

A New Approach for Labeling the Class of Bank Credit Customers via Classification Method in Data Mining

Hamid Eslami Nosratabadi, Sanaz Pourdarab and Ahmad Nadali

Abstract—ONE of the most important parts in credit scoring is determining the class of customers to run the Data Mining algorithms. The purpose of this research is estimating the Label of Credit customers via Fuzzy Expert System. Here the class of customers has been specified by a Fuzzy Expert System and then the Data Mining Algorithms have been run on the final data with Clementine software. The presented steps have been studied in an Iranian Bank as empirical study.

Index Terms—Data Mining, Fuzzy Expert System, Credit Scoring, Labeling, Clementine.

I. INTRODUCTION

There is various conventional loan assessment approaches where the criteria used vary from one financial establishment to another. The structure of the evaluating algorithms used in conventional loan assessment systems provide a result that relays on statistical data. Two of the most commonly used statistical techniques are Linear Discriminant Analysis and Logistic Regression [1]. These techniques leads to an uncertainty described as the classification problem which should be improved [2]. Credit scoring models have been widely used by financial institutions to determine if loan customers belong to either a good applicant group or a bad applicant group .The advantages of using credit scoring models can be described as the benefit from reducing the cost of credit analysis, enabling faster credit decision, insuring credit collections, and diminishing possible risk[3,4]. Since an improvement in accuracy of a fraction of a percent might translate into significant savings [4], a more sophisticated model should be proposed to significantly improve the accuracy of the credit scoring model in this paper. In order to obtain a satisfied credit scoring model, numerous methods have been proposed. Roughly, these methods can be classified to parametric statistical methods (e.g. discriminant analysis and logistic regression), non-parametric statistical methods (e.g. k nearest neighbor and decision trees), and soft-computing approaches (e.g. artificial neural network (ANN) and rough sets). Genetic programming (GP) is used in

a survey to build credit scoring models. It was concluded that GP can provide better performance than other models [2]. Fuzzy Adaptive Network (FAN) has been used in another paper to model the credit rating of small financial enterprises. The main advantages of the proposed network are the ability for linguistic representation, linguistic aggregation and the learning ability of the neural network [5]. There is another paper which considers the question of whether inaccurate self-assessment of credit is associated with undesirable financial market outcomes [6]. Another one has suggested the design a N-tier model integrated with the idea of Model-View-Controller to help to finalize the scoring models quickly and easily alter the embedded scoring models later[7]. For managing credit risk, another paper upgrades the quantitative analysis used in the financial performance modules of state-of-the-art credit scoring methodologies[8]. There is a summary for the various quantitative methods such as Markov chains, discriminant analysis, decision trees, expert systems, the various mathematical programming methods and neural networks used in credit management. They discussed applications of neural networks to corporate credit decisions and fraud detection[9]. Neural network has been used in three researches which one of them is related to the case of Chase Manhattan Bank to evaluate corporate loan risk and loan approval [10,11] and in one of them the results are benchmarked against the more traditional methods such as discriminant analysis, k nearest neighbor, decision trees, etc[12]. Another study investigates the usefulness of Decision Trees, Neural Networks and Bayesian Belief Networks in the identification of fraudulent financial statements[13]. Three strategies have been used in a survey to construct the hybrid Support Vector Machine (SVM)-based credit scoring models to evaluate the applicant's credit score from the applicant's input features[24]. Another study , proposes an integrated data mining and behavioral scoring model to manage existing credit card customers in a bank which demonstrates that identifying customers by a behavioral scoring model is helpful characteristics of customer [14]. Another survey applies data mining techniques to CRM and gets result that classification and association models are the two commonly used models for data mining in CRM [15].

Here, an integrated Fuzzy expert system with data mining has been used. The reasons to use FIS are as:1) An usual criterion like delay's months or days can not cover the whole matter and considering multi criteria, is more close to real world and judgment of experts and the quality degree of the customer's legal credit is more trustable .2) Multi criteria, in spite of the above mentioned advantages, is more complicated and Fuzzy Expert System can solve this problem. Since there are many different numerical states

First Author, Hamid Eslami Nosratabadi is with the Department of Information Technology Management, Science and Research Branch, Islamic Azad University, Tehran, Iran (Corresponding author's email is Hamideslami.na@gmail.com).

