



Analysis of Taxpayers with a Data Mining Approach

Ghazaleh Behdadfar¹, Farid Asgari^{2*}, Babak Hajikarimi³, Abdollah Nazari⁴

¹Ph.D student, Department of Industrial Management, Abhar Branch, Islamic Azad University, Abhar, Iran.

²Assistant Professor, Department of Industrial Management, Abhar Branch, Islamic Azad University, Abhar, Iran (Corresponding author).

^{3,4}Assistant Professor, Department of Industrial Management, Abhar Branch, Islamic Azad University, Abhar, Iran.

abstract

Analysis of taxpayers' behavior is important in maintaining tax justice and strengthening the foundations of the tax system. By better understanding the behavior of taxpayers, appropriate measures can be taken to prevent tax evasion. In this study, tax payers were analyzed with data mining approach. In this regard, X-Means and K-Means algorithms were used to cluster taxpayers through RapidMiner software and the information of 9994 taxpayers in several business categories. Based on the results, the number of optimal clusters was seven in the Kamiangin method, and the average number of optimal clusters was three in the X method. Examining the clusters in the Kamyangin method shows that in cluster (7), the amount of tax contribution expressed is lower than other clusters. In this cluster, there were businesses related to laboratories, radiology, physiotherapy, etc. On the other hand, the highest share of tax expressed from diagnosis also belonged to cluster (1). In this cluster of money changers; Guild of cloth bankers; bags and shoes, bags and suitcases; notary offices; audio and video equipment; There were sweets, nuts and ice cream. Based on the X-mean results, taxpayers were classified into three clusters, and the largest share of the studied indicators in cluster one includes money changers; Guild of cloth bankers; bags and shoes, bags and suitcases; notary offices; audio and video equipment; Sweets, nuts and ice cream; vehicles and spare parts; Dentists. Also, the decision tree classification method was used to predict the final tax. Based on the results in the prediction of the final tax, the highest weight is assigned to the variable of zero declaration ratio, declared tax share and finally the number of taxpayers. Also, the accuracy of the artificial neural network has been obtained at 97%, which shows that it is a more characteristic approach to clustering with an accuracy of 66.67%.

Key Words: taxpayers, data mining approach, tax

1. Introduction

Income tax, along with other taxes, has long been an important source of government revenue, which is used to support public services, pay government debts, and contribute to a country's development goals. Taxes are levied by the government on many types of income of individuals and businesses, and vary depending on the amount earned from wages, salaries, and the individual circumstances of the business (Smelser and Baltes, 2011). Despite the fact that many



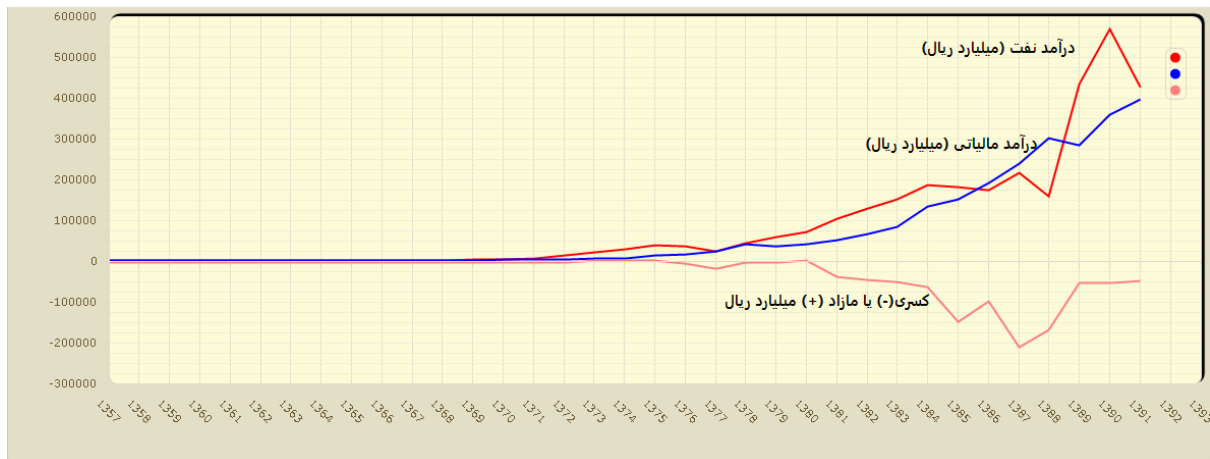
people understand the need to pay income tax, taxpayers are increasingly developing new tax evasion methods, which naturally make it difficult to detect and require the use of modern and robust tax fraud detection methods. As Freya (2018) stated, although income tax fraud is a hot topic in Rwanda, there are not enough studies to determine the extent of the problem and the amount of money lost as a result. According to the study of Mororonkur et al. (2022), more than 20 billion Rov (currency) were lost due to tax fraud in the five years from 2016.

Most of these cases were discovered using traditional random audit methods and sometimes whistle-blowing and incentive methods. KPMG research showed that 750 frauds were investigated between March 2013 and August 2015, and that only 3% of these fraud cases were detected using data analysis and machine learning, while 44% were detected using intuition and methods. Traditional were discovered. According to the study of Koltor and Glayan (2017), the most common fraud detection methods are rule-based systems, which are no longer effective enough. Auditors review the income tax records of millions of individuals and businesses and rely on knowledge, experience, and intuition to determine whether tax fraud has occurred. As discussed in González and Valásquez's (2013) review, this method has two drawbacks: it relies heavily on past experience, which means it often misses new fraud tactics; And it is based on knowledge which makes it expensive to maintain and update. While fraudsters are constantly developing new tactics to avoid paying taxes, it's no longer possible to track them down using human intuition alone or hoping auditors will notice. Fortunately, machine learning has changed the way things are done in many fields, including fraud detection. Although they did not focus on income taxes, several studies (González and Valásquez, 2013; Dias et al., 2016; Wu et al., 2012; Mororonkur et al., 2022) applied supervised and unsupervised machine learning methods, and While many researches used unsupervised machine learning due to the lack of labeled data, supervised machine learning produced better and more accurate results in tax fraud detection.

Considering the above, the need for optimal and correct management of government revenues, including taxes, is felt more than in the past. Although tax revenue is important, tax administration is more important (Gosh and Douglas, 1994). Due to the fact that tax evasion has become very common in the current era, it seems that tax management has been challenged to a great extent and has even become ineffective in most cases (Mubarak and Ashraf, 2017). This phenomenon with different intensity and weakness is common in both developed and developing countries (Bonucci et al., 1999) and is often considered as modern financial crime (De Roux et al., 2018). Due to the existence of a large informal economy in developing countries, the loss of tax revenue caused by tax evasion is higher in these countries compared to developed countries. On the other hand, in countries with natural resources such as oil resources, the increasing importance of tax income compared to the income from natural resources has attracted the attention of researchers to conduct more studies (Perez et al., 2019). In graph (1), the amount The income from oil, taxes and budget deficit in Iran are shown:



Chart 1- Oil revenues, taxes and government budget deficit



Source: Central Bank

In Iran, like other countries, tax evasion has become one of the main challenges of governments (Rahal, 2017). Rahal (2017) suggests that the failure to collect taxes has a significant effect on deepening the budget deficit, and the budget deficit, in turn, has a negative effect on economic growth (Eschinger and Anste, 2000). Therefore, paying attention to this ominous economic phenomenon and investigating the factors affecting it requires more attention and investigation. Many studies inside and outside the country have dealt with the phenomenon of tax evasion (for example, Rahal, 2017; Rahi, 2004; Al-Khatib and Abdul Jabbar, 2004; Abu Umaria, 2019; Al-Sharuf, 2019; Kobaja and Omar, 2019). But the predictors of evasion Taxation is much wider than it can be covered by studies in this field, so there are still many research gaps to better understand the determinants of tax evasion (Khalif and Acek, 2015). Existing studies such as (Jackson and Milliron, 1986; Richardson and Sawyer, 2001) consider the key factors affecting tax evasion to be a combination of demographic, economic and behavioral factors. Other studies have only focused on some behavioral factors (Alm and Turgler, 2019; Bank et al., 2015; Krichler et al., 2008), but they are very limited. Based on this, in this study, an attempt has been made to evaluate tax evasion from the perspective of various types of businesses.

