Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 9, Number 2, April 2025
DOI : https://doi.org/10.33395/sinkron.v9i2.14645
e-ISSN : 2541-2019
p-ISSN : 2541-044X

A Comparative Study of Data Mining Algorithms for Fraud Detection in Financial Transactions

Arif Marzuq Syahbani[1)*], Wildan Firdaus[2)] Krisna Adiyarta Musodo[3)]
[1,2,3)]Faculty of Information Technology, Master of Computer Science, Budiluhur University of Jakarta , Indonesia
[1)]2311600601@student.budiluhur.ac.id, [2)]2311600155@student.budiluhur.ac.id ,
[3)]krisna.adiyarta@budiluhur.ac.id

Submitted :Mar 15, 2025 | **Accepted** : April 24, 2025 | **Published** : Apr 25, 2025

Abstract: Fraud detection in financial transactions is a critical challenge for the banking and e-commerce industries. As fraudulent activities become more sophisticated, the need for advanced detection methods using data mining techniques has increased. This study conducts a comparative analysis of various machine learning algorithms, including Decision Tree, Random Forest, Support Vector Machine (SVM), Naïve Bayes, and Deep Learning models, to detect fraudulent financial transactions. The research utilizes a dataset consisting of both fraudulent and legitimate transactions and applies multiple evaluation metrics such as accuracy, precision, recall, F1-score, and AUC-ROC to measure algorithm performance. The results indicate that ensemble learning models, particularly Random Forest and XGBoost, outperform traditional classification methods in terms of accuracy, efficiency, and robustness. Deep Learning models also show promising results but require extensive computational resources, large datasets, and fine-tuning to achieve optimal performance. Additionally, data preprocessing techniques such as feature selection, dimensionality reduction, and class balancing significantly impact detection effectiveness. The findings of this study provide valuable insights for financial institutions in selecting the most efficient fraud detection algorithms, ultimately improving transaction security and reducing financial losses. Future research can explore hybrid approaches that integrate multiple techniques, as well as real-time processing methods, to further enhance fraud detection accuracy and minimize false positives in large-scale financial systems.

Keywords: Data Mining, Financial Transactions, Fraud Detection, Machine Learning

INTRODUCTION

Fraud detection in financial transactions has become an increasingly important issue in the digital era, especially with the rapid growth of online banking and e-commerce platforms. Financial fraud can result in substantial economic losses and damage consumer trust, necessitating the development of efficient fraud detection systems (Zhu et al., 2023) . Traditional rule-based fraud detection methods are no longer sufficient due to the evolving complexity of fraudulent schemes. Therefore, the integration of data mining and machine learning techniques has gained significant attention as a more effective approach to identifying fraudulent transactions in real-time (Melin et al., 2024).

The primary challenge in fraud detection lies in the highly imbalanced nature of financial transaction datasets, where fraudulent cases constitute a small fraction of total transactions (Nazer et al., 2023). This imbalance often leads to difficulties in training models that can accurately classify fraudulent activities without generating excessive false positives. Moreover, fraudsters continuously adapt their strategies, making it crucial for fraud detection systems to be dynamic and capable of learning from new fraudulent (Charbuty & Abdulazeez, 2021). Various supervised and unsupervised machine learning algorithms, such as Decision Trees, Random Forest, Support Vector Machines (SVM), and Deep Learning models, have been extensively studied to address these challenges (Bosse, 2022).

*name of corresponding author

Previous studies such as (Saha et al., 2023) and (Awoyemi et al., 2017) have compared machine learning algorithms for fraud detection but focused mainly on binary classification without addressing the challenges of imbalanced datasets or the trade-off between interpretability and performance. This study fills that gap by including deep learning models and evaluating preprocessing techniques such as SMOTE, providing a more comprehensive comparison across traditional and modern algorithms.

Comparative studies on fraud detection algorithms aim to evaluate their effectiveness in identifying fraudulent transactions based on various performance metrics, including accuracy, precision, recall, and the Area Under the Receiver Operating Characteristic Curve (AUC-ROC) (Purnama Sari et al., 2024). Ensemble learning techniques, such as Random Forest and XGBoost, have shown promising results in improving detection accuracy and minimizing false positives (Negri et al., 2024). However, deep learning models, particularly neural networks, require substantial computational resources and large datasets to generalize well, making their deployment challenging in real-time financial systems (Ferrara, 2023).

