Credit Card Fraud Detection Model Based on Correlation Feature Selection

Ahmad Salim[1,*], Salah N. Mjeat[1], Daniah Abul Qahar Shakir[2], Mohammed Awad Alfwair[3]

[1]Middle Technical University, Baghdad, Iraq
[2]University of Anbar, Anbar, Iraq
[3]The University of Jordan, Amman, Jordan
Emails: ahmadsalim@mtu.edu.iq; salahnoori@mtu.edu.iq; daniahashakir@uoanbar.edu.iq; m.alfoair1990@gmail.com

Abstract

Credit card fraud is a widespread cybercrime that threatens financial security. Effective cybersecurity measures are essential to mitigate these risks. Machine learning has shown promising results in detecting credit card fraud by analyzing transaction data and identifying patterns of suspicious behavior. Feature selection is crucial in machine learning because it simplifies the model, improves its performance, and prevents overfitting. This research introduces a machine learning model designed for credit card fraud detection. The model makes use of three types of correlations. Pearson, Spearman, and Kendall, to identify features and enhance the fraud detection process. Testing on datasets yielded impressive results achieving category accuracies of 99.95% and 99.58% surpassing alternative approaches. Also, the results showed that Kendall correlation is the best among the three types of correlation in selecting attributes in all approved datasets.

Keywords: Cybersecurity; Credit card fraud detection; Machine learning; Feature selection; Correlation

1. Introduction

Credit cards have grown to be a crucial device for modern-day transactions. With the arrival of digital charge systems, credit playing cards have grown to be a convenient manner to pay for goods and offerings [1-2]. However, credit cards also include inherent dangers, especially the ones related to fraud [3]. Unauthorized purchases can be made via criminals the usage of stolen credit information card statistics, putting the cardholder within the function of resolving the problem. Credit cards can be very handy, and when handled responsibly and securely, they offer many benefits. Fraud, on the other hand, is an attempt to manipulate a system for personal gain, usually by using false information or deception. Credit card fraud detection (CCFD) systems are designed to identify and prevent fraud in credit card transactions, thus protecting consumers and financial institutions in particular it is a means of detecting whether the transaction is valid or not [4-5].

Skimming, phishing, identity theft, chargeback fraud, and counterfeiting are the common forms of existing fraudulent credit card transactions. Each of these has the potential to cause financial losses and negatively impact credit scores [6]. The resulting financial losses and credit score setbacks highlight the urgent necessity for vigorous preventive measures. Methods and techniques for detecting credit card fraud encompass of statistical analysis, machine learning algorithms and human expertise [7-8]. The goal from applying CCFD is to identify anomalous patterns and behaviours in practices to prevent or avoid fraudulent activity and the increasing importance of machine learning (ML) in credit card fraud detection, protecting the interests of consumers and financial institutions is evident [9]. ML algorithms are trained to identify complex network patterns, enabling financial institutions to quickly and accurately identify potentially fraudulent activity. The scalability of these algorithms facilitates continuous learning from new data, and through increases the effectiveness of developing deceptive

DOI: https://doi.org/10.54216/JCIM.140224
Received: January 28, 2024 Revised: April 06, 2024 Accepted: July 12, 2024

behaviour detection. More importantly, rapid selection of suitable features is an important step in machine learning algorithms [10]. The importance of feature selection in machine learning, particularly in the realm of credit card fraud detection, cannot be overstated [11-12]. This process involves identifying pertinent and informative features or variables from a dataset, thereby boosting the accuracy and efficiency of predictive models [13-14]. Various techniques employed for feature selection encompass filter methods, wrapper methods, meta-heuristic algorithms, and embedded methods [15-16]. Among these techniques, correlation feature selection stands out as a method designed to pinpoint highly correlated input features within a dataset, subsequently eliminating redundancy to enhance model performance [17][18]. Proper feature selection can reduce overfitting, improve the model's interpretability and save computational resources, leading to better credit card fraud detection.

The main aim of this work is to build a machine-learning model that enhances the credit card fraud detection process by using correlation coefficients to select the most important features in the classification process. The rest of the manuscript is organized as follows: the second section reviews the most important similar works. The third section describes the feature selection techniques; the fourth section clarifies the proposed method; the fifth section illustrates the supported datasets in this work; the results are highlighted and discussed in the sixth part and the paper is concluded in the last part.