2- Theoretical foundations and research background

Governments can improve tax compliance through effective tax measures that influence taxpayers' behavior; Because the gap between potential and actual tax revenue is due to incomplete and ineffective tax management (Louis et al., 2019; Mbarato et al., 2020).

Tax non-compliance is an important challenge that has become increasingly common around the world (Kafila and Girmai, 2009). According to Shalo (2018), many factors, including non-compliance with tax regulations, cause low tax collection. Tax non-compliance generates significant revenue for the government, as many taxpayers default on their tax obligations and



are prosecuted for failure to pay taxes on time. The tax system, despite the government's incentives for voluntary adoption, relies heavily on the impact force.

Tax compliance refers to the willingness of taxpayers to follow tax laws to achieve economic balance in a country. This process is the method of persuading taxpayers to comply with relevant tax laws (Oladipo et al., 2022). On the other hand, tax non-compliance is a major challenge for many tax authorities and it is difficult to convince taxpayers to comply. According to the economic deterrence model, the behavior of taxpayers is affected by the tax rate, which determines the benefits of tax avoidance and the probability of detecting and punishing fraud and the costs of tax avoidance. Another compliance theory, financial exchange theory, suggests that incentives can promote compliance and that governments can increase compliance by providing more effective and affordable products. The third theory of compliance, social influence, claims that the compliance behavior and attitudes of individuals towards the tax system are influenced by the behavior and social norms of the individual reference group (Snaveley, 1990). The theory of adaptive behavior emphasizes that the fairness and reasonableness of the tax system affects tax compliance behavior (Walsh, 2012). The political accountability model states that tax compliance depends on the level of citizens' trust in their governments (Krishler et al., 2008). The theory of planned behavior also shows that perceived behavioral control depends on beliefs (related to attitudes towards behavior and mental norms), that is, control of beliefs (perceptions about acquiring skills, resources and opportunities) (Saad, 2011).

A country's tax system strives to collect taxes in an orderly and professional manner to support the government's tax policy. While recent experiences show significant progress in certain regions of the world, tax administrators have faced challenges in achieving this goal. Taxpayer non-compliance is the main cause of developing countries' problems in effectively collecting tax revenues (Okpiyo et al., 2019). Tax evasion and avoidance are characteristics of tax offenses that are defined as the inability to file tax returns, report income, accurately calculate deductions, and pay on time (Jenkins and Forlemo, 1993). Both tax avoidance and evasion pose significant economic challenges to a country. For example, tax avoidance can lead to investment diversion, lead individuals and firms to undervalue their assets, or even exempt some from taxation. On the other hand, tax evasion can destroy business ethics and morals, as people look for loopholes in the system, which can lead to higher dividend reporting and increased internal profits (Dalow et al., 2012).

Therefore, this will have a negative impact on the economy and lead to national inflation. One of the measures that the government or tax authorities take to reduce the rate of tax evasion and compliance is to increase the level of voluntary or mandatory compliance of taxpayers.

Although several studies have been conducted on factors affecting tax compliance behavior, they have found different results for the same variables. For example, (Jambria, 2020, Engda and Bisa, 2014 and Diganto, 2018) conclude that fairness and justice do not have a significant impact on tax compliance behavior. Adam and Simert (2020) propose that fairness, justice,



equity rights, and fairness in the tax system have a significant impact on tax compliance behavior. The researchers mentioned above also found differences in their results when it came to another variable, tax knowledge. For example, Assaf and Sabat (2019), Jambrieh (2020), and Diganto (2018) concluded that tax knowledge significantly affects tax compliance behavior. On the other hand, Adam and Simert (2020), who found that tax compliance behavior is not affected by tax knowledge and that tax compliance and knowledge are not related. Due to these conflicting results, more research is needed based on different theories of tax acceptance. This study is organized as a theoretical foundation for tax compliance behavior and the development of hypothesis, research material, results, and conclusion.

tax compliance

Various factors can affect taxpayers' attitudes towards tax compliance and subsequently affect their tax compliance behavior from cultural components and other socio-cultural factors; Factors affecting tax compliance and non-compliance are different in different countries (Okpiyo et al., 2019). Considering this analysis, Barbuta Miso (2011) classifies the determinants of tax compliance into three non-economic categories and seven economic categories. The amount of real income, tax rate, fines, penalties, tax benefits, and the probability of tax audit, seven economic determinants of tax compliance were considered. Non-economic factors include the individual's view of the fairness of the tax system; how to deal with taxes; and national, social and individual norms. However, Palil and Mustafa (2011) argue that the legal system, ethics, and other contextual factors influence tax planning to some extent.

Financial exchange theory

According to financial exchange theory, governments can improve compliance by providing desired products to citizens in more effective and accessible ways (Cole and Gordon, 1988). Alam et al. (1992) showed that the perception of the availability of public goods and services is positively related to compliance. Therefore, taxpayers are primarily concerned with what they get in return for paying taxes in the form of public services. This view considers taxation and the provision of public goods and services as a kind of contract between the government and the people who pay taxes for it. People value the products and services provided by the government and are willing to pay taxes because they understand that their contributions are necessary to finance these products and services and to encourage others to do so. Consequently, positive payoffs can increase the likelihood of voluntary compliance without direct coercion. This means that taxpayers' behavior may depend on how satisfied or dissatisfied they are with the conditions of their relationship with the government. If taxpayers believe that the tax system is unfair, they may try to change their position with the government through tax evasion (Halhel and Ahmed, 2014).



Social influences, comparative performance theory

According to the social impact theory, the individual's behavior and attitude towards the tax system is influenced by social norms and the behavior of their reference group (Sanoli, 1990). Taxes, like other aspects of behavior, can affect human behavior. An individual's reference group, which includes friends, neighbors, and family, can influence compliance behavior and attitudes toward the tax system. Thus, knowing that some members of important groups are the ones to reduce taxes weakens the taxpayer's obligation to comply. Additionally, people with social connections may be deterred from committing fraud because they fear the consequences and exposure. In addition, Sa (1991) argues that social influences can affect the perceived likelihood of detection.

The adaptive behavior model suggests that better compliance can result from eliminating imbalances in the relationship between taxpayers and the government. As Feldelstad et al. (2012) state, perceptions of how the government treats them compared to others can significantly influence their views of both parties and the government. If a group receives preferential treatment by the government, this can affect both the group receiving benefits and the relationship of citizens with the government. Consequently, how the government treats an individual in relation to other members of its larger national community is as important as what the individual receives from the government. The way each taxpayer is treated and the relationship between the tax burden of others and compliance behavior are only two ways that show that the perception of the fairness of the tax system affects compliance decisions. According to Walsh (2012), if people believe that others pay taxes, they are more likely to do so.

Treating people equally under the same conditions is the most understandable need for justice (Jayawardane, 2015). Horizontal and vertical equity, also known as tax equity, are the two main elements of tax equity. Vertical equity indicates that taxpayers who perform better should provide the same percentage of their income as those who perform better, horizontal equity supports tax collection based on financial status (Sahu, 2021). Individuals with different income levels are characterized by vertical equity (Barjaya, 1987).

The theory of political legitimacy

The theory of political legitimacy shows that citizens' trust in the government affects their tax compliance (Kerkler et al., 2008). Political scientists have studied the processes leading to political legitimacy and civic identity. Legitimacy can be defined as the belief or trust that the government, institutions and social arrangements are appropriate and fair and serve the interests of the general public (Jeldstad et al., 2012). Trust is essential for taxpayers to willingly pay their taxes. Controlled expectations and trust in an uncertain environment relate to the relationship between government and power. Taxpayers with less trust in the government see events negatively, those with more trust see events positively. Taxpayers who do not trust the government are likely to be more skeptical about the use of tax revenue collected by the



government. Taxpayers' commitment to the tax system and tax payment increases when they trust the government (Jimens and Eyre, 2016).