Second Author, Sanaz Pourdarab is with the Department of Information Technology Management, Science and Research Branch, Islamic Azad University, Tehran, Iran (Email:Pourdarab.sanaz@yahoo.com).

Third Author, Ahmad Nadali is with the Department of Information Technology Management, Science and Research Branch, Islamic Azad University, Tehran, Iran (Email:Nadali.ahmad@gmail.com).

which are exist for the three criteria effective on the customer's class, decision making will be more complicated for the experts. According to the experts' ideas which consider the linguistic values for the variables and some rules for declaring the conditions to make decision for labeling the record, FIS is a useful tool to extract the different states existing for each customer from the Model. The outputs are considered as the Fuzzy Labels for customers and can be used in next step. As a whole, In classification problems, like Credit Scoring, we have a database which includes some fields and records and one field will be specified as the Class field which is being predetermined by the use of multi criteria. In other words, It is a multi attributes field and those considered criteria effect on the values of this field. In addition, for any business, for assigning a Class state to each row (Labeling the row), there are two situations: 1) there is a clear definition of the Class field according to the facts and it is being reported as the fact for the record. In fact, There is no need to any extra effort or to experts' views to label the record and the data is being saved like other fields. 2) To determine the type of the class is done according to the different values of some specified criteria. The Class field is a multi attributes field. This research is going to study these types of classification matters and to suggest the solutions. There are some challenges such as: a) Some classification problems are the kind of multi attributes ones. b) The multi attributes classification problems are being faced with many challenges for working with any kind of algorithms and techniques, because of different states for each criterion and this matter also makes vagueness for experts which increases with the enhancement of the number of criteria and the Expert System can overcome to these kinds of problems. c) Any mistake in determination of the class type makes problem for labeling and creates wrong form in the model. So, labeling task is so sensitive issue in classification problems specially the kind of multi criteria. The scope of the current paper, however, is limited to the evaluation of considered Bank¹. Section 2 of this paper gives an explanation about Data Mining & Credit Scoring & FIS as a survey in literature of them. Section 3 presents the research methodology and section 4 provides an illustrative example for considered Bank, while Section 5 argues the relative advantages and the managerial implications of the proposed methodology as the conclusion.

II. LITERATURE REVIEW

A. Data Mining & Credit Scoring

Data Mining (DM) is an iterative process within which progress is defined by discovery, either through automatic or manual methods. DM is most useful in an exploratory analysis scenario in which there are no predetermined notions about what will constitute an "interesting" outcome. The application of Data Mining techniques for financial classification is a fertile research area [13]. As [16] defines data mining as "the process that uses Statistical,

Mathematical, Artificial Intelligence and Machine-Learning techniques to extract and identify useful information and subsequently gain knowledge from large databases". [13] Also provide a similar definition regarding data mining as being the process of extracting or detecting hidden patterns or information from large databases. With comprehensive customer data, data mining technology can provide business intelligence to generate new opportunities. Each data mining technique can perform one or more of the following types of data modeling: Association, Classification, Clustering, Forecasting, Regression, Sequence discovery, Visualization. The above seven models cover the generally mentioned data mining models in various articles. There are numerous machine learning techniques available for each type of data mining model. Choices of data mining techniques should be based on the data characteristics and business requirements. Here are some examples of some widely used data mining algorithms: Association rule, Decision tree, Genetic algorithm, Neural networks, K-Nearest neighbor, Linear/logistic regression. Association aims to establishing relationships between items which exist together in a given record. Common tools for association modeling are statistics and Apriori algorithms. Classification is one of the most common learning models in data mining. Common tools used for classification are neural networks, decision trees and if then-else rules. Clustering is the task of segmenting a heterogeneous population into a number of more homogenous clusters. Common tools for clustering include neural networks and discrimination analysis. Forecasting estimates the future value based on a record's patterns. It deals with continuously valued outcomes. Common tools for forecasting include neural networks and survival analysis. Regression is a kind of statistical estimation technique used to map each data object to a real value provide prediction value. Common tools for regression include linear regression and logistic regression. Sequence discovery is the identification of associations or patterns over time. Common tools for sequence discovery are statistics and set theory. Visualization refers to the presentation of data so that users can view complex patterns. It is used in conjunction with other data mining models to provide a clearer understanding of the discovered patterns or relationships. Examples of visualization model are 3D graphs, "Hygraphs" and "SeeNet" [13].