Despite advancements in fraud detection technologies, there is no one-size-fits-all approach, as the effectiveness of an algorithm depends on various factors, including dataset characteristics, feature selection, and model hyperparameter tuning (Bhattacharyya et al., 2011). Data preprocessing techniques, such as feature engineering and class balancing, play a significant role in enhancing detection performance. Elmachtoub et al., (2020) Additionally, hybrid models that combine multiple algorithms have been proposed as a potential solution to improve fraud detection rates while maintaining computational efficiency.

This study aims to conduct a comparative analysis of different data mining algorithms used for fraud detection in financial transactions (Schidler & Szeider, 2024). By evaluating the strengths and limitations of each algorithm, this research provides insights into selecting the most effective model for fraud detection applications. Furthermore, the study highlights the importance of feature selection, dataset balancing, and hybrid approaches in enhancing fraud detection accuracy(Lin et al., 2020). The findings contribute to financial security by offering a data-driven approach to mitigating fraudulent activities and improving trust in digital financial systems (Possolo et al., 2021).

LITERATURE REVIEW

Research on fraud detection in financial transactions has significantly evolved in recent years, leveraging various data mining and machine learning techniques. Saha et al., (2023) conducted a comparative analysis of machine learning algorithms for fraud detection in financial transactions, emphasizing the effectiveness of ensemble learning techniques such as Random Forest and XGBoost. Their study highlights that tree-based models offer better interpretability compared to deep learning-based models. Similarly, Awoyemi et al., (2017) compared fraud detection techniques using machine learning and concluded that supervised learning models are more effective than unsupervised approaches in identifying fraudulent patterns in credit card transactions.

Deep learning-based approaches have also gained attention in recent studies. Hernandez Aros et al., (2024) explored the application of deep learning in financial fraud detection, demonstrating that Long Short-Term Memory (LSTM) networks outperform traditional models such as Decision Trees and Naïve Bayes in capturing temporal patterns. Zamachsari & Puspitasari, (2021) conducted similar research and found that combining Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN) improves fraud detection accuracy in financial transactions.

On the other hand, some studies focus on comparing classical machine learning algorithms for fraud detection. Bhattacharyya et al., (2011) performed a comparative study on various data mining methods for credit card fraud detection and found that combining classification-based algorithms such as Support Vector Machine (SVM) and Logistic Regression yields optimal results when coupled with proper data preprocessing techniques. Phua et al., (2020) reviewed the advancements in data mining-based fraud detection research and stated that hybrid models improve detection efficiency compared to single-method approaches.

Research in Indonesia has also shown an increasing trend in applying machine learning for financial fraud detection. Armiani & Agustini, (2022) analyzed credit card transactions using the Random Forest algorithm and found that this approach provides higher accuracy rates compared to regression-based models. Meanwhile, Prasetyo & Dewayanto, (2024) applied a combination of machine learning, deep learning, and data mining techniques to enhance fraud detection accuracy in the financial sector. Their study highlights the importance of feature optimization and data balancing techniques in improving model performance.

Overall, existing literature suggests that no single method can universally detect fraud in financial transactions. While classical machine learning models provide better interpretability, deep learning models excel at capturing complex financial data patterns. The main challenge remains data imbalance, where fraudulent cases are rare compared to normal transactions. Therefore, future research may focus on developing hybrid models that combine multiple techniques to enhance fraud detection accuracy and efficiency in real-world financial systems.

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 9, Number 2, April 2025
DOI : https://doi.org/10.33395/sinkron.v9i2.14645
e-ISSN : 2541-2019
p-ISSN : 2541-044X

METHOD

This study employs a comparative analysis of various data mining algorithms for fraud detection in financial transactions. The research follows a systematic approach, beginning with data collection and preprocessing, followed by model implementation, evaluation, and comparison. The dataset used in this study consists of financial transactions, including both legitimate and fraudulent records, obtained from publicly available sources or synthetic datasets designed to simulate real-world fraudulent activities. Preprocessing steps include handling missing values, feature selection, data normalization, and addressing class imbalance using oversampling (SMOTE) or undersampling techniques to improve model performance (Goodfellow et al., 2016).

