International Journal of Integrated Science and Technology (IJIST)
Vol. 3, No. 7, 2025: 2163-2172

Implementation of Classification Algorithm for Anomaly Detection in Credit Card Transactions with RapidMiner

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

A R T I C L E I N F O

Keywords: Random Forest, Credit Card, Anomaly Detection, Feature Engineering.

Received : 12, May
Revised : 20, June
Accepted: 25, July

©2025 Fitria, Iryanie, Heldalina (s): This is an open-access article distributed under the terms of the Creative Commons Atribusi 4.0 Internasional.

A B S T R A C T

This study evaluates the application of the Random Forest algorithm in the classification of credit card transactions to detect transaction types such as cash_out, cash_in, payment, transfer, and debit. The dataset used is derived from Kaggle.com and includes attributes such as the number of transactions, sender and recipient balances, and transaction types. The results showed an accuracy of 80.63%, with the best performance in cash_out and payments, but difficulties in classifying debit and transfer transactions due to class imbalances. Classroom balancing using smote or undersampling, as well as unsupervised learning techniques, can improve model performance. Improving the model through feature engineering and hyperparameter tuning is also needed to improve the effectiveness of fraud detection.

DOI: https://doi.org/10.59890/ijist.v3i7.126
https://ntlmultitechpublisher.my.id/index.php/ijist

INTRODUCTION

Credit card transaction fraud has become one of the significant issues affecting the financial industry globally. In recent decades, the increasing volume of credit card transactions has created a major challenge for financial service providers in detecting and preventing fraudulent acts. In response to this problem, many efforts have been made to develop an effective and efficient fraud detection system. One widely used approach is programming algorithms, specifically classification algorithms, which have great potential in detecting anomalies in credit card transaction data.

Fraud detection systems on credit card transactions require methods that can handle large and highly varied volumes of data. This is because credit card transaction data generally includes very complex information, including patterns that are difficult for humans to predict or recognize. Therefore, the use of machine learning algorithms, such as Random Forest, Support Vector Machine (SVM), and others, has become a top choice in the analysis of credit card transaction data to identify suspicious behavior or anomalies that may indicate fraudulent activity (1–4).

This study aims to explore and implement classification algorithms, especially Random Forest, in detecting anomalies in credit card transaction data using RapidMiner. Using RapidMiner, which is one of the popular software for data analysis and machine learning programming, the study is expected to provide new insights into the effectiveness of algorithms in detecting transaction fraud. In addition, this study also aims to identify the strengths and weaknesses of algorithms used in the context of large and unbalanced credit card transaction data (5–9).

The importance of this research lies in its ability to improve the accuracy and efficiency in detecting credit card fraud. As technology evolves and transaction patterns become more complex, the classification algorithms used in this study will help credit card service providers to identify suspicious transactions faster and more accurately. In addition, a better detection system can reduce financial losses resulting from fraud as well as increase consumer confidence in the use of credit cards.

Some previous literature shows that the use of algorithms such as Random Forest can provide more accurate results compared to other algorithms such as Logistic Regression and Decision Tree. For example, research by Balogun et al., Dai, and Guo et al., (10–12) emphasizes Random Forest's advantages in handling unbalanced datasets, which is one of the key challenges in credit card fraud detection. In addition, feature selection techniques used in conjunction with these algorithms also improve the performance of fraud detection systems by filtering out the most relevant features (13,14).

However, while algorithms like Random Forest have proven effective, it is important to evaluate whether these techniques are still relevant and can be optimally applied in credit card fraud detection with more dynamic and larger data. Therefore, this study will assess and compare the performance of classification algorithms in anomaly detection of credit card transactions using

RapidMiner, as well as provide recommendations for the development of more efficient detection systems .

The study also seeks to fill in existing research gaps, which often do not discuss in detail how feature selection techniques can contribute to improving algorithm performance in fraud detection. Thus, this study will not only provide new insights in terms of the implementation of classification algorithms but also contribute to the development of better methods in detecting fraud in credit card transaction data.

METHODS

This study aims to implement a classification algorithm, especially Random Forest, in the detection of anomalies in credit card transaction data using RapidMiner software. The methods used in this study include research design, sample selection, research instruments, data collection procedures, and analysis methods. Here is a detailed explanation of each aspect in this research method.

