ISSN No:-2456-2165

Taxes and Finance Field Using Machine Learning Techniques: A Survey

Abeer A Shujaaddeen[1*]; Fadl Mutaher Ba-Alwi1[2]; Abdulkader M. Al-Badani[3]
[1]Computer Science Department, Faculty of Computer and Information Technology, Sana'a University, Yemen
[2]Computer Science Department, Faculty of Computer and Information Technology, Sana'a University, Yemen
[3]Faculty of Science and Engineering, Department of Computers,Aljazeera University, Ibb, Yemen

Corresponding Author:- Abeer A Shujaaddeen[1*]

Abstract:- Taxes are considered one of the most important revenues for developed and undeveloped countries alike, because of their importance in raising the level of the country. Taxes are an amount that the state imposes on companies and individuals. However many taxpayers try to evade tax by not paying their taxes in several ways, such as lying on the declaration form, hiding part of the data for tax fraud, and other ways and methods. Therefore, many countries have implemented many procedures and regulations to reduce tax evasion. Recently, it has resorted to artificial intelligence techniques such as machine learning (ML) and deep learning (DL) such as neural networks, decision trees, random forests, clustering techniques such as K-Mean, and others to reduce tax evasion. In this paper, we will present a summary of a group of countries in their trying to detect tax and financial evasion and fraud.

Keywords:- *Taxes, Tax Fraud, Taxpayers, Machine Learning, and Deep Learning.*

I. INTRODUCTION

Tax can be defined as a monetary payment made to the government by individuals and organizations. Its purpose is to provide funding for public sectors that are under the administration of the government. These sectors include education, which encompasses schools, teachers' salaries, and the salaries of workers in ministries and government institutions. Additionally, tax revenue is used to support various aspects such as hygiene, economic policies, and the maintenance of state infrastructure, including sanitation, dam construction, and unemployment insurance [1].

Tax fraud is a broad term used to describe the intentional actions taken by individuals or organizations to unlawfully evade tax payments. It involves concealing the true financial status of the taxpayer from tax authorities to minimize the amount of taxes owed. This can include submitting false tax reports, such as underreporting profits or providing inaccurate information. Tax fraud is commonly associated with activities conducted in the informal economy. One way to measure tax fraud is through the "tax gap," which represents the discrepancy between the income that should be reported to tax authorities and the actual amount reported. In essence, tax fraud involves providing false information on a tax return form to reduce tax liability [2].

The persistent issue of aggressive tax avoidance and the reluctance of certain tax practitioners to collaborate with tax administrations continue to pose significant challenges. Concurrently, business leaders in some developing countries express concerns about being held to higher standards compared to other taxpayers [3].

Local tax authorities, who are responsible for developing cost-effective solutions to address this issue, place a high priority on identifying and preventing tax fraud. The use of machine learning algorithms has been at the forefront of several recent efforts to detect tax fraud.

Machine learning and artificial intelligence play a crucial role in combating tax and financial evasion. They achieve this by leveraging algorithms to detect potential wrongdoing and conducting real-time transaction analysis, thereby reducing fraud. The use of machine learning and deep learning techniques is crucial for these systems to function well.

- *The Researches and Approaches that tackle using Machine Learning and Deep Learning Techniques in Trying Tax and Finance Fraud Detecting as follows:*

According to [4], supervised machine learning techniques face challenges in detecting tax fraud, particularly in the Colombian context, due to the limited availability of historically labeled data. Auditing requires significant time and resources, making it difficult to generate labeled data for training supervised algorithms. Consequently, the generalization power of these algorithms is hindered, limiting their effectiveness. The researchers in the paper propose a technique that enables tax authorities to prioritize audits based on data-driven methods, without relying on historically labeled data.

The results produced using this method show that the model can identify questionable tax declarations and flag them as suspicious without the need for past labeled data, improving operational efficiency.

According to tax invoices, researchers in [5] propose a CNN-RNN structure that is compositional and incorporates an attention mechanism to classify transaction behavior.. This classification provides a fresh viewpoint for examining the regional industrial structure and is essential for tax oversight.

According to preliminary research, transaction behavior can be classified with an overall accuracy of 75%.

In [6], the paper examines the tax planning landscape in the context of artificial intelligence and big data. It addresses tax planning issues within the framework of big data and suggests utilizing these technologies to optimize tax planning. A new model is created when big data and tax planning are combined.

The paper given in [7] reflects on the preliminary findings of a collaborative scientific research initiative between the Tax Administration and the Faculty of Sciences at the University of Novi Sad. The project's goal is to create algorithms for detecting the risk of tax evasion using advanced big data analytics and artificial intelligence techniques, as well as machine learning. The presented approach is based on an indicator that compares a legal entity's income distribution to the average income distribution in the relevant business sector. The results illustrate the effectiveness of the developed indicator.

In [8,] the researchers propose a universal architecture termed the unsupervised conditional adversarial network (UCAN) for identifying tax evasion. This approach is the first attempt to address audit tasks in unlabeled target domains via inter-region transfer. The architecture makes use of an adversarial neural network and incorporates label information into the distribution adapter, which allows for fine-grained adaption of the data's joint probability distribution. The model applies a constraint based on the retrieved features' conditional maximum mean discrepancy (CMMD) to align the conditional probability distribution (CPD) for deep representation. The model combines the distribution adapter and the label predictor to allow for end-to-end learning of unsupervised feature transfers. Experimental results illustrate the model's remarkable performance in numerous migration tasks compared to the state-of-the-art.approaches.

The research in [9] focuses on identifying tax fraud in Spanish personal income tax returns (IRPF). The study makes use of cutting-edge machine learning-based forecasting techniques, notably Multilayer Perceptron neural network (MLP) models. Using neural networks, the researchers were able to divide up the taxpayers and assess the probability that a particular taxpayer would attempt to evade taxes. The chosen model outperformed previous tax fraud detection models, with an efficiency rate of 84.3%. The suggested method might be expanded to measure a person's propensity for tax fraud in regard to various sorts of taxes. These models can help tax offices make defensible choices.

There are two goals for the study in [10]. Its primary goal is to find out how Small and Medium Businesses (SMEs) view the existing situation and strategies for cutting administrative expenses. Second, it examines the connection between the costs of tax policies and entrepreneurial activity using descriptive statistics and hierarchical cluster analysis. Datasets from Slovenia and the European Union (EU) are analyzed independently. The results indicate that the total amount of early-stage entrepreneurial activity and the density of new businesses in EU nations are more significantly impacted by tax administrative costs.

Furthermore, the findings for Slovenia highlight the need of a reliable tax system, with an emphasis on information technology and procedural measures.

In [11], the researchers explore the challenge of financial fraud and how financial organizations are using mining tools to counter it. The paper presents an overview of fraud strategies, with a particular emphasis on machine learning, data mining, and preventative techniques like clustering, classification, and regression. The goal is to use mining techniques to create remedies for financial fraud.

The study provided in [12] addresses the issue of establishing the strategy of a self-interested, risk-averse tax body. The study uses Q-learning and new advances in Deep Reinforcement Learning to achieve approximate solutions. The research entails identifying the expected tax evasion behavior of taxpayer entities, establishing the risk aversion level of the "average" entity using empirical tax evasion estimates, and evaluating sample tax plans. The model serves as a testbed for tax policies and makes various policy recommendations based on the outcomes.

In [13], the study discusses known strategies for identifying tax evasion in databases utilizing expert systems. It compares the suggested expert system to various strategies for improving tax evasion detection. The study proposes an abstract solution based on an expert system in the domain of tax evasion, complete with performance modeling. The expert system builder acts as an interface for personnel working with the defined expert system. The results show that the suggested expert system detects tax evasion trends with a high level of accuracy.

The study described in reference [14] introduces a conceptual framework that aims to establish a solid methodological and theoretical basis for employing Data Analytics in the field of taxation. The research primarily concentrates on the utilization of operational data by tax authorities and identifies machine learning techniques that prove effective in detecting particular forms of fraud.

