Analyzing and predicting taxation rate with emphasis on the role of accounting measures and the governance system of the artificial intelligence approach of choosing the variable of neighborhood analysis and hyperdisk (NCA-HDLMC)

Ahmad Safari
Phd student , Department of Accounting, South Tehran Branch Islamic Azad University ,Tehran, Iran
safari_754@yahoo.com

Roya Darabi
Associate Professor, Department of Accounting, South Tehran Branch Islamic Azad University, Tehran, Iran
(Corresponding author)
royadarabi110@yahoo.com

Masoumeh Latifi Benmaran
Department of Accounting, South Tehran Branch, Islamic Azad University, Tehran, Iran
m.latifi@gmail.com

Submit: 27/07/2020 Accept: 12/09/2020

ABSTRACT

The purpose of this study is to analyze the factors affecting the taxation rate on companies listed on the Tehran Stock Exchange. This analysis was performed using the Neighborhood Analysis Algorithm and the Hyperdisk in a linear and nonlinear manner. The initial independent variables in this study include the variables of the governance system and the variables of accounting. To measure the taxation rate, the taxes rate paid by the company has been used. Experimental findings related to the study of 150 companies listed on the Tehran Stock Exchange in the period from 2012 to 2019, show that using the neighborhood analysis algorithm as a method of switching accounting criteria (quick ratio, return on assets, sales returns, ratio of sales to total assets) have the greatest impact on the company's taxation rate. Other research findings also showed that the nonlinear algorithm has a higher power than the linear algorithm in predicting taxation.

Keywords:

Taxation, Variables of Accounting and Governance System, Artificial Intelligence Algorithm.

1. Introduction

One of the most important and widespread challenges in most countries, especially in developing countries, is high tax evasion and low power of government in taxation. The size of the underground economy, which is closely linked to tax evasion, is between 14 and 16 percent of GDP in OECD member countries and 35 to 44 percent in developing countries [16]. This issue can be considered from two perspectives. The macro perspective is that the high tax rate is the main cause of tax evasion, and the micro perspective is where taxpayers make rational decisions about the expected benefits and costs of tax evasion. In these decisions, variables such as the probability of tax evasion, evasion penalty, taxpayer income level, cultural and social factors and human behavioral norms play a key role [5]. Since the government's budget to cover current expenditures must come from tax revenue, it is important to accelerate tax collection. Accelerating tax collection, regardless of other factors, requires the registration and maintenance of financial information and the reporting of the status and performance of each of the steps involved in the process. Prolonged taxation in Iran and uncertainty about timely compliance with legal regulations have always been one of the most important challenges in on time tax collection in the Iranian tax system; Therefore, this study tries to analyze and predict the rate of tax collection from the company.

2. Review of the Related Literature

Income tax is actually a tax paid by taxpayers and is subject to principles such as legality, fairness and transparency, and apart from these principles, taxes have goals such as distributive justice. Privatization, on the one hand, should not be used as an excuse to gain windfall wealth and increase socio-economic inequalities, and on the other hand, causes to neglect the long-term benefits of privatization, which are the outcome of tax revenues [14]. Khayat Behbahan and Zarei [6] studied the effect of corporate governance on corporate tax avoidance. For this purpose, 91 companies were selected from the Tehran Stock Exchange through the screening technique during 10 years of 2006-2015. The results show that the centralized ownership structure obtained through the right to vote is positively related to tax avoidance. Cash and cash flows are not related to tax avoidance. Nazemi and Pour Angha [3] in examining the supervisory role of the board of directors in tax avoidance showed that there is a significant positive relationship between the percentage of shares owned by the CEO in family companies and the effective tax rate and the role of the moderator and the board of directors moderates this relationship. Thus, in family-owned companies, the independence of the board of directors and the size of the board of directors intensify the dual role of the CEO and chairman of the board and weakens the relationship between the percentage of shares owned by the CEO and the effective tax rate.

Coster et al. [3] showed that a reduction in tax payments does not have an undesirable effect on a company's performance, while a reduction in operating costs (such as marketing, manufacturing, labor, etc.) can adversely affect a company's performance. Armstrong et al. [1] examined the effect of corporate governance on tax avoidance. They found that the effect of corporate governance on tax avoidance was mostly on both sides of the spectrum, in other words, the sequence of the statistical table of tax avoidance. Notably, the researchers observed a positive relationship between the percentage of non-executive members as well as financial skills and expertise of the board at the top end of the tax avoidance function. At the bottom of the spectrum was no tax avoidance function. Also, they did not observe a significant relationship between corporate governance and tax avoidance, neither in the median nor in the mean of distribution function. They argued that corporate governance seeks a balance in tax avoidance and tends to decline in companies that evade very high taxes, and in companies with very low taxes, the company's governance tends to increase tax avoidance. Khurana and Moser [8] discuss whether institutional ownership is effective in tax avoidance. They used the effective 5-year annual cash tax rate from 1995 to 2008 to measure tax avoidance, and concluded that companies with higher institutional ownership levels generally refused to pay more taxes. According to the objectives and theoretical foundations of the research, the following hypotheses have been developed:

1) Accounting measures have a higher ability to explain the power of taxation from the company than the measures of governance system.