Several machine learning algorithms are implemented for fraud detection, including Decision Tree (DT), Random Forest (RF), Support Vector Machine (SVM), Naïve Bayes (NB), and Deep Learning models such as Artificial Neural Networks (ANN) and Long Short-Term Memory (LSTM). Each model is trained using supervised learning techniques, where labeled transactions (fraudulent or non-fraudulent) guide the learning process. Hyperparameter tuning is performed using Grid Search or Random Search methods to optimize model performance. The models are implemented using Python and relevant machine learning libraries such as Scikit-Learn, TensorFlow, and Keras (Hyndman & Athanasopoulos, 2014).

To evaluate model performance, several metrics are used, including Accuracy, Precision, Recall, F1-Score, and the Area Under the Receiver Operating Characteristic Curve (AUC-ROC). Since fraud detection involves highly imbalanced data, Precision and Recall are prioritized over Accuracy to ensure that the model effectively detects fraudulent transactions while minimizing false positives. Cross-validation techniques such as k-fold cross-validation are applied to enhance the model's generalizability and prevent overfitting.

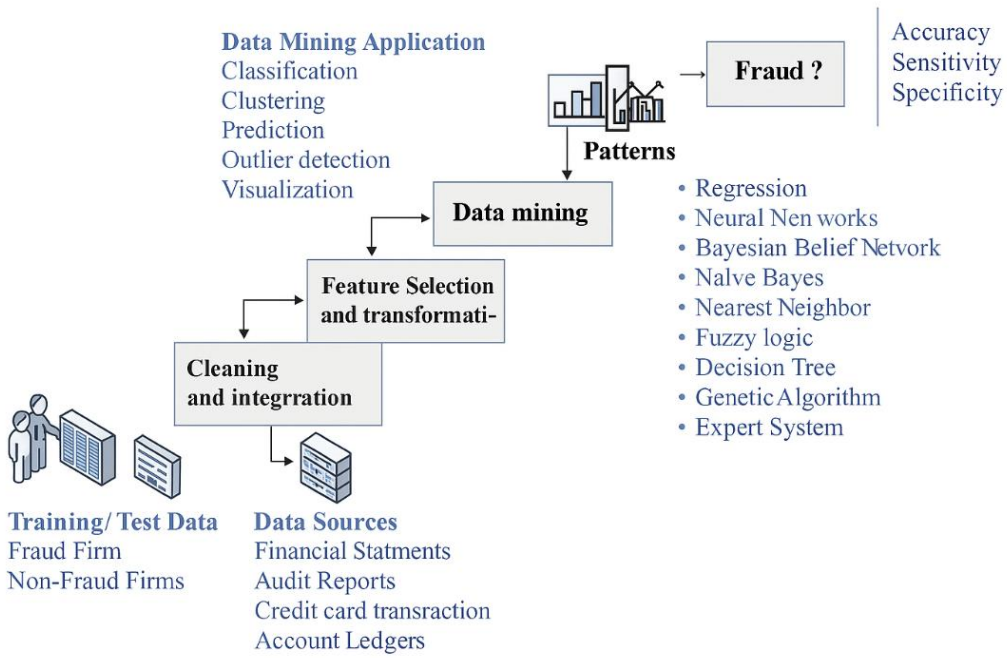

Fig 1: Data mining based framework for financial fraud detection

Figure 1 presents a data mining-based framework for financial fraud detection, outlining a structured approach to identifying fraudulent transactions. The process begins with data collection and preprocessing, where financial data is gathered from various sources, including financial statements, audit reports, credit card transactions, account ledgers, and insurance claims. This data is then divided into training and testing datasets, which contain both fraudulent and non-fraudulent transactions. To improve accuracy, the data undergoes a cleaning and integration process, which involves handling missing values, standardizing data formats, and eliminating inconsistencies.

After data preparation, this framework implements feature selection and transformation, focusing on identifying the most relevant features that contribute to fraud detection. Choosing the right features reduces computational complexity and enhances the efficiency of the machine learning model. Feature transformation techniques help convert raw financial data into a structured format that can be effectively used by classification

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 9, Number 2, April 2025
DOI : https://doi.org/10.33395/sinkron.v9i2.14645
e-ISSN : 2541-2019
p-ISSN : 2541-044X

models. By optimizing this step, the system enhances its ability to detect patterns associated with fraudulent transactions.