In [15], the researchers utilize data mining tools to detect fraud in banking by leveraging the data already collected by the bank. They employ supervised machine learning techniques, specifically support vector machines, to detect fraudulent transactions based on intentional and unintentional client reactions and new transactions. The support vector machine algorithm successfully identifies customers engaged in fraudulent transactions, using a database of credit card transactions to combat banking fraud.

The study in [16] addresses the economic impact of unpaid taxes by suggesting an automated system for forecasting tax defaults. The researchers use a variety of feature transformation techniques as well as cutting-edge machine learning algorithms. The prediction algorithm is validated using a dataset containing information on tax

defaults and non-defaults in Finnish limited liability enterprises.

In [17], the researchers look on the use of unsupervised and semi-supervised machine learning approaches to detect abnormal tax returns for the Norwegian Tax Administration. They investigate the capabilities of these strategies and examine how different dataset aspects affect their performance. The goal is to discover appropriate ways for detecting new types of errors, resulting in a reduction in tax errors that affect tax revenue.

The research discussed in reference [18] tackles the issue of having a scarcity of labeled data in the domain of tax fraud detection. To overcome this challenge, the researchers utilize unsupervised anomaly detection methods, which are not commonly employed in tax fraud detection studies. They examine a distinctive dataset that incorporates VAT declarations and client listings for all VAT numbers in Belgium across ten sectors.

The study in [19] seeks to review the body of research on audit and tax from the perspective of developing technology while also establishing a research agenda for the future. By combining text analysis and bibliometrics, the researchers use a meta-literature technique to assess 154 notable English papers published in Scopus journals during the last 35 years. The programs utilized in the study included RStudio, VOS Viewer, and Microsoft Excel.

In [20], social planners and economic agents are trained via model-free reinforcement learning (RL) in AI-based economic simulations. The fundamental advantage of model-free RL is its flexibility, which allows the planner to employ any social purpose as a reward function. Furthermore, no prior world knowledge is required to design a successful tax policy.

[21] introduces a revolutionary method called MALDIVE for assisting tax authorities in tax risk assessment to find tax evasion and avoidance. The network model used by MALDIVE to describe the numerous connections amongst taxpayers. To help public servants identify problematic taxpayers, an approach that combines data mining and visual analytics methodologies has been developed. The paper provides a four-step implementation process for MALDIVE.

The study in [22] analyzes tax evasion detection as a critical function of tax administration and develops a model for estimating the likelihood of tax evasion that incorporates quantitative and qualitative markers. The study employs research techniques such as systematic analysis, scientific abstraction, logical generalization, expert review, and statistical analysis. The study evaluates the chance of identifying tax evasion in the Republic of Azerbaijan using the proposed model, and the results show a 29% probability. The findings suggest the need for improvements in the tax administration mechanism in Azerbaijan, emphasizing the practical significance of the proposed model in enhancing the effectiveness of tax institutions and impacting state budget revenue through the determination of the probability of detecting tax evasion.

In [23], the study focuses on modeling tax behavior in the expatriate community. The researchers analyze survey results from the "Ethical Obligation to Pay Fair Taxation Survey" to identify possible combinations, resulting in the identification of 18 structures. Using a big data strategy, data on these 18 structures is collected, resulting in 2090 pages of data containing 377,783 words related to tax evasion. The data is pre-processed and analyzed using KH Coder, a text analysis tool. The interpretation of the data leads to the reduction of the 18 structures to seven comprehensive structures. A literature review is conducted based on these seven "basic" structures. The data is analyzed using KH Coder and machine learning techniques, resulting in a new tax evasion model with seven dimensions: 1. Taxation of the Rich, 2. Implementation Strategies, 3. Business Tax Planning, 4. Capital Gains Tax, 5. Inequality of Wealth and Power, 6. Economic Effects of Taxes, and 7. Audits and Materiality.

In [24] proposes to apply machine learning for decision-making in fiscal audit plans related to service taxes in the municipality of São Paulo. The researchers use machine learning, specifically Random Forests, to forecast crimes against the tax system. The findings show that Random Forests outperform other learning algorithms in terms of tax crime prediction. Random Forests also have strong generalization ability. Improved projections result in more efficient audit strategies, more tax income, and taxpayer compliance with tax regulations.

In [25] examines how artificial intelligence (AI) is used in the Indian revenue system. They take into account variables like tax expertise, tax education, tax complexity, legal penalties, interactions with tax authorities, ethics, perceptions of the tax system's fairness, feelings about paying taxes, knowledge of offenses and penalties tax compliance, tax education, and the likelihood of an audit. The goal of the study is to comprehend how AI might affect these variables and possibly improve the Indian taxation system.

The study [26] proposes a novel hybrid machine learning-based technique for mitigating the risk of tax fraud. The approach incorporates domain information into the model, resulting in an explainable DT model that domain experts can verify. It also contains an anomaly validation function that employs two separate anomaly detection methods (K-means and autoencoder). The method is intended to detect tax fraud involving personal income and makes use of big data techniques to improve tax fraud detection.

In [27], the researchers demonstrate the use of machine learning and network science tools to automatically identify patterns of tax evaders. This has potential applications in various areas such as bribery practices, money laundering, and other illegal activities, benefiting society. However, caution should be exercised when applying these methods, and their limitations should be considered.

The paper [28] describes a machine learning-based approach for detecting tax evasion in Espírito Santo, Brazil. Four classifiers (Random Forest, k-nearest Neighbors, Neural Network, and Support Vector Machine) are trained using tax and financial data from diverse organizations. The Random Forest classifier performs the best, with a macro-averaged F1 score of 92.98%. The study illustrates Random Forest's ability to produce reliable outcomes.

In [29], the researchers discuss financial statement fraud, which is becoming a major issue for governments, businesses, and investors. They offer a hybrid system that includes a support vector machine, an upgraded ID3 decision tree, multilayer perceptron neural networks, and a genetic algorithm to improve accuracy and performance. The model was evaluated on financial statements from Tehran Stock Exchange-listed companies, and it predicted financial statement fraud with a high accuracy (about 80%).

The study in [30] explores the range of applications of machine learning, including recommendation systems fraud detection, customer behavior prediction, image recognition, speech recognition, black & white movie colorization, and accounting fraud detection. The focus is on the use of neural networks in finance, accounting, and research fields. The researchers emphasize that machine learning in accounting research has not yet reached its full potential.

In [31], the researchers discuss the increasing threat of financial fraud and the need for solutions in the financial sector. They present an overview of different fraud techniques and emphasize the importance of continually improving fraud detection systems. Machine learning and data mining techniques, such as classification, clustering, and regression, have been widely used in recent studies for fraud prevention.

In reference [32], researchers employed machine learning approaches to solve the difficulty of detecting fraud among a varied set of taxpayers. They created a fraud prediction model with gradient boosting as the core method. Despite working with a limited sample size and dealing with widely defined fraud, the study was able to identify key elements from tax returns with little further information. The results showed that the projected fraud rate among the top cases was almost 1.85 times higher than the average observed rate. This study demonstrates the usefulness of the proposed model in predicting and identifying potential cases of fraud within the taxpayer community.

In [33], used powerful machine learning techniques to detect tax evasion. To find optimal weights, the researchers modified the multilayer perceptron neural network with an improved particle swarm optimization (IPSO) technique. They also improved support vector machine (SVM) classifiers by adjusting their settings. The suggested IPSO-MLP model beat the IPSO-SVM, logistic regression, SVM, Naive Bayes, k-nearest neighbor, AdaBoost, and C5.0 decision tree models in terms of accuracy. The IPSO-MLP model obtained 93.68% accuracy, whereas the IPSO-SVM model achieved 92.24%.

In [34], the researchers proposed machine learning-based predictive analytics as a decision support system for exploiting latent tax opportunities. They developed three machine learning models: decision tree, random forest, and logistic regression. Using trigger data and other predictors, they analyzed 5,562 samples of potential tax income. The random forest model produced the most precise prediction outcomes.

