International Journal of Innovative Engineering Applications

Journal homepage: https://dergipark.org.tr/ijiea

A SUSPICIOUS COMPANY DETECTION SYSTEM BASED ON MACHINE LEARNING WITH A BIBLIOMETRIC ANALYSIS FOR ACCOUNTING AND AUDITING

Yunus Korkmaz[*1], *Sadık Serçek*[2], *Mukaddes Korkmaz*[3]

[1]*Dicle University, Faculty of Engineering, Dept. of Computer Engineering, Diyarbakir, Turkey.*
[2]*Dicle University, Faculty of Economics and Administrative Sciences, Dept. of Business Administration, Diyarbakir, Turkey.*
[3]*Accounting Unit, Dicle Electricity Distribution Inc., Diyarbakir, Turkey.*

Abstract

Original scientific paper

In accounting, auditing involves reviewing the financial statements and records of an organization by independent auditors in order to guarantee their precision and comply with relevant statutes and rules. Audits are typically conducted by external or internal auditors who review financial statements, assess internal controls, and verify the precision of financial data. The aim of audit is to provide confidence to stakeholders that the pecuniary data presented by the organization is reliable and trustworthy. In this work, role of the artificial intelligence (AI) was examined with a comprehensive literature review for auditing in accounting area. In technical aspect of this work, we built a fraudulent company detection system based on machine learning (ML) classification algorithms like Decision Tree (DT), Bagged Tree, Ensemble KNN (K-Nearest Neighbors), Linear Support Vector Machines (SVM) etc. Decision Tree and Bagged Tree reached AUC value of 1 which means that they are the perfect classifiers. The AUC values of 0.99, 0.99, 0.9998 and 0.9963 were obtained for Linear SVM, Logistic Regression, Subspace KNN and Naïve Bayes, respectively. The result proved that any machine learning based solution can help auditors to easily have an idea about the fraudulent companies before on-site auditing.

Keywords: *Machine learning, suspicious company, decision tree, bibliometric analysis, accounting and auditing.*

1 Introduction

The increasing application of the AI in accounting causes automation to become more involved in auditing processes. This situation changes the way accounting and auditing professionals do business and their requirements. The proliferation of artificial intelligence and automation can perform repetitive and time-consuming tasks quickly and accurately, while offering accounting and auditing professionals the opportunity to concentrate on more imaginative and beneficial activities like analytical thinking, strategic planning and consulting. Therefore, the outlook for the accounting and auditing industry hinges on the skills of its practitioners to understand and interact with technology. Professionals must learn how to use technology to improve customer service and create value by transforming their business. With the advancement of technology, automation in the labor market is expected to increase. However, certain professions are observed to be resistant to automation. Approximately 47% of the workforce in the United States are waited to be exposed to the effects of automation [1]. This suggests that within white-collar occupations, occupations such as auditors, accountants, budget analysts, and loan officers may be more susceptible to automation, while occupations such as surgeons, therapists, and CEOs are thought to be more resistant. In blue-collar sectors, occupations that rely on physical strength are expected to be more exposed to the effects of automation. However, professions such as computer scientists, lawyers, writers, editors and designers are observed to be more resistant to automation [2]. These findings are important for assessing the consequences of labor market automation and understanding future job trends. It is also emphasized that changes in the labor market should be considered from a broader perspective, not just specific professions. By examining the impacts of the AI in area of accounting and auditing, this paper aims to present how these disciplines can be affected by artificial intelligence technologies and provides information regarding the potential advantages that can be obtained and the potential risks regarding the study. Additionally, importance and significance of the AI in digital conversion in accounting and auditing is emphasized. The motivation for situating this study within the accounting and auditing domain stems from the growing need for data-driven tools that can support auditors in identifying potentially suspicious companies at an early stage. Unlike prior research that mostly focuses on isolated fraud-detection models or narrow ratio-based analyses, this study integrates a bibliometric review with a comparative machine-learning framework. This combined approach provides a more comprehensive and methodologically distinct contribution to the auditing literature.

*Corresponding author.

E-mail address: yunus.korkmaz@dicle.edu.tr (Y. Korkmaz

Received 23 September 2025; Received in revised form 21 November 2025; Accepted 25 November 2025

Doi: https://doi.org/10.46460/ijiea.1789267

1.1 Artificial Intelligence (AI)

John McCarthy took an important step in 1956 by introducing the term Artificial Intelligence, a technology that can be programmed to empower computers or robots to execute tasks typically undertaken by human being. This introduction marked the beginning of the history of Artificial Intelligence, and from this period onwards, scientists, mathematicians and philosophers began to adopt the concepts of artificial intelligence [3]. Artificial Intelligence, as defined by McCarthy, is the science and practice of developing smart devices. It is a kind of software application focused on creating and operating smart agents having ability to understanding their environment and making decisions to optimize their likelihood of success. AI involves the ability to hold multiple ideas simultaneously and operate accordingly. The field aims to incorporate learning from past experiences, decision-making, reasoning, deduction and quick reaction capabilities. Furthermore, artificial intelligence must be able to prioritize and navigate complexity and uncertainty. Early discussions about AI in the 1950s not only highlight its longstanding existence but also raise ethical questions related to effects of machine capability over decision-making of people [4].

Artificial intelligence is a model implemented by machines by imitating human characteristics such as reasoning, inferring, generalizing and learning [5]. Great progress has been made in the area of machine learning (ML), and the reasons for this include scientific developments with cheap computer power and the use of big data [6]. Artificial intelligence can be considered an umbrella term that includes “Big Data” approaches and sophisticated machine learning methods. Artificial intelligence is utilized to analyze dataset and predict future outcomes [7]. It consists of subcomponents like deep learning, image processing, machine learning, natural language processing, robotics and automatic decision making [8-10]. The process, which started with benefiting from the artificial intelligence tools in the field of accounting, may evolve into complete automation in the future, and this may also bring ethical issues. Artificial Intelligence refers to software and hardware that have human-like thinking and decision-making capabilities through methods like deep learning, natural language processing and machine learning. These modules can make decisions based on data and information, improve their processes and understand human language. Some techniques commonly used among artificial intelligence systems can be categorized as following sub-titles.

1.1.1 Machine Learning (ML)

It is an artificial intelligence technology based on data, information and knowledge. The ML can augment the decisions based on human behavior, interaction patterns, trends, and relationships between behaviors with new data inputs. It can also improve its own processes by recoding old decisions with new data. When an image is shown, the machine can identify what the image is with the data it acquired [11].

