

## Research Article

# Analysis of Changes in Market Shares of Commercial Banks Operating in Turkey Using Computational Intelligence Algorithms

M. Fatih Amasyali,<sup>1</sup> Ayse Demirhan,<sup>2</sup> and Mert Bal<sup>3</sup>

<sup>1</sup> Department of Computer Engineering, Yildiz Technical University, Davutpasa Campus, Esenler, 34220 Istanbul, Turkey

<sup>2</sup> Department of Business Administration, Yildiz Technical University, Yildiz Campus, Besiktas, 34349 Istanbul, Turkey

<sup>3</sup> Department of Mathematical Engineering, Yildiz Technical University, Davutpasa Campus A-116, Esenler, 34220 Istanbul, Turkey

Correspondence should be addressed to Mert Bal; mert.bal@gmail.com

Received 15 November 2013; Revised 24 February 2014; Accepted 13 March 2014; Published 15 April 2014

Academic Editor: Wolfgang Faber

Copyright © 2014 M. Fatih Amasyali et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This paper aims to model the change in market share of 30 domestic and foreign banks, which have been operating between the years 1990 and 2009 in Turkey by taking into consideration 20 financial ratios of those banks. Due to the fragile structure of the banking sector in Turkey, this study plays an important role for determining the changes in market share of banks and taking the necessary measures promptly. For this reason, computational intelligence methods have been used in the study. According to the research results, it is seen that it was not able to properly anticipate the data for the banking sector in the periods of financial crises (2000-2001 and 2008-2009). However, it is seen that, Simple Linear Regression is distinguished as a good algorithm among the computational intelligence algorithms for all periods between the years 1990 and 2009.

## 1. Introduction

As a natural result of the financial liberalization in the economy and the banking industry in Turkey after 1980s, the competition in the banking industry increased significantly due to the reasons such as many new domestic and foreign players in the banking industry, release of the fund transfers especially from international markets, enabling the banks to make transactions in foreign currencies, advances in the technology, and introducing new services by the banks in the industry. Therefore, a bank, operating in the banking industry, can differentiate itself from the other banks only if it can develop new strategies.

In recent years, because of economic and financial crisis, some of the public and private banks were bankrupted and some of them are merged and therefore they were forced to change how they operate. In this instance, a serious competition occurred among the surviving banks to take the market shares of the banks that have left the industry. The banks, which have evaluated the present circumstances, used

cutting edge technology, and improved the scope of their products and services, were able to advance forward significantly. Thus, these advances create a necessary environment for such banks to improve their market shares. Therefore, evaluating their position in the market and developing new strategies in accordance with their positions became much more important.

The presence of a tough competition between the banks besides the fragile structure of the banking sector in Turkey makes it important to determine the change in the market shares of banks and to take the necessary measures. For this reason, goal-oriented estimations that would be made by using computational intelligence methods are of great importance for the industry.

## 2. General Situation of Turkish Banking Sector

The banks during the period following 1980s are mostly multibranch banks which were financed by public sector

and operating in deposit banking field and have oligopolistic structure. In private banking point of view, the banks belonging to holdings have significance in the sector. In the banking sector between 1980 and 1990, with the help of the politics towards financial liberalization and resulting environment, the banks could adapt themselves to meet the international criteria [1]. During this period of financial liberalization process in Turkey, Capital Markets, Istanbul Stock Exchange (ISE), the Interbank Money Market (Interbank), and Foreign Exchange Market were established and Istanbul Gold Exchange started to operate [2]. In addition to this, Savings Deposit Insurance Fund (SDIF) is also established. SDIF is an institution in the banking sector which guarantees the rights of depositors and the regulating system. Gaining functionality of the free economy mechanism and rearrangements towards liberalization of the market have created significant impacts on the banking system [3]. The competition in the market increased due to new foreign/domestic players in the sector and liberating the deposit/credit interest rates [4].

As a result of the new legal and institutional regulations, helping new players to enter the market, compete, and expand in the sector, the number of the banks in Turkish banking sector has experienced a rapid expansion in employment, service diversification, and technological infrastructure points [5].

In 1999, a significant economic contraction in the economic activities and rising real interest rates caused a decline in private sector credit demand, a deterioration of the asset quality, and an increase in the need of liquidity of public banks, in particular banks with weak financial situation. In the late 1990s, macroeconomic instability in the banking sector, financial risks encountered, high costs of the resources, unfair competition conditions, lack of equity capital, oligopolistic structure, the rapid advancements in technology, holdings, and shallow capital market are some of the known basic problems [6, 7].

At the beginning of 2000, a comprehensive economic program has been implemented in order to reduce inflation and restore economic growth environment. In general, since the second half of the year, the conditions such as delays in the structural adjustment arrangements, inflation which did not fall as quickly as expected, and increasing costs of public goods and services were increased at the same level as the inflation; uncontrolled domestic demand caused deterioration of the economic outlook, and as a result of those reasons a serious concussion occurred in November 2000 and February 2001 in the banking sector.

Even though their arising points are quite different from each other, the nature of both crises which occurred in November 2000 and February 2001 is financial [8]. The starting circumstances of the crises affect the duration and depth of the crisis. The circumstances in question usually show the fragility of the economic structure in the country experiencing the crisis [9]. The basis of these two crises experienced in Turkey is due to the fragile nature of the banking sector [10]. November 2000 crisis was due to the liquidity problem in the banking sector. Increasing demand for liquidity during this period could not be met and this caused accelerated increase in the interest rates and exchange rates and the crisis in the financial market was deepened [11].

In addition, if we look at the crisis in February 2001, with the impact of rising financial fragility after November, it was a direct attack on Turkish Lira. This crisis that started in economic and financial system has quickly spread to the real sector.

November 2000 and February 2001 crisis intensified the problems happening in the banking sector further and introduced new problems. The banking sector, which was having serious difficulties due to liquidity and interest risk after the November crisis, faced significant losses arising from foreign exchange risk with the February crisis [12].

The significant increase in the interest rates, as a result of the crisis occurred in the November 2000, the financial structure of the public banks in the need of excessive overnight borrowing and the banks under the coverage of SDIF [13].

In order to solve the problems permanently and increase the competitiveness of the banking sector and after the financial crises in November 2000 and February 2001, a program has been established; a special emphasis was given to financial sector and the measures to be taken in this regard were determined.

The financial sector in Turkey was also affected negatively to a large extent by the global crises that affected all the world economies negatively in 2008. When it is considered from the point of deposit in investment and development banks, particularly in the last quarter of the year 2008 and in the beginning of the year 2009, the balance sheet risks increased rapidly, foreign funding opportunities became fewer, and the need for liquidity increased [14].

In line with the international developments, the income in the Turkish economy decreased, the private sector demand shrank, and the volume of foreign trade decreased whereas the budget deficit and unemployment rate increased and direct investments decreased. Therefore, in order to soften and restrict the negative impacts of the global crisis, many measures were taken by the related institutions in order to expand both the money and the financial politics during that period [15].

Computational intelligence algorithms that we have used in the paper will be described in Section 3.

### 3. Computational Intelligence Algorithms

In computational intelligence area, the aim is prediction of output values based on input values. If the output values are labels, this problem is named as a classification problem. If the output values are numeric values, it is named as a regression problem. Both of classification and regression problems construct a prediction model from the training set. The model performance is evaluated on the test set. For our problem the outputs are numeric. So, we have used the regression algorithms.

The regression approximation addresses the problem of estimating a function  $y = f(x)$  based on a given dataset  $G = \{(x_i, d_i)\}_{i=1}^N$ , where,  $x_i = [x_{i1}, x_{i2}, \dots, x_{im}]$  is input vector,  $d_i$  is the real value,  $y_i$  is the estimated value,  $f$  is the

estimation function,  $N$  is the number of observations, and  $m$  is the number of features.

In literature, several regression algorithms are proposed for building the estimation function. In this section, the used ones are briefly introduced.

**3.1. Zero Rule (ZeroR).** Zero Rule (ZeroR, 0-R) is a trivial classifier, but it gives a lower bound on the performance of a given dataset which should be significantly improved by more complex classifiers. As such it is a reasonable test on how well the class can be predicted without considering the other attributes [16].

**3.2. M5 Model Rules (M5R).** Generates a decision list for regression problems using separate-and-conquer. In each iteration it builds a model tree using M5 and makes the “best” leaf into a rule [17].