The study in [35] aimed to establish a fraud detection system in tax. The researchers employed predictive techniques and feature extraction to identify fraud trends and anticipate future tax payments. They were able to use the random algorithm to anticipate the amount of future tax each individual should pay.

The purpose of [36] was to identify tax fraud features with a supervised machine learning model. The researchers compared numerous models, including Gaussian NB, XG Boost, Random Forest, Decision Tree, and Logistic Regression. The evaluation metrics showed that artificial neural networks were the most accurate model for predicting tax fraud.

The primary goal of the study in [37] was to improve the effectiveness of detecting tax fraud in Lithuania by utilizing data mining technologies. The researchers created models for segmentation, behavioral templates, risk assessment, and tax criminal detection. The findings proved the capacity of data mining tools to detect tax evasion and access confidential data, which can assist reduce revenue losses due to tax evasion. The study's findings can help scientists, professionals, and decision-makers anticipate tax fraud detection in developing countries.

The researchers offer a paradigm for identifying tax fraud in [38]. There are four modules in the framework:

Monitored Module: A tree-based model is used in this module to draw knowledge from the data. It uses labeled data to train the model in a supervised learning method. The objective is to identify data correlations and trends that may point to probable fraud.

Unsupervised Module: The unsupervised module is responsible for determining anomaly scores. It identifies patterns that deviate significantly from the norm or exhibit unusual behavior. These anomalies can be indicative of fraudulent activities.

Behavioral Module: The behavioral module calculates a taxpayer's compliance score. It assesses the taxpayer's historical behavior, such as past compliance with tax regulations, timely filing of returns, and acc Prediction Module: To ascertain the possibility of fraud for each tax return, the prediction module makes use of the outputs from the previous modules. To produce a thorough fraud prediction score, it incorporates the findings from the behavioral, unsupervised, and supervised modules. Accuracy of reported information. A low compliance score may suggest a higher likelihood of fraud.

The effectiveness of the framework was demonstrated by testing it on actual tax returns provided by the Saudi tax administration. The researchers evaluated its performance in detecting tax fraud based on the framework's outputs and compared them to known instances of fraud. The results showed the framework's ability to effectively identify potential cases of tax fraud.

Overall, this study presents a comprehensive framework that combines the techniques of supervised and unsupervised learning with behavioral analysis for detecting tax fraud. By integrating multiple modules, it provides a holistic approach to identifying potentially fraudulent activities in tax returns.

II. THE ALGORITHMS AND TECHNIQUES USED IN TAX AND FINANCIAL FRAUD DETECTION

In the previous papers the researchers used many ML and DL techniques and different kinds of learning supervised and un supervised learning as follow:

A. Naive Bayes (NB).
B. Decision tree.
C. Random Forest.
D. Neural network.
E. K-mean clustering.
F. Self-Organizing Map (SOM).

A. Naive Bayes (NB)

Each pair of features is presumed to be independent by the naive Bayes technique, which is based on the Bayes theorem. It works well and may be used for both binary and multi-category applications, such as text or document classification, spam filtering, and so on. The NB classifier can be used to build a reliable prediction model and classify noisy occurrences in the data. The primary advantage is that it takes less training data than more involved approaches, allowing for faster estimation of the parameters. However, because it makes such strong assumptions about the independence of features, its performance may be compromised. The most common NB classifier modifications are Gaussian, Complement, Multinomial, Categorical, and Bernoulli[38].

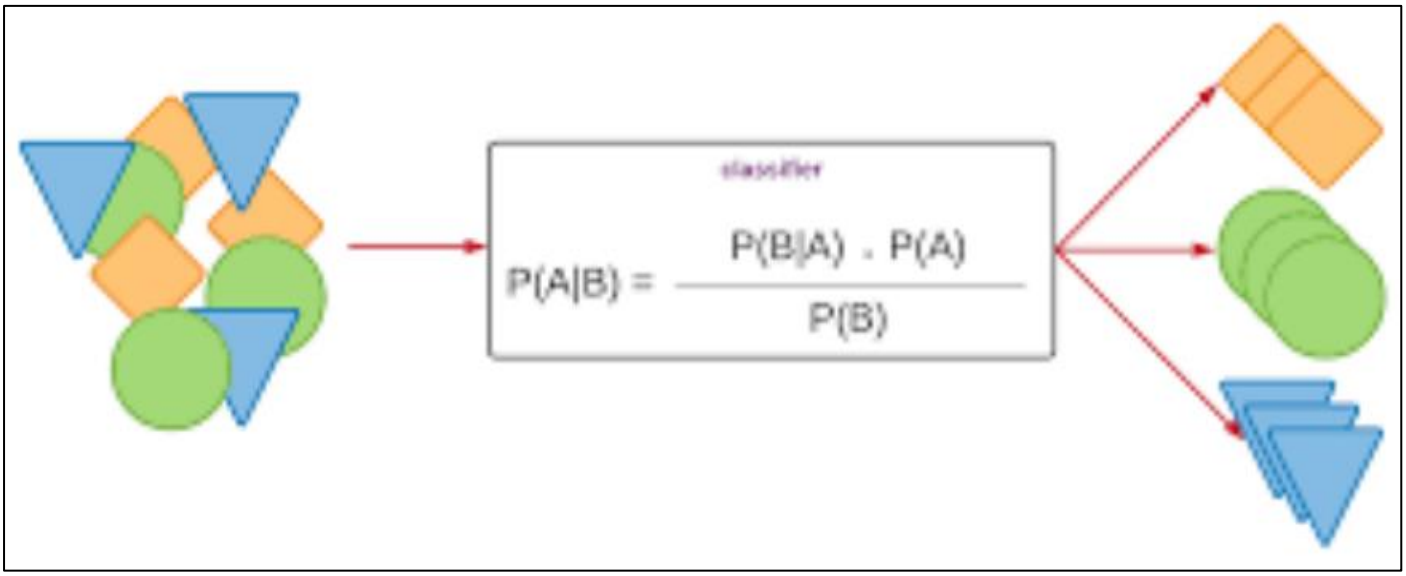

Fig 1 Naive Bayes

B. Decision Tree (DT)

Decision trees are a popular nonparametric supervised learning approach. It applied DT learning techniques to both classification and regression problems. The most prevalent DT algorithms are CART, ID3, C4.5, and regression. Furthermore, Sarker et al.'s newly developed Intrude Tree and Behav DT are effective in the relevant application fields of user behavior analytics and cybersecurity analytics, respectively. DT classifies occurrences by organizing the tree's nodes from root to leaf. Examining the tree's root node and moving along the branch that corresponds to the attribute value in order to sort the instances according to their defined features. Two typical splitting criteria are "gini" for the Gini impurity and "entropy" for achieving.