2. Related Work

In the realm of credit card fraud detection, researchers have introduced numerous methodologies predominantly rooted in machine learning and deep learning paradigms. A significant emphasis within these research endeavours revolves around the identification and selection of pivotal database features. This emphasis seeks to mitigate dimensions while concurrently enhancing classifier performance. There are multiple academic papers highlighted the importance and focusing on this point of feature selection to strengthen the effectiveness of fraud detection systems.

Rtayli et al. [19] introduced a modern approach to credit card fraud detection. They hired the support vector machine (SVM) algorithm alongside recursive feature elimination (RFE) to parent out and prioritize vital dataset factors. Also, they first-class-tuned the SVM model's settings using grid search to make it work better. Moreover, the paintings greater the behaviour hyper-parameter for the SVM model, by way of the usage of a grid search method to examine the maximum most beneficial mixture of parameters, accordingly make model performance better. The have a look at further encompasses sensitivity evaluation, evaluating the method's resilience throughout more than one level of fraudulent activity. The complete framework delineated within the paper amalgamates feature selection, hyper-parameter exceptional-tuning, and system getting to know strategies, supplying a promising avenue for credit score card fraud detection.

The article outlined in reference [20] introduced an advanced framework harnessing deep learning methodologies for credit card fraud detection. To improve the precision of fraud identity, researchers put forth a version amalgamating an attention mechanism along a long short-term memory (LSTM) network. To focus on the important features, this new method aims to analysing a sequences of credit card transactions and applies the attention technique. The LSTM community, on the other hand, is instrumental in delineating temporal dependencies inside the transaction collection, thereby uncovering capability fraudulent styles. Notably, a unique feature extraction method leveraging statistical measures was proposed to encapsulate transaction-unique characteristics. In addition, provided a promising technique in CCFD, capitalizing on the synergy among deep learning methodologies, statistical function extraction, and interest mechanisms.

In [21], researchers presented a method for CCFD using adaptive feature selection methods. The proposed method selects the most relevant features from the credit card transaction dataset and uses machine learning algorithms to detect fraud. The study suggests that adaptive feature selection methods can improve credit card fraud detection and enhance the security of electronic payment systems. The paper lacks a detailed explanation of the feature selection method used. In [22], researchers proposed an approach to address class imbalance and sparse feature selection problems in credit card fraud detection. The authors introduce a new metric called overlapping score, which measures the degree of overlap between feature distributions of positive and negative classes. The proposed framework aims to reduce the degree of overlap and select the most useful features for fraud detection. The results show that the proposed framework significantly outperforms several state-of-the-art methods in terms of precision, recall, F1 score, and precision.

Furthermore, some researchers have used optimization algorithms to select features in CCFD models. In [23], researchers proposed a technique that uses genetic algorithm (GA) as feature selection and feature classification through three different machine learning algorithms: Naive Bayes, Random Forest, and SVM for CCFD. The authors propose a framework that combines GA feature selection and the three algorithms to identify the most

DOI: https://doi.org/10.54216/JCIM.140224
Received: January 28, 2024 Revised: April 06, 2024 Accepted: July 12, 2024

relevant features for fraud detection. The results show that the proposed framework with GA feature selection outperforms the three algorithms without feature selection in terms of all evaluation metrics. Random forest with GA feature selection achieved the highest accuracy, precision, and F1 score among the three algorithms.

Padhi et al [24], proposed an innovative feature selection strategy using the Rock Hyrax Swarm Optimization (RHSO) algorithm specifically designed for credit card fraud detection (CCFD). Their approach combines the RHSO algorithm with a support vector machine classifier, strategically identifying the most important features critical for fraud detection. Inspired by the collective behaviour of rock hyrax swarms, the RHSO algorithm seeks a balance between exploring and exploiting the search space, with the goal of ultimately arriving at the optimal solution. Experimental results confirm the effectiveness of the proposed framework, demonstrating superior performance across multiple metrics including precision, recall, F1 score, and precision when compared to alternative methods.