**3.3. Decision Table (DT).** A decision table consists of a hierarchical table in which each entry in a higher level table gets broken down by the values of a pair of additional attributes to form another table. The structure is similar to dimensional stacking. Given a training sample containing labeled instances, an induction algorithm builds a hypothesis in some representation. The representation investigated here is a decision table with a default rule mapping to the majority class, when it is abbreviated as DTM. A DTM has two components:

- (1) a *schema*, which is a set of features, and
- (2) a *body*, which is a multiset of labeled instances.

Each instance consists of a value for each of the features in the schema and a value for the label. Given an unlabelled instance  $\vec{x}$ , the label assigned to the instance by a DTM classifier is computed as follows.

Let  $I$  be the set of labeled instances in the DTM exactly matching the given instance  $\vec{x}$ , where only the features in the schema are required to match and all other features are ignored. If  $I = \emptyset$ , return the majority class in the DTM. Otherwise, return the majority class in  $I$ . Unknown values are treated as distinct values in the matching process [18].

**3.4. Reduced Error Pruning Tree (REP Tree).** Reduced Error Pruning (REP) was introduced by Quinlan [19] in the context of decision tree learning. It has been subsequently adapted to rule set learning as well [20]. It produces an optimal pruning of a given tree—the smallest tree among those with minimal error with respect to a given set of pruning examples [20, 21]. The REP algorithm works in two phases: first the set of pruning examples  $S$  is classified using the given tree  $T$  to be pruned. Counters that keep track of the number of examples of each class passing through each node are updated simultaneously. In the second phase—a bottom-up pruning phase—those parts of the tree that can be removed without increasing the error of the remaining hypothesis are pruned away [22]. The pruning decisions are based on the node statistics calculated in the top-down classification phase.

**3.5. Conjunctive Rule (CR).** This class implements a single conjunctive rule learner that can predict for numeric and nominal class labels.

A rule consists of antecedents “AND”ed together and the consequent (class value) for the classification/regression. In this case, the consequent is the distribution of the available classes (or mean for a numeric value) in the dataset. If the test instance is not covered by this rule, then it is predicted using the default class distributions/value of the data not covered by the rule in the training data. This learner selects an antecedent by computing the Information Gain of each antecedent and prunes the generated rule using Reduced Error Pruning or simple prepruning based on the number of antecedents. For classification, the information of one antecedent is the weighted average of the entropies of both the data covered and not covered by the rule. For regression, the information is the weighted average of the mean-squared errors of both the data covered and not covered by the rule. In pruning, weighted average of the accuracy rates on the pruning data is used for classification while the weighted average of the mean-squared errors on the pruning data is used for regression [23].

**3.6. Gaussian Process Regression (GPR).** Probabilistic regression is usually formulated as follows.

Given a training set  $D = \{(\vec{x}_i, y_i), i = 1, 2, \dots, n\}$  of  $n$  pairs of (vectorial) inputs  $\vec{x}_i$  and noisy (real, scalar) outputs  $y_i$ , compute the predictive distribution of the function values  $f_*$  (or noisy  $y_*$ ) at test locations  $\vec{x}_*$ . In the simplest case we assume that the noise is additive, independent, and Gaussian, such that the relationship between the (latent) function  $f(\vec{x})$  and the observed noisy targets  $y$  is given by

$$y_i = f(\vec{x}_i) + \varepsilon_i, \quad \text{where } \varepsilon_i \sim \mathcal{N}(0, \sigma_{\text{noise}}^2), \quad (1)$$

where  $\sigma_{\text{noise}}^2$  is the variance of the noise, and we use the notation  $\mathcal{N}(\vec{a}, A)$  for the Gaussian distribution with mean  $\vec{a}$  and covariance  $A$  [24].

The GP is a popular nonparametric model for supervised learning problems [25]. A Gaussian Process (GP) is a collection of random variables, any finite number of which have consistent joint Gaussian distributions. Gaussian Process (GP) regression is a Bayesian approach which assumes a GP prior over functions; that is, a priori the function values behave according to

$$p(\vec{f} | \vec{x}_1, \vec{x}_2, \dots, \vec{x}_n) = \mathcal{N}(\vec{0}, K), \quad (2)$$

where  $\vec{f} = [f_1, f_2, \dots, f_n]^T$  is a vector of latent function values  $f_i = f(\vec{x}_i)$  and  $K$  is a covariance matrix, whose entries are given by the covariance function  $K_{ij} = k(\vec{x}_i, \vec{x}_j)$ . Valid covariance functions give rise to semidefinite covariance matrices. In general, positive semidefinite kernels are valid covariance functions. The covariance function encodes our assumptions about the function we wish to learn by defining a notion of similarity between two function values as a function of the corresponding two inputs.

A very common covariance function is the Gaussian, or squared exponential:

$$K_{ij} = k(\vec{x}_i, \vec{x}_j) = \nu^2 \exp\left(-\frac{\|\vec{x}_i - \vec{x}_j\|^2}{2\lambda^2}\right), \quad (3)$$

where  $\nu^2$  controls the prior variance, and  $\lambda$  is an isotropic lengthscale parameter that controls the rate of decay of the covariance that determines how far away  $\vec{x}_i$  must be from  $\vec{x}_j$  for  $f_i$  to be unrelated to  $f_j$  [24].

**3.7. Isotonic Regression (IR).** The Isotonic Regression (IR) problem is considered in the following least distance setting. Given a vector  $y \in R^n$ , a strictly positive vector of weights  $w \in R^n$ , and a directed acyclic graph  $G(N, E)$  with the set of nodes  $N = \{1, 2, \dots, n\}$ , we find  $x^* \in R^n$  that solves the problem

$$\begin{aligned} \min \quad & \sum_{i=1}^n w_i (x_i - y_i)^2 \\ \text{s.t.} \quad & x_i \leq x_j \quad \forall (i, j) \in E. \end{aligned} \quad (4)$$

Since this is a strictly convex quadratic programming problem its solution  $x^*$  is unique. The optimality conditions and an analysis of the typical block (or cluster) structure of  $x^*$  can be found in [26–28]. The monotonicity constraints defined by the acyclic graph  $G(N, E)$  imply a partial order of the components  $x_i, i = 1, 2, \dots, n$ .

A special case of Isotonic Regression problem arises when there is a complete order of the components. This problem, referred to as IRC problem, is defined by a directed graph  $G(N, E)$  and is formulated as follows:

$$\begin{aligned} \min \quad & \sum_{i=1}^n w_i (x_i - y_i)^2 \\ \text{s.t.} \quad & x_1 \leq x_2 \leq \dots \leq x_n. \end{aligned} \quad (5)$$

IR problem has numerous important applications, for instance, in machine learning, data mining, statistics, operations research, and signal processing. These applications are characterized by a very large value of  $n$ . For such large-scale problems, it is of great practical importance to develop algorithms whose complexity does not rise with  $n$  too rapidly. The existing optimization-based algorithms [29] and statistical IR algorithms have either high computational complexity or low accuracy of the approximation to the optimal solution they generate.

The most widely used method for solving IRC problem (5) is the so-called Pool-Adjacent-Violator (PAV) algorithm [30–32]. This algorithm is of computational complexity  $O(n)$  [11–13]. The PAV algorithm has been extended by Pardalos and Xue [28] to the special case of IR problem (4), in which the partial order of the components is presented by a directed tree [33].

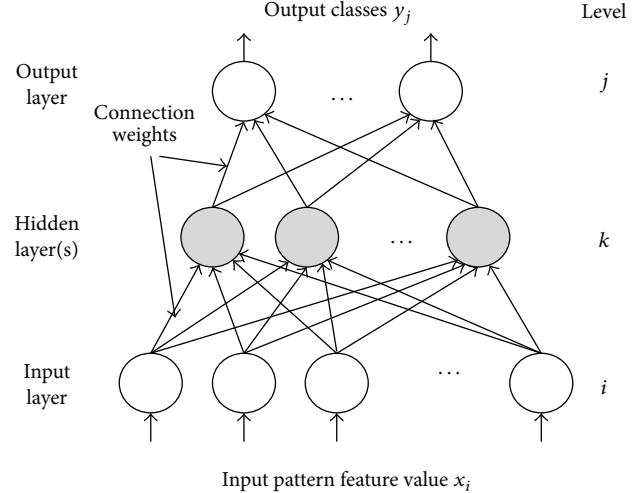


FIGURE 1: Structure of a Multilayer Perceptron [36].

**3.8. Linear Regression (LR).** The final model is a linear regression of a subsample of the attributes. The subsample is selected by iteratively removing the one with the smallest standardized coefficient until no improvement is observed in the estimate of the error given by the Akaike Information Criterion (AIC) [34]

$$\vec{y} = \vec{w}_0 + \vec{w}_i \cdot \vec{x}_i + \varepsilon. \quad (6)$$

The AIC value, with the assumption of eventual normally distributed errors, is calculated as follows:

$$\text{AIC} = 2k + N \ln\left(\frac{\text{RSS}}{N}\right), \quad (7)$$

where  $k$  is the number of parameters in the model, RSS is the residual of sum of the squares, and  $N$  is the number of observations [35].