Theory of planned behavior

The theory of planned behavior is an influential theory in social psychology that aims to explain people's behavior. The theory of rational action was proposed by Isk-Ajzen and Fishbein (1970) to explain conscious behavior. This idea postulates that certain elements, which begin for various reasons in terms of method and evolution, influence the behavior of people in society. However, a person's ability to perform a certain behavior depends on his motivation to do so. Three elements of mental norms, behavioral characteristics and cognitive-behavioral control define the purpose of a behavior (Bobek and Hatfield, 2003). According to Soad (2011), the theory of planned behavior shows that beliefs are a necessary prerequisite to control behavior as well as attitude towards behavior and mental standards. As stated by Matheson (1991), controller beliefs are the recognition of the acquisition of opportunities, resources, and capabilities, and the understanding of the importance of these resources in achieving goals. A person's ability to control his behavior is highly dependent on his skills, expertise and social support. One of the factors affecting tax compliance is the ability of taxpayers to understand tax laws and their willingness to follow them. Tax knowledge refers to the general level of tax knowledge, knowledge of avoidance options, general academic qualifications, or knowledge of tax law (Bornman and Ramutumba, 2019). Taxpayers' knowledge of their rights, obligations, and tax payment methods, as well as the consequences of non-compliance, is obtained through tax education. Taxpayer education can have a positive attitude towards tax compliance and provide the necessary knowledge to comply with tax laws.

Theory of economic deterrence

According to Allingham and Sandmo (1972), the economic deterrence model assumes that factors such as tax rates, tax evasion benefits, the probability of detection of fraud, and the severity of punishment can affect taxpayers' behavior. Because, rational decisions are made in the midst of uncertainty and tax evasion can lead to tax penalties. Consequently, the more likely it is that tax evasion will be detected and punished, the fewer people will participate in it. Conversely, when the probability of an audit is low and the penalty is low, the expected return from evasion is high. According to Halkhil and Ahmad (2014), this model predicts significant non-compliance. In certain circumstances, the fear of detection and arrest can act as an effective deterrent to encouraging honest behavior. For example, fear of being caught is known to be an effective strategy for eliciting honest behavior; Although there are criticisms that this model only considers the mandatory aspect of compliance. Influenced by concepts of economic deterrence, tax administrations have developed enforcement strategies that focus primarily on punishment and fear of detection, as well as the associated time and financial cost, including the burden of tax payment itself.



The economic costs of taxes include not only the actual tax payment and the additional burden related to it, but also the time and money spent on tax compliance and tax planning (Blafus et al., 2011). Legal and regulatory obligations from authorities and tax laws are called compliance costs. Neither actual tax payments nor any costs related to tax distortions are included in these costs (Aragbe and Modogo, 2014). Compliance costs are lost when taxes are collected. In addition to the costs incurred to acquire and maintain the knowledge required for the position, such as understanding legal responsibilities and penalties, it also includes the costs of collecting, disclosing, and filing taxes on the company's products and income, and the salaries and wages of its employees. Aragbe and Modogo, (2014) divided compliance costs into two categories: tax planning and accounting. To maintain an accurate accounting system, the costs associated with data collection and tax liability calculations are called computational costs. However, planning costs increase when taxpayers attempt to legally reduce or avoid taxes.

Many studies have been conducted regarding the behavior of taxpayers, some of the most important of which have been discussed below:

Among domestic studies, Barikani (1400) in his study entitled "Analysis of people's disobedience and acceptance of paying taxes" with a qualitative approach has come to the conclusion that several reasons such as discrimination in receiving taxes, the invisibility of the effects of tax revenues in the society There is a lack of transparency about where tax revenues are spent, for tax evasion. The existence of widespread tax evasion by some people reduces the motivation of others to pay taxes and causes the government to have a budget deficit from tax revenues and critical sectors such as the health and treatment sector, the education sector, the country's construction and development projects, etc. are not provided with resources. . In this situation, the country's economy will be weakened and dependence on other countries will be the result. Namazian et al. (1400) in a study entitled "Evaluation of the effect of the effective tax rate and the quality of corporate governance on tax evasion and tax corruption" using financial variables, at the same time the effect of the effective tax rate and the quality of corporate governance on tax evasion and corruption They investigated with multivariate regression method. Based on the results of this research, there is a positive and significant relationship between the effective tax rate and tax evasion and corruption; But with the addition of the variable of corporate governance quality to the above relationships, no significant change was observed on the relationship between the effective tax rate and tax evasion and corruption. Samati et al. (1400) in their study entitled "Determining the factors affecting tax evasion using the meta-analysis method" using the meta-analysis method, investigated the factors affecting tax evasion and also the severity of their relationship with tax evasion for the country of Iran. contracted The results of this research showed that the variables of economic factors, per capita income, technology factors, social factors, legal factors, inflation, cultural factors, trade restrictions, tax burden, unemployment and the size of the government are the factors affecting tax evasion and cultural factors are the most influential factors. And per capita income has the least impact on tax evasion. Gholami (1400) in his study entitled "Economic security considerations of tax arrears and evasion" suggests that in many countries, the government's



focus and reliance is on tax revenues to provide budget, but in countries that mainly get their budget from oil revenues. they bring, the focus on tax revenues and management of the provincial tax system is less. This problem also exists in Iran due to the existence of oil revenues and the control and management of the tax system is not optimal. Ignoring tax revenues as the main source of funding has led to the lack of proper monitoring of provincial taxes, the existence of extensive tax exemptions, the lack of recognition of tax authorities, and the creation of many tax arrears and the spread of tax evasion. Mohseni (1400) in his study titled "Effective factors on tax evasion using the structural equation technique - the moderating role of tax knowledge" examined the factors affecting tax evasion by considering the moderating role of tax knowledge. The statistical population of this study is taxpayers of various types of tax jobs during 2018 And it was the first half of 2019. 8 cases affecting tax evasion were investigated according to Fisher's modified 4-factor questionnaire (1992). The results of this research showed that the variables of tax rate and income level and detection probability, tax crimes, the complexity of the tax system and the understandable roles of the government have a positive and significant relationship with the variable of tax evasion. While the influence and attitude of peers has not been an influencing factor in tax evasion. Also, tax knowledge, as a moderating variable, has an effect on the relationship between significant factors and tax evasion. Ahmadi et al. (1400) in their study entitled "Evaluation of tax evasion measurement models from neoclassical economics to behavioral economics: Hierarchical analysis approach in Iranian economy", examined the evolution of different tax evasion models from neoclassical approach to behavioral economics and the evolution of the models. and discussed the challenges and experimental results obtained from these models and extracted and analyzed the assumptions of the variables affecting tax evasion in each of the models. Then, using the hierarchical analysis method, according to the opinions of experts in the tax field, they ranked the factors affecting tax evasion in different models. The results showed that the instability of preferences, loss aversion, and ambiguity aversion are the main factors affecting tax evasion. Also, the assumption of rational behavior and behavioral consistency have the lowest explanatory power of tax evasion in Iran.

Belah et al. (2023) in their study titled predicting tax fraud in Rwanda using supervised machine learning models. In this study, various models such as artificial neural networks, logistic regression, decision tree, etc. were evaluated to identify the most effective model for predicting tax fraud. The findings showed that artificial neural networks are the best model for predicting tax fraud and factors such as business start time, domestic businesses, owners who import and export goods, businesses that do not incur losses, and the characteristics of taxpayers' behavior should be considered in this method. take In his study, Nasser (2023) proposed a fraud detection framework that uses supervised and unsupervised models to exploit the entire set of tax returns. The framework consists of four modules: a supervised module, which uses a tree-based model to extract knowledge from data; An unsupervised module that calculates anomaly scores. A behavioral module, which determines the compliance score for each taxpayer. and a prediction module, which uses the output of previous modules to output



the probability of fraud for each tax return. In his study, Manuel (2022) has provided a framework for investigating tax fraud, which aims to promote the following: (1) a coherent and rational procedural approach to deal with tax fraud at the national and European Union levels. (ii) common methods and techniques for investigating and prosecuting tax fraud; (3) a harmonized standard for work in different agencies (multi-agency work) at national and EU levels; (4) the most appropriate practices to prevent and combat tax fraud by discouraging criminals from expanding into new fraudulent ventures; (5) saving time and resources during interjurisdictional legal analysis and thus increasing the initial investigation. (6) Effective pathways for information conversion and an effective response to transnational VAT fraud. (7) Information on the different types of evidence required to secure a conviction in selected EU member states.

