2nd International Conference on Information Science and Technology Innovation (ICoSTEC)
February, 25 2023. Yogyakarta, Indonesia

# Data Mining in Auditing: Challenges and Opportunities

Aditya Arisudhana, Khaula Lutfiati Rohmah

*Universitas Respati Yogyakarta*

[1]adityaarisudhana@respati.ac.id
[2]khaulalutfiati@respati.ac.id

***Abstract* — Audit is a process of evaluation of a subject matter with a view to express an opinion on whether the subject matter is fairly presented carried out by an independent party. Assessment of the fairness of financial statements is very important because of the conflict of interest between the preparers of financial reports and stakeholders. The audit usually uses sampling to obtain data related to the fairness assessment. The auditor uses sampling because of time and cost constraints. Sampling also has a risk that can lead to errors in assessing the fairness of financial statements. Therefore we need a method that can improve accuracy in data collection and processing, for example data mining. Data mining is a method that can be used to collect and process data more quickly and accurately. The use of data mining techniques that may develop in the future can have an impact on the audit process. Data mining in audits will be both an opportunity and a challenge for auditors in the future**

***Keywords* — Include at least 5 keywords or phrases**

## I. INTRODUCTION

Financial report audit is one of the most important activities in today's business activities. Audit of financial statements provides an assessment of the fairness of a company's financial statements. Audits are usually carried out by independent parties so that the level of objectivity is higher. Today, audit reports are the basis for the belief that the published financial reports are appropriate and appropriate in actual circumstances. The audit is usually carried out through special procedures that have been made by the Association of Auditors from the beginning to the end of the work. Auditors can also adjust implementation techniques because the conditions and situations in each company are different. Audit activities will usually compare the conformity between evidence of activities, reports prepared by management and certain standards or criteria. The end result of these activities is an audit report.

The implementation of audit activities is also inseparable from risks. The risk that may occur is that the auditor fails to make an objective audit report according to actual conditions. One of the causes is the collection of data using sampling techniques. Auditors will find it difficult to retrieve all of the company's financial data which is very large in number. The auditor uses a sampling technique because of time and cost constraints. One possible method to use is data mining related to the company's financial statements. Data mining will be able to improve the quality of audit work starting from the data collection process to drawing conclusions

The use of data mining techniques will also have its own impact on auditors in the future. The use of data mining will bring changes in the audit process and the work of the auditor itself. On the one hand, Data Mining will simplify the work of the auditor and increase the accuracy of the audit work. On the other hand, data mining is also a challenge for auditors. Data mining will be a challenge starting from the adoption process. When it is successful, data mining will also be able to shift some of today's auditor jobs.

## II. LITERATURE REVIEW

### 1. Auditing

Auditing is a process of evaluation of a subject matter with a view to express an opinion on whether the subject matter is fairly presented. Audits can also be divided into audit of financial statement, compliance audits and audit of internal control over financial reporting. Audit of financial statements is a process for evaluating conformity between evidence of transactions, management financial reports and applicable financial standards. Compliance audit is a process to evaluate the company's compliance with certain regulations. Operational audit is a procedure for evaluating the efficiency and effectiveness of standard, operational activities of a company (Mgregson, 2017). Audit sampling is the selection of a portion of audit evidence from the total required audit evidence. Inappropriate sample selection will have an impact on audit activities. Auditors can make erroneous conclusions or audit work becomes inefficient (IAASB, 2021).