2) Hyperdisk nonlinear algorithm has a higher ability to predict taxation from the company than the linear hyperdisk algorithm.

3. Research methodology

3.1. Statistical Population

This research is applied in terms of purpose and in terms of the type of study is a field-library study that is done using historical data as a post-event. The statistical population of this study is all companies listed on the Tehran Stock Exchange that comply with the following criteria: 1. Lack of change in the fiscal period in the period under review. 2) is not of investment companies, financial intermediaries, banks, insurance and leasing. 3) Their data is available. 4) The end of their fiscal year is March 20. Finally, due to the above-mentioned restrictions, 150 companies were selected as statistical population during the years 2012 to 2019. All companies were evaluated as statistical samples due to the availability of information.

4. Research variables

The initial independent variables of this study were considered in two groups of accounting and governance system measures as described in Table (1).

Table 1: Research variables

Accounting measures			
Sales return	**Quick ratio**	**The ratio of current assets to total asset**	**Financial distress**
Stock return	Tobin's Q ratio	The ratio of cash to asset	The ratio of operational return to sales
Company size	Financial leverage	The ratio of working capital	The ratio of dividend on asset
The ratio of current asset to current debt	Sales growth	The ratio of sales to total assets	Return on equity
Systematic risk	Asset return	The ratio of operational cash	The ratio of held cash
Governance system measures			
change of CEO	**Dual role of CEO**	**The size of the board of directors**	**The size of audit committee**
internal auditor unit	Proportion of non-executive directors	Percentage of institutional owners	Financial expertise of the audit committee
	Ownership concentration	Audit committee independence	Governmental ownership of company
Dependent variable: If the company's paid tax has increased this year compared to the previous year, the taxation rate has also been considered an increase and the artificial variable 1 has been used and otherwise the zero artificial variable has been used.			

5. Research findings

5.1. Descriptive Statistics

In order to examine the general characteristics of the variables and to analyze them accurately, it is necessary to be familiar with the descriptive statistics related to the variables. The table below shows the descriptive statistics of the variables used in the research, which include central indicators and scattering, for a sample consisting of 150 companies in the time period between 2012 to 2019

Selection of independent variables by analysis of neighboring components

The analysis of neighborhood component algorithm (NCA) is a filtering approach for selecting independent variables. NCA is a non-parametric and embedded method for selecting a feature to maximize the predictive accuracy of regression and classification algorithms. To do this, consider the classification of several classes with a set of tutorials, including n observations- each observation is a company-year-:

(1)

$$S = \{(x_i, y_i)\}, i = 1,2,\dots,n$$

In which $x_i \in \mathbb{R}^p$ is the vectors of the independent variables of the company (feature), $y_i \in \{1,2,\dots,c\}$ is the dependent variable of taxation (label) and c is the number of classes. The goal is to learn the classification $f: \mathbb{R}^p \rightarrow \{1,2,\dots,c\}$ so that by giving the independent variables to the company, ie f (x), the dependent variable predicts taxation rate. Consider a random classification that:

-Coincidentally, a company-year x is selected from the set S (Ref (x)).

-The selected company-year label, ie x, is chosen.

This method is similar to the 1-NN classification, in which reference point 3 is chosen as the nearest neighbor to the new point x; but in the NCA method, the reference point is randomly selected and all points in S have the same probability of being selected as the reference point; therefore, the probability of P (Ref (x) = x_j | S) that the company-year $x_j$is selected as the reference point x of the set S is greater than that, the point $x_j$is chosen as the reference point x based on the minimum distance $d_w$below.