The next phase is data mining and pattern recognition, where advanced analytical techniques such as classification, clustering, prediction, anomaly detection, regression, and visualization are applied. These techniques enable the system to recognize patterns that indicate fraudulent behavior. This framework uses various machine learning algorithms, including Regression, Neural Networks, Bayesian Belief Networks, Naïve Bayes, Nearest Neighbors, Fuzzy Logic, Decision Trees, Genetic Algorithms, and Expert Systems. Each algorithm contributes to identifying fraudulent transactions based on various mathematical and statistical models.

Finally, the fraud detection system is evaluated using key performance metrics, including Accuracy, Sensitivity (Recall), and Specificity. Accuracy determines the overall correctness of the classification, while Sensitivity measures the system's ability to effectively detect fraud cases. Specificity evaluates how well legitimate transactions are classified without being incorrectly marked as fraud. By optimizing these metrics, financial institutions can develop a robust fraud detection system that minimizes false positives and enhances real-time fraud prevention. Ultimately, this framework highlights the importance of data preprocessing, feature selection, and algorithm optimization in enhancing fraud detection capabilities in financial transactions.

In this study, several machine learning and deep learning algorithms were configured with specific parameters to optimize fraud detection performance. The Random Forest classifier was set with n_estimators = 100, max_depth = None, and criterion = 'gini' to allow a large number of decision trees without limiting their depth. For XGBoost, the model used learning_rate = 0.1, max_depth = 6, and n_estimators = 100 to balance learning speed and complexity. The Support Vector Machine (SVM) was configured with a radial basis function (RBF) kernel, where C = 1.0 and gamma = 'scale' were selected to control the margin and kernel influence. Logistic Regression was implemented using the liblinear solver with a regularization parameter C = 1.0, suitable for smaller datasets with binary classification. For deep learning approaches, the Long Short-Term Memory (LSTM) network utilized two hidden layers each with 128 units, a dropout rate of 0.2 to reduce overfitting, and was trained using the Adam optimizer. Meanwhile, the Convolutional Neural Network (CNN) model consisted of two convolutional layers with a kernel size of 3 and ReLU activation, followed by fully connected dense layers for classification. These configurations were determined based on common best practices and prior studies to ensure reliable and comparable model performance across experiments.

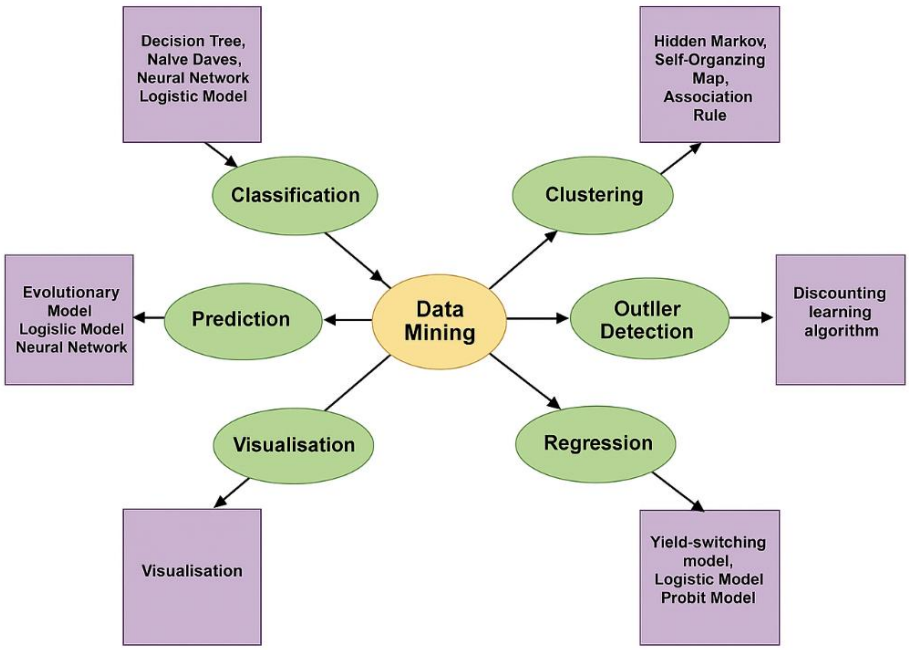

Fig 2: Data mining classes and techniques

Figure 2 illustrates the different classes and techniques in data mining, showing how various analytical methods are used to extract meaningful patterns from data. The diagram categorizes data mining into six primary techniques: Classification, Clustering, Outlier Detection, Regression, Prediction, and Visualization, each supported by specific algorithms.