(Singh et al., 2022) ‘audit sampling’ as utilization of audit systems to under 100 percent ofthings inside a populace of audit significance to such an extent that all sampling units get an opportunity of determination to give the auditor a reasonable basis on which to reach inferences about the wholepopulace. The objective of the auditor when using audit sampling is to provide a reasonable basis for theauditor to draw conclusions about the population from which the sample is selected. Audit sampling can be applied using either nonstatistical or statistical sampling approaches. The factors that should beconsidered for

deciding upon the extent of checking on sampling plan are size of the organization underaudit, state of the internal control, adequacy and reliability of books and records, tolerable error rangeand degree of the desired confidence. The goal of the auditor while utilizing review sampling is to give asensible premise to the auditor to make inferences about the populace from which the sample is chosen (IAASB, 2021) Sample size can be determined by applying statistical formulas or professional judgment. When conditions are similar the it will make similar effects on sample size regardless of the choice of statistical or non-statistical methods. With statistical sampling sample items are selected in such a way that each sample unit has a probability of being selected. Judgments are used to select sample items for non-statistical sampling. Since the purpose of sampling is to provide auditors with a reasonable basis on which to draw conclusions about the population from which they will select samples it is wise for auditors to select representative samples to avoid bias by selecting samples with the following characteristics: important.

Characteristics of the project sample population. **Definition of Data Mining:**

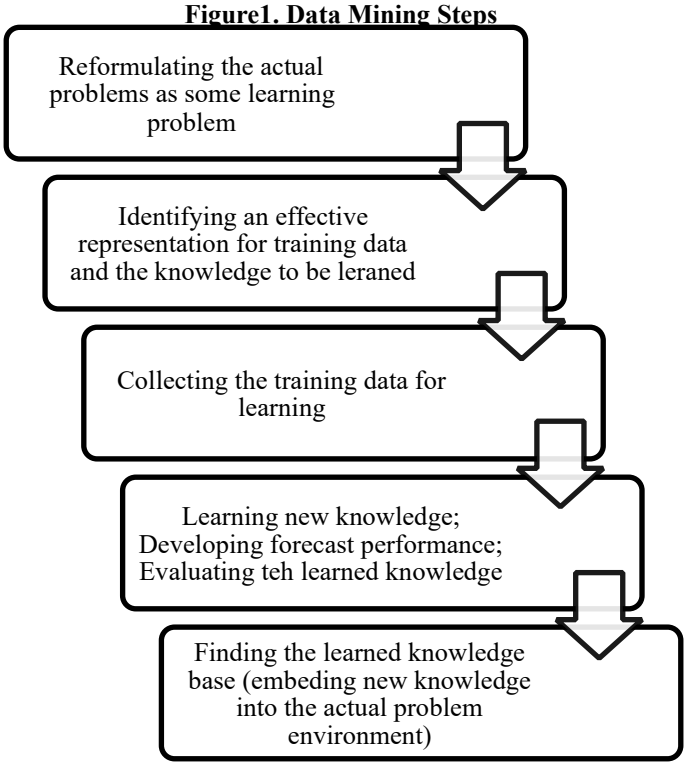
Data mining emerged in late 80's by using concepts and methods from the fields of artificial intelligence, Pattern Recognition, Database System and Statistics (Kirkos & Manolopoulos, 2004). Data mining, sometimes referred to as data or knowledge discovery, derives its name from searching for valuable information in a large database, data warehouse or data mart, that can be used to establish relationships between variables, and then to validate the findings by applying the detected patterns to new subjects of data (Rostami et al., 2011). Even more, Rostami explained, data mining is a way to gain market intelligence from a huge amount of data. Data mining is used to search for valuable information from the mounds of data collected over time, which could be used in decisionmaking. Data mining is analysis process of tremendous volume of data for finding useful, obvious and important information, patterns by using automatic and semiautomatic developing tools.

Data mining techniques are now applied to discover hidden trends and patterns in financial data base, e.g., in stock market data for market prediction (Kovalerchuk & Vityaev, 2000). Data mining draws from two major sources data base and machine learning technologies (Fayyad, Piatetsky-Shapiro, Smyth, 1996). The goal of machine learning is to construct computer programs that automatically improve with experience (Mitchell.1997). Detecting fraudulent credit card transactions is one of the successful applications of machine learning (Kovalerchuk & Vityaev, 2000). Many other are known in finance and other areas (Mitchell, 1999).