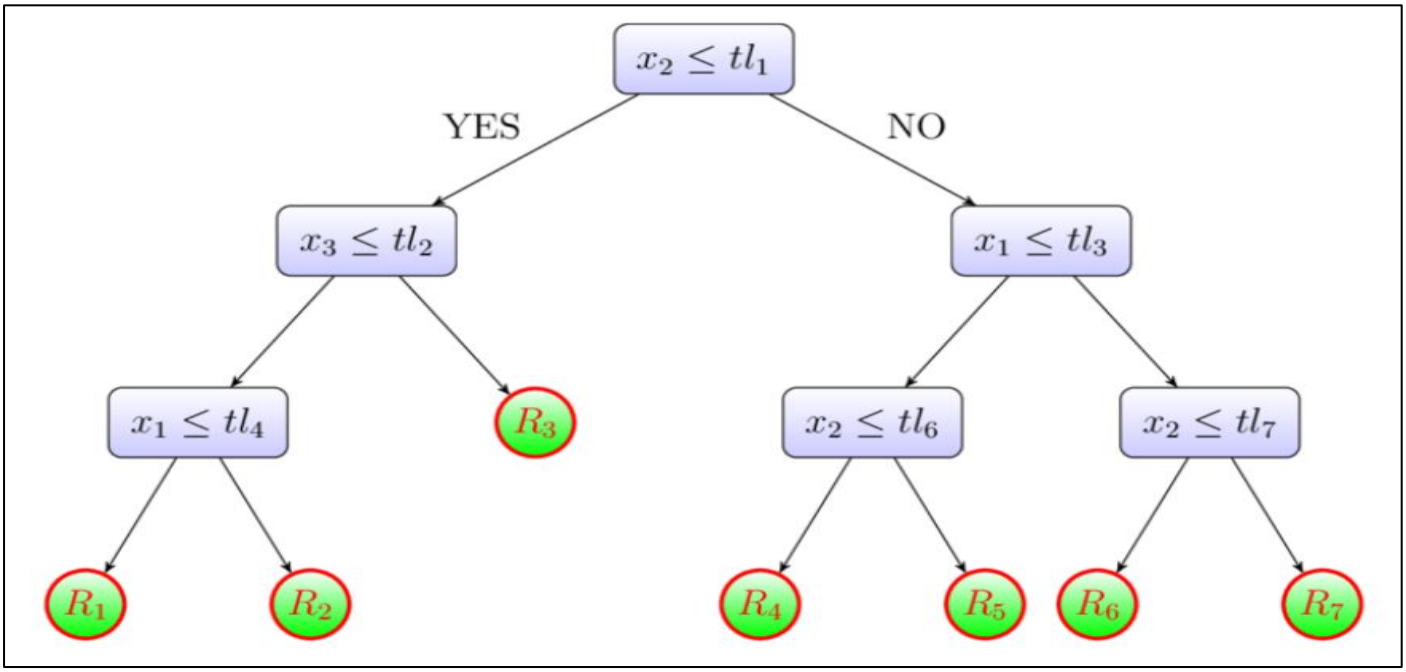

Fig 2 Decision Tree

C. Random Forest (RF)

A random forest classifier is an ensemble classification strategy that is utilized in a variety of machine learning and data science applications. This method employs "parallel ensembling," which parallelizes the fitting of many decision tree classifiers to different dataset subsamples and uses averages to reach the conclusion, final choice, or majority vote. As a result, it decreases overfitting while also improving prediction and control precision. As a result, an RF learning model with several decision trees outperforms a model with only one decision tree. Bootstrap aggregation (bagging) is paired with random feature selection to generate a collection of decision trees with controlled variation.

It may be used to tackle classification and regression problems, and it is effective with both continuous and categorical data [39]. A random forest, an ensemble approach that generates multi-decision trees, is a variant of the Decision Tree. In a random forest, each decision tree is built from a subset of features rather than every feature, which would necessitate using every feature. The final class prediction is based on a majority vote among the trees, and the trees forecast the class outcome[40]

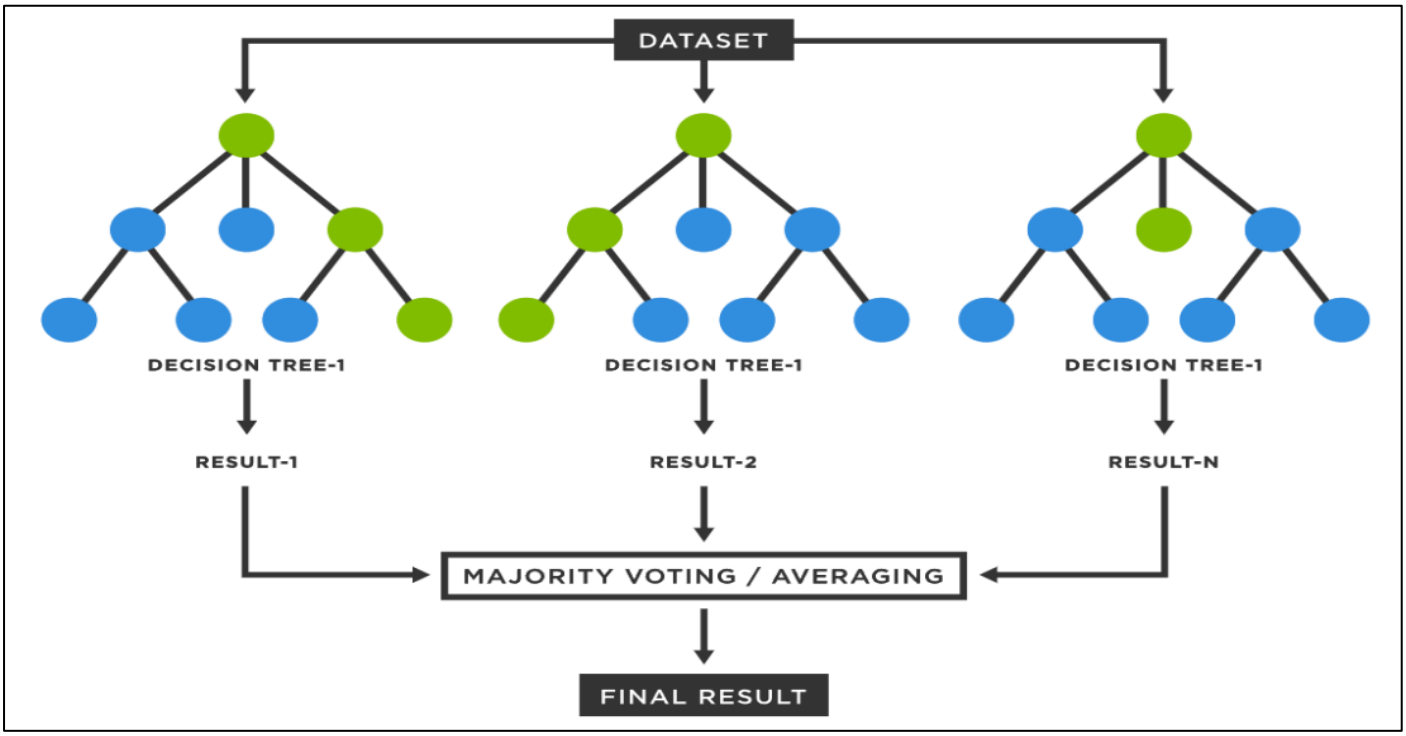

Fig 3 Random Forest

D. K-means Clustering

Clustering using K-means. When datasets are scattered, the resilient, rapid, and simple K-means clustering method yields accurate results. This method distributes a cluster's data points in a way that minimizes the squared distance between the data points and the cluster's centroid. To put it another way, the K-means algorithm calculates the k number of centroids and then allocates each data point to the nearest cluster possible. The findings may be unequal because the selection process begins with randomly chosen cluster centers. The K-means clustering approach is susceptible to outliers because extreme numbers can easily change the mean[39].

