techniques is of great interest for organizations, as it facilitates tactical and strategic decision making, generating a competitive advantage. In the special case of credit granting organizations, it is important to clearly define rejection/approval criteria. In this direction, classification rules are an appropriate tool, provided that the rule set has low cardinality and that the antecedent of the rules has few conditions. This paper analyzes different solutions based on Particle Swarm Optimization (PSO) techniques, which are able to construct a set of classification rules with the aforementioned characteristics using information from the borrower and the macroeconomic environment at the time of granting the loan. In addition, to facilitate the understanding of the model, fuzzy logic is incorporated into the construction of the antecedent. To reduce the search time, the particle swarm is initialized by a competitive neural network. Different variants of PSO are applied to three databases of financial institutions in Ecuador. The first institution specializes in massive credit placement. The second institution specializes in consumer credit and business credit lines. Finally, the third institution is a savings and credit cooperative. According to our results, the incorporation of fuzzy logic generates rule sets with greater precision.

Keywords: particle swarm optimization; fuzzy classification rules; credit risk

JEL Classification: C38; C45; D81

1. Introduction

Economic decisions are closely linked to risk. Several types of risk can affect the survival of financial institutions. Among these risks, we can mention market risk, credit risk, and operational risk. Credit risk is the most common risk and is defined as the probability of loss due to the default of the borrower on the required payments, and its severity is determined by the amount of defaulted debt. Credit risk assessment is the evaluation done by financial institutions on the future ability of borrowers to meet their financial obligations. The variables that must be analyzed will depend on the type of credit to be granted. This situation leads to the collection of many variables for the purpose of analyzing the borrowers. Some variables depend on the personal situation of the customer, whereas others depend on the general economic environment. Thus, it is important to analyze, at the time of granting a loan, the macroeconomic environment in which the client operates, in addition to the borrowers' specific conditions. In this way, a more comprehensive analysis of the debtor can be carried out. The goal is to recommend a more precise answer to the credit official.

Thanks to information technology, numerous processes automatically register their operations, producing large repositories of historical information. These records contain not only the data upon which the decision was made (approval/rejection), but also the repercussions of the decision (default/repayment). Therefore, financial institutions have information about instances of past granted credit and the success or failure of the repayment. Our goal is twofold: (i) to identify and generalize the criteria used to grant credits and (ii) to identify the common features of successes and mistakes.

It is important to indicate that in most countries there are government institutions that oversee financial institutions and enforce suitable regulations. One of these regulations requires financial institutions to be responsible for establishing the methodologies and processes to identify, measure, control, and monitor credit risk. In addition, financial institutions must set a system in place to monitor credit risk levels permanently. In general, these regulations leave some room for discretion. This situation means that each entity, for each credit type (commercial, consumption, housing, and microcredit) selects the best model, within some general principles and criteria, for credit risk assessment. Implemented methodologies can consider the combination of quantitative and qualitative criteria, in accordance with the experience and strategic policies of the firm. Therefore, it is relevant to the financial industry to develop new and better models. Such enhanced models should be based not only on the historical information of the clients but also on other internal and external information in order to produce better risk-controlled decisions.

The CEO of a financial institution aims to increase the volume of operations granted, without affecting global risk decision making, based on expert analytical methodologies that allow the expected loss to be determined based on the probability of default, the level of exposure, and the severity of the loss, which helps to control overall financial risk. Another issue that must be considered is that, in order to attract more customers, the granting process must be quick. Consequently, financial institutions must improve their prediction accuracy, make decisions in a short period of time and provide the credit officer with understandable criteria for the acceptance/rejection of credit applications.

This article proposes the construction of a predictive model, based on the personal and microeconomic information of borrowers, as well as contextual macroeconomic information, in order to aid in the decision-making process of credit granting. Among the different classification techniques existing in the data mining area, classification rules are extremely attractive due to their explanatory capacity (Kotsiantis 2007).

A classification rule is a conditional expression in the form:

IF A THEN B

Where *A* is the antecedent of the rule and *B* is its consequent. The antecedent of the rules used in this article is formed by a conjunction of expressions of the form *attribute = value*. If this conjunction is true, then the rule produces the consequent as a result. This consequent will be of the form (*attribute_to_predict = class_value*).

In the literature, different methods exist for generating classification rules, using either qualitative or quantitative attributes (Fürnkranz and Kliegr 2015). However, the high cardinality of the generated rule set, as well as the rigor with which each antecedent is expressed, leads to the application being impractical. Although different metrics can be applied to identify which rules are the most relevant, the thresholds calculated on the numerical attributes will continue to be rigid. This makes the application of the rules rather difficult.

In this article we propose the application of Particle Swarm Optimization (PSO), a well-known optimization technique, to generate a list of classification rules. We consider different variants of cluster initialization, in order to extend the representation of the solutions. Classification rules operate on fuzzy numerical information, thus avoiding the use of rigid thresholds when determining the conditions that form the background of the rules. This process is carried out using fuzzy attributes, whose values are determined in linguistic terms.

For example, if *Income* is a numeric attribute, its participation in the antecedent of a rule requires setting a range, e.g., (*Entry < 1000*). Instead of using it in this way, the algorithm presented in this

paper proposes to convert it, for example, into a diffuse attribute that takes Low, Medium, and High values and allows them to participate in the conjunction with an expression of the form (*Enter = Low*). Consequently, numerical thresholds will not be used in the precedent of the rule, thus facilitating its understanding by the credit officer. As a result, a set of rules that may or may not be diffuse, of low cardinality, with a background formed by few conditions, and offering acceptable classification accuracy will be obtained.

In this article, we propose: (i) the application of an optimization technique based on particle swarms to generate a list of classification rules, taking into account different variants when initializing the clusters; (ii) to operate on the number of particles that will carry out the search; and (iii) to extend the representation of the solutions found to express fuzzy conditions. The result will be a set of rules that may be fuzzy or crisp, is of low cardinality, has an antecedent formed by few conditions, and offers an acceptable classification accuracy.

This work is organized in the following way: Section 2 briefly describes the relevant related work; Section 3 presents the different PSO variants used; Section 4 shows the results obtained; and Section 5 summarizes the conclusions and describes some future research lines.

2. Related Literature

There are different techniques that allow for improvement of credit scoring models. The seminal work by Altman (1968) opened a research line that began working with statistical methods, including logistic regression and linear discriminants (Mahmoudi and Duman 2015; Zhichao et al. 2017).

