PRIZREN SOCIAL SCIENCE JOURNAL
VOLUME: 8
ISSUE: 3
SEPTEMBER- DECEMBER, 2024

ISSN: 2616-387X
DOI: 10.32936/PSSJ

www.prizrenjournal.com
info@prizrenjournal.com

PROACTIVE APPROACH IN TAX RISK MANAGEMENT: DATA ANALYSIS TECHNIQUES TO IDENTIFY HIGH-RISK TAXPAYERS

SCAN ME

Funda KARAKOYUN

Dokuz Eylul University, İzmir Vocational School, Department of Accounting and Taxation, Turkey, funda.karakoyun@deu.edu.tr

Article history:

Submission 5 September 2024
Revision 19 November 2024
Accepted 20 December 2024
Available online 31 December 2024

Keywords:

Tax Risk Management,
Data Mining,
Machine Learning,
Risk Analysis Techniques,
Tax Audit,
Proactive Approach.

DOI:

https://doi.org/10.32936/pssj.v8i3.593

Abstract

There is a transition from labour-intensive traditional methods to the use of technology-intensive systems to identify risky taxpayers in tax audits. With this process, which has been accelerating for decades, statistical models that will reduce tax loss and evasion of states and increase efficiency in collection have been diversified in the field of machine learning. While innovations in software programs with digitalisation transformation reduce the manual workload of the tax administration, machine learning algorithms are used with experts employed in the field in continuously developing risk analysis studies. For tax administrations that fall behind these developments, the identification of high-risk taxpayers, the effectiveness of collection, and tax compliance are becoming more difficult to audit in the context of rapidly increasing global data.

The influence of a country's distinctive historical, economic, and sociological characteristics is a significant determinant of taxpayer behaviour. Establishing relationships across taxpayer characteristics and classifying behaviours in the tax risk management process affects the success of tax scenarios. Although, the work of experts in traditional methods and data analysis in auditing is limited, the effectiveness of data mining methods and the combination of the results obtained with tax scenarios approach the desired accuracy rate in risk detection. For this reason, it is observed that different methods are tested in experimental studies since a single technique that audits all business records cannot be sufficient and reliable. This study elucidates the deployment of information technology and digital tools in tax risk management, with consideration of the nuances inherent to the diverse national contexts.

1. Introduction

Tax evasion is recognised as a universal phenomenon. Technology-based risk analyses are seen as a necessity in the processing of big data for the detection of tax frauds, which are the erosion of the fundamental financial resources of the state, with an early warning system. Considering both time saving and personnel workload and expenses, different rule-based machine learning methods make a strong contribution to the solution of this problem. For the detection of high-risk taxpayers, models are being developed which evaluate both non-relationship based financial data and relationship based behavioural characteristics of the business. Classifier algorithms are trained and tested to increase further the accuracy of the model.

Technological developments have gained a different dimension with Industry 4.0. In almost every field of economic and social life, digital transformation and digital technologies and tools have been implemented. The application area of this process in auditing in the public sector and tax administrations is developing rapidly. Instead of reactive methods based on manual labour and reporting limited to certain rules, audit methods that obtain warning signals with a proactive approach are used. Each sector develops its systems in the fields of computer-based software, utilisation of networking features, Machine Learning, artificial intelligence, Internet of Things (IoT), in line with its own needs. The public sector also increases its investments in this direction in the global competition.

In this study, digital methods such as anomaly or outlier detection, risky taxpayer valuation and data mining that increase efficiency in risk management are discussed. In tax risk management, the models developed by various country practices are evaluated in terms of accurate detection of risky taxpayers in the evaluation of the relationships between the attributes in big data, and the preferred models for ensuring audit efficiency are explained.

1.1 Digital Transformation in Tax Administrations

Digital transformation is the management of efficiency and risks by applying digital competencies to processes, products and assets, and thereby utilising opportunities to create new resources (Eyüpgiller, 2023, p. 71). New business models, products and services are developed in the use of information technology with the widespread use of the Internet, especially the Internet of Things (IoT), the development of Machine Learning, Big Data, Cloud Computing, Artificial Intelligence tools like “chatbot”, and the participation of robots with artificial intelligence in the workforce. This situation leads to digital transformation with systems where new processes requiring digital knowledge and skills emerge (Biyan, 2022, p.3329). In addition, the simultaneous use of digital products by multiple users has led to an increase in the need for data storage. This situation has resulted in the necessity for the creation of flexible, accessible, and easy-to-manage information services such as "Cloud Computing" technology, which is defined as a technology for storing assets such as data, information, documents, software, and applications owned by organisations or individuals in a virtual warehouse on the internet (Köse, and Taner, 2024, p.182). Since everything can be digitally connected to Cloud Computing with the digital transformation, businesses and individuals are provided with faster and easier access to computing resources through advantages such as scalability, flexibility, and low-cost infrastructure (Gölçek, 2023, pp.547-548; Bulutistan, 2024A).

Transition to electronic system in taxation during the preparation process for digital transformation; realisation of many applications such as payments, tax return audits, notifications, minutes, reporting transactions, book and document ready declaration system, electronic attendance system in electronic environment, has constituted the important steps of Internet Tax Office applications, and E-finance applications. Thus, the system enables taxpayers to store a large amount of data on all tax-related debt, lien, accrual, and payment information in the tax administration. (Rüzgar, 2022, pp.5-6). Data analyses of a large number of stored data become possible with digital tools. It might be suggested that digitalisation is being achieved through the operations performed, processes followed, and systems established.

Another type of Artificial Intelligence that has the potential to transform in accounting and auditing is the automation of intelligent processes, called robotic processes. This process automation is supported by a range of technologies, including Machine Learning, Natural Language Processing, Structured Data Interaction, Intelligent Document Processing, Computer Vision, and Cognitive Automation (Köse, and Taner, 2024, p.180). With the development of information technology tools, tax administrations apply automation activities to almost every field and carry its development to a new dimension with digitalisation. The widespread use of the internet has initiated a new era in which digitalisation is reflected in the activities of tax administrations with applications such as Computer-Based Programming, Robotic Process Automation, Block Chain Technology and Artificial Intelligence (Durmuş, and Erdem, 2023, p.227). In Parallel, Software and hardware, which are digitalisation tools, are rapidly developed in product diversity through the internet (Biyan, 2022, p.3328).

In the presence of very large data packages, accessing accurate and information-generating data sources becomes more important in terms of the principle of tax security. In this regard, a different tax administration model suitable for increasing economic digitalization was named as “Tax Administration 3.0” at the OECD Tax Administration Forum (FTA) in 2019. Accordingly, the following elements are listed as necessary for the digitalisation of tax administration (Yayman, 2021, pp.2783,2784).

- A single information technology (IT) database should be created,
- Tax authorities should be trained by considering previous methods and experiences,
- The transfer and sustainability of financial resources should be ensured,
- The tax administration should be accurate, consistent, and professional in its communication, announcements, explanations, and guidance, and should receive and evaluate feedback from the parties in this context.
- Tax risk management uses information technology programmes that receive data directly; real-time compliance transaction data is transmitted to the tax administrations in real time without the need for separate submission to the tax administrations.

The following benefits are expected from digital transformation in the tax system (Eyüpgiller, 2023, p.72).

- Digitalisation combites the informal economy and keeps it under control.
- Digitalisation increases the efficiency and productivity of the administration. In public administration, as in all businesses, effectiveness and efficiency have come to the fore.
- Digitalisation is seen as improving tax compliance of taxpayers, reducing the costs of taxpayers, and thus reducing their resistance to tax.

In Turkey, within the scope of the Digital Tax Office, models are being developed that are standards-based, predictable, prevent unfair competition and encourage voluntary compliance (Revenue Administration, Online). The Digital Tax Assistant (DTA) called GIBI, developed by the Revenue Administration (GIB) and powered by Artificial Intelligence-supported Machine Learning, provides answers to the frequently asked questions of all citizens regarding the legislation. The single administration model is defined in issues such as the timely payment of public debts, the use of the *"mutual agreement method"* in solving tax problems involving more than one country in the double taxation avoidance agreement, the popularisation of advance pricing agreements in transfer pricing, etc. **(**Tax Inspection Board, 2023; Rüzgar, 2022, p.2-5).