However, most researchers used only one dataset to test their models, and this may not give a complete explanation of the efficiency of the model in detecting fraud. Furthermore, one of the most important challenges facing researchers in producing a machine learning model for CCFD more accurately is that the datasets used for this purpose are unbalanced, as the number of fraud transactions within the databases is much less than the number of normal operations. Besides, there are features that negatively affect the classification of fraudulent processes as normal processes. Therefore, an efficient technique must be used to select the most important features only in machine learning models to reduce dominance and improve the performance of the model.

3. Feature Selection

Feature selection plays a crucial role in CCFD models by identifying relevant and informative features and improving model accuracy and efficiency. By reducing the number of features, feature selection can improve model performance, reduce overfitting and enhance model interpretability [25-13]. In machine learning, feature selection can be performed using different approaches, such as filter methods, wrapper methods and embedded methods. The process of feature selection encompasses distinct methodologies: filter methods relying on statistical measures, wrapper methods evaluating feature subsets via a specific model, and embedded methods integrating feature selection within the model training phase [13].

Among these, statistical feature selection stands out as a prevalent technique in machine learning, spotlighting the identification of pivotal features grounded in their statistical attributes, such as correlation, mutual information, or significance [26]. This technique operates with the objective of dimensionality reduction while upholding the most informative features intact, thereby enhancing both the accuracy and efficiency of the model [26-27].

This paper delves into the utilization of three distinct types of correlation to discern critical features, examining their efficacy within Credit Card Fraud Detection (CCFD) models. The aim is to measure and contrast the performance of these correlation types in optimizing CCFD model outcomes.

4. Proposed Model

Machine learning models for CCFD use supervised learning algorithms to learn patterns of fraudulent behaviour from transactional data. In this manuscript, a machine learning model is proposed for credit card fraud detection. The model consists of three main stages, where after extracting the data, the pre-processing process takes place, then the most important features are selected using correlation coefficients, and finally, the classification process takes place based on machine learning classifiers. These stages are shown in Figure 1.

In the initial processing process, the values of the features are normalised, meaning that the values are confined between 0 and 1. Normalization is used in machine learning to scale input features to a common range, which can improve the performance and stability of models. It helps prevent features with large values from dominating those with smaller values and ensures that each feature contributes equally to the model's predictions. In the second stage, correlation is used to select the most important features.

Correlation serves as a statistical measure that measures the relationship between two variables. In feature selection, examining correlations between features helps identify redundant or irrelevant elements that can be pruned from the data set. This scale ranges from -1 which indicates a perfect negative correlation to 1 which represents a perfect positive correlation, while 0 indicates no correlation between the variables. Different types of correlation techniques can be used in feature selection, including:

DOI: https://doi.org/10.54216/JCIM.140224
Received: January 28, 2024 Revised: April 06, 2024 Accepted: July 12, 2024

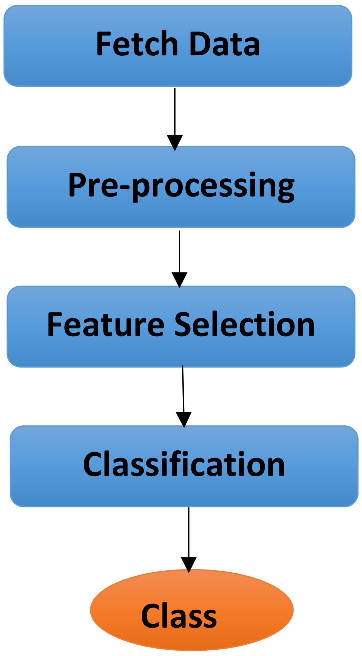


Figure 1. Proposed Model Stages

- Pearson correlation: This measures the linear relationship between two continuous variables. The Pearson correlation is given by Eq. (1).

$$r = \frac{n(\sum xy) - (\sum x)(\sum y)}{\sqrt{[n\sum x^2 - (\sum x)^2][n\sum y^2 - (\sum y)^2]}} \quad (1)$$

- Spearman correlation: This measures the monotonic relationship between two variables. It is used when the relationship between two variables is not necessarily linear, but still has a constant trend. Spearman's correlation formula shown in Eq. (2).

$$r = 1 - \frac{6\sum d_i^2}{n(n^2-1)} \quad (2)$$

- Kendall correlation: This measures the ordinal association between two variables. It is used when the variables are ranked or ordered. Eq. (3) illustrates the Kendall correlation formula.

$$r = \frac{3*T\sqrt{n(n-1)}}{\sqrt{2(2n-5)}} \quad (3)$$

Where r is a correlation, n is the number of pairs and x and y represent the features. These three distinct types of correlation are applied individually to sift through the dataset, selecting features to be incorporated into classifiers.