After decades of research, machine learning techniques emerged as powerful alternatives. In this line, Support Vector Machines (SVM) (Bellotti and Crook 2009; Harris 2013; Li et al. 2017) and multiperptron neural networks are capable of improving the performance of credit models (Tavana et al. 2018; Zhao et al. 2015). There are hybrid models that combine fuzzy logic with SVM (Wang et al. 2005) or fuzzy logic with optimization techniques (Chen 2006). There are also hybrid models that select the most salient previous features and reduce the dimension of the input space, which in turn enhance the results (Leo et al. 2019; Malhotra and Malhotra 2003; Oreski and Oreski 2014). The authors in Leo et al. (2019) provide an excellent review on the risk management techniques used by financial institutions. In spite of the fact that these machine learning models present good accuracy, they are not considered particularly useful, as explaining the response obtained is difficult (Baesens et al. 2003). In other words, these machine learning models provide an answer that is hard to relate to specific characteristics of the credit or the borrower.

Recently, the authors in Millán-Solarte and Caicedo-Cerezo (2018) compared the results of logistic regression, linear discriminants, neural networks, and decision trees. The classical logistic regression model is frequently used to estimate the probability that the client of a financial institution will default on his or her payments. However, logistic regression alone is not as efficient as other techniques. Consequently, the authors in Millán-Solarte and Caicedo-Cerezo (2018) proposed hybrid models that combine logistic regression with decision trees. Their results were much better. Nevertheless, these types of models do not use fuzzy logic, taking away the advantage in the interpretation of the rules obtained.

In contrast to the previous techniques, classification rules are a widely accepted model when seeking to justify the responses obtained because they formalize, in a clear way, the knowledge discovered. They are considered more natural and understandable, providing experts (in our case, credit officers) with the possibility of analyzing the criteria used when giving an answer.

There are several methods for generating rules, which generally use an incremental criterion to define the composition of the antecedent. During the process, available cases are inspected, and conditions are added to the antecedent so that the rule gains accuracy at the cost of losing support. In this sense, a wide variety of solutions can be found in the literature, such as methods based on pruned trees (e.g., PART algorithm by Frank and Witten (1998)) or specific metrics, such as PRISM

algorithm by [Cendrowska \(1987\)](#). Regardless of the option chosen, the rules obtained tend to have an antecedent with a large average length.

On the other hand, another type of alternative can be found to build classification rules. This is the case for competitive neural networks. Once the neural network is trained, the centroids (represented by the positions of the neurons) can be used to identify the most representative characteristics or attributes, in order to explain patterns or relationships in the input information. This approach has been used as a departing point to obtain classification rules ([Hung and Huang 2010](#); [Pateritsas et al. 2007](#)). However, this type of response does not present good precision, as the characterization obtained is representative of a set of cases in an unsupervised manner. That is, the associations identified by the neural network are equivalent to those that could be obtained with other clustering techniques and do not obey a class distribution ([Reyes et al. 2013](#)).

Support Vector Machines (SVM) can be also used as a starting point for rule generation. However, they are considered “black boxes”, since the generated rules are frequently complex and difficult to understand ([Barakat and Bradley 2010](#); [Núñez et al. 2002](#)). Other methods suitable for rule extraction are population based techniques that allow relationships among items and, as a result, obtain a set of rules ([Gandhi et al. 2010](#); [Kennedy 2010](#); [Wang et al. 2006 2007](#)). These techniques can be defined as processes of search and optimization, which improve the available solutions by means of iterative strategies until achieving an optimal value.

Different strategies based on population optimization techniques have also been proposed to determine sets of rules. In most cases, as in the solutions analyzed in this article, the result is a list of rules. The construction of such a list requires repeated executions of the selected technique. Modeling the extraction of rules as an optimization problem involves defining the representation to be used, the way to measure the performance of individuals throughout the process, and the operators that will guide the search to the best solutions. In [Al-Maqaleh and Shahbazkia \(2012\)](#) a Genetic Algorithm (GA) was used to evolve the conditions that make up the conjunction that determines the antecedent. An alternative proposal, known as Ant Colony Optimization (ACO), is presented in [Medland et al. \(2012\)](#). In the latter case, the results obtained are very good, although the computational time can be prohibitive for very large data sets. ACO performs better with qualitative attributes, while GA achieves better results with numerical attributes. Unfortunately, most of the existing methods cover historical information using large and complex sets of rules that become difficult to understand and consume a large amount of resources ([Carvajal Montealegre 2015](#)).

In previous studies, we used methods such as C4.5 and PART ([Lanzarini et al. 2015, 2017](#)). In the case of C4.5, the method of finding classification rules is through a pruned tree whose branches are mutually exclusive and allow classification of the examples. In the case of PART, a list of rules equivalent to those generated by the proposed classification method is given but in a deterministic way. The operation of PART is based on the construction of partial trees. Each tree is created in a similar way to the one proposed by C4.5; however, during the process, errors are calculated in each expansion and pruning is performed when increments are detected. A detailed description of this algorithm is in [Frank and Witten \(1998\)](#).

Another technique used is Optimization by PSO particle cluster, ([Lanzarini et al. 2015, 2017](#)), which is a search and optimization technique that can be applied in different contexts. PSO is used to select, simultaneously, the conditions where the conjunction will give rise to the antecedent of each rule. The use of PSO in the extraction of rules requires some particular considerations. Although all aspects are important, the definition of the aptitude function, which measures the performance of the rules as they are formed, is the central aspect in achieving an efficient set of rules. If the process of searching for solutions (in this case rules) starts from a position close to the optimum, the time of obtaining it will be considerably reduced. This starting point of the process was solved using a competitive neural network. Additionally, [Shihabudheen and Pillai \(2018\)](#), concludes that methods using PSO achieve more accuracy than gradient-based techniques or SVM.

Although partition algorithms provide greater accuracy, this is achieved through a greater number of rules, which makes understanding more difficult. In fact, the difference in accuracy between both types of methods is within the range of 1 to 3 percentage points. The accuracy of the classification based on PSO is very good and is comparable to the other methods, however, in relation, the number of rules is between 10 and 20 times greater in the methods of partition. More detailed discussion on this feature can be found in [Lanzarini et al. \(2015\)](#) and [Jimbo et al. \(2016\)](#).