2. Proactive Approach in Tax Audit

Tax losses, as a common problem of developed and developing countries, have multicomponent and layered negative consequences, such as hindering governments' revenue collection optimisation and making the legal and financial system insecure. In a study on the magnitude of the loss, "The Association of Certified Fraud Examiners" (ACFE) analysed 1921 case studies from 138 countries, including Turkey, and found that $3.1 billion is "lost" each year worldwide due to internal fraud and corruption in 2023 (ACFE, 2024). According to the study, each company loses 5 per cent of its average revenue in this way. The most common corruption in companies is asset misappropriation with 86 per cent. Misappropriation of assets can be defined as *"reducing business assets"* and/or *"preventing asset growth"*. This definition can be exemplified as follows:

-Withdrawal and concealment of available cash from the business,

- Collecting receivables but concealing them,

-Presenting as if a payment has been made even though no payment has been made,

-Costing of personal expenses to the enterprise,

-Embezzlement and concealment of economic assets.

According to the report, the sectors with the highest asset losses are listed as real estate, trade, transport and storage, construction, and service sectors (Şahin, 2024).

In order to minimise these losses, new tools, practices and policies are necessary for countries to make their own risk assessments and to formulate strategies on identifying potential areas of abuse, and to combat crimes. Thus, taxpayer typologies have been created with the understanding developed by the behavioural economic theory. While risks can be taken into account with rational behaviour as an influencing factor in tax audits, tax non-compliance can be demonstrated with intentional or unintentional errors and omissions. There are studies indicating that automated systems such as electronic declaration method and audits increase taxpayer compliance compared to traditional methods. It is expected to give a positive impetus to the compliance process and effective analysis in the analyses made with the "risk map" extracted according to the taxpayer typology (Çiçek, et al. 2019, pp.231-234).

Reactive approach, also known as traditional methods, is expressed as taking measures after transactions and events have occurred. In our country, applications such as e-invoice, e-archive invoice and BA-BS forms are used to provide data based on cross audit technique (Kara, and Karan, 2023, p.79). In the audit, based on the proactive approach, analytical tools and audit methods have been developed on the basis of taking measures to prevent these frauds from occurring by identifying the frauds that may occur before the fraud-related indicators that allow preventive measures (Ertikin, 2017, p. 78).

In the proactive approach, data mining, Artificial Neural Networks, Machine Learnings, whispering method, data analysis and Benford's Law are widely used (Kara, and Karan, 2023, p.78). According to Benford's Law, since the vast majority of data are the amounts that occur in real life, they have numerical distributions in accordance with the expected systematics. Human behaviours, on the other hand, may not conform to this systematic. When there are amounts in the accounting data that do not arise from real transactions, completely random or made up by people, a deviation from the expected distribution occurs in the numerical distribution of the data in general. Therefore, Benford's Law can be utilised to reveal any errors or fraud in the data. During the analysis of the probability distributions of the figures, auditors may need a criterion to compare the observed distributions of the figures. Benford's Law provides auditors with this criterion by giving the expected distributions of figures as a mathematical rule. Benford's Law is a technique that directs the auditor to the right points by revealing items and transactions that are likely to contain fraud and error rather than being used as a stand-alone audit tool (Rasgen, 2016. p.70). As an analytical procedure, analysing numerical distributions is suitable for making comparisons between different variables, like previous

years, periods, conditions, collection sales invoices, etc. Therefore, these analyses are reported to risk managers with the help of computer-aided audit programs. For example, when US income taxpayers report their income, expenditures, and savings to the tax administration with the same return, the data pools automatically update themselves every period and become suitable for cross-checking. By using the cross-checking technique, the accuracy of their declarations can be mutually checked (Uğur, 2016, pp.122 - 145).

Some of the proactive methods used in detecting fraud are statistical methods, financial analysis techniques, computer-aided audit techniques (CAAT's), Artificial Intelligence (AI) techniques, fuzzy logic and expert systems digital analysis methods (Benford's Law) (Yıldırım 2014, p.96). In proactive methods, unlike traditional audit methods, there is an audit process with the expectation that there may always be a fraudulent transaction without any allegation or denunciation of fraud. In this method, the audit is carried out over the entire database of the enterprises, rather than the sample, and the information technologies used provide the opportunity to analyse the database without time costs (Sağlar and Uluç 2022, p.89).

3. Data Analysis Techniques in Risky Taxpayer Identification

Tax is the last item in an entity's income statement before net income. The loss arising from tax risk differentiates the net income to be declared by the taxpayer. Risks may occur in areas such as incorrect records, illegal transactions or controversial areas arising from legislation (Söğütlüoğlu, 2024). Since risk management is a continuous process involving risk identification, analysis, assessment, prioritisation, solution seeking and model evaluation, risk analysis is related to "reasoning" and "decision making" skills**.** Within the scope of risk analysis, applications such as converting data into information, planning tax audits, proposing solutions, assigning risk scores to taxpayers in terms of determining tax evasion or grouping taxpayers, determining, and analysing the relationship networks between taxpayers can be mentioned (Ruzgas, 2023, p.290; Durmuş, and Erdem, 2023, p.245).

Traditional methods are manual, report-based, rule-based methods that depend on the labour force of finance and tax experts, where risk points in the data are identified. Over time, data mining directs tax audit through continuous learning of historical data. In "Tax Risk Detection", input data are examined in two classifications: contextual and behavioural attributes. While contextual attributes classify taxpayers based on data diversity such as record information, financial statements, business information (number, numbering, text information), in behavioural attributes, statistical information, upward and downward information are taken into account. "Tax Risk Detection" methods based on Artificial Intelligence and data mining are handled in two categories based on relationship and non-relationship methods, the details of which are given in Table 1. In non-relationship based methods, contextual features such as registration information, financial data and business information are used without considering the interaction between taxpayers. However, in tax scenarios, there are various types of relationships between different entity objects. Therefore, relationship-based risk assessment methods have emerged to take advantage of behavioural characteristics. Relationship-based data mining case selection methods are developed to utilise complex relationships in tax networks to identify tax risks by taking advantage of developing technology. In tax risk identification, risks are identified by classifying relationship-based and non-relationship-based features based on the interaction between taxpayers. Here, taxpayers' personal and business characteristics related to taxation, tax paid, documents and declaration information are selected, and classifiers based on these characteristics are trained, for example, support vector machines (SVM). In this way, relationship-based methods using both contextual attributes and behavioural attributes are obtained. With these methods, rich semantic information can be extracted from complex structural data more easily and more deeply, thus improving the accuracy of risk detection (Zheng, et al., 2024, pp.46-49).

The conceptual model of tax liability risk management is based on the principles in the "*Organisation for Economic Co-operation and Development*" (OECD) guidelines. In this model, stages are followed in determining the tendency to avoid liabilities. The tax system is considered as a whole for the effectiveness of monitoring and control measures. In taxpayer grouping, taxpayers are divided into groups or classes according to the types of taxes administered, taxpayer characteristics and behaviour, and taxation risk factors. In addition to large, medium and small scale, the scope of activity, number of employees, and performance criteria (whether or not to carry out economic activities) are taken into account in legal entities (Ruzgas, 2023, p.292; Adamov, 2019). Although the areas, where tax risk is seen, vary according to the type of taxpayer, region, and the effectiveness of tax fraud determinants in the country conditions, the following issues are generally accepted (Güneş 2024);

- Transfer pricing and gain transfer through disguised capital.
- Different opinions on the application of exemptions and exceptions.
- Open transactions in sales and wages.
- Increasing costs through misleading documents and expenses not related to the work.

- Structuring and invoicing between related companies for value added tax (VAT) and special consumption tax (SCT) planning purposes.
- Failure to pay the stamp tax that should be calculated on the contracts.
- Shifts in income and expense accruals.
- Errors in cost, depreciation, and valuation.
- Inaccuracies in withholding tax applications.
- Errors in tax rates, tax base calculations, deductions, and additions.

Table 1. Some Methods Used in Tax Risk Detection

Tax Risk Assessment	**Non-relationship Based**	**Shallow Model**	**Association Rule**
			Decision Tree-Based
			Support Vector Machine
			Bayesian Classifier
			Logistic Regression
			Clustering Model
			Hybrid Model
		Deep Model	**Artificial Neural Networks**
			Hybrid Model
		Reinforced Learning	**Value Based**
		Other	**Evolutionary Algorithm**
			Agent Based
	Relationship Based	**Graph Pattern Matching**	
		Graphical representation Learning	
		Visual analysis	

Source: Zheng et al., 2024, p.49

The elements considered when separating tax-related data according to their nature are as follows (Ruzgas, 2023, p.293)

- Type of economic activity,
- Type and amounts of tax to be paid (income tax, VAT, SCT, etc.),
- Taxable person characteristics (different tax rates or taxation rules for corporations, small businesses, social enterprises, free economic zones, etc.)

