European Journal of Operational Research 222 (2012) 168–178

Contents lists available at SciVerse ScienceDirect

European Journal of Operational Research

journal homepage: www.elsevier.com/locate/ejor

Innovative Applications of O.R.

An empirical comparison of conventional techniques, neural networks and the three stage hybrid Adaptive Neuro Fuzzy Inference System (ANFIS) model for credit scoring analysis: The case of Turkish credit card data

Soner Akkoç *

Department of Banking and Finance, School of Applied Sciences, Dumlupınar University, Central Campus, 43100 Kütahya, Turkey

ARTICLE INFO

Article history:
Received 11 March 2011
Accepted 19 April 2012
Available online 27 April 2012

Keywords:
OR in banking
Credit scoring
Neuro fuzzy
ANFIS
Artificial neural networks

ABSTRACT

The number of Non-Performing Loans has increased in recent years, paralleling the current financial crisis, thus increasing the importance of credit scoring models. This study proposes a three stage hybrid Adaptive Neuro Fuzzy Inference System credit scoring model, which is based on statistical techniques and Neuro Fuzzy. The proposed model's performance was compared with conventional and commonly utilized models. The credit scoring models are tested using a 10-fold cross-validation process with the credit card data of an international bank operating in Turkey. Results demonstrate that the proposed model consistently performs better than the Linear Discriminant Analysis, Logistic Regression Analysis, and Artificial Neural Network (ANN) approaches, in terms of average correct classification rate and estimated misclassification cost. As with ANN, the proposed model has learning ability; unlike ANN, the model does not stay in a black box. In the proposed model, the interpretation of independent variables may provide valuable information for bankers and consumers, especially in the explanation of why credit applications are rejected.

1. Introduction

Recently, the number of Non-Performing Loans (NPLs) has rapidly increased, due to both the effects of global crisis and the appetite for increased risk, in parallel with banks granting more loans without enough evaluation. The global financial crisis has also affected the Turkish banking sector, and as of December 2009, the ratio of NPL to Gross Loans increased to 5.3% from 3.6%. When analyzed the consumer loans of the Turkish banking sector, can be seen in the following ratios; NPL in housing, 2.1%, in consumer loans, 4.4%, in vehicle loans, 10.7% and in credit card loans, 10.8% (BRSA, 2009). This situation has increased the credit risk of the financial sector and has brought about discussions relating to the credit scoring models' efficiency.

Nowadays, credit scoring models using statistical techniques, Operational Research, and Artificial Intelligence (AI) technologies are being developed (Thomas, 2000). Credit scoring models help credit institutions evaluate credit applications with respect to consumers' characteristics such as age, income, and marital status (Chen and Huang, 2003). The objective of credit scoring models is to sort the applications: those that have a high probability of performing financial obligations are assigned to a "good credit" group, and those that have a low probability of performing financial obligations are assigned to a "bad credit" group. Therefore, credit scoring models are basically classification problems (Hand, 1981; Hsieh, 2004; Lee et al., 2006). If these assignments are made accurately, more creditworthy applicants are granted credit, thereby increasing profits; non-creditworthy applicants are denied credit, thus decreasing losses (West, 2000).

In parallel with the growing credit volume of the financial sector, many different credit scoring models have been developed by banks and researchers in order to evaluate credit applications, including Linear Discriminant Analysis (LDA), Logistic Regression Analysis (LRA), Multivariate Adaptive Regression Splines (MARS), Classification and Regression Tree (CART), Artificial Neural Network (ANN), Support Vector Machines (SVM) and Genetic Algorithm (GA) (Abdou et al., 2008; Abdou, 2009; Angelini et al., 2008; Bellotti and Crook, 2009; Chen and Huang, 2003; Chuang and Lin, 2009; Cinko, 2006; Desai et al., 1996; Hsieh and Hung, 2010; Hsieh, 2004, 2005; Huang et al., 2006, 2007; Lee and Chen, 2005; Kim and Sohn, 2010; Lee et al., 2002, 2006; Lee, 2007; Li et al., 2006; Luo et al., 2009; Malhotra and Malhotra, 2003; Nanni and Lumini, 2009; Ong et al., 2005; Paleologo et al., 2010; Sustersic et al., 2009; Tong et al., 2012; Tsai and Wu, 2008; Tsai et al., 2009; West, 2000; West et al., 2005; Yu et al., 2008). Both LDA and LRA have been widely used in credit scoring (Chuang and Lin, 2009; Crook et al., 2007; Desai et al., 1996; Lee et al., 2006; Thomas, 2000; West, 2000). However; LDA makes some assumptions: a linear relationship among the independent variables and a normal

* Tel.: +90 274 2652031x4631; fax: +90 274 2652155.
E-mail addresses: akkocsoner@hotmail.com, sakkoc@dpu.edu.tr

0377-2217/$ - see front matter © 2012 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.ejor.2012.04.009

distribution of the variables. LDA is criticized because it is unable to provide justification for these assumptions (Thomas, 2000; West, 2000). LRA is used for making prediction on a data set with dichotomous outcomes. In contrast to LDA, LRA does not require the normality assumption. But both models assume that there is a linear relationship among variables, so both of these models may not have enough predictive accuracy in credit scoring (Lancher et al., 1995; Lee and Chen, 2005; Thomas, 2000; West, 2000).

When we look at the last two decades, ANN comes out as an important alternative in financial prediction studies, and draws attention from many researchers with its high prediction accuracy. ANN depends mainly upon transferring the processes of human brain to the computer environment. Unlike statistical techniques, ANN does not require any assumptions, and in research about credit scoring, ANN performs better than both LDA and LRA (Abdou et al., 2008; Chen and Huang, 2003; Desai et al., 1996; Lee and Chen, 2005; Lee et al., 2002; Malhotra and Malhotra, 2003; Sustersic et al., 2009; Tsai et al., 2009; West, 2000). However, ANN is also criticized; (i) for its long training process in developing the optimal network's architecture, (ii) for its inability to identify the relative importance of potential input variables, and (iii) because the model acts as a black box without logic or rule-based explanations for the input–output approximation; in other words, for its inability to explain the underlying principle for the decision to reject applications (Chen and Huang, 2003; Piramuthu, 1999; Trippi and Turban, 1996; West, 2000).

Neuro Fuzzy (NF) systems are relatively new hybrid AI technologies developed by using ANN and Fuzzy Logic (FL) simultaneously; there is little research in applying them to credit scoring models. The purpose of this study is to investigate the ability of the three stage hybrid Adaptive Neuro Fuzzy Inference System (ANFIS) credit scoring model, which is based on statistical techniques and NF, by using the credit card data of an international bank operating in Turkey. The performance of the proposed model is also compared with LDA, LRA, and ANN. The main contribution of this paper is that the three stage hybrid ANFIS credit scoring model is a competitive modeling approach for credit card evaluation. In addition to these, we try to interpret with 3D graphs how the proposed model makes credit decisions, which may provide valuable information for bankers and consumers.

The rest of the paper is organized as follows. We will review the literature of credit scoring in Section 2. Section 3 gives a brief outline of LDA, LRA, ANN and NF in building credit scoring model. Section 4 presents the data and the empirical results of the credit scoring models. Finally, concluding remarks are given in Section 5.

2. Literature review

Because of the vast volume of existing literature focusing on credit scoring, we will review only credit scoring studies using commonly used statistical techniques, ANN and NF, in this section. Durand (1941) first actualized credit scoring by LDA, by searching the differences between good and bad credit groups. Since then, statistical techniques, primarily LDA and LRA, have been used in financial prediction studies (Altman, 1968; Martin, 1977; Meyer and Pifer, 1970; Sinkey, 1975; West, 1985).

Since the 1990s, ANN has been also used in modeling credit scoring. Desai et al. (1996) developed credit scoring models with ANN on a data set of 1962 credit consumers, obtained from three different credit unions. Among the subjected models (ANN, LDA and LRA), the performance of ANN was the best, especially in predicting bad credit. Malhotra and Malhotra (2003) got similar results on a set of 1078 data, obtained from twelve different credit unions. West (2000) compared the performance of five different ANN models on credit scoring with LDA, LRA, k nearest neighbor, Kernel Destiny Estimation and CART. West (2000) stated that different ANN models can be used successfully in credit scoring and that LRA can be an alternative to ANN.