3- Research method

In this research, data-mining method was used for data analysis. Data mining is a methodology for dealing with a large amount of data systematically, which uses various statistical and machine learning tools to discover hidden knowledge in the data. In this study, two widely used methods in data mining, namely classification and clustering, are discussed and investigated. Clustering is one of the methods in which no labels are considered for the records and the records are grouped into a set of clusters based on the similarity criteria that are introduced. The absence of labels makes any clustering algorithm an unsupervised algorithm. In unsupervised methods, the algorithm does not have steps under the names of training and evaluation, and at the end of the clustering operation, the built model (the same clusters created) along with its efficiency is presented as an output. One of the most important clustering techniques is the k-Means and X-Means method, which were examined in this research and their results were compared.

Also, classification is one of the methods in which for each of the records of the data set to be explored, there is a label that represents the truth in the problem. This tag makes each classification algorithm to be considered a supervised algorithm. Decision tree is one of the most famous methods of building a classification model. In decision tree-based classification algorithms, the output knowledge is presented as a tree of different states of feature values. In line with the purpose of this research, the information related to the taxpayers of the country has been used. The total number of observations is 9,994 cases for the year 1402.

4-Results and discussion

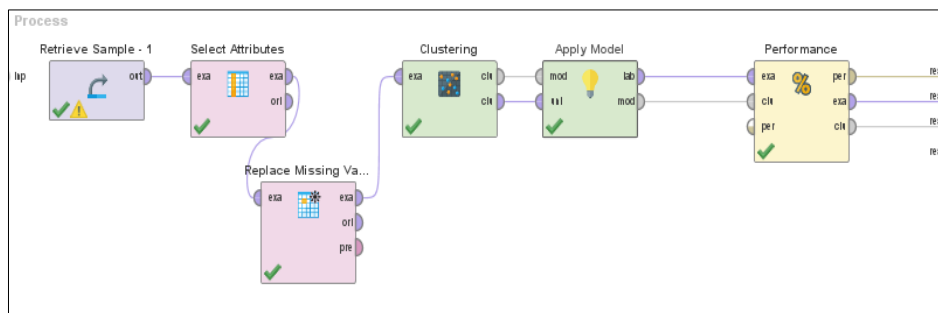
Tax payers' behavior clustering through X-Means and K-Means algorithms

In this section, RapidMiner version 9.9 software is used to cluster the behavior of taxpayers through X-Means and K-Means algorithms. In this regard, the information of 9994 taxpayers in several business categories has been used. In order to evaluate the clusters obtained by the clustering algorithms and find the optimal clustering, the Davis Boldin (DB) index was used.



which is based on the maximum distance between clusters and the minimum distance within clusters (Ahmadi and Ahadzadeh, 2015). The first step in this method is choosing the optimal number of clusters. In this regard, firstly, on the input data, clustering techniques with different number of clusters are implemented and finally, the number of clusters with appropriate Davis-Bouldin criterion is specified. In this regard, the evaluation of clusters with the mentioned values of K was done in two clustering methods, and the method with the lowest DB value was selected as the appropriate clustering method.

Figure 1- K-Means clustering operator in RapidMiner software



Source: present research

As emphasized, Davis-Bouldin index was used to select the appropriate number of clusters, the results of which are presented for each value of k in Table (1). The lower the values of the mentioned evaluation indices, the more suitable the cluster separation criteria and the better intra-cluster continuity (Ahmadi and Ahadzadeh, 2015), based on the obtained results, in the K-means method, the number of optimal clusters is seven and in the X-method The average number of optimal clusters is three clusters identified.

Table 1- DB index values in K-MEANS and X-MEANS methods

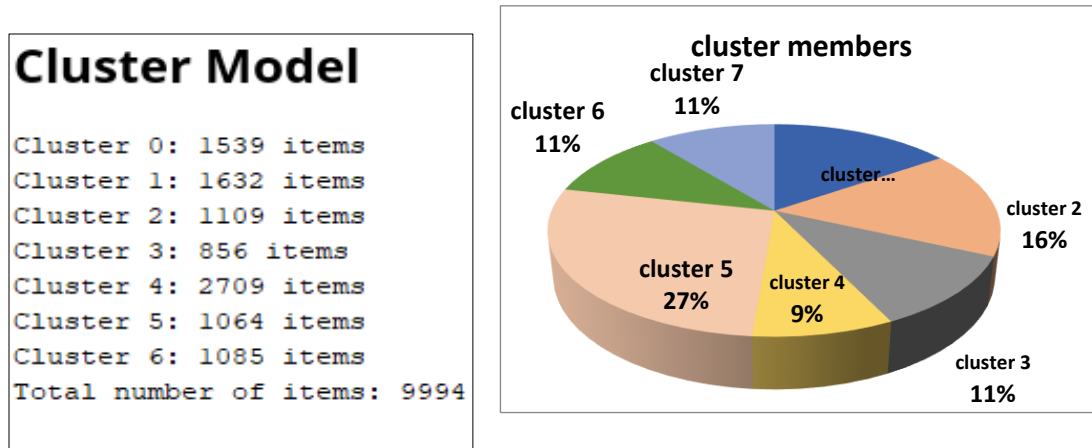
X-MEANS	K-MEANS	Number of clusters
0/459	0/08	1
0/478	0/092	2
0/402	0/096	3
0/469	0/094	4
0/553	0/086	5
0/509	0/087	6
0/466	0/076	7
0/532	0/091	8

Source: present research

Also, in order to know the members of each cluster, the percentage of the members in the two mentioned methods is shown in graphs (2) and (3):



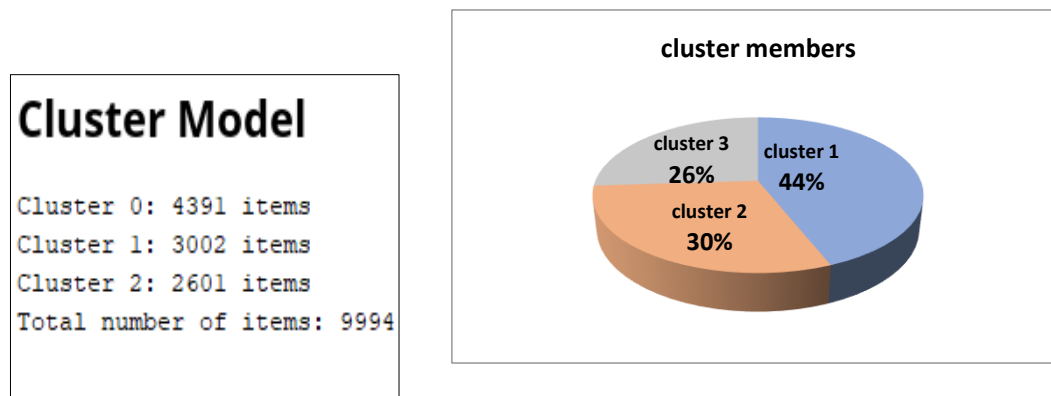
Chart 2- The percentage of members of each cluster in the K-Means method



Source: Research calculations

As can be seen from the graph (1), the most members in the K-means clustering method are related to the fifth cluster, which accounts for 27% of the total number of observations. In the fifth cluster, based on the Ka-Mean method, the share of declared tax from the assessment (0.2961), the share of definitive tax from the assessment (0.4727), the share of declarations from the number of taxpayers (0.6527), the share of zero declarations from the total (25) 57) and fixed tax (286). The businesses in the fifth cluster include ironmonger guild; land, air, sea transportation; car shows; It is a class of halls and reception halls and other businesses. The rest of the clusters are listed in table (2). The members of each cluster in the X-mean clustering method are also shown in diagram (2).

Chart 3- The percentage of members of each cluster in the X-Means method



Source: Research calculations

As can be seen from the graph (3), most members belong to the first cluster (44%). In this cluster, mainly the businesses of money changers; Guild of cloth bankers; bags and shoes, bags



and suitcases; notary offices; audio and video equipment; There are sweets, dried fruit and ice cream. In the following, considering the identification of optimal clusters, the results of clustering by the two mentioned methods will be examined. Based on this, Table (2) shows the distribution of customers in K-Means clustering algorithm clusters with K=7.

