

# Application of Artificial Intelligence Techniques for Credit Risk Evaluation

Ahmad Ghodselahi and Ashkan Amirmadhi

**Abstract**—Credit risk is the most challenging risk to which financial institution are exposed. Credit scoring is the main analytical technique for credit risk evaluation. Application of artificial intelligence has lead to better performance of credit scoring models. In this paper a hybrid model for credit scoring is designed which applies ensemble learning for credit granting decisions. Ten classifier agents are utilized as the members of ensemble model. Support vector machine, Neural Networks and Decision Tree as base classifiers were compared based on their accuracy in classification. Since even a small improvement in credit scoring accuracy causes significant loss reduction, then the utilization of best classification model is of a great importance. A real dataset was used to test the model and classifiers. The test results showed that proposed hybrid ensemble model has better classification accuracy and performance when compared to other credit scoring methods. In addition, among three classifiers, the support Vector Machine had the best performance and accuracy.

**Index Terms**—Credit Risk, Ensemble learning, Hybrid Model, Artificial Intelligence techniques.

## I. INTRODUCTION

Banking is special industry that deals with capital and risk for making profit. The bank success is directly pertaining to its capability of controlling and managing related risks. Banks are exposed to different kinds of risk, but the most challenging risk which can cause a bank to full failure is credit risk. The recent world's financial crisis has aroused remarkable consideration of financial institutions and banks on credit risk. Credit risk is an important and widely studied topic in the bank industry lending decisions and profitability. For all banks, credit remains the single largest risk being difficult to compensate. Credit risk is a general term which implies to future losses. Credit risk is the loss of bank's profit, since the customer does not adhere to his or her loan refund commitment [1]. Application of statistical and intelligent techniques in credit risk evaluation and bankruptcy prediction research has been an area of interest since 7th decade. Usually, the generic approach of credit risk evaluation is to apply some classification techniques on similar data of previous customers, both faithful and irresponsible customers, in order to find a relation between the characteristic and potential failures [2]. Credit scoring has become one of the main analytical ways for financial institutions to assess credit risk. The purpose of credit scoring

is to classify the applicants into two groups: applicants with good credit and applicants with bad credit. Applicants with good credit have great possibility to repay financial obligation while, applicants with bad credit have high possibility of defaulting. Credit scoring process is an independent evaluation whose aim is to find out how an object is capable and willing to meet its payable obligations, specifically based on complex analysis of all the known risk factors of the assessed object. It is realized by a scoring agency. A higher credit scoring shows a low credit risk. According to the assessed object, credit ratings of the state, company, municipality, financial institution, single bond, etc. exist. Credit scoring is a result of a credit scoring process. It is represented by a rating class defined on a rating scale. Rating classes are assigned to assessed objects. Credit scorings are used by bond investors, debt issuers, and governmental officers as a measure of the risk of a company. They provide a means of determining risk premiums and marketability of bonds, allowing firms issuing debt to estimate the likely return investors require. Bankers and companies considering providing credit rely on credit scorings to make important investment decisions, many regulatory requirements for financial decisions are based on credit ratings, etc. The accuracy of credit scoring is critical for financial institution's profitability. Even 1% of improvement on the accuracy of credit scoring of applicants, will decrease a great loss for financial institutions. Usually credit score is a value that reveals the credit of the customer based on quantitative analysis of customer's credit history and characteristics [3]. The credit scoring model identifies financial variables that have statistical explanatory power in differentiating bad customers from good ones. The benefits obtained by developing a reliable credit scoring system are [4]:

- reducing the cost of credit analysis
- enabling faster decision
- insuring credit collections and diminish possible risk

Credit scoring was originally evaluated subjectively according to personal experiences, and later it was based on 5Cs: the character of the consumer, the capital, the collateral, the capacity and the economic conditions. But with the tremendous increase of applicants, it is impossible to conduct the work manually. Many organizations in the credit industry are developing new models to support the credit decisions. The objective of these new credit scoring models is to improve the accuracy, which means more credit worthy applicants are granted credit, and consequently, increasing profits. The first credit scoring model was designed by Altman [5]. The credit scoring models can be divided into two categories: traditional models and novel models. The most common and utilized traditional models are Linear Discriminant Analysis (LDA) and Logistic Regression (LR)

Manuscript received July 24, 2011; revised August 8, 2011.

A. Ghodselahi is with Tarbiat Modares University, Tehran, Iran. He is IACSIT member. (E-mail: Ahmad.Ghodselahi@Gmail.com).

A. Amirmadhi is with the Industrial Management Department, Islamic Azad University, Science & research branch, Tehran, Iran. (E-mail: Ashkan.Amirmadhi@Yahoo.com).