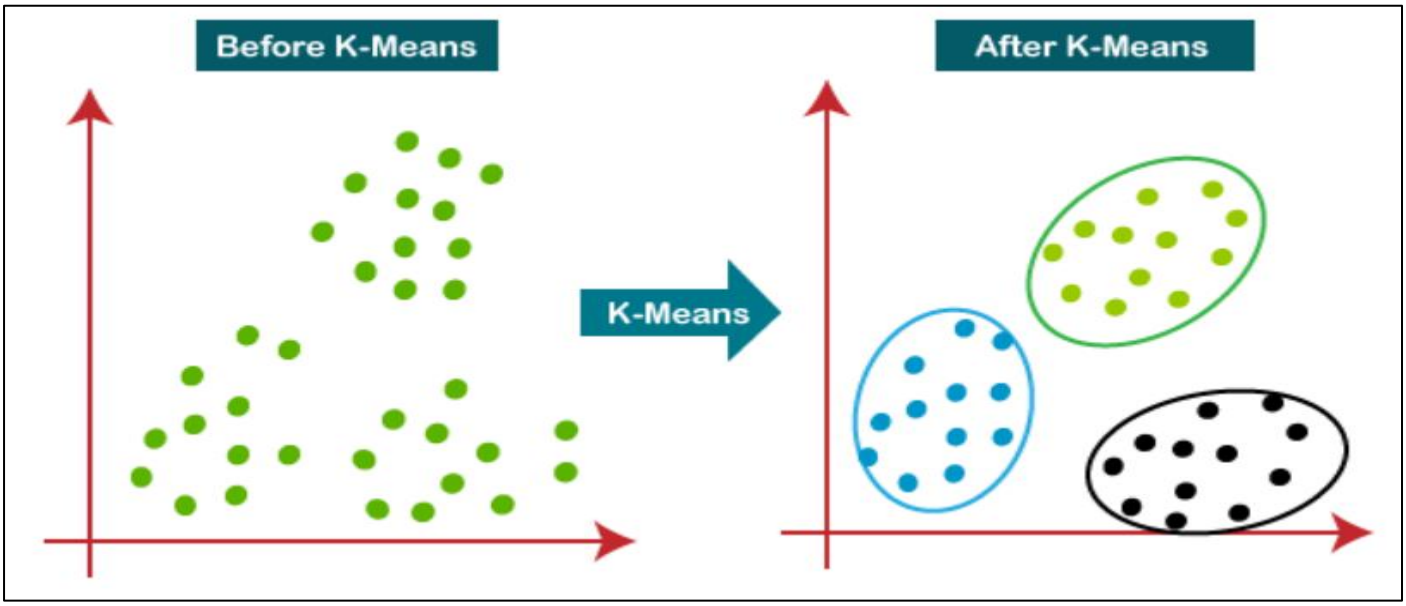

Fig 4 K-means Clustering

E. Artificial Neural Network and Deep Learning

A broad family of artificial neural networks (ANN) that rely on machine learning and representation learning approaches includes deep learning. Deep learning offers a computational framework for data learning by combining many processing levels, including input, hidden, and output layers. Deep learning's primary advantage is that it outperforms other methods in many circumstances, especially when learning from massive datasets.

Deep learning methods commonly utilized include Convolutional Neural Networks (CNN), Long Short-Term Memory Recurrent Neural Networks (LSTM-RNN), and Multi-Layer Perceptron (MLP) [39, 40].

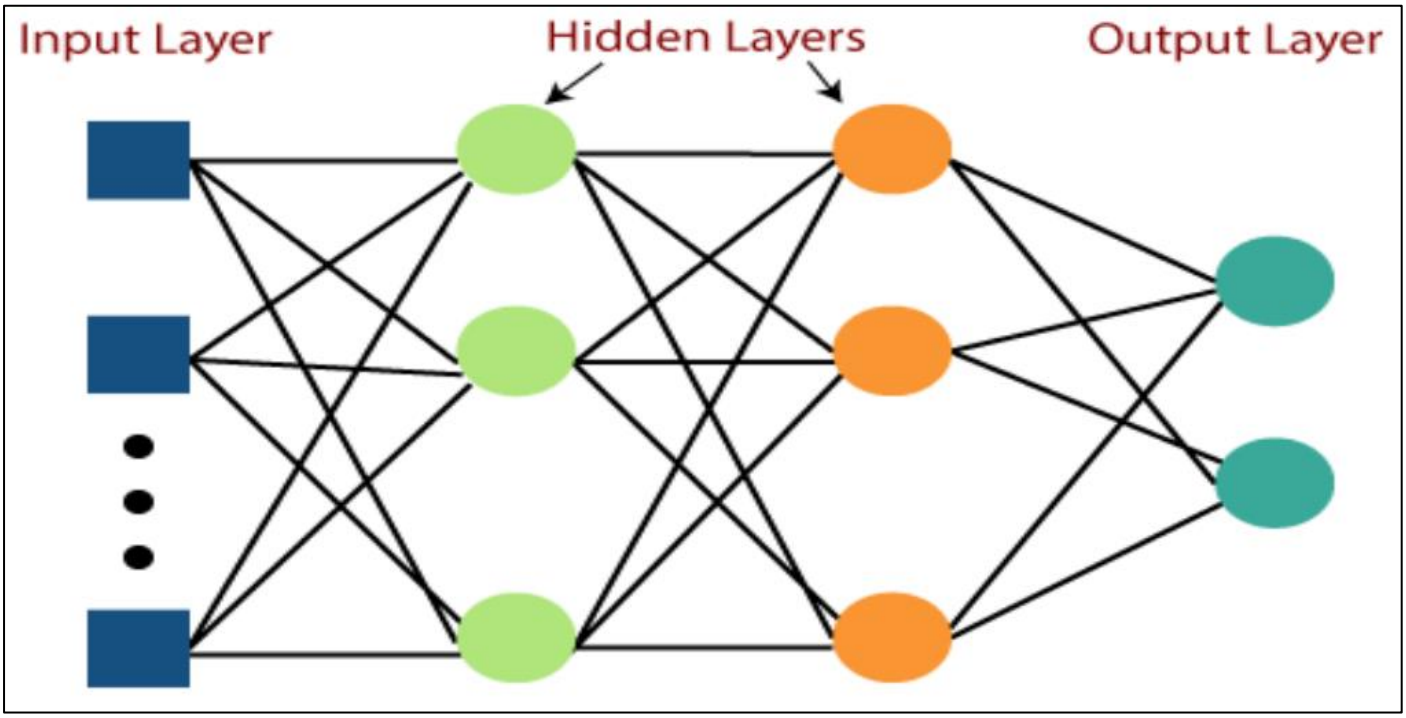

Fig 5 Multilayer Perceptron

ISSN No:-2456-2165

F. A Self-Organizing Map (SOM)

A self-organizing map (SOM) is an artificial neural network (ANN) trained using unsupervised learning to construct a discretized, low-dimensional (usually two-dimensional) representation of the training samples' input space. This representation, known as a map, serves to reduce dimensionality. Instead of employing error corrective learning (e.g., backpropagation with gradient descent), as other ANNs do, SOMs use competitive learning, which uses a neighborhood function to preserve the input space's topological features[41].