RESEARCH DESIGN

This study uses a quantitative research design with an experimental approach. The experiment was conducted to evaluate the performance of the classification algorithm in detecting anomalies in credit card transaction data. This research focuses on testing and comparing various algorithm models, particularly Random Forest, in detecting suspicious transactions or fraud. RapidMiner was chosen as a software for performing data processing and machine learning algorithm implementation due to its ease in workflow visualization and big data analysis.

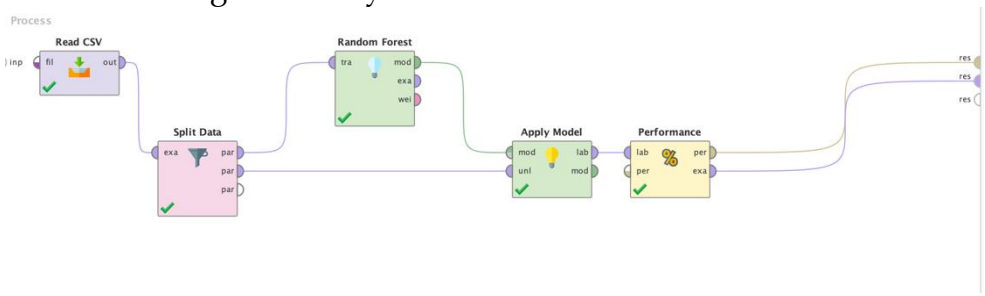

Figure 1. Modelling

Sample/Research Subject

The sample in this study consists of credit card transaction data obtained from credit card service providers or relevant public databases. The data used contains information about credit card transactions, including attributes such as transaction time, number of transactions, location, and customer ID. This dataset usually contains data on legitimate transactions and transactions that are suspected of fraud. The number of samples used in this study is expected to reach thousands of entries to provide a considerable representation of transaction patterns. This data has an unbalanced characteristic, where fraudulent transactions are only a small fraction of the total transactions, creating challenges in terms of analysis and classification.

Research Instruments

The main instrument used in this study is the RapidMiner software, which provides a wide range of tools for machine learning programming and data analysis. RapidMiner allows users to import datasets, process data, select features, train model classification algorithms, and evaluate model performance. The classification algorithm that will be tested in this study is Random Forest. In addition, tools for feature selection in RapidMiner will be used to ensure that only the most relevant features are used in the fraud detection model.

Data Collection Procedure

The data collection procedure in this study involves several steps. First, credit card transaction data is collected from available sources, both from credit card service providers and relevant public datasets. This dataset is then processed to prepare it for analysis, including data cleansing to address missing or inconsistent data issues. Furthermore, the relevant features are selected using the feature selection techniques available in RapidMiner. This process is crucial because selecting the right features can improve the performance of a fraud detection model. Once the data is ready, the data is divided into two sets: training data and test data. Training data is used to train the model, while test data is used to evaluate the model's performance.

Analysis Method

The analysis method in this study involves several stages. First, the training data is used to train the classification algorithm model. RapidMiner allows users to set model parameters, such as the depth of trees in a Random Forest. Once the model has been trained, the next step is to evaluate the model's performance using the test data. The evaluation was carried out by measuring several performance metrics, including accuracy, precision, recall, and F1-score. This metric is used to assess the model's ability to detect fraudulent transactions without producing too many false positive results.

Data processing and analysis will be carried out in a structured manner using RapidMiner, which allows for more efficient modeling and faster testing of various classification algorithms. The study also noted common challenges in credit card fraud detection, such as data imbalances and data privacy issues. Therefore, this study will ensure that the resulting model is not only accurate but can also handle class imbalances in the data.

RESULTS

This study aims to analyze credit card transaction data obtained from Kaggle.com, using the Random Forest algorithm to predict the type of transaction. The data used includes several attributes such as the number of transactions, the type of transaction, the sender and receiver balances, and the IDs of each entity (sender and receiver). This dataset is processed and analyzed using the RapidMiner analysis tool. This study does not use explicit labels like isFraud but focuses more on the classification of transaction types (e.g.: cash_out, cash_in, payment, transfer, debit).

The sample size in this study includes 3,000 transactions that have been processed and divided into training data and test data. The classification process

was carried out using the Random Forest model, which is trained on some of the data to generate predictions of transaction types and tested on other data.

