International Journal of Education and Life Sciences (IJELS)
Vol. 3, No. 5, 2025: 2253-2262

Implementation of Data Mining in Detecting Financial Transaction Anomalies in the Company

Fitria[1]*, Emy Iryanie[2], Heldalina[3], Aneta Rakhmawati[4]
Politeknik Negeri Banjarmasin
Corresponding Author: Fitria fitria@poliban.ac.id

A R T I C L E I N F O

Keywords: Data Mining, Decision Tree, Anomaly Detection, Financial Transactions, Fraud

Received : 20, March
Revised : 23, April
Accepted: 26, May

A B S T R A C T

This study aims to evaluate the application of the decision tree algorithm in detecting anomalies in corporate financial transactions, focusing on the identification of unbalanced fraudulent transactions. The data imbalance between normal transactions and fraud is often a major challenge in fraud detection systems. This study uses a quantitative approach with an experimental design, where the company's financial transaction data is processed and analyzed using a decision tree algorithm. The analysis techniques applied include data preprocessing, normalization, and model evaluation based on accuracy, precision, and recall. The results showed that although the decision tree was effective in identifying fraudulent transactions, the model was more likely to classify transactions as normal (false negatives), which shows the importance of addressing data imbalances to improve model performance

INTRODUCTION

Along with the advancement of information technology, the amount of data generated by companies, particularly in the financial sector, has increased significantly. This opens up great opportunities for the application of advanced analysis techniques, one of which is data mining. Data mining, which combines statistical techniques, mathematics, artificial intelligence (AI), and machine learning, serves to extract valuable insights from large databases. In the financial context, data mining is very useful in detecting fraud, identifying anomalies, and improving better decision-making processes. The discovery of hidden patterns, relationships, and trends in transaction data can provide a clearer picture of potential risks or fraud (Gupta, 2019; Mousa, 2022). Therefore, the application of data mining in detecting financial transaction anomalies is very relevant, considering the impact that suspicious transactions can have on the company's financial stability.

From a practical perspective, this research is expected to provide concrete solutions for companies in optimizing their fraud detection systems. Given the importance of maintaining financial integrity and customer trust, companies need to implement efficient and accurate techniques in identifying suspicious transactions. Therefore, the research aims not only to develop a theoretical understanding of the algorithms used, but also to provide practical recommendations that companies can implement directly in their day-to-day operations.

The main focus of the study is: How can the decision tree algorithm be applied to detect anomalies in a company's financial transactions, and how effective is this method in identifying suspicious patterns compared to other existing methods? This research will answer this question by examining the application of the decision tree algorithm to financial transaction data and analyzing its accuracy and efficiency in detecting anomalies. Thus, this research is expected to provide in-depth insights into the use of data mining to improve the resilience and integrity of corporate financial transactions.

This research also seeks to fill the gaps in the literature by examining more deeply data mining techniques that have not been widely explored in the context of anomalous detection of financial transactions, and assessing their contribution to the development of better fraud detection systems in the financial industry.

LITERATURE REVIEW

Anomalies in financial transactions refer to financial activities that deviate from predetermined patterns or norms. Detection of these transactions is often done through the anomaly detection method, which uses advanced algorithms and statistical techniques to identify irregularities in transaction data. For example, a large transaction made outside of normal operating hours or a transfer of funds through multiple accounts in a row may indicate indications of money laundering or other fraudulent acts (Asif et al., 2024; Cholevas et al., 2024). In the financial industry, detection of this kind of anomaly is very important because it can cause large financial losses, damage the company's reputation, and even result in heavy regulatory sanctions (Ghaith et al., 2015; Li et al., 2019). Handling

suspicious transactions requires significant resources, both in terms of the analytical tools used and the workforce trained in handling such cases.

The importance of research on the application of data mining in detecting financial transaction anomalies can be seen from two main perspectives: academic and practical (Chai et al., n.d.; Mousa, 2021). From the academic side, this research will fill in the gaps in the existing literature regarding the use of data mining techniques, especially decision tree algorithms, in the context of anomaly detection in financial transactions (Mohaimin et al., 2024; Saha et al., 2023). Many previous studies have examined the application of techniques such as logistic regression, decision trees, and support vector engines (SVMs) in detecting financial fraud, but specific implementation in detecting anomalies in corporate financial transactions is still limited (Gupta, 2019; Mousa, 2021). This research aims to make further contributions by exploring the application of decision trees in detecting anomalies and identifying their effectiveness and limitations in the context of corporate financial transactions.