Nowadays, data mining (DM) techniques are widely used in financial and business area. Data mining resources are from financial data which collected by many organizations like banks, stock exchange authorities, taxation authorities, big accounting and auditor offices specialized databases, etc and in some cases are publicly available (Kirkos & Manolopoulos, 2004). The American Institute of Chartered Public Accountants has identified data mining as one of the top ten technologies for tomorrow and the Institute of Internal Auditors has listed DM as one of the four research priorities (Koh, 2004). DM can be used to solve typical problem in accounting such as corporate bankruptcy, credit risk estimation, going concern reporting, financial distress and corporate performance prediction (Kirkos & Manolopoulos, 2004).

Data mining learning paradigms have been derived from machine learning paradigms, which is using three components namely knowledge representation, forecast performer and learning mechanism (Kovalerchuk & Vityaev, 2000). DM steps is described in the picture below.

**Figure1. Data Mining Steps**




Data mining is a procedure of seek and find several models, summarizing that has below stages (Rostami et al., 2011):

1. Hypothesis formulation in this stage, one modeler usually points one set of variables for dependency function, in this stage, it can be proposed several hypos for one theorem.
2. Data collecting. This stage is about making and collecting data. Totally, there are two possibilities. First is when possibility that is data making process under modeler control. Second possibility is shown when professional cannot effect the production in procedure. This method is called visual method.
3. Data preprocessing. In visual method, data usually collected from data marts in data warehouses and data centers. Data preprocessing usually contains below two stages: Omitting unusual data Characters, encrypting and selecting
4. Model evaluation. The principle work of data mining in this stage is selecting suitable techniques. This process is

not usually obvious. Its' implementing is upon several models and selecting best of them is an additional role and work.

5. Model interpretation and making conclusion.

Data mining techniques in Audit:

Rules of Data Mining Extraction explain below (Ott et al., 2008):

1. The IT audit must ensure that the source of the data is extracted as early as possible in the data creation process.
2. Auditor must fully understand all the data elements. The auditor should consult with business analysis to document each data element, including its significance to the critical success factors of the enterprise

Initiating Data Mining Technology (Ott et al., 2008): The methods employed by the IT audit group to initiate a data mining exercise could result in a full-fledged continuous auditing process requiring scheduled hours. Audit management should properly plan and have a reasonable perspective before embarking on a data mining exercise. As with all projects, adequate controls should be established, including project management and system development life cycle controls. Audit management should ensure that all auditors utilize data mining to better understand potential risks within the various financial and operational processes.

Direct analysis in search of questionable occurrences of values within the data is the most common data analysis method employed in data mining. Specifically, data analysis usually begins by searching the data files for of data indicating potential fraud, waste and abuse. Example of specific occurrences is risk associated with revenue, inventory, and disbursement. The statistical methods are more adaptive to problems, whereas, computational methods are quite inflexible to few problems in financial fraud domain (Mandal et al., 2016).

**CHALLENGE**

According to previously discussed, data mining (DM) had been used in financial and auditing area. The explanation about challenge will be presented below.

1. The data miners' challenge is to find trends quickly while they are valid, as well as to recognize the trends are no longer effective. Auditors must choose the appropriate data mining algorithm based on the purpose and problem to be solved. Choosing the wrong algorithm can produce irrelevant or useless information. Therefore, auditors must have sufficient knowledge of different data mining algorithms and how they can be used in different situations.
2. (Kovalerchuk & Vityaev, 2000). Intellectual challenges in data mining: The effective knowledge representation is an important problem for the success of data mining. Thus, the conceptual challenges in data mining are: Proper formulation of the problems and crafting the knowledge representation to make learning meaningful and tractable. (Vasarhelyi et al., 2015)Auditors face challenges when conducting audits due to the increasing volume and complexity of data. Data mining can help auditors to analyze large datasets and identify patterns and anomalies that may indicate fraud or errors. Auditors need to have knowledge and skills in data mining and analytics to be able to effectively identify these risks. That knowledge can improve the potential benefits of using data mining in auditing, such as improved audit quality, increased efficiency, and better risk management.
3. DM methods most likely to detect these indicators. The use of DM for financial audits is a management decision that requires consideration of its cost to the business (Saglar & Kefe, 2021). Accurate Interpretation of Results: After the data mining process is complete, auditors must be able to interpret the results correctly. Auditors need to understand what the results mean and their implications for the audit work being conducted. Misinterpreting the results can lead to incorrect conclusions and can jeopardize the integrity of audit work.