Table 2- Clustering results with K-Means technique

Cluster 7	Cluster 6	Cluster 5	Cluster 4	Cluster 3	Cluster 2	Cluster 1	Feature
0/01059	0/08789	0/296189	0/3908	0/54806	0/65426	1/1369	Share tax expressed from diagnosis
0/2112	0/34701	0/4727	0/59349	0/61883	0/66511	1/11355	Definitive tax share of the diagnosis
0/786	0/6099	0/6527	0/6697	69/38	0/7171	0/74108	Declaration share of the number of taxpayers
44	54	57/25	61/66	64	69/4	83/666	The contribution of the declaration is zero from the total
107	196	286	375	543	854	2566	Fixed tax

Source: Research calculations

To provide a better understanding of clustering based on the K-Means criterion in this section, the results related to the share of definitive tax from diagnostic tax are presented in Table (3).

Table 3- Share of fixed tax of businesses from diagnostic tax

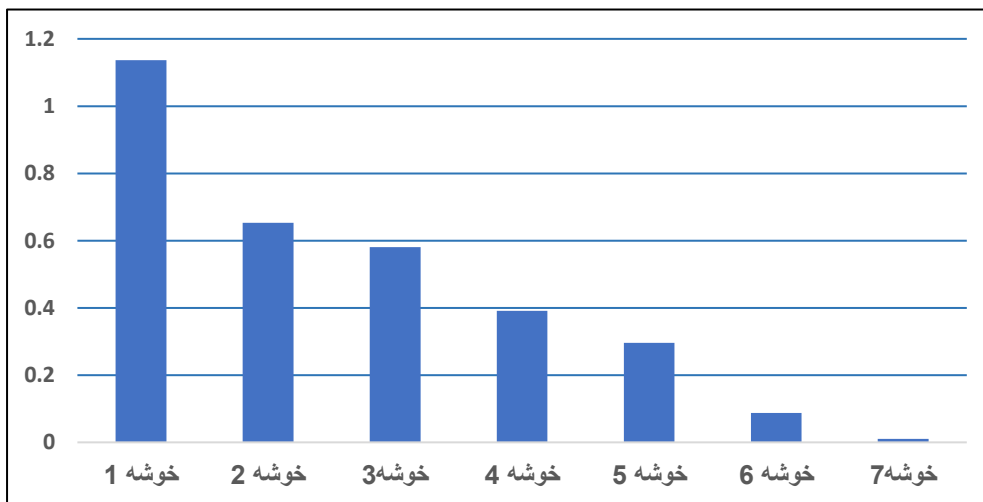
Business	contribution	cluster
money changers; Guild of cloth bankers; bags and shoes, bags and suitcases; notary offices; audio and video equipment; Sweets, nuts and ice cream	1/11355	one
vehicles and spare parts; dentists; Cell phone and landline phone sellers guild; Association of restaurants and food sellers	0/66511	two
household appliances; office machines and computers; doctors guild; Food bankers guild	0/61883	three



Pharmacies and sellers of medical and health supplies, etc.; jewelers, watchmakers; Chemical, petroleum and medical industries	0/59349	four
ironmongers guild; land, air, sea transportation; car shows; Halls and reception halls and..	0/4727	five
real estate brokers and consultants; Lawyers, marriage and divorce offices, judicial experts and...	0/34701	six
Laboratories, radiology, physiotherapy and...	0/2112	Seven

Also, the results related to the share of definitive tax from diagnostic tax are presented in graph (4).

Chart 4 - The share of tax expression from the diagnosis based on the K-Means criterion



Source: Research calculations

Based on the presented results, the examination of the clusters shows that in cluster (7), the amount of tax contribution is considered lower than other clusters. In this business cluster, it is related to laboratories, radiology, physiotherapy, etc. On the other hand, the highest share of tax expressed from diagnosis also belongs to cluster (1). In this business cluster of money changers; Guild of cloth bankers; bags and shoes, bags and suitcases; notary offices; audio and video equipment; There are sweets, dried fruit and ice cream. In justification of this result, it can be said that businesses in group one have more tax compliance and group seven have lower tax compliance. The table below shows the distribution of customers in the clusters of the K-Means clustering algorithm with K=3 and the coverage percentage of each cluster of the total number of taxpayers.



Table 4- Clustering results with K-Means technique

Cluster 3	Cluster 2	Cluster 1	Feature
0/125803	0/322615	0/67015	Share tax expressed from diagnosis
0/472967	0/63722	0/910976	Definitive tax share of the diagnosis
0/544316	0/64837	0/700398	Declaration share of the number of taxpayers
55/333	63/25	71/166	The contribution of the declaration is zero from the total
298/944	723/166	1995/833	Fixed tax

Source: Research calculations

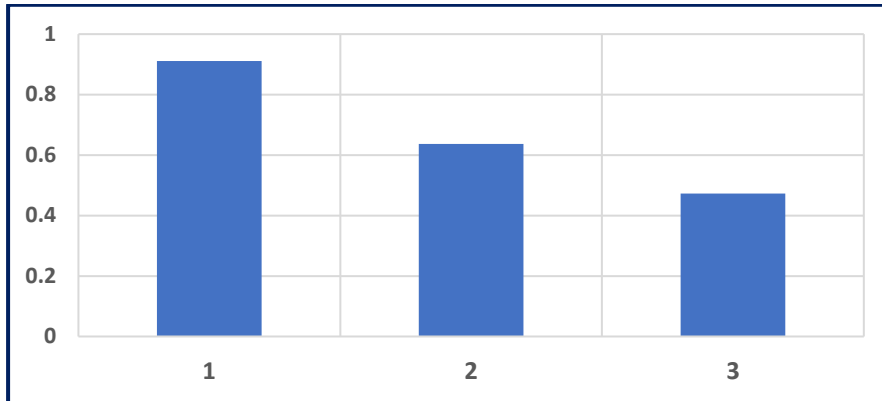
Based on the results of the table, taxpayers were classified into three clusters, and the largest share of the examined indicators is in cluster one. The classification of businesses based on clusters is presented in table (5). As it shows, most of the businesses in cluster one include money changers; Guild of cloth bankers; bags and shoes, bags and suitcases; notary offices; audio and video equipment; Sweets, nuts and ice cream; vehicles and spare parts; Dentists.

Table 5- Share of fixed tax of businesses from diagnostic tax

Business	Coefficient	cluster
money changers; Guild of cloth bankers; bags and shoes, bags and suitcases; notary offices; audio and video equipment; Sweets, nuts and ice cream; vehicles and spare parts; Dentists	0/910976	one
Cell phone and landline phone sellers guild; Association of restaurants and food vendors; household appliances; office machines and computers; doctors guild; Food bankers guild; Pharmacies and sellers of medical and health supplies, etc.; jewelers, watchmakers; Chemical, petroleum and medical industries	0/63722	two
ironmongers guild; land, air, sea transportation; car shows; Halls and reception halls; real estate brokers and consultants; Lawyers, marriage and divorce offices, judicial experts	0/472967	three

The chart below shows the share of the final tax from diagnosis in three classified clusters:

Diagram 5- Contribution of definitive tax from diagnosis based on clustering with K-Means technique

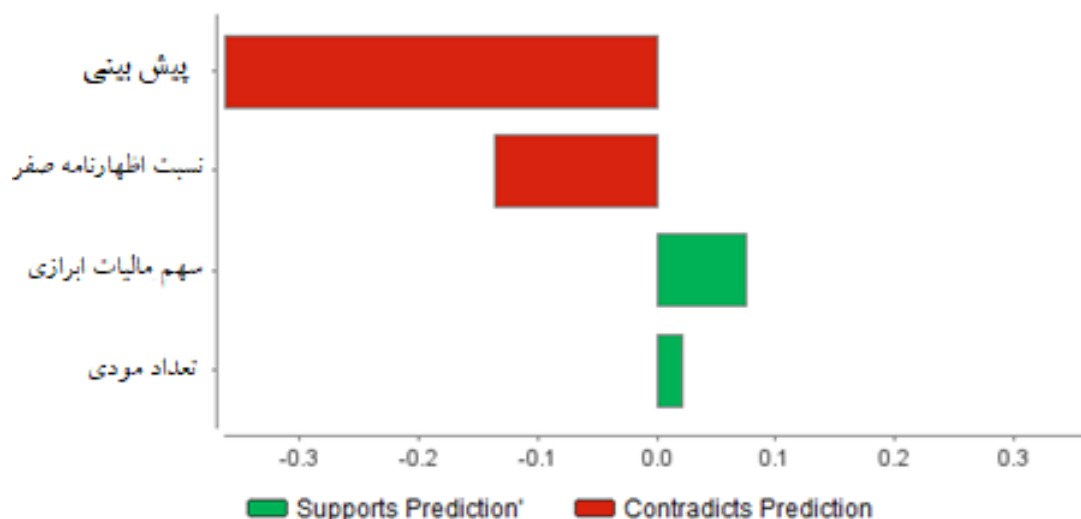


The results of clustering based on the X-Means method also confirm the previous results of the research based on the Ka-Means approach. Based on these results, the lowest share of definitive tax is related to the third cluster.