A credit score is a numerical expression based on a statistical analysis of a person's credit files, to represent the creditworthiness of that person. A credit score is primarily based on credit report information typically sourced from credit bureaus. Lenders, such as banks and credit card companies, use credit scores to evaluate the potential risk posed by lending money to consumers and to mitigate losses due to bad debt. Lenders use credit scores to determine who qualifies for a loan, at what interest rate, and what credit limits. Lenders also use credit scores to determine which customers are likely to bring in the most revenue. The use of credit or identity scoring prior to authorizing access or granting credit is an implementation of a trusted system. There are three popular models used in building credit scoring models. The first model is logistic regression, which is mostly used for classification problems in the area of

¹ Since the information of considered bank is confidential, The Authors have not been authorized to present the name of considered Bank.

statistics. The second model is ANN, which is known for its excellent ability of learning non-linear relationships in a system. The third model is rough sets, which is one kind of induction based algorithms, and has been widely used in classification problems since 1990s [2]. Logistic regression model is one of the most popular statistical tools for classification problems and is more suitable for the credit scoring problems.[2] Artificial neural networks were developed to mimic the neurophysiology of the human brain to be a type of flexible non-linear regression, discriminant, and clustering models[2] Rough sets is a mathematical tool used to deal with vagueness or uncertainty Compared to fuzzy sets[2]

B. Fuzzy EXPERT System

A fuzzy expert system is simply an expert system that uses a collection of fuzzy membership functions and rules, instead of Boolean logic, to reason about data The fuzzy Inference Systems (FIS) are very good tools as they hold the nonlinear universal approximation. They are suitable to handle experimental data as well as a priori knowledge on the unknown solution, which is expressed by inferential linguistic rules in the form IF THEN whose antecedents and consequents utilize fuzzy sets instead of crisp numbers. Fuzzy inference systems can express human expert knowledge and experience by using fuzzy inference rules represented in “if-then” statements. The fuzzy inference process has five steps: Fuzzify inputs, apply fuzzy operator, apply implication method, aggregate all outputs and Defuzzify. In order to obtain a good FIS it is necessary that the researchers possess domain knowledge; the knowledge has to be represented in a symbolic form, be complete, correct and consistent [17]. In other words, Fuzzy Inference System incorporates fuzzy inference and rule-based expert systems. Fuzzy inference in this system refers to the use of computer programs to execute inference work resembling what humans do daily. The input constitutes some ambiguous linguistic semantics or unclear concepts for a specific event. Following the fuzzy inference mechanism, the output can be a fuzzy set or a precise set of certain features. Fuzzy inference infers the results from the existing knowledge base. 1) Fuzzy concept base: This contains the terminology and relevant predicate of a linguistic expression. Terminology is in the domain of the fuzzy set, possesses many pre-defined dismemberment values denoted by predicates. 2) Fuzzy proposition base: Membership functions accrue to the fuzzy proposition, which was induced from fuzzy concept base. There are numerous types of membership functions, such as S-shape, Z-shape, and P-shape, all easily definable with equations and parameters. For example, if the general fuzzy set is expressed as

$$A = \{(x, \mu_A(x))\}, x \in X \quad (x, \mu_A(x)) \quad [i]$$

where i denotes the membership function, and is a singleton, then a fuzzifier given by

$$\mu(X) = 1 / (1 + (x/K_2)^{K_1}) \quad x \in X \quad [ii]$$

Produces an S-shaped curvature. K1 and K2 are called the exponential and denominational fuzzifiers, respectively. By having controllable parameters such as K1 and K2, adaptive fuzzy algorithms can be developed. 3) Fuzzy rule base: The fuzzy proposition is then presented in IF–THEN format and