1.1.2 Deep Learning (DL)

It is a subcomponent that enables learning complex patterns and relationships. It can show high performance on large data sets. It is a part of machine learning which refers how hardware learns to judge benefiting architecture inspired by the people’s brain. It is an algorithm that has many inputs and outputs and contains many layers. It focuses on producing meaningful results by processing previous data. It can especially contribute to audit processes and management decision processes through the processing of financial statement data [12].

1.1.3 Image Processing

It enables information extraction by performing various operations on digital images. The main purpose in this field is to analyze image data and perform tasks such as detecting patterns, recognizing objects, and recognizing faces. Image processing tries to simulate the visual ability of the computer, so that automatic decisions can be made and data can be analyzed [14].

1.1.4 Robotic Process Automation (RPA)

It is an application that facilitates the development and management of software robots that copy human movements. These robots can perform tasks that humans can do faster, differently and more consistently. In this way, it allows the staff to prioritize more worthy activities by taking over mundane duties [15].

1.1.5 Neural Networks (NN)

It is a computer software system that imitates the structure of people’s thinking and can change its structure to better perform the task it is learning. This system is used in risk evaluation, that is an important component within audit process [16].

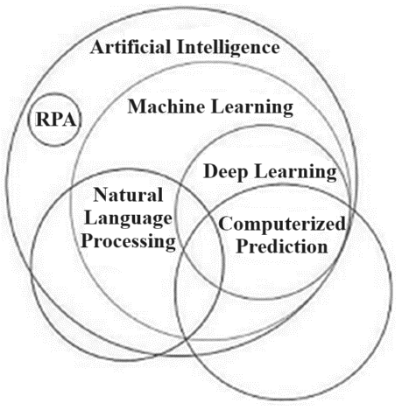

Figure 1. Relation of the AI with its sub-categories.

Figure 1 shows the relationship between the subsets with artificial intelligence [17] and also with robotic process automation (RSO), a popular technology solution that mimics human actions. Understanding the relationships between subsets of AI and the importance of technologies such as robotic process automation can provide efficiency and competitive advantage in business.

A correct understanding of these relationships can help companies take more conscious and effective steps in their strategic decisions.

1.2 The AI in Accounting

Accountants have a crucial role in meticulously preparing and verifying a firms' financial data to guarantee their precision and adherence to tax laws. It is also necessary to verify the accuracy of records and to see that tax payments are made promptly and in correct amounts globally. With the rapid pace of AI development, it is anticipated that robots will increasingly perform a wide array of tasks across industries, involving low-level production, transactions, and even accounting, possibly replacing human workers in many roles. Consequently, basic accounting practitioners are among those who will be significantly affected by these changes brought about by AI. [18]. In the coming two decades, accountants will experience a significant shift in their roles, requiring them to adeptly combine their expertise with the use of advanced technologies [19]. The presence of technology can be intimidating for new accountants, particularly those from the millennial generation, who may feel threatened due to their lack of skills and knowledge in operating these technologies [20]. Companies such as Deloitte, PwC, Earnst & Young and KPMG, the big four accounting firms, use artificial intelligence technologies in accounting and finance fields. Deloitte uses an artificial intelligence application called "Kira Systems" to improve security activities [21]. PwC facilitates the preparation of tax returns by using RPA technology in the data collection and conversion to tax base processes [22]. Earnst & Young brought together different experts by establishing an artificial intelligence center in India [21]. KPMG has developed the tax analysis software "K-analyzer", which can analyze thousands of transactions. It is observed that these companies increase efficiency and reduce errors by using artificial intelligence [22]. Additionally, in the financial sector, companies such as HSBC Bank and Morgan Stanley use artificial intelligence in areas such as fraud detection and investment consultancy [23]. Artificial intelligence has an important role in these sectors and is used as a powerful tool to optimize business processes and make smarter decisions. These developments show that artificial intelligence technologies in accounting will be more widespread at the upcoming days and will accelerate the transformation in the sector.

1.3 The AI in Auditing

Digitizing the audit process can increase security by digitally recording the access history of each document. In this way, auditors will save time as they will not have to search through filing cabinets to access documents. They will be able to carry out a more effective audit process by taking advantage of digital files. Digital audits can offer the opportunity to audit all financial transactions of a company instead of a sample, while increasing efficiency and accuracy [24]. Kokina and Davenport emphasize that artificial intelligence relies heavily on substantial amounts of information and computational capacity to function effectively. Integration of machine learning into audit processes has streamlined the analysis of vast datasets for auditors, enabling them to gain comprehensive insights into client business operations. [25]. Using artificial intelligence applications in audit processes allows auditors to focus on valuable activities by reducing the time they spend on repetitive tasks. This time saving shows that audit firms will turn to new audit techniques in the near future [26]. Artificial intelligence-supported audit processes can make audit activities more effective with advantages such as digitalization of documents and big data analysis. In this way, auditors can focus on valuable activities and save time by making more in-depth analysis. Additionally, by leveraging artificial intelligence applications, it can be predicted that audit firms will turn to new audit techniques in the near future.

1.4 Pros & Cons of the AI Usage

The Pros of the AI in accounting and auditing are pivotal in reshaping financial processes, offering novel opportunities, and enhancing sector efficiency. These advantages can be summarized as follows [27-32]:

- Artificial intelligence rapidly changes the functioning of financial institutions, providing cost savings and operational efficiency.
- Situations such as increased accuracy and speed, improved reporting, and reduced paper use positively affect the performance of the AI in accounting.
- The AI ensures the reliability of information by reducing human errors and increasing efficiency in accounting and auditing processes.
- Using automated AI analytics in auditing provides more accurate and efficient results.
- Big data analysis and machine learning offer auditors new opportunities.
- Analyzing datasets by the help of natural language processing (NLP) and text classification greatly enhances effectiveness of audit procedures.
- Artificial neural networks improve data quality and reduce costs by storing auditors' experiences.
- Traditional neural networks are used in areas such as evaluating management fraud.
- Artificial intelligence applications enhance audit efficiency and effectiveness by integrating new forms of evidence.
- In a country where tax collection is intense, artificial intelligence can contribute to generating more tax revenue.
- It may be possible to revise the curricula of institutions providing accounting education in accordance with current conditions.

Benefiting the AI in area of accounting offers significant advantages in pecuniary processes, increasing efficiency in the sector and creating new opportunities. Effective and strategic use of these technologies increases efficiency in accounting and auditing processes and ensures the reliability of information. Besides the advantages brought by artificial intelligence, some disadvantages should not be ignored. The drawbacks of benefiting artificial intelligence in accounting and auditing areas can be listed as follows [33,34].