Deterministic tax prediction from decision tree classification method

In the following, the decision tree classification method is used to predict the final tax. In this approach, the definitive tax variable is considered as a dependent variable. Based on the results obtained in the prediction of the final tax, the highest weight is assigned to the zero declaration ratio variable, the declared tax share and the number of taxpayers, respectively. Chart (6) represents the important factors for predicting the final tax under consideration.

Chart 6- Factors affecting the prediction of definitive tax

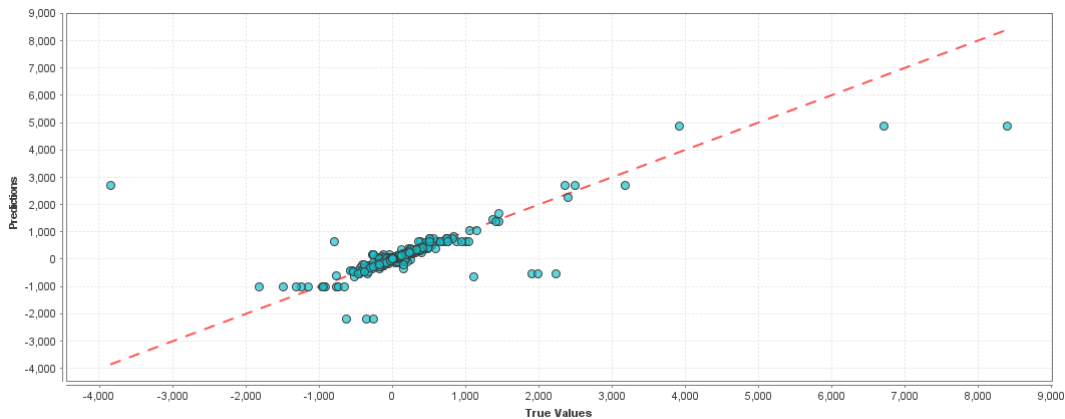


Source: Research calculations



The definitive tax prediction chart is also presented as follows:

Diagram 7- Decision tree approach-prediction diagram



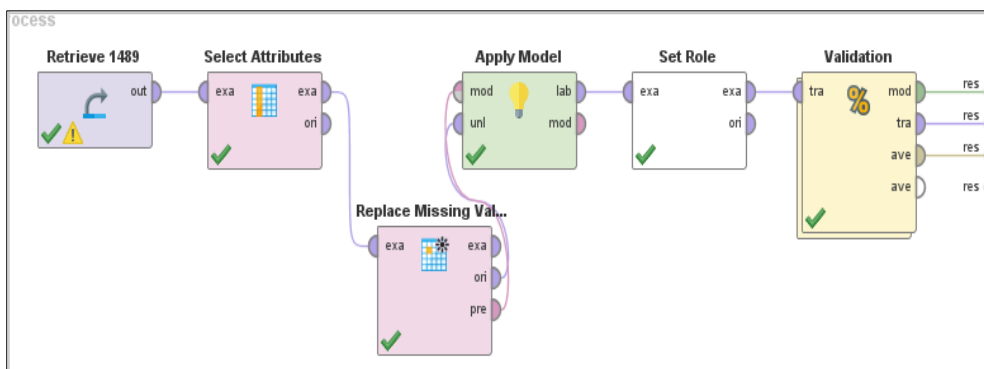
Source: Research calculations

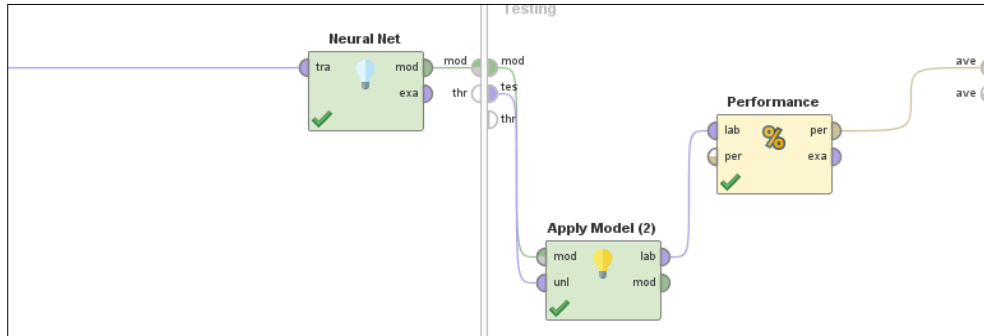
According to the performance results, the decision tree method is a suitable method for predicting the definitive tax due to the accuracy of its measurement in this research (67/66).

Examining the behavior of taxpayers with the approach of artificial neural network

Neural networks are made up of a set of non-linear nodes that are connected in parallel. The weights of this connection with their changes help the neural networks to observe the appropriate training to solve a specific problem. Normally, the nodes that exist in a column are called a layer. Therefore, every neural network has at least two layers (one input and one output), although in most of these networks, hidden layers are used between the input and output layers to increase the accuracy. It is also important that this method, like the decision tree method, is not easily interpretable and does not provide a set of rules to the user. Figure (2) shows the artificial neural network operator in the Rapidminer software:

Figure 2- Artificial neural network operator in RapidMiner software

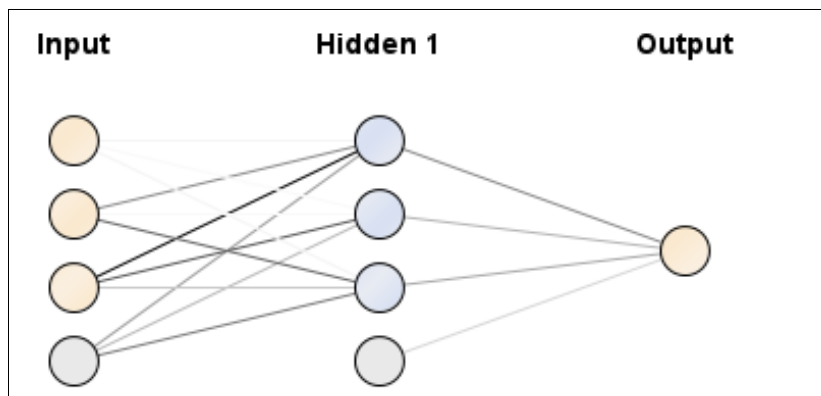




Source: research findings

The artificial neural network in this research includes the inputs of the number of taxpayers, the ratio of zero declarations and declared taxes, and the output of definitive taxes. This network is presented with a hidden layer in Figure (3):

Figure 3- Artificial neural network topology



Source: research findings

As the results in the figure show, this network has a middle layer which contains 3 nodes. The weights of these nodes are presented in table (5):

Table 6- Weights related to the middle layer of the artificial neural network

Nood 3	Nood 2	Nood 1	Ninety entries
0/184	0/082	0/152	Number of taxpayers
2/76	0/081	2/072	Zero declaration ratio
-0/951	-2/996	-4/517	Express tax

Source: research findings



According to the results of the research, in the tax payers' behavior section, using the information of the tax affairs organization, the accuracy of the artificial neural network has been obtained at 97%, so the more efficient approach to clustering is 66.67%.