Lee et al. (2002) proposed a hybrid credit scoring model which integrates ANN with LDA. The performance of this proposed model has been found more successful than that of LDA, LRA, or ANN separately. In a study using Egypt's personal loan data, Abdou et al. (2008) found that ANN is more successful than LDA, LRA and Probit Analysis. Cinko (2006) obtained successful results using ANN and CART on credit card data. Angelini et al. (2008) achieved a 7% average error rate with ANN on a credit data set consisting of small and medium sized enterprises (SMEs) obtained from an Italian bank. Sustersic et al. (2009) found that credit scoring models developed with ANN are more successful than LRA on consumer credit data, after reducing independent variables with GA and Principal Component Analysis. Chen and Huang (2003) used GA in the process of transferring three rejected credit applications to the conditional acceptance group, and found that the ANN model is more successful than LDA and CART. Lee et al. (2006) have found the models which were developed using CART and MARS on credit card data were more successful than those using LDA, LRA and ANN. Lee and Chen (2005) compared the performance of the LDA, LRA, ANN, MARS, and MARS-ANN models on a data set of mortgage loans obtained from a local Taiwan bank. The best performance was obtained with the ANN model that used the variables found to be more important by MARS. Chuang and Lin (2009) obtained 76%, 76.5%, 77.5%, 79.5%, and 82.5% prediction performance from the LDA, LRA, CART, ANN, and MARS-ANN models respectively, on the German credit data. In the last part of the study, when the data that were transferred to the bad credit group were reevaluated with Case Based Reasoning (CBR), the accuracy rate of the model went up to 86%. Tsai et al. (2009) found that the Data Envelopment Analysis-LDA and ANN models were more successful than LDA and LRA on Taiwanese consumer credit data.

In recent years, ensemble classifiers have been proposed to improve the performance of credit scoring models. The main idea of ensemble classifiers is to combine a number of classifiers into one multiple classifier (Nanni and Lumini, 2006). West et al. (2005) ascertained that ensemble classifier-ANN models reduced the error rate of single classifiers by 3% or 5%. Yu et al. (2008) found that while ANN and SVM are more successful than LRA among single classifiers, the best performance was obtained from ensemble classifiers-ANN. Similarly, in the study done by Nanni and Lumini (2009), ANN was determined to be the best among single classifiers, but the best performance was generally obtained from the random subspace ensemble of classifier with the Levenberg–Marquardt neural net model. In the study of Tsai and Wu (2008), ensemble classifiers-ANN performed better in only one of the three datasets. Hsieh and Hung (2010) developed ensemble classifier credit scoring models after they separated German credit data into good, bad, and borderline groups with Cluster Analysis. Finlay (2011) compared the performance of multiple classifiers and found that Error Trimmed Boosting outperformed all other multiple classifiers on UK credit data.

Most of the reviewed studies focus on whether or not banks should grant credit to consumers who apply to them. On the other hand, the models, which help to make decisions on how to evaluate some demands of the existing customers like raising credit limits, are called behavioral scoring models (Thomas, 2000). Hsieh (2004) has developed behavioral scoring models on the credit card data with self-organizing map ANN. In this research, bank consumers are classified into three major profitable groups, and the results of this study can be used in developing marketing strategies. In another study, Hsieh (2005) drew the conclusion that cluster analysis raised the performance of credit scoring models based on ANN.

In credit scoring, hybrid models, which are formed by using more than one AI technology, are also used. The advantage of hybrid credit scoring models is that each AI technology brings its own strengths. NF systems are relatively new hybrid AI technologies; in credit scoring, there is little research that has used these systems. Piramuthu (1999) found that ANN was more successful than NF in credit scoring, but he also stated the necessity of using NF when it is important to know how credit decisions were made. In the study by Malhotra and Malhotra (2002), NF credit scoring models performed better than models using LDA. At the same time they interpreted why credit applications were accepted or not. Odeh et al. (2010) built LRA, ANN and ANFIS credit scoring models on data from US bank loans, and stated that benchmarking banks' internal rating system may be necessary.

When credit scoring models in literature are examined, it is seen that researchers focus on a high accuracy rate. Credit scoring models have been developed by the credit institutions and researchers with various credit data such as credit cards, consumer loans, mortgage loans, SME loans and corporate loans. It is hard to say which model gives the best result in every case when we take into consideration differently formed credit datasets as well as different datasets used in many countries. So, the benefit of credit scoring models can be raised by using more than one model as a decision support system.

3. Research methodology

3.1. Linear discriminant analysis

LDA was first proposed by Fisher (1936) as a discrimination and classification technique. LDA investigates whether there is an explicit difference or not among two or more groups, depending on a group of independent variables. LDA is an appropriate statistical technique when the dependent variable is categorical and the independent variables are metric (Malhotra and Malhotra, 2003). The model that LDA has is stated in Eq. (1):

$$Z_i = b_0 + b_1 x_{i1} + b_2 x_{i2} + \cdots + b_m x_{im} \quad (1)$$

where Z_i is a discriminant score, b_0 is the intercept term, and b_i ($i = 1, \ldots, m$) represents the estimated regression coefficient associated with the corresponding independent variables x_i ($i = 1, \ldots, m$). LDA has been widely applied in financial prediction studies (Altman, 1968; Deakin, 1972; Desai et al., 1996; Lee et al., 2002; Malhotra and Malhotra, 2002, 2003).

3.2. Logistic regression analysis

LRA is one of the most popular statistical modeling techniques for classification problems in which the probability of a dichotomous outcome depends on a group of independent variables. The model that LRA has is stated in Eq. (2):

$$\ln[\mathrm{pi}/(1 - \mathrm{pi})] = b_0 + b_1 x_{i1} + b_2 x_{i2} + \cdots + b_m x_{im} \quad (2)$$

where pi is the probability of the outcome of interest, b_0 is the intercept term, and b_i ($i = 1, \ldots, m$) represents the coefficient associated with the corresponding variables x_i ($i = 1, \ldots, m$). The dependent variable is the logarithm of the odds, $\{\ln[\mathrm{pi}/(1 - \mathrm{pi})]\}$, which is the logarithm of the ratio of two probabilities of the outcome of interest (Lee et al., 2002). The objective of a LRA model in credit scoring is to determine the conditional probability of a specific observation belonging to a class, given the values of the independent variables of that credit applicant (Lee and Chen, 2005). LRA does not necessarily require the assumptions of LDA. However, Harrell and Lee (1985) found that LRA is as efficient and accurate as LDA even though the assumptions of LDA are satisfied.

LRA has also been widely applied in financial prediction studies (Aziz et al., 1988; Foreman, 2003; Gentry et al., 1985; Henley, 1995; Keasey and Watson, 1987; Laitinen, 1999; Laitinen and Laitinen, 2000; Ohlson, 1980; Tseng and Lin, 2005). Although the LRA model can perform well in many applications, the accuracy of LRA decreases when the relationships of the system are non-linear. ANN has been proposed to deal with this problem (Ong et al., 2005).

3.3. Artificial neural networks

ANN, which was developed by simulating working principles of the human brain, is a flexible non-linear modeling tool. The human being's learning ability is transferred to a computer environment with ANN. In other words, ANN has ability to learn from examples. ANN is composed of a number of processing elements, which come together within the frame of particular rules which are called neurons or nodes (Haykin, 1999; Zhang et al., 1998). An ANN generally consists of three layers of interconnected neurons. A three layer ANN is shown in Fig. 1. The first layer is called the input layer, where external information, corresponding to independent variables in statistics, is received. Each neuron in the input layer sends signals to the hidden layer. Information received from the input layer is processed in the hidden layer. The output layer transmits the information outside of the network which corresponds to a dependent variable in statistics.

Since the 1990s, ANNs have been widely used in financial prediction studies (Akkoç, 2007; Alam et al., 2000; Bell, 1997; Davalos et al., 1999; Desai et al., 1996; Jensen, 1992; Jo et al., 1997; Lee and Chen, 2005; Lee et al., 2002; Lee and Chen, 2005; Leshno and Spector, 1996; Malhotra and Malhotra, 2003; Odom and Sharda, 1990; Piramuthu, 1999; Ravi and Pramodh, 2008; Salchenberger et al., 1992; Swicegood and Clark, 2001; Tam, 1991; Tam and Kiang, 1992; Tan and Dihardjo, 2001; Tsukuda and Baba, 1994; West, 2000; Wilson and Sharda, 1994; Yang et al., 1999; Yildiz, 2001; Zhang et al., 1999). The majority of these studies report that prediction accuracies of ANNs are better than conventional statistical techniques. Although ANN can be applied successfully in many fields, it has some disadvantages. ANN requires a long training process in developing the optimal model. ANN has also been criticized for lack of theory. There is no opportunity to explain the result produced by ANN, in other words, the model acts as a black box (Chen and Huang, 2003; Piramuthu, 1999; Trippi and Turban, 1996; West, 2000). Hybrid models that are formed by using at least two AI technologies such as NF together, can remove these disadvantages and produce promising results.