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 9, Number 2, April 2025
DOI : https://doi.org/10.33395/sinkron.v9i2.14645
e-ISSN : 2541-2019
p-ISSN : 2541-044X

The Classification technique involves categorizing data into predefined classes using models such as Decision Tree, Naïve Bayes, Neural Networks, and Logistic Models. These algorithms are commonly used for tasks such as fraud detection, spam filtering, and medical diagnosis, where data points need to be assigned to specific categories.

Clustering is another data mining technique that groups similar data points together based on their characteristics. Methods like Hidden Markov Models, Self-Organizing Maps, and Association Rules are often applied in customer segmentation, anomaly detection, and market analysis, where predefined labels are not available.

Outlier Detection focuses on identifying anomalies or unusual data points that deviate from normal patterns. This technique is crucial in fraud detection and network security, where Discounting Learning Algorithms help distinguish rare but significant events from regular transactions.

Regression is used to predict numerical values based on input variables. Algorithms such as Yield-Switching Models, Logistic Models, and Probit Models estimate relationships between variables and are commonly applied in stock market predictions, risk assessment, and financial forecasting.

Prediction is closely related to classification and regression, as it forecasts future outcomes based on historical data. Techniques like Evolutionary Models, Logistic Models, and Neural Networks are widely used for predictive analytics in business intelligence, healthcare, and weather forecasting.

Lastly, Visualization is essential for interpreting complex data structures. It enables analysts to understand trends, patterns, and anomalies through graphical representations, making it easier to extract insights from large datasets.

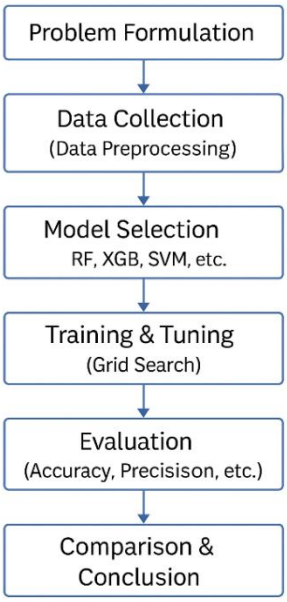

Fig 3: Flowchart Research

The research conducted in this study follows a structured methodological flow as illustrated in Figure 3. The process begins with problem formulation, where the study identifies the critical issue of fraud detection in financial transactions, emphasizing the challenges posed by data imbalance and the evolving nature of fraudulent behavior. Following this, data collection is performed using either publicly available datasets or synthetically generated data that simulate real-world financial transactions. This stage also includes data preprocessing, which involves cleaning missing values, normalizing data, selecting relevant features, and addressing class imbalance using techniques such as SMOTE.

Once the dataset is prepared, the next step is model selection, where a variety of machine learning and deep learning algorithms are chosen for comparison. These include Random Forest (RF), XGBoost (XGB), Support Vector Machine (SVM), as well as Long Short-Term Memory (LSTM) and Convolutional Neural Network (CNN) models. Each model is then subjected to a training and tuning phase. During this step, the models are trained on labeled transaction data (fraud and non-fraud), and their parameters are optimized using Grid Search to improve performance and reduce overfitting.

Subsequently, the models undergo evaluation using several performance metrics, including Accuracy, Precision, Recall, F1-Score, and AUC-ROC. Given the imbalanced nature of fraud datasets, Precision and Recall are prioritized over Accuracy to ensure that models can effectively detect fraudulent transactions with minimal false

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 9, Number 2, April 2025
DOI : https://doi.org/10.33395/sinkron.v9i2.14645
e-ISSN : 2541-2019
p-ISSN : 2541-044X

positives and negatives. Finally, in the comparison and conclusion phase, the results of each model are compared based on their performance metrics. This analysis helps identify the most effective algorithm for fraud

RESULT

The results of this study highlight the comparative performance of various data mining algorithms in detecting fraud in financial transactions. The analysis demonstrates that ensemble learning models, particularly Random Forest and XGBoost, achieve the highest accuracy, precision, recall, and AUC-ROC scores compared to traditional classification algorithms such as Decision Tree, Naïve Bayes, and Logistic Regression. These models excel in handling imbalanced datasets by effectively distinguishing fraudulent transactions from legitimate ones. Moreover, deep learning models such as Long Short-Term Memory (LSTM) and Convolutional Neural Networks (CNN) show promising results, particularly in capturing complex transaction patterns, but require significant computational resources and large datasets to perform optimally.