(2)

$$d_w(x_i,x_j) = \sum_{r=1}^{p} w_r^2 |x_{ir} - x_{jr}|$$

$w_r$is the weight of the properties. Assume that

(3)

$$P(Ref(x) = x_j|S) \propto k\left(d_w(x_j,x)\right)$$

Where k is some kernel function or similarity function that when the distance $d_w(x_j,x)$ is small, it shows a large value as a sign of the high similarity of these two points (company-year). If the similarity function is considered as:

(4)

$$k(z) = \exp\left(-\frac{z}{\sigma}\right)$$

the reference point x is selected from the set S, and therefore the set P (Ref (x) = x_j | S) is equal to one for all j, so it can be argued that:

(5)

$$P(Ref(x) = x_j|S) = \frac{k\left(d_w(x_j,x)\right)}{\sum_{j=1}^{n} k\left(d_w(x_j,x)\right)}$$

Now, with the help of the leave-one-out strategy of the company-years, one by one they are randomly selected from the set , that is, the prediction of the $x_i$label is obtained using the data in S^{-i} in which S^{-i} is the set of training data S except the sample (x_i,y_i). The probability that the point $x_j$will be selected as the reference point for $x_i$is calculated as follows.

(6)

$$P_{ij} = P(Ref(x_i) = x_j|S^{-i}) = \frac{k\left(d_w(x_i,x_j)\right)}{\sum_{j=1,i\neq j}^{n} k\left(d_w(x_i,x_j)\right)}$$

The average probability of leave-one-out correct classification is equal to the probability of P_i classified randomly and correctly classifies observation i using S^{-i}.

$$P_i = \sum_{j=1,i\neq j}^{n} \text{probability that } x_j \text{ is the reference point for } x_i \times \text{I (label of } j \text{ matches } i))$$

(7)

$$P_i = \sum_{j=1,i\neq j}^{n} \text{probability that } x_j \text{ is the reference point for } x_i \times \text{I (label of } j \text{ matches } i))$$

in which $I(True) = 1$ and $I(False) = 0$, So,

(8))

$$P_i = \sum_{j=1,i\neq j}^{n} P_{ij}y_{ij}$$

When

(9)

$$y_{ij} = \begin{cases} 1, y_i = y_j \\ 0, otherwise \end{cases}$$

Average probability of leave-one-out correct classification using random classification can be written as follows

(10)

$$F(W) = \sum_{i=1}^{n} p_i$$

(W) is dependent on the weight vector, and the NCA's goal is to maximize this expression toward the weight vector. Considering a penalty sentence, it uses the following objective function as a function.

(11)

$$F(W) = \sum_{i=1}^{n} p_i - \lambda \sum_{r=1}^{p} w_r^2$$
$$= \sum_{i=1}^{n} \sum_{j=1, i \neq j}^{n} P_{ij} y_{ij} - \lambda \sum_{r=1}^{p} w_r^2$$
$$= \sum_{i=1}^{n} F_i(W)$$

In which $F_i(W) = \sum_{j=1, i\neq j}^{n} P_{ij} y_{ij} - \lambda \sum_{r=1}^{p} w_r^2$ and λ is the penalty coefficient.

The penalty term leads most of w to zero. After selecting the Kernel parameter σ in σ, as 1, discovering the weight vector w, as the following minimization problem, can be stated as below for the specified value of λ .

(12)

$$\widehat{W} = \arg\min_{W} f(W) = \sum_{i=1}^{n} \alpha_i f_i(W)$$
$$f(W) = -F\,(W)$$
$$f_i(W) = -F_i(W)$$

Therefore, it finds weights that minimize classification errors.

7. Hyperdisk algorithm based on large margin classification

The hyperdisk algorithm, based on the large margin classification (HDLMC), is a linear binary classification framework with large margin that approximates each class with a hyperdisk-common cover affine and a hypersphere that includes training samples in the feature space and then finds the linear classification that maximizes the distance between the two separators. This method has been developed for nonlinear feature spaces using kernel tricks, and multi-class problems with combination of binary classifiers are treated in a similar way as SVM. Consider the issue of binary classification with training data $\{x_i, y_i\}, i = 1,\ldots,n, y_i \in \{-1,1\}, x_i \in R^d$. As mentioned above, SVM can be considered as a method that first approximates a convex cover for each class using training samples of that class and then finds a superpage that minimizes the margin between the two covers. The convex cover includes all the points that:

(13)

$$H_c^{convex} = \left\{ x = \sum_{i=1}^{n_c} \alpha_i x_{ci} \mid \sum_{i=1}^{n_c} \alpha_i = 1, \alpha_i \geq 0 \right\}$$