- Risk that biased algorithms may cause financial and reputational damages.
- Risk of decreasing labor need or causing income inequality.
- The trend of technology and accounting firms to hire artificial intelligence experts.
- Risk of being used against the auditor in court.
- It should be detailed enough that "machine learning", artificial intelligence training and software updates are not needed due to often changing accounting regulations. For ex: in Turkey, areas such as tax legislation, Social Security Institution legislation, business and trade legislation undergo frequent changes.
- Lack of regularity of documents, size of the informal economy.

These disadvantages summarize the potential risks brought by the utilization of AI at the realms of accounting and auditing.

2 The ML techniques for Accounting and Auditing

Machine learning, a sub-field within artificial intelligence, involves creating algorithms and mathematical models which let machines to extract meaning from datasets and make presage or decisions with no any direct coding [35]. Machine learning methods have become increasingly popular in accounting and auditing for their capacity to analyze vast datasets, detect patterns, and make predictions. They offer numerous benefits compared to traditional approaches, such as enhanced efficiency, greater accuracy, and power to handle complex and unstructured datasets. In this section, we explore various fundamental machine learning methods employed in area of accounting and auditing.

2.1 Decision Trees (DT)

The DT is a widely used machine learning method employed for tasks including categorization processes together with some regression steps. In the realms of accounting and auditing, decision trees are utilized to categorize financial transactions, detect fraudulent behavior, and forecast financial results. One of their key strengths lies in their interpretability, which is especially valuable in auditing contexts where transparency and explain ability are paramount [36]. The splitting criterion in decision trees is often based on information gain (IG) or Gini impurity (GI). For a node N, the IG is calculated as Equation 1.

$$IG(N) = H(N) - \sum_i \frac{N_i}{N} H(N_i) \qquad (1)$$

where; H(N) is the entropy of node N, N_i is instance number in child node I and N is total instances in node N.

2.2 Support Vector Machines (SVM)

Support Vector Machines (SVM) represent another prevalent machine learning algorithm preferred in accounting and auditing. SVMs excel in classifying intricate and multidimensional data sets. Within auditing, SVMs prove invaluable in detecting anomalies within financial data, pinpointing irregular patterns, and categorizing transactions according to their associated risk levels [37]. In SVM, the objective is to determine the hyper-plane with the greatest edge space which divides data locations of distinct classes. Equation 2 shows the decision function to estimate the class of a new data point x.

$$f(x) = sign\,(\sum_{i=1}^{n} a_i y_i K(x_i, x) + b) \qquad (2)$$

where; a_i are the learned Lagrange multipliers, y_i are the class labels, K is the kernel function, b is bias term.

2.3 Artificial Neural Networks (ANN)

Neural networks, a category of machine learning methods, take cues from the complex mechanisms of the human brain. Within accounting and auditing, neural networks serve diverse functions, such as fraud detection, risk evaluation, and financial prediction. Their strength lies in their capacity to identify complex structures in data, rendering them highly suitable for analyzing extensive and varied datasets [38]. The ANN models were first developed by simple perceptron as demonstrated in Figure 2. The basic feedforward artificial/software-based NN with one hidden layer can be represented mathematically by Equation 3.

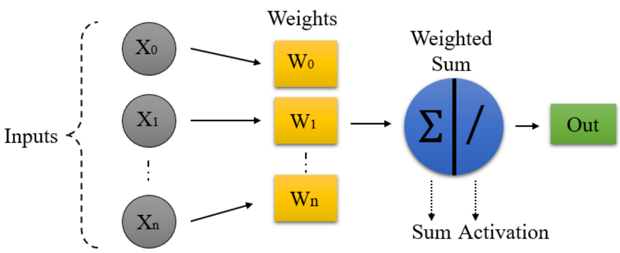

Figure 2. A simple perceptron structure in ANN.

$$y = f_2\,(\sum_{j=1}^{M} w_{2j} . f_1\,(\sum_{i=1}^{D} w_{1i} x_i + b_1) + b_2) \qquad (3)$$

where; x_i are the input features, w_{1i} and w_{2j} are the weights, b_1 and b_2 are the biases, f_1 and f_2 are the activation functions (e.g., sigmoid or ReLU).

2.4 Random Forests (RF)

Random forests represent a technique in ensemble learning that merges numerous decision trees to enhance classification precision. Within accounting and auditing, this technique proves useful for spotting trends within financial data, categorizing transactions, and forecasting future results. Random forests exhibit resilience against overfitting and can manage extensive datasets featuring high dimensionality [39]. In random forests, individual trees are trained on random samples of the dataset. The last estimation is determined via aggregating majority vote from all trees. Forecast for a novel instance x is shown by Equation 4.

$$y = mode\ (f_1(x), f_2(x), \ldots, f_T(x)) \quad (4)$$

where; $f_T(x)$ is the prediction of tree t and T is the total number of trees.

2.5 K-Nearest Neighbors (KNN)

The K-Nearest Neighbors (KNN) algorithm is a straightforward and intuitive piece of software utilized for classification and regression purposes. In the realm of accounting and auditing, the K-Nearest Neighbors (KNN) algorithm finds application in classifying financial transactions, spotting anomalies within data, and recognizing patterns in financial datasets. KNN operates by identifying K closest data points in training set to data location under classification and assigning most prevalent label among those neighbors to the data point [40]. K-Nearest Neighbors (KNN) is a straightforward machine learning algorithm that classifies a new data location based on class most frequently found within its k nearest neighbors. Prediction for a new instance x is formulated by Equation 5.

$$y = mode\ (y_{N1}, y_{N2}, \ldots, y_{NK}) \quad (5)$$

where; y_{NK} is the class label of the Kth nearest neighbor of x. the "mode" function refers to the statistical mode that is the value which appears most frequently in dataset.

2.6 Logistic Regression (LR)

Logistic Regression is a statistical method that is commonly used for two-class classification tasks. In accounting and auditing, logistic regression can be used to predict the likelihood of an event occurring, such as the likelihood of fraud in a financial transaction. Logistic regression tries to learn the probability of the binary outputs using logistic function [41]. The probability of the positive class for an input vector x is given by Equation 6. In the LR, the sigmoid function generally is used to map the output of the linear regression model (which can be numeric) to a value between 0 and 1, representing the probability of a given input that falls into a specific class (usually 0 or 1).

$$P(Y = 1 \mid X = x) = \frac{1}{1 + e^{-(w_0 + w_1 x_1 + \cdots + w_n x_n)}} \quad (6)$$

where; w_0, w_1, w_2, …, w_n are the model parameters (coefficients), x_1, x_2, …, x_n are the input features, e is base for logarithm.