Machine learning classifiers form integral components within models, facilitating automated data classification by identifying patterns and utilizing features to categorize information. In the final phase of the study, an array of machine learning classifiers was employed to ensure precise and dependable predictions and classifications based on the selected features.

DOI: https://doi.org/10.54216/JCIM.140224
Received: January 28, 2024 Revised: April 06, 2024 Accepted: July 12, 2024

5. Datasets

Datasets are decisive in training machine learning models, providing a representative sample of real-world data for accurate predictions and decision-making. To accomplish this work, two global datasets were adopted. The first approved dataset (dataset 1) for model testing was obtained from the Kaggle website and contained 284,807 European cardholders' transactions occurring in 2013, 492 of which were frauds [28]. The database consists of 30 features, and the transactions are of two types: natural or fraudulent.

The second dataset (dataset 2) contains 594,643 transactions collected within six months from 2012 to 2013, of which 7,200 were fraudulent. This data was collected using the BankSim simulation tool [28]. Tables 1 and 2 show the selected features by type of correlation from dataset 1 and dataset 2, respectively.

Table 1: Selected Features from Dataset 1

Correlation Type	Selected Features
Pearson	V2, V4, V8, V11, V19, V20, V21, V22, V25, V26, V27, V28, and Amount
Spearman	V2, V4, V8, V19, V20, V21, V22, V25, V26, V27, and V28
Kendall	V2, V4, V8, V11, V19, V20, V21, V22, V25, V26, V27, and V28

Table 2: Selected Features from Dataset 2

Correlation Type	Selected Features
Pearson	Merchant, category, and amount
Spearman	Customer, gender, merchant, category, and amount
Kendall	Customer, merchant, category, and amount

6. Results and Discussion

The study harnesses the potency of machine learning in detecting credit card fraud, leveraging algorithms to discern fraudulent transactions based on data patterns and behaviours. For this study, we used a computer comprising a Core i7 processor and 16 GB of random-access memory facilitated the model's execution.

The primary objective of the proposed method centres on reducing dimensionality by selecting the most impactful features, thereby enhancing the model's classification accuracy. This endeavour involves utilizing three distinct types of correlation to gauge their effectiveness. The study employed four distinct machine learning classifiers: the Random Forest Classifier (RFC), Extra Trees Classifier (ETC), Gradient Boosting Classifier (GBC), and Support Vector Machine (SVM). The assessment of the proposed method's performance relies on accuracy as the primary criterion, evaluating its efficacy in enhancing fraud detection accuracy.

Table 3 presents the results of the proposed method based on the three types of correlation select features from dataset 1 and using four classifiers. It is clear from the last row, which contains the average use of each type of correlation and all approved classifiers, that the Kendall correlation achieved the best results compared to the other two types, while the Pearson correlation achieved the worst results. Moreover, the last column shows that the RFC classifier was the most accurate in classifying the selected features. Additionally, the method achieved the best result (99.95) when using the Kendall correlation to select features and adopting RFC as a classifier.

Table 3: The Proposed Model Results (dataset 1)

Classifier	**Pearson**	**Spearman**	**Kendall**	**Avg**
RFC	99.92	99.93	**99.95**	99.93
ETC	99.93	99.92	99.93	99.92
GBC	99.86	99.89	99.89	99.88
SVM	99.89	99.92	99.93	99.91
Avg	99.9	99.91	99.92	

DOI: https://doi.org/10.54216/JCIM.140224
Received: January 28, 2024 Revised: April 06, 2024 Accepted: July 12, 2024

Table 4 shows the accuracy of the classification of the transactions in dataset 2. It is also clear from the table that selecting features using Kendall gave the best results compared to the other two types of correlation, while GBC achieved the best results as a classifier compared to other approved classifiers.