They can be expressed as convex combinations of training samples (for example, linear combinations with non-negative coefficients with the sum value equaling to one). If $\{x_{ci}\}_{i=1,\ldots,n_c}$ are the samples of class C, the convex cover will be as follows. Consider several covers, the separator superpage s obtained by finding the points on each cover so that these points minimize the distance between the two covers, and in this case the superpage separating the two classes is equal to the perpendicular bisector of the line connecting the two nearest class points. The convex cover is the smallest convex set that includes all samples. The HDLMC method is based on a freer convex cover approximation, in which affine cover subscriptions of the samples and the smallest hypersphere include the samples in the input space. It can be said that this convex cover is the smallest hypersphere which includes the area covered by the affine cover of the samples. The affine cover is the smallest space including the samples that is calculated using the following equation:

(14)

$$H_c^{affine} = \left\{ x = \sum_{i=1}^{n_c} \alpha_i x_{ci} \mid \sum_{i=1}^{n_c} \alpha_i = 1 \right\}$$

The convex cover is unlimited, so large margin classifiers must be very sparse to be realistic in order to separate them in issues with the dimensions of the convex cover-based methods. In this case, it seems that making the convex cover sparse, without the need to add more parameters to the model, will be useful. For this purpose, in this article, hyperdisk approximation has been used. By limiting the model to a supersphere that includes training samples within the convex cover, it provides better results in localization of the classes inside the affine cover without losing the simplicity and stability of the affine method. The obtained model may estimate the value of the class a little higher or lower, but it has better answers than the convex cover in high-dimensional problems. See Figure (1). Hypersphere models are suitable for outlier detection.

Some classifiers of the nearest-convex-model type have been introduced, but previous studies have shown that such applications perform better than convex, affine, and hyperdisk covers, which is why hyperdisk is used instead of the hypersphere in this paper.

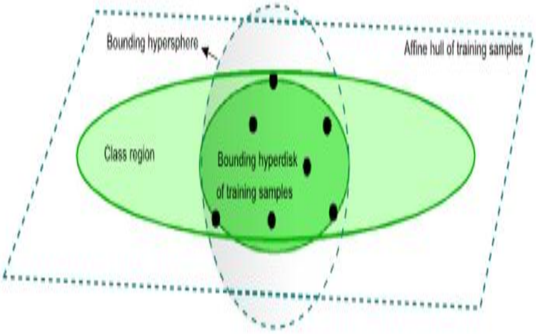

Figure 1: Display of hyperdisk cover for a Class {Cevikalp, 2013 # 16}

The smallest hyperdisk containing a class can be written as follows:

(15)

$$H_c^{disk} = \left\{ x = \sum_{i=1}^{n_c} \alpha_i x_{ci} \mid \sum_{i=1}^{n_c} \alpha_i = 1, \left\| \sum_{i=1}^{n_c} \alpha_i x_{ci} - S_c \right\|^2 \le r_c^2 \right\}$$

In which S_c is the center of the class and its radius r_c can be determined using the following equations or its dual equivalent:

(16)

$$\min_{s_c, r_c} \left(r_c^2 + \gamma \sum_i \xi_i \right)$$
$$s.t.\ \|x_{ci} - s_c\|^2 \le r_c^2,\ i = 1, 2, \cdots n_c$$
$$or\ \ its\ \ dual$$
$$\min_{\alpha} \left(\sum_{i,j} \alpha_i \alpha_j \langle x_{ci}, x_{cj} \rangle + \sum_i \alpha_i \langle x_{ci}, x_{ci} \rangle \right)$$
$$s.t.\ \sum_i \alpha_i = 1,\ \ \forall i\ \ 0 \le \alpha_i \le \gamma,\ \ i, j = 1, 2, \cdots n_c$$

In which $\langle - \rangle$ shows the internal multiplication between the sample. Here, α_i is the Lagrangian coefficients for the center $s_c = \sum_i \alpha_i x_{ci}$ and $\gamma \in [q / n_c, 1]$ is the roof parameter that by setting a limited value for it, outliers can be removed. For each x_{ci} the corresponding radius is $r_c = \|x_{ci} - s_c\|$,and $0 < \alpha_i < \gamma$. In a d dimensional space, there is a maximum of α_i ,d + 1, which is usually non-zero.

To find the separator maximum between two hyperdisks, you need to find the closest points between them (each point on a disk) and obtain the perpendicular bisector of the line connecting them. If we show the points with x_+ and x_-, the bisector will be $\langle w, x \rangle + b = 0$, and w and b are calculated as follows:

(17)

$$w = \frac{x_+ - x_-}{\|x_+ - x_-\|}$$

(18)

$$b = -\langle w, (x_+ - x_-)/2 \rangle$$

The HDLMC algorithm uses a quadratically constrained quadratic program to find these points. Solving a quadratically constrained quadratic program is possible (QCQP). Suppose that X_+ and X_- are matrices with columns that are positive and negative class samples, respectively. As before, first the center and the radii of two-class superspheres are calculated. Then the closest points x_+ and x_- on the hyperdisks can be rewritten similarly to the following optimization problem, in which $\alpha_{\pm}$ is the vector of weight expansion as for the nearest points ie x_+ and x_-, we have:

(19)

$$x_{\pm} = X_{\pm}\alpha_{\pm}$$

Bahrani, Yusuf Ebn Ahmad, Al-Hadaiq Nazera

To obtain a vote x_+ and x_- the following equation must be solved in order to find $\alpha_{\pm}$ and as a result x_+ and x_-.