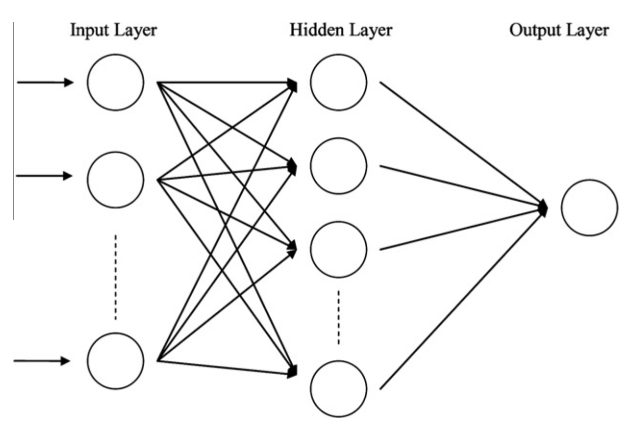

Fig. 1. A three layer ANN.

3.4. Neuro fuzzy systems

Zadeh (1965) introduced FL, a mathematical system which deals with modeling imprecise information in the form of linguistic terms. The point, where FL emerges becomes the crisis about the classic set theory. In classic set theory, there are two possibilities: an individual is a member of a set or is not a member of a set. But the human reasoning process works in a different, less dichotomous way. So in FL, it is possible that one individual can be a member of more than one set in a certain degree by means of membership functions. A is defined as a fuzzy set in Eq. (3).

$$A = \{(x, \mu A(x)) | x \in X\}, \tag{3}$$

This equation $\mu A(x)$ shows a membership function which gets a value between 0 and 1, and x shows a member of the set A. FL represents models using if-then rules. For example;

If *age* is **average** and *current work duration* is **high**, then *risk* is **low**

where *age, current work duration* and *risk* are linguistic variables; **low, average** and **high** are linguistic values that are identified by membership functions.

Each AI technology has a unique ability. ANN carries out machine learning by simulating the human being's neural system. FL is very similar to a human being's reasoning. But these technologies have some unique disadvantages. Not being able to make a comment on how the produced solution regarding a problem is produced by ANN; in other words, the information that stays in the black box is an important disadvantage for ANN. An important disadvantage of FL is not having the ability to learn. In parallel to AI technology development, combinations of these technologies have come into question. So the disadvantages of these technologies are removed when they are combined into one model. To take advantage of the learning capability of ANN and the modeling superiority of FL, these technologies are used simultaneously; this is called NF. Recently NF systems have gained a lot of interest, because they have the advantages of both ANN and FL. NF systems have the ability to apply human experience to problems by using fuzzy rules (Jang et al., 1997). The most important advantage for NF models over other non-linear AI technologies is their learning capability by using verbal variables (Abonyi, 2003). Another advantage of NF is the ability to make a comment on how the model produced the output value; in other words, the NF model would not remain a black box. NF has been applied to few researches for financial prediction (Akkoç, 2007; Chen et al., 2009; Malhotra and Malhotra, 2002; Piramuthu, 1999; Yildiz and Akkoç, 2009).

3.4.1. Adaptive neuro fuzzy inference system (ANFIS)

ANFIS, which was developed by Jang (1993), is a kind of NF system. ANFIS utilizes human expertise in the form of fuzzy if-then rules. ANFIS has the ability to construct models with only target sample data and exhibits fault tolerance. In other words, ANFIS determines itself appropriate parameters to provide the best learning by describing membership functions and entering data into the system (Jang et al., 1997). The most important feature that separates ANFIS from ANN is the obtained model that can be presented with rules like "If...then..." To give two fuzzy if-then rules example, for a first order Sugeno model the two rules will be below:

$$\begin{aligned} &\text{Rule} - 1: \text{ If } x = A_1 \text{ and } y = B_1 \text{ then } f_1 = p_1x + q_1y + r_1 \\ &\text{Rule} - 2: \text{ If } x = A_2 \text{ and } y = B_2 \text{ then } f_2 = p_2x + q_2y + r_2 \end{aligned} \tag{4}$$

where x and y are independent variables, A_i and B_i are fuzzy sets (linguistic labels like; low and high), p_i, q_i, r_i are the parameters of dependent variable. The corresponding equivalent ANFIS architecture which consists of five layers is shown in Fig. 2.

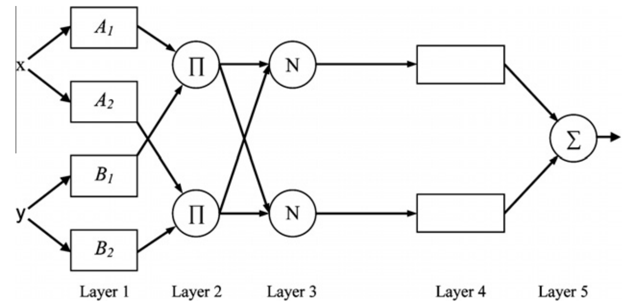

Fig. 2. ANFIS architecture.

Layer 1. Every node i in this layer, is an adaptive node with a node function described by;

$$\begin{aligned} O_{1,i} &= \mu A_i(x), \quad \text{for } i = 1, 2 \text{ and} \\ O_{1,i} &= \mu B_{i-2}(y), \quad \text{for } i = 3, 4 \end{aligned} \tag{5}$$

where x is the input node i, A_i and B_i are the linguistic label (low, medium, high, etc.) associated with this node function $O_{1,i}$ and $O_{1,i-2}$ are the membership function of A_i and B_i respectively. This research utilized $\mu A_i(x)$ and $\mu B_i(y)$ to be bell-shaped membership function with maximum and minimum equal to 1 and 0 respectively, such as:

$$A_i(x) = 1/1 + [(x - c_i/a_i)^2]^{b_i}$$

where a_i, b_i and c_i are the premise parameters of the membership function.

Layer 2. Every node in this layer is a fixed node labeled Π which multiplies the incoming signals and sends the product out. The outputs of this layer which represents firing strength of the rules can be represented as:

$$O_{2,i} = w_i = A_i(x) \text{ x }_{Bi}(y), \quad i = 1, 2. \tag{6}$$

Layer 3. Every node in this layer is a fixed node. Every node is labeled N. The ith node calculates the ratio of the ith rules firing strength to the sum of all rule's firing strength. In other words, this layer normalizes firing strength of the node i.

$$O_{3,i} = \overline{w_i} = \frac{w_i}{w_1 + w_2}, i = 1, 2. \tag{7}$$

Layer 4. This layer calculates the consequent. Every node in this layer is an adaptive node, with a node function,

$$o_{4,i} = \bar{w}_i f_i = \bar{w}(p_i x + q_i y + r_i), \tag{8}$$

where $\bar{w}_i$ is the output of layer 3 and p_i, q_i and r_i are the parameter set.

Layer 5. The single node in this layer is a fixed node labeled Σ that computes the overall output as the summation of all incoming signals.

$$O_{5,i} = \sum_i \bar{w}_i f_i = \frac{\sum_i w_i f_i}{\sum_i w_i} \tag{9}$$

3.5. Cross-validation

To provide a reliable estimate and minimize the impact of data dependency in developing credit scoring models, k-fold cross validation is used among the researchers (Desai et al., 1996; Hsieh, 2005; Lee and Chen, 2005; Sustersic et al., 2009; West, 2000). In this procedure, the entire credit data set is divided randomly into k mutually exclusive and approximately equal size subsets (also called folds). The classification model is trained and tested k times. Each time the model is trained using the $k-1$-folds of samples and a single fold is retained as the validation data for testing the model. The training sample is used to estimate the credit scoring model's parameters while the validation sample is used to test the generalization capability of the credit scoring model. The overall credit scoring accuracy is reported as an average across all k folds. In this research, we used 10-fold cross validation which is commonly used. An estimate from 10-fold cross validation is likely to be more reliable than an estimate from a common practice of using a single validation set.