In grouping according to the level of compliance, taxpayers are classified according to their common characteristics, the amount of tax they pay and their behaviour. In grouping, the following levels are taken into account (Ruzgas, 2023, p.293);

- Those who honestly fulfil their obligations,
- Those who try to fulfil their obligations fairly but do not always succeed,
- Those who tend not to comply until their obligations are addressed,
- Those who avoid fulfilling obligations in bad faith.

Tax harmonisation risks based on OECD and European Union (EU) practices are as follows (Ruzgas, 2023, p.295);

- Taxpayer registration risk,
- Risk of incorrect tax declaration,
- Risk arising from non-declaration,
- Risk of non-payment of taxes.

Data-driven risk identification analysis can be used through a detailed analysis of taxpayers' returns, specific enquiries with taxpayers and, in some cases, specific and/or comprehensive audits of selected cases. For example, k-mean clustering of a large dataset can identify the optimal number of clusters and observations that are significantly distant from a cluster, however, distance from a cluster, itself is not a compliance risk. Instead, data needs to be analysed to identify the underlying tax compliance risk, for example, it may be that the distance represents trade deduction claims for a particular industry and income segmentation. In this case, the underlying compliance risk is the over-claiming of business deductions to reduce tax (Brondolo 2022, p.42)

3.1. Data Mining

The increasing use of digital platforms and smart devices has led to a rapid increase in data volume. In parallel with this increase, storage capacity, computing power and algorithms have gradually been improved. These developments offer potential opportunities to help processing data more efficiently and make better decisions. Big data is at the forefront of these technologies (Gölçek 2023, p.547). However, big data does not make sense on its own. They gain a meaningful dimension from inferences based on certain patterns in the data. There are data mining and Machine

Learning techniques that enable the identification of patterns, suitable for the research to be conducted and the type of data. With these techniques, rules, patterns and relationships buried inside data are found (Ersöz, and Çınar, 2021, pp. 397-414). In other words, a data warehouse is created by processing and manipulating raw data. The data warehouse is a repository for historical data, which is stored, optimised, and subjected to complex data modelling and interactive analysis. These processes render the data warehouse an effective tool for decision support and strategic planning. Data maps consisting of data subsets improve the decision-making process. The data transformed into information are used in accordance with the needs by technology combinations (Bulutistan, 2024B).

Data mining can be briefly defined as a decision support analysis application in the literature. In fact, it is a tool that supports the decision process to be made to reach a solution in the problem by providing the information necessary to solve the problem (Terzi, 2012, p.54). In other words, data mining is a multidisciplinary field that combines computer science and statistics to extract information from a data set and transform it into an understandable form for future use. This technique is widely used in various fields including banking, finance, health, communication, medicine and engineering. Thus, data mining includes various operations such as association, clustering, prediction and classification (Şencan, 2023, p.15).

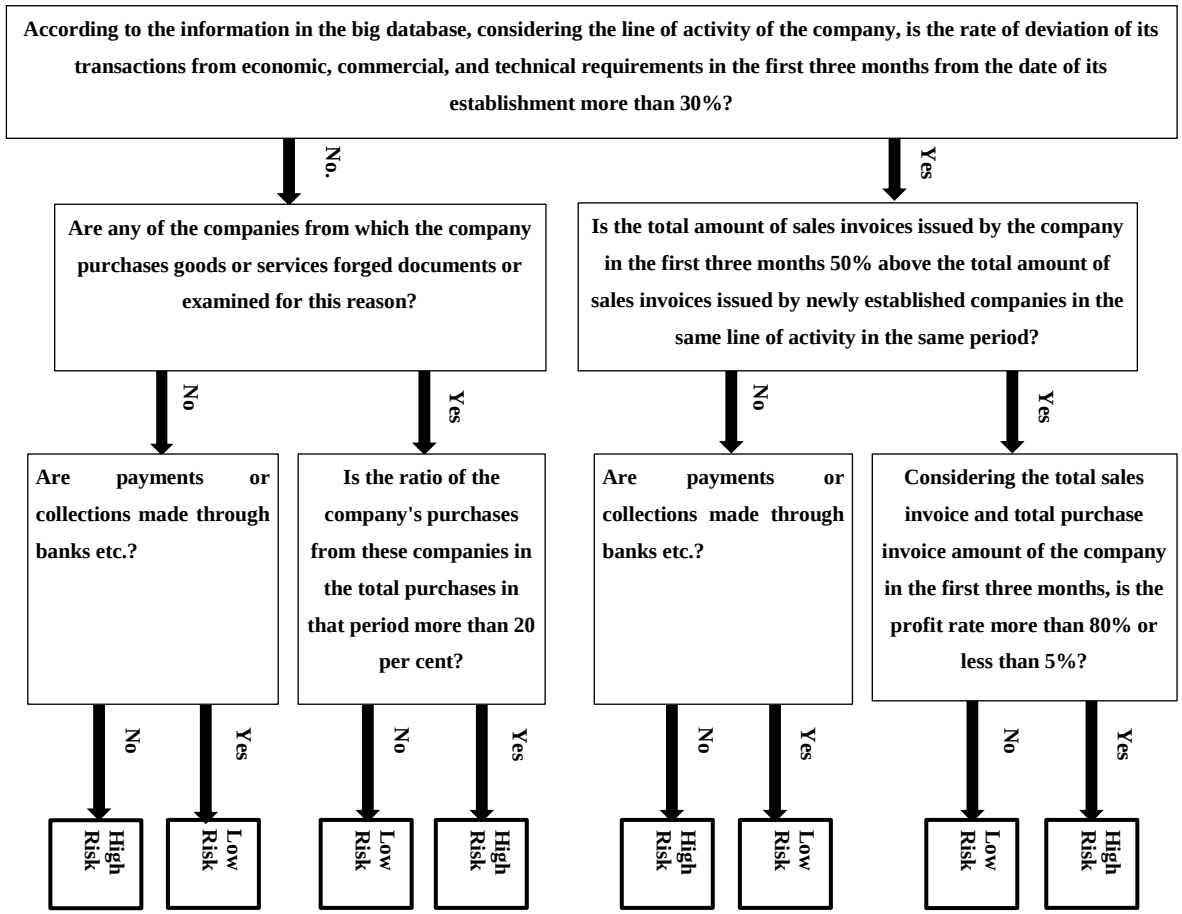
In data mining, predictive and descriptive models are created. In predictive models, a model is developed based on data with known results (Çelik, 2009, p.7.). For example, a predictive model is developed from the data obtained from tax evaders and their characteristics. This model, the results of which are known, is used to predict tax evasion in new taxpayers joining the system. The big data to be evaluated for tax administration can be used differently according to the defined problem. All documents, declarations and book records kept in electronic environment, which are obligatory to be kept or notified by taxpayers, are an important part of the tax audit process. In addition, databases are cross-linked with information collected from transactions such as property records, bank transactions, loans, vehicle licences, insurance policies, title deed data from other institutions and organisations (İlgün, 2020, p.14). Here, the role of "Cloud Computing" is to enable cross-linking of tax records with other databases and data mining in social network as a technology that provides access to networks, servers, or applications at any time with minimal effort without the intervention of service providers (Güneş, and Arslan, 2023, p.48).

A common method of using data mining in tax audit risk management is the Machine Learning. With this method, it is aimed to obtain meaningful inferences from the available big data. Since tax evasion and unfair VAT refunds through the issuance of false documents are frequently encountered in practice, an example of tax scenario using this method for the identification of high-risk taxpayers who issue forged documents is shown in Figure 1. In the scenario, a newly established company is tested whether it carries out an activity in accordance with economic, commercial and technical requirements. For the preparation, firstly, certain questions were asked, and the information contained in the big data was classified and interpreted. In this process, while determining the works and transactions that are contrary to the normal flow of commercial life and are not continuous, the method of using economic value, the wastage and yield rates, and the works done without payment were tested on the average of other companies in the sector, assuming that the first three months of the company's notifications, statements and records have accumulated and matured enough to be able to make meaningful inferences. According to the answers obtained, the goods and services purchased or sold by the taxpayer, the status of invoices, payment and collection methods, profitability level, and ratio analyses were examined to determine the situation about the company and the risk level of the taxpayer was concluded. In this way, risky taxpayers are subjected to more detailed testing (Güneş, and Arslan, 2023, p.54-55).