[6, 7, 8]. The weakness of the LDA is the assumption of linear relationship between variables, which is usually nonlinear and the sensitivity to the deviation from the multivariate normality assumption. The LR is predicting dichotomous outcomes and linear relationship between variables in the exponent of the logistic function, but does not require the multivariate normality assumption. Because of the deficiency, the linear relationship between variables, both LDA and LR are stated to have lack of accuracy [9]. Advances in information technology have lowered the cost of acquiring, managing and analyzing data, in an effort to build more robust and strong financial systems [10]. Recently, new approaches were applied for developing robust credit scoring systems. Recent studies have revealed that emerging artificial intelligent techniques, such as Decision Tree (DT), Support Vector Machine (SVM), Genetic Algorithm (GA) and Artificial Neural Networks (ANN) are advantageous to statistical models and optimization technique for credit risk evaluation. In contrast with statistical methods, AI methods do not assume certain data distributions. These methods automatically extract knowledge from training samples. According to previous studies, AI methods are superior to statistical methods in dealing with corporate credit risk evaluation problems, especially for nonlinear pattern classification. Application of aforementioned techniques had been investigated by several works. Baesens et al. [14] conducted a study for benchmarking of 17 different classification techniques on eight different real-life credit datasets. They used SVM and Least Squared-SVM with linear and Radial Basis Function (RBF) kernels and adopted a grid search mechanism to tune the hyper parameters in their study. Their experimental results indicated that SVM has the highest average ranking on performance. Schebesch and Stecking [15] used a standard SVM with linear and RBF kernel for applicant credit scoring and used a linear-kernel-based SVM to divide a set of labeled credit applicants into subsets of typical and critical patterns, which can be used for rejected applicants. In [16] SVMs were used for bankruptcy prediction and better accuracy was generated by SVM when compared to other methods. Gestel et al. [17] used LS-SVM for credit rating of banks, and compared the results with ordinary least squares, LR and multilayer perceptron (MLP). Min et al. [18] proposed methods for improving SVM performance in two aspects: feature subset selection and parameter optimization. Bencic et al. [19] designed credit scoring model for small businesses in Italia. They used three techniques namely, LR, NNs and DT. Abdou et al. [20] investigated the ability of NNs, such as Probabilistic Neural Networks (PNN) and multi-layer feed-forward nets, and traditional techniques such as discriminant analysis, probit analysis and LR in evaluating credit risk in Egyptian banks by applying credit scoring models. The results of their investigation have shown that NNs models have more accurate classification rate in comparison with other techniques. In [21], an application of NNs to credit risk evaluation related to Italian small businesses was described. This work presents two neural network systems, one with a feed-forward network, and the other one with special purpose architecture. This paper suggested that both NNs could be very successful in learning and estimating the default tendency of borrower, provided

that careful data analysis, data pre-processing and training are performed properly. Pang and Gong [22] had applied C 5.0 algorithm for credit risk. They stated that DT is good techniques for these kinds of problem. Cho et al. [23] proposed a hybrid approach based on the combination of variable selection using DT and case-based reasoning. Although almost all classification methods can be used to evaluate credit risk, some hybrid approaches have shown higher correctness of predictability than any individual methods. In machine learning, the hybridization approach has been an active research area to improve classification or prediction performance over single learning approach. In general, it is based on combining two different machine learning techniques. For example, a hybrid classification model can be composed of one unsupervised learner to pre-process the training data and one supervised learner to learn the clustering result or vice versa. In [24], a hybrid mining approach was presented for credit scoring. Due to unrepresentative data, a two-stage approach was used which utilized self organizing map for clustering and NN to construct credit scoring model. Huang et al. [25] designed hybrid SVM-based credit scoring models for assessing the credit scores of applicants. In [26] a hybrid credit scoring model was developed applying genetic programming and SVM. The accuracy of their hybrid model was better when compared with SVM, genetic programming, DT, LR, and back propagation neural networks. Chen et al. [27] designed a SVM based hybrid model which mainly has three strategies. First using Cart, then using Mart and at last using grid search for model variable improvement. Tsai and Chen [28] present four hybrid credit scoring model and compare the performance of these hybrid models. They used DT, ANN, naïve bays classification and LR as classification methods, while K-means and expectation maximization were used as clustering methods. The result showed that the combination of LR and NN can provide the highest accuracy and maximize the profit. Motivated by the hybrid model, integrating multiple classifier into aggregated output, ensemble learning, has been turned out to be an efficient method for achieving high classification performance. There is a growing interest that existing application of single classifier can be further improved by ensemble methods. The works of [3, 29, 30] have shown that ensemble methods have performed better than single classifier. Tsai and Wu [4] investigated the performance of a single classifier as the baseline classifier to compare with multiple classifiers and diversified multiple classifiers by NN based on three datasets. Zhou and Lai [31] developed multi-agent ensemble model for credit risk evaluation. Each agent is acted by a weighted LS-SVM. The results showed that the proposed model has better accuracy than other methods with which the proposed model was compared. Nanni and Lumini [30] compared the performance of ensemble of classifiers with single ones. The result showed applying ensemble method lead to better performance of classification. In [32], a comparative assessment of the performance of three popular ensemble methods - Bagging, Boosting and Stacking - and four base learners - LR analysis, DT, ANN, and SVM - on credit scoring problem was conducted. Better performance of ensemble methods rather than individual methods, was revealed by the results.