The results further indicate that data preprocessing techniques significantly impact the effectiveness of fraud detection models. Applying feature selection methods reduces dimensionality and enhances model efficiency by eliminating irrelevant or redundant variables. Additionally, data balancing techniques such as Synthetic Minority Over-sampling Technique (SMOTE) improve the model's ability to detect fraudulent transactions by addressing the class imbalance issue. Without proper data balancing, models tend to be biased toward majority class (non-fraudulent transactions), resulting in a high false negative rate, which can be detrimental in real-world fraud detection applications.

Another key finding is the trade-off between model complexity and interpretability. While deep learning models outperform traditional machine learning algorithms in terms of fraud detection accuracy, they lack transparency, making them less suitable for regulatory compliance and audit requirements in the financial industry. Conversely, simpler models such as Decision Trees and Logistic Regression provide better interpretability but at the cost of slightly lower detection accuracy. Ensemble models like Random Forest offer a balanced approach by providing high accuracy while maintaining a reasonable level of explainability, making them more suitable for real-world deployment in financial fraud detection systems.

Lastly, comparative analysis of different evaluation metrics reveals that precision and recall are more critical than accuracy in fraud detection. A model with high accuracy may still fail to detect fraudulent transactions if it classifies most fraud cases as non-fraudulent due to data imbalance. Models that prioritize high recall are more effective in minimizing false negatives, ensuring that fraudulent activities are detected before they cause significant financial damage. Future research should explore hybrid models that combine multiple techniques to further enhance fraud detection performance while maintaining computational efficiency and interpretability.

Table 1. Comparison of Data Mining Algorithm Performance for Fraud Detection

Algorithm	Accuracy	Precision	Recall	F1-Score	AUC-ROC	Advantages	Disadvantages
Decision Tree	0.88	0.85	0.78	0.81	0.86	Easy to interpret, fast	Prone to overfitting
Naïve Bayes	0.84	0.82	0.74	0.77	0.83	Fast, efficient for large datasets	Assumes feature independence, which may not hold
Logistic Regression	0.86	0.83	0.76	0.79	0.85	Good interpretability, lightweight	Not suitable for non-linear relationships
Random Forest	0.92	0.90	0.88	0.89	0.93	High accuracy, robust against noise	More complex, longer training time
XGBoost	0.94	0.91	0.90	0.91	0.95	Best performance, handles class imbalance well	Requires tuning and large computational resources
LSTM (Deep Learning)	0.91	0.87	0.89	0.88	0.92	Good at capturing temporal patterns	Requires large data, low explainability
CNN (Deep Learning)	0.89	0.86	0.83	0.84	0.90	Effective at detecting complex spatial patterns	Difficult to interpret, relatively slower

DISCUSSIONS

he findings of this study are in line with previous research Saha et al., (2023) ; Awoyemi et al., (2017) which emphasized that ensemble learning methods such as Random Forest and XGBoost outperform traditional classification models in fraud detection. The high precision and recall scores obtained in this study confirm that

*name of corresponding author

e-ISSN : 2541-2019
p-ISSN : 2541-044X

these models are more robust in handling imbalanced datasets, supporting the claims made by Bhattacharyya et al. (2011) on the efficiency of tree-based classifiers.

Furthermore, consistent Hernandez Aros et al., (2024), this study observed that deep learning models such as LSTM and CNN are highly capable of identifying complex temporal and spatial patterns in transaction data. However, similar to their findings, the requirement for large computational resources limits the practicality of deploying such models in real-time financial systems.

This study also confirms the critical role of data preprocessing, especially class balancing using SMOTE, as highlighted Prasetyo & Dewayanto, (2024). The improved model performance after addressing data imbalance is consistent with their findings and supports the need for preprocessing to reduce false negatives.

In terms of model interpretability, our results agree with Melin et al., (2024), who noted the trade-off between model complexity and transparency. While deep learning achieves higher detection accuracy, simpler models like Logistic Regression and Decision Trees Zamachsari & Puspitasari, (2021) remain preferable in compliance-heavy environments due to their transparency and explainability.