In order to detect the forged document issuer, the nodes in the sample Machine Learning method given above can be diversified as amount of paid-in capital, number of employees, erroneous payment detection, asset information of the taxpayer or institution, rounded invoice amounts, issuance of invoices with consecutive numbers to the same person or institution, collection or payment on public holidays, payment by cheque and collection rate control (Güneş and Arslan, 2023, p.54).

The most important feature of data mining is that multiple methods can be compared. The correct classification success may vary according to the data set, the selected features, and the type of problem. Accordingly, as a result of the comparison, the method with the highest accuracy rate is selected and the optimisation in the next step is obtained (Kırda, and Özçelik, 2021, p.630). If a good result is to be obtained from data mining, it is more useful to convert the data into numerical values for the linear regression model and categorical values for the Machine Learning algorithm (Gür, 2023, p.522).

Figure 1. Proposed Machine Learning application for detecting the taxpayer issuing forged documents.



Source: Güneş ve Arslan, 2023, s.54

3.2. Machine Learning

Storing large and ever-increasing amounts of data with a new technology instead of traditional data storage methods, using algorithms instead of traditional data analysis tools, and enabling computers to make predictions for the future by giving them the ability to think like humans (Artificial Intelligence, AI) are referred to as Machine Learning. Machine Learning algorithms operate by creating a model based on examples known as "training data" to make predictions or decisions without explicit programming. Owing to the information technologies that connect data with applications such as the Internet of Things (IoT) and Cloud Computing that collect data from different channels, decision makers use the collected data according to the needs (Durmuş, and Erdem, 2023, p.239; Demirci, 2018, pp.9-10). Therefore, data mining combines statistical science, Machine Learning and Artificial Intelligence to discover big data. They are decision support applications that optimise accurate predictions for the future by linking past data to the present. In this case, tax administrations should correctly construct algorithms to serve the desired purpose from big data. Many integrations in the data warehouse make warehouse operations more efficient with Artificial Intelligence technology. When data mining software and new algorithms are examined, all algorithms that may be necessary in Machine Learnings, Artificial Neural Networks, association, regression, time series analyses are included in both Statistical Package for the Social Sciences (SPSS), and Statistical Analysis Software (SAS) (Çelik, 2009, p.15).

Machine Learning is the automatic extraction of information from various types of data using computational and statistical methods in almost every field of public and private sectors with Artificial Intelligence and data science technologies. For example, data mining and Machine Learning techniques are used in the government, especially in analysing sensor data, customer prediction analyses, financial services in risk identification, fraud detection, etc. (Ersöz, and Çınar, 2021, p.398).

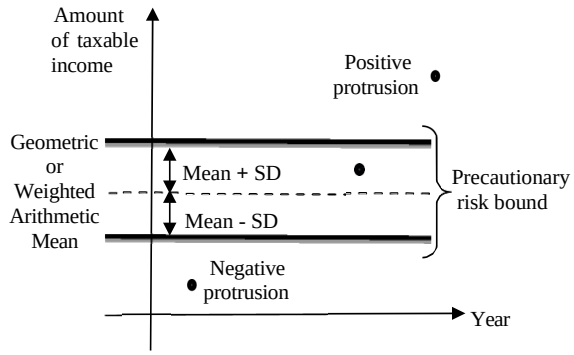
Machine Learning with data mining methods is divided into two main categories: supervised and unsupervised. Supervised expression is used when there is a well-defined or precise target. In supervised learning, according to the criteria determined, various examples are given, and the class formed by expressing with rule sentences is separated. When the accuracy test of the

separated test set is deemed sufficient, the model is applied to the new examples. Therefore, complex data models associated with successful case results are selected, while unsuccessful ones are eliminated. Thus, it is used to extract information and conclusions from data, for example, a Machine Learning. If there is no specific definition for the desired result or if there is uncertainty, the term unsupervised is used. In this method, the classes are not predetermined, and the classification technique is applied according to the characteristics of the data in the data set. The algorithm does not have a training dataset to influence its development. Therefore, it is more difficult to evaluate the performance of models based on unsupervised approaches. Unsupervised approaches are often used in exploratory analyses such as clustering. This method is mostly used to understand, recognise and explore the data (Şencan, 2023, p.15; Brondolo, 2022, p.42). For the analysis to predict high-risk taxpayers, the system based on Machine Learning that enables the classification of financial data and taxpayers is trained and tested. Determining the rules of classification varies according to the model tools. Different Machine Learning algorithms are compared in the applied techniques. When the accuracy rate of the results reaches the desired success, the model tool with the best performance is determined **(**Zheng, et al., 2024, p.52). In experimental studies conducted in different countries, the performance of the model is evaluated from the findings of the effect of the number of variables, balanced/unbalanced distribution of data, and analysis of risky taxpayer behaviours. The system based on the training of classifiers using Machine Learning can be combined with classification methods to provide high accuracy in the detection of taxpayer non-compliance. For the use of data analysis techniques by tax administrations, various models have been used in the detection of tax evasion acts with transactions between related persons (Durmuş, and Erdem, 2023, p.240-241).

Among the non-relationship based shallow models given in Table 1, association rule is a rule-based Machine Learning method, which is one of the most important techniques in the field of data mining. It is used to discover correlations and patterns between certain attributes in big data (Zheng, et al., 2024, p.50). In relationship-based data mining case selection methods, there is usually a correlation between the features that reflect the tax risk, and then, it is used in tax scenarios.

In a study on the Logistic Regression example, the number of high-risk taxpayers determined by regression algorithms based on support vector machine in the data received from the Iran Tax Office between 2010-2020 was determined. Here, there is a relationship between the standard deviation (SD) and the distance of the annual final taxable income amount to the average. The precautionary bound is calculated by addition and subtraction of SD from geometric or weighted mean and the model is formed. In the model, suspicious financial behaviour is identified as outliers laying outside the precautionary risk boundary, as depicted in Figure 2. In addition, high-risk occupations can also be classified by using occupational coefficients.

Figure 2: Calculation of the mean and standard deviation (SD) of probable taxpayer revenues.



Source: Rad, and Shahbahrami, 2016, p.6

Machine Learning technology has the potential to increase the effectiveness and efficiency of financial auditing with its features such as analysing large amounts of data, quickly detecting inconsistencies and anomalies, warning about potential risks, reducing error rates and extracting meaningful information from complex data sets, and improving the performance of algorithms by constantly feeding them with new data (Köse, and Taner, 2024, p.203).

Support Vector Machine (SVM), one of the effective machine-based audit mechanisms, is used to classify taxpayers if taxpayers change the data for tax evasion. In classification analysis of data mining, patterns are obtained from existing data by classification methods, and these are used to classify unseen units (Kırda, and Özçelik, 2021, p.611).

Real-time data acquisition and analysis can also enable a shift from a reactive approach to a proactive approach in auditing by identifying situations where tax evasion is likely to occur and comparing these parameters with existing taxpayer data. Therefore, a new dimension is seen in the preventive audit of undeclared events resulting in tax loss. First of all, within the scope of taxpayer services aimed at preventing illegal activities, assistance services that provide information and guidance etc. are offered to the system (İlgün, 2020, p.9).

With data mining, it is possible to separate abnormal data hidden in big data and detect fraud by inductive (e.g., Benford's Law) and deductive methods (Terzi, 2012, p.54). For example, when significant deviations from Benford's Law regarding trade receivables accounts are identified, the auditor (Machine

Learning) can increase the level of verification, perform additional transaction scanning, and further examine the documents supporting the recorded credit sales (Sağlar, and Uluçan, 2022, p.90). In fraud detection, a method called "Red Flags" is used to identify suspicious points. Symptoms and indicators that enable fraud to be recognised in advance are considered to be warning signs of fraudulent transactions, although they are not conclusive evidence of fraud (Terzi, 2012, p.56). As the “red flags” are listed in detail in the literature, the entire database can be analysed by using information technologies when the criteria considered as fraud indicators in the application of the method are encountered. For an effective query in the execution of the analysis, both people who can perform fraud audits and people who are experts in information technologies are needed (Çalış, et al., 2014, p.100). With the data analysis techniques (social network analysis, transaction network analysis, etc.) developed in the data mining analysis software, tax evasion acts and networks where fraudulent commercial transactions take place can be detected through related persons. In various country examples, this technique is frequently used in the detection of tax fraud in value added tax (VAT) (Durmuş, N.K and Erdem, 2023, p.240).