**OPPORTUNITY**

According to previously discussed, data mining (DM) had been used in financial and auditing area. In auditing, two major benefits from DM that mostly applied are forecasting and fraud detection purposes. The explanation will be presented below. 1. Data Mining for Financial Forecasting Purposes Data mining is a solution to solve financial forecasting difficulties faced by financial professional or financial researcher (Kovalerchuk & Vityaev, 2000). Financial forecasting is important information for the investors and creditors in decision making process. Data mining try to enhance the study about time series prediction (one of financial forecasting ultimate challenge) with new approach In auditing financial statement report, auditor must have to make statement about company (business organization) sustainability or going concern ability. This statement expected to enhancing investor and creditor trust to the company financial performance in the future. The auditors need relevant internal and external information to build company going concern opinion. External information, include news, research articles, website materials, conversation, and social media, provides massive amounts of material so auditors need to make a pattern. This situation named Big Data, and one way to analyzed it is by DM method. DM techniques have been employed to facilitate the auditing process, to predict corporate performance, and to facilitate credit risk estimation (Kirkos & Manolopoulos, 2004). Auditing techniques using DM are developing as a promising way to improve audit quality. DM modern trends in auditing are to embrace the concept of business risk, which emphasizes the strategic objectives of a business enterprise (Kirkos & Manolopoulos, 2004). To take advantage of DM techniques, the auditor must first understand the strategic objectives and apply them to business processes. DM techniques such as Neural Networks, Genetic Algorithms,

Case Base Reasoning, and fuzzy logic may facilitate this new risk-based auditing approach

2. Data Mining for Fraud Detection Purposes

Fraudulent financial statement is a deliberate mistake with a specific purpose, such as covering up poor financial performance. With the detection of financial fraud, possible destructive effects of financial fraud are prevented (Saglar & Kefe, 2021). There are hopes that DM can be tools to facilitate the auditor's performance in detecting fraud.

DM methods provide benefits in many aspects during the auditing of financial statements.

First, the increased emphasis on fraud detection in audits by regulators and standard setters provides motivation to identify and use tools to increase auditor efficiency.

Second, the increasing use of DM tools as a forensic tool in accounting firms means that there is a growing population of people in firms with DM experience and a general awareness of DM (Saglar & Kefe, 2021).

Data mining is a technique that provides specific information that can detect weaknesses in controls (Ott et al., 2008). Internal control has important role to prevent financial fraud such as fraudulent financial statement. Furthermore, Ott at al described that an objective of data mining techniques is to uncover patterns indicating a broken process and/or develop predictive patterns in business information. The first objective is for the auditors to know the purpose of each data element, including how collective data patterns play a role in business decision (Ott et al., 2008).

Gray and Debreceny (2014) emphasized four points for the use of DM techniques in fraud detection (Saglar & Kefe, 2021). These are analysis of fraud risks, identification of possible fraudulent methods, determination of fraud indicators and fraudulent methods used selection of appropriate DM methods most likely to detect these indicators. By using the available data as a cache, fraudulent transactions and fraud can be detected and costs can be reduced (Saglar & Kefe, 2021). According to literature studies, Artificial Neural Network (ANN) and Decision Tree (DT), which are data mining methods, are mostly preferred during audit activities (Saglar & Kefe, 2021).

In the literature review, it was seen that artificial neural networks, logistic regression, decision trees, support vector machines, genetic algorithms, and text mining are used intensively in auditing activities. While the detection of fraud and fraudulent financial reports was examined within the scope of internal audit and external audit activities, it was determined that artificial neural networks and decision trees methods were mostly used.