The rest of this paper is organized as follows. In Section II, an overview of ensemble learning is presented. In section III, the details of experimental design are presented. Section IV reports experimental results. Based on the observations and results of these experiments, section V draws conclusions.

## II. OVERVIEW OF ENSEMBLE LEARNING

Ensemble learning is a machine learning paradigm where multiple learners are trained to solve the same problem. In contrast to ordinary machine learning approaches that try to learn one hypothesis from training data, ensemble methods try to construct a set of hypotheses and combine them to use [32]. This method is used to improve the performance and accuracy of classification task. The multiple classifier systems are based on the aggregation of a pool of classifiers such that their fusion achieves higher performance than the single classifiers. The key idea of most methods for building ensemble of classifiers is to modify the training dataset, builds classifiers on these  $n$  new training sets and then combines them into a final decision rule [30]. The rationale is that it may be more difficult to optimize the design of a single complex classifier than to optimize a combination of relatively simple classifiers. In ensemble models the error and deviation of one classifier is compensated by the other members of ensemble on classification task. On the generalization ability of ensemble method is usually much stronger than that of a single classifier. Dietterich [33] gave three reasons by viewing the nature of machine learning as searching a hypothesis space for the most accurate hypothesis. The first reason is that, the training data might not provide sufficient information for choosing a single best classifier. For example, there may be many classifiers performing equally well on the training set. Thus, combining these classifiers may be a better choice. The second reason is that, the search processes of the classifier algorithms might be imperfect. For example, even if there may exist a unique best hypothesis, it might be difficult to achieve since running the algorithms results in sub-optimal hypotheses. Thus, ensembles can compensate for such imperfect search processes. The third reason is that, the hypothesis space being searched might not contain the true target function, while ensembles can give some good approximation. For example, it is well-known that the classification boundaries of DTs are linear segments parallel to coordinate axes. If the target classification boundary is a smooth diagonal line, using a single DT cannot lead to a good result but a good approximation can be achieved by combining a set of DTs.

## III. THE DESIGN OF THE METHODOLOGY

This section introduces the process of credit scoring model proposed by this study. While there is no overall best AI techniques used in building credit scoring models, for what is best depends on the details of the problem, the data structure, the characteristics used, the extent to which it is possible to segregate the classes by using those characteristics, and the objective of the classification. In this paper a comparison of methods applied for credit scoring, was conducted. A hybrid model is used for better classification which employs two machine learning techniques, clustering and classification techniques. For clustering task, fuzzy C-Means clustering is utilized. For classification task, three popular classifiers;

SVM, NN and DT, is utilized. The whole process consists of Fuzzy C-Means clustering, normalization, building classifier agents and finally, defining a method to combine the results generated by each agent. In this study, 10 classifiers are employed as ensemble members. The aim of the proposed model is to make full use of knowledge and intelligence of the members of group to make a rational decision over a pre-determined set of criteria. Each part of the hybrid credit scoring model is briefly described in following sub sections.

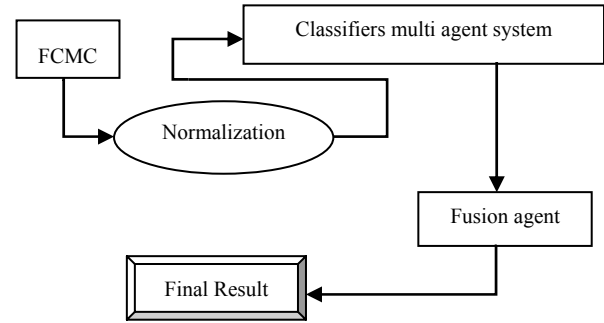


Fig I. Proposed hybrid model

### A. Clustering

The first phase of the model is fuzzy clustering of dataset. This phase is as a pre-process for building classifier agents that generates homogeneous clusters with same features. This pre-process will lead to better training of classifier agents and as a result, better classification model is made and the probability of misclassification is reduced which is caused by inapt training data. Sometimes, even with a correct classification model, the ability of a model for predicting a new instance is limited. Such limitations are because of improper classification patterns which arise from training data. In addition, data uncertainty leads to more complex learning process of classifier agents. Therefore, the higher quality of training data causes the higher ability of classifier agents for correct classifications. The proposed model utilized FCM clustering to generate 10 clusters associated with their classifier agents.