4. Empirical study

4.1. Turkish credit card market and data

Although credit cards are widely used nowadays, they are relatively new when compared to the generally acceptable instrument of payment in Turkey. The usage of credit cards started to increase especially after 2001, in parallel with high GDP growth rate. This situation brought fierce competition environment to the Turkish banking sector. Concordantly, to have a greater market share, banks started to deliver credit cards without any serious evaluation. Namely, they did not ask for critical independent variable such as total income, total debt, and past credit behavior of the applicants. As a result, the volume of credit card transactions increased from about 12 billion dollars to 158 billion dollars in last decade (Fig. 3). On the other hand, the NPL rate in credit cards reached to 10.8% as of December 2009 (BRSA, 2009). As it is seen Fig. 3, the number of credit cards also increased and reached about 47 million in this process. But we should note that recently banks started to evaluate the credit card applicants seriously, especially due to the recent global financial crisis. Owing to the above mentioned reasons, in Turkey many credit cards data lacks important independent variables. Inherently, this situation affects credit scoring models' performance negatively. So it can be said that our models' performance is low when we compare it with the related literatures.

Credit cards have an important position in banks' credit portfolio in Turkey. The credit card portfolio forms 28% of personal loan portfolio of banks with about 24 billion dollars and takes the third place after mortgage and consumer loans in personal loan portfolio (BRSA, 2009). Therefore, it is seen that credit scoring models become more important for Turkish banking sector. When it is considered that credit card monthly interest rate varied between 2% and 7% during the past decade (http://www.tcmb.gov.tr/yeni/eng/), it is noted that the importance of credit scoring models increased a little more over the last decade.

In this study, credit scoring models are developed by using credit card data of an international bank operating in Turkey. The data set consists of a set of loans given in 2003 to a total of 2000 credit card applicants 1000 are good credit consumers while the remaining 1000 are bad credit consumers. A bad credit is defined as, at least three missed payments for the first year otherwise it is defined as a good credit. In order to minimize the possible bias associated with the random sampling, we used 10-fold cross validation. Each bank consumer in the credit dataset contains 11 independent variables which are shown in Table 1 and the dependent variable is the credit status of the consumer. Average age, average educational level, average total work duration, average total current work duration, and average home duration are higher for good applicants. Owing to having no missing data, we used all data without any adjustment while building credit scoring model. The SPSS version 13.0 for Windows is used for constructing the LDA and LRA credit scoring model. The neural network simulator Thinks Pro (version 1.05) is utilized to develop the ANN credit scoring model. The ANFIS Editor Graphical User Interface of Matlab (version 7.7) is utilized to develop the three stage ANFIS credit scoring model.

4.2. Credit scoring results of LDA

The overall LDA credit scoring accuracy is presented as an average across 10-folds. We have built 10 discriminant functions for each fold due to the reason of having 10-folds data. Seven significant independent variables of eleven were included in the discriminant functions, namely; marital status, age, education level, total work duration, total current work duration, car exist and working type. However, four significant variables were included in each discriminant function, namely; marital status, education level, total work duration and total current work duration. The credit scoring results of the validation sample using the obtained discriminant function from training data are shown in Table 2. As is seen in Table 2, the average classification rate of LDA is 57.2%. The model correctly classified 49.4% of good credit card applicants and 65% of the bad credit card applicants. In other words, with 50.6% (35%) class 0 (1) applicants misclassified as class 1 (0).

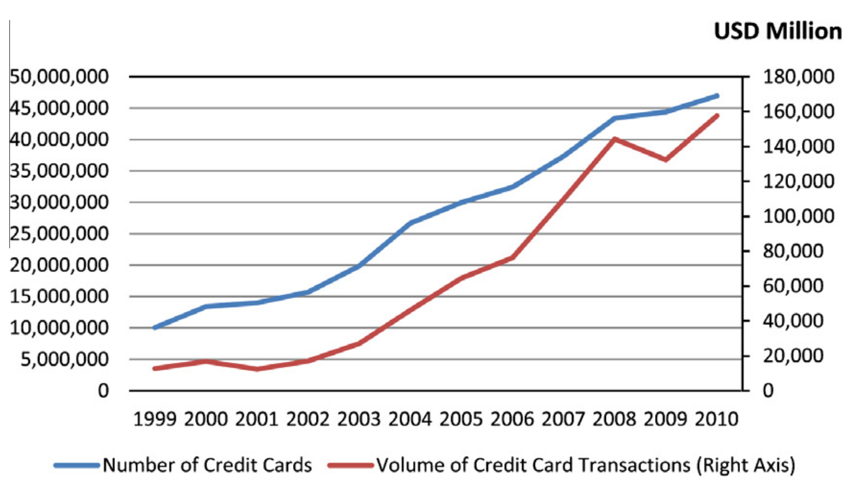

Fig. 3. Statistics about Turkish credit card market. Source: The Interbank Card Center (http://www.bkm.com.tr/bkm-en/)

Table 1
List of variables in constructing the credit scoring models.

Variables	Good				Bad			
	Mean	Std. dev.	Min.	Max.	Mean	Std. dev.	Min.	Max.
X_1 Gender	1.11	0.313	1	2	1.12	0.319	1	2
X_2 Marital status	1.37	0.502	1	3	1.49	0.52	1	3
X_3 Age	32.58	9.631	18	69	30.85	8.828	18	69
X_4 Education level	2.49	1.126	1	6	2.38	1.015	1	6
X_5 Basic working type	5.45	1.393	1	10	5.27	1.087	3	10
X_6 Total work duration	11.54	7.916	0	40	11.13	7.244	0	45
X_7 Total current work duration	4.94	5.39	0	40	3.28	3.937	0	35
X_8 Home duration	11.84	10.323	0	51	11.15	10.865	0	64
X_9 Residence type	2.78	1.318	1	5	2.97	1.189	1	5
X_{10} Car exist	1.8	0.399	1	2	1.89	0.316	1	2
X_{11} Region	4.41	1.837	1	7	4.3	1.913	1	7
Y Good-bad credit	0	0	0	0	1	0	1	1

Table 2
Credit scoring using discriminant analysis.

Actual class	Classified class (%)	
	0 (Good credit)	1 (Bad credit)
0 (Good credit)	49.4	50.6
1 (Bad credit)	35	65
Average correct classification rate: 57.2%		

4.3. Credit scoring results of LRA

The overall LRA credit scoring accuracy is presented as an average across 10-folds. We have built 10 regression functions for each fold due to the reason of having 10-folds data. Eight significant independent variables of eleven were included in the regression functions, namely; marital status, age, education level, total work duration, total current work duration, car exist, working type and region. However, four significant variables were included in each regression function, namely; marital status, education level, total work duration and total current work duration. The credit scoring results of the validation sample using the obtained regression function from training data are shown in Table 3. As is seen in Table 3, the average classification rate of LRA is 57.75%. The model correctly classified 50.54% of good credit card applicants and 65% of the bad credit card applicants. In other words, with 49.5% (35%) class 0 (1) applicants misclassified as class 1 (0).

4.4. Credit scoring results of ANN

In this research, there are 11 input nodes in the input layer and one output node in the output layer. The feedforward multilayer perceptrons are commonly used to solve the classification problems (Piramuthu, 1999). Since (Vellido et al., 1999) point out that more than 75% of business applications using ANN will adopt the BPN training algorithm, in constructing credit scoring model in this stage; this study also uses the feedforward multilayer perceptrons with the BPN training algorithm which is a supervised type of learning. Generally, the learning rate is set between 0.01 and 0.4, the momentum is set between 0.8 and 0.99 and the training lengths ranging from 1000 to 10,000 epochs (Chuang and Lin, 2009). Learning rate is crucial since smaller learning rates tend to slow down the learning process before convergence while larger ones, especially greater than 0.4, may cause network oscillation and convergence difficulty (Lee et al., 2006; Lee and Chen, 2005). As recommended by Cybenko (1989), Hornik et al. (1989) and Zhang et al. (1998) the single hidden layer network is sufficient to model any complex system, therefore, the designed network will have only one hidden layer. Determining the number of hidden nodes is generally associated with input nodes. The most commonly used way in determining the number of hidden nodes is via experiments or trial and error process. The number of hidden

Table 3
Credit scoring using logistic regression analysis.