Table 4: The Proposed Model Results (dataset 2)

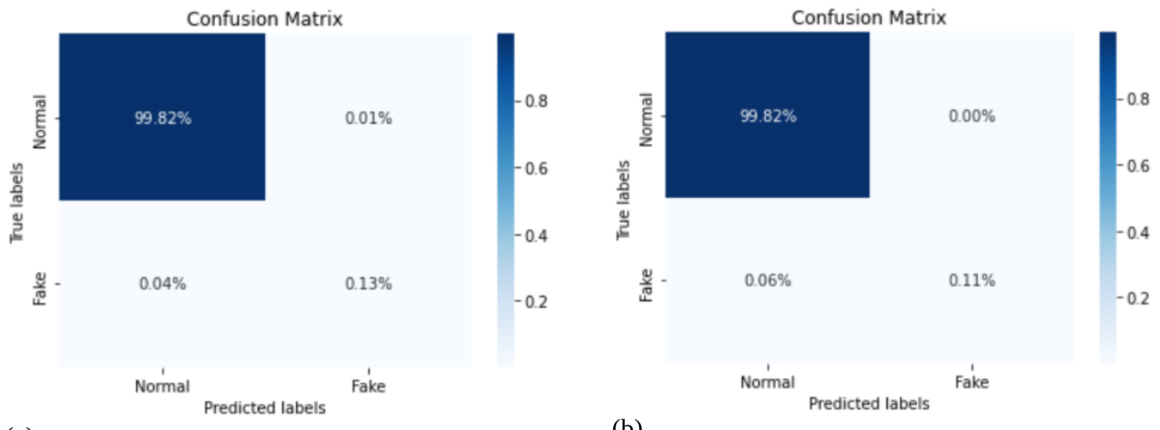
Classifier	**Pearson**	**Spearman**	**Kendall**	**Avg**
RFC	99.44	99.55	99.56	99.51
ETC	99.42	99.53	99.55	99.5
GBC	99.57	99.55	**99.58**	99.56
SVM	99.43	99.41	99.44	99.42
Avg	99.46	99.51	99.53	

Additionally, in order to evaluate the performance of the suggested method, it is compared with other methods designed for the same purpose and using the same dataset. Table 5 illustrates a comparison of the proposed method in this work with similar methods, as it is clear that our method has achieved advantageous performance compared to the methods proposed for the same purpose.

Table 5: Comparison with Similar Methods

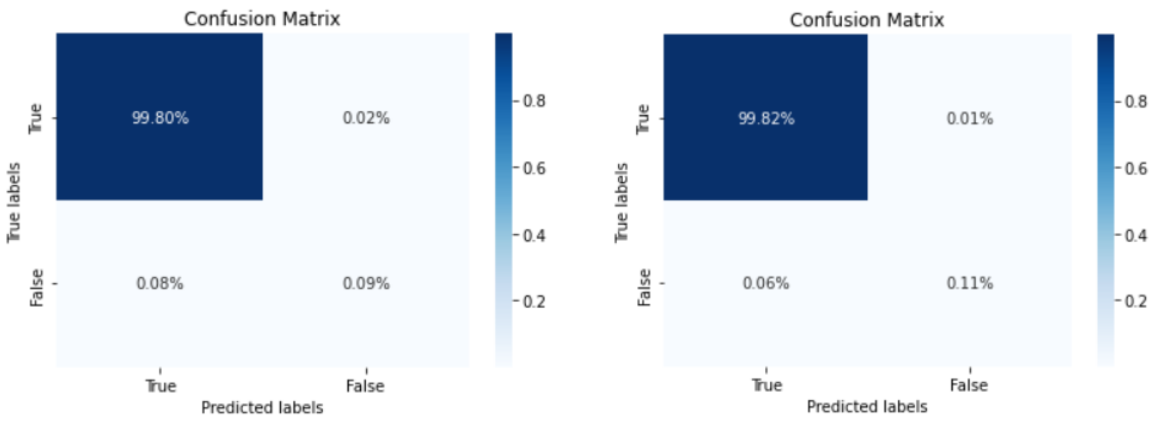
Dataset	**Method**	**Accuracy**
	Our method	**99.95%**
Dataset 1	Method [19]	99%
	Method [20]	96.72%
	Method [24]	99.8%
	Our Method	**99.58%**
Dataset 2	Method [20]	97.48%
	Method [29]	99.56%