**3.9. Multilayer Perceptron (MLP).** Multilayer Perceptron (MLP) also referred to as multilayer feedforward neural networks is the most used and popular neural network method. It belongs to the class of supervised neural network. The MLP topology consists of three sequential layers of processing nodes: an input layer, one or more hidden layers, and an output layer which produces the classification results. A MLP structure is shown in Figure 1 [36].

The principle of the network is that when data are presented at the input layer, the network nodes perform calculations in the successive layers until an output value is obtained at each of the output nodes. This output signal should be able to indicate the appropriate class for the input data. A node in MLP can be modeled as one or more artificial neurons, which computes the weighted sum of the inputs at the presence of the bias and passes this sum through

the nonlinear activation function. This process is defined as follows [36]:

$$\begin{aligned}\mu_j &= \sum_{i=1}^N w_{ji}x_i + \theta_j, \\ y_j &= \varphi_j(\mu_j),\end{aligned}\quad (8)$$

where  $\mu_j$  is the linear combination of inputs  $x_1, x_2, \dots, x_N$ ,  $\theta_j$  is the bias (adjustable parameter),  $w_{ji}$  is the connection synaptic weight between the input  $x_i$  and the neuron  $j$ ,  $\varphi(\cdot)$  is the activation function (usually nonlinear function) of the  $j$ th neuron, and  $y_j$  is the output. Here, hyperbolic tangent and logistic sigmoid function can be used for the nonlinear activation function. But, in most of the application widely used logistic sigmoid function is applied as follows:

$$\varphi(\lambda) = \frac{1}{1 + e^{-\lambda}}, \quad (9)$$

where  $\lambda$  represents the slope of the sigmoid [37].

The bias term  $\theta_j$  contributes to the left or right shift of the sigmoid activation function, depending on whether  $\theta_j$  takes a positive or negative value.

Learning in a MLP is an unconstrained optimization problem, which is subject to the minimization of a global error function depending on the synaptic weights of the network [38].

The first backpropagation learning algorithm for use with MLP structures was presented by Rumelhart et al. [39]. The backpropagation algorithm is one of the simplest and most general methods for the supervised training of MLP. This algorithm uses a gradient descent search method to minimize a mean square error between the desired output and the actual outputs. Backpropagation algorithm is defined as follows [36–40]:

- (I) initialize all the connection weights  $w$  with small random values from a pseudorandom sequence generator;
- (II) repeat until convergence (either when the error  $J$  is below a preset value or until the gradient  $\partial J / \partial w$  is smaller than a preset value);
  - (i) compute the update using  $\Delta w(m) = -\xi (\partial J(m) / \partial w)$ ,
  - (ii) iterative algorithm requires taking a weight vector at iteration  $m$  and updating it as  $(m+1) = w(m) + \Delta w(m)$ ,
  - (iii) compute the error  $J(m+1)$ ,

where  $m$  is the iteration number,  $w$  represents all the weights in the network, and  $\xi$  is the learning rate and merely indicates the relative size of the change in weights. The error  $J$  can be chosen as the mean square error function between the actual output  $y_j$  and the desired output  $d_j$  and  $d$  and  $y$  are the desired and the network output vector of length  $N$ :

$$J(w) = \frac{1}{2} \sum_{j=1}^N (d_j - y_j)^2 = \frac{1}{2} (d - y)^2. \quad (10)$$

**3.10. Simple Linear Regression (SLR).** It is the process of fitting straight lines (models) between each attribute and output. In (11), the values of  $\vec{w}$  and  $\vec{w}_0$  are estimated by the method of least squares. Consider

$$\vec{y} = \vec{w} \cdot \vec{x} + \vec{w}_0. \quad (11)$$

The model having lowest squared error is selected as the final model among each parameter model [35].

**3.11. Support Vector Machines (SVMs).** The Support Vector Machines (SVMs) [41] are types of learning machines based on statistical learning theory. SVMs are supervised learning methods that have been widely and successfully used for pattern recognition in different areas [42]. Especially in recent years SVMs with linear or nonlinear kernels have become one of the most promising learning algorithms for classification as well as regression [43]. The problem that SVMs try to solve is to find an optimal hyperplane that correctly classifies data points by separating the points of two classes as much as possible [44].

Let  $x_i$  (for  $1 \leq i \leq N_x$ ) be the input vectors in input space, with corresponding binary labels  $y_i \in \{-1, 1\}$ . Let  $\vec{X}_i = \Phi(x_i)$  be the corresponding vectors in feature space, where  $\Phi(x_i)$  is the implicit kernel mapping, and let  $K(x_i, x_j) = \Phi(x_i) \cdot \Phi(x_j)$ .  $\Phi(x_j)$  be the kernel function, implying a dot product in the feature space [45].  $K(x, y)$  represents the desired notion of similarity between data  $x$  and  $y$ .  $K(x, y)$  needs to satisfy Mercer's condition in order for  $\Phi$  to exist [44]. There are a number of kernel functions which have been found to provide good generalization capabilities [46]. The kernel function that has been used in SVM is a linear function and the details of the function are given below

$$\text{Linear Kernel: } K(x_i, x_j) = x_i^T x_j.$$

The optimization problem for a soft-margin SVM is

$$\min_{\vec{w}, b} \left\{ \frac{1}{2} \|\vec{w}\|^2 + C \sum_i \xi_i \right\}, \quad (12)$$

subject to the constraints  $y_i(\vec{w} \cdot \vec{x} + b) = 1 - \xi_i$  and  $\xi_i \geq 0$ , where  $\vec{w}$  is the normal vector of the separating hyperplane in feature space, and  $C > 0$  is a regularization parameter controlling the penalty for misclassification. Equation (12) is referred to as the primal equation. From the Lagrangian form of (12), we derive the dual problem:

$$\max_{\alpha} \left\{ \sum_i \alpha_i - \frac{1}{2} \sum_{i,j} \alpha_i \alpha_j y_i y_j K(x_i, x_j) \right\} \quad (13)$$

subject to  $0 \leq \alpha_i \leq C$ . This is a quadratic optimization problem that can be solved efficiently using algorithms such as Sequential Minimal Optimization (SMO) [47]. Typically, many  $\alpha_i$  go to zero during optimization, and the remaining  $x_i$  corresponding to those  $\alpha_i > 0$  are called support vectors. To simplify notation, from here on we assume that all nonsupport vectors have been removed, so that  $N_x$  is now the number of support vectors, and  $\alpha_i > 0$  for all  $i$ . With this

formulation, the normal vector of the separating plane  $\vec{w}$  is calculated as

$$\vec{w} = \sum_{i=1}^{N_x} \alpha_i y_i \vec{x}_i. \quad (14)$$

Note that because  $\vec{X}_i = \Phi(x_i)$  is defined implicitly,  $\vec{w}$  exists only in feature space and cannot be computed directly. Instead, the classification  $f(\vec{q})$  of a new query vector  $\vec{q}$  can only be determined by computing the kernel function of  $\vec{q}$  with every support vector

$$f(\vec{q}) = \text{sign} \left( \sum_{i=1}^{N_x} \alpha_i y_i K(\vec{q}, \vec{x}_i) + b \right), \quad (15)$$

where the bias term  $b$  is the offset of the hyperplane along its normal vector, determined during SVM training [45].

**3.12. IB1 (One Nearest Neighbor).** IB1 [33] is an implementation of the simplest similarity based learner, known as nearest neighbor. IB1 simply finds the stored instance closest (according to Euclidean distance metric) to the instance to be classified. The new instance is assigned to the retrieved instance's class. Equation (16) shows the distance metric employed by IB1. Consider

$$D(x, y) = \sqrt{\sum_{i=1}^n f(x_i, y_i)}. \quad (16)$$

Equation (16) gives the distance between two instances  $x$  and  $y$ ;  $x_i$  and  $y_i$  refer to the  $i$ th feature value of instances  $x$  and  $y$ , respectively. For numeric valued attributes  $f(x_i, y_i) = (x_i - y_i)^2$ , for symbolic valued attributes  $f(x, y) = 0$ , if the feature values  $x_i$  and  $y_i$  are the same, and 1 if they differ [48].