(20)

$$\begin{aligned}&\min_{\alpha_+,\alpha_-}\|X_+\alpha_+ - X_-\alpha_-\|^2\\ &s.t.\ \sum_i \alpha_{+i} = 1,\quad \sum_j \alpha_{-j} = 1,\quad i(j) = 1,2,\cdots n_+(n_-)\\ &\quad \|X_+\alpha_+ - s_+\|^2 \le r_+^2,\quad \|X_-\alpha_- - s_-\|^2 \le r_-^2\end{aligned}$$

Now, by defining $\alpha \equiv \begin{pmatrix}\alpha_+\\ \alpha_-\end{pmatrix}$ and $X = [X_+, -X_-]$, and also considering $e_{\pm}$ as a column vector of valves 1 with the appropriate dimension $\begin{bmatrix}1\\1\\ \vdots\\1\end{bmatrix}$, the above problem can be rewritten as follows:

(21)

$$\begin{aligned}&\min_{\boldsymbol{\alpha}} \boldsymbol{\alpha}^T\mathbf{G}\boldsymbol{\alpha}\\ &\text{s.t.}\boldsymbol{\alpha}_+^T\mathbf{e}_+ = 1,\ \boldsymbol{\alpha}_-^T\mathbf{e}_- = 1,\\ &\quad \boldsymbol{\alpha}_+^T\mathbf{G}_+\boldsymbol{\alpha}_+ - 2\boldsymbol{\alpha}_+^T\mathbf{X}_+^T\mathbf{s}_+ + \mathbf{s}_+^T\mathbf{s}_+ \le r_+^2,\\ &\quad \boldsymbol{\alpha}_-^T\mathbf{G}_-\boldsymbol{\alpha}_- - 2\boldsymbol{\alpha}_-^T\mathbf{X}_-^T\mathbf{s}_- + \mathbf{s}_-^T\mathbf{s}_- \le r_-^2,\end{aligned}$$

Where we have:

(22)

$$G = X^TX,\ G_+ = X_+^TX_+,\ and\ \ G_- = X_-^TX_-$$

This QCQP issue is a quadratically constrained quadratic program. It is also convergent because G, G_+ *and* G_- are all positive definite. Therefore, by solving it, we can obtain $\alpha_{\pm}$ and as a result x_+ and x_-. The HDLMC algorithm is a kernel in the form (23). Suppose that $\phi(.)$ is the explicit feature space embedding and $k(x_i, x_j) = <\phi(x_i), \phi(x_j)>$ is the kernel function corresponding it. We want to plot x points inside the affine cover of a set from the known samples $\{x_i\}_{i=1,\dots,n}$. Suppose $\Phi = [\phi(x_1),\dots,\phi(x_n)]$ is the explicit feature space embedding matrix, $K = \phi^T\phi = [k(x_i, x_j)]$ is the Kernel matrix $n \times n$, and $K_x = \Phi^T\phi(x) = [k(x_i, x)]$, $n \times 1$ is the Kernel vector x of the samples. The mean of samples feature space is $\mu = (1/n)\Phi 1_n$ (1_n is a n-dimensional vector of 1). The QCQP algorithm is now stated as an internal multiplier, so it is easy to kernelize it. Assume that the kernel $k(x_i, x_j) = <\phi(x_i), \phi(x_j)>$ and $\Phi_- = [\phi(x^-_1),\dots,\phi(x^-_n)]$ are explicit matrices of feature vectors of the set of positive and negative training samples. Consider the explicit kernel matrices that correspond to them as $K_+ = \Phi_+^T\Phi_+$, $K_- = \Phi_-^T\Phi_-$ and $K_+ = \Phi_+^T\Phi_-$.