FCM is a method of clustering which allows one piece of data to belong to two or more clusters. This method is frequently used in pattern recognition. It is based on minimization of the following objective function:

$$J_m = \sum_{i=1}^N \sum_{j=1}^C u_{ij}^m \|x_i - c_j\|^2, \quad 1 \leq m \leq \infty \quad (1)$$

where  $m$  is any real number greater than 1,  $u_{ij}$  is the degree of membership of  $x_i$  in the cluster  $j$ ,  $x_i$  is the  $i$ th of  $d$ -dimensional measured data,  $c_j$  is the  $d$ -dimensional center of the cluster, and  $\|*\|$  is any norm expressing the similarity between any measured data and the center. Fuzzy partitioning is carried out through an iterative optimization of the objective function (1), with the update of membership  $u_{ij}$  and the cluster  $c_j$  by:

$$u_{ij} = \frac{1}{\sum_{k=1}^C \left( \frac{\|x_i - c_j\|}{\|x_i - c_k\|} \right)^{\frac{2}{m-1}}}, \quad c_j = \frac{\sum_{i=1}^N u_{ij}^m x_i}{\sum_{i=1}^N u_{ij}^m} \quad (2)$$

This iteration will stop when,

$$\max_{ij} \{ |u_{ij}^{(k+1)} - u_{ij}^{(k)}| \} < \epsilon, \quad (3)$$

The algorithm is composed of the following steps [34, 35]:

1. Initialize  $U = [u_{ij}]$  matrix,  $U^{(0)}$
2. at  $k$ -step: calculate the centers vectors  $C^{(k)} = [c_j]$  with  $U^{(k)}$
3. Update  $U^{(k)}, U^{(k+1)}$

$$u_{ij} = \frac{1}{\sum_{k=1}^c \left( \frac{\|x_i - c_j\|}{\|x_i - c_k\|} \right)^{\frac{2}{m-1}}}$$

4. If  $\|U^{(k+1)} - U^{(k)}\| < \varepsilon$  then STOP; otherwise return to step 2

### B. Normalization

Data normalization should be performed in order to feed the classifier agents with data ranging in same interval for each input node. In credit assessment the numerical values representing the attributes of an applicant vary significantly in value and if a simple normalization process is applied to whole dataset, some useful information may be lost. At the normalization phase, the input data are separately normalized to values between 0 and 1. This is achieved by finding the maximum or highest value within each input attribute for all 1000 instances in dataset and dividing all the values within that same attribute by the obtained maximum value. This is a simple but efficient normalization.

### C. Classification

As it was mentioned, three classifiers is used and compared in this paper. A brief description of SVM, NN and DT methods is presented in the following subsections.

#### C.1 Support Vector Machine

Support Vector Machine (SVM) technique is a classification technique that as an AI technique has proven its performance in many fields, such as text categorization, credit risk, and bankruptcy prediction. SVM is based on the idea of Structural Risk Minimization (SRM) to build a model of a given system. A SVM employs structural risk minimization rather than the empirical risk minimization used by conventional neural networks. SVMs use a linear model to implement nonlinear class boundaries via the nonlinear mapping of input vectors into a high-dimensional feature space. In this high-dimensional space, the maximum margin hyper plane is found so that the separation between decision classes can be maximized. Support vectors are defined as the training examples closest to the maximum margin hyper plane. The methodology is becoming renowned due to many useful features and promising empirical performance. SVM is an optimization technique in which prediction error and model complexities are simultaneously minimized. The strength of this technique lies with its capability to model nonlinearity and resulting in complex mathematical models. SVMs are used to find an optimal hyper plane which maximizes the margin between itself and the nearest training examples in the new high-dimensional space and minimizes the expected generalization error.

#### C.2 Neural Networks

Neural networks (NNs) are defined as massively parallel processors, which tend to preserve the experimental knowledge and enable their further use. They simulate the human brain with the intent to collect the empirical evidence during the learning process, and inter-neural connections

(synapses) are used to store the knowledge. An important feature of NNs, in addition to the ability of learning, is the ability to generalize the learned knowledge. Currently, there exist many structures and learning algorithms of NNs, including a number of their applications. In the economic field, NNs are used primarily in problems where the variables are in non-linear relationships. An artificial neural network is composed of a group of neural nodes that link with the weighted nodes. Every node can simulate a neuron of creatures, and the connection among these nodes is equal to the synaptic that connects among the neurons. The most common type of neural networks consists of three layers of units: input layers, hidden layers, and output layers. It is called multilayer perceptron (MLP). A layer of input units is connected to a layer of hidden units, which is connected to a layer of output units.