Figures 2 and 3 show the confusion matrix of features from dataset 1 and dataset 2 when using the Kendall correlation and through all four classifiers, where the proposed method achieved the best results in most of them. Despite the high accuracy achieved by the proposed method, the correlation matrix shows that the classification fails to focus more on classifying fraud operations as normal operations, and this is due to the imbalance of data in the approved dataset and others that are used for the same purpose, as in most databases used to detect fraud, where the percentage of fraud operations does not exceed 1% of the total number of transactions. Figure 4 shows the receiver operating characteristic (ROC) curve of the best classification case using the proposed method.



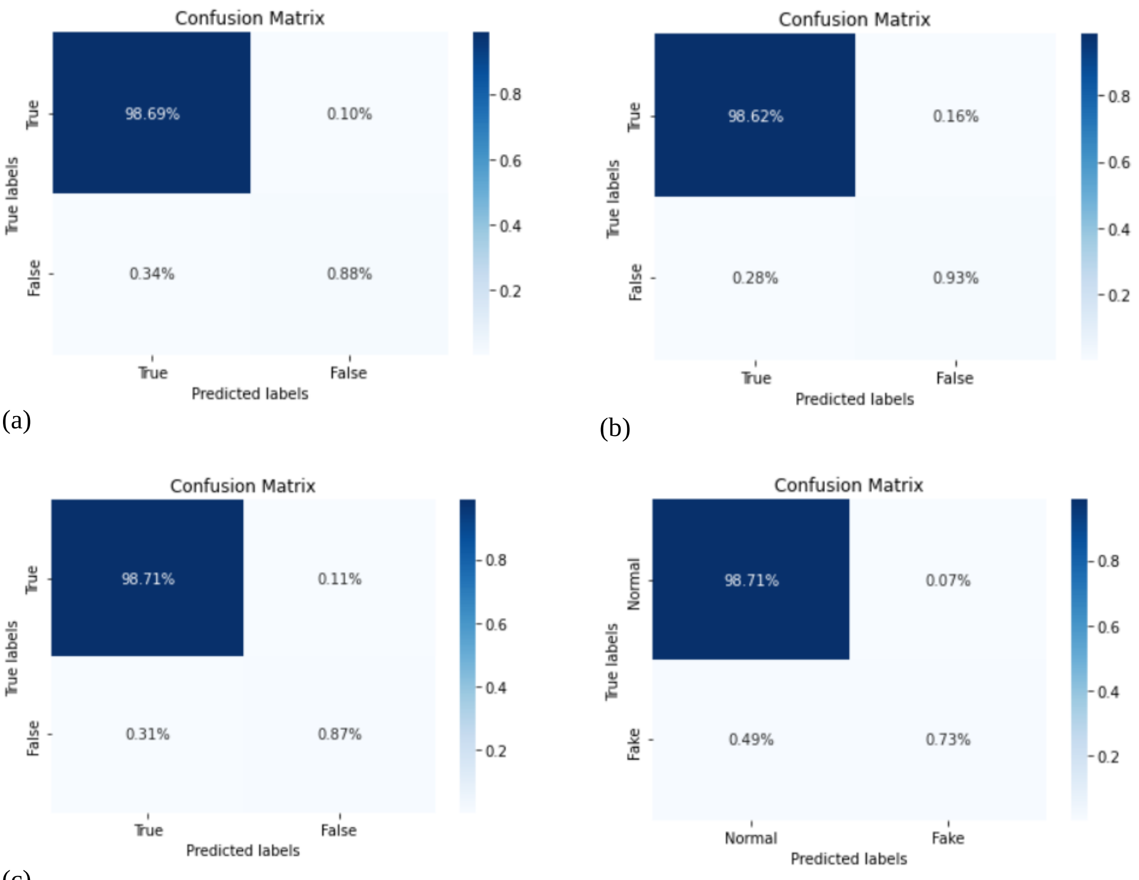
(a) (b)

DOI: https://doi.org/10.54216/JCIM.140224
Received: January 28, 2024 Revised: April 06, 2024 Accepted: July 12, 2024



(c) (d)