5- conclusion

In this section, RapidMiner version 9.9 software is used to cluster the behavior of taxpayers through X-Means and K-Means algorithms. In this regard, the information of 9994 taxpayers in several business categories was used. Davis Boldin (DB) index was used to evaluate the clusters obtained by clustering algorithms and to find the optimal clustering. In this regard, the evaluation of clusters with the mentioned values of K was done in two clustering methods, and the method that has the lowest DB value has been selected as the appropriate clustering method. Based on the obtained results, in the Ka-mean method, the number of optimal clusters is seven, and in the X-mean method, the number of optimal clusters is three. Based on the results presented in the Ka-Mean method, the examination of the clusters shows that in cluster (7), the amount of tax contribution expressed is considered lower than other clusters. In this business cluster, there were laboratories, radiology, physiotherapy, etc. On the other hand, the highest share of tax expressed from diagnosis also belongs to cluster (1). In this business cluster of money changers; Guild of cloth bankers; bags and shoes, bags and suitcases; notary offices; audio and video equipment; There are sweets, dried fruit and ice cream. In other words, businesses in group one have more tax compliance and group seven have lower tax compliance. Based on the results of the X-mean table, taxpayers were classified into three clusters, and the largest share of the examined indicators is in cluster one. The results show that most businesses in cluster one include money changers; Guild of cloth bankers; bags and shoes, bags and suitcases; notary offices; audio and video equipment; Sweets, nuts and ice cream; vehicles and spare parts; Dentists. In the following, the decision tree classification method is used to predict the final tax. In this approach, the definitive tax variable is considered as a dependent variable. Based on the results obtained in the prediction of definitive tax, the highest weight is allocated to the variable of zero declaration ratio, declared tax share and finally the number of taxpayers respectively. Also, based on the results of the artificial neural network research, in the tax payers' behavior section, using information from the Tax Administration, the accuracy of the artificial neural network has been obtained at 97%, so the more efficient approach to clustering is 66.67%. In this regard, according to the obtained results, it is suggested that the country's tax commissioner should develop strategies for tax compliance and tax evasion prevention in accordance with each business. For example, the strategies related to tax compliance of the seventh or third cluster (laboratories, radiology, physiotherapy, etc.) should be used from the following strategies:

A- Necessary measures to facilitate compliance: providing information through various communication channels regarding the taxpayer's legal duties, submitting declarations, etc.; posting notices regarding taxpayers' duties at the workplace; The requirement to use a sealed card reader by the organization and the name of the centers; Taxpayer training



B- Necessary measures to deal with non-compliance: matching the information received from the taxpayers of this branch with the information of pharmacies, insurances; Applying regulations related to money laundering for persons who rent out their bank accounts; Use of special investigators; Using pre-filled declarations for this class; Estimating the taxable income of this sector with scientific techniques; Establishing a specialized tax unit for laboratories, radiology, and physical therapy services and requiring the guild activists to submit quarterly declarations.

References

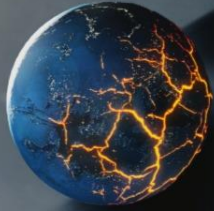
1. Ahmadi Vahidah, Shahnazi Rooh Elah, Islamluiyan Karim, Sadraei Javaheri Ahmad. (1400). Evaluation of tax evasion measurement models from neoclassical economics to behavioral economics: Hierarchical analysis approach in Iranian economy. *Planning and Budgeting Quarterly* 1400; 26 (1): 115-141
2. Barikani, Ali, (1400), analysis of non-compliance and acceptance of tax payment by the people, <https://civilica.com/doc/1341300>
3. -Samti, Majid, Izadi, Afsana, Fathi, Saeed, (1400), determining factors affecting tax evasion using the meta-analysis method, *stable economy*, summer 1400, second period - number 3 (1 to 22)
4. Gholami, Atefeh, (1400), economic security considerations of arrears and tax evasion, *economic security period*, July 9, 1400, number 4 (87 series)
5. -Mohsani Maleki, Bahram, (1400), factors affecting tax evasion using the structural equation technique - the moderating role of tax knowledge, *Accounting and Auditing Surveys*, 28 Fall 1400, Number 3
6. Namazian, Ali, Zainli, Hadith, Pourheidari, Omid, (1400), study of the effect of the effective tax rate and the quality of corporate governance on tax evasion and tax corruption, *accounting and auditing reviews of the fall of 1400*, 28th period - number 3, rank A. Ministry of Science/ISC (26 pages - from 507 to 532)
7. -Abuamria, F. (2019). The Effect of Deterrence Factors on Discourage Shadow Economy Level and Tax Evasion. *International Journal of Academic Research in Accounting, Finance and Management Sciences*, 9(1), 62-67.
8. -Ademe, H, & Simret, D (2020) Determinants of Tax Compliance Behavior in Ethiopia: Evidence from South Gondar Zone
9. -Ajzen I, Fishbein M (1970) The prediction of behavior from attitudinal and normative variables. *J J Exp Soc Psychol* 6(4):466-487
10. -Alkhatib, A, A., & Abdul-Jabbar, H. (2017). An Economic Framework for Tax Evasion in -Palestine. *American-Eurasian Journal of Scientific Research*, 12(6), 300-304.
11. -Alm, J., & Torgler, B. (2011). Do ethics matter? Tax compliance and morality. *Journal of Business Ethics*, 101(4), 635-651.



12. -Alshrouf, M. (2019). The effect of tax audit using the computer on tax non-compliance in Palestine. *International Journal of Academic Research in Business and Social Sciences*, 9(3), 296–304.
13. -Allingham MG, Sandmo A (1972) Income tax evasion: a theoretical analysis
14. -Assfaw, AM, & Sebhat, W (2019). Analysis of tax compliance and its determinants: evidence from Kaffa, bench Maji and Sheka zones category B tax payers, SNNPR, Ethiopia
15. -Asha, R.B.; Suresh Kumar, K.R. Credit card fraud detection using Artificial Neural Networks. *Glob. Transitions Proc.* 2021, 2, 35–41.
16. -Barbuta-Misu, N (2011) A review of factors for tax compliance
17. -Barjoyai, B (1987) Taxation: Principle and Practice in Malaysia (Pencukaaian Prinsip dan Amalan di Malaysia. J Kuala Lumpur, Dewan Bahasa dan Pustaka
18. Belle, Fille, Murorunkwere., Dominique, Haughton., Joseph, Nzabanita., Francis, Kipkogei., Ignace, Habimana, Kabano. (2023). Predicting tax fraud using supervised machine learning approach. *African Journal of Science, Technology, Innovation and Development*, 1-12. doi: 10.1080/20421338.2023.2187930
19. -Benk, S., McGee, R. W., & Yüzbaşı, B. (2015). How religions affect attitudes toward ethics of tax evasion? A comparative and demographic analysis. *Journal for the Study of Religions and Ideologies*, 14(41), 202–223.
20. -Bobek DD, Hatfield RC (2003) An investigation of the theory of planned behavior and the role of moral obligation in tax compliance. *Behav Res Account* 15(1):13–38
21. -Bornman M, Ramutumbu PJMAR (2019) A conceptual framework of tax knowledge
22. -Bonchi, F.; Giannotti, F.; Mainetto, G.; Pedreschi, D. Using data mining techniques in fiscal fraud detection. In *Proceedings of the International Conference on DataWarehousing and Knowledge Discovery, Berlin/Heidelberg, Germany, 30 August 1999*; pp. 369–376
23. -Blaufus K, Eichfelder S, Hundsdoerfer J (2011) The hidden burden of the income tax: compliance costs of German individuals, of the income tax: Compliance costs of German individuals, *Diskussionsbeiträge*, No. 2011/6, Freie Universität Berlin, Fachbereich Wirtschaftswissenschaft, Berlin
24. -Cowell FA, Gordon J (1988) Unwillingness to pay: tax evasion and public good provision. *J Public Econ* 36(3):305–321
25. -Dalu T, Maposa VG, Pabwaungana S, Dalu T (2012) The impact of tax evasion and avoidance on the economy: a case of Harare, Zimbabwe. *African. J Econ Sustain Dev* 1(3):284–296
26. -de Roux, D.; Perez, B.; Moreno, A.; Villamil, M.D.P.; Figueroa, C. Tax fraud detection for under-reporting declarations using an unsupervised machine learning approach. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, London, UK, 19–23 August 2018*; pp. 215–222.
27. -Dias, A.; Pinto, C.; Batista, J.; Neves, E. Signaling tax evasion, financial ratios and cluster analysis. *BIS Q. Rev.* 2016.