constitutes the rule base. Specifically, a finite fuzzy logic implication statement in the rule base was described by a set of general fuzzy IF–THEN rules containing only the fuzzy logical AND operation, in the form ‘IF a1 is A1 AND a2 is A2 THEN b1 is B1.’ 4) Fuzzy strategy base: This contains the algorithms for computing the condition part and the conclusion part. A proposition might encompass many conditions. An appropriate fitness of a rule had to be found so that the conclusion can be drawn. This is carried out by a process of implication. A membership function that defines the implication relation can be expressed in a number of ways. To illustrate the operation, we assume that we have the following simple conditional proposition (canonical rule):

IF X is A THEN Y is B

The implication relation is defined by

$$R(x,y) = \int_{x,y} \mu(x,y) / (x,y) \quad [iii]$$

Where linguistic/fuzzy variable X and Y take the value of A and B, respectively, and l(x, y) is the membership function of the implication relation. The membership function is denoted by

$$\mu(x,y) = \mu_A(x) \wedge \mu_B(y) \quad [iv]$$

The symbol \wedge corresponds to intersection operation [18].

Numeric analysis approach of fuzzy system was first presented by Takagi and Sugeno and then a lot of studies have been made [17]. On the other hand, studies of fuzzy neural networks that combine both advantages of the fuzzy systems and the learning ability of the neural networks have been carried out. These techniques can alleviate the matter of fuzzy modeling by learning ability of neural networks and have been reported since around the beginning of 1990s [19, 6]. Fuzzy neural networks can be applied not only for simple pattern classification but also meaningful fuzzy if-then rules creation; therefore, they can be put into practice for various applications [17].

III. RESEARCH METHODOLOGY

The cross-industry standard process for data mining (CRISP-DM) is well-known process model to develop Data Mining projects[20]. As it was mentioned before, since its introduction in 1996, CRISP-DM has been the most favored methodology in data mining domain [21]. Therefore, we have chosen it as our reference model. According to this methodology, the steps of research can be described as follows:

1) Business understanding: This initial phase focuses on understanding the project objectives and requirements from a business perspective, then converting this knowledge into a data mining problem definition and a preliminary plan designed to achieve the objectives.

2) Data understanding: The data understanding phase starts with an initial data collection and proceeds with activities in order to get familiar with the data, to identify data quality problems, to discover first insights into the data or to detect interesting subsets to form hypotheses for hidden information.

3) Data preparation: The data preparation phase covers all activities to construct the final dataset (data that will be fed into the modeling tool(s)) from the initial raw data. Data preparation tasks are likely to be performed multiple times

and not in any prescribed order. Tasks include table, record and attribute selection as well as transformation and cleaning of data for modeling tools. In this step, there is a problem for labeling to the class of customers which we try to solve it with a fuzzy inference system. The steps of the designed system consist of:

- a) To gather the data that their Class field is the score of credit customers and has been assigned as a number to each record.
- b) To specify the effective criteria on estimation of the credit **score** of each customer as the input of FIS model and the output is the credit score (the quality degree of each customer).
- c) To **specify** the linguistic values and Fuzify them.
- d) To **extract** the rules related to the criteria and the credit score.
- e) To implement the model in FIS. (vi)To estimate the Class field for each customers' record via FIS model.

4) Modeling: In this phase, various modeling techniques are selected and applied and their parameters are calibrated to optimal values. Typically, there are several techniques for the same data mining problem type. Some techniques have specific requirements on the form of data. Therefore, stepping back to the data preparation phase is often necessary.

5) Evaluation: At this stage in the project you have built a model (or models) that appear to have high quality from a data analysis perspective. Before proceeding to final deployment of the model, it is important to more thoroughly evaluate the model and review the steps executed to construct the model to be certain it properly achieves the business objectives. A key objective is to determine if there is some important business issue that has not been sufficiently considered. At the end of this phase, a decision on the use of the data mining results should be reached.