**3.13. K-Star.** K-star [49] can be considered as a variation of instance based learning which uses an entropic distance measure. To compute the distance between two samples, the concept of "complexity of transforming from one sample into another sample" is introduced. A K-star distance is then defined by summing over all possible transformation paths between two distances. This approach can be applied on real valued attributes, as well as symbolic attributes [50].

**3.14. Additive Regression (AR).** It is initially focused on the regression problem where the response  $y$  is quantitative, and it is aimed to model the mean  $E(y | x) = F(x)$ . The additive model has the following form:

$$F(x) = \sum_{i=1}^N f_i(x_i). \quad (17)$$

Here there is a separate function  $f_i(x_i)$  for each of the  $N$  input variables  $x_i$ . More generally, each component  $f_i$  is a function of a small prespecified subset of the input variables. The

"backfitting algorithm" is a convenient modular algorithm for fitting additive models. A backfitting update is

$$f_i(x_i) \leftarrow E \left[ y - \sum_{k \neq i} f_k(x_k) | x_i \right] \quad (18)$$

Any method or algorithm for estimating a function of  $x_i$  can be used to obtain an estimate of the conditional expectation in (18). In particular, this can include nonparametric smoothing algorithms, such as local regression or smoothing splines. In the right hand side, all the latest versions of the functions  $f_i$  are used in forming the partial residuals. The backfitting cycles are repeated until convergence [51]. Under fairly general conditions, backfitting can be shown to converge to the minimizer of  $E(y - F(x))^2$  [52].

**3.15. Bagging.** Bagging [53] predictors is a method for generating multiple versions of a predictor and using these to get an aggregated predictor. The aggregation averages over the versions when predicting a numerical outcome and does a plurality vote when predicting a class. The multiple versions are formed by making bootstrap replicates of the learning set and using these as new learning sets. Tests on real and simulated data sets using classification and regression trees and subset selection in linear regression show that bagging can give substantial gains in accuracy. The vital element is the instability of the prediction method. If perturbing the learning set can cause significant changes in the predictor constructed, then bagging can improve accuracy [53].

## 4. Purpose of Study

The study aims to determine which parameters (factors) affect the change in the market shares of the public and private commercial banks and the foreign banks established in Turkey and have branches operating in Turkey between 1990 and 2009, in parallel with the economic and social developments that occurred during this period in the country.

Three market shares, which can be considered as the indicator of all market shares, are taken into account in the study. These figures ( $Y_1$ ,  $Y_2$ , and  $Y_3$ ) are the market shares as an indicator of the market share in total deposits ( $Y_3$ ), as an indicator of the success of the bank in deposit collection, the market share in total credits ( $Y_3$ ), as an indicator of the contribution to the total credits, and finally the market share in the total assets ( $Y_1$ ) as an indicator of the change in the reliability and soundness of the bank.

Because of the long-term effects of the Russian and Asian crises encountered in 1997 and 1998, November 2000 and February 2001 crises were experienced successively with deeper financial strength in Turkey. These crises affected mostly the banks which do not have solid financial structures and cannot act according to the conditions experienced. Therefore, 2001 crisis with regard to the banking sector should not only be considered economically. After 2001, the banking sector has been renewed with new regulations. Therefore, another purpose of the study is to determine if the crises affect the market shares of the banks.

TABLE 1: The data of the banks for the period of 1990–2009.

Number	Bank name
1	Z. Bankası
2	H. Bankası
3	Vakıfbank
4	İş Bankası
5	Akbank
6	Alternatifbank
7	Şekerbank
8	G. Bankası
9	Citibank
10	Credit Agricole Indosuez
11	T.E. Bankası
12	Finansbank
13	Anadolu
14	Adabank
15	Oyakbank (ING Bank)
16	Arap-Türk
17	Tekstilbank
18	Turkish Bank
19	Türk Dış Ticaret (FORTIS)
20	H. Bank
21	Bank M.
22	HSBC
23	Banca Di Roma
24	ABN-Amro
25	West LB AG
26	Y. K. Bankası
27	Societe Generale (S.A)
28	Bayindirbank
29	Bank Europa (Sitebank)
30	Pamukbank (BFB)

Resource: The Banks Association of Turkey, Banks in Turkey 1990–2009  
<http://www.tbb.org.tr/>.

## 5. Dataset and Parameters Used in Study

*5.1. Scope of Study.* The public and private commercial banks and the foreign banks established in Turkey between 1990 and 2009 are included in the study. The bankrupted banks, the banks that were transferred to SDIF and the way of operation has been changed and development and investment banks are excluded from the study.

For the period 1990–2009, the data belonging to 30 banks listed in Table 1 are used in the study.

The market shares of the banks can be determined using three different criteria: market shares in total assets, total credits, and total deposits [54]. Total assets, total credits, and total deposits factors, which can be used as indicators of the change in the market share, shall be explained using the following ratios: which is important for banks to continue their operations; capital adequacy ratios, profitability ratios which is important to measure the profitability of the banks, liquidity ratios which show the status of short-term debt

repayment, income-expenditure structure rates, which give information about which items are playing important role in total income and expenditure of the banks and correction of the profit of the bank changing which income and expenditure items, and asset quality and the ratios showing the group shares of the banks. All of those figures show the situation of the banks as well as giving illustrative information about if the banks will face a problem in the future. Therefore, following those ratios will help to take necessary precautions in case of negative developments. The precautions that will be taken immediately will not only prevent the decrease in the market shares but also will help the banks to catch the opportunities that cause their market shares to increase. The ratios showing the market shares of the banks used in the study are dependent variables and capital adequacy, asset quality, liquidity, profitability, and the ratios showing the group shares are explanatory variables. An average value is calculated as a dependent variable that indicates the market shares of banks by using total assets ( $Y_1$ ), total credit ( $Y_2$ ), and total deposit ( $Y_3$ ) rates of all the banks in the sector. In determination of the market shares of the banks in the study during the period 1990 and 2009, analysis will be developed using computational intelligence algorithms and 20 financial ratios relating to capital adequacy, asset quality, liquidity, profitability, and the ratios showing the group shares.

The data used in the study consists of secondary data which were taken from the book “Our Banks” and prepared annually by Banks Association of Turkey. 20 financial ratios prevailing for the period between 1990 and 2009 are shown in Table 2.

The liquidity ratios are defining the power of repayment of the short-term debt of businesses, and the profitability ratios are the indicators for the businesses’ ability to provide a satisfactory profit for the invested capital and the effectiveness of management in well-functioning financial markets. To those areas where the profitability is limited, financial investors do not put their money and the entrepreneurs do not make investments. In addition, the credit institutions are reluctant to give credits to the businesses which have low profitability. The ratios relating to capital adequacy, asset quality, liquidity, profitability, and the ratios showing the group shares are important to the businesses and therefore they are crucial to banks. With the calculation of those ratios, the banks can determine both their situation and their position in the sector more efficiently. Therefore, all of the ratios above are included as explanatory variables.

## 6. Experimental Results

The bank dataset collection includes 20 features/inputs (see Table 2) and an output feature (changes in market shares) for each year (1990–2009). For each year, a dataset is formed including input features and an output feature. Each dataset consists of 30 samples. All features were normalized between 0 and 1.

We have applied 15 machine learning algorithms from WEKA library [16] to predict the output of samples. In Table 3 the algorithms and IDs are given.

TABLE 2: 20 financial ratios that are used between the period of 1990 and 2009.

Code	Ratios
	<i>Capital adequacy ratios (CA)</i>
CA1	(Shareholders' equity + profit)/total assets
CA2	(Shareholders' equity + profit)/(deposit + nondeposit sources)
CA3	(Net working capital)/total assets
CA4	(Shareholders' equity + profit)/(total assets + noncash loans)
	<i>Assets quality ratios (AQ)</i>
AQ1	Total loans/total assets
AQ2	Loans under follow-up/total loans
AQ3	Permanent assets/total assets
	<i>Liquidity ratios (L)</i>
L1	Liquid assets/total assets
L2	Liquid assets/(deposit + nondeposit sources)
L3	TC liquid assets/total assets
	<i>Profitability ratios (P)</i>
P1	Net profit (losses)/total assets (ROA)
P2	Net profit (losses)/total shareholders' equity (ROE)
P3	Profit (losses) before taxes after continuing operations/total assets
	<i>Income-expenditure structure ratios (IE)</i>
IE1	Net interest income after receivable under follow-up/total assets
IE2	Interest income (net)/interest expenses
IE3	Noninterest income (net)/noninterest expenses
IE4	Personnel expenses/other operating income
	<i>Group shares ratios (GS)</i>
GS1	Total assets
GS2	Total loans
GS3	Total deposits
	<i>Market shares ratio (MS)</i>
MS	MS = (Y1 + Y2 + Y3)/3 Y1: total assets; Y2: total loans; Y3: total deposits

Resource: The Banks Association of Turkey, Banks in Turkey 1990–2009  
<http://www.tbb.org.tr/>

The default design parameters were selected for Zero Rule, M5 rules, decision table, conjunctive rule, REP Tree, Gaussian Processes, Isotonic Regression, linear regression, Multilayer Perceptron, Simple Linear Regression, support vector machine regression, one nearest neighbor, and Kstar algorithms. For the meta algorithms (additive regression and bagging), the ensemble sizes were selected as 100.