Boosting is a method for cumulatively training the predictors. AdaBoost is the most frequently used model. In this algorithm, sequencing is important because a sequential computation is a cumulative structure. Here, the training data is taught to the machine and thus it is called the Machine Learning technique. When it is desired to make inferences for new incoming samples, the test samples are reduced to the correct decision regions by using the alpha values and results of the decision trees, and hence, the label information is predicted (Güzel, 2020).

One of the important problems in practice is that the psychological and social effects of tax regulations cannot be fully calculated. At this point, big data analytics can be used to evaluate the behavioural effects of tax policies by processing and interpreting data collected from different sources (İlgün, 2020, p.11). These algorithms can analyse large datasets, identify anomalies and detect irregularities in tables with high accuracy. Machine Learning algorithms, which learn from datasets and whose performance can be improved with more data over time, offer auditors the opportunity to recognise risks in advance and to manage them effectively by using big data analysis to identify potential risks in financial reports and business processes. This capability not only increases the efficiency of the audit process, but also improves the detection of potential fraud and forgery activities. Namely, Machine Learning models can make the audit process more robust with the ability to recognise patterns that indicate financial fraud (Köse, and Taner, 2024, pp.215-216).

In Turkey, technical analyses were applied for taxpayers issuing forged documents. Information such as invoice flow, partnership relations, and accounting relations among and between other taxpayers from which these taxpayers obtain invoices were identified through the Tax Inspection Board -Risk Analysis System (VDK-RAS) (Tax Inspection Board (VDK), 2023, pp.29-30). In particular, the Tax Audit Analysis System (VEDAS) is one of the important steps towards digital/electronic auditing for the Turkish Tax Inspection Board (VDK). Based on the taxpayers' declarations, electronic domestic audit software that can read e-books and e-invoices has been created, which measures the risk of possible errors, fraud, etc... With this analysis, the entire financial status of the enterprise, the level of compliance with tax legislation, the validity of e-documents and records in technical terms can be audited (Konar, 2023, p.7; Tax Inspection Board (VDK), 2019, pp. 36-37).

4. Risk Analysis Determination Methods in Some Country Practices

Mark Nigrini's (1992) study on "Analysing Income Tax Losses with Digital Distribution Method" is one of the first studies in the literature on how the frauds in the accounting data of 7 different companies were revealed using digital analysis and the application of computer technologies in fraud auditing (Çalış, et al., 2014, p.94). Algorithm methods, which have been developed and diversified today, and the places where they are applied are given in Table 2.

The prominent feature in Table 2 is that all of them used Machine Learning, except for regression, which is generally used in applications that can be considered as specialised. Because in Machine Learning, the change of the relations between "input" and "output" according to the applied model is monitored and then subsequently applied further to the similar situations. Thus, the hybrid models with the highest detection scores are formed by combining Random Forest and Artificial Neural Networks or the methods using Artificial Intelligence. In this method briefly, classifiers are first trained, and then experimental studies and tests are conducted to determine which model is the most accurate classifier. In addition to reducing the manual workload in assessing tax risk with the most accurate model, the model is developed and updated for other factors affecting the risk. When evaluating the choice of optimal control and monitoring tools, frequent violations by the taxpayer, tax risks arising from specific taxpayer groups or sectors, and their behavioural modelling are taken into account (Ruzgas, et al., 2023, p.11).

Table 2. Tax Risk Analysis Application Methods in Country Examples

Method	Researchers /Countries	Results
Decision Tree, Neural Network, Bayesian Network	Gonzales and Velasquez, 2013 /Chile	Clustering algorithms were used to group taxpayer behaviours. Decision trees, Neural Networks and Bayesian networks were used to identify tax evasion and related behaviour patterns, and to identify variables. Using multilayer perception Neural Network models in small- and large-scale enterprises, fraud cases were detected by 92% and 89%, respectively.
K-Nearest Neighbours, Random Forest, SVM, Neural Network	Andrade et al., 2021 /Brazil	Company financial data were classified to detect tax evasion. These classifiers (k-nearest neighbours, Random Forest, SVM and Neural Network) were trained and tested. As a result, an accuracy rate of 92.98% was obtained for the Random Forest classifier and it was concluded that the manual effort in eruption loss detection can be reduced by 81%.
Decision Tree	Bonchi et al., 1999 / Italy	By using deductive methods supported by logic database languages and inductive methods supported by decision trees, effective control strategies are designed.
Random Forest, Neural Networks, Naive Bayes, Logistic Regression	Ippolito and Lozano, 2020 /Brazil	Data partitioning, model training, testing, evaluation and validation were used to predict the crimes committed against the service tax system of the Municipality of Sao Paulo. As a result, the Random Forest method was found to be superior to other learning algorithms (like, Neural Networks, Naive Bayes, Logistic Regression) in terms of tax crime prediction performance.
Artificial Neural Network, Random Forest, XGBoost, Decision Tree, Logistic Regression, GaussianNB	Murorunkwere et al., 2022 /Rwanda.	Machine Learning models used in tax evasion detection are compared. The success rates obtained were Artificial Neural Network (89.9%), Random Forest (82.5%), XGBoost (82.4%), Decision Tree (81.2%), Logistic Regression (81%), GaussianNB (72.2%).
Random Forest	An and Suh, 2020 / South Korea	To solve imbalance problem, data was divided into sub-datasets on which a Decision Tree algorithm for each, and then classification was obtained by removing the trees whose performance were less than the average accuracy of all trees in the set. Thus, the best cluster on the basis of number was obtained. It was called Modified Random Forest model, and the highest accuracies were achieved in the applications.
Random Forest, Heterogenous Neural Network, Multilayer Neural Network	Xavier et al., 2022 /Brazil	Hetero-Graphic Neural Network, Random Forest and Multilayer Natural Network applications were analysed for tax evasion using company's data open to public. Especially with the Random Forest model and Hetero-Graphic Neural Network, the identification score over 98% was achieved.
Bayesian dialysis, Hybrid Machine Learning	Garcia and Gaballero, 2023 / Spain	By using Bayesian Dialysis and Hybrid Machine Learning Method, fraud detection was increased to 96% in VAT, and from 3.6% to over 23% in Customs.
Logistic Regression, Decision tree, Artificial Neural Network	Lin et al. (2015) /Taiwan	Fraud was analysed using data mining methods and expert opinions. Logistic Regression, Decision Trees and Artificial Neural Networks were used for data mining. In the study, 129 fraud cases and 447 non-fraud cases were selected, and 90% success was achieved for Decision Trees and 91% for artificial Neural Networks. The success of Logistic Regression was quite low.
Artificial Neural Network, Mapping	Assylbekov et al. (2016)/ Kazakhstan	A type of Artificial Neural Network trained using unsupervised learning, and Self-Organising Map were successfully used to detect VAT evasion behaviour.
Logistic Regression	Ruzgas et al. (2023) /Lithuania	The Logistic Regression method, which is used in cases where the VAT taxpayer deregisters without the initiative of the tax authorities, proved to be a reliable method for identifying important variables that can predict the avoidance of future tax obligations.