Model Test Results

The following figure shows some of the results and evaluations of the model performed:

Prediction Results of Transaction Type

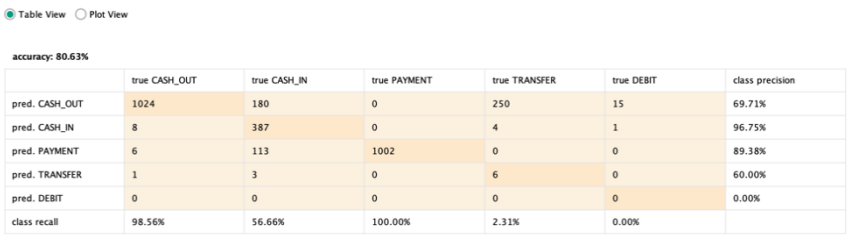

Table View Plot View

accuracy: 80.63%

	true CASH_OUT	true CASH_IN	true PAYMENT	true TRANSFER	true DEBIT	class precision
pred. CASH_OUT	1024	180	0	250	15	69.71%
pred. CASH_IN	8	387	0	4	1	96.75%
pred. PAYMENT	6	113	1002	0	0	89.38%
pred. TRANSFER	1	3	0	6	0	60.00%
pred. DEBIT	0	0	0	0	0	0.00%
class recall	98.56%	56.66%	100.00%	2.31%	0.00%	

Figure 2. Prediction Results

This table shows the prediction results using the Random Forest model for the credit card transaction dataset. From this table, the model has an accuracy of 80.63%, with the best performance in classifying cash_in and payments, with an accuracy of 96.75% and 89.38%, respectively. However, there is difficulty in classifying debit transactions, resulting in 0.00% accuracy, indicating that the model has difficulty recognizing these types of transactions.

a. Confusion Matrix for Classification Models

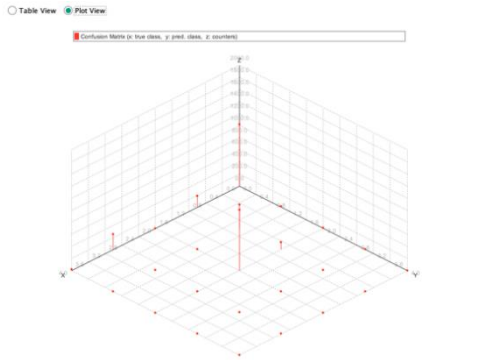

Figure 3. Confusion Matrix

This Confusion Matrix describes the model's performance in classifying different types of transactions, with the X-axis indicating the prediction and the Y-axis indicating the correct class. The model performs better in classifying cash_out types of transactions, but difficulties in classifying debits and transfers, which is reflected in the low number of the associated lines.

Visualization of Scatter Plot Prediction Results

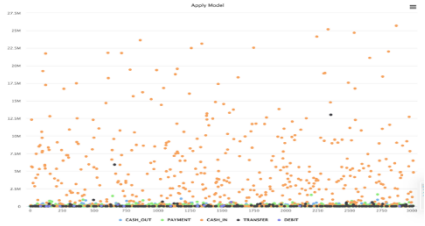

Figure 4. Visualization of Scatter Plot

This scatter plot visualization shows the prediction distribution of transaction types in the form of points based on relevant features (such as the number of transactions and balances). Each color represents a different type of transaction, with more cash_out representing most transactions, while debits and transfers appear less frequently.

Model Accuracy and Evaluation Results

Table 1. Model Accuracy and Evaluation Results

Predicted/ Actual	Cash_Out	Cash_In	Payment	Transfer	Debit
Cash_Out	1024	180	0	250	15
Cash_In	8	387	0	4	1
Payment	6	113	1002	0	0
Transfer	1	3	0	6	0
Debit	0	0	0	0	0

This table shows the distribution of model accuracy across different types of transactions. These accuracy metrics show that the Random Forest model has high accuracy for some types of transactions, but has difficulty in classifying rare transactions, such as debit.

DISCUSSION

In this study, the researcher implemented the Random Forest algorithm for the classification of credit card transactions with the aim of detecting different types of transactions, such as cash_out, cash_in, payment, transfer, and debit. Based on the results described earlier, there are several important findings that need to be discussed further to provide a deeper understanding of the model's performance, the challenges faced, and the implications of these findings in the field of research and industry.