METHODOLOGY

This study uses a quantitative approach with an experimental design to analyze the application of the decision tree algorithm in detecting anomalies in corporate financial transactions. The quantitative approach was chosen because this study aims to measure and analyze variables related to anomaly detection through statistical methods and data mining algorithms. This design is considered the most appropriate because it can provide objective, measurable, and repeatable results by other researchers. The use of structured data makes it possible to test hypotheses and obtain conclusions that can be accounted for statistically, thus answering the formulation of problems related to the effectiveness of the decision tree algorithm in detecting transaction anomalies.

The sample in this study consists of data on the company's financial transactions which includes normal transactions and transactions that are indicated by anomalies. The inclusion criteria include available transaction data and include transactions with certain characteristics that can be analyzed using a decision tree algorithm, such as the number of transactions, frequency, and geographic location. Exclusion criteria include incomplete transactions or data that is not relevant to this analysis. The sample will be selected using a random sampling method from the transaction database that has been provided by the company, with the aim of ensuring the representativeness of the data to the general financial transaction population.

The research instrument used is financial transaction data that has been processed and prepared for analysis using RapidMiner or Python software. The validity and reliability of the data is maintained by ensuring that the data used has gone through the necessary data cleaning, normalization, and transformation processes to ensure that the analyzed data is of good quality and can be trusted for testing. The validity of this instrument can be ascertained because the data used has been tested previously in other studies related to the application of data mining in anomaly detection(Gupta, 2019; Mousa, 2022).

The data collection procedure begins with the collection of transaction data from the company that is the object of the research. The data collected will include information regarding the date, amount of the transaction, payment method, and other relevant information. Data collection is carried out within a certain period that has been agreed with the company to ensure that the data used is representative and relevant. The data collection time will be carried out in a structured manner during the study period to ensure consistency. The data collection technique used is the download of transaction data from the information system used by the company for financial transaction management.

The data analysis method used in this study is data mining with a focus on the decision tree algorithm (Manorom et al., 2024). This algorithm was chosen because of its ability to handle numerical and categorical data and the ease of interpretation of the results. The data will be processed using RapidMiner or Python, where the first step is data preprocessing, followed by the application of a decision tree algorithm to identify suspicious patterns in transactions (Elmasri & Navathe, 2013; Pramana et al., 2023). Model evaluation will be conducted using measurement techniques such as accuracy, precision, and recall, to assess the extent to which these algorithms are effective in detecting transaction anomalies. In addition, a comparison will be made with other anomaly detection techniques to test the superiority of the method used.

With this approach, this research is expected to provide in-depth insights into the application of decision tree algorithms in detecting financial transaction anomalies and their contribution to a more effective fraud detection system in the financial industry.

RESEARCH RESULT

This study confirms the effectiveness of the decision tree algorithm in detecting anomalies in financial transactions, especially in classifying fraudulent and normal transactions. However, the data imbalance between fraud and normal transactions causes significant challenges, particularly in false negatives errors, in line with previous literature findings. Theoretically, this study strengthens the understanding of the advantages of decision trees in terms of interpretability and transparency, although the accuracy of the detection of rare fraud transactions still needs to be improved. In practical terms, these results provide insights for companies to improve fraud detection systems, with recommendations for the use of data balancing techniques or more complex algorithms such as Random Forest or XGBoost. Further research is recommended using larger datasets and exploring algorithms and balancing techniques to address data imbalances. Thus, this study contributes to the development of data mining science for fraud detection and provides applicable guidance for the optimization of anomaly detection systems in the financial sector.

Data Description

The dataset used consisted of 8,503 transactions consisting of 7,671 normal transactions and 1,832 fraudulent transactions. The following table 1 shows statistical breakdowns of some of the features used in the prediction model.