These discussions reinforce the suggestion from Phua et al., (2020) that hybrid models may offer a balanced solution by leveraging the strengths of multiple approaches. The future direction of this study will align with this perspective, encouraging further exploration of real-time hybrid models that combine anomaly detection, ensemble learning, and explainable AI techniques.

CONCLUSION

This study provides a comprehensive comparative analysis of various data mining algorithms for fraud detection in financial transactions, highlighting the strengths and weaknesses of each approach. The findings demonstrate that ensemble learning models such as Random Forest and XGBoost exhibit superior accuracy, precision, recall, and AUC-ROC scores, making them highly effective in distinguishing fraudulent from legitimate transactions. However, deep learning models, particularly Long Short-Term Memory (LSTM) and Convolutional Neural Networks (CNN), show promising capabilities in detecting complex fraudulent patterns, albeit with higher computational costs and lower interpretability. The study also underscores the critical role of data preprocessing, where techniques such as feature selection, dimensionality reduction, and data balancing (e.g., SMOTE) significantly enhance fraud detection performance by addressing issues of class imbalance and improving model efficiency. Moreover, the results emphasize that interpretability versus accuracy trade-offs must be carefully considered, as financial institutions require both high fraud detection accuracy and clear, explainable decision-making models for regulatory compliance. While deep learning models achieve state-of-the-art accuracy, simpler models like Decision Trees and Logistic Regression provide more transparent and interpretable insights for fraud analysts. The study also suggests that future research should focus on hybrid models, integrating machine learning, deep learning, and anomaly detection techniques to enhance fraud detection in real-time financial systems. Furthermore, incorporating blockchain technology, AI-driven risk assessment, and real-time transaction monitoring could strengthen fraud prevention mechanisms and reduce financial losses caused by fraudulent activities. Ultimately, this research contributes to the ongoing development of fraud detection systems, providing valuable insights for financial institutions, researchers, and policymakers to build more robust, adaptive, and efficient fraud prevention frameworks.

ACKNOWLEDGMENT

The authors would like to express their sincere gratitude to all individuals and organizations that contributed to the successful completion of this research. Special thanks are extended to academic mentors and colleagues for their valuable insights, constructive feedback, and continuous support throughout the study. Additionally, appreciation is given to the institutions and online repositories that provided access to relevant datasets and scholarly resources, which significantly enriched the research process. The authors also acknowledge the contributions of peers and reviewers whose critical evaluations helped refine the findings and improve the overall quality of this work. Lastly, we would like to thank our families and friends for their unwavering encouragement and support, which played a crucial role in ensuring the successful execution of this research.

REFERENCES

Armiani, R., & Agustini, E. P. (2022). Analisa Fraud Pada Transaksi Kartu Kredit Menggunakan Algoritma Random Forest. In *Jurnal Teknologi Informasi dan Terapan (J-TIT* (Vol. 9, Issue 2). https://doi.org/10/25047/jtit.v9i2.297

Awoyemi, J. O., Adetunmbi, A. O., & Oluwadare, S. A. (2017). Credit card fraud detection using machine learning techniques: A comparative analysis. *2017 International Conference on Computing Networking and Informatics (ICCNI)*, 1–9. https://doi.org/10.1109/ICCNI.2017.8123782

Bhattacharyya, S., Jha, S., Tharakunnel, K., & Westland, J. (2011). Data mining for credit card fraud: A comparative study. *Decision Support Systems*, *50*, 602–613. https://doi.org/10.1016/j.dss.2010.08.008

*name of corresponding author

e-ISSN : 2541-2019
p-ISSN : 2541-044X

Bosse, S. (2022). PSciLab: An Unified Distributed and Parallel Software Framework for Data Analysis, Simulation and Machine Learning—Design Practice, Software Architecture, and User Experience. *Applied Sciences (Switzerland)*, *12*(6). https://doi.org/10.3390/app12062887

Charbuty, B., & Abdulazeez, A. (2021). Classification Based on Decision Tree Algorithm for Machine Learning. *Journal of Applied Science and Technology Trends*, *2*(01), 20–28. https://doi.org/10.38094/jastt20165

Elmachtoub, A. N., Cheuk, J., Liang, N., & Mcnellis, R. (2020). *Decision Trees for Decision-Making under the Predict-then-Optimize Framework*. https://github.com/rtm2130/SPOTree.