#### C.3 Decision Tree

A decision tree (DT) is a model of the data that encodes the distribution of the class label in terms of the predictor attributes; it is a directed, acyclic graph in a form of a tree. The root of the DT does not have any incoming edges. Every other node has exactly one incoming edge and zero or more outgoing edges. If a node  $n$  has no outgoing edges we call  $n$  a leaf node, otherwise we call  $n$  an internal node. Each leaf node is labeled with one class label; each internal node is labeled with one predictor attribute called the splitting attribute. Each edge  $e$  originating from an internal node  $n$  has a predicate  $q$  associated with it where  $q$  involves only the splitting attribute of  $n$ . A DT can be used to predict the values of the target or class attribute based on the predictor attributes. To determine the predicted value of an unknown instance, you begin at the root node of the tree. Then decide whether to go into the left or right child node based on the value of the splitting attribute. You continue this process using the splitting attribute for successive child nodes until you reach a terminal or leaf node. The value of the target attribute shown in the leaf node is the predicted value of the target attribute. A DT can also be converted into rules which could be used for prediction tasks such as credit default and bankruptcy.

#### D. Fusion agents

Majority vote is the most common and used method for combining the group members' result in ensemble models. Despite the good capability of this method for combining, another method have been used which led to better classification accuracy than the aforementioned fusion method. Every agent is assigned a weight, according to the sum of its members' membership degrees. The weight of agent 1 is the sum of Cluster 1 members' membership degrees.

## IV. EXPERIMENTAL ANALYSIS

In order to test the performance of hybrid model of this paper, the real world German dataset is used which is presented in follow.

#### A. Real World Dataset

The German dataset is available at UCI Machine Learning Repository. It contains 1000 instances, with 700 cases were granted credit and 300 cases were refused. In these instances,

each case is characterized by 20 decision attributes, 7 numerical and 13 categorical.

**B. Experimental Result**

The evaluation criteria used to compare the tested methods are type I, type II and total accuracy which are calculated based on the following formulas:

$$\text{Type I accuracy} = \frac{\text{Number of classified and also observed bad}}{\text{Number of observed bad}} \quad (4)$$

$$\text{Type II accuracy} = \frac{\text{Number of classified and also observed good}}{\text{Number of observed good}} \quad (5)$$

$$\text{Total accuracy} = \frac{\text{Number of correct classification}}{\text{Numbr of total evaluation}} \quad (6)$$

TABLE I. SVM KERNELS' TOTAL ACCURACY

SVM Kernel	Fusion method	
	Majority vote	Membership degree
	Total accuracy (%)	Total accuracy (%)
RBF	77.5	78.93
Polynomial	79.64	81.42
Sigmoid	49.64	56.42
Linear	71.78	71.07

TABLE II. RESULT OF CREDIT SCORING MODELS (GERMAN DATASET)

Method	Accuracy (%)			AUC
	Type I	Type II	Total	
DA	67.49	64.91	65.91	66.2
LR	48.68	84.67	71	66.67 5
Decision tree	49.93	78.55	70.35	64.24
RBFN individual	39.47	86.29	68.5	62.88
SVM individual	27.63	97.58	71	62.60 5
MLP individual	55.26	85.48	74	70.37
Bagging DT*	48.28	86.34	74.92	67.31
Bagging NN*	48.40	87.20	75.56	67.8
Bagging SVM*	43.42	89.86	75.93	66.64
Boosting DT*	40.89	82.96	72.77	61.92 5
Boosting NN*	49.20	83.63	73.3	66.41 5
Boosting SVM*	45.62	89.44	76.3	67.53
Stacking*	45	89.24	75.97	67.12
Ensemble DT**	43.67	90.67	76.07	67.16
Ensemble DT***	47.12	89.11	76.07	68.16
Ensemble NN**	26.43	96.37	74.67	61.4
Ensemble NN***	43.67	92.22	77.14	67.95
Ensemble SVM**	67.81	84.97	79.64	76.39
Ensemble SVM***	66.66	88.08	81.42	77.37

\* Ref [32]

\*\* Majority vote Fusion method

\*\*\* Membership degree fusion method

Besides accuracy criterion, the area under the receiver operating characteristic curve is also calculated for each method. In order to rank all models, the area under the receiver operating characteristic (ROC) graph is used as another performance measurement. The ROC graph is a useful technique for ranking models and visualizing their performance. Usually, ROC is a two-dimensional graph in which sensitivity is plotted on the Y-axis and 1-specificity is

plotted on X-axis. Actually, the sensitivity is equal to type II accuracy and the specificity is equal to type I accuracy. To perform the model ranking task, a common method is to calculate the area under the ROC curve, abbreviated as AUC. Since the AUC is a portion of the area of the unit square, its value is always between 0 and 1 [36].