It is important to note that although smart systems have several advantages over traditional linear programming or time series calculation techniques, for financial institutions, their use does not replace a credit expert; instead, it provides support so that they can more easily evaluate a loan.

The rules express, through their respective antecedents, conditions that must be met in order to obtain the desired response. These conditions can be relaxed using fuzzy sets ([Zadeh 1965, 1996](#)). Credit classification is a problem, where the knowledge of the credit expert is generally available. Therefore, it is feasible to obtain adequate fuzzy propositions that interpret the rules of the loan officer.

The rationale for using an optimization technique, such as PSO, that allows obtaining fuzzy rules not only makes possible a trade-off between the accuracy and simplicity of the rules ([Bagheri et al. 2014](#)) but also deals with the granting of credits as a more successful task on the part of the loan officers. This is because the risk to the financial institution decreases due to the analysis carried out, leading them to make appropriate decisions with greater accuracy by verifying a reduced number of rules in the shortest time possible.

This article analyzes different techniques, based on particle clusters, capable of obtaining classification rules operating over qualitative and quantitative attributes. Emphasis is placed on the importance of the use of fuzzy sets when expressing the conditions that involve numerical attributes. This setting not only facilitates the understanding and application of the rules but also offers a significantly higher precision *vis-à-vis* crisp conditions.

3. PSO for Rule Extraction

Particle Swarm Optimization (PSO) is a population metaheuristic proposed by [Eberhart and Kennedy \(1995\)](#), where each individual of the population (called a particle) represents a possible solution to the problem and is adapted following three factors: the particle's current knowledge (its ability to solve the problem), its historical knowledge or previous experience (its memory), and the historical knowledge or previous experiences of the individuals located in its neighborhood (social knowledge). To measure the performance of the particle as a solution to the problem, a fitness function is used. This technique works by iteratively improving the fitness value of all the particles of the population by combining the three aforementioned factors. [Algorithm 1](#) contains the pseudocode of a basic PSO algorithm.

In this paper, we analyzed different variants of the PSO model. We applied them in order to obtain classification rules. This is an iterative process that starts with the selection of the class that determines the consequent, as the one that with the largest number of uncovered examples and adapts a population of particles following [Algorithm 1](#). Each individual of the population corresponds to the antecedent of a different rule. Therefore, all the rules determined by the set of particles have the same consequent: the a priori selected class. Upon completion of the adaptation process, the best individual in the population will result in a new rule that will be added to the list. The examples correctly covered by this last rule will be removed from the data set. The process is repeated starting with the selection of the new class with the largest number of examples, for which the most suitable antecedent will be searched. The process ends when enough cases have been covered.

Algorithm 1 Pseudocode of basic PSO method

```

Set initial population1
Determine the initial velocity of particles
while end condition is not met do
  Adjust the velocity of particles
  for each particle of the population do
    Evaluate its fitness
    Register its current location if it is the best solution found so far
  end for
  for each particle of the population do
    Identify the best global solution found within the neighborhood
    Calculate the movement direction of the particle
    Move the particle
    Verify that the particle does not move beyond the search space.
  end for
end while
Return: the best solution found

```

The different variants analyzed in this paper extend the particle representation by combining the continuous PSO, defined by [Eberhart and Kennedy \(1995\)](#) to determine the limits of the numerical attributes with the binary PSO by [Kennedy and Eberhart \(1997\)](#), to select which attributes will be part of the antecedent. In all cases, the neighborhood used to obtain the best solution is global, i.e., the entire population. Regarding the initial position of the particles within the population, and with the aim of reducing the construction time of the rule set, a competitive neural network has been used to identify the most promising areas of the search space. In other words, the algorithm performs, first, a grouping of the examples. Then, taking into account the centroids, particles are placed in order to generate the rules. The competitive neural networks used were Self Organization Map (SOM) and Learning Vector Quantization (LVQ). Both networks are intended to make a grouping of the input data, offering centroids as a result. Each centroid seeks to represent a set of input data based on a previously defined similarity or distance measure. The difference between SOM and LVQ lies on the learning mechanism. Whereas SOM is a unsupervised learning network, LVQ is a supervised learning network. SOM aims to provide additional information regarding the way in which groupings are organized. This is better than any other ‘winner-take-all’ type clustering technique, such as k-means. In contrast, as LVQ uses (during training) information on the expected response, it presents a better fit to the centroids. In this article, we use both architectures, with LVQ as the best performer.

In this way, particles begin the adaptation process within the most promising area of the search space, requiring fewer iterations in order to get a good solution. The size of the population is an important factor in the performance of the algorithm. Therefore, fixed and variable population PSO methods were analyzed. Fixed population PSO defines the population size before starting the search process. Population size definition puts certain limitations on the efficacy and efficiency of the algorithm. On the one hand, if few individuals are used, several areas of the search space will remain unexplored. On the other hand, if too many individuals are used, convergence time will increase.

This inconvenience could be bypassed using the variable population PSO (VarPSO) proposed by [Lanzarini et al. \(2008\)](#). This method allows one to control the number of particles throughout the adaptive process, using concepts such as lifetime and neighborhood. VarPSO eliminates the necessity to define beforehand the quantity of solutions to be used. The population size variation is based on a modification of the adaptive process allowing the addition and/or elimination of individuals based on their ability to solve the problem. This is mainly done through the concept of life time, which allows the determination of the time that each element belongs to the population. PSO tends to quickly populate

the explored areas with good fitness. To avoid overpopulation over a restricted area, each individual's neighborhood is analyzed, and the worst solutions of very populated areas are eliminated.

Among the variants analyzed using PSO, the best results were obtained with Fuzzy Rules Variable Particle Swarm Optimization (FRvarPSO) that uses a cluster of particles of variable size, initialized through a competitive neural network and incorporating fuzzy attributes to express the conditions related to numerical variables. The goal of this method is to obtain a low cardinality and easy to interpret set of classification rules with adequate accuracy. We consider two important aspects. First, the ability of the method to operate with fuzzy attributes. Second, the addition of information based on membership degrees both in the fitness evaluation and in the way the search is carried out through the optimization technique.