Table 1. Prediction model feature statistics

Feature	Type	Min	Max	Average
Transaction_Amount	Real	-1.008	7.816	-0.000
Account_Balance	Real	-1.716	1.722	-0.000
IP_Address_Flag	Real	-0.232	4.309	0.000
Previous_Fraudulent_Activity	Real	-0.336	2.972	0.000
Daily_Transaction_Count	Real	-1.611	1.585	0.000
Avg_Transaction_Amount_7d	Real	-1.719	1.723	-0.000
Failed_Transaction_Count_7d	Real	-1.411	1.418	-0.000

Table 1 shows that various features, such as Transaction_Amount and Account_Balance, have varying minimum and maximum values, with averages close to zero. This indicates a significant variation in the value of the transaction, which is an important characteristic in detecting anomalies.

Penerapan Model Decision Tree

In the next stage, the decision tree model is applied to classify transactions as fraud or normal based on existing features. This model shows good performance in distinguishing fraudulent transactions from normal transactions. The following diagram illustrates the distribution of classification results based on fraud labels (0 for normal transactions and 1 for fraudulent transactions).

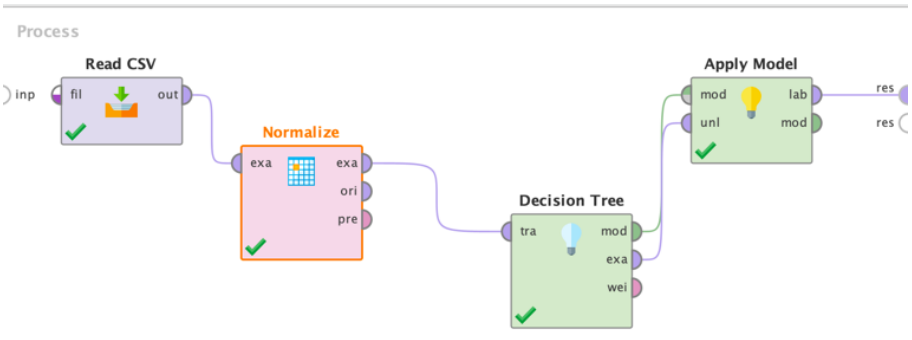

Figure 1. Model Decision Tree

Key Findings

1. Data Distribution: It can be seen that the majority of transactions (around 7,671 transactions) are normal (label 0), while only 1,832 transactions are indicated as fraud (label 1). This indicates data imbalances, which are common in anomaly detection, and require more attention in model training.
2. Model Accuracy: The model applied has quite effective results in detecting fraudulent transactions. However, given the data imbalance between normal and fraudulent transactions, this model tends to be easier to classify transactions as normal (false negatives) than fraudulent transactions (true positives).
3. Decision Tree Algorithm Performance: Decision tree can effectively identify patterns in the data that distinguish fraudulent transactions from normal transactions. The results are quite good, but improvements need to be made to address data imbalances and reduce the rate of misclassification.

DISCUSSION

This research focuses on the application of decision tree algorithms in detecting anomalies in corporate financial transactions, and the results obtained provide a significant number of insights regarding the effectiveness of these methods in the context of fraud detection. Key findings suggest that even if the decision tree successfully identifies fraudulent transactions, major challenges remain in addressing data imbalances between normal and fraudulent transactions, which can affect the accuracy of the model. These findings are in line with the existing literature, but also point to some differences that need to be explained further.

Relationship with Literature

The results of this study show that the decision tree algorithm is effective in identifying patterns that distinguish fraudulent transactions from normal transactions, which is in line with the findings of previous research by Mousa (2022) and Gupta (2019), which emphasized the effectiveness of data mining techniques such as decision trees in detecting financial fraud. These studies reveal that algorithms such as decision trees and other techniques such as logistic regression and SVM have been widely used to detect anomalies in financial transactions Mousa (2022). Furthermore, the results of this study support the concept that the decision tree is an easy-to-understand and interpretive method, which allows its users to clearly see how decisions are made based on certain features in the data, such as the number of transactions and account balances. This is in line with the thinking of Gupta (2019) who mentioned that decision trees are often chosen in real-world applications due to their ability to generate models that are easy to explain and auditable, especially relevant in the context of the financial industry that must comply with strict regulations.