		According to the publicly available data of the Lithuanian tax administration, after the introduction of risk assessment models of taxpayers, the informal economy rates started to decrease.
A rule-based Machine Learning method - Association rule	Wu et al. (2012)/ China.	The association rule of the data mining technique was used on the VAT database to uncover specific patterns and relationships found in VAT evasion reports. This model can be used to select cases suspected of having VAT reports that are not eligible for further audit checks, helping tax auditors to perform tax evasion investigations more effectively
Random Forest,	Mittal et al. (2018) / India	A Random Forest-based classifier is used as a training set to identify businesses that cheat in the VAT system by using taxation data and reports from tax authorities. As a result of working on retrospective data, it is revealed that tens of millions of dollars of tax losses caused by fraudulent transactions can be recovered.
Hybrid Unsupervised Outlier Detection	Savic et al. (2022) /Serbia	The HUNOD (Hybrid Unsupervised Outlier Detection) method is a combination of Outlier Detection with a Clustering algorithm (K-means) and Outlier Detection with a representative learning algorithm (Learning Autoencoders). With this method, income tax risks can be detected between 90% and 98% depending on the clustering configuration and learning mechanism.
Association rule, Multidimensional classification	Matos et al. (2015) /BRAZIL	Tax fraud patterns were identified using association rule and multidimensional classification. Since common fraud patterns can be identified, tax evaders can be detected with 80% accuracy with this relatively fast model.
Multilayer Perceptron Neural Network	Lopez et al., (2019) /Spain	A Multi-Layer Perceptron Neural Network model was applied to income tax declaration data. In addition to the benefit of classification according to the probability of tax evasion of taxpayers, the model showed higher accuracy than the other models with a rate of 84.3%.
Regression, SVM	Rad et al. (2016)/Iran	By examining the fluctuations and noise detection in the amounts paid by taxpayers in previous years and combining linear regression analysis with SVM on the data from previous years, risk levels were discovered, and the prediction system was created. The method gave priority to high-income taxpayers.
Unsupervised Clustering	De Roux et al. (2018) / Columbia	Experimental results on 1367 tax returns showed that the operational efficiency of the tax audit process can be improved without the need for historically labelled data

The other feature that emerges in the table is the Random Forest model, which is a multiple Machine Learning model resulting from combining the Machine Learning method with Machine Learning. In this model, many Machine Learning results are used in Machine Learning to select the most appropriate one. Therefore, this tree-based model is one of the best known Machine Learning algorithms. The tree-based model usually contains a root node, several internal nodes, and several leaf nodes. The leaf nodes represent decision outcomes, the other nodes represent judgements about attributes, and each branch represents the output of a judgement (Zeng, et al., 2024, p,49). In studies with statistical models, it is stated that tree-based models perform well at a general level, and the Random Forest model has the highest accuracy among tax evasion detection methods (Zeng, et Al., 2024, p.51). On the other hand, SVM, as a classification model, has a good classification effect on small-scale data and strong generalisation ability, but when dealing with large-scale data, it is seen that it has difficulty in performing well in datasets with a large imbalance rate as well as being sensitive to missing data (Zeng, et Al., 2024, pp.8-9).

In addition to Machine Learning algorithms, these technological tools, which enable taxpayers to comply with tax laws better and more efficiently, are rapidly being developed in the service of tax administrations (Yayman, 2021, p.2786). In a study conducted by the OECD in 2023, it was measured that 65% of tax administrations use Artificial Intelligence technology and 40% of countries utilise virtual assistants. Some examples of virtual assistant applications used by tax administrations are given below (Çerik, 2024, p.50-51)

- *CRA - Canada: "CHARLIE", developed in March 2020 to answer questions that taxpayers will ask, especially during*

the preparation of their tax returns, answered 5 million questions in the first year of its implementation.

- *IRAS - Singapore: "ASK JAMIE" is an application that assists taxpayers in the preparation of income tax for taxi drivers and car hire businesses. It has answered over 15 million questions from taxpayers within 5 years of its launch.*
- *SAT - Mexico: Introduced in 2020, "ORIENTASAT" aims to assist in the preparation of income tax, especially for individual taxpayers. It uses a knowledge base of 1,149 standardised answers and 17,776 question types.*
- ALEX- Advanced Artificial Intelligence-based software used by the Australian tax administration or databases that allow taxpayers to access data critical to complying with tax law (Yayman, 2021, p.2787).
- The chatbot named SyRi (Systeem Risico Indicatie) in Netherlands has been implemented by many municipalities after 2014 to profile taxpayers in terms of tax evasion risk (Durmuş, and Erdem, 2023, p.238).
- GİBİ- is a Digital Tax Assistant that works with AI supported Machine Learning in Turkey. All citizens are provided with answers to frequently asked questions about the legislation. A single administration model is defined for issues such as the payment of public debts in due time, the use of the "mutual agreement method" in solving tax problems involving more than one country in the double taxation avoidance agreement, the dissemination of advance pricing agreements in transfer pricing, etc. **(**Revenue Administration, Online; Rüzgar, 2022, pp.2-5).

5. Legal Issues and Determinations

Traditional tax systems have difficulty in coping with new business models and value creation processes emerging in the digital economy. Therefore, new digital products and technologies should be actively used in taxation (Gölçek, 2023, pp.543-573).

It should be ensured that countries' online digital services are used in an accurate, reliable manner: in communication announcements, explanations and guidance should show accuracy, consistency, and professionalism (Yayman, 2021, p.2784). The tax administration should not take the risk of providing false information in its jurisdiction. Differences and grey areas in the regulations should be focused on, and legislation should be carefully created for autonomous systems that are able to follow international regulations and to make judgements on all administrative decisions and court cases. Lawmakers need to clearly determine the rules for specific problems such as who will be held responsible for a wrong advice generated by autonomous systems, such as the taxpayer, system developer or service provider. (İlgün, 2020, pp.253-254). In Artificial Intelligence technology, big data support is needed to identify, match, and predict the right data. Collecting more data over time poses the risk of not being able to guarantee the protection of the privacy of personal information and the risk of information disclosure (Rüzgar, 2022, p.10).

In order to achieve the tax targets of the administrations in terms of tax security and to ensure healthy public finance, it has been seen that the data obtained should be harmonized with the entire system, and modern and interdisciplinary applications such as digital systems should be integrated into the financial system for real-time monitoring and control of simultaneous, reliable, fast and accurate results in a transparent manner for the taxpayer and the administration (Tolkun, and Tekin, 2022, p.291).

Today, not all regional and international tax administrations have the same level of digitalisation and the same capacity and capabilities to use Machine Learning and data analytics. The governance models they use will have structural limitations in terms of fiscal violations, tax compliance, and administrative costs, etc. Developing such a capacity is a long process that includes, but is not limited to, the following components (Adamov, 2019, p.4).

- Understanding how taxation can benefit from operational data,
- Understanding own data and having the ability to transform data into quality data,
- Deployment of the right applications and provision of online services to collect the right data,
- Making all data accessible (using specialised distributed platforms),
- Ability to apply Machine Learning and data analytics at the scale,
- Development of data science (scientific) skills.

New concepts and business models encountered in the digital transformation process may create new problems. Legal systems should be organised with legal regulations to respond to these problems and tax compliance should be encouraged in this area.

6. Conclusion and Evaluations

Instead of time-consuming and labour-intensive traditional auditing tools, tax authorities want to benefit from the advantages obtained by computer-based tax risk analysis of developing technologies. In countries where the informal economy is more effective with their fragile structures, governments are struggling with more tax revenue losses. Therefore, there is a need for models developed for the detection of tax evasion in risk assessment based on predicting the behaviour of high-risk taxpayers. Analyses based on statistical models and software

programmes are effective in tracking and revealing confidential information**.**

The financial information monitored in the follow-up of taxation processes, tax compliance studies, data generated by global multinational companies and their subsidiaries constitute a great source, and there is an increasing need for real-time tax technologies in tax-related analyses. For this purpose, one of the main objectives of the reforms in the field of digital technology is to "strengthen the effectiveness of risk management and tax audit"**.** With the development in information technologies, big data emerges with the fact that the documents, books and declarations of accounting records for businesses in particular are now kept completely electronically, and in this direction, big data analytics is inevitable in auditing.

In this study, it is emphasised that the methods used in the detection of high-risk taxpayers are researched and the contribution of these methods to the audit mechanism in reducing tax losses is emphasised. In risk analysis management, countries develop preventive measures against tax losses and evasions by testing the taxpayers to be examined and determining statistical methods having high accuracy rates before the tax inspection process is initiated. Thus, unnecessary and ineffective tax examinations that cause waste of time are eliminated and risky taxpayers are identified with high accuracy. When creating tax scenarios, experts take into account possible interactions between different entities. However, some methods may be insufficient to reveal the behavioural characteristics that define these interactions between taxpayers. In this case, relationship-based risk identification methods are utilised.

In the risk analysis model, taxpayers are grouped on the basis of sector, region and size criteria and scored according to the risks they have for the relevant analysis period. Taxpayers with high-risk scores are considered as risky taxpayers and may be referred for examination. The definition of "Risk Analysis" in the inspection letters sent to taxpayers increases taxpayer awareness, activates the role of risk management in the audit, and therefore, as an action that increases the deterrent power for taxpayers against tax fraud, it becomes a driving force that encourages tax compliance.

Socio-economic and demographic structure of the country as well as regional differences can be effective in the choice of risk analysis method, so that the specific structure of the tax region and policies may play an important role. While Machine Learning techniques based on learning and problem solving provide speed, cost, and time savings in tax risk management, they also create the need for specialisation in this field. In risk detection methods, while digitalisation in taxation facilitates the process, it also requires the follow-up of updated statistical models. Therefore, in parallel with the improvement of tax automation systems, integrations are made with institutions and organisations that provide data richness.