Figure 2. Confusion Matrix of Kendall Correlation Features (dataset 1): (a) Kendall with RFC (b) Kendall with ETC (c) Kendall with GBC (d) Kendall with SVM



(a) (b) (c) (d)

Figure 3. Confusion Matrix of Kendall Correlation Features (dataset 2): (a) Kendall with RFC (b) Kendall with ETC (c) Kendall with GBC (d) Kendall with SVM

DOI: https://doi.org/10.54216/JCIM.140224
Received: January 28, 2024 Revised: April 06, 2024 Accepted: July 12, 2024

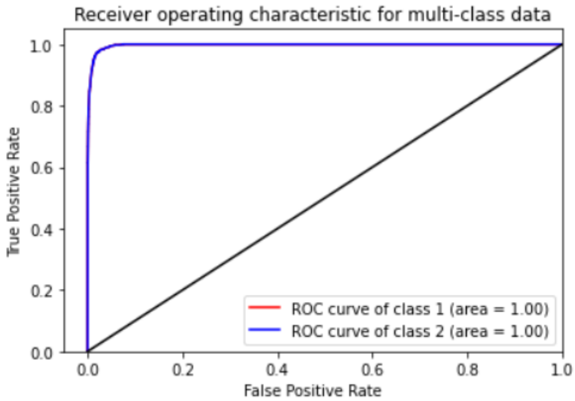


Figure 4. ROC Curve of Kendall with ETC (dataset 1)

7. Conclusion

This paper proposes a machine-learning model for detecting credit card fraud. The model's primary objective is to identify the most critical features, using Peterson, Spearman, and Kendall correlation methods to determine correlation. The proposed model was tested using transaction from two global datasets and based on four classifiers: RFC, ETC, GBC and SVM. The results showed that the Kendall correlation is the best among the three types of correlation in selecting features from dataset 1 and dataset 2 for the CCFD model; RFC achieved the best results as a classifier in dataset 1, whereas GBC gave the highest accuracy in dataset 2. The model obtained an accuracy of 99.95%. Furthermore, the model achieved competitive results compared to other methods proposed for the same purpose.

Funding: “This research received no external funding”

Conflicts of Interest: “The authors declare no conflict of interest.”

References

[1] Darch Abed Dawar, A. (2024). Enhancing Wireless Security and Privacy: A 2-Way Identity Authentication Method for 5G Networks. International Journal of Mathematics, Statistics, and Computer Science, 2, 183–198. https://doi.org/10.59543/ijmscs.v2i.9073

[2] G, N. , Y. Jessinda, A. RSupraja., S. "Personnel Monitoring System Using Mobile Application during the COVID 19," Journal of Journal of Cognitive Human-Computer Interaction, vol. 2, no. 2, pp. 40-49, 2022. DOI: https://doi.org/10.54216/JCHCI.020201

[3] G. Baader and H. Krcmar, “Reducing false positives in fraud detection: Combining the red flag approach with process mining,” Int. J. Account. Inf. Syst., vol. 31, pp. 1–16, 2018.

[4] S. B. E. Raj and A. A. Portia, “Analysis on credit card fraud detection methods,” in 2011 International Conference on Computer, Communication and Electrical Technology (ICCCET), 2011, pp. 152–156.

[5] S. Makki, Z. Assaghir, Y. Taher, R. Haque, M.-S. Hacid, and H. Zeineddine, “An experimental study with imbalanced classification approaches for credit card fraud detection,” IEEE Access, vol. 7, pp. 93010–93022, 2019.

[6] J. West and M. Bhattacharya, “Intelligent financial fraud detection: a comprehensive review,” Comput. \& Secur., vol. 57, pp. 47–66, 2016.

[7] Aziz, A. Mirzaliev, S. Maqsudjon, Y. "Enhancing Malware Detection in Cybersecurity through Optimized Machine Learning Technique," Journal of International Journal of Advances in Applied Computational Intelligence, vol. 4, no. 2, pp. 26-32, 2023. **DOI:** https://doi.org/10.54216/IJAACI.040203

[8] R. Van Belle, B. Baesens, and J. De Weerdt, “CATCHM: A novel network-based credit card fraud detection method using node representation learning,” Decis. Support Syst., vol. 164, p. 113866, 2023.