2.7 Naïve Bayes (NB)

Naive Bayes is a straightforward probabilistic classifier that relies on applying Bayes' theorem while assuming significant autonomy between the features. In accounting and auditing, Naive Bayes can be used for classification tasks, such as classifying financial transactions or documents based on their content. Naive Bayes is computationally efficient and works well with large datasets [42]. This algorithm relies on Bayes' theorem and operates under assumption that features are unrelated to each other, provided class label. Probability of class label y given an input vector x is given by Equation 7.

$$P(Y = y \mid X = x) = \frac{P(X = x \mid Y = y)P(Y = y)}{P(X = x)} \quad (7)$$

where; P (X=x | Y=y) is the likelihood of investing x given that class is y, P (Y=y) is prior probability of class y, P (X=x) is marginal probability of investing x.

2.8 Clustering Algorithms

Clustering algorithms are utilized to categorize comparable data points together according to their attributes. In the realms of accounting and auditing, these algorithms find application in segmenting customers based on their behavior, group financial transactions based on their attributes, and identify patterns in financial data. Clustering algorithms can help auditors gain insights into the underlying structure of financial data and detect anomalies or outliers [43]. A frequently employed algorithm is K-means, which minimizes the total squared distances from every point to its respective cluster center. This algorithm is widely used in unsupervised machine learning to partition a dataset into K clusters, assigning each data point to the cluster with the closest mean. K-means can be used to assess risk by clustering financial data based on risk factors. This can help identify high-risk groups or transactions that require further investigation.

2.9 Natural Language Processing (NLP)

Natural Language Processing (NLP) is a sector of artificial intelligence (AI) that concentrates on the communication between machine and people using natural language. It tries to understand the purpose of human speech with different artificial intelligence approaches. Traditional and related approaches are used. Flexible algorithms are developed with machine learning. It aims to automatically discover connection in the computer's input and output. It is used in online environments in the area of research and technological developments by analyzing and coding the language [13]. In accounting and auditing, NLP can be applied to process text data from financial reports, regulatory filings, and other documents. NLP techniques can help auditors extract relevant information, identify key trends, and detect anomalies in textual data [44]. One common task of NLP algorithm is sentiment analysis, where the goal is to determine the sentiment (positive, negative, neutral) of a piece of text. Various techniques, such as bag-of-words or word embeddings, can be used for the NLP.

3 Suspicious Company Detection System

The fraudulent firm classification proposed in this work was done using three main classification algorithms which are Decision Tree, Bagged Tree and Linear SVM. A decision tree is a hierarchical structure resembling a tree, where each internal node signifies a "test" conducted on an attribute. Each branch indicates the test outcome, while each leaf node signifies a class label or continuous value. Decision trees are constructed by iteratively dividing the dataset according to attribute values to create a tree that can be used for classification or regression.

To mathematically explain bagged tree: Let T be the number of trees in the ensemble, N be the number of data points in the training set, and X be the input feature matrix of size $N \times D$, where D is the number of features. The algorithm first samples a bootstrap sample of the training set (with replacement) to create a new training set X_t of size N from X. Then it trains a decision tree f_t using the new training set X_t. Then it makes a prediction for a new input x by aggregating the predictions of all trees as shown in Equation 8.

$$y = \frac{1}{T}\sum_{t=1}^{T} f_t(x) \quad (8)$$

Bagged trees, also known as bootstrap aggregating or bagging, involve training multiple decision trees on bootstrapped samples of the dataset (sampling with replacement). Each tree is trained independently, and during prediction, the output of the bagged model is the average (for regression) or majority vote (for classification) of the predictions of individual trees [45]. Linear Support Vector Machine (Linear SVM) is a variation of the SVMs that uses a linear kernel function to identify the best hyperplane that distinguishes data points of distinct classes within a multi-dimensional space. Linear SVM is particularly useful when the data is linearly separable, meaning a single hyperplane can perfectly separate the data points of different classes. It is commonly employed in binary classification tasks and can be expanded to multiclass classification through methods like one-vs-rest or one-vs-one. SVM can use different kernel functions, such as linear, polynomial, radial basis function (RBF), and sigmoid, to map the input features into a higher-dimensional space where the classes are separable by a hyperplane. Linear SVM specifically uses a linear kernel function, which means it looks for a linear decision boundary in the original feature space without explicitly transforming the data into a space with more dimensions. Linear SVM generally differentiates from SVM in terms of kernel function used, the complexity of the decision boundary they can model, computational complexity, and interpretability. General structure of suspicious company detection system was given in Figure 3. Input data was first analyzed by using Exploratory Data Analysis (EDA). Then unnecessary and meaningless data was eliminated in Data Pre-Processing step. Classification phase was applied over that clean data to robustly implement the machine learning algorithms.

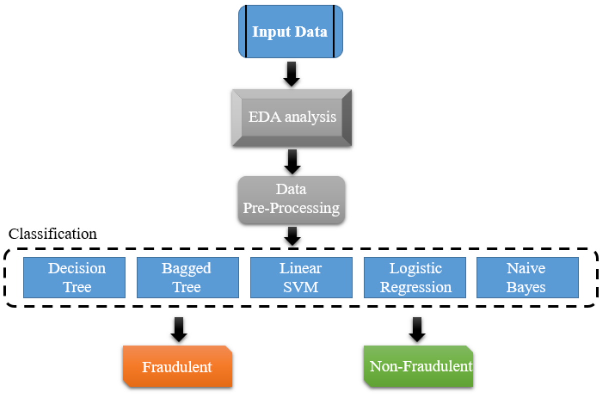

Figure 3. Pipeline of proposed system.

Dataset used in this work was collected from "University of California, Irvine (UCI) - Machine Learning Repository" [46]. The dataset aims to assist auditors in identifying fraudulent firms by creating a classification model based on current and previous risk factors. It contains comprehensive non-confidential data from 2015 to 2016, collected from the Auditor Office of India, to develop a predictor for classifying suspicious firms. The dataset was supplied by an external government audit firm, which also acts as the external auditor for government entities in India. In the audit planning phase, auditors evaluate the business activities of different government offices to identify those most likely to contain significant misstatements. This assessment is based on the risk associated with achieving the financial reporting objectives in [47]. There are 777 observations (records) from 14 different sectors in dataset as shown by Table 1.