6) Deployment: Creation of the model is generally not the end of the project. Even if the purpose of the model is to increase knowledge of the data, the knowledge gained will need to be organized and presented in a way that the customer can use it[21].

The detail of models and the results has been explained in the rest of paper.

IV. RESULT AND DISCUSSION

In this section, CRISP-DM methodology has been implemented on real data set. Focus of this research is on Data Preparation and Modeling phases of this Standard for a Classification problem in order to credit scoring for an Iranian bank . The followings are the explanation for both of these steps:

Data Preparation Step:

The data preparation phase covers all activities to construct the final dataset (data that will be fed into the modeling tool) from the initial raw data. At First, the researchers have removed the noisy data from all the available records and have selected the main attributes and then a database is created which contains 120 records and 12 fields.

The customer's attributes include 12 fields in the final database: Activity time, Number of employees, Age of

managing director, Education level of managing director, Years of cooperation with the bank, Type of Collateral, amount of loan, Current rate, Debit ratio, Period of the collection of claims, Activity area, Profit Rate or rate of return.

We can compare the results with the ones from other models as well. Having the good results with this Hybrid model is not so important in this survey and to find a solution to forecast the problems that the output field (Class field) is being determined by multi criteria, is a proper work and overcomes the exist challenges and vagueness. Here we have considered three criteria effective on the quality degree (as the output) of bank credit customers and the judgment about the credit status of customers depends on the three criteria .The criteria include: i)Sum of the days of Delay to settle the payments(an usual criterion for delay), ii)Sum of the times when delay happened for payments in addition to the delay in paying back period.(The high repetitive times proves the Delay is not random), iii)Sum of the amounts of payments with delay which the increase of this criterion will face the bank and Its capital in turn with the high level of risk and as a result will cause problems for giving loan to new customers. Here, we have considered the sum function for all of the three mentioned criteria here, since there is a possibility to have many times for delay in payments. Here is the FIS model based on the 15 rules obtained from Experts' opinions(Table 1).

TABLE 1. THE RULES OF FUZZY EXPERT SYSTEM

	Days of Delay	Number of payments with Delay	The amount of payments with Delay	Degree of credit Quality for customer
1	Very High	High	Very High	Very low
2	Very High	Low	low	low
3	Very High	Medium	Medium	Very low
4	High	High	Very low	low
5	High	Low	low	Medium
6	High	Medium	Medium	low
7	Medium	High	High	low
8	Medium	Low	low	Medium
9	Medium	Medium	Very High	low
10	low	High	low	Medium
11	low	Low	low	High
12	low	Medium	Medium	Medium
13	Very low	High	Very low	High
14	Very low	Low	Very High	High
15	Very low	Medium	High	Very High

The membership function for each variable which is according to experts' opinions has been shown in Fig1-4 and Fig 5 has presented the rule viewer of labeling.

According to the result of FIS model, we have estimated in this step the quality of credit degree for each customer based on the three considered criteria. The 120 outputs as the degree of credit quality for 120 customers, are the outputs of the main database and the input variables, independent ones, are customer attributes. Labeling each record or determining

the value of class has been done according to the degree of credit quality and three spectrums have been considered as being shown in Table 2. The credit degree between (0-0.399) is considered as Bad class label, credit degree between (0.4-0.699) as Medium and (0.7-1) as Good class label.