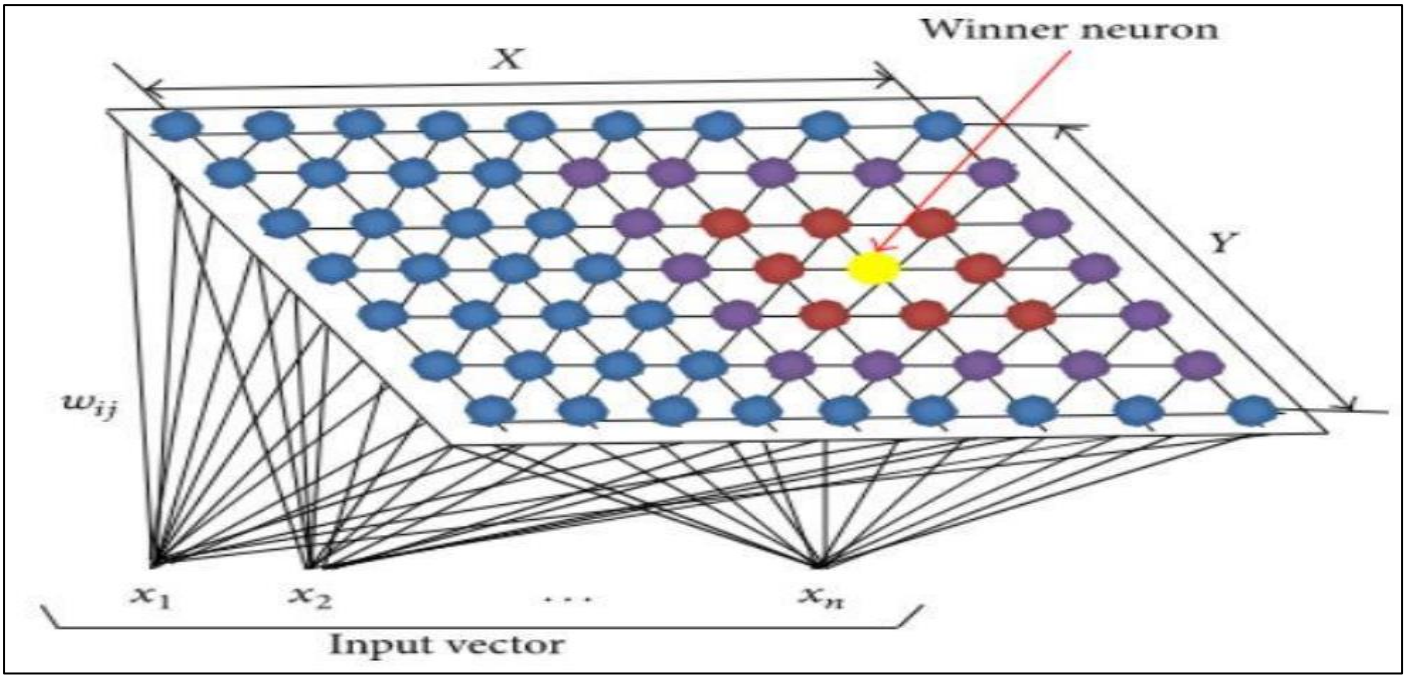

Fig 6 Self-Organizing Map (SOM)

G. Autoencoder

An autoencoder is made consisting of an encoder and a decoder, both symmetrical. according to Figure 3. The encoder extracts features from the raw data. The decoder reconstructs the data from the features it has extracted. During training, the divergence between the encoder input and decoder output gradually decreases. When the decoder successfully reconstructs the data from the features collected. It indicates that the encoder's features, which represent the data's content, are those features. It is critical to understand that no part of this procedure requires monitored information. There are other varieties of autoencoders, such as sparse and sparsely noisy ones[7].

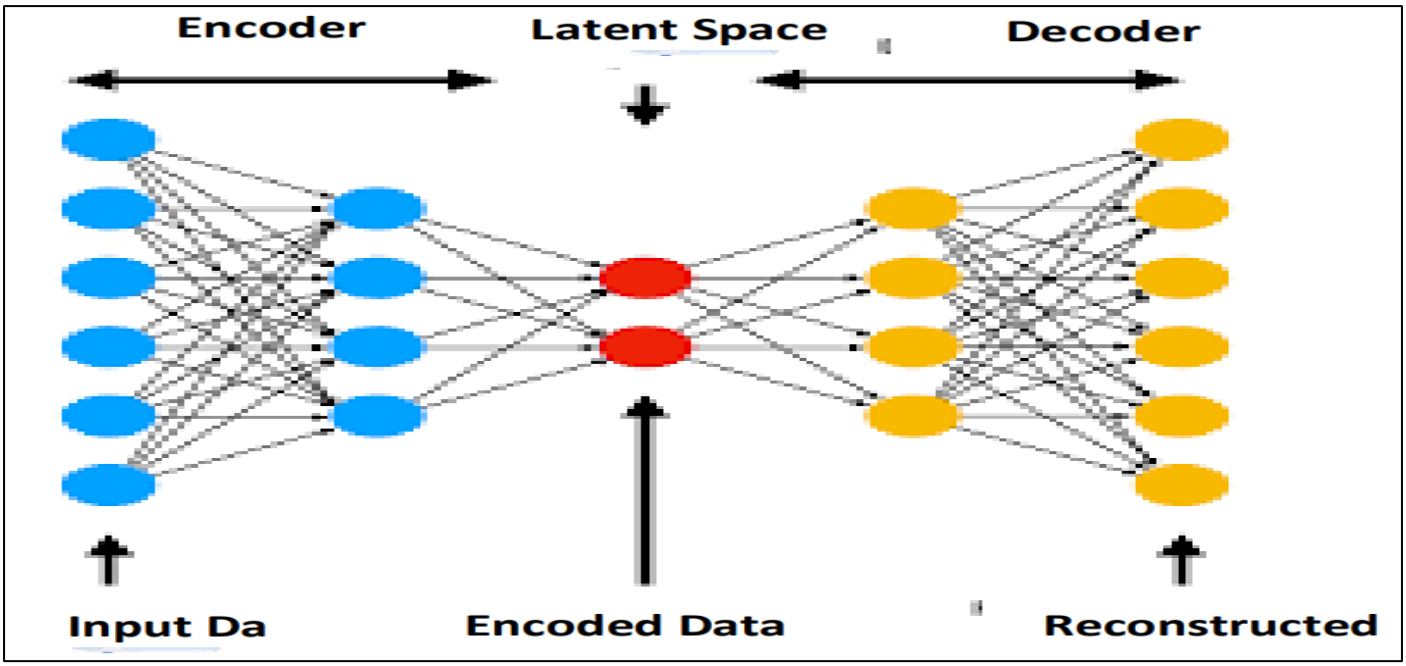

Fig 7 An Autoencoder

ISSN No:-2456-2165

Table1 Summary of Previous Studies to Trying Detecting Tax and Financial Evasion using Artificial Intelligence and Machine Learning

Ref	Year	Technology	Study description
[4]	2018	Supervised ML techniques	The researchers proposed a system that allows tax authorities to prioritize audits based on data rather than previously classified data.
[5]	2018	Compositional CNN-RNN model framework	Based on the official transaction code seen on tax bills, researchers proposed a compositional CNN-RNN model architecture with an attention mechanism for describing transaction behavior.
[6]	2019	AI and Big Data	The paper looked into the issues of tax planning in the age of artificial intelligence and big data
[7]	2019	Advanced methodologies in big data analytics, as well as the development of artificial intelligence using machine learning	In this scientific research project, the Tax Administration and the University of Novi Sad's Faculty of Sciences worked together to create algorithms that use advanced big data analytics techniques and machine learning to create artificial intelligence in order to identify the risk of tax evasion. The project's initial results were featured in the article.
[8]	2019	Unsupervised conditional adversarial network (UCAN)	The researchers suggested the unsupervised conditional adversarial network (UCAN) as a universal architecture for detecting tax evasion. This is the first solution for addressing audit tasks in unlabeled target domains via inter-region transfer.
[9]	2019	Machine Learning improved predictive tools by utilizing Multilayer Perceptron neural network (MLP) models.	They used Multilayer Perceptron neural network (MLP) models and powerful machine learning predictive approaches to aid in the detection of tax fraud using personal income tax returns (IRPF, in Spanish) filed in Spain.
[10]	2019	Hierarchical clustering and descriptive statistics	This study has a dual research goal. It first sought to understand how SMEs perceived the current situation and the necessary steps to cut back on the corresponding red tape. Secondly, it made an effort to establish a link between tax burdens and entrepreneurship using hierarchical cluster analysis and descriptive statistics.
[11]	2019	Mining techniques	The study aimed to use mining techniques to provide solutions to financial fraud
[12]	2020	Q-learning combined with recent Deep Reinforcement Learning advancements	The problem of figuring out a risk-averse, self-interested tax entity's approach was investigated in the paper. They were able to get approximations of the answers by combining Q-learning with new discoveries in Deep Reinforcement Learning .
[13]	2020	An expert system on tax evasion and a performance model.	The study used a performance model and an expert system in the area of tax evasion to provide an abstract solution.
[14]	2020	The recognition of some predictive modeling techniques	It was investigated how tax authorities could benefit from their operational data. The purpose of this work was to uncover machine learning algorithms that are good at detecting a specific form of fraud[10].
[15]	2020	SVM, a type of supervised ML technology	To leverage the data to identify the fraud occurring at the bank, the researchers used data mining methods.
[16]	2020	ML approaches	Using a dataset of tax defaults and non-defaults at Finnish limited liability businesses, the proposed prediction system was validated.
[17]	2020	ML methods: unsupervised and semi-supervised.	The Norwegian Tax Administration is interested in understanding how to choose unsupervised and semi-supervised machine learning approaches that are effective at detecting abnormal tax returns. Additionally, they have looked into the detection techniques and how the various dataset characteristics affect how well they work.
[18]	2020	Unsupervised anomaly detection techniques	The researchers' key argument in this work was that sample selection bias causes the small number of labeled data points (known fraud/legal cases) in the tax fraud detection domain to not be representative of the population as a whole.
[19]	2020	AI, Big Data, and Blockchain	In this study, the existing research on audit and tax in relation to developing technologies was reviewed. In addition, presents a future research agenda.
[20]	2020	Model-free RL in AI-based	
[21]	2020	This approach combined suitably, through a variety of DM and visual analytics techniques	To help tax administrations uncover tax avoidance and evasion, this study introduced an innovative approach named MALDIVE. .