Table 1. The sector and the number of observations.

Sector	# of Records
Agricultural Watering	114
Community Health	77
Infrastructure and Highways	82
Woods	70
Business	47
Livestock Farming	95
Intercommunication	1
Electrical	4
Terrain	5
Knowledge and Engineering	3
Travel	1
Aquaculture	41
Manufacturing	37
Farming	200

After obtaining the dataset, some settings were applied to tidy up the columns and rows. This process can be seen as data pre-processing. Dataset was passed over using Exploratory Data Analysis (EDA). Exploratory Data Analysis (EDA) is a methodology for examining datasets to outline their principal attributes, frequently employing visual techniques. The primary goal of EDA is to understand the data and its underlying structure, patterns, and relationships between variables. Typical EDA involves some set of operations: *Data Cleaning*: it involves recognizing and rectifying errors or disparities in the data, like absent values or anomalies. *Summary Statistics*: it entails computing descriptive metrics like average, median, standard deviation, and range to encapsulate the data. *Data Visualization*: it encompasses generating visual depictions of the data, such as histograms, box plots, scatter plots, and heat maps, to investigate patterns and correlations. *Identifying Patterns*: Using statistical methods to identify trends, patterns, or anomalies in the data. *Formulating Hypotheses*: Based on the findings from EDA, formulating hypotheses about the data that can be further tested using statistical methods. Null values were filled by average(mean) of the related columns in terms of EDA for the mentioned dataset. Also, One-Hot encoding was applied for the non-numeric columns. The dataset used is imbalanced including approximately 35% of observations belonging to class 1, i.e. fraudulent companies, others belong to class 0.

4 Results on Bibliometric Analysis and Classification

In this section, we provide results for suspicious company detection system based on machine learning. The main focus of this study is to examine usage of artificial intelligence on accounting area in order to help auditors in before on-site auditing processes. After a deep literature search, we want to point out some studies contributing highly to the researchers working on this area. Systematic literature review was preferred in the methodological approach of the Bibliometric Analysis. The systematic screening method has been accepted as a logical method for determining known boundaries and clarifying unknown boundaries [48]. Within the scope of the study, a detailed search was first made in the literature in Web of Science, Scopus, ScienceDirect and Elsevier databases for the dates between 2018-2024. Because of the study's restricted scope, academic studies registered in national (for Turkey) databases such as “Yök Tez”, “Dergi Park” and international databases such as Google Academy were also examined as databases in the literature review. The keywords of “Accounting”, “Audit” and “Artificial Intelligence” were used in the research. However, the topics of “Artificial intelligence in accounting”, “Artificial intelligence in auditing”, “Pros and cons of artificial intelligence in accounting and auditing” were also researched in Turkish and English. Research data was obtained from 2018 to 2024 (including the first quarter).

At this point, research and articles discuss the effects of artificial intelligence on the accounting and auditing professions. As seen in Table 3, researchers from different geographies, from Turkey to Indonesia, from India to Vietnam, are examining the role of artificial intelligence in the accounting profession and evaluating the effects of this technology on professional practice. The digital transformation of the researches and the role of artificial intelligence in the accounting are discussed by studies in Table 3 which examines the studies in the field of artificial intelligence and accounting. Additionally, it focused on the potential of artificial intelligence to prevent financial fraud in the field of accounting and stated that artificial intelligence could transform the accounting industry. These studies reveal the need for professional accountants to adapt to artificial intelligence technologies and adapt to developing technology.

It is also important to show the results of the fraudulent company classification system developed by machine learning algorithms. Train and test accuracy of Decision Tree, Bagged Tree, Linear SVM and other algorithms were provided in Table 2. 5k-fold cross validation was used during the classification phase. Cross-validation in classification is a technique used to evaluate the performance of a machine learning (ML) model by splitting the dataset into dividing the data into subsets, training the model on one subset, and testing it on the other subsets. This aids in evaluating the model's ability to generalize to new data and in estimating its performance on a dataset that is separate from the one used for training. Also, dataset was splitted into train and test sets with a rate of 30%.

Table 2. Performance of machine learning models (in %).

Algorithm	Train Accuracy	Test Accuracy
Decision Tree	100	100
Bagged Tree	100	100
Linear SVM	99.45	99.57
Logistic Reg.	98.90	99.57
Naïve Bayes	96.13	98.71
Subspace KNN	98.90	99.14

Among all machine learning algorithms, Decision Tree and Bagged Tree have reached up to 100% classification accuracy for both train and test phases. This proved that the artificial intelligence can efficiently be used in detection of suspicious firms before on-site auditing using companies’ risk information.

Confusion matrices of best six classification methods with 232 test observations in test phase were shown in Figure 4. A confusion matrix is a table frequently employed to illustrate how well a classification model performs on a dataset where the true values are known. It provides a visual representation of the algorithm's performance by contrasting predicted outcomes with actual outcomes.

It can easily be derived from confusion matrices that ML algorithms of Decision Tree and Bagged Tree correctly predicted the true class of all 232 test observations. Naïve Bayes misclassified 3 of test observations and Subspace KNN missed two of them while Linear SVM and Logistic Regression classified incorrectly only one of test observations.

The Receiver Operating Characteristic (ROC) curve is a visual indication that shows how well a classifier can distinguish between classes. It plots the true positive rate (TPR), also known as sensitivity, against the false positive rate (FPR), which is 1 minus the specificity. The curve is created by varying the threshold for classifying observations as positive or negative and calculating the TPR and FPR at each threshold. The Model Operating Point is a specific point on the ROC curve that corresponds to the threshold used by the classifier to make classifications. It shows the FPR and TPR at that threshold, indicating how well the classifier is performing at that particular threshold. For instance, a false positive rate of 0.3 implies that the classifier erroneously identifies 30% of the negative class observations as positive. Likewise, a true positive rate of 0.7 signifies that the classifier accurately identifies 70% of the positive instances like positives.

Table 3. Research on artificial intelligence in accounting and auditing.