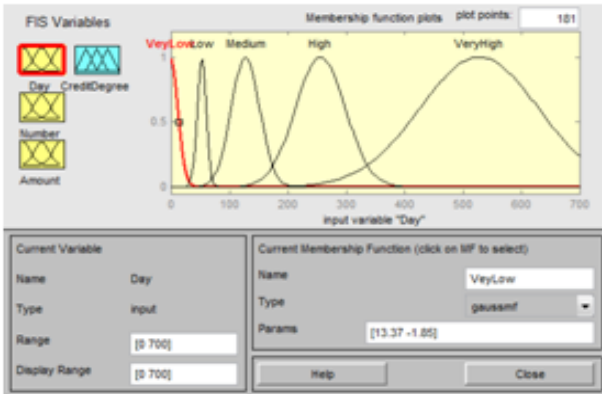


Fig 1. Five Gaussian Membership function for Days of Delay

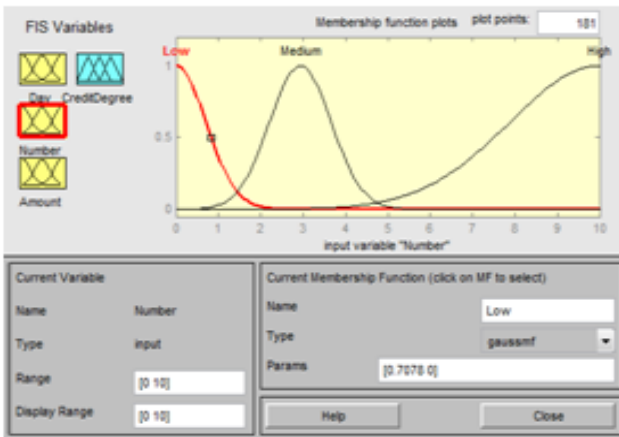


Fig 2. Three Gaussian Membership function for Number of payments with Delay

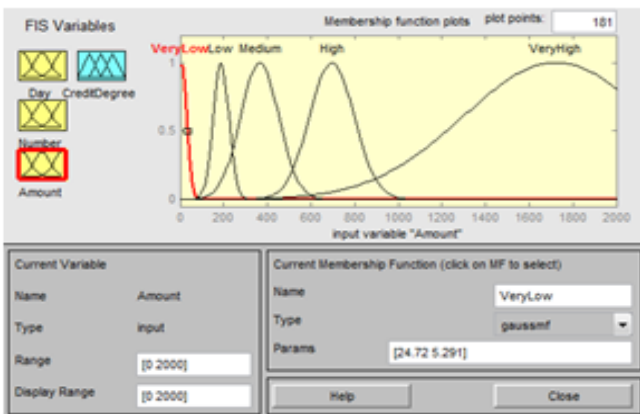


Fig 3. Five Gaussian Membership function for The amount of payments with Delay

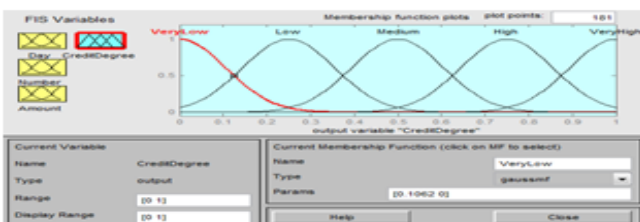


Fig 4. Five Gaussian Membership function for Degree of credit Quality for customer

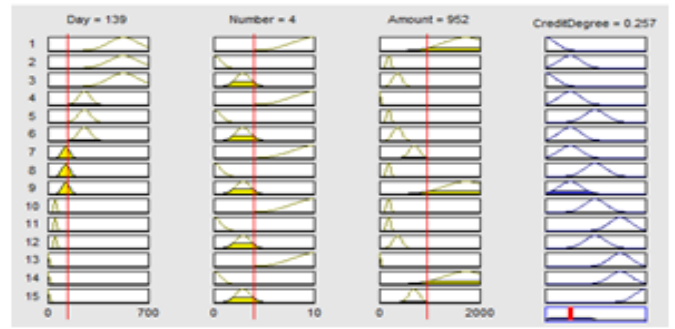


Fig 5. The rule viewer for labeling to class of customer

Table 2. Three spectrums of each class

Class label	Credit Degree
Bad	0-0.399
Medium	0.4-0.699
Good	0.7-1

Modeling Step:

In this phase, various modeling techniques are available to be applied and their parameters are calibrated to optimal values.

Here, we have done the Classification Data Mining Algorithms with the Clementine tool on these data which include 120 records and 12 fields or customers' attributes and 1 class field. Clementine data stream is shown in Fig6.