The i -th particle of the population (for models FRvarPSO) is represented in the following way:

- $pBin_i = (pBin_{i1}, pBin_{i2}, \dots, pBin_{in})$ is a binary vector that stores the current position of the particle and indicates which items or conditions make up the antecedent of the rule according to PSO.
- $v1_i = (v1_{i1}, v1_{i2}, \dots, v1_{in})$ and $v2_i = (v2_{i1}, v2_{i2}, \dots, v2_{in})$ are combined to determine the direction towards which the particle will move.
- $pBestBin_i = (pBestBin_{i1}, pBestBin_{i2}, \dots, pBestBin_{in})$ stores the best solution found by the particle so far.
- $fitness_i$ is the fitness value of individual i .
- $fitness_pBest_i$ is the fitness value of the best local solution found (vector $pBestBin_i$).
- $L_i = (L_{i1}, L_{i2}, \dots, L_{in})$ is a binary vector that indicates the possible values taken by each linguistic variable.
- $Gp_i = (Gp_{i1}, Gp_{i2}, \dots, Gp_{in})$ is a vector of real values that stores the average membership degrees of the examples that meet the rule, for each value of the linguistic variable.
- $v3_i = (v3_{i1}, v3_{i2}, \dots, v3_{in})$ indicates the change in direction of L_i , with membership degree Gp_i .
- $pBestGp_i = (pBestGp_{i1}, pBestGp_{i2}, \dots, pBestGp_{in})$ stores the best solution found by the particle for the membership degrees of the linguistic variable.
- $sopBin_i = (sopBin_{i1}, sopBin_{i2}, \dots, sopBin_{in})$ indicates which items or conditions make up the rule's antecedent that actually represent the particle, and whose fitness is in $fitness_i$.
- TV is an integer number that indicates the remaining life time of the particle. This parameter is only used in variable population models.

The movement of the i -th particle is controlled using a variant of PSO directed by the velocity vectors $v1_i$ and $v2_i$, where $pBin_i$ is the result of applying the sigmoid function. The binary individual that selects the conditions that make up the antecedent of the rule is expressed in $sopBin_i$ and emerges from $pBin_i$ after removing the invalid solutions. In order to decide the value with which each linguistic variable may participate in the condition, vector Gp_i is added according to the average membership degrees of each linguistic variable in the different fuzzy sets. This average is calculated according to the membership degrees of the examples that comply with the antecedent of the rule when evaluating fitness. This vector is the one used to modify the velocity vector $v3_i$, as indicated in the real PSO version. Then, L_i is the result of applying the sigmoid function to $v3_i$. Thus, it is a binary vector that indicates the possible values that each linguistic variable, if selected, can take. The fitness function of the PSO version that gives the best result, and whose representation was recently described, is the following.

$$Fitness = support \times confidence \times factor1 - factor2 \times \frac{NumAtribsAntecedent}{MaxAtribs} \quad (1)$$

where *support* is a numerical value that measures the representativeness of the rule, and *confidence* is another numerical value that represents the accuracy. *Support* is calculated as the ratio of the number of cases or credit applications that the rule correctly predicts to the total number of studied cases. *Confidence* is the quotient between the number of times the rule response is correctly divided by the

number of times the rule is applied. In both cases, these are numerical values belonging to the interval $[0, 1]$. If a rule has 0 support, that rule is incorrect in all cases: No credit application fulfills the rules, either because it does not verify the antecedent, or the consequent, or both. A rule with confidence 1 has always responded correctly. *factor1* is a penalty value in case the support is not within the ranges established in the algorithm. The second term in the fitness function reflects the importance given to the number of attributes included in the antecedent, and *factor2* is a constant.

Each of the variants presents specific details. In the case of LVQ+PSO a full description is in Lanzarini et al. (2015) and Lanzarini et al. (2017). In the case of FRvarPSO, a detailed analysis is found in Jimbo et al. (2018) and Jimbo et al. (2018).

The main difference of this work compared to Jimbo et al. (2018), is the detail obtained with the three optimization techniques used: (i) basic PSO, (ii) variable population PSO, and (iii) variable population PSO with fuzzy rules. Additionally, in this work we detail the variables that are used, emphasizing role of fuzzy variables in credit scoring problems.

4. Data and Results

This paper uses data from three financial institutions from Ecuador, regarding credit operations between January 2012 and December 2016. We analyze each institution separately. The data set includes not only personal and economic information about the clients, but also the history of more than granted 129,000 credit applications. Granted loans have several associated characteristics at the personal, microeconomic, and macroeconomic levels.

We analyze three data structures. The first one is related to credit subjects. This structure includes information from the borrowers, guarantors, and co-debtors of the credit and contingent operations, as well as the cardholders with overdue balances. The second structure includes all operations granted. This structure includes credit and contingent operations that have been granted, renewed, refinanced, or restructured each month. The last structure is related to transaction balances. It is updated monthly and includes the details of the balances of credit and contingent operations that are still active, indicating overdue operations and detailing the number of days.

In the case of microeconomic variables, the behavior of the client can be identified through three types of variables:

1. *Descriptive variables*: These define the general characteristics of the client, such as age, province of residence, economic activity, equity, gender, marital status, level of education, occupation, housing ownership, family responsibilities, income, and expenses.
2. *Credit history*: This includes a summary of the customer's previous operations. A representative list of variables is: quantity of previous credit operations, amount of such operations, interest rate, lines of credit, fulfillment of previous payments, collateral, and purpose of the loan.
3. *Characterization of the loan*: This is information about the repayment of previous loans. Loan unperformance is classified according to the length of late payment (30, 90, 180, 360, and +360 days).

The advantage of this data analysis is that the credit subject's behavior over a time horizon is reviewed, including the customer's past behavior.

In order to reduce the cardinality of the input data set, we conduct a correlation analysis in order to eliminate less representative variables. In order to reduce the cardinality of the input data set, we filter possible input variables by means of correlational analysis. The correlation matrix measures the relationships between pairs of variables. When the value of the relationship approaches one the variables, they are strongly related. On the contrary, when the value of the relationship is close to zero, the relationship is weaker. This allows us to eliminate the least representative variables. Several fields were also unified to include the income, expenses, and debts of the credit subject. The variables that were considered after the transformation were: year, month, province, destination of the loan,

overdue days, value of the operation, cash and equivalent, total income, total assets, total expenses, and total debts.

Among the macroeconomic variables studied were: the consumer price index (CPI), Unified Basic Remuneration (UBR), and the stock market price of Corporación Favorita¹. Additionally, we took into account the type of credit, province, and date of the credit.