Ultimately, the real value of data mining is educating the business process owner on the means and methods of identifying fraud, waste and abuse, so it can be embedded within the organization’s management controls (Ott et al., 2008).

Other opportunities of data mining in auditing are from financial benefit perspective. Potential Financial Benefits of Using Data Mining Techniques are explain below (Ott et al., 2008):

a. Reduce external audit fees. For example, the IT internal auditor may use data mining to validate interface software that performs data transfers between systems.
b. At the request of management, data mining can be used to validate a known control such as a preventive and detective duplicate payment control within the accounts payable system (disbursement process).
c. Reduce the need for auditors to travel to a work site, thus reducing travel expenses for the company. In addition, time is saved by not requesting business management to supply unnecessary supporting documentation when the process is efficient and effective based upon the values noted from performing a data mining analysis.

## CONCLUSION

The main focus of Data Mining is to use the data assets of the business to obtain useful financial and non-financial information (Saglar & Kefe, 2021). DM methods allow for the efficient processing of data with missing values, irrelevant data or highly correlated datasets automatically (Saglar & Kefe, 2021). Thus opportunities are the evolution of audit technique. It is relevance to big data that auditor must face on auditing works.

Data mining methods used in the audit field are mostly focused on prediction (Saglar & Kefe, 2021). Further, Saglar and Kefe explained that assurance and compliance-oriented DM practices primarily focus on three main topics namely auditing, business health, and forensic accounting. Auditing mainly includes engagement, planning, conducting, and post-auditing phases. Business health comprises financial viability, bankruptcy, and going concern.

Data mining methods widely used by auditor to detecting financial fraud include fraudulent statement during their work in financial statement auditing. Support vector machine, Neural network are to be popular among researchers due to their adaptability to new techniques, and evolving tactics of committing frauds (Mandal et al., 2016).

Other opportunities from practicing data mining in audit are that this technique gives auditor and company to reduce audit cost efficiently. The patterns uncovered using data mining help organizations make better, timelier and more profitable decisions (Ott et al., 2008).

## REFERENCES

IAASB. (2021). *Handbook of international quality control, auditing, review, other assurance, and related services pronouncements.*

Kirkos, E., & Manolopoulos, Y. (2004). Data Mining in Finance and Accounting : a Review of Current Research Trends. *On Enterprise Systems and Accounting*, 63–78.

Kovalerchuk, B., & Vityaev, E. (2000). *Data*

*Mining in Finance Advances in Relational and Hybrid Methods*. Kluwer Academic Publisher.
Mandal, P., Mahata, A., Biswas, B., Pal, U., Sarfaraj, Md. A., & Barman, S. (2016). A complete literature review on financial fraud detection applying data mining techniques. *International Journal of Trust Management in Computing and Communications*, *3*(4), 336. https://doi.org/10.1504/ijtmcc.2016.10005490
Mgregson. (2017). *Understanding a financial statement audit | 1 Understanding a financial statement audit*. www.pwc.com
Ott, B. J., Macleod, A., & Fan, K. M. (2008). *Computer-assisted Audit Techniques: Value of Data Mining for Corporate Auditors*. 3–6.
Rostami, K. H., . H. O., . A. K. M., . H. A., & . H. N. (2011). Data Mining and Application in Accounting and Auditing. *Journal of Education and Vocational Research*, *2*(6), 211–215. https://doi.org/10.22610/jevr.v2i6.42 Saglar, J., & Kefe, İ. (2021). A Review on Data Mining Methods Used in Internal Audit and External Audit. *Ekev Akademi Dergisi*, *88*, 259–274. https://doi.org/10.17753/ekev2157 Singh, A., Singh, B., & Singh, R. B. (2022). *Sampling Approach in Accounting: Audit Sampling* (pp. 67–76).
Vasarhelyi, M. A., Kogan, A., & Szegletes, M. N. (2015). Data Mining in Auditing: An Overview and Future Directions". *Journal of Emerging Technologies in Accounting*.