It is envisaged to increase the human capital power of tax administrations that can produce analytical solutions in data analysis and establish links between patterns, to make strategies to expand training outside the risk analysis centre in this field, to strengthen the technology infrastructure related to digitalisation and risk management in audit and to allocate investment budgets in this field. The following findings are obtained by analysing large data sets:

- The processes transferred to tax inspection reports are prepared in a more understandable format,
- Suggestions and scenarios developed with the findings identified from past data enable more inclusive studies in the audit,
- It works at a speed that people cannot reach during the creation of an audit planning for the new period and helps the audit process to be completed faster by utilising the past audit planning and risk identification.
- The efficiency of audit activities can be enhanced, audit results can be reported more effectively, error rates and costs can be reduced.

As a result, in studies conducted on the use of data mining, Machine Learning and Artificial Intelligence combined technology in tax auditing, it is frequently noted that Machine Learning and Artificial Neural Network algorithms have a higher accuracy rate in detecting risky taxpayers and fake document issuers. In addition, it is emphasized that tax scenarios that differ according to taxpayer group, tax type and sector should be diversified by taking into account human behaviour characteristics and correlation between relationships.

In studies focused on risk analysis, it may not be possible to find the law based on the distribution of probabilities. In this case, collecting as much information as possible based on solution options and applying Machine Learning algorithms in the Random Forest method can produce accurate results. The correct flow and combination of these data to the tax authorities directly affects the efficiency of the report in risk analysis.

In the evaluation made in terms of the limitations of the research; in the Turkish legal system, due to the issues related to tax privacy in the Tax Procedure Law and the rules regarding data security within the scope of the Personal Data Protection Law, no response could be received to the survey statements requested to be

conducted with the Risk Management Centre in Turkey. However, the risk management studies declared in the annual activity reports of the Tax Audit Board were utilised.

It is seen that there is a need for information engineering of big data for the explainability and reliability of risk detection results, and that there are increasing efforts to transition from risk detection informatics to intelligence.

References

1. Adamov, A.Z., (2019). Machine Learning and Advanced Analytics in Tax Fraud Detecti*on.* IEEE 13th International Conference on Application of Information and Communication Technologies (AICT), *B*aku, Azerbaijan, 2019, pp. 1-5, DOI: 10.1109/aict47866.2019.8981758
2. An B., Suh Y., (2020). Identifying Financial Statement Fraud with Decision Rules Obtained from Modified Random Forest. Data Technol Appl., 54(2), pp.235–55. DOI:10.1108/dta-11-2019-0208
3. Andrade JPA, Paulucio LS, Paixao TM, Paixao TM, Berriel RF, Carneiro TCJ, et al. (2021). A Machine Learning-Based System for Financial Fraud Detection. In: Proceedings of the 18th National Meeting on Artificial and Computational Intelligence (ENIAC 2021); 2021 Nov 29– Dec 3; online. São Leopoldo: Sociedade Brasileira de Computação (SBC); 2021. pp. 165–76.
4. Assylbekov, z., Melnykov, I., Bekishev, R., Baltabayeva, A., Bissengaliyeva D. and Mamlin, E., (2016). Detecting Value-Added Tax Evasion by Business Entities of Kazakhstan. Proceedings of the 8th KES International Conference on Intelligent Decision Technologies, KES-IDT 2016, Smart Innovation, Systems and Technologies, 56, pp. 37-49
5. Biyan, Ö., Altundal Biyan, N. and Işıklı, C. (2022). The Effect of Web-Based Software from Digital Transformation Tools on The Performances of Facilities on Staff: Apsiyon Software Case. Alanya Academic Review, 6(3), pp.3327-3347. https://doi.org/10.29023/alanyaakademik.1059141
6. Bonchi F, Giannotti F, Mainetto G, Pedreschi D. (1999). Using Data Mining Techniques in Fiscal Fraud Detection. In: Proceedings of the 1st International Conference on Data Warehousing and Knowledge Discovery; 1999 Aug 30–Sep 1; Florence, Italy. Berlin: Springer, pp.369–76. https://doi.org/10.1007/3-540-48298-9_39
7. Brondolo, J., Chooi, A., Schloss, T., and Siouclis, A., (2022). Compliance Risk Management: Developing Compliance Improvement Plans. Technical Notes and Manuals (Vol. 2022, Issue 001), International Monetary Fund, Washington, DC. DOI: https://doi.org/10.5089/9798400205910.005
8. Bulutistan, (2024A) Cloud Computing. Available from: https://bulutistan.com/blog/bulut-bilisim/. (12. 05. 2024).
9. Bulutistan, (2024B) What is a Data Lake? Differences Between a Data Warehouse and a Data Lake. Available from: https://bulutistan.com/blog/data-lake-veri-golu-nedir-veri-ambari-ve-veri-golu-arasindaki-farklar/. (20. 07. 2024).
10. Çalış, Y.E, Keleş, E., and Engin, A., (2014). The Importance of Information Technologies in Detecting Fraud and a Practice. Journal f Accounting and Finance, 63, pp. 93-108. https://doi.org/10.25095/mufad.396479
11. Çelik, M., (2009). Classification Methods Used in Data Mining and an Application (Veri Madenciliğinde Kullanılan Sınıflandırma Yöntemleri ve Bir Uygulama), Master's Thesis, İstanbul University, Institute of Social Sciences, Department of Econometrics, Turkey, p.7.
12. Çerik, M.C., (2024). A New Era in Taxation: Artificial Intelligence Applications (Vergilemede Yeni Bir Dönem: Yapay Zekâ Uygulamaları). Master's Thesis, Aksaray University, Institute of Social Sciences, Department of Finance and Financial Management, Turkey, pp.50-51
13. Çiçek, S. & Çiçek H.G., and Sahin-İpek E.A.S., (2019). Behavioral Approach to Tax Compliance Process: Taxpayer Behaviors and Typologies. Sosyoekonomi, Vol. 27(39), pp.231-234
14. de Roux D, Perez B, Moreno A, del Pilar Villamil M, Figueroa C. (2018). 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; 2018 Aug 19–23; London, UK. New York City: Association for Computing Machinery (ACM); 2018. pp.215–22.
15. Demirci, Y., (2018). Data Mining in Public Services: A Research on Municipality Help Desk Data. Doctoral Dissertation, Sakarya University Institute of Social Sciences, Turkey.
16. Durmuş, N.K and Erdem, I.A, (2023). Tax Administration 3.0: An Analysis in Terms of