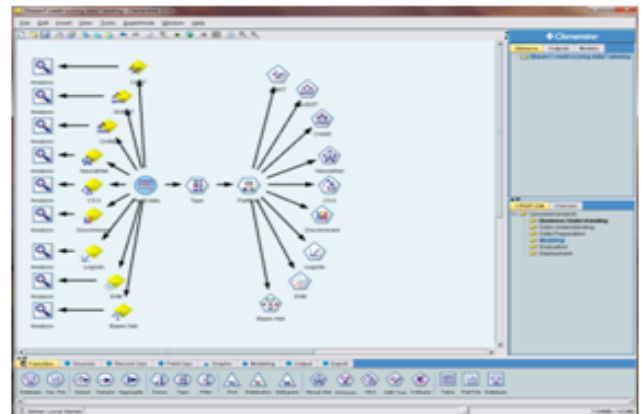


Fig6. Clementine Data Stream for Implementing Classification Algorithms

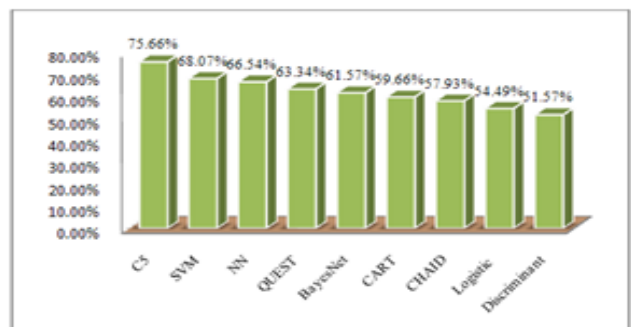


Fig 7. The Accuracy of obtained results of Data mining Algorithms

As it is shown in Fig 7, the final results from Data Mining show that the C5 Algorithm has the highest Accuracy. Thus, the obtained rules have been presented to the considered bank as the outputs.

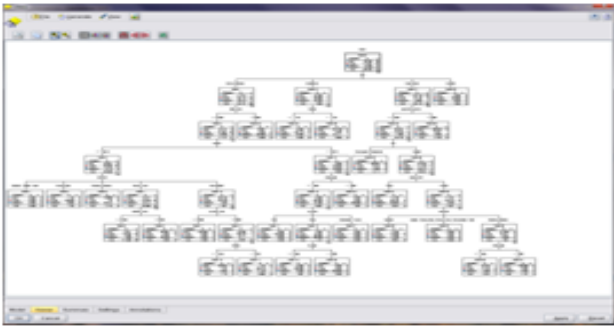


Fig 8. The result of Data Mining Algorithms

In Fig 8 is presented the result of Data Mining Algorithms that obtained with C5 Algorithm.

V. CONCLUSION

In this paper, with the use of Fuzzy Expert System, we assigned the Class to the records of a database and then the Data Mining algorithms have been done on the records with Clementine software. Correct Labeling of records is so essential in Data Mining, otherwise the Data Mining is not an integrated solution and we can not trust on the results. In addition, to label each record (to estimate the credit level of customer) was done based on the states of some important criteria from the viewpoint of experts and according to linguistic variables in the format of rules. Then, some Data Mining algorithms have been done on Bank Database which the results show the Decision tree algorithm (C5) has more accuracy than other algorithms. This research is useful for Classification problems such as Credit Scoring that first, to determine the class for each record should be done according to the experts ideas and second, this viewpoint, is in addition to consider the states of some criteria and third, these opinions should be presented with linguistic variables. So the using Fuzzy conclusion in these kinds of problems is effective.

APPENDIX

Here, some useful MATLAB commands to work with the proposed fuzzy inference system (FIS) which is based on Mamdani are presented:

```
[System]
Name='Labeling'
Type='mamdani'
Version=2.0
NumInputs=3
NumOutputs=1
NumRules=15
AndMethod='min'
OrMethod='max'
ImpMethod='min'
AggMethod='max'
DefuzzMethod='centroid'
```