$K = \begin{pmatrix}k_+ & -k_{+-}\\ -k_{+-}^T & k_-\end{pmatrix}$ is to be defined. Suppose the centers of the hypersphere is in the feature space $K_+\beta_+$ and $K_-\beta_-$. Then QCQP is as follows

(23)

$$\begin{aligned}&\min_{\alpha} \alpha^TK\alpha\\ &s.t.\alpha_+^Te_+ = 1, \alpha_-^Te_- = 1\\ &\quad \alpha_+^TK_+\alpha_+ - 2\alpha_+^TK_+^T\beta_+ + \beta_+^TK_+\beta_+ \le r_+^2,\\ &\quad \alpha_-^TK_-\alpha_- - 2\alpha_-^TK_-^T\beta_- + \beta_-^TK_-\beta_- \le r_-^2.\end{aligned}$$

Note that β_+ and β_- may be sparse because hyperspheres centers depend only on the training points that are exactly on the sphere. This can be used to reduce the required calculations. Consider the weight of α, is the classifier $\alpha_+^T K_{+x} - \alpha_+^T K_+^T K_{-x} + b > 0$, and $b = -(\alpha_+^T K_{+x} \alpha_+ + \alpha_-^T K_- \alpha_-)/2$, K_{+x} and K_{-x} are the kernel vectors of the classification samples with positive and negative training samples.

8. Proposed approach

In this section, the proposed filter approach for selecting independent variables is presented along with the presentation of the taxation prediction model. This approach consists of two steps: the first step is to select a subset of the independent variables by the NCA algorithm, and the second step is to build the model in order to predict the taxation based on the hyperdisk algorithm based on the large margin classification. The perspective of the proposed method is shown in Figure (2). First, the data are divided into two categories, training and evaluation, using the 10-item validation method. The algorithm for analyzing neighboring components suggests a subset of independent variables. Hyperdisk algorithm based on large margin classification is learned using training data. The best independent variables are selected along with the hyperdisk model based on large margin classifier as the output of the proposed approach and now the evaluation data that hyperdisk model based on large margin classifier have not observed them yet, are given to the model and the degree of prediction error is reported.

Variable selection by analyzing neighboring components

Company-year data were given to variable to the algorithm of analysis and analysis of neighboring components. 4 independent variables were selected. These variables include quick ratio, asset return, sales return, and ratio of sales to total asset.

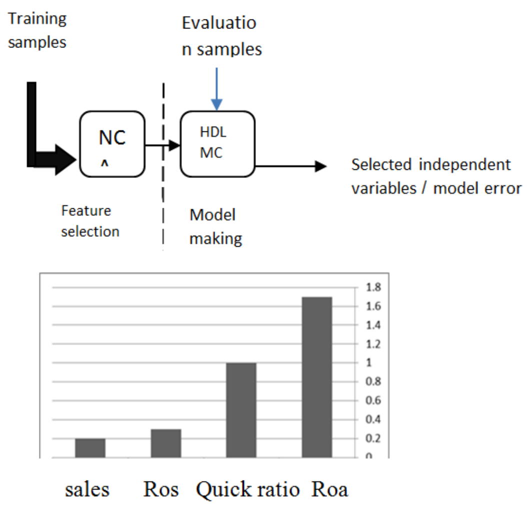

Figure 2: Selection of independent variables and training process

9. Hyperdisk prediction results based on large margin classification

To evaluate the hyperdisk models based on large margin identification, the identification rate and confusion matrix have been used. Also, in order to implement it fairly and to investigate the phenomenon of over-fitting, the 10-fold cross validating method has been used. Company-years are divided into two categories: training and evaluation by 10-fold cross validating method. The training data is given to the hyperdisk algorithm based on the large margin classification. The hyperdisk algorithm, based on the margin classifier, has learned the parameters of the model; this model will be linear if it does not use the Kernel and will be non-linear if it does and the model is not visible. In this dissertation, tht two models are learned and the coefficients of the linear model are shown in the results section. Solving the problem of optimization of hyperdisk based on the large margin classification has been done with the help of two toolboxes mlcv_QuadProg and MOSEK, and the MOSEK academic license has been used for implementation. After implementing the learning process of this algorithm, in order to examine how successfully the hyperdisk model based on the large margin classifier has gone through the learning process, the same training data that was previously given to the algorithm is given to the model again, this time the difference is that hyperdisk model based on the large margin classifier, predicts the value of the dependent variable, then the average of the 10 identification rate of the 10-fold cross validation method is calculated and reported in the table below. The closer this identification rate is to 100, the better the model learns. This year, the identification rate of HDLMC algorithm linear is 62.42 and nonlinear is 92.90, and next year it is 60.77 for linear model and 89.13 for nonlinear. These percentages indicate the learning rate of the model. This year's identification rate is higher than next year's identification rate, so the hyperdisk algorithm based on large margin classifier has learned this year more accurately than next year. On the other hand, the accuracy of the nonlinear model is significantly better than that of hyperdisk based on linear large margin classifier.