The performance of each classification algorithm was evaluated using 5-fold cross-validation. In each 5-fold cross-validation, each dataset is randomly split into 5 equal sized segments and results are averaged over 5 trials. The performance evaluation is based on root mean-squared error (RMSE) defined in (19). Consider

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^N (d_i - y_i)^2}, \quad (19)$$

TABLE 3: The used algorithms and names.

Algorithm ID	Algorithm name
(1)	ZeroR
(2)	M5 rules
(3)	Decision table
(4)	Conjunctive rule
(5)	REP Tree
(6)	Gaussian Processes
(7)	Isotonic Regression
(8)	Linear regression
(9)	Multilayer Perceptron
(10)	Simple Linear Regression
(11)	SMOreg
(12)	IBk (one nearest neighbor)
(13)	Kstar
(14)	Additive regression
(15)	Bagging

where  $y^i$  is the estimated output for  $i$ th test sample,  $d^i$  is the real output value of  $i$ th test sample, and  $N$  is the number of the test samples. Table 4 shows the root mean-squared error results.

According to Table 4, the best algorithm (having the most minimum errors) is Simple Linear Regression. This very simple algorithm generally performs better than all the other complicated algorithms.

According to Table 4, the best predicted years are 2006 and 2007. The worst predicted years are 2001, 2002, 2008, and 2009.

Due to the financial crisis experienced in the beginning of 2000s in Turkey and the global crisis that affected our country as well as all the world economies in the period 2008–2009, the estimations for those periods (2001, 2002, 2008, and 2009) did not lead to good results. It is seen that this situation is based on the fact that economic and financial data are much more fluctuating in the crisis periods in comparison to stagnation periods.

Two questions were raised when we want to have more understanding on the dynamics of the datasets.

- (1) The datasets includes 20 features. It is known that the irrelevant features can badly affect results. Is it possible to get better results with fewer features?
- (2) Which features affected the results the most?

To answer these related questions we applied Correlation Based Feature Selection (CFS) [50] algorithm for each year's dataset separately. CFS method will be explained briefly below.

CFS evaluates the worth of a subset of attributes by considering the individual predictive ability of each future along with the degree of redundancy between them. Correlation coefficients are used to estimate correlation between subset of attributes and class, as well as intercorrelations between the features. Relevance of a group of features grows with the correlation between features and classes and decreases

TABLE 4: The RMSEs of the compared algorithms over 20 bank datasets.

Year/dataset	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	Min. RMSE	Best algorithm ID
1990	0.19	0.07	0.16	0.16	0.18	0.14	0.1	5.15	0.16	0.05	0.56	0.11	0.11	0.19	0.15	0.05	10
1991	0.2	0.09	0.14	0.18	0.18	0.15	0.11	0.47	0.13	0.06	0.15	0.13	0.14	0.17	0.15	0.06	10
1992	0.18	0.12	0.16	0.16	0.17	0.15	0.11	0.13	0.12	0.07	0.07	0.13	0.12	0.16	0.15	0.07	10
1993	0.19	0.12	0.14	0.18	0.19	0.16	0.12	0.16	0.16	0.07	0.08	0.14	0.14	0.16	0.15	0.07	10
1994	0.2	0.09	0.15	0.17	0.17	0.15	0.1	0.21	0.17	0.06	0.13	0.12	0.12	0.13	0.14	0.06	10
1995	0.19	0.08	0.17	0.18	0.18	0.16	0.11	0.14	0.19	0.07	0.09	0.15	0.17	0.18	0.15	0.07	10
1996	0.18	0.07	0.13	0.16	0.17	0.15	0.11	0.21	0.14	0.07	0.16	0.13	0.12	0.13	0.14	0.07	10
1997	0.2	0.13	0.17	0.17	0.18	0.16	0.13	0.29	0.16	0.08	0.18	0.12	0.13	0.18	0.15	0.08	10
1998	0.22	0.12	0.17	0.17	0.17	0.17	0.1	0.2	0.18	0.09	0.13	0.15	0.12	0.13	0.13	0.09	10
1999	0.22	0.11	0.2	0.19	0.18	0.18	0.14	0.35	0.15	0.09	0.17	0.18	0.13	0.19	0.14	0.09	10
2000	0.23	0.11	0.14	0.19	0.17	0.19	0.09	0.2	0.13	0.09	0.13	0.19	0.16	0.15	0.14	0.09	7
2001	0.27	0.24	0.17	0.24	0.24	0.23	0.21	0.36	0.33	0.26	0.34	0.23	0.17	0.27	0.21	0.17	13
2002	0.28	0.29	0.26	0.23	0.23	0.23	0.21	1.01	0.41	0.24	0.52	0.29	0.23	0.29	0.21	0.21	15
2003	0.27	0.12	0.15	0.18	0.16	0.15	0.1	0.22	0.11	0.11	0.06	0.09	0.18	0.16	0.12	0.06	11
2004	0.26	0.07	0.14	0.13	0.11	0.15	0.1	0.19	0.07	0.08	0.1	0.08	0.18	0.13	0.1	0.07	9
2005	0.3	0.16	0.21	0.16	0.17	0.16	0.11	0.32	0.19	0.1	0.09	0.09	0.17	0.15	0.13	0.09	11
2006	0.31	0.06	0.19	0.13	0.12	0.14	0.07	0.03	0.03	0.04	0.02	0.06	0.15	0.12	0.1	0.02	11
2007	0.33	0.05	0.19	0.16	0.13	0.17	0.05	0.02	0.04	0.06	0.02	0.06	0.14	0.14	0.11	0.02	11
2008	0.37	0.48	0.38	0.41	0.38	0.36	0.4	0.7	0.56	0.43	0.54	0.54	0.47	0.4	0.36	0.36	15
2009	0.34	0.39	0.41	0.37	0.36	0.35	0.46	0.39	0.63	0.37	0.48	0.48	0.44	0.54	0.36	0.34	1

TABLE 5: The selected features according to the years.

Selected features			
1990	P3	IE3	GS3
1991	L1	P3	GS3
1992	L2	GS2	GS3
1993	L2	P3	IE2 GS1 GS3
1994	AQ2	L1	GS1 GS2 GS3
1995	AQ1	AQ2	IE3 GS1
1996	CA3	GS1	GS2 GS3
1997	L1	P3	GS1 GS3
1998	L2	IE1	GS3
1999	L1	IE2	GS1 GS2 GS3
2000	L1	GS3	
2001	AQ2	L3	GS1
2002	CA2	L1	IE2 GS1
2003	L1	P2	GS1 GS2 GS3
2004	P2	GS1	GS2 GS3
2005	IE3	GS1	GS2 GS3
2006	GS1	GS2	GS3
2007	CA3	P2	GS1 GS2 GS3
2008	L3	IE3	
2009	CA3	L2	P2

TABLE 6: The most selected features and selection frequencies.

Feature name	Selection frequency
GS3	15
GS1	13
GS2	9
L1	7
AQ2	4
IE3	4
P2	4
P3	4
L2	4
IE2	3
CA3	3
L3	2
IE1	1
CA2	1

bidirectional search, best-first search, and genetic search. Equation formalizes the heuristic:

$$r_{zc} = \frac{t\bar{r}_{zi}}{\sqrt{t + t(t-1)\bar{r}_{ii}}}, \quad (20)$$

with growing intercorrelation [50]. CFS is used to determine the best feature subset and is usually combined with search strategies such as forward selection, backward elimination,

where  $r_{zc}$  is the correlation between the summed feature subsets and the class variable,  $t$  is the number of subset features,  $\bar{r}_{zi}$  is the average of the correlations between the

TABLE 7: The RMSEs of the compared algorithms over 20 dimensionally reduced bank datasets.