1. Model Performance and Class Imbalance

One of the key findings that emerged in this study was the high accuracy of the model in classifying more common types of transactions, such as cash_out and payments, but limitations in classifying infrequent types of transactions, such as debits and transfers. The Random Forest model was able to achieve an overall accuracy of 80.63%, with the highest accuracy for cash_in (96.75%) and payment (89.38%). Nonetheless, rarer types of transactions, such as debits and transfers, have a very low accuracy of 0.00% and recalls, indicating that the model has difficulty in recognizing and classifying such transactions well. In the analysis using the Random Forest algorithm for the classification of transaction types, one of the key findings that is often reported is the high accuracy in classifying more common transactions. Research by Fernández-Delgado et al. (15) underscores the effectiveness of Random Forest in providing more precise results than other algorithms in classifying complex datasets, including financial transaction data. In addition, Guo et al. (10) showed that Random Forest was able to identify patterns in credit card transactions with a very high degree of accuracy, thanks to its ability to address the problem of high and low data balance.

Another study by Cheng et al. (16) found that the use of more selective features in the Random Forest model can optimize its performance, allowing for better classification of common transaction types such as retail shopping and bill

payments, as well as fraud. This suggests that the selection of the right features and the incorporation of modern machine learning techniques can significantly improve the results of classification models.

This class imbalance is a major challenge in classifying the transaction types in this dataset. Transaction classes such as cash_out occur more frequently in a dataset compared to transfers or debits, which makes it easier for the model to recognize more common transactions and less effective at detecting infrequent transactions. This is a common problem in many real-world datasets, where class distributions are often unbalanced.

The use of class balancing such as SMOTE (Synthetic Minority Over-sampling Technique) or undersampling to balance these classes, as well as the application of a weighted loss function, can be a solution to improve the model's performance in identifying infrequent transactions.

2. *Implications and Significance of Research Results*

This research has important implications for the financial and banking industry in terms of fraud detection and security of credit card transactions. The results show that while Random Forest can provide good accuracy in identifying transactions that occur more frequently, there is great potential to improve the model in detecting more hidden or less frequent frauds. This is a challenge that the industry must face in developing a more effective and efficient fraud detection system. It was also emphasized in the study of Zhang et al. (17) the need for advanced data processing techniques and additional features to improve the model's detection capabilities against unusual anomalies. They found that the integration of data balancing methods such as SMOTE (Synthetic Minority Over-sampling Technique) can be helpful in improving fraud detection, which is less common in unbalanced datasets.

This model, while effective for some types of transactions, still has limitations in terms of generalizability or the ability to handle more complex and fewer transactions. Therefore, an ensemble-based approach, which combines various fraud detection algorithms, may be a better solution. For example, combining Random Forest with anomaly detection techniques such as Isolation Forest or Autoencoders can help in better identifying unusual transactions, even when the data is unbalanced.

In addition, this research opens the potential to integrate blockchain technology or other methods to protect highly sensitive data in financial transactions. These technologies can help reduce the risk of data leaks, which is a major concern in the use of fraud detection technology.

CONCLUSIONS

This study aims to explore and evaluate the application of the Random Forest algorithm in the classification of credit card transactions for fraud detection, using datasets obtained from Kaggle.com. Based on the results obtained, this study provides some insights into the effectiveness of the model in predicting the types of credit card transactions, as well as the challenges faced in handling unbalanced datasets.

In general, the random forest model showed an overall accuracy of 80.63%, which reflects the model's ability to classify more common types of transactions such as cash_out and payments quite well. However, low accuracy for some types of transactions, such as debits and transfers, suggests that the model has difficulty handling less frequent or more complex transactions. This is due to class imbalances in the dataset, where the rarer types of transactions are not represented well enough during the training process.

The confusion matrix and scatter plots presented show that the model is easier to recognize transactions that occur more frequently but fails to recognize the types of transactions that occur less frequently. This emphasizes the importance of balancing classes in the dataset to improve the model's performance in detecting unusual transactions. Therefore, the use of SMOTE (Synthetic Minority Over-sampling Technique) or majority-class undersampling can help balance the dataset and improve the model's ability to identify rare transaction types.

In addition, the high accuracy and recall for more common transactions indicates that the features used in the model, such as the number of transactions, sender and receiver balances, and transaction types, are highly influential in accurately classifying transactions. However, for rarer types of transactions, more sophisticated feature engineering, such as adding transaction time variables or customer behavior patterns, is needed to improve the model's performance.