Actual class	Classified class (%)	
	0 (Good credit)	1 (Bad credit)
0 (Good credit)	50.5	49.5
1 (Bad credit)	35	65
Average correct classification rate: 57.75%		

Table 4
Result of ANN with various hidden nodes.

Architecture	Average accuracy rate %
11-20-1	56.7
11-21-1	56.6
11-22-1	57.1[a]
11-23-1	55.5
11-24-1	56.95

[a] Best network architecture.

Table 5
Average accuracy rate of various learning parameters for the 11-22-1 architecture (%).

Learning rate	Momentum rate		
	0.7	0.8	0.9
0.001	54.1	53.9	53.75
0.005	55.55	55.55	57.35
0.01	55.85	57.75	58.6[a]
0.05	55.85	57.5	57.9
0.1	57.85	57.4	57.95
0.2	58.25	57.1	58
0.3	57.1	57.05	57.15
0.4	57	56.25	56.1
0.5	56.05	56.55	56.4

[a] Best average accuracy rate.

Table 6
Credit scoring using ANN.

Actual class	Classified class (%)	
	0 (Good credit)	1 (Bad credit)
0 (Good credit)	75	25
1 (Bad credit)	57.8	42.2
Average correct classification rate: 58.6%		

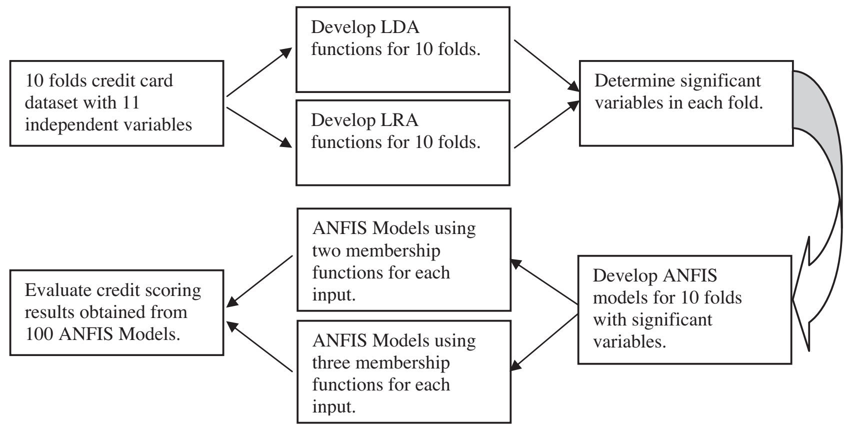

Fig. 4. The processes of proposed three stage hybrid ANFIS credit scoring model.

nodes to be tested $2n$, $2n \pm 1$ and $2n \pm 2$, n denotes input nodes (Hecht-Nielsen, 1990). To determine the optimal number of hidden nodes, 20, 21, 22, 23 and 24 was tested in every 10-fold, when the learning rate, momentum and training epochs are set to 0.2, 0.8 and 2000, respectively. In other words, we developed 50 models for this purpose.

The accuracy rate of each architecture is an average of 10 repetitions for every fold and is shown at Table 4. The network architecture with the highest accuracy rate (11-22-1) is considered as the optimal network architecture. After determining the optimal network architecture, various learning parameters were applied on this architecture. Nine different learning rates (from 0.001 to 0.5) and three different momentum rates (from 0.7 to 0.9) were tested, so here we have also developed 270 models. Average accuracy rate of various learning parameters for the 11-22-1 architecture is shown in Table 5. The average classification rate of the ANN model is 58.6% when the learning rate and momentum are set to 0.1 and 0.9 respectively. As is seen in Table 6, the model correctly classified 75% of good credit card applicants and 42.2% of the bad credit card applicants. In other words, with 25% (57.8%) class 0 (1) applicants misclassified as class 1 (0).

4.5. Credit scoring results of proposed three stage hybrid ANFIS model

Because of having too much rules and parameters with 11 independent variables, we do not have any chance to built ANFIS model. So this paper proposes LDA and LRA as feature selection methods. The process of proposed credit scoring model is shown in Fig. 4. As is seen in Fig. 4, before ANFIS models were developed we determined significant variables. Four variables were found significant by LDA and LRA in every fold namely; marital status, education level, total work duration and total current work duration. We used these significant variables as an input in the ANFIS models. We used *anfisedit* command to create the ANFIS *.fis* file. We loaded the data and developed the ANFIS with membership functions. Firstly, we developed ANFIS models by assigning two "tri", two "gbell", two "gauss", two "gauss2" and two "pi" membership functions for each input respectively. We repeated a similar process by assigning three membership functions for each input. So, we developed 100 models, because of having 10-folds credit card data. After assigning membership functions, ANFIS models build up the rules which are used to train the models. Every ANFIS model was trained with 500 epochs. When the training process of the models has been completed, the models make the credit scoring decision. Finally, by using *evalfis* command, we evaluate the results of developed models.

Average accuracy rate of various ANFIS architecture across 10-folds are shown in Table 7. The ANFIS architecture with the highest accuracy rate is considered as the optimal model. The optimal ANFIS model with a 60% average accuracy rate has three Gauss membership functions in each input. As is seen in Table 8, the proposed model correctly classified 51.4% of good credit card applicants and 68.6% of the bad credit card applicants. In other words, with 48.6% (31.4%) class 0 (1) applicants misclassified as class 1 (0).

As we mentioned before, the most important feature that separates ANFIS models from ANN is the set model that does not stay in a black box. In the ANFIS model, decision making by using the independent variables can be interpreted. The non-linear functions which the ANFIS model produces and allows us to interpret are presented in Figs. 5 and 6. The non-linear functions were developed on the k9[1] fold with the highest accuracy rate of 63.5%. In these Fig. 3D graphs demonstrate complicated relationships between the dependent variable and the independent variables. In the Fig. 5, functions, which are developed relating to Marital Status and Educational Level, are presented. Marital Status has three possible values, namely; 1-Married, 2-Single, 3-Widowed. Educational Level has six possible values, 1 indicates the lowest and six indicates the highest. Fig. 5 indicates that the probability of becoming default increases

Table 7
Result of three stage hybrid ANFIS model with different membership functions.

Type of membership functions	Average accuracy rate (%)	Average accuracy rate (%)
	Number of membership functions (2–2–2–2)	Number of membership functions (3–3–3–3)
Tri	59.2	59.75
Gbell	59.4	59.75
Gauss	58	60[a]
Gauss2	58.55	58.75
Pi	56.75	57.95

[a] Best average accuracy rate.

Table 8
Credit scoring using three stage hybrid ANFIS model.

Actual class	Classified class (%)	
	0 (Good credit)	1 (Bad credit)
0 (Good credit)	51.4	48.6
1 (Bad credit)	31.4	68.6
Average correct classification rate: 60%		

[1] We obtained similar functions from other folds.

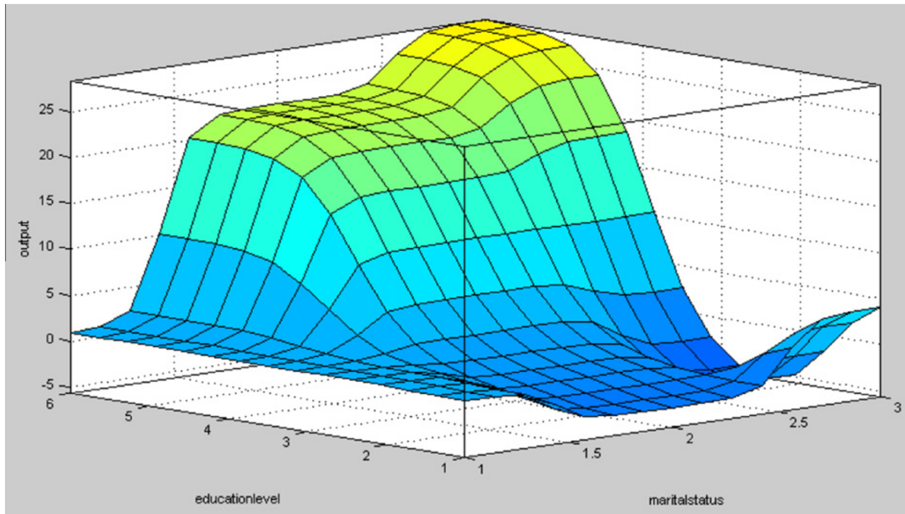

Fig. 5. The 3D graph for marital status and educational level.