For each SVM agent, 4 types of kernel, namely RBF, polynomial, sigmoid and linear, are available. A kernel function can be interpreted as a kind of similarity measure between the input objects. Although some kernels are domain specific, there is in general no best choice. Since each kernel has some degree of variability in practice, there is nothing else for it but to experiment with different kernels. In this paper, four aforementioned kernel types were set in order to achieve best kernel for credit risk assessment model. For each agent the parameter C was set to 10 and gamma for RBF kernel was set to 0.1 and for other kernels, it was set to 1. Moreover, for each agent implementation with different kernel, the results of SVM agents were combined with both fusion methods which are majority vote and proposed method, membership degree. In this way the best fusion method will be defined based on the correct classification results. For NN agent, MLP structure was set. For building MLP classification models, each MLP agent needs training and testing data. Each MLP agent was trained by the defined training dataset, assigned to associated cluster. As the last classifier, the DT classification ensembles were built by applying C 5.0. In order to compare the performance of the two mentioned fusion methods namely, majority vote and membership degree, the results of each ensemble member in were combined by these two fusion method. As the total accuracy of the hybrid model in each implementation with 4 types of kernels was shown in table I, it is obvious the best total accuracy was generated when the SVM kernel is set to polynomial. Moreover, as it is showed in table I, the polynomial and RBF kernels have generated highest total accuracy, respectively, when compared to other kernels. Linear kernel did not lead to good classification accuracy, for the relationship between the features, is not linear. The comparison between two fusion methods' results showed that except in linear kernel that there is a slight difference between experimented fusion methods, in other kernel types, the membership degree fusion method resulted in better accuracy than majority vote; therefore, the new proposed ensemble member result combining method has better performance than the common combining method. Also the result of the DT ensemble and MLP ensemble verified the better performance of membership degree fusion method in that the accuracy was improved for MLP agents and the performance (AUC) was improved for DT agents. Application of membership degree fusion method did not lead to better accuracy, but it improved the performance by enhancing the type I accuracy. In order to evaluate the performance of proposed model with three base classifiers, other method for credit scoring is employed. The result of [32] for comparing common and popular methods of ensemble- bagging, boosting and stacking- with the result of model, is also presented. Among three tested base classifiers with proposed methodology, the SVM ensemble had better accuracy than ensemble of MLP and DT, even with majority vote fusion method. Principally, the performance of base

classifiers was improved by using ensemble method rather than individual one. It showed that ensemble learning and making use of different classifiers leads to better classification task. In comparison with other mentioned and popular ensemble methods, except in DT ensemble which is slightly lower than boosting SVM, ensemble of SVM and MLP, applying membership degree fusion method, reached the better results. It is obvious that the proposed hybrid ensemble model utilizing SVM classifier has better accuracy and performance. Its total accuracy is the best among mentioned methods and it surpasses other common ensemble methods. It also got better result than other classifiers, DT and MLP, with same proposed model, which revealed the great capability of SVM as precise classification technique. When it comes to performance, the AUC is a good measurement criterion for classifiers. As it is presented in the table II, The AUC of SVM ensemble applying membership degree fusion method is superior to other methods' performance. According to the AUC results, the ensemble of SVM has the best performance which makes it the first classifier in credit scoring models performance ranking.

#### V.CONCLUSION

In this paper a hybrid model was developed which applies ensemble learning method to improve the performance of classification in the field of credit risk assessment. The model combined a classification and clustering method. for clustering fuzzy C-Means was applied while for the classification task, three base classifiers - SVM, NN and DT- were utilized as base learners. Ten agents as an ensemble member were designed in the proposed model. In addition two fusion methods were used as one of them was a common method, majority vote, have been used in several work and the other one was a new method proposed in this paper. This new method outperformed the common method used in credit scoring ensemble models. In term of the model result, application of ensemble of base classifiers are prior to the individual ones in that better accuracy and performance is generated. The proposed hybrid ensemble model had better results than other popular ensemble namely, bagging, boosting and stacking. Applying the hybrid model, the performance of three classifiers were compared and based on the result SVM ensemble generated better results than MLP ensemble and DT ensemble, which showed that SVMs are apt classifier resulting in accurate classification. The credit scoring results measured in this research support the hypothesis that ensemble of SVM can be used in credit scoring applications to improve the overall accuracy from a fraction of a percent to several percent. In an overall view, the proposed hybrid ensemble model with SVM as base classifiers that uses membership degree method for combining the results of ensemble members, has the best accuracy and performance.

#### REFERENCES

[1] P. Ya-qiong, "A study on evaluation of consumer credit's risks of commercial banks," Proc. International Conference on Wireless Communications (WiCom 2007), IEEE, pp. 4531-4534, Sept. 2007, doi: 10.1109/WICOM.2007.1115.

[2] L. Yu, S. A. Wang, and K. K. Lai, "Credit risk assessment with a multistage neural network ensemble learning approach," *Expert systems with applications*, vol. 34, pp. 1434-1444, Feb. 2008, doi: 10.1016/j.eswa.2007.01.009.