In addition to the value of the operation, we introduce, in our model, the following variables regarding the loan applicant's financial situation: cash and equivalents, total income, total assets, total expenses, and total debts. These are not considered as numerical variables, but as linguistic ones. For example, the borrower's income can be considered as "low", "medium", or "high", with a degree of membership between 0 and 1. This assigned value, with its degree of membership was granted by the credit expert, according to the economic reality of Ecuador. For example, if low income is considered for the range of [100, 1200], the variable $Income = 1000$, would be defined as $Income = low$ with a membership degree of 0.8.

Fuzzy membership functions are considered triangular. The values of fuzzy attributes were determined based on Unified Basic Remuneration (UBR), that is, the current value of the basic salary for each year. Figure 1 shows the membership function of the variables "value of the operation" and "cash and equivalents", related to the data of the Savings and Credit Cooperative². In the case of the variable, "value of the operation" (in dollars), we defined three ranges of values: low, for operations in the interval [0, 25000]; medium, for operations in the interval [20000, 50000]; and high for operations in the interval [40000, 80000]. In the case of the variable "cash and equivalents", the intervals are: Low [0, 10000], Medium [8000, 35000], and High [30000, 90000]. The fuzzy sets were defined with the help of an expert in the area of credit risk, according to the economic situation of Ecuador. For a detailed review of how fuzzy sets works, see Klir and Yuan (1995).

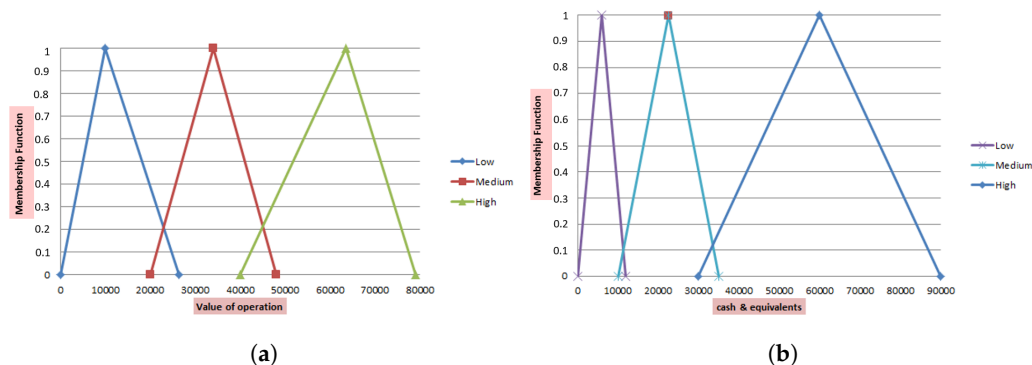


Figure 1. Triangular membership function of the variables (a) 'value of operation' and (b) 'cash and equivalents'.

The proposed FRvarPSO method was compared with the performance of several methods that combine fixed and variable population PSO, initialized with two different competitive neural networks: LVQ and SOM (Kohonen 2012).

The result of the classification rules of the fuzzy models also indicates the level of risk in granting the credit. The antecedent of the classification rules is formed by nominal and/or linguistic variables, facilitating the interpretation by the credit officer.

¹ This is the largest non-state company in Ecuador (<http://www.corporacionfavorita.com/>). Corporación Favorita is a business group comprising several business units such as grocery stores, home appliances stores, real estate, industrial companies, etc. In 2017 the corporate equity was worth USD 1197 million, and the sales were USD 1906 million. The company's stock price behavior is seen by economic specialists as an indicator of the country's economic strength.

² This is an institution classified as segment 2 by the regulation authority (Superintendencia of Popular and Solidarity Economy), with assets in the range of \$20 million–\$80 million

Classification rules are characterized in that their consequent, also called the conclusion of the rule, identifies the value of the class that applies to the examples covered by those rules. Therefore, the consequent of all the rules that make up the set of rules must refer to the same attribute or characteristic, which corresponds to the expected response. In the case of our credit granting problem, the consequent refers to the granting or denial of credit. An example of a classification rule applied to credit risk is:

IF (Income is high) AND (Expenses is medium) AND (delinquency is low) THEN (grant the credit).

We performed 25 independent runs for each method and computed the mean and standard deviations of the classification accuracy and the resulting number of rules. One important feature of our method is that the algorithm uses random values in order to ensure the movement of the particle is not excessively deterministic. The most important characteristic of this paper's results is the accuracy improvement with respect to what is reported in Lanzarini et al. (2008). This accuracy improvement is due to set of fuzzy classification rules, which stems from micro and macroeconomic variables.

The performance of the methods SOM+PSO, SOM+varPSO, LVQ+PSO, and LVQ+varPSO, and their fuzzy versions SOM+Fuzzy PSO, SOM+Fuzzy varPSO, LVQ+Fuzzy PSO, and LVQ+Fuzzy varPSO, is detailed in Table 1 and graphically displayed in Figure 2. We observe that for all three financial institutions, the method FRvarPSO exhibits the best accuracy with the lowest number of rules. In other words, FRvarPSO offers the most straightforward model for any level of accuracy. As a consequence, this provides the credit analyst easier arguments to understand the reason for granting or denying credit.

Table 1. Comparison of the accuracy and number of rules of hybrid and fuzzy models considered in this paper.

Database	Method	Accuracy	# Rules
Savings and credit cooperative	SOM+PSO	0.7718	6.2966
	SOM+varPSO	0.7701	5.6899
	LVQ+PSO	0.7871	5.4772
	LVQ+varPSO	0.7894	5.3535
	SOM+fuzzy PSO	0.7857	5.9874
	SOM+fuzzy varPSO	0.7891	5.6920
	LVQ+fuzzy PSO	0.7922	5.4958
	FRvarPSO	0.7988	5.2990
Bank in Ecuador dedicated to massive microcredits	SOM+PSO	0.9688	7.4939
	SOM+varPSO	0.9700	7.6936
	LVQ+PSO	0.9737	7.6993
	LVQ+varPSO	0.9778	7.3894
	SOM+fuzzy PSO	0.9819	6.9930
	SOM+fuzzy varPSO	0.9840	6.1888
	LVQ+fuzzy PSO	0.9869	6.7588
	FRvarPSO	0.9880	6.3972
Bank in Ecuador dedicated to retail consumer's credit and firm loans	SOM+PSO	0.8133	8.4994
	SOM+varPSO	0.8264	8.2951
	LVQ+PSO	0.8256	6.4926
	LVQ+varPSO	0.8315	6.4926
	SOM+fuzzy PSO	0.8316	6.4415
	SOM+fuzzy varPSO	0.8385	6.3990
	LVQ+fuzzy PSO	0.8455	6.6904
	FRvarPSO	0.8501	5.9901