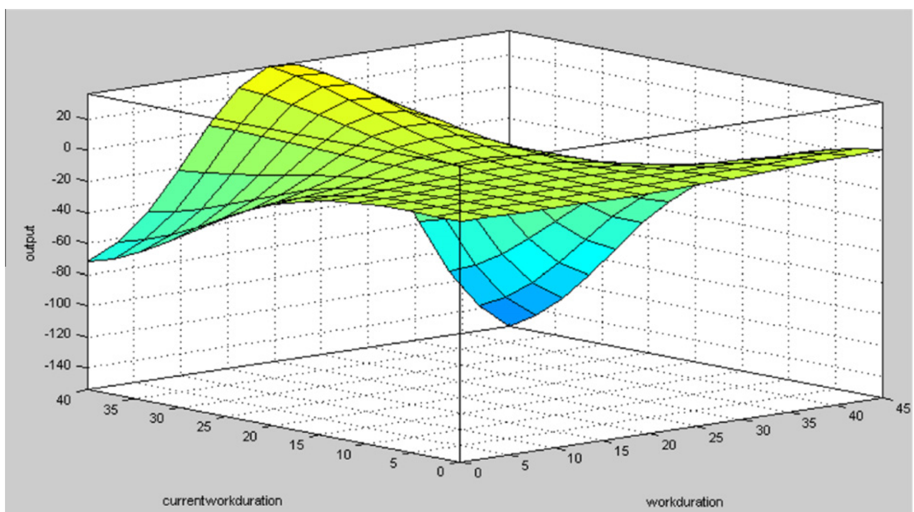

Fig. 6. The 3D graph for work duration and current work duration.

when Educational Level is low and Marital Status is Single or Widowed. In Fig. 5, this is shown in relatively dark places. On the other hand Fig. 5 also indicates that the probability of becoming financially responsible consumers increases in parallel with the increase of Educational Level, especially when the Educational Level above 3.

In the Fig. 6, functions, which are developed relating to Work Duration and Current Work Duration, are presented. Fig. 6 indicates that the probability of becoming default increases when Work Duration and Current Work Duration are very high. On the other hand, Fig. 6 also indicates that the probability of becoming financially responsible consumers increases when Current Work Duration is higher than 10 years and Work Duration is between 10 and 30 years. Briefly, according to the proposed model consumers who have high Education Level and Average Work Duration reduces credit risk of banks.

4.6. Evaluate the performance of different credit scoring models

4.6.1. Credit scoring result between constructed models

The credit data cannot easily be separated because the data at the time of application cannot capture the complexities which may lead to default in each applicant's life. So, low accuracy rates arise from credit data inherently (Bellotti and Crook, 2009). The quality of a rating system is measured by the Accuracy Ratio (AR) (Engelmann et al., 2003; BCBS, 2005). In order to evaluate the classification performance of four constructed credit scoring models, the classification accuracy rates are summarized in Table 9 and Area Under Curve (AUC) and AR in Table 10. Table 9 gives the percentage of correctly classified credit card loans in every fold. From the results, we can conclude that the proposed three stage hybrid ANFIS credit scoring model which is based on LDA, LRA and ANFIS, has the best classification capability in terms of average classification rate and AR. The performance of constructed models can be ranged below,

Three stage hybrid ANFIS > ANN > LRA > LDA.

This result also confirmed using a paired *t*-test. As the *p* values indicate, the proposed three stage hybrid ANFIS model is roundly better than LDA and LRA, while the difference is not as significant when compared to ANN.

4.6.2. Type I, Type II errors and estimated misclassification cost of constructed models

In order to evaluate the performance of credit scoring models, beside classification accuracy rates and AR, the misclassification

Table 9
Comparing classification results for constructed models.

Folds	LDA (%)	LRA (%)	ANN (%)	Three stage hybrid ANFIS (%)
1	59.5	60	60	60
2	56.5	55.5	58	56.5
3	59	59.5	60	61.5
4	59	61.5	60.5	61.5
5	59	59.5	61.5	62
6	57	59.5	54	61
7	51	50	60.5	53.5
8	55.5	55	53.5	60
9	64	65	59	63.5
10	51.5	52	59	60.5
Average	57.2	57.75	58.6	60
p-Value[a]		0.193	0.344	0.01
p-Value[b]			0.609	0.036
p-Value[c]				0.305

[a] The *p* values are for a two-tailed paired *t* test comparing the LDA results with the other three methods.
[b] The *p* values are for a two-tailed paired *t* test comparing the LRA results with the other two methods.
[c] The *p* value is for a two-tailed paired *t* test comparing the ANN results with the other method.

Table 10
Area Under Curve and Accuracy Ratio for constructed models.

Fold	LDA		LRA		ANN		Three stage hybrid ANFIS	
	AUC (%)	AR (%)	AUC (%)	AR (%)	AUC (%)	AR (%)	AUC (%)	AR (%)
1	67.8	35.6	67.9	35.8	64.1	28.2	67.1	34.2
2	59.6	19.2	59.5	19	57.4	14.8	59.1	18.2
3	64.1	28.2	64.5	29	62.4	24.8	66	32
4	64	28	64.9	29.8	67	34	62.5	25
5	62.3	24.6	62.4	24.8	69.3	38.6	60.9	21.8
6	65.3	30.6	63.4	26.8	54.3	8.6	64.8	29.6
7	53.3	6.6	54.4	8.8	61	22	56.7	13.4
8	60.1	20.2	60.2	20.4	60.7	21.4	64.1	28.2
9	66.5	33	66.5	33	66.6	33.2	66.8	33.6
10	58.1	16.2	58.7	17.4	62.1	24.2	63.3	26.6
Average	62.1	24.2	62.2	24.5	62.5	25	63.1	26.3

costs also can be taken into account. Mainly, classification problems have two types of errors, Type I (good credit is misclassified as bad credit) and Type II (bad credit is misclassified as good credit). In credit scoring applications, it is believed that the cost of Type I and Type II errors are significantly different. In other words, it is more important to classify a bad applicant correctly than it is to classify a good applicant correctly. As recommended Dr. Hofmann, who compiled the German credit data, stated that the relative costs of misclassification for Type I error is 1 and Type II error is 5 (West, 2000). This study also uses this relative cost ratio in order to estimate misclassification costs of the constructed credit scoring models with the following equation:

$$\text{Misclassification Cost} = C(1/0) \times P(1/0) \times \pi_1 + C(0/1) \times P(0/1) \times \pi_2 \tag{10}$$

where C (1/0) and C (0/1) are the corresponding misclassification cost of Type I and Type II errors. P (1/0) and P (0/1) represent the probabilities of Type I and Type II errors. π_1 and π_2 are the prior probability of good and bad credit. These values are set to 0.5 and 0.5, respectively, using the ratio of good and bad credit in the empirical dataset.

Table 11 shows that Type I errors are higher than Type II errors in LDA, LRA and the three stage hybrid ANFIS whilst Type I error is

Table 11
Type I, Type II errors and misclassification costs of credit scoring models.

	Type I (%)	Type II (%)	Estimated misclassification cost
LDA	50.6	35	1.128
LRA	49.5	35	1.1225
ANN	25	57.8	1.57
Three stage hybrid ANFIS	48.6	31.4	1.028

lower than Type II error in ANN. Estimated misclassification cost results are generally consistent with the previous analysis. The proposed three stage hybrid ANFIS credit scoring model has the lowest estimated misclassification cost. Because of high Type II error rate, the estimated misclassification cost of ANN is highest. The estimated misclassification costs of constructed models can be ranged below,

Three stage hybrid ANFIS < LRA < LDA < ANN.

As a result, we can conclude that the credit scoring result of the proposed three stage hybrid ANFIS model outperforms the conventional and commonly used LDA, LRA and ANN models. Although the proposed model performs best, we should note that, the accuracy rates and AR were unsatisfactory. We believe that low accuracy rates arise from the independent variables. In the meanwhile, this study shows that, without important variables such as, income, past credit behavior, debt, mortgage payment, it is very hard to built fulfilling credit scoring model.

5. Conclusions

The credit volume of financial industry has increased rapidly in recent years. At the same time, the NPL volume has also increased, in parallel with the global financial crisis. So it has become more and more important for credit institutions to find good consumers capable of fulfilling their financial obligations. To do that, financial institutions use credit scoring models as a decision support system to evaluate credit applications. Many credit scoring models have been developed for better credit approval processes.