REFERENCES

- [1] Umit İlhan, "Neuro Fuzzy Type 2, Client Assessment System", Ph.D. Thesis, Department of Computer Engineering Nicosia – 2007.
- [2] C. S. Ong, J. J. Huang, G. H. Tzeng "Building credit scoring models using genetic programming", *Expert Systems with Applications*, 2005, 29 (1), pp. 41-47.
- [3] T. S. Lee, C. C. Chiu, C. J. Lu, I. F. Chen, "Credit scoring using the hybrid neural discriminant technique". *Expert Systems with Applications*, 2002, 23(3), pp.245–254.
- [4] D. West. "Neural network credit scoring models". *Computers and Operations Research*, 2000, 27(11-12), pp.1131–1152.
- [5] Y. Jiao, Y. R. Syau, E. S. Lee, "Modelling credit rating by fuzzy adaptive network", *Mathematical and Computer Modelling*, 2007, 45(5-6), pp.717-731.
- [6] M. Courchane, A. Gailey, P. Zorn, "Consumer credit literacy: What price perception?", *Journal of Economics and Business*, 2008, 60 (1-2), pp. 125-138.
- [7] R. Tsaih, Y. J. Liu, W. Liu, Y. L. Lien, "credit scoring system for small business loans", *Decision Support Systems*, 2004, 38 (1), pp. 91-99.
- [8] A. B. Emela, M. Oral, A. Reisman, R. Yolalana "A credit scoring approach for the commercial banking sector", *Socio-Economic Planning Sciences*, 2003, 37 (2), pp.103-123.
- [9] E. Rosenberg, A. Gleit, "Quantitative methods in credit management: a survey", *Operations Research*, 1994, 42, pp.589–613, Key: citeulike: 651925, Springer.
- [10] D.J. Hand, W.E. Henley, "Statistical Classification Methods in Consumer Credit Scoring: A Review", *Journal of Royal Statistical Society*, 1997, 160 pp.523–541, Key: citeulike:708451, Springer.
- [11] B. Widrow, D.E. Rumelhart, M.A. Lehr, "Neural networks: application in industry, business and science", *Communications of the ACM*, 1994, 37 (3) pp. 93–105.
- [12] D. West, "Neural network credit scoring models", *Computers and Operations Research*, 2000, 27 (11-12), pp. 1131-1152.
- [13] E. Kirkos, C. Spathis, Y. Manolopoulos. "Data Mining techniques for the detection of fraudulent financial statements", *Expert Systems with Applications*, 2007, 32 (4), pp. 995-1003.
- [14] N. C. Hsieh, "An integrated data mining and behavioral scoring model for analyzing bank customers", *Expert Systems with Applications*, 2004, 27 (4), pp. 623-633.
- [15] E.W.T. Ngai, L.Xiu, D.C.K. Chau, "Application of data mining techniques in customer relationship management: A literature review and classification", *Expert Systems with Application*, 2008, 36 (2 PART 2), pp. 2592-2602.
- [16] Turban, E., Aronson, J. E., Liang, T. P., & Sharda, R. "Decision support and business intelligence systems" (Eighth ed.). Pearson Education, 2007.
- [17] H. Iyatomi, M. Hagiwara, "Adaptive fuzzy inference neural network", *Pattern Recognition*, 37 (10), pp. 2049-2057.
- [18] Y. S. Juang, S. S. Lin, H. P. Kao, "Design and implementation of a fuzzy inference system for supporting customer requirements", *Expert Systems with Applications*, 2007, 32 (3), pp. 868-878.
- [19] Y. S. Kim and S. Y. Sohn "Managing loan customers using misclassification patterns of credit scoring model", *Expert Systems with Applications*, 2004, 26 (4), pp. 567-573.
- [20] O. Marban, E. Menasalvas, C. Fernandez-Baizan, "A cost model to estimate the effort of data mining projects (DMCoMo)", *Information Systems*, 33, 2008, pp. 133–150
- [21] E. Chapman and et al., "CRISP-DM 1.0 Step-by- Step Data Mining Guide, SPSS", <http://www.crispdm.org/CRISPWP-0800.pdf>, 2000.

Hamid Eslami Nosratabadi, photograph and biography not available at the time of publishing.

Sanaz Pourdarab, photograph and biography not available at the time of publishing.

Ahmad Nadali, photograph and biography not available at the time of publishing.