The Authors	Article
Özçetin, N. [49]	Artificial Intelligence in Accounting Auditing
Lai, C.S. [50]	Artificial Intelligence in Accounting and Auditing: The Way Forward
Göl, M. [51]	Transition to Artificial Intelligence Technology in Accounting Audit
Yardımcıoğlu, M. [52]	Reflections of Artificial Intelligence Technology on the Accounting Field: Literature Review
Chouhan,V. [53]	Measuring Accounting Professionals Perception on use of AI Based Accounting Practices in India
Haridasan, V. [54]	Evaluating Artificial Intelligence's Effect On Accounting Information Systems For Small And Medium-Sized Enterprises.
Kuncoro, E. A. [55]	Artificial Intelligence and the Role of External Auditor in Indonesia.
Anh, N. T. M. [56]	The Effect of Technology Readiness on Adopting Artificial Intelligence in Accounting and Auditing in Vietnam
Avcı, V.E. [57]	Detecting Financial Statement Fraud: An Analysis of Machine Learning Techniques
Kuo, Y.F. [58]	Fraud Detection in Forensic Accounting: A Neural Network Approach
Karahan, A. [59]	Detection and Prevention of Fraud: An Internal Audit Perspective

The studies listed in Table 3 generally examine the role of artificial intelligence in accounting and auditing from conceptual, technological, or perception-based perspectives. These works explore topics such as the digital transformation of audit practices, the effects of AI on accounting information systems, professional attitudes toward AI adoption, and the potential of AI-supported methods for fraud detection. Collectively, the literature highlights the increasing relevance of AI technologies in audit processes, yet most studies provide descriptive evaluations rather than integrated analytical frameworks. Unlike the existing literature, which predominantly focuses on conceptual discussions or isolated algorithmic fraud-detection models, this study combines a bibliometric analysis with a comprehensive machine-learning-based suspicious-company detection framework. By integrating a systematic mapping of AI-related audit research with an empirical evaluation of multiple classification algorithms, the present study provides both a broader intellectual overview and a practical predictive model. This dual methodological approach distinguishes the study from prior research in the accounting and auditing domain.

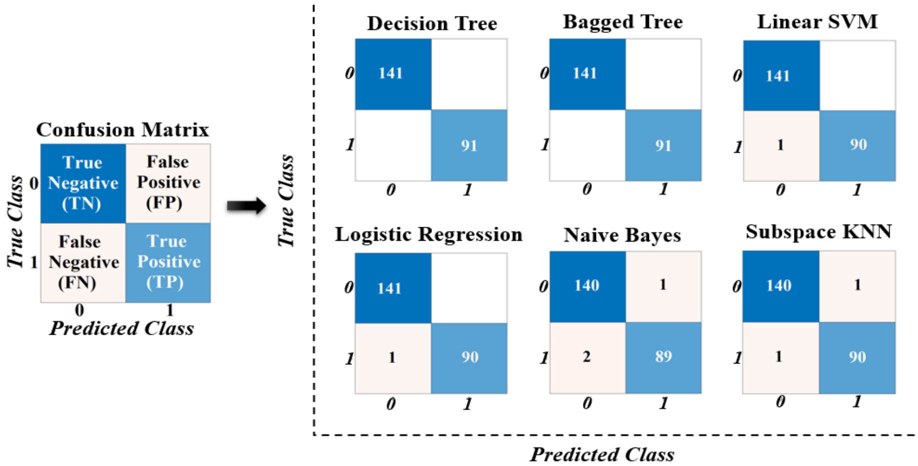

Figure 4. Confusion matrix for the best six classification methods.

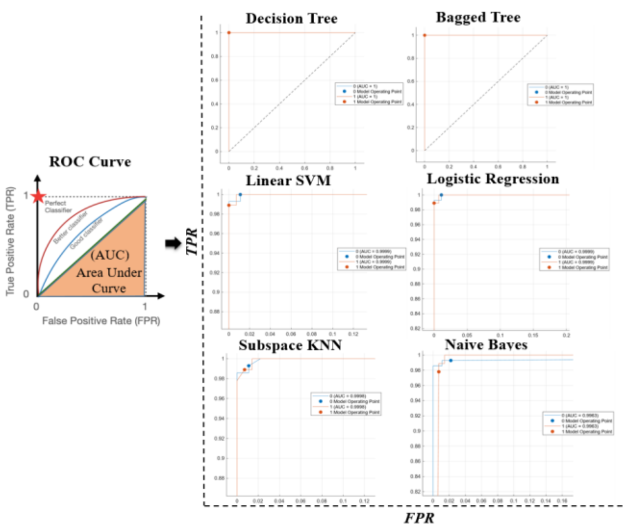

Figure 5. ROC (Receiver Operating Characteristic) curve for the best six classification methods.

The ROC curve for the best classification methods were shown by Figure 5. In ROC curve, the best classifiers are expected to have area under curve (AUC) value of 100% (or 1). As Figure 5 implies, Decision Tree and Bagged Tree has AUC value of 1 which means that they are the perfect classifiers. The AUC values of 0.9999, 0.9999, 0.9998 and 0.9963 were obtained for Linear SVM, Logistic Regression, Subspace KNN and Naïve Bayes, respectively. Also, feature importance scores were extracted to see which column (feature) is most critical for the performed classification. ANOVA, Chi2, Kruskal Wallis and MRMR algorithms were used for this purpose. Chi2, Kruskal Wallis and MRMR pointed that the feature of "Audit Risk" is the most valuable feature while ANOVA found that "Score" column is most critical feature for classification.

5 Discussion

In this work, the use of the Artificial Intelligence (AI) in Accounting and Auditing was examined in terms of a comprehensive bibliometric analysis. Moreover, a suspicious company detection system was developed based on Machine Learning (ML) algorithms which are Decision Tree, Bagged Tree, Linear Support Vector Machine (SVM), Logistic Regression, Subspace K-Nearest Neighbors (KNN) and Naïve Bayes. The results of classification system proved that an auditor can use such classification system with 100% accuracy in order to have an idea about fraudulent companies before on-site auditing. With the increasing use of artificial intelligence in accounting and auditing, the effects of this technology on accounting and auditing processes are gaining importance. Our findings include that artificial intelligence can provide significant benefits in areas such as financial fraud prevention, automated audit processes, improved decision-making capabilities and time savings. However, the risks brought by the use of artificial intelligence should not be ignored. Issues such as data privacy, algorithmic justice, ethical issues and ignoring the human factor are points that need to be considered regarding the use of artificial intelligence in accounting and auditing processes. From the perspective of the accounting and auditing profession, these risks have direct implications for audit quality and professional responsibility. Algorithmic bias may lead to inappropriate risk assessments or misclassification of companies, potentially affecting audit conclusions. Excessive automation can also weaken professional judgment, which is a core requirement of auditing standards. Additionally, the dynamic nature of financial regulations raises concerns about whether AI models can remain up-to-date and compliant without continuous expert supervision. Confusion matrices and ROC curves together with Table 2 were provided to highlight the success of ML-based classification algorithms. There are also more types of evaluation metrics in such classification cases. The most commonly used evaluation metrics are calculated by the following formulas in Table 4. TP, TN, FP and FN can be fetched from confusion matrices.