Table 2: Mean of Identification Rate for Evaluating Hyperdisk Model Learning Based on Large Margin Classification for Taxation This Year and Next Year

	Current year		Next year	
	HDLMC		HDLMC	
	Linear	NonLinear	Linear	NonLinear
1	63.22	92.85	60.51	88.78
2	63.04	92.62	60.85	89.48
3	62.64	92.92	60.58	89.14
4	62.7	92.85	60.73	89.22
5	61.04	92.92	60.58	89.04
6	61.22	93	61.25	88.78
7	62.61	92.55	61.03	89.3
8	63.16	92.77	60.66	89.14
9	62.61	93.14	59.99	89.3
10	62	93.29	61.55	89.04
Avg.	62.42	92.9	60.77	89.13

To investigate the occurrence of a phenomenon called over-fitting, company-years of the test that have not been seen yet by the algorithm, are given to the hyperdisk obtained model based on the large margin classification. Hyperdisk's algorithm, based on the large margin classifier predicts the taxation for theses company-years. It is obtained by comparing the predicted value with the actual value of the efficiency of the obtained model for company-years that have not been seen by the algorithm. This process is performed by a 10-fold cross validation method so that all companies are tested at least once and we can reply more on the result. The mean of this identification rate is shown below. The HDLMC algorithm for company-years that has never seen, has a prediction accuracy close to the company-years of training, and only some errors have been added to it. Since the difference between the training and test data identification rates is small, there is no over-fitting. This year, the identification rate for predicting the linear HDLMC algorithm is 61.27 and is 92.89 for the non-linear, and next year it is 60.63 for linear model and 89.12 for the nonlinear.

Table 3: Mean of Identification Rate for Evaluating the Efficiency of Hyperdisk Model Based on Large Margin Classification for Taxation this Year and Next Year

	Current year		Next year	
	HDLMC		HDLMC	
	Linear	NonLinear	Linear	NonLinear
1	55.47	93.29	63.09	92.19
2	57.81	95.33	58	85.94
3	59.06	92.62	63.09	88.98
4	57.03	93.29	59.73	88.28
5	71.88	92.62	59.73	89.84
6	64.84	91.95	58.39	92.19
7	60.16	95.97	59.73	87.5
8	62.99	93.96	62.42	88.98
9	60.16	90.6	66.44	87.5
10	63.28	89.26	55.7	89.84
Avg.	61.27	92.89	60.63	89.12

10. Conclusion and Recommendations

Collecting tax revenue has always been difficult compared to non-tax revenue. Iran is no exception, and in upstream development laws and plans, the growth of tax revenues is recognized as one of the main goals of Article 17 of the General Economic Policy of the Resistance Economy, which refers to the reform of the government revenue system; that is, the reform of the income system should be done by increasing the share of tax revenues. It is clear that reducing the incentives of evasion in taxpayers is one of the main prerequisites for the society to achieve the above goals and promote tax justice. This, in turn, requires the tax organization to avoid tax evasion and to avoid access to information and also heavy penalties if any tax evasion is discovered Since the government's budget to cover current expenditures must come from tax revenue, it is important to accelerate tax collection. Accelerating tax collection, regardless of other factors, requires the registration and maintenance of financial information and the reporting of the status and performance of each of the steps involved in the process. Prolonged taxation in Iran and uncertainty about timely compliance with legal regulations have always been one of the most important challenges in on time tax collection in the Iranian tax system; Therefore, this study tries to analyze and predict the rate of tax collection from the company.