ANN and FL gained a lot of interest in the last two decades. ANN has learning ability, but cannot be interpreted because the decision-making process stays in a black box. FL systems have interpretable linguistic rules, but they have no learning ability. NF systems use ANN and FL simultaneously, and have the advantages of both. This study proposes a three stage hybrid ANFIS credit scoring model, based on statistical techniques and NF. To demonstrate the performance of the proposed model, credit scoring tasks were performed on credit card data of an international bank operating in Turkey. The performance of the three stage hybrid ANFIS model is also compared with LDA, LRA and ANN. Our empirical results show that the proposed three stage hybrid ANFIS model has the best credit scoring capability in terms of both the classification accuracy rate and accuracy ratio and estimated misclassification cost. With the proposed three stage hybrid ANFIS credit scoring model, we can also interpret the credit scoring decision process. This may provide valuable information for bankers and consumers, especially in explanations of why credit applications are rejected. In this context, 3D graphs show that an applicant with a lower education level and very high work duration may not be able to fulfill financial obligations. In other words, the probability of falling into default increases. On the other hand, an applicant with a higher educational level and average work duration (especially above 10 years) may be able to fulfill financial obligations. So according to our results, it can be recommended that to reduce credit risk, financial institutions operating in Turkey should take these properties into account in the credit evaluation process.

Future research may aim to find important predictor variables by using GA and MARS, and constructing a credit scoring model with these variables. Fuzzy credit scoring models may be developed with the rules suggested by the bankers and experts. A NF credit scoring model can be constructed by using different independent variables. The progress of consumer credibility can be evaluated with a time series. After investigating the characteristics of different credit data such as credit cards, consumer loans, mortgage loans, SME loans and corporate loans, effective credit scoring models can be developed for every type of credit data separately.

Acknowledgments

This research was supported by The Council of Higher Education of Turkey and Dumlupınar University. The author thanks Richard Skolnik and Eric Tsai SUNY-Oswego, Yildiray Yildirim Syracuse University, Bekir Mumyakmaz Dumlupınar University, and an anonymous referee for their constructive and helpful comments, and the anonymous bank for providing the credit card data that makes this research possible.

References

Abdou, H.A., 2009. Genetic programming for credit scoring: the case of Egyptian public sector banks. Expert Systems with Applications 36, 11402–11417.
Abdou, H.A., Pointon, J., Ahmed, E.-M., 2008. Neural nets versus conventional techniques in credit scoring in Egyptian banking. Expert Systems with Applications 35, 1275–1292.
Abonyi, J., 2003. Fuzzy Model Identification for Control. Birkhauser.
Akkoç, S., 2007. Bankruptcy Prediction Using Neurofuzzy Modeling and an Empirical Analysis. Doctoral Dissertation. The Dumlupınar University, Kütahya, Turkey.
Alam, P., Booth, D., Lee, K., Thordarson, T., 2000. The use of fuzzy clustering algorithm and self-organizing neural network for identifying potentially failing banks: an experiment study. Expert Systems with Applications 18, 185–199.
Altman, E.I., 1968. Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. Journal of Finance 23 (4), 589–609.
Angelini, E., Tollo, G., Roli, A., 2008. A neural network approach for credit risk evaluation. Quarterly Review of Economics and Finance 48, 733–755.
Aziz, A., Davit, C.E., Gerald, H.L., 1988. Bankruptcy prediction – an investigation at cash flow based models. Journal of Management 25, 419–437.
Banking Regulation and Supervision Agency, 2009. Financial Market Report, Issue:16 Ankara. <http://www.bddk.org.tr/WebSitesi/english/Reports/Financial_Markets_Report/8127FMR_Dec_2009.pdf>.
Basel Committee on Banking Supervision (BCBS), 2005. Studies on the Validation of Internal Rating Systems. Working Paper 14, Basel.
Bell, T.B., 1997. Neural nets or the logit model: a comparison of each model's ability to predict commercial bank failures. Intelligent Systems in Accounting, Finance and Management 6, 249–264.
Bellotti, T., Crook, J., 2009. Support vector machines for credit scoring and discovery of significant features. Expert Systems with Applications 36, 3302–3308.
Chen, H.J., Huang, S.Y., Lin, C.S., 2009. Alternative diagnosis of corporate bankruptcy: a neuro fuzzy approach. Expert Systems with Applications 36, 7710–7720.
Chen, M.C., Huang, S.H., 2003. Credit scoring and rejected instances reassigning through evolutionary computation techniques. Expert Systems with Applications 24, 433–441.
Chuang, C-L., Lin, R-H., 2009. Constructing a reassigning credit scoring model. Expert Systems with Applications 36, 1685–1694.
Cinko, M., 2006. Comparison of credit scoring techniques. Istanbul Commerce University Social Science Journal 9, 143–153.
Crook, J.N., Edelman, D.B., Thomas, L.C., 2007. Recent developments in consumer credit risk assessment. European Journal of Operational Research 183, 1447–1465.
Cybenko, G., 1989. Approximation by superpositions of a sigmoidal function. Mathematical Control Signals Systems 2, 303–314.
Davalos, S., Gritta, R.D., Chow, G., 1999. The application of a neural network approach to predicting bankruptcy risks facing the major US air carriers: 1979–1996. Journal of Air Transport Managements 5 (2), 81–86.
Deakin, E.B., 1972. A dicriminant analysis of predictors of business failure. Journal of Accounting Research 10, 167–168.
Desai, V.S., Crook, J.N., Overstreet, G.A., 1996. A comparison of neural networks and linear scoring models in the credit union environment. European Journal of Operational Research 95, 24–37.
Durand, D., 1941. Risk Elements in Consumer Instalment Financing. National Bureau of Economic Research, New York.
Engelmann, B., Evelyn, H., Dirk, T., 2003. Testing rating accuracy. Risk January, 82–86.
Finlay, S., 2011. Multiple classifier architectures and their application to credit risk assessment. European Journal of Operational Research 210, 368–378.
Fisher, R.A., 1936. The use of multiple measurements in taxonomic problems. Annual Eugenics 7, 179–188.
Foreman, R.D., 2003. A logistic analysis of bankruptcy within the US local telecommunications industry. Journal of Economics & Business 55, 135–166.
Gentry, J.A., David, T.W., Paul, N., 1985. Classifying bankrupt firms with fund flow components. Journal of Accounting Research 23, 146–160.
Hand, D.J., 1981. Discrimination and Classification. Wiley, New York.
Harrell, F.E., Lee, K.L., 1985. A comparison of the discrimination of discriminant analysis and logistic regression. In: Se, P.K. (Ed.), Biostatistics: Statistics in Biomedical, Public Health, and Environmental Sciences. North-Holland, Amsterdam.
Haykin, S., 1999. Neural Networks: A Comprehensive Foundation, second ed. Prentice Hall Int. Inc., New Jersey.
Hecht-Nielsen, R., 1990. Neurocomputing. Addison-Wesley, Menlo Park, CA.
Henley, W.E., 1995. Statistical Aspects of Credit Scoring. Doctoral Dissertation. The Open University, Milton Keynes, UK.
Hornik, K., Stinchcombe, M., White, H., 1989. Multilayer feedforward networks are universal approximators. Neural Networks 2, 359–366.
Hsieh, N.C., 2005. Hybrid mining approach in the design of credit scoring models. Expert Systems with Applications 28, 655–665.
Hsieh, N.C., Hung, L.P., 2010. A data driven ensemble classifier for credit scoring analysis. Expert Systems with Applications 37, 534–545.
Hsieh, N.C., 2004. An integrated data mining and behavioral scoring model for analyzing bank customers. Expert Systems with Applications 27, 623–633.
Huang, C.L., Chen, M.C., Wang, C.J., 2007. Credit scoring with a data mining approach based on support vector machines. Expert Systems with Applications 33 (4), 847–856.
Huang, J., Tzeng, G., Ong, C.S., 2006. Two-stage genetic programming (2SGP) for the credit scoring model. Applied Mathematics and Computation 174 (2), 1039–1053.
Jang, J-S.R., 1993. ANFIS: adaptive-network-based fuzzy inference system. IEEE Transactions on Systems, Man, and Cybernettics 23 (3), 665–685.
Jang, J.-S.R., Sun, C.T., Mizutani, E., 1997. Neuro-Fuzzy and Soft Computing. Prentice-Hall Inc..
Jensen, H.L., 1992. Using neural networks for credit scoring. Managerial Finance 18, 15–26.
Jo, H., Han, I., Lee, H., 1997. Bankruptcy prediction using case-based reasoning, neural networks, and discriminant analysis. Expert Systems with Applications 13 (2), 97–108.
Keasey, K., Watson, R., 1987. Non-financial symptoms and the prediction of small company failure: a test of Argenti's Hypotheses. Journal of Business Finance & Accounting 14, 335–354.
Kim, S.H., Sohn, Y.S., 2010. Support vector machines for default prediction of SMEs based on technology credit. European Journal of Operational Research 201, 838–846.
Laitinen, E.K., 1999. Predicting a corporate credit analyst's risk estimate by logistic and linear models. International Review of Financial Analysis 8 (2), 97–121.
Laitinen, E.K., Laitinen, T., 2000. Bankruptcy prediction: application of the Taylor's expansion in logistic regression. International Review of Financial Analysis 9 (4), 327–349.
Lancher, R.C., Coats, P.K., Shanker, C.S., Fant, L.F., 1995. A neural network for classifying the financial health of a firm. European Journal of Operational Research 85 (1), 53–65.
Lee, T.S., Chen, I.F., 2005. A two-stage hybrid credit scoring model using artificial neural networks and multivariate adaptive regression splines. Expert Systems with Applications 28, 743–752.
Lee, T.S., Chiu, C.C., Lu, C.J., Chen, I.F., 2002. Credit scoring using the hybrid neural discriminant technique. Expert Systems with Applications 23 (3), 245–254.
Lee, T.S., Chiu, C.C., Chou, Y.C., Lu, C.J., 2006. Mining the customer credit using classification and regression tree and multivariate adaptive regression splines. Computational Statistics and Data Analysis 50, 1113–1130.
Lee, Y.C., 2007. Application of support vector machines to corporate credit rating prediction. Expert Systems with Applications 33 (1), 67–74.
Leshno, M., Spector, Y., 1996. Neural network prediction analysis: the bankruptcy case. Neurocomputing 10, 125–147.
Li, S.T., Shiue, W., Huang, M.H., 2006. The evaluation of consumer loans using support vector machines. Expert Systems with Applications 30 (4), 772–782.
Luo, S.T., Cheng, B.W., Hsieh, C.H., 2009. Prediction model building with clustering-launched classification and support vector machines in credit scoring. Expert Systems with Applications 36, 7562–7566.
Malhotra, R., Malhotra, D.K., 2002. Differentiating between good credits and bad credits using neuro-fuzzy systems. European Journal of Operational Research 136, 190–211.
Malhotra, R., Malhotra, D.K., 2003. Evaluating consumer loans using neural networks. Omega 31 (2), 83–96.
Martin, D., 1977. Early warning of bank failure. Journal of Banking and Finance 1, 249–276.
Meyer, P.A., Pifer, H.W., 1970. Prediction of bank failures. Journal of Finance 25, 853–858.
Nanni, L., Lumini, A., 2009. An experimental comparison of ensemble of classifiers for bankruptcy prediction and credit scoring. Expert Systems with Applications 36, 3028–3033.
Odeh, O.O., Featherstone, M.A., Das, S., 2010. Predicting credit default: comparative results from an artificial neural network, logistic regression and adaptive neuro-fuzzy inference system. International Research Journal of Finance and Economics 42, 7–18.