Table 4. Common evaluation metrics in classification problems.

Evaluation Metric	Formula
Accuracy	$\frac{TP+TN}{TP+TN+FP+FN}$
Precision	$\frac{TP}{TP+FP}$
Recall/Sensitivity/True Pos.Rate (TPR)	$\frac{TP}{TP+FN}$
Specificity	$\frac{TN}{TN+FP}$
F1 Score	$\frac{2.Precision.Recall}{Precision+Recall}$
False Pos.Rate (FPR)/False Alarm Rate (FAR)/1-Specificity	$\frac{FP}{FP+TN}$
False Neg.Rate (FNR)/Miss Rate	$\frac{FN}{FN+TP}$

The formulations of the evaluation metrics presented in Table 4 are based on standard definitions commonly used in machine learning and data mining literature [60]. These metrics provide essential tools for assessing the performance of classification models in various applications. They are widely used to evaluate the effectiveness of classifiers, including those used in fraudulent company detection.

6 Conclusion

This paper demonstrates that machine learning algorithms, particularly Decision Tree and Bagged Tree, can effectively identify suspicious companies with high accuracy, providing auditors with a reliable preliminary assessment before on-site auditing. For the business world and auditing practitioners, such AI-based systems can be integrated into regular audit workflows to improve fraud detection efficiency, reduce manual workload, and support data-driven decision-making. Professional chambers and regulatory authorities can use these findings to guide the development of AI adoption frameworks, auditing standards, and training programs that ensure ethical and effective use of AI in financial auditing. For researchers and academics, the study highlights several avenues for future work. These include exploring hybrid models that combine multiple machine learning algorithms, investigating real-time AI monitoring systems for fraud detection, and examining the ethical, legal, and data privacy implications of AI in auditing. Additionally, bibliometric insights from this study suggest opportunities to expand AI adoption research across different industries and geographic regions by providing a richer understanding of its impact on accounting practices. To sum up, the integration of AI in auditing not only enhances fraud detection but also opens new directions for research, regulation, and professional practice in the accounting field.

Declaration

The authors declare that the ethics committee approval is not required for this study.

References

[1] Kaplan, J. (2016). Artificial intelligence: What everyone needs to knowR. *Oxford University Press*.
[2] Pavaloiu, A. (2016). The Impact of Artificial Intelligence on Global Trends. *Journal of Multidisciplinary Developments*, 1(1), 21-37.
[3] McCarthy, J., Minsky, M. L., Rochester, N., & Shannon, C. E. (2006). A proposal for the dartmouth summer research project on artificial intelligence, august 31, 1955. *AI magazine*, 27(4), 12-12.
[4] Singh, G., Mishra, A., & Sagar, D. (2013). An overview of artificial intelligence. *SBIT journal of sciences and technology*, 2(1), 1-4.
[5] Hutter, M. (2005). Universal Artificial Intelligence: Sequential Decisions based on Algorithmic Probability. *Springer*.
[6] LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, 521(7553), 436-444.
[7] Russell, S. J., & Norvig, P. (2016). Artificial Intelligence: A Modern Approach. *Pearson*.
[8] Domingos, P. (2012). A few useful things to know about machine learning. *Communications of the ACM*, 55(10), 78-87.
[9] Schmidhuber, J. (2015). Deep learning in neural networks: An overview. *Neural Networks*, 61, 85-117.
[10] Manning, C. D., Raghavan, P., & Schütze, H. (2008). Introduction to Information Retrieval. *Cambridge University Press*.
[11] Nadas, E.(2021). *Artificial Intelligence Applications in Accounting and Auditing*, (Master's dissertation Istanbul Bilgi University).
[12] Zhang, Y., Xiong, F., Xie, Y., Fan, X., & Gu, H. (2020). The impact of artificial intelligence and blockchain on the accounting profession. *IEEE Access*, *8*, 110461-110477.
[13] Albayrak, A. (2020). Preparation of interdisciplinary graduate course content using natural language processing techniques. *Bilişim Teknolojileri Dergisi*, *13*(4), 373-383.
[14] Gonzalez, R. C., & Woods, R. E. (2017). Digital Image Processing (4th ed.). *Pearson*.
[15] Çivak, H. (2022). *Robotic Process Automation: An Application Example,* (Master's dissertation Karabük University).
[16] Bengio, Y. (2009). Learning deep architectures for AI. *Foundations and trends® in Machine Learning*, 2(1), 1-127.
[17] Struthers-Kennedy, A., & Nesgood, K. (2020). Artificial Intelligence and Internal Audit: A Pragmatic Perspective, Protivity, *Knowledge Leader*.
[18] Susskind, R., & Susskind, D. (2015). The Future of the Professions: How Technology Will Transform the Work of Human Experts. *Oxford University Press*.
[19] Greenman, C. (2017). Exploring the impact of artificial intelligence on the accounting profession. *Journal of Research in Business, Economics and Management*, 8(3), 1451.
[20] Handoko, B. L., Mulyawan, A. N., Samuel, J., Rianty, K. K., & Gunawa, S. (2019). Facing industry revolution 4.0 for millennial accountants. *International Journal of Innovative Technology and Exploring Engineering* (IJITEE), 9(1), 1037-1041.
[21] Yaninen, D. (2017). Artificial intelligence and the accounting profession in 2030. *J. Account. Finance*, 3-29.
[22] Davenport, T. H., & Ronanki, R. (2018). Artificial intelligence for the real world. *Harvard Business Review*, 96(1), 108-116.
[23] Wilson, H. J., & Daugherty, P. R. (2018). İşbirliğine dayalı zeka: İnsanlar ve yapay zeka güçlerini birleştiriyor. *Harvard Business Review Türkiye*.
[24] Krahel, J. P., & Titera, W. R. (2017). The audit of information systems: A review of the past 25 years. *Journal of Information Systems*, 31(1), 47-73..
[25] Kokina, J., & Davenport, T. H. (2017). The emergence of artificial intelligence: How automation is changing auditing. *Journal of emerging technologies in accounting*, *14*(1), 115-122.
[26] Luo, J. X., Meng, Q. J. ve Cai, Y. (2018). Analysis of the impact of artificial intelligence on the development of accounting industry. *Open Journal of Business and Management*, 6, 850-856.
[27] Odoh, L. C., Echefu, S. C., Ugwuanyi, U. B., & Chukwuani, N. V. (2018). Effect of artificial intelligence on the performance of accounting operations among accounting firms in South East Nigeria. *Asian Journal of Economics, Business and Accounting*, *7*(2), 1-11.
[28] Zemankova, A. (2019, December). Artificial intelligence in audit and accounting: Development, current trends, opportunities and threats-literature review. In *2019 International Conference on Control, Artificial Intelligence, Robotics & Optimization (ICCAIRO)* (pp. 148-154). IEEE.
[29] Chen, C. X., Hsu, A. W., & Yang, Y. (2020). Artificial intelligence in accounting and auditing: An overview of technology, applications, and implications. *Journal of Emerging Technologies in Accounting*, 17(2), 1-12..
[30] Brandas, C., Muntean, M., & Didraga, O. (2018). Intelligent decision support in auditing: Big Data and machine learning approach. In *17th International conference on ınformatics in economy (IE 2018) education, research & business technologies. The Bucharest University of Economic Studies, Bucharest, Romania*.
[31] Mirzaey, M., Jamshidi, M. B., & Hojatpour, Y. (2017). Applications of artificial neural networks in information system of management accounting. *International Journal of Mechatronics, Electrical and Computer Technology*, *7*(25), 3523-3530.
[32] Glover, S. M., Prawitt, D. F., & Drake, A. A. (2020). The future of audit: A research synthesis. Auditing: *A Journal of Practice & Theory*, 39(1), 163-191.
[33] Alles, M. G., Brennan, G., Kogan, A., & Vasarhelyi, M. A. (2019). Artificial intelligence and the future of the accounting profession. *Journal of Emerging Technologies in Accounting*, 16(2), 1-14..
[34] Omoteso, K. (2012). The application of artificial intelligence in auditing: Looking back to the future. *Expert Systems with Applications*, *39*(9), 8490-8495.
[35] Alpaydin, E. (2010). Introduction to machine learning. *MIT press*.