The results of this study show that using the neighborhood analysis algorithm as a method for variable selection of accounting criteria (quick ratio, return on assets, sales returns, ratio of sales to total assets) have the greatest impact on the taxation rate from the company. Other research findings have also shown that nonlinear algorithm has a higher power than linear algorithms in predicting taxation. In analyzing the results, it can be stated that the role of governance system as a supervisory, leading and decision-making element in areas such as the present study can be very low compared to accounting criteria and this issue still confirms the inefficiency of the governance system in some areas. The results of this study show that using the neighborhood analysis algorithm as a method for variable selection of accounting criteria (quick ratio, return on assets, sales returns, ratio of sales to total assets) has the greatest impact on the taxation rate from the company. Other results of this research also showed that the nonlinear algorithm has a higher power than the linear algorithm in predicting taxation. In analyzing the results, it can be stated that the role of governance system as a supervisory, leading and decision-making element in areas such as the present study can be very low compared to accounting criteria and this issue still confirms the inefficiency of the governance system in some areas. The results of this study is consistent with the researches by Khodamipour et al. [9]; Samimi et al. [15]; Kordestani and Jahangiri [10]; Gapta and Newbury [4]; Cracker et al. [3]; Maboodi and Darabi [11]; Mehrabanpour et al. [12] and Iers et al. [2] that examined the relationship between tax and financial performance and accounting criteria and financial ratios and is somewhat consistent with the results obtained from the hypothesis testing in this study. One of the reasons for the research results could be that the governance system in Iranian companies is complex and has its own specific items (such as government interference in the selection of the board of directors and the influence of political elites in the corporate governance system) and lack of expressing costs, especially tax costs has caused the results of the present study do not meet the theoretical foundations and what we expect from the governance system. Based on the results of the research, tax decision makers are recommended to be aware of changes in criteria (quick ratio, sales and return on assets) and to consider changes in these items as a model because it changes the tax rate in companies listed in the exchange. Analyzing the relationship between taxation and the company's performance during the company's life cycle and by dividing it into high and low growth opportunities can be presented as a proposal for future researches. It is also recommended to use other criteria of the management system as independent criteria to explain the provincial tax and compare it with the results of this study.

References

1) Armstrong, C.,Blouin, J. Jagolinzer, A.,Larcker, D. (2015). Corporate governance, incentives,and tax avoidance, Journal of Accounting and Economics, 60 (1), 1 – 17.
2) Ayers, C. , Benjamin,L,Casey S . (2011)."Does Tax Deferral Enhance Firm Value?"Working Paper.5(4),13-19.
3) Crocker, K. , Slemrod, J. (2005). Corporate tax evasion with agency costs, Journal of Public Economics.89 (9-10),1593-1610.
4) Gupta,S.,Newberry, K. (1977)."Determinants of the Variability In CorporateEffectiveTax Rates:Evidence From Longitudinal Data", Journal Of Accounting And Public Policy 16(5),1-34.
5) Hindriks, J, & Myles, G. (2004). Intermediate Public Economics E-book
6) KhayatBehbahan,Z,Zarei,S.(2018). Investigating the effect of corporate governance on corporate tax avoidance in companies listed on the Tehran Stock Exchange, Horizons of the Humanities. 3(24). 17-38. (in Persian).
7) Khodami pour ,A , Amininia, M , Fadavi, M, H . (2015). Investigating the Relationship between Tax Avoidance and Cash Holding whit Considering the Financial Constraint Condition. ،Financial Accounting Knowledge Quarterly.2(1) ,82-67. (in Persian).
8) Khurana, I. K. and Moser, W. J. (2012).'Institutional Shareholders' Investment Horizons and Tax Avoidance', TheJournal of the American Taxation Association. American Accounting Association. 35 (1), 111–134.
9) Koester, A., Shevlin, T., Wangerin, D. (2017). The role of managerial ability in corporate tax avoidance. 63(10) , 3147-3529.
10) Kordestani, Gh , Jahangiri Livari, B. (2016). The Impact of Tax Avoidance on the Relationship between Product Market Power with Returns and Stock Price Informati، Financial Accounting Quarterly.1(27), 19-38. (in Persian).
11) Maboudi M, Darabi R. (2014). The Effect of Tax Deferral on Firm Value in the Companies Listed in Tehran Stock Exchange . taxjournal. 22 (23), 65-74. (in Persian).
12) Mehrabanpour . M, R , Vaghfi S,H , mamsalhi ,P. (2017). Investigating the Impact of the Effective Factors on tax avoidance and its impact on Performance Evaluation Criteria using Structural Equation.، Financial Accounting Knowledge Quarterly.4(3), 86-65. (in Persian).
13) Nazemi , A , Poorangha , M, R. (2018). Examining the Monitoring Role of the Board of Directors in Tax Aggressiveness of Family Firms Listed in Tehran Stock Exchange. Journal of Accounting Reviews. 5(18) , 106-83 . (in Persian).
14) RostamiV,KamarkhaniA. (2016). Competition, Taxation & Improvement of Privatization. The Journal of Planning and Budgeting. 21 (2), 34-45 . (in Persian).
15) Samimi Momen, M, Dehghan,A, Nobari ,Tabrizi ,A(2017) Determining the relationship between tax avoidance with performance and cash held in companies listed on the Tehran Stock Exchange, National Conference on World Scientific Research in Management, Accounting, Law and Social Sciences .6(4) , 66-78.(in Persian).
16) Schneider, F., Enste, D.H., (2000). "Shadow Economies: Size, Causes, and Consequences", Journal of Economic Literature,3(8),77–114