[3] L. Yu, S. Wang, and K. K. Lai, "An intelligent-agent-based fuzzy group decision making model for financial multicriteria decision support: the case of credit scoring," *European journal of operational research*, vol. 195, pp. 942-959, June. 2009, doi: 10.1016/j.ejor.2007.11.025.

[4] C.-f. Tsai, and J.-w. Wu, "Using neural network ensembles for bankruptcy prediction and credit scoring," *Expert systems with applications*, vol. 34, pp. 2639-2649, May. 2008, doi: 10.1016/j.eswa.2007.05.019.

[5] I. E. Altman, "Financial ratios, discriminant analysis and the prediction of corporate bankruptcy," *The journal of finance*, vol. 23, pp. 589-611, 1968.

[6] B. Baesens, R. Setiono, C. Mues, and J. Vanthienen, "Using neural network rule extraction and decision tables for credit-risk evaluation," *Management science*, vol. 49, pp. 312-329, March. 2003, doi: 10.1287/mnsc.49.3.312.12739.

[7] Lee, T. S., and I. F. Chen, "A two-stage hybrid credit scoring model using artificial neural networks and multivariate adaptive regression splines" *Expert systems with application*, vol. 28, pp. 743-752, May. 2005, doi: 10.1016/j.eswa.2004.12.031.

[8] D. West, "Neural network credit scoring models," *Computers and operations research*, vol. 27, pp. 1131-1152, Oct. 2000, doi: 10.1016/S0305-0548(99)00149-5.

[9] M. Sustersic, D. Mramor, and J. Zupan, "Consumer credit scoring models with limited data," *Expert system with application*, vol. 36, pp. 4736-4744, April. 2009, doi: 10.1016/j.eswa.2008.06.016.

[10] E. Angelini, G. D. Tollo, and A. Roil, "A neural network approach for credit risk evaluation," *The quarterly review of economics and finance*, vol. 48, Nov. 2008, pp. 733-755, doi: 10.1016/j.qref.2007.04.001.

[11] Y.-q. Wang, "Building credit scoring systems based on support-based support vector machine," *Proc. Fourth international conference on natural computation, (ICNC 2008)*, Oct. 2008, pp. 323-327, doi: 10.1109/ICNC.2008.763.

[12] L. Huang, and Y. Dai, "A support vector machine approach for prediction of T cell epitopes," *Proc. Third asia-pacific bioinformatics conference, (APBC 2005)*, Jan. 2005, pp. 312-328.

[13] J. H. Min, and Y.-C. Lee, "Bankruptcy prediction using support vector machine with optimal choice of kernel function parameters," *Expert systems with applications*, vol. 28, May. 2005, pp. 603-614, doi: 10.1016/j.eswa.2004.12.008.

[14] B. Baesens, T. Van Gestel, S. Viaene, M. Stepanova, J. Suykens, and J. Vanthienen, "Benchmarking state-of-art classification algorithm for credit scoring," *Journal of operational research society*, vol. 54, June. 2003, pp. 627-635, doi: 10.1057/palgrave.jors.2601545.

[15] K. B. Schebesch, and R. Stecking, "Support vector machine for classifying and describing credit applicants: Detecting typical and critical regions," *Journal of the operational research society*, vol. 56, Sep. 2005, pp. 1082-1088, doi: 10.1057/palgrave.jors.2602023.

[16] K. S. Shin, T. S. Lee, and H. Kim, "An application of support vector machines in bankruptcy prediction model" *Expert systems with applications*, vol. 28, Jan. 2005, pp. 127-135, doi: 10.1016/j.eswa.2004.08.009.

[17] T. V. Gestel, B. Baesens, J. A. Suykens, D. Van den Poel, D.-E. Baestaens, and B. Willekens, "Bayesian kernel based classification for financial distress detection," *European journal of operational research*, vol. 172, Aug. 2006, pp. 979-1003, doi: 10.1016/j.ejor.2004.11.009.

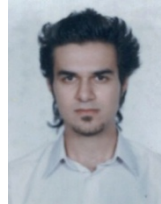
[18] S. H. Min, J. Lee, and I. Han, "Hybrid genetic algorithms and support vector machines for bankruptcy prediction," *Expert systems with applications*, vol. 31, Oct 2006, pp. 652-660, doi: 10.1016/j.eswa.2005.09.070.

[19] M. Bencic, N. Sarlija, and M. Zekic-Susac, "Modeling Small-Business Credit Scoring By Using Logistic Regression, Neural Networks and Decision Trees," *J. Intelligent systems in accounting, finance and management*, vol. 13, no. 3, pp.133-150, July.2005, doi: <http://dx.doi.org/10.1002/isaf.v13:3>.