Artificial Intelligence. Finance Magazine, 184, pp.225-253

17. Ersöz, F., and Çınar, Y., (2021). Comparison of Data Mining and Machine Learning Approaches: An Application in Textile Industry. European Journal of Science and Technology, (29), pp. 397-414. https://doi.org/10.31590/ejosat.1035124
18. Ertikin, K. (2017). Fraud Auditing: Proactive Approaches Used for Red Flags Detection. The Journal of Accounting and Finance, (75), pp. 71-94. https://doi.org/10.25095/mufad.399313
19. Eyüpgiller, S., (2023) The Analysis of a Critical Period in the Digital Transformation of the Turkish Tax, System: 2017-2021. Social Review of Technology and Change, vol. 1 no. 1, pp. 68 – 100.
20. Gölçek A.G., (2023) Tax Planning in Digital Economy: Challenges and Opportunities in Taxation. Academic Review of Economics and Administrative Sciences (OHUIIBFD), 16(2), pp. 543–573. http://doi.org/10.25287/ohuiibf.1231016
21. Güneş, A.Ş., and Arslan, Y., (2023). A Study on the Use of Data Mining and Artificial Intelligence in Tax Auditing: A Decision Tree Example for Detecting Forged Document Makers. Vergi Raporu, 289, pp.45-57.
22. Garcia, I.G., Caballero, A.M., (2023). A Comparison Between Bayesian Dialysis and Machine Learning to Detect Tax Fraud and Its Causes: The Case of Vat, Corporate Tax and Customs Duties in Spain. SN Computer Science, 4(80). https://doi.org/10.1007/s42979-022-01483-5
23. González PC, Velásquez JD. (2013). Characterization And Detection of Taxpayers with False Invoices Using Data Mining Techniques. Expert Syst., 40(5), pp. 1427–36. DOI:10.1016/j.eswa.2012.08.051
24. Gür, Ö., (2023). Decision Tree Supported Fraud Detection and an Application. Alanya Academic Review, 7(1), pp.511-528. https://doi.org/10.29023/alanyaakademik.1196078
25. Güzel, K., (2020)., What is Boosting? AdaBoost Algorithm Step by Step. Available from: https://kadirguzel.medium.com/boosting-nedir-ad%C4%B1m-ad%C4%B1m-adaboost-algoritmas%C4%B1-439cce20ab9a (12. 06. 2024)
26. İlgün, M.F., 2020. Big Data Analytics in Tax Audit. ournal of Politics, Economics and Management Studies (SEYAD), Volume 8, Issue 1, pp. 1-24.
27. Kara, S., and Karan, L., (2023). Fraud Audit Approaches and A Case Study. Denetişim, 28, pp. 75-87. https://doi.org/10.58348/denetisim.1233039
28. Kırda, K., and Özçelik, M. K. (2021). Identifying The Companies with The Risk of Financial Statement Fraud by Data Mining. Journal of Accounting and Taxation Studies, 14 (2), pp.609-639. https://doi.org/10.29067/muvu.802703
29. Konar, S., (2023). Risky Situations and Opinions of The Administration in End of The Term Transactions. GİB, İstanbul Revenue Administration (seminar notes), İstanbul. Available from: https://istanbulymmo.org.tr/dosyalar/sunumlar/20231228/DS%C4%B0%20Serdar%20KONAR.pdf (28. 07. 2024)
30. Köse, H.Ö, and Taner, A., (2024) Kamu Yönetiminde Mali denetim (Financial Auditing in Public Administration), T.C. Sayıştay Başkanlığı (Turkish Court of Accounts), Ankara. Available online: https://www.sayistay.gov.tr/files/3077_KAMU_YONETIMINDE_MALI_DENETIM_FUL.son.pdf (05. 08. 2024)
31. Lin CC, Chiu AA, Huang SY, Yen DC. (2015) Detecting the Financial Statement Fraud: The Analysis of The Differences Between Data Mining Techniques and Experts' Judgments. Knowl. Base Syst., 89, pp.459–70. https://doi.org/10.1016/j.knosys.2015.08.011
32. Matos T, de Macedo JAF, Monteiro JM. (2015). An Empirical Method for Discovering Tax Fraudsters: A Real Case Study of Brazilian Fiscal Evasion. In: Proceedings of the 19th International Database Engineering & Applications Symposium; 2015 Jul 13–15; Yokohama, Japan. New York City: Association for Computing Machinery (ACM); 2015. pp.41–8.
33. Mittal S, Reich O, Mahajan A. (2018). Who is bogus? Using one-sided labels to identify fraudulent firms from tax returns. In: Proceedings of the 1st ACM SIGCAS Conference on Computing and Sustainable Societies; 2018 Jun 20–22; Menlo Park and San Jose, CA, USA. New York City: Association for Computing Machinery (ACM); 2018. pp. 1–11.
34. Murorunkwere, B.F., Haughton, D., Nzabanita, J., Kipkogei, F., Kabano, I., (2023). Predicting Tax Fraud Using Supervised Machine Learning Approach. African Journal of Science, Technology, Innovation and Development, DOI: 10.1080/20421338.2023.2187930

35. Rad, M.S. and Shahbahrami, A., (2016) Detecting High Risk Taxpayers Using Data Mining Techniques. 2nd International Conference of Signal Processing and Intelligent Systems, 2016 (ICSPIS), Tehran, Iran. IEEE Publishing.
36. Rasgen, M., (2016). An Application Related to The Use of Benford's Law in Fraud Auditing, Master's Thesis, Dokuz Eylül University, Institute of Social Sciences, Department of Business Administration, Division of Accounting, Turkey.
37. Revenue Administration, [Online]. Online Tax Assistant (GİBİ). Available from: https://x.com/gibsosyalmedya/status/1844369795434807473 (02. 08. 2024)
38. Ruzgas, T.; Kižauskiene˙, L., Lukauskas, M., Sinkevic˘ius, E., Frolovaite˙, M., and Arnastauskaite˙, J. (2023). Tax Fraud Reduction Using Analytics in an East European Country. Axioms, 12(3), p. 288. https://doi.org/10.3390/axioms12030288
39. Rüzgar, M.T., (2022). Establishment Of Digital Tax Office And Digital Tax Assistant System In Turkey Within The Scope Of Economy Reform Package, Turkuaz International Journal of Socio-Economic Strategic Research, Volume: 4, Issue: 1, pp.5-6
40. Sağlar, J., & Uluç, S. M. (2022). An Application on The Availability of Analytical Review Procedures in Fraud Auditing. Socrates Journal of Interdisciplinary Social Studies, 20, pp. 82–107. https://doi.org/10.51293/socrates.315
41. Savic, M., Atanasijevic, J., Jakovetic, D., Krejic, N., (2022) Tax Evasion Risk Management Using a Hybrid Unsupervised Outlier Detection Method. Expert Systems with Applications, Volume 193, p.116409. https://doi.org/10.1016/j.eswa.2021.116409
42. Söğütlüoğlu, G., 2024, Tax Risk and Opportunity Analysis (Vergisel Risk ve Fırsat Analiz). Deloitte Times. Available from: https://www.verginet.net/dtt/5/Vergiselriskvefirsatanaliz.aspx (02. 08. 2024)
43. Şahin, A., (2024). Internal Losses and Audits (Şirket İçi Kayıplar ve Denetimi, 28.02.2024). Available from: https://www.muhasebetr.com/yazarlarimiz/aydinsahin/005/ (20. 07. 2024)
44. Şencan, D. (2023). Estimation of Tax Loss and Evasion in Turkey with Data Mining Process. Erzurum Technical University Journal of Social Sciences Institute (ETUSBED), 17, pp. 11-24. https://doi.org/10.29157/etusbed.1292006
45. Tax Inspection Board (VDK), (2023). Activity Report 2022. Tax Inspection Board, Republic of Turkey Ministry of Treasury and Finance, Ankara. Available from: https://ms.hmb.gov.tr/uploads/sites/17/2023/03/Vergi-Denetim-Kurulu-Baskanligi-2022-Yili-Faaliyet-Raporu.pdf (15. 06. 2024)
46. Tax Inspection Board (VDK), (2019). Activity Report 2018. Tax Inspection Board, Republic of Turkey Ministry of Treasury and Finance, Ankara. Available from: https://ms.hmb.gov.tr/uploads/sites/17/2023/03/Vergi-Denetim-Kurulu-Baskanligi-2022-Yili-Faaliyet-Raporu.pdf (12. 06. 2024)
47. Terzi, S. (2012). Detection of Fraud and Financial Impropriety with Data Mining. The Journal of Accounting and Finance, (54), pp. 51-64.
48. Tolkun, S., and Tekin, A., (2022). Digital Transformation in The Turkish Tax System and The Role of Tax Security on Efficiency. Anadolu Akademi Social Sciences Journal, Vol.4, Iss.2, pp.290-314.
49. Uğur, A.A., (2016). The Applicability of the Risk Analysis System in Tax Audit Effectiveness (A Sample Application on Gaziantep Carpet Sector). IJPF, Vol. 1, Issue 2, pp.122– 145. https://doi.org/10.30927/ijpf.319881
50. Xavier, OC, Pires, SR, Marques, TC, Soares, AS. (2022). Tax Evasion Identification Using Open Data and Artificial Intelligence. Rev Adm Pública, 56(3), pp. 426–40. http://dx.doi.org/10.1590/0034-761220210256x
51. Wu, R.S., Ou, C.S., Lin, H., Chang, S.I., Yen, D.C., (2012). Using Data Mining Technique to Enhance Tax Evasion Detection Performance. Expert Syst Appl., 39(10), pp. 8769–77. https://doi.org/10.1016/j.eswa.2012.01.204
52. Yayman, D., (2021). Future Scenarios of Tax Administrations. Journal of Social and Humanities Sciences Research, 8(76), pp. 2872-2895. https://doi.org/10.26450/jshsr.2775
53. Zheng, Q., Xu, Y., Liu, H., Shi, B., Wang, J., Dong, B., (2024). A Survey of Tax Risk Detection Using Data Mining Techniques. Engineering, Vol. 34, pp. 43-59. https://doi.org/10.1016/j.eng.2023.07.014