[36] Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J. (1984). Classification and regression trees. *CRC press*.

[37] Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine learning*, 20(3), 273-297.

[38] Bishop, C. M. (1995). Neural networks for pattern recognition. *Oxford university press*.

[39] Breiman, L. (2001). Random forests. *Machine learning*, 45(1), 5-32.

[40] Cover, T., & Hart, P. (1967). Nearest neighbor pattern classification. *IEEE transactions on information theory*, 13(1), 21-27.

[41] Cox, D. R. (1958). The regression analysis of binary sequences. *Journal of the Royal Statistical Society: Series B (Methodological)*, 20(2), 215-242.

[42] Domingos, P., & Pazzani, M. (1997). On the optimality of the simple Bayesian classifier under zero-one loss. *Machine learning*, 29(2-3), 103-130.

[43] Jain, A. K. (2010). Data clustering: 50 years beyond K-means. *Pattern recognition letters*, 31(8), 651-666.

[44] Manning, C. D., Raghavan, P., & Schütze, H. (2008). Introduction to information retrieval. *Cambridge University Press*.

[45] Breiman, L. (1996). Bagging predictors. *Machine learning*, 24(2), 123-140.

[46] Hooda, N. (2018). Audit Data. *UCI Machine Learning Repository*. https://doi.org/10.24432/C5930Q

[47] Houston, Richard W. and Peters, Michael F. and Pratt, James H. (1999). *The Audit Risk Model, Business Risk, and Audit Planning Decisions.* Retrieved 20 June, 2025, from *https://ssrn.com/abstract=163219*

[48] Pickering, C., & Byrne, J. (2014). The benefits of publishing systematic quantitative literature reviews for PhD candidates and other early-career researchers. *Higher Education Research & Development*, 33(3), 534-548

[49] Özçetin, N. (2022). Artificial Intelligence in Accounting Auditing. *Uşak Üniversitesi Uygulamalı Bilimler Fakültesi Dergisi*, 2(1), 29-41.

[50] Lai, C. S., & Cheng, H. (2021). Artificial Intelligence in Accounting and Auditing: The Way Forward. *International Journal of Business and Society*, 22(2), 541-554.

[51] Göl, M. (2023). Transition to Artificial Intelligence Technology in Accounting Auditing. *Accounting and Auditing on the Axis of Current Developments* (pp. 203-220). Özgür Yayın Dağıtım Ltd. Şti.

[52] Yardımcıoğlu, M., & Şıtak, B. (2020). Reflections of artificial intelligence technology on accounting: literature review. *Bilecik Şeyh Edebali Üniversitesi Sosyal Bilimler Dergisi*, 5(2), 342-353.

[53] Chouhan, V., Shakdwipee, P., Vasita, M. L., & Chand, P. (2020). Measuring Accounting Professionals Perception on use of AI Based Accounting Practices in India. *International Journal of Engineering and Advanced Technology (IJEAT)*, 9(3).

[54] Haridasan, V., Muthukumaran, K., Usha, K., Vasu, S. B., & Jhansi, V. (2023). Evaluating Artificial Intelligence's Effect On Accounting Information Systems For Small And Medium-Sized Enterprises. *Migration Letters*, 20(S13), 680-693.

[55] Kuncoro, E. A., Lindrianasari, L., & Fatmasari, A. (2023). Artificial Intelligence and the Role of External Auditor in Indonesia. *In E3S Web of Conferences* (Vol. 426, p. 02122). EDP Sciences.

[56] Anh, N. T. M., Hoa, L. T. K., Thao, L. P., Nhi, D. A., Long, N. T., Truc, N. T., & Ngoc Xuan, V. (2024). The Effect of Technology Readiness on Adopting Artificial Intelligence in Accounting and Auditing in Vietnam. *Journal of Risk and Financial Management*, *17*(1), 27.

[57] Avcı, V. E., & Gökgöz, F. (2021). Detecting Financial Statement Fraud: An Analysis of Machine Learning Techniques. *Journal of Accounting and Finance*, 21(1), 122-140.

[58] Kuo, Y. F., & Wu, C. M. (2018). Fraud Detection in Forensic Accounting: A Neural Network Approach. *Journal of Forensic Accounting Research*, 3(2), 80-94.

[59] Karahan, A. (2022). Detection and Prevention of Fraud: An Internal Audit Perspective. Journal of History School, 2022-LVIII, pp.1554-1580.

[60] Han, J., Kamber, M., & Pei, J. (2012). Data mining: Concepts and techniques (3rd ed.). *Morgan Kaufmann*.