[9] Akhmetshin, E., et al. "Intelligent Data Analytics using Hybrid Gradient Optimization Algorithm with Machine Learning Model for Customer Churn Prediction," in Fusion: Practice and Applications, vol. 14, no. 2, pp. 159–59, 2024.

DOI: https://doi.org/10.54216/JCIM.140224
Received: January 28, 2024 Revised: April 06, 2024 Accepted: July 12, 2024

[10] J. Chaquet-Ulldemolins, F.-J. Gimeno-Blanes, S. Moral-Rubio, S. Muñoz-Romero, and J.-L. Rojo-Álvarez, "On the Black-Box Challenge for Fraud Detection Using Machine Learning (I): Linear Models and Informative Feature Selection," Appl. Sci., vol. 12, no. 7, p. 3328, 2022.

[11] D. Varmedja, M. Karanovic, S. Sladojevic, M. Arsenovic, and A. Anderla, "Credit card fraud detection-machine learning methods," in 2019 18th International Symposium INFOTEH-JAHORINA (INFOTEH), 2019, pp. 1–5.

[12] Noori, N. Muhammed, and Omar Saber Qasim, "Deep Features Selections with Binary Marine Predators Algorithm for Effective Classification of Image Datasets," Fusion: Practice and Applications, 2023, 86-6.

[13] Nagamalla, V. karkee, J. Kumar, R. "Integrating Predictive Big Data Analytics with Behavioral Machine Learning Models for Proactive Threat Intelligence in Industrial IoT Cybersecurity," Journal of International Journal of Wireless and Ad Hoc Communication, vol. 7, no. 2, pp. 08-24, 2023. DOI: https://doi.org/10.54216/IJWAC.070201

[14] S. Khalid, T. Khalil, and S. Nasreen, "A survey of feature selection and feature extraction techniques in machine learning," in 2014 science and information conference, 2014, pp. 372–378.

[15] J. Cai, J. Luo, S. Wang, and S. Yang, "Feature selection in machine learning: A new perspective," Neurocomputing, vol. 300, pp. 70–79, 2018.

[16] A. O. Adewumi and A. A. Akinyelu, "A survey of machine-learning and nature-inspired based credit card fraud detection techniques," Int. J. Syst. Assur. Eng. Manag, vol. 8, pp. 937–953, 2017.

[17] H. F. Eid, A. E. Hassanien, T. Kim, and S. Banerjee, "Linear correlation-based feature selection for network intrusion detection model," in Advances in Security of Information and Communication Networks: First International Conference, SecNet 2013, Cairo, Egypt, September 3-5, 2013. Proceedings, 2013, pp. 240–248.

[18] I. Jain, V. K. Jain, and R. Jain, "Correlation feature selection based improved-binary particle swarm optimization for gene selection and cancer classification," Appl. Soft Comput., vol. 62, pp. 203–215, 2018.

[19] N. Rtayli and N. Enneya, "Enhanced credit card fraud detection based on SVM-recursive feature elimination and hyper-parameters optimization," J. Inf. Secur. Appl., vol. 55, p. 102596, 2020.

[20] I. Benchaji, S. Douzi, B. El Ouahidi, and J. Jaafari, "Enhanced credit card fraud detection based on attention mechanism and LSTM deep model," J. Big Data, vol. 8, pp. 1–21, 2021.

[21] A. Singh and A. Jain, "Adaptive credit card fraud detection techniques based on feature selection method," in Advances in Computer Communication and Computational Sciences: Proceedings of IC4S 2018, 2019, pp. 167–178.

[22] B. Omar, F. Rustam, A. Mehmood, G. S. Choi, and others, "Minimizing the overlapping degree to improve class-imbalanced learning under sparse feature selection: application to fraud detection," IEEE Access, vol. 9, pp. 28101–28110, 2021.

[23] Y. K. Saheed, M. A. Hambali, M. O. Arowolo, and Y. A. Olasupo, "Application of GA feature selection on Naive Bayes, random forest and SVM for credit card fraud detection," in 2020 international conference on decision aid sciences and application (DASA), 2020, pp. 1091–1097.

[24] B. K. Padhi, S. Chakravarty, B. Naik, R. M. Pattanayak, and H. Das, "RHSOFS: Feature Selection Using the Rock Hyrax Swarm Optimization