Year/data set	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	Min. RMSE	Best algorithm ID
1990	0.19	0.05	0.12	0.15	0.16	0.13	0.09	0.05	0.09	0.05	0.05	0.09	0.12	0.13	0.14	0.05	10
1991	0.2	0.05	0.12	0.15	0.17	0.13	0.1	0.05	0.09	0.05	0.06	0.12	0.15	0.13	0.13	0.05	10
1992	0.18	0.07	0.13	0.15	0.17	0.13	0.11	0.06	0.1	0.07	0.05	0.1	0.1	0.13	0.14	0.05	11
1993	0.19	0.08	0.13	0.18	0.18	0.14	0.12	0.08	0.11	0.07	0.06	0.12	0.08	0.13	0.14	0.06	11
1994	0.2	0.06	0.13	0.17	0.17	0.13	0.1	0.06	0.11	0.06	0.06	0.11	0.11	0.12	0.13	0.06	10
1995	0.19	0.07	0.16	0.17	0.18	0.14	0.11	0.07	0.12	0.06	0.08	0.15	0.16	0.13	0.14	0.06	10
1996	0.18	0.07	0.11	0.16	0.16	0.13	0.11	0.07	0.12	0.07	0.07	0.1	0.1	0.11	0.13	0.07	10
1997	0.2	0.1	0.14	0.16	0.18	0.14	0.13	0.09	0.12	0.08	0.08	0.1	0.09	0.14	0.15	0.08	10
1998	0.22	0.1	0.14	0.16	0.16	0.15	0.1	0.1	0.15	0.09	0.09	0.11	0.11	0.11	0.12	0.09	10
1999	0.22	0.13	0.17	0.19	0.18	0.14	0.14	0.09	0.13	0.09	0.08	0.09	0.1	0.16	0.14	0.08	11
2000	0.23	0.08	0.1	0.14	0.13	0.13	0.09	0.08	0.1	0.09	0.09	0.08	0.09	0.09	0.1	0.08	8
2001	0.27	0.23	0.14	0.22	0.22	0.21	0.2	0.2	0.26	0.23	0.2	0.25	0.2	0.26	0.19	0.14	3
2002	0.28	0.24	0.25	0.22	0.23	0.23	0.22	0.24	0.38	0.23	0.23	0.3	0.27	0.24	0.2	0.2	15
2003	0.27	0.06	0.14	0.17	0.15	0.11	0.1	0.07	0.06	0.11	0.04	0.08	0.1	0.12	0.11	0.04	11
2004	0.26	0.05	0.13	0.11	0.1	0.11	0.1	0.05	0.06	0.08	0.03	0.06	0.09	0.09	0.08	0.03	11
2005	0.3	0.06	0.2	0.16	0.14	0.11	0.11	0.07	0.07	0.1	0.04	0.06	0.08	0.11	0.11	0.04	11
2006	0.31	0.03	0.19	0.12	0.11	0.08	0.07	0.01	0.02	0.04	0.01	0.05	0.05	0.08	0.09	0.01	8
2007	0.33	0.04	0.19	0.15	0.11	0.11	0.05	0.02	0.02	0.06	0.01	0.04	0.08	0.1	0.09	0.01	11
2008	0.37	0.39	0.33	0.36	0.38	0.36	0.33	0.38	0.44	0.37	0.42	0.45	0.36	0.42	0.35	0.33	3
2009	0.34	0.37	0.37	0.35	0.35	0.34	0.4	0.37	0.44	0.35	0.37	0.42	0.4	0.52	0.36	0.34	1

subset features and the class variable, and  $\bar{r}_{ii}$  is the average intercorrelation between subset features [48–55]. Equation (20) is, in fact, Pearson's correlation, where all variables have been standardized. In Table 5, the selected features are given.

In Table 6, the selection frequencies of the features are given.

According to Table 6 the most selected features are among GS groups. In our study, domestic and foreign banks other than development and investment banks operating in Turkey are examined. According to our study, the share of each single bank within its own group has an influence on its share in the whole banking sector to a certain extent. As it may be derived from the analysis results, the variable representing the group shares had mostly a priority and was effective for the forecasts made with regard to the market shares of banks. Therefore, it is seen that such situation may have a positive impact on the banks, which are able to come to the fore in terms of technological developments, wide product range, and competitive advantage and increase their shares within their own groups, towards increasing their shares throughout the whole industry banks.

With the dimensionally reduced datasets, the same experiments are repeated with the same 15 algorithms in Table 7. The results are given at Table 7.

According to Table 7, the best algorithms (having the most minimum errors) are Simple Linear Regression and Support Vector Regression.

The best and the worst predicted years are the same with all featured datasets.

To determine the effect of feature selection Figure 2 was drawn. In Figure 2,  $x$ -axis shows the years and  $y$ -axis shows the minimum error of the 15 algorithms. The solid line shows the minimum errors of all featured datasets. The dotted line shows the minimum errors of dimensionally reduced datasets.

According to Figure 2, the feature selection process has positive effects over almost all years. Because of the fact that the obtained errors with dimensionally reduced datasets are lower than all featured datasets, the generated rules by the best algorithms over dimensionally reduced datasets are given at Table 8.

When Tables 8 and 5 are investigated together, it can be seen that the selected features by the feature selection algorithm and the used features in the rules are very similar.

The performance differences among algorithms were also investigated with  $t$ -test.  $t$ -test says whether there is a significant difference between two distributions or not. In our experiments, each algorithm produces 5 RMSEs for each dataset, because we used 5-fold cross-validation. The distributions which are compared are these errors of algorithms. In Table 9, the number of datasets where there is significant difference between the algorithms at the rows and columns is given. The most successful 4 algorithms (decision table, linear

TABLE 8: The rules generated by the best algorithm for each year (dataset).

Year	Algorithm ID	Rule
1990	10	$MS = 0.97 * GS3 - 0.02$
1991	10	$MS = 0.99 * GS3 - 0.02$
1992	11	$MS = -0.0292 * (\text{normalized } L2) + 0.3947 * (\text{normalized } GS2) + 0.5996 * (\text{normalized } GS3) + 0.0068$
1993	11	$MS = -0.1056 * (\text{normalized } L2) + 0.0094 * (\text{normalized } P3) + 0.0052 * (\text{normalized } IE2) + 0.4708 * (\text{normalized } GS1) + 0.523 * (\text{normalized } GS3) + 0.0044$
1994	10	$MS = 1.04 * GS1 - 0.02$
1995	10	$MS = 0.99 * GS1 - 0.02$
1996	10	$MS = 1 * GS1 - 0.01$
1997	10	$MS = 0.97 * GS3 + 0$
1998	10	$MS = 1.06 * GS3 + 0.01$
1999	11	$MS = -0.0237 * (\text{normalized } L1) - 0.0225 * (\text{normalized } IE2) + 0.296 * (\text{normalized } GS1) + 0.3055 * (\text{normalized } GS2) + 0.3722 * (\text{normalized } GS3) + 0.0272$
2000	8	$MS = -0.1229 * L1 + 0.9571 * GS3 + 0.0769$
2001	3	If $GS1 \in (-\infty, -0.1]$ then $MS = 0.03137$ If $GS1 \in (0.3, -0.4]$ then $MS = 0.59189$ If $GS1 \in (0.4, -0.5]$ then $MS = 0.041797$ If $GS1 \in (0.5, -0.6]$ then $MS = 0.441048$ If $GS1 \in (0.8, -0.9]$ then $MS = 0.093575$ If $GS1 \in (0.9, -\infty)$ then $MS = 1.0$
2002	15, 11	The best performed algorithm is bagging. But the rules generated by bagging are very long. Instead of giving these long rules, a rule generated by the second best algorithm (SMOreg) is given. $MS = -0.2014 * (\text{normalized } CA2) - 0.0889 * (\text{normalized } L1) - 0.0859 * (\text{normalized } IE2) + 0.7895 * (\text{normalized } GS1) + 0.1961$
2003	11	$MS = -0.0024 * (\text{normalized } L1) + 0.0009 * (\text{normalized } P2) + 0.3401 * (\text{normalized } GS1) + 0.2689 * (\text{normalized } GS2) + 0.4899 * (\text{normalized } GS3) - 0.0007$
2004	11	$MS = -0.0015 * (\text{normalized } P2) + 0.342 * (\text{normalized } GS1) + 0.2547 * (\text{normalized } GS2) + 0.4802 * (\text{normalized } GS3) - 0.0001$

TABLE 8: Continued.