[22]	2020	logical generalization, expert evaluation, statistical analysis	According to the study, identifying tax evasion is one of tax administration's primary responsibilities. It also developed a methodology for predicting the risk of tax evasion using both quantitative and qualitative data.
[23]	2020	Big data strategy	The study's goal was to simulate expats' tax-related behavior.
[24]	2020	ML Random Forest	The key driving force behind this endeavor was the use of machine learning to enhance decision-making in fiscal audit plans for service taxes for the municipality of Sao Paulo.
[25]	2021	Artificial Intelligence (AI)	Based on criteria including the complexity of the tax system, tax education, legal sanctions, and relationships with the tax authorities, researchers seek to determine the function of AI in the Indian taxation system.
[26]	2021	A hybrid ML-based approach DT model, K-means	The researchers developed a novel strategy for managing the risk of tax fraud utilizing a hybrid approach based on machine learning that has numerous characteristics.
[27]	2021	Use tools from ML and network science	The researchers proved that utilizing machine learning and network science technologies, it is possible to automatically detect tax evasion tendencies that are comparable to those previously recognized by humans.
[28]	2021	Different classifiers RFKNN, NN, and SVM.	This investigation introduced a machine learning-based technique. that can determine if a corporation is engaging in fraud or not. Four distinct classifiers were used to analyze tax and financial data from diverse organizations.
[29]	2021	SVM with an improved ID3 decision tree is used as a hybrid approach, and also for improving the accuracy and performance of the multilayer perceptron neural networks and genetic algorithm.	The present study's objective was to offer a hybrid method for detecting financial fraud that combines a support vector machine with an enhanced ID3 decision tree.
[30]	2021	An artificial neural network	The researchers cited fraud detection in recommendation systems, consumer behavior forecasting, picture and speech recognition, colorization of black and white films, and accounting fraud detection as some examples of the wide range of applications in which ML is used.
[31]	2021	Clustering, classification, and regression.	.In this paper, the researchers proposed a state of art on different fraud techniques, also, they were forced to continually improve fraud detection systems
[32]	2022	Gradient Boosting Machine Learning Tools	The researchers created a machine learning-based fraud prediction model, employing gradient boosting as their first choice.
[33]	2022	particle swarm optimization (IPSO) algorithm	Robust machine learning methods are used in this study to solve the identification of the tax evasion problem. In this study, we established the best weight and ideal parameters for support vector machine (SVM) classifiers by optimizing the multilayer perceptron neural network using an advanced optimization of particle swarms (IPSO) technique.
[34]	2022	. The researchers developed three machine learning models: decision tree, random forest, and logistic regression.	The research in this study suggested integrating trigger data from taxpayers as a decision support system along with machine learning-based predictive analytics to take advantage of realizing latent tax opportunities. This study provided more specific predictive analytics algorithms that can accurately identify which potential taxpayers are most likely to pay their fair share.
[35]	2022	feature extraction and the random algorithm	The objective of this study establish a fraud detection system for tax
[36]	2023	Supervised machine learning models include GaussianNB,XGBoost, Random Forest, Decision Tree, and Logistic Regression.	The purpose of this study was to identify symptoms of tax fraud using the most reliable supervised machine learning model. .
[37]	2023	Data Mining Techniques	This study's main objective is to better detect tax fraud by using data mining tools to investigate the effects of wealth in Lithuania.
[38]	2023	A supervised module, an unsupervised module, a behavioral module, and a prediction module	The work done in this study was focused on suggesting a framework for detecting tax fraud

III. CONCLUSION

Many governments around the world rely heavily on taxes ,but tax administrations are facing many problems and challenges, especially with regard to tax fraud by taxpayers therefore, tax administrations around the world seek to develop their tax systems supported by machine learning such as clustering methods like k-mean, Self-Organizing Map (SOM), and classification methods as decision tree, random forest, naive bayes, and neural network, to increase its efficiency and eliminate fraud and tax evasion.

REFERENCES

[1]. S. Mills, "Chapter 1 Taxation Principles and Theory," Found. Tax. Law, no. 1908, 1925,[Online]. Available: https://www.oup.com.au/data/assets/file/0014/132062/9780190318529_SC.pdf

[2]. http:\\ar,wikipedia.or

[3]. Hartnett D," Tax Administration Challenges in Developing Countries", 4/4/ 2016.

[4]. D. De Roux, B. Pérez, A. Moreno, M. Del Pilar Villamil, and C. Figueroa, "Tax fraud detection for under-reporting declarations using an unsupervised machine learning approach," Proc. ACM SIGKDD Int. Conf. Knowl. Discov. Data Min., pp. 215–222, 2018, doi: 10.1145/3219819.3219878.

[5]. J. Yu, Y. Qiao, K. Sun, H. Zhang, and J. Yang, "Poster: Classification of transaction behavior in tax invoices using compositional CNN-RNN model," *UbiComp/ISWC 2018 - Adjun. Proc. 2018 ACM Int. Jt. Conf. Pervasive Ubiquitous Comput. Proc. 2018 ACM Int. Symp. Wearable Comput.*, pp. 315–318, 2018, doi: 10.1145/3267305.3267597.

[6]. J. Shan, "Optimization Strategy of Tax Planning System in the Context of Artificial Intelligence and Big Data," in *Journal of Physics: Conference Series*, 2019, vol. 1345, no. 5, doi: 10.1088/1742-6596/1345/5/052006.

[7]. J. Atanasijević, D. Jakovetić, N. Krejić, N. Krklec-Jerinkić, and D. Marković, "Using big data analytics to improve the efficiency of tax collection in the Tax Administration of the Republic of Serbia," *Ekon. Preduz.*, vol. 67, no. 1–2, pp. 115–130, 2019, doi: 10.5937/ekopre1808115a.

[8]. R. Wei, B. Dong, Q. Zheng, X. Zhu, J. Ruan, and H. He, "Unsupervised Conditional Adversarial Networks for Tax Evasion Detection," *Proc. - 2019 IEEE Int. Conf. Big Data, Big Data 2019*, pp. 1675–1680, 2019, doi: 10.1109/BigData47090.2019.9005656.

[9]. D. Rodr, "Tax Fraud Detection through Neural Networks : An Application Using a Sample of Personal Income Taxpayers," 2019, doi: 10.3390/fi11040086.

[10]. "Tax-Related Burden on SMEs in the European Union : The Case of Slovenia Dejan Ravšelj Polonca Kovač Aleksander Aristovnik," vol. 2117, pp. 69–79, 2019, doi: 10.2478/mjss-2019-0024.

[11]. N. Sael and F. Benabbou, "ScienceDirect ScienceDirect Performance of machine learning techniques in the detection of Performance of machine learning techniques in the detection of financial frauds financial frauds," *Procedia Comput. Sci.*, vol. 148, no. Icds 2018, pp. 45–54, 2019, doi: 10.1016/j.procs.2019.01.007.

[12]. N. D. Goumagias, D. Hristu-varsakelis, and Y. M. Assael, "Using deep Q-learning to understand the tax evasion behavior of risk-averse firms," pp. 1–29,2020.