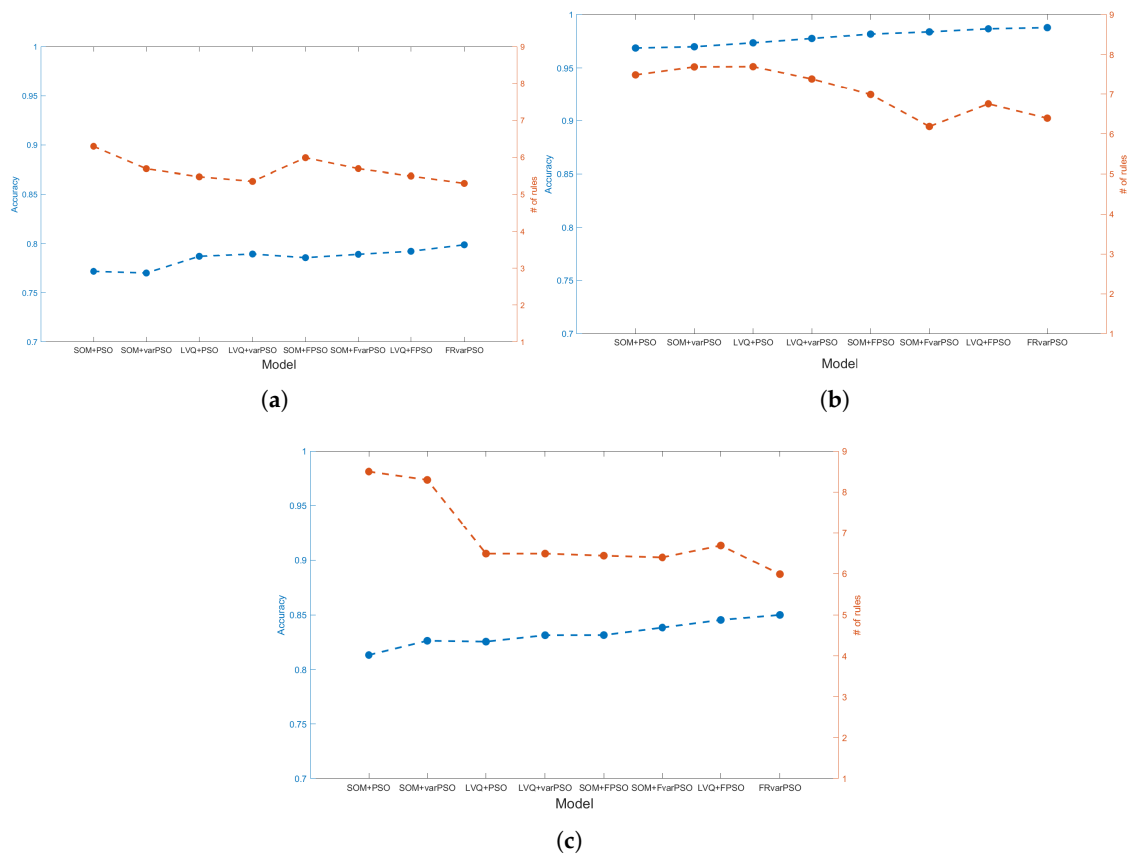


Figure 2. Comparison of the accuracy and number of rules of the different models used in this paper. (a) Savings and credit cooperative; (b) Bank dedicated to massive microcredits; (c) Bank dedicated to retail consumer’s credit.

5. Conclusions

In the present article, we carried out a comparative analysis of several classification methods. We applied these methods to real credit risk analysis, based on the combination of a competitive neural network (either SOM or LVQ) and an optimization technique (either PSO or varPSO). We compared the performance of crisp and fuzzy versions of the models. The results obtained in the fuzzy versions were satisfactory.

We detect that the proposed FRvarPSO method combines very good precision with a low cardinality rule set. A financial institution could use this technique in order to manage their credit risk. At the same time, this could help credit officers improve their understanding of the decision model, as the rules are formed by fuzzy variables. This situation could improve the solvency of financial institutions, as well as enhance the transparency in loan granting decisions.

The use of a fuzzy logic approach in a credit scoring problem, in order to reduce the risk in granted loans, reinforces the viability and sustainability of microfinance institutions. Fuzzy logic is useful where there is uncertainty, as in the case of granting credit. Working variables are characterized by their conceptual vagueness and imprecision. Their treatment as fuzzy variables is suitable and it is fully justified. The use of linguistic terms contextualizing observed values through membership functions allows the expression and interpretation of rules in a more reasonable way.

Considering that FRvarPSO is a hybrid method that uses fuzzy logic, neural networks, and optimization techniques, it represents an alternative with greater performance than those models that use a single technique, such as linear discriminant methods, logistic regression methods, or neural networks. Additionally, our method has the advantage of allowing the identification of the risk in the granting of credit, with considerable precision.

The empirical application of this paper was conducted in the financial market of Ecuador. However, our results can be extrapolated to other markets (not only in Latin America), as credit scoring relies not only on the customer's financial health, but on the overall strength of the economy. Future research will include a deeper study on the methods for optimizing the variables membership functions. This could lead to an improvement in the FRvarPSO model. Studies could also consider the incorporation of a defuzzification procedure for the output variable, to indicate the percentage of risk that is taken by granting a given credit. This measurement will provide the credit committee with strategic information on the consequences of lowering the threshold for granting credit.

Author Contributions: Conceptualization, L.L. and P.J.S.; methodology, L.L. and P.J.S.; software, L.L. and P.J.S.; formal analysis, A.F.B., L.L. and P.J.S.; investigation, A.F.B., L.L. and P.J.S.; data curation, L.L. and P.J.S.; writing and editing, A.F.B., L.L. and P.J.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Conflicts of Interest: The authors declare no conflict of interest.