[20] H. Abdou, J. Pointon, and A. Elmasry, "Neural Nets Versus Conventional Techniques in Credit Scoring in Egyptian Banking," *J. Expert systems with applications*, vol. 35, no. 3, pp. 1275-1292, Oct. 2008, doi: <http://dx.doi.org/10.1016/j.eswa.2007.08.030>.

[21] E. Angelini, G. D. Tollo, and A. Roil, "A Neural Network Approach for Credit Risk Evaluation," *The quarterly review of economics and finance*, vol. 48, no. 4, pp. 733-755, Nov. 2008, doi: 10.1016/j.qref.2007.04.001.

- [22] S. Pang and J. Gong, "C5.0 Classification Algorithm and Application on Individual Credit Evaluation of Banks," *Systems Engineering - Theory & Practice*, vol. 29, no. 12, pp. 94-104, Dec. 2009, doi:10.1016/S1874-8651(10)60092-0.
- [23] S. Cho, H. Hong and B. Ha, "A hybrid approach based on the combination of variable selection next term using previous term decision trees and case-based reasoning next term using the Mahalanobis distance: For bankruptcy prediction," *Expert systems with applications*, vol. 37, no. 4, 2010, doi:10.1016/j.eswa.2009.10.040.
- [24] N. C. Hsieh, "Hybrid mining approach in design of credit scoring model," *Expert systems with applications*, vol. 28, May. 2005, pp. 655-665, doi: 10.1016/j.eswa.2004.12.022.
- [25] C. L. Huang, M. C. Chen, and C. J. Wang, "Credit scoring with a data mining approach based on support vector machines," *Expert systems with applications*, vol. 33, Nov. 2007, pp. 847-856, 10.1016/j.eswa.2006.07.007.
- [26] D. Zhang, M. Hifi, Q. Chen, and W. Y., "A hybrid credit scoring model based on genetic programming and support vector machines," *Proc. The fourth international conference on natural computation, (ICNC 2008)*, IEEE, Oct. 2008, pp. 8-12, doi: 10.1109/ICNC.2008.205
- [27] W. Chen, C. Ma, and L. Ma, "Mining the customer credit using hybrid support vector machine technique," *Expert systems with applications*, vol. 36, May. 2009, pp. 7611-7616, doi: 10.1016/j.eswa.2008.09.054.
- [28] C.-F. Tsai, and M. L. Chen, "Credit rating by hybrid machine learning techniques," *Applied soft computing*, vol. 10, March. 2010, pp. 374-380, doi: 10.1016/j.asoc.2009.08.003.
- [29] M. F. Amin, M. M. Islam, and K. Murase, "Ensemble of single-layered complex-valued neural networks for classification tasks," *Neurocomputing*, vol. 72, June. 2009, pp. 2227-2234, doi: 10.1016/j.neucom.2008.12.028
- [30] L. Nanni, and A. Lumini, "An experimental comparison of ensemble of classifiers for bankruptcy prediction and credit scoring," *Expert systems with applications*, vol. 36, March. 2009, pp. 3028-3033, doi: 10.1016/j.eswa.2008.01.018.
- [31] L. Zhou and K. K. Lai, "Multi-agent ensemble models based on weighted least square SVM for credit risk assessment," *Proc. Global congress on intelligent systems, (GCIS 2009)*, IEEE, Aug. 2009, pp. 559-563, doi: 10.1109/GCIS.2009.283.
- [32] G. Wang, J. Hao, J. Ma, and H. Jiang, "A comparative assessment of ensemble learning for credit scoring," *Expert systems with applications*, vol. 38, Jan. 2011, pp. 223-230, doi: 10.1016/j.eswa.2010.06.048.
- [33] T. G. Dietterich, "Machine learning research: Four current directions," *AI Magazine*, vol. 18, no.4, pp. 97-136.
- [34] J. C. Bezdek, "Pattern recognition with fuzzy objective function algorithm," Newyork: Plenum press.
- [35] J. C. Dunn, "A fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters," *Journal of cybernetics and systems*, vol. 3, 1973, pp. 32-57, doi: 10.1080/01969727308546046.
- [36] T. Fawcett, "ROC Graphs: Notes and Practical Considerations for Researchers. HP laboratories Palo Alto: Intelligent enterprise technologies laboratory, 2004.



**Ahmad Ghodselahi** was born in September 1985 in Tehran, Iran. He received his B.Sc degree in industrial engineering from Azad university of Karaj, Iran, in 2008. He received his M.Sc. degree in Information Technology (IT) management from Tarbiat Modares University, Tehran, Iran, in March 2011. This author became a Member of IACSIT. His research interest is in Artificial Intelligence (AI).



**Ashkan Amirmadhi** was born in April 1984 in Tehran, Iran. He received his B.Sc degree in industrial engineering from Azad university of Karaj, Iran, in 2008. He is M.Sc. student of industrial management Islamic Azad University, Science & research branch, Tehran, Iran. His research interest is in Quality evaluation.