28. -Deyganto KO (2018b) Factors influencing taxpayers' voluntary compliance attitude with tax system: evidence from Gedeo zone of Southern Ethiopia. *Univers J Account Financ* 6(3):92–107
29. -Engida TG, Baisa GA (2014) Factors influencing taxpayers' compliance with the tax system: an empirical study in mekelle city, Ethiopia. *eJTR* 12:433
30. -Eragbhe E, Modugu KP (2014) Tax compliance costs of small and medium scale enterprises in Nigeria. *Int J Account Tax* 2(1):63–87
31. -Feria, R. *Tax Fraud and the Rule of Law*; Oxford University Centre for Business Taxation: Oxford, UK, 2018.
32. -Fjeldstad O-H, Schulz-Herzenberg C, Hoem Sjursten I (2012) People's views of taxation in Africa: a review of research on determinants of tax compliance
33. -Ghosh, S.; Douglas, L.R. Credit card fraud detection with a neural-network. In *Proceedings of the Twenty-Seventh Hawaii International Conference*, Wailea, HI, USA, 4–7 January 1994
34. -González, P.C.; Velásquez, J.D. Characterization and detection of taxpayers with false invoices using data mining techniques. *Expert Syst. Appl.* 2013, 40, 1427–1436. [CrossRef]
35. -Helhel Y, Ahmed Y (2014) Factors affecting tax attitudes and tax compliance: a survey study in Yemen. *J Eur J Bus Manag* 6(22):48–58
36. -Jackson, B. R., & Milliron, V. C. (1986). Tax compliance research: Findings, problems, and prospects. *Journal of Accounting Literature*, 5(1), 125–165.
37. -Jayawardane D (2015) Psychological factors affect tax compliance. A review paper. *Int J Arts Commer* 4(6):131–141
38. -Jemberie DB (2020) Determinants of tax compliance: a case of nekemte town category 'C' Business profit tax payers. *Int J Econ Financ Manag Sci* 8(3):89
39. -Jenkins, GP, & Forlemu, EN (1993). Enhancing voluntary compliance by reducing compliance costs: a taxpayer service approach: Harvard Institute for International Development. Harvard University
40. -Jimenez P, Iyer GS (2016b) Tax compliance in a social setting: the influence of social norms, trust in government, and perceived fairness on taxpayer compliance. *Adv Account* 34:17–26
41. -Kefela D, Ghirmai T (2009) Reforming tax polices and revenue mobilization promotes a fiscal responsibility: a study of east and West African states. *J Law Confl Resolut* 1(5):98–106
42. -Khlif, H., & Achek, I. (2015). The determinants of tax evasion: A literature review. *International Journal of Law and Management*, 57(5), 486–497.
43. -Kirchler, E., Hoelzl, E., & Wahl, I. (2008). Enforced versus voluntary tax compliance: The "slippery slope" framework. *Journal of Economic Psychology*, 29(2), 210–225.
44. -Kültür, Y.; Çağlayan, M.U. Tax fraud and the rule of law. *Expert Syst.* 2017, 34, 12191. [CrossRef]



45. -Lois P, Drogalas G, Karagiorgos A, Chlorou A (2019) Tax compliance during fiscal depression periods: the case of Greece. *EuroMed J Bus* 14(3):274–291
46. -Manuel, Monge., Carlos, Poza., Sofia, Borgia. (2022). A proposal of a suspicion of tax fraud indicator based on google trends to foresee Spanish tax revenues. *International Economics*, doi: 10.1016/J.INTECO.2021.11.002
47. -Mathieson K (1991) Predicting user intentions: comparing the technology acceptance model with the theory of planned behavior. *Inf Syst Res* 2(3):173–191
48. -Mebratu AA, Leyku F, Meroune L (2020) Determinant of tax revenue effort in sub-Saharan African countries: a stochastic frontier analysis. *Int J Sustain De World Policy* 9(1):47–71
49. -Mubarek, A.M.; E_sref, A. CMultilayer perceptron neural network technique for fraud detection. In *Proceedings of the S2017 International Conference on Computer Science and Engineering (UBMK), Antalya, Turkey, 5–8 October 2017*.
50. -Murorunkwere, B.F.; Tuyishimire, O.; Haughton, D.; Nzabanita, J. Fraud etection Using Neural Networks: A Case Study of Income Tax. *Future Internet* 2022, 14, 168. <https://doi.org/10.3390/fi14060168>
51. Nasser, Alsadhan. (2023). A Multi-Module Machine Learning Approach to Detect Tax Fraud. *Computer Systems: Science & Engineering*, doi: 10.32604/csse.2023.033375
52. -Okpeyo ET, Musah A, Gakpetor ED (2019a) Determinants of tax compliance in Ghana. *J Appl Account Tax* 4(1):1–14
53. -Oladipo O, Ogunjobi J, Eluyela D (2022) The impact of ethical tax behaviour on tax compliance of tax authority and corporate taxpayers of listed manufacturing companies in Nigeria. *Academia Letters*, Article 4698. <https://doi.org/10.20935/AL4698>
54. -Qubbaja, A., & Omar, S. A. (2019). The impact of political events on Palestine securities exchange returns: An empirical study between (1997-2016). *International Journal of Academic Research in Accounting, Finance and Management Sciences*, 9(3), 287–294.
55. -Saad N (2011) Fairness perceptions and compliance behavior: taxpayers’ judgments in self-assessment environments. *Procedia - Social and Behavioral Sciences* Vol 109, pp 1069–1075. <https://doi.org/10.1016/j.sbspro.2013.12.590>
56. -Sah RK (1991) Social osmosis and patterns of crime. *J Polit Econ* 99(6):1272–1295
57. -Sahu P (2021) Fairness dimension of goods and services tax: evidence from Indian MSMEs. *Int J Account, Bus Financ* 1(1):40–47
58. -Shallo L (2018) Determinants of tax compliance in ethiopia: case study in revenue and customs authority, Hawassa Branch. *Res J Financ Account* 9(21):95–104
59. Smelser, N.J; Baltes, P.B. *International Encyclopedia of the Social & Behavioral Sciences*, 11th ed.; Elsevier: Amsterdam, The Netherlands, 2011.
60. -Snavelly K (1990) Governmental policies to reduce tax evasion: coerced behavior versus services and values development. *Policy Sci* 23:57–72. <https://doi.org/10.1007/BF00136992>



61. -Pérez López, C.; Delgado Rodríguez, M.; de Lucas Santos, S. Tax fraud detection through neural networks: An application using a sample of personal income taxpayers. *Future Internet* 2019, 11, 86. [CrossRef]
62. -Rahhal, A. I. (2017). Effectiveness of Palestinian income tax rates in facing tax evasion. *Global Journal of Engineering Science and Research Management*, 4(1), 8–14.
63. -Richardson, M., & Sawyer, A. (2001). A taxonomy of the tax compliance literature: Further findings, problems and prospects. *Australian Tax Forum*, 16(2), 137-320.
64. -Riahi-Belkaoui, A. (2004). Relationship between tax compliance internationally and selected determinants of tax morale. *Journal of International Accounting, Auditing and Taxation*, 13(2), 135–143.
65. -Snively K (1990) Governmental policies to reduce tax evasion: coerced behavior ersus services and values development. *Policy Sci* 23:57–72. <https://doi.org/10.1007/BF00136992>
66. -Wu, R.S.; Ou, C.S.; Lin, H.; Chang, S.I.; Yen, D.C. Using data mining technique to enhance tax evasion detection performance. *Expert Syst. Appl.* 2012, 10, 8769–8777. [CrossRef]
67. -Walsh K (2012) Understanding taxpayer behaviour–new opportunities for tax administration. *Econ Soc Rev* 43(3):451–475