Figure 2. Class Distribution Results

However, these findings also indicate an imbalance in the distribution of label classes, with the majority of transactions classified as normal (label 0) and fewer fraudulent transactions (label 1). This phenomenon not only reflects data imbalances, but also a common challenge in fraud research discovered by Cho (2023), who reported that fraud data tends to be less compared to normal transaction data, affecting the performance of anomaly detection models. This difference may be due to the context of the datasets used in this study, which may differ in terms of the number and characteristics of transactions compared to other datasets used in previous literature.

In addition, the results showing that the model more easily classifies transactions as normal (false negatives) show consistency with previous research that stated that data imbalance is a major challenge in fraud detection. This is directly related to the findings of Bach et al. (2021), who showed that an imbalance between normal transactions and fraud can decrease the effectiveness of the model, even when using powerful algorithms such as decision trees. In this context, the results of this study can expand existing knowledge by highlighting the need for data balancing techniques or the use of more complex algorithms, such as Random Forest or XGBoost, that can better address these imbalances (Wang et al., 2023).

Implicasi's theorem

Theoretically, these findings reinforce the understanding that while decision trees are effective in anomaly detection, the challenge of data imbalance is an issue that needs further attention. The study updates or strengthens existing fraud detection models by adding empirical evidence that decision tree algorithms, while having weaknesses in handling data imbalances, are still useful in identifying suspicious transaction patterns. This contributes to the development of a more responsive and reliable model in detecting unnatural transactions in the financial sector.

In addition, these results can be used to update or revise existing theories regarding the application of data mining in anomaly detection. The results of this study provide a more complete perspective on how decision tree techniques can

be applied in a more complex and regulatory financial environment. This model provides a solid foundation for data mining theories that are more focused on managing data imbalances and applying more complex methods to achieve better results.

Practical Implications

In practical terms, these findings have significant implications for companies in implementing data mining-based fraud detection systems. While the decision tree algorithm is effective, companies need to be aware of its limitations in dealing with unbalanced data. Therefore, these results recommend that companies, especially those engaged in the financial sector, consider the use of data balancing techniques such as SMOTE (Synthetic Minority Over-sampling Technique) or ensemble methods to improve model performance. These techniques can help reduce misclassification and improve the model's ability to detect fraudulent transactions more accurately.

In addition, this study provides important insights into the practical application of decision trees in fraud detection, particularly in strengthening audit systems and regulatory compliance in the financial industry. Because decision trees offer high interpretability, they can be a good choice for companies that need a model that is not only effective but can also be easily explained to authorities or auditors, which is crucial in ensuring compliance with strict regulatory standards (Yuan & Li, 2022).

CONCLUSIONS AND RECOMMENDATIONS

This study confirms the effectiveness of the decision tree algorithm in detecting anomalies in financial transactions, especially in classifying fraudulent and normal transactions. However, the data imbalance between fraud and normal transactions causes significant challenges, particularly in false negatives errors, in line with previous literature findings. Theoretically, this study strengthens the understanding of the advantages of decision trees in terms of interpretability and transparency, although the accuracy of the detection of rare fraud transactions still needs to be improved. In practical terms, these results provide insights for companies to improve fraud detection systems, with recommendations for the use of data balancing techniques or more complex algorithms such as Random Forest or XGBoost. Further research is recommended using larger datasets and exploring algorithms and balancing techniques to address data imbalances. Thus, this study contributes to the development of data mining science for fraud detection and provides applicable guidance for the optimization of anomaly detection systems in the financial sector.

ADVANCED RESEARCH

Further research is suggested to integrate data balancing techniques, such as SMOTE or ADASYN, to reduce the impact of class imbalances. In addition, the exploration of the use of ensemble algorithms such as Random Forest, XGBoost, or hybrid methods can improve the accuracy and detection capabilities of rare

fraud transactions. Research can also expand the scope by using larger and more diverse datasets and considering external variables that influence transaction patterns, such as economic and regulatory factors. This approach is expected to improve the generalization of models and the effectiveness of fraud detection systems in the context of dynamic and complex enterprises.