Odom, M.D., Sharda, R., 1990. A neural network model for bankruptcy prediction. In: Proceedings of the International Joint Conference on Neural Networks, II, pp. 163–167.

Ohlson, J.A., 1980. Financial ratios and probabilistic prediction of bankruptcy. Journal of Accounting Research 18 (1), 109–131.

Ong, C., Huang, J., Tzeng, G., 2005. Building credit scoring models using genetic programming. Expert Systems with Applications 29 (1), 41–47.

Paleologo, G., Elisseeff, A., Antonini, G., 2010. Subagging for credit scoring models. European Journal of Operational Research 201, 490–499.

Piramuthu, S., 1999. Financial credit-risk evaluation with neural and neuro fuzzy systems. European Journal of Operational Research 112, 310–321.

Ravi, V., Pramodh, C., 2008. Threshold accepting trained principal component neural network and feature subset selection: application to bankruptcy prediction in banks. Applied Soft Computing 8 (4), 1539–1548.

Salchenberger, L., Mine, C., Lash, N., 1992. Neural networks: a tool for predicting thrift failures. Decision Sciences 23, 899–916.

Sinkey, J.F., 1975. A multivariate statistical analysis of the characteristics of problem banks. Journal of Finance 30, 21–36.

Sustersic, M., Mramor, D., Zupan, J., 2009. Consumer credit scoring models with limited data. Expert Systems with Applications 36, 4736–4744.

Swicegood, P., Clark, J.A., 2001. Off-site monitoring for predicting, bank under performance: a comparison of neural networks, discriminant analysis and professional human judgment. Intelligent Systems in Accounting, Finance and Management, 10, 169–18.

Tam, K.Y., Kiang, M., 1992. Predicting bank failures: a neural network approach. Decision Sciences 23, 926–947.

Tam, K.Y., 1991. Neural network models and the prediction of bank bankruptcy. Omega 19 (5), 429–445.

Tan, N.W., Dihardjo, H., 2001. A study on using artificial neural networks to develop an early warning predictor for credit union financial distress with comparison to the probit model. Managerial Finance 27 (4), 56–77.

Thomas, L.C., 2000. A survey of credit and behavioural scoring: forecasting financial risk of lending to consumers. International Journal of Forecasting 16, 149–172.

Tong, N.C.E., Mues, C., Thomas, C.L., 2012. Mixture cure models in credit scoring: if and when borrowers default. European Journal of Operational Research 218, 132–139.

Trippi, R.R., Turban, E., 1996. Neural Networks in Finance and Investing. Irwin Professional Pub., Chicago.

Tsai, C.F., Wu, J.W., 2008. Using neural network ensembles for bankruptcy prediction and credit scoring. Expert Systems with Applications 34, 2639–2649.

Tsai, M.C., Lin, S.P., Cheng, C.C., Lin, Y.P., 2009. The consumer loan default predicting model – an application of DEA-DA and neural network. Expert Systems with Applications 36, 11682–11690.

Tseng, F.M., Lin, L., 2005. A quadratic interval logit model for forecasting bankruptcy. Omega 33, 85–91.

Tsukuda, J.S., Baba, I., 1994. Predicting Japanese corporate bankruptcy in terms of finance data using neural network. Computers and Industrial Engineering 27 (1–4), 445–448.

Vellido, A., Lisboa, P.J.G., Vaughan, J., 1999. Neural networks in business: a survey of applications (1992–1998). Expert Systems with Applications 17, 51–70.

West, R.C., 1985. A factor analytic approach to bank condition. Journal of Banking and Finance 9, 253–266.

West, D., 2000. Neural network credit scoring models. Computers and Operations Research 27, 1131–1152.

West, D., Dellana, S., Qian, J., 2005. Neural network ensemble strategies for financial decision applications. Computers and Operations Research 32, 2543–2559.

Wilson, R.L., Sharda, R., 1994. Bankruptcy prediction using neural networks. Decision Support Systems 11, 545–557.

Yang, Z.R., Platt, M.B., Platt, H.D., 1999. Probabilistic neural networks in bankruptcy prediction. Journal of Business Research 44 (2), 67–74.

Yildiz, B., Akkoç, S., 2009. Predicting bank bankruptcies with neuro fuzzy method. Journal of BRSA Banking and Financial Markets 3 (1), 9–35.

Yildiz, B., 2001. Prediction of financial failure with artificial neural network technology and an empirical application on publicly held companies. Istanbul Stock Exchange Review 5 (17), 47–62.

Yu, L., Wang, S., Lai, K.K., 2008. Credit risk assessment with a multistage neural network ensemble learning approach. Expert Systems with Applications 34, 1434–1444.

Zadeh, L., 1965. Fuzzy sets. Information and Control 8 (3), 338–353.

Zhang, G., Hu, M.Y., Patuwo, B.E., Indro, D.C., 1999. Artificial neural networks in bankruptcy prediction: general framework and cross-validation analysis. European Journal of Operational Research 116, 16–32.

Zhang, G., Patuwo, B.E., Hu, M.Y., 1998. Forecasting with artificial neural networks: the state of the art. International Journal of Forecasting 14, 35–62.