Year	Algorithm ID	Rule
2005	11	$MS = -0.0028 * (\text{normalized } IE3) + 0.3304 * (\text{normalized } GS1) + 0.3268 * (\text{normalized } GS2) + 0.4712 * (\text{normalized } GS3) + 0.0002$
2006	8	$MS = 0.1241 * GS1 + 0.3862 * GS2 + 0.6142 * GS3 + -0.0026$
2007	11	$MS = 0.0049 * (\text{normalized } CA3) + 0.0012 * (\text{normalized } P2) + 0.3388 * (\text{normalized } GS1) + 0.3281 * (\text{normalized } GS2) + 0.471 * (\text{normalized } GS3) - 0.0056$
2008	3	If $L3 \in (-\infty, -0.1]$ then $MS = 0.1942307$ If $L3 \in (0.1, -0.2]$ then $MS = 0.479808$ If $L3 \in (0.2, -0.3]$ then $MS = 0.073077$ If $L3 \in (0.3, -0.4]$ then $MS = 0.019231$ If $L3 \in (0.4, -0.5]$ then $MS = 0.980769$ If $L3 \in (0.5, -0.6]$ then $MS = 0.2230765$ If $L3 \in (0.6, -0.7]$ then $MS = 0.0826925$ If $L3 \in (0.7, -0.8]$ then $MS = 0.042308$ If $L3 \in (0.9, -\infty)$ then $MS = 0.242308$
2009	1	$MS = 0.403535$

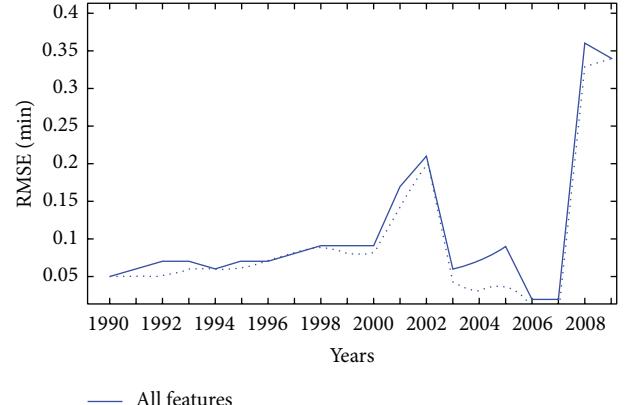


FIGURE 2: The effect of feature selection.

regression, Simple Linear Regression, and Support Vector Regression) were compared with all the other algorithms.

In Table 9, the values in the cells has the  $(X/Y/Z)$  format.  $X$  shows that the number of datasets of the algorithm at the column has statistically significant lower error than the algorithm at the row.  $Z$  shows that the number of datasets of the algorithm at the column has statistically significant higher error than the algorithm at the row.  $Y$  shows the number of datasets the algorithm at the column has no statistically significant difference error than the algorithm at the row.

According to Table 9, decision table algorithm has no better performance than the other 3 compared algorithms (see third row). There is not any significant difference between

TABLE 9: Comparison of the best 4 algorithms with all the other algorithms by  $t$ -test.

Algorithm ID	3—Decision table	8—Linear regression	10—Simple linear regression	11—Support vector regression
1	(12/8/0)	(16/4/0)	(15/5/0)	(16/4/0)
2	(0/17/3)	(0/20/0)	(0/20/0)	(1/19/0)
3		(4/16/0)	(4/16/0)	(5/15/0)
4	(1/19/0)	(10/10/0)	(6/14/0)	(11/9/0)
5	(0/20/0)	(11/9/0)	(9/11/0)	(14/6/0)
6	(0/18/2)	(2/18/0)	(1/19/0)	(2/18/0)
7	(0/18/2)	(2/18/0)	(1/19/0)	(4/16/0)
8	(0/16/4)	—	(0/20/0)	(0/20/0)
9	(0/17/3)	(0/20/0)	(0/20/0)	(2/18/0)
10	<b>(0/16/4)</b>	<b>(0/20/0)</b>	—	<b>(2/17/1)</b>
11	(0/15/5)	(0/20/0)	(1/17/2)	—
12	(1/15/4)	(3/17/0)	(2/18/0)	(4/16/0)
13	(0/16/4)	(2/18/0)	(0/20/0)	(2/18/0)
14	(2/18/0)	(4/16/0)	(3/17/0)	(8/12/0)
15	(0/20/0)	(2/18/0)	(0/20/0)	(4/16/0)

Simple Linear Regression and linear regression. Support Vector Regression has 2 better and one lower performance than linear regression.

In other words, there are very small differences between the best performed 4 algorithms over 20 datasets. Simple Linear Regression generates very simple and understandable rules. It has very good performance among other 14 algorithms. Because of its simplicity, the prediction process time is very fast. Because of these reasons we can say that Simple Linear Regression is the best solution to predict market shares values in bank datasets.

## 7. Conclusion

The study uses the data of 30 domestic and foreign banks other than development and investment banks, which have been operating between the years 1990 and 2009 in Turkey. 20 financial ratios of those banks, which are gathered by the Banks Association of Turkey, were taken into consideration in order to model and anticipate the change in their market shares.

This study would serve to determine the changes in market share of banks and to take measures quickly when it is necessary due to the fragile structure of the banking sector in Turkey. For this reason, computational intelligence algorithms were used in the study so that the algorithm that leads to the best estimation could be determined. As a result of the analysis, it is seen that Simple Linear Regression is the best algorithm for all periods. However, due to the financial crisis experienced in the beginning of 2000s in Turkey and the global crisis that had big impacts on our country as well as all the world economies in the period 2008-2009, the estimations for those periods (2001, 2002, 2008, and 2009) did not lead to good results.

As a result of the analysis, it is seen that the share of each bank within its own group was more effective with regard to the estimations for the market shares of banks throughout the

whole sector. For this reason, it is seen that such situation may have a positive impact on the banks, which are able to come to the fore in terms of technological developments, wide product range, and competitive advantage and increase their shares within their own groups, towards increasing their shares throughout the whole industry banks.

It is seen that using some specific features rather than all features leads to better results in the analysis. During the selection of features, it is seen that the specified features are in line with the features of the rules created.

## Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

## References

- [1] F. Çolak and A. Yigidim, *Türk Bankacılık Sektöründe Kriz*, Nobel Yayın Dağıtım, Ankara, Turkey, 2001 (Turkish).
- [2] S. Oksay, "Finansal piyasalarda yeni yasal düzenlemeler ihtiyacı ve türk finans sistemi," *Sosyal Bilimler Enstitüsü Öneri Dergisi*, 2000 (Turkish).
- [3] M. Ural, "bankacılık sistemimizde verimlilik," *D.E.Ü.İ.İ.B.F. Dergisi*, vol. 2, pp. 147–157, 1999 (Turkish).
- [4] Türkiye Bankalar Birliği, "50. Yılında Türkiye Bankalar Birliği ve Türkiye'de Bankacılık Sistemi (1958–2007)," İstanbul, Turkey, 2008 (Turkish).
- [5] Bankacılık Düzenleme ve Denetleme Kurumu, "Yıllık Rapor 2001," 2002 (Turkish).
- [6] H. Seyidoğlu, *Uluslararası İktisat, Geliştirilmiş*, Güzem Yayıncılık, İstanbul, Turkey, 13th edition, 1999 (Turkish).
- [7] B. Tunay, *Finans Sisteminde Yeni Yönetmelipler: Türk Finans Piyasalarının Bugünü ve Geleceği*, Beta Basım yayım Dağıtım, İstanbul, Turkey, 2001 (Turkish).
- [8] H. Tunç, "Finansal Kriz ve Türkiye Ekonomisi," *ISO Dergisi* 421, 2001 (Turkish).