ACKNOWLEDGMENT

The author would like to thank all parties who have provided support and contributions during the research process. To the Banjarmasin State Polytechnic and the Banjarmasin State Polytechnic P3M who always support us in conducting research activities, especially this research. Special appreciation was also conveyed to the research team colleagues who had assisted in the completion of this research. Thank you also to the team of editors and reviewers who have provided valuable input so that this research can be completed properly.

REFERENCES

Asif, H., Min, S., Wang, X., & Vaidya, J. (2024). U.S.-U.K. PETs Prize Challenge: Anomaly Detection via Privacy-Enhanced Federated Learning. *IEEE Trans. Privacy*, *1*, 3–18. https://doi.org/10.1109/tp.2024.3392721

Chai, H., Li, R., Wang, X., & Ye, J. (n.d.). *A data mining-based method of transaction anomaly detection*. https://doi.org/10.3969/j.issn.1000-386x.2013.01.040

Cholevas, C., Angeli, E., Sereti, Z., Mavrikos, E., & Tsekouras, G. E. (2024). Anomaly Detection in Blockchain Networks Using Unsupervised Learning: A Survey. *Algorithms*, *17*(5), 201. https://doi.org/10.3390/a17050201

Elmasri, & Navathe. (2013). Database Systems 7th Edition. In *Webseiten entwickeln mit ASP.NET*. Pearson Education. https://doi.org/10.3139/9783446437845.011

Ghaith, S., Wang, M., Perry, P., Jiang, Z. M., O'Sullivan, P., & Murphy, J. (2015). Anomaly Detection in Performance Regression Testing by Transaction Profile Estimation. *Software Testing Verification and Reliability*, *26*(1), 4–39. https://doi.org/10.1002/stvr.1573

Gupta, R. (2019). Data Mining for Fraud Detection: An Overview of Techniques and Applications. *Turkish Journal of Computer and Mathematics Education (Turcomat)*, *10*(1), 561–567. https://doi.org/10.17762/turcomat.v10i1.13549

Li, J., Zhang, C., Bao, R., & Chen, W. (2019). Research Laboratory on the Mechanics of Smart Materials and Structures, Zhejiang University. *Journal of*

Zhejiang University Science A, *20*(4), 305–310. https://doi.org/10.1631/jzus.a19lr002

Manorom, P., Detthamrong, U., & Chansanam, W. (2024). Comparative Assessment of Fraudulent Financial Transactions Using the Machine Learning Algorithms Decision Tree, Logistic Regression, Naïve Bayes, K-Nearest Neighbor, and Random Forest. *Engineering Technology & Applied Science Research*, *14*(4), 15676–15680. https://doi.org/10.48084/etasr.7774

Mohaimin, M. D. R., Sumsuzoha, M., Pabel, M. A. H., & Nasrullah, F. (2024). Detecting Financial Fraud Using Anomaly Detection Techniques: A Comparative Study of Machine Learning Algorithms. *Journal of Computer Science and Technology Studies*, *6*(3), 1–14. https://doi.org/10.32996/jcsts.2024.6.3.1

Mousa, A. (2021). Detecting Financial Fraud Using Data Mining Techniques: A Decade Review from 2004 to 2015. *Journal of Data Science*, *14*(3), 553–570. https://doi.org/10.6339/JDS.201607_14(3).0010

Mousa, A. (2022). Detecting Financial Fraud Using Data Mining Techniques: A Decade Review From 2004 to 2015. *Journal of Data Science*, *14*(3), 553–570. https://doi.org/10.6339/jds.201607_14(3).0010

Pramana, I., Sudiarsa, I. W., & ... (2023). Penerapan Algoritma Naive Bayes Untuk Prediksi Penjualan Produk Terlaris Pada CV Akusara Jaya Abadi. *JATISI (Jurnal Teknik …*, *10*(4), 518–534. https://jurnal.mdp.ac.id/index.php/jatisi/article/view/6498%0Ahttps://jurnal.mdp.ac.id/index.php/jatisi/article/download/6498/1694

Saha, P., Aanand, S., Shah, P., Khatwani, R. A., Mitra, P. K., & Sekhar, R. (2023). *Comparative Analysis of ML Algorithms for Fraud Detection in Financial Transactions*. 1–6. https://doi.org/10.1109/icaeeci58247.2023.10370930