ADVANCED RESEARCH

Suggestions for future research include the development of more effective techniques to deal with unbalanced data, the application of ensemble-based models, as well as the exploration of more sophisticated anomaly detection algorithms to improve the accuracy and reliability of fraud detection. Further research also needs to address the privacy and ethical issues associated with the use of credit card transaction data, as well as integrate fraud detection systems with more advanced security technology solutions to protect sensitive information.

ACKNOWLEDGMENT

The author would like to express his deepest gratitude to all parties who have provided support and contribution in the completion of this research, especially the research team who always helps the implementation of the research until it is completed. The author also expressed his gratitude to the Banjarmasin State Polytechnic and P3M Poliban who always gave us support to carry out research activities. Hopefully, the results of this research can make a positive contribution in the field of fraud detection and improve our understanding of the application of the Random Forest algorithm in the world of finance.

REFERENCES

Ismawati, Fatah Z. Penggunaan Data Mining Untuk Mendeteksi Penipuan Transaksi Kartu Kredit Algoritma Decision Tree. Jamastika. 2025;4(1):95–101.

Syahbani AM, Firdaus W, Musodo KA. A Comparative Study of Data Mining Algorithms for Fraud Detection in Financial Transactions. Sinkron. 2025;9(2):814–21.

Eldo H, Ayuliana A, Suryadi D, Chrisnawati G, Judijanto L. Penggunaan Algoritma Support Vector Machine (SVM) Untuk Deteksi Penipuan Pada Transaksi Online. J Minfo Polgan. 2024;13(2):1627–32.

Paramitha AF, Arimbi YD, Riyanto S, Apriani NF, Siagian AHAM. Classification of Suspicious Financial Transactions Using Light Gradient Boosting Machine Method (LGBM) Based on Social Network Analysis (SNA) Indicators. Sistemasi. 2024;13(2):572.

Aslam F. Advancing Credit Card Fraud Detection: A Review of Machine Learning Algorithms and the Power of Light Gradient Boosting. Ajcst. 2024;

Jiang Y. Credit Card Transaction Fraud Detection Based on Machine Learning. Appl Comput Eng. 2024;51(1):73–80.

Hassan H, Ahmad MA, Mustapha R. An Enhanced Feature Engineering Technique for Credit Card Fraud Detection. Fudma J Sci. 2024;8(4):8–16.

Esenogho E, Mienye ID, Swart TG, Aruleba K, Obaido G. A Neural Network Ensemble With Feature Engineering for Improved Credit Card Fraud Detection. Ieee Access. 2022;10:16400–7.

Leevy JL, Hancock J, Khoshgoftaar TM. Comparative Analysis of Binary and One-Class Classification Techniques for Credit Card Fraud Data. J Big Data. 2023;10(1).

Guo L, Song R, Wu J, Xu Z, Zhao F. Integrating a Machine Learning-Driven Fraud Detection System Based on a Risk Management Framework. Appl Comput Eng. 2024;87(1):80–6.

Balogun O, Kupolusi JA, Akomolafe A. Credit Card Fraud Detection Using Machine Learning Algorithms. Br J Comput Netw Inf Technol. 2024;7(3):1–35.

Dai S. Research on Detecting Credit Card Fraud Through Machine Learning Methods. 2023;1030–7.

Khatun MS, Alam BR, Taslim M, Hossain MA. Handling Class Imbalance in Credit Card Fraud Using Various Sampling Techniques. Am J Multidiscip Res Innov. 2022;1(4):160–8.

Akazue MI, Debekeme IA, Edje AE, Asuai CE, Osame UJ. UNMASKING FRAUDSTERS: Ensemble Features Selection to Enhance Random Forest Fraud Detection. J Comput Theor Appl. 2023;1(2):201–11.

Fernandes RJ. Queue Fundamentals, Implementation and Its Applications in Round Robin Scheduling. Int J Adv Sci Eng. 2022;9(1):2556–66.

Cheng Y, Yao X. Carbon intensity reduction assessment of renewable energy technology innovation in China: A panel data model with cross-section dependence and slope heterogeneity. Renew Sustain Energy Rev [Internet]. 2021;135. Available from: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85089659603&doi=10.1016%2Fj.rser.2020.110157&partnerID=40&md5=778fae3935ff2174137e18c37789ec20

Zhang Z, Hu Q, Yin J. Maritime-Accident-Induced Environmental Pollution and Economic Loss Analysis Using an Interpretable Data-Driven Method. Sustainability. 2025;17(7):3023.