- [9] K. Duman, "Finansal kriz ve bankacılık sektörünün yeniden yapılandırılması," *Akdeniz Üniversitesi İ.I.B.F. Dergisi*, vol. 4, 2002 (Turkish).
- [10] I. Sayim, K. Duman, and A. Korkmaz, "Türkiye ekonomisinde finansal krizler: bir faktör analizi uygulaması," *D.E.Ü.I.I.B.F. Dergisi*, vol. 1, pp. 45–69, 2004 (Turkish).
- [11] R. Karluk, *Türkiye Ekonomisi: Tarihsel Gelişim Yapısal ve Sosyal Değişim*, Beta Basım, İstanbul, Turkey, 2004 (Turkish).
- [12] S. Uyar, *Bankacılık Krizleri*, Ziraat Matbaacılık, Ankara, Turkey, 2003 (Turkish).
- [13] Bankacılık Düzenleme ve Denetleme Kurumu, "Bankacılık Sektörü Yeniden Değerlendirme Programı," 2001 (Turkish).
- [14] Türkiye Bankalar Birliği, "Bankalarımız 2009," İstanbul, Turkey, 2010, (Turkish).
- [15] Türkiye Bankalar Birliği, "Bankalarımız 2010," İstanbul, Turkey, 2011, (Turkish).
- [16] Weka Manual for Version 3-6-3.
- [17] M. Hall, G. Holmes, and E. Frank, "Generating rule sets from model trees," in *Proceedings of the 12th Australian Joint Conference on Artificial Intelligence*, pp. 1–12, Springer, Sydney, Australia, 1999.
- [18] R. Kohavi, "The power of decision tables," in *Proceedings of the European Conference on Machine Learning (ECML '95)*, N. Lavrac and S. Wrobel, Eds., pp. 174–189, Springer, Berlin, Germany, 1995.
- [19] J. R. Quinlan, "Simplifying decision trees," *International Journal of Man-Machine Studies*, vol. 27, no. 3, pp. 221–234, 1987.
- [20] T. Elomaa and M. Kääriäinen, "An analysis of reduced error pruning," *Journal of Artificial Intelligence Research*, vol. 15, pp. 163–187, 2001.
- [21] F. Esposito, D. Malerba, and G. Semeraro, "A comparative analysis of methods for pruning decision trees," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 5, pp. 476–491, 1997.
- [22] M. Kääriäinen, T. Malinen, and T. Elomaa, "Selective rademacher penalization and reduced error pruning of decision trees," *Journal of Machine Learning Research*, vol. 5, pp. 1107–1126, 2004.
- [23] I. H. Witten and E. Frank, *Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations*, Morgan Kaufmann, San Francisco, Calif, USA, 1999.
- [24] J. Q. Candela, C. E. Rasmussen, and C. K. I. Williams, "Approximation methods for Gaussian process regression," Tech. Rep., <http://research.microsoft.com/pubs/70486/tr-2007-124.pdf>.
- [25] C. Walder, K. I. Kim, and B. Schölkopf, "Sparse multiscale Gaussian process regression," in *Proceedings of the 25th International Conference on Machine Learning (ICML '08)*, W. W. Cohen, A. McCallum, and S. T. Roweis, Eds., pp. 1112–1119, ACM Press, Helsinki, Finland.
- [26] M. J. Best and N. Chakravarti, "Active set algorithms for isotonic regression; a unifying framework," *Mathematical Programming*, vol. 47, no. 3, pp. 425–439, 1990.
- [27] C. I. C. Lee, "The min-max algorithm and isotonic regression," *The Annals of Statistics*, vol. 11, no. 2, pp. 467–477, 1983.
- [28] P. M. Pardalos and G. Xue, "Algorithms for a class of isotonic regression problems," *Algorithmica*, vol. 23, no. 3, pp. 211–222, 1999.
- [29] V. de Simone, M. Marina, and G. Toraldo, "Isotonic regression problems," in *Encyclopedia of Optimization*, C. A. Floudas and P. M. Pardalos, Eds., Kluwer Academic, Dordrecht, The Netherlands, 2001.
- [30] M. Ayer, H. D. Brunk, G. M. Ewing, W. T. Reid, and E. Silverman, "An empirical distribution function for sampling with incomplete information," *The Annals of Mathematical Statistics*, vol. 26, no. 4, pp. 641–647, 1955.
- [31] R. E. Barlow, D. J. Bartholomew, J. M. Bremner, and H. D. Brunk, *Statistical Inference under Order Restrictions*, Wiley, New York, NY, USA, 1972.
- [32] D. L. Hanson, G. Pledger, and F. T. Wright, "On consistency in monotonic regression," *The Annals of Statistics*, vol. 1, no. 3, pp. 401–421, 1973.
- [33] O. Burdakov, O. Sysoev, A. Grimvall, and M. Hussian, "An  $O(n^2)$  algorithm for isotonic regression," in *Nonconvex Optimization and Its Applications*, G. di Pillo and M. Roma, Eds., vol. 83 of *Large-Scale Nonlinear Optimization Series*, pp. 25–33, Springer, 2006.
- [34] H. Akaike, "A new look at the statistical model identification," *IEEE Transactions on Automatic Control*, vol. 19, no. 6, pp. 716–723, 1974.
- [35] S. Ekinci, U. B. Celebi, M. Bal, M. F. Amasyali, and U. K. Boyaci, "Predictions of oil/chemical tanker main design parameters using computational intelligence techniques," *Applied Soft Computing Journal*, vol. 11, no. 2, pp. 2356–2366, 2011.
- [36] H. Yan, Y. Jiang, J. Zheng, C. Peng, and Q. Li, "A multilayer perceptron-based medical decision support system for heart disease diagnosis," *Expert Systems with Applications*, vol. 30, no. 2, pp. 272–281, 2006.
- [37] V. Havel, J. Martinovic, and V. Snasel, "Creating of conceptual lattices using multilayer perceptron," in *Proceedings of the International Workshop on Concept Lattices and Their Applications (CLA '05)*, R. Belohlavek and V. Snasel, Eds., pp. 149–157, Olomouc, Czech Republic, 2005.
- [38] A. B. Goktepe, E. Agar, and A. H. Lav, "Role of learning algorithm in neural network-based backcalculation of flexible pavements," *Journal of Computing in Civil Engineering*, vol. 20, no. 5, pp. 370–373, 2006.
- [39] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning representations by back-propagating errors," *Nature*, vol. 323, no. 6088, pp. 533–536, 1986.
- [40] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*, John Wiley & Sons, New York, NY, USA, 2001.
- [41] V. N. Vapnik, *The Nature of Statistical Learning Theory*, Springer, New York, NY, USA, 1995.
- [42] B. Keshari and S. M. Watt, "Hybrid mathematical symbol recognition using support vector machines," in *Proceedings of the 9th International Conference on Document Analysis and Recognition (ICDAR '07)*, pp. 859–863, IEEE Computer Society, Curitiba, Brazil, September 2007.
- [43] C.-M. Huang, Y.-J. Lee, D. K. J. Lin, and S.-Y. Huang, "Model selection for support vector machines via uniform design," *Computational Statistics and Data Analysis*, vol. 52, no. 1, pp. 335–346, 2007.
- [44] E. Frias-Martinez, A. Sanchez, and J. Velez, "Support vector machines versus multi-layer perceptrons for efficient off-line signature recognition," *Engineering Applications of Artificial Intelligence*, vol. 19, no. 6, pp. 693–704, 2006.
- [45] T. Benyang and D. Mazzoni, "Multiclass reduced-set support vector machines," in *Proceedings of the 23rd International Conference on Machine Learning (ICML '06)*, pp. 921–928, ACM, Pittsburgh, Pa, USA, June 2006.
- [46] J. N. S. Kwong and S. Gong, "Learning support vector machines for a multi-view face model," in *Proceedings of the British*

*Machine Vision Conference (BMVC '99)*, T. P. Pridmore and D. Elliman, Eds., pp. 503–512, British Machine Vision Association, Nottingham, UK, 1999.

- [47] J. Platt, “Fast training of support vector machines using sequential minimal optimization,” in *Advances in Kernel Methods-Support Vector Learning*, B. Schölkopf, C. Burges, and A. Smola, Eds., MIT Press, 1999.
- [48] M. A. Hall, *Correlation-based feature selection for machine learning [Ph.D. thesis]*, The University of Waikato, Hamilton, New Zealand, 1999.
- [49] J. G. Cleary and L. E. Trigg, “K\*: an instance-based learner using an entropic distance measure,” in *Proceedings of the 12th International Conference on Machine Learning (ICML '95)*, A. Prieditis and S. J. Russell, Eds., pp. 108–114, Morgan Kaufmann, Tahoe City, Calif, USA, 1995.
- [50] Y. Zhao, “Learning user keystroke patterns for authentication,” *International Journal of Mathematics and Computer Science*, vol. 1, pp. 149–154, 2005.
- [51] J. Friedman, T. Hastie, and R. Tibshirani, “Additive logistic regression: a statistical view of boosting,” *The Annals of Statistics*, vol. 28, no. 2, pp. 337–407, 2000.
- [52] A. Buja, T. Hastie, and R. Tibshirani, “Linear smoothers and additive models,” *The Annals of Statistics*, vol. 17, no. 2, pp. 453–555, 1989.
- [53] L. Breiman, “Bagging predictors,” *Machine Learning*, vol. 24, no. 2, pp. 123–140, 1996.
- [54] Türkiye Bankalar Birliği, “Bankalarımız 2005,” Türkiye Bankalar Birliği Yayıncıları, İstanbul, Turkey, 2006, (Turkish).
- [55] A. G. Karegowda, A. S. Manjunath, and M. A. Jayaram, “Comparative study of attribute selection using gain ratio and correlation based feature selection,” *International Journal of Information Technology and Knowledge Management*, vol. 2, pp. 271–277, 2010.