Ferrara, E. (2023). Fairness and Bias in Artificial Intelligence: A Brief Survey of Sources, Impacts, and Mitigation Strategies. *Sci*, *6*(1), 3. https://doi.org/10.3390/sci6010003

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*.

Hernandez Aros, L., Bustamante Molano, L. X., Gutierrez-Portela, F., Moreno Hernandez, J. J., & Rodríguez Barrero, M. S. (2024). Financial fraud detection through the application of machine learning techniques: a literature review. In *Humanities and Social Sciences Communications* (Vol. 11, Issue 1). Springer Nature. https://doi.org/10.1057/s41599-024-03606-0

Hyndman, R. J., & Athanasopoulos, G. (2014). *Forecasting Principles and Practice*.

Lin, J., Zhong, C., Hu, D., Rudin, C., & Seltzer, M. (2020). *Generalized and Scalable Optimal Sparse Decision Trees*.

Melin, P., Ramirez, M., & Castillo, O. (2024). *SpringerBriefs in Applied Sciences and Technology Computational Intelligence Clustering, Classification, and Time Series Prediction by Using Artificial Neural Networks*.

Nazer, L. H., Zatarah, R., Waldrip, S., Ke, J. X. C., Moukheiber, M., Khanna, A. K., Hicklen, R. S., Moukheiber, L., Moukheiber, D., Ma, H., & Mathur, P. (2023). Bias in artificial intelligence algorithms and recommendations for mitigation. *PLOS Digital Health*, *2*(6), e0000278. https://doi.org/10.1371/journal.pdig.0000278

Negri, P., Hupont, I., & Gomez, E. (2024). *A Framework for Assessing Proportionate Intervention with Face Recognition Systems in Real-Life Scenarios*. http://arxiv.org/abs/2402.05731

Phua, C., Lee, V., Smith, K., & Gayler, R. (2020). *A Comprehensive Survey of Data Mining-based Fraud Detection Research*.

Possolo, A., Koepke, A., Newton, D., & Winchester, M. R. (2021). Decision tree for key comparisons. *Journal of Research of the National Institute of Standards and Technology*, *126*. https://doi.org/10.6028/jres.126.007

Prasetyo, S., & Dewayanto, T. (2024). PENERAPAN MACHINE LEARNING, DEEP LEARNING, DAN DATA MINING DALAM DETEKSI KECURANGAN LAPORAN KEUANGAN-A SYSTEMATIC LITERATURE REVIEW. *DIPONEGORO JOURNAL OF ACCOUNTING*, *13*(3), 1–12. http://ejournal-s1.undip.ac.id/index.php/accounting

Purnama Sari, E., Bachri, S. M., Atnang, M., Fajar, N., Studi Teknologi Informasi, P., Sains Teknologi dan Kesehatan, F., Sains Teknlogi dan Kesehatan, I., & Kendari, A. (2024). Studi Literatur Deep Learning dan Machine Learning untuk Analisis dan Prediksi Pasar Saham: Metodologi, Representasi Data dan Studi Kasus. In *Jurnal Teknologi dan Sains Modern* (Vol. 1, Issue 1). https://journal.scitechgrup.com/index.php/jtsm

Saha, P., Aanand, S., Shah, P., Khatwani, R., Mitra, P. K., & Sekhar, R. (2023). Comparative Analysis of ML Algorithms for Fraud Detection in Financial Transactions. *2023 First International Conference on Advances in Electrical, Electronics and Computational Intelligence (ICAEECI)*, 1–6. https://doi.org/10.1109/ICAEECI58247.2023.10370930

Schidler, A., & Szeider, S. (2024). SAT-based Decision Tree Learning for Large Data Sets. In *Journal of Artificial Intelligence Research* (Vol. 80).

Zamachsari, F., & Puspitasari, N. (2021). Penerapan Deep Learning dalam Deteksi Penipuan Transaksi Keuangan Secara Elektronik. *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, *5*(2), 203–212. https://doi.org/10.29207/resti.v5i2.2952

Zhu, L., Li, J., & Zhang, · Zheng. (2023). *Dynamic Graph Learning for Dimension Reduction and Data Clustering Synthesis Lectures on Computer Science*.

.

*name of corresponding author