References

- Al-Maqaleh, Basheer M., and Hamid Shahbazkia. 2012. A genetic algorithm for discovering classification rules in data mining. *International Journal of Computer Applications* 41: 40–44. [\[CrossRef\]](#)
- Altman, Edward I. 1968. Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. *The Journal of Finance* 23: 589–609. doi:10.1111/j.1540-6261.1968.tb00843.x. [\[CrossRef\]](#)
- Baesens, Bart, Tony Van Gestel, Stijn Viaene, Maria Stepanova, Johan Suykens, and Jan Vanthienen. 2003. Benchmarking state-of-the-art classification algorithms for credit scoring. *Journal of the Operational Research Society* 54: 627–35. doi:10.1057/palgrave.jors.2601545. [\[CrossRef\]](#)
- Bagheri, Ahmad, Hamed Mohammadi Peyhani, and Mohsen Akbari. 2014. Financial forecasting using anfis networks with quantum-behaved particle swarm optimization. *Expert Systems with Applications* 41: 6235–50. doi:10.1016/j.eswa.2014.04.003. [\[CrossRef\]](#)
- Barakat, Nahla, and Andrew P. Bradley. 2010. Rule extraction from support vector machines: A review. *Neurocomputing* 74: 178–90. doi:10.1016/j.neucom.2010.02.016. [\[CrossRef\]](#)
- Bellotti, Tony, and Jonathan Crook. 2009. Support vector machines for credit scoring and discovery of significant features. *Expert Systems with Applications* 36(Pt 2): 3302–8. doi:10.1016/j.eswa.2008.01.005. [\[CrossRef\]](#)
- Carvajal, Montealegre, and Carlos Javier. 2015. Extracción de reglas de clasificación sobre repositorio de incidentes de seguridad informática mediante programación genética. *Tecnura* 19: 109. [\[CrossRef\]](#)
- Cendrowska, Jadzia. 1987. Prism: An algorithm for inducing modular rules. *International Journal of Man-Machine Studies* 27: 349–70. doi:10.1016/S0020-7373(87)80003-2. [\[CrossRef\]](#)
- Chen, Chia-Chong. 2006. A pso-based method for extracting fuzzy rules directly from numerical data. *Cybernetics and Systems* 37: 707–23. doi:10.1080/01969720600886980. [\[CrossRef\]](#)
- Eberhart, R., and J. Kennedy. 1995. A new optimizer using particle swarm theory. Paper presented at Sixth International Symposium on Micro Machine and Human Science, MHS'95, Nagoya, Japan, October 4–6, pp. 39–43. doi:10.1109/MHS.1995.494215. [\[CrossRef\]](#)
- Frank, Eibe, and Ian H. Witten. 1998. Generating accurate rule sets without overall optimization. Paper presented at ICML '98 the Fifteenth International Conference on Machine Learning, San Francisco, CA, USA, July 24–27.
- Fürnkranz, Johannes, and Tomáš Kliegr. 2015. A brief overview of rule learning. In *Rule Technologies: Foundations, Tools, and Applications*. Edited by Nick Bassiliades, Georg Gottlob, Fariba Sadri, Adrian Paschke and Dimitru Roman. Cham: Springer International Publishing, pp. 54–69.
- Gandhi, Kalji R., Marcus Karnan, and Senthamarai K. Kannan. 2010. Classification rule construction using particle swarm optimization algorithm for breast cancer data sets. Paper presented at 2010 International Conference on Signal Acquisition and Processing, Bangalore, India, February 9–10, pp. 233–37. doi:10.1109/ICSAP.2010.58. [\[CrossRef\]](#)
- Harris, Terry. 2013. Quantitative credit risk assessment using support vector machines: Broad versus narrow default definitions. *Expert Systems with Applications* 40: 4404–13. doi:10.1016/j.eswa.2013.01.044. [\[CrossRef\]](#)

- Hung, Chihli, and Lynn Huang. 2010. Extracting rules from optimal clusters of self-organizing maps. Paper presented at 2010 Second International Conference on Computer Modeling and Simulation, Bali, Indonesia, September 28–30, vol. 1, pp. 382–86. doi:10.1109/ICCMS.2010.92. [CrossRef]
- Jimbo, Patricia, Laura Lanzarini, and Aurelio F. Bariviera. 2018. Extraction of knowledge with population-based metaheuristics fuzzy rules applied to credit risk. In *Advances in Swarm Intelligence*. Edited by Yin Tan, Yuhui Shi and Qirong Tang. Cham: Springer International Publishing, pp. 153–63.
- Jimbo, Patricia, Lanzarini Laura, and Bariviera Aurelio F. 2018. Fuzzy credit risk scoring rules using frvarpso. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems* 26: 39–57. doi:10.1142/S0218488518400032. [CrossRef]
- Jimbo, Patricia, Augusto. Villa Monte, Enzo. Rucci, Lanzarini Laura., and Aurelio F. Bariviera. 2016. An exploratory analysis of methods for extracting credit risk rules. In *XXII Congreso Argentino de Ciencias de la Computación (CACIC 2016)*. San Luis: Nueva Editorial Universitaria, pp. 834–41.
- Kennedy, James. 2010. Particle swarm optimization. In *Encyclopedia of Machine Learning*. Edited by Claude Sammut and Geoffrey I. Webb. Boston: Springer, pp. 760–66. doi:10.1007/978-0-387-30164-8_630. [CrossRef]
- Kennedy, James., and Russell. C. Eberhart. 1997. A discrete binary version of the particle swarm algorithm. Paper presented at 1997 IEEE International Conference on Systems, Man, and Cybernetics. Computational Cybernetics and Simulation, Orlando, FL, USA, October 12–15, vol. 5, pp. 4104–8. doi:10.1109/ICSMC.1997.637339. [CrossRef]
- Klir, George J., and Bo Yuan. 1995. *Fuzzy Sets and Fuzzy Logic: Theory and Applications*. Upper Saddle River: Prentice Hall PTR.
- Kohonen, Teuvo. 2012. *Self-Organizing Maps*. Springer Series in Information Sciences. Berlin/Heidelberg: Springer.
- Kotsiantis, Sotiris. B. 2007. Supervised Machine Learning: A Review of Classification Techniques. In *Emerging Artificial Intelligence Applications in Computer Engineering*, Ebook ed. Edited by Iias Maglogiannis, Kostas Karpouzis, Manolis Wallace and John Soldatos. Amsterdam: IOS Press, vol. 160, Chapter 1, pp. 3–24.
- Lanzarini, Laura, Victoria Leza, and Armando De Giusti. 2008. Particle swarm optimization with variable population size. In *Artificial Intelligence and Soft Computing—ICAISC 2008*. Edited by Leszek Rutkowski, Ryszard Tadeusiewicz, Lotfi A. Zadeh and Jacek M. Zurada. Berlin/Heidelberg: Springer, pp. 438–49.
- Lanzarini, Laura, Augusto Villa-Monte, Aurelio Fernández-Bariviera, and Patricia Jimbo-Santana. 2015. Obtaining classification rules using lvq+pso: An application to credit risk. In *Scientific Methods for the Treatment of Uncertainty in Social Sciences*. Edited by Jaime Gil-Aluja, Antonio Terceño-Gómez, Joan C. Ferrer-Comalat, José M. Merigó-Lindahl and Salvador Linares-Mustarós. Cham: Springer International Publishing, pp. 383–91.
- Lanzarini, Laura Cristina, Augusto Villa Monte, Aurelio F. Bariviera, and Patricia Jimbo Santana. 2017. Simplifying credit scoring rules using lvq + pso. *Kybernetes* 46: 8–16. doi:10.1108/K-06-2016-0158. [CrossRef]
- Leo, Martin, Suneel Sharma, and Koilakuntla Maddulety. 2019. Machine learning in banking risk management: A literature review. *Risk MDPI* 7: 29. doi:10.3390/risks7010029. [CrossRef]
- Li, Zhiyong, Ye Tian, Ke Li, Fanyin Zhou, and Wei Yang. 2017. Reject inference in credit scoring using semi-supervised support vector machines. *Expert Systems with Applications* 74: 105–14. doi:10.1016/j.eswa.2017.01.011. [CrossRef]
- Mahmoudi, Nader, and Ekrem Duman. 2015. Detecting credit card fraud by modified fisher discriminant analysis. *Expert Systems with Applications* 42: 2510–16. doi:10.1016/j.eswa.2014.10.037. [CrossRef]
- Malhotra, Rashmi, and Davinder K. Malhotra. 2003. Evaluating consumer loans using neural networks. *Omega* 31: 83–96. doi:10.1016/S0305-0483(03)00016-1. [CrossRef]
- Medland, Matthew, Fernando E. B. Otero, and Alex A. Freitas. 2012. Improving the cant-minerpb classification algorithm. In *Swarm Intelligence*. Edited by Marco Dorigo, Mauro Birattari, Christian Blum, Anders L. Christensen, Andries P. Engelbrecht, Roderich Groß and Thomas Stützle. Berlin/Heidelberg: Springer, pp. 73–84.
- Millán-Solarte, Julio César, and Edinson Caicedo-Cerezo. 2018. Modelos para otorgamiento y seguimiento en la gestión del riesgo de crédito. *Revista de Métodos Cuantitativos para la Economía y la Empresa* 25: 23–41.
- Núñez, Haydemar, Cecilio Angulo, and Andreu Catalá. 2002. Rule extraction from support vector machines. Paper presented at ESANN'2002 proceedings - European Symposium on Artificial Neural Networks, Bruges, Belgium, April 24–26, pp. 107–112.