[13]. I. S. Conference and E. Sarajevo, "Expert Systems as a Means in Detecting Tax Evasion," no. September, pp. 18–20, 2020.

[14]. A. Z. Adamov, "Machine Learning and Advanced Analytics in Tax Fraud Detection," no. October 2019, 2020, doi: 10.1109/AICT47866.2019.8981758.

[15]. C. Reviews, "An income tax fraud detection using AI," vol. 7, no. 16, pp. 119–124, 2020.

[16]. M. Z. Abedin, H. Mohammad, D. Science, N. Science, and G. Bishwabidyalay, "Tax Default Prediction using Feature Transformation-Based Machine Learning," no. December, 2020, doi: 10.1109/ACCESS.2020.3048018.

[17]. N. Gedde, I.-S. Sandvik, and J. Andersson, "Unsupervised Machine Learning on Tax Returns Investigating Unsupervised and Semisupervised Machine Learning Methods to Uncover Anomalous Faulty Tax Returns",2020.

[18]. V. Jellis, M. David, P. Bruno, J. Vanhoeyveld, D. Martens, and B. Peeters, "This item is the archived peer-reviewed author-version of : Value-added tax fraud detection with scalable anomaly detection techniques Reference :," vol. 86, 2020.

[19]. O. F. Atayah, "Audit and tax in the context of emerging technologies : A retrospective analysis , current trends , and future opportunities," vol. 21, no. November 2020, pp. 95–128, 2021, doi: 10.4192/1577-8517-v21.

[20]. S. Zheng *et al.*, "The AI Economist: Improving Equality and Productivity with AI-Driven Tax Policies," Apr. 2020, [Online]. Available: http://arxiv.org/abs/2004.13332.

[21]. W. Didimo, L. Grilli, G. Liotta, F. Montecchiani, and D. Pagliuca, "Combining Network Visualization and Data Mining for Tax Risk Assessment," pp. 16073–16086, 2020.

[22]. A. Musayev and M. Gazanfarli, "Modeling the Probability of the Detection Process of Tax Evasion Taking into Account Quality and Quantity Indicators," *Asian J. Econ. Bus. Account.*, vol. 18, no. 4, pp. 28–37, 2020, doi: 10.9734/ajeba/2020/v18i430291.

[23]. A. H. Miller and C. Republic, "Using Database Approach , With Big Data And Unsupervised Machine Learning To Model Tax Behavior In The Expatriate Community," no. October, 2020.

[24]. A. Ippolito and A. C. G. Lozano, "Tax crime prediction with machine learning: A case study in the municipality of São Paulo," *ICEIS 2020 - Proc. 22nd Int. Conf. Enterp. Inf. Syst.*, vol. 1, no. Iceis, pp. 452–459, 2020, doi: 10.5220/0009564704520459.

[25]. A. Rathi, S. Sharma, G. Lodha, and M. Srivastava, "A Study on Application of Artificial Intelligence and Machine Learning in Indian Taxation System," no. February, 2021, doi: 10.17762/pae.v58i2.2265.
[26]. J. Atanasijevi, "Tax Evasion Risk Management Using a Hybrid Unsupervised Outlier Detection Method," no. 451, p. 30, 2021.
[27]. M. Zumaya *et al.*, "Identifying Tax Evasion in Mexico with Tools from Network Science and Machine Learning," *Underst. Complex Syst.*, pp. 89–113, 2021, doi: 10.1007/978-3-030-81484-7_6.
[28]. J. P. A. Andrade *et al.*, "A Machine Learning-based System for Financial Fraud Detection," pp. 165–176, 2021, doi: 10.5753/eniac.2021.18250.
[29]. A. Javadian, A. Ali, P. Aghajan, and M. Hosseini, "A Hybrid Model Based on Machine Learning and Genetic Algorithm for Detecting Fraud in Financial Statements," *J. Optim. Ind. Eng.*, vol. 14, no. 2, pp. 169–186, 2021, doi: 10.22094/JOIE.2020.1877455.1685.
[30]. X. Zhang, "Construction and Simulation of Financial Audit Model Based on Convolutional Neural Network," *Comput. Intell. Neurosci.*, vol. 2021, pp. 1–11, 2021.
[31]. M. Vlad and S. Vlad, "The Use of Machine Learning Techniques in Accounting . A Short," *J. Soc. Sci. Fascicle*, vol. 4, pp. 1–5, 2021.
[32]. V. Baghdasaryan, H. Davtyan, and A. Sarikyan, "Improving Tax Audit Efficiency Using Machine Learning : The Role of Taxpayer ' s Network Data in Fraud Detection Improving Tax Audit Efficiency Using Machine Learning : The Role of Taxpayer ' s Network Data in Fraud Detection," *Appl. Artif. Intell.*, vol. 00, no. 00, pp. 1–23, 2022, doi: 10.1080/08839514.2021.2012002.
[33]. H. Mojahedi, A. Babazadeh Sangar, and M. Masdari, "Towards Tax Evasion Detection Using Improved Particle Swarm Optimization Algorithm," Math. Probl. Eng., vol. 2022, 2022, doi: 10.1155/2022/1027518.
[34]. J. Perbendaharaan, K. Negara Dan Kebijakan Publik, R. David Febriminanto, and M. Wasesa, "Indonesian Treasury Review Machine Learning for Predicting Tax Revenue Potential," Keuangan Negara dan Kebijakan Publik, 2022. [Online]. Available: www.pajak.com
[35]. A. Menon, D. Khator, D. Prajapati, and A. Ekbote, "IPL Prediction Using Machine Learning," Indian J. Comput. Sci., vol. 7, no. 3, pp. 274–276, 2022, doi: 10.17010/ijcs/2022/v7/i3/171267
[36]. B. F. Murorunkwere, D. Haughton, J. Nzabanita, and I. Kabano, "Predicting tax fraud using supervised machine learning approach," African J. Sci. Technol. Innov. Dev., vol. 0, no. 0, pp. 1–12, 2023, doi: 10.1080/20421338.2023.2187930.
[37]. T. Ruzgas, L. Kižauskienė, M. Lukauskas, E. Sinkevičius, M. Frolovaitė, and J. Arnastauskaitė, "Tax Fraud Reduction Using Analytics in an East European Country," Axioms, vol. 12, no. 3, p. 288, Mar. 2023, doi: 10.3390/axioms12030288.
[38]. N. Alsadhan, "A Multi-Module Machine Learning Approach to Detect Tax Fraud," Comput. Syst. Sci. Eng., vol. 46, no. 1, pp. 241–253, 2023, doi: 10.32604/csse.2023.033375.
[39]. I. Sadgali, N. Sael, and F. Benabbou, "Performance of machine learning techniques in the detection of financial frauds," Procedia Comput. Sci., vol. 148, no. Icds 2018, pp. 45–54, 2019, doi: 10.1016/j.procs.2019.01.007.
[40]. I. H. Sarker, "Machine Learning: Algorithms, Real-World Applications and Research Directions," SN Comput. Sci., vol. 2, no. 3, 2021, doi: 10.1007/s42979-021-00592-x.
[41]. A. . Shujaaddeen, F. M. . Ba-Alwi, and G. Al-Gaphari, "A New Machine Learning Model for Detecting levels of Tax Evasion Based on Hybrid Neural Network ", Int J Intell Syst Appl Eng, vol. 12, no. 11s, pp. 450–468, Jan. 2024.
[42]. T. Germano, "Self Organizing Maps @ davis.wpi.edu," p. 4, 1999, [Online]. Available: http://davis.wpi.edu/~matt/courses/soms/.
[43]. A. M. Ozbayoglu, M. U. Gudelek, and O. B. Sezer, "Deep learning for financial applications: A survey," *Appl. Soft Comput. J.*, vol. 93, pp. 1–52, 2020, doi: 10.1016/j.asoc.2020.106384.