- Oreski, Stjepan, and Goran Oreski. 2014. Genetic algorithm-based heuristic for feature selection in credit risk assessment. *Expert Systems with Applications* 41 (Pt 2): 2052–64. doi:10.1016/j.eswa.2013.09.004. [CrossRef]
- Pateritsas, Christos, Stylianos Modes, and Andreas Stafylopatis. 2007. Extracting rules from trained selforganizing maps. Paper presented at IADIS International Conference on Applied Computing, Salamanca, Spain, February 18–20. Edited by Nuno Guimarães and Pedro Isaías, pp. 183–90.
- Reyes, Jorge, Antonio Morales-Esteban, and Francisco Martínez-Álvarez. 2013. Neural networks to predict earthquakes in Chile. *Applied Soft Computing* 13: 1314–28. doi:10.1016/j.asoc.2012.10.014. [CrossRef]
- Shihabudheen, K.V., and Gopinatha N. Pillai. 2018. Recent advances in neuro-fuzzy system: A survey. *Knowledge-Based Systems* 152: 136–62. doi:10.1016/j.knosys.2018.04.014. [CrossRef]
- Tavana, Madjid, Amir-Reza Abtahi, Debora Di Caprio, and Maryam Poortarigh. 2018. An artificial neural network and bayesian network model for liquidity risk assessment in banking. *Neurocomputing* 275: 2525–54. doi:10.1016/j.neucom.2017.11.034. [CrossRef]
- Wang, Yongqiao, Shouyang Wang, and Kin Keung Lai. 2005. A new fuzzy support vector machine to evaluate credit risk. *IEEE Transactions on Fuzzy Systems* 13: 820–31. doi:10.1109/TFUZZ.2005.859320. [CrossRef]
- Wang, Ziqiang, Xia Sun, and Dexian Zhang. 2006. Classification rule mining based on particle swarm optimization. In *Rough Sets and Knowledge Technology*. Edited by Guo-Ying Wang, James F. Peters, Andrzej Skowron and Yiyu Yao. Berlin/Heidelberg: Springer, pp. 436–41.
- Wang, Ziqiang, Xia Sun, and Dexian Zhang. 2007. A pso-based classification rule mining algorithm. In *Advanced Intelligent Computing Theories and Applications. With Aspects of Artificial Intelligence*. Edited by De-Shuang Huang, Laurent Heutte and Marco Loog. Berlin/Heidelberg: Springer, pp. 377–84.
- Zadeh, Lotfi A. 1965. Fuzzy sets. *Information and Control* 8: 338–53. doi:10.1016/S0019-9958(65)90241-X. [CrossRef]
- Zadeh, Lotfi A. 1996. Fuzzy logic = computing with words. *IEEE Transactions on Fuzzy Systems* 4: 103–11. doi:10.1109/91.493904. [CrossRef]
- Zhao, Zongyuan, Shuxiang Xu, Byeong Ho Kang, Mir Md Jahangir Kabir, Yunling Liu, and Rainer Wasinger. 2015. Investigation and improvement of multi-layer perceptron neural networks for credit scoring. *Expert Systems with Applications* 42: 3508–16. doi:10.1016/j.eswa.2014.12.006. [CrossRef]
- Zhichao, Jin, Guo Lili, and Gao Daqi. 2017. Advanced pseudo-inverse linear discriminants for the improvement of classification accuracies. Paper presented at 2017 International Joint Conference on Neural Networks (IJCNN), Anchorage, AK, USA, May 14–19, pp. 25–30. doi:10.1109/IJCNN.2017.7965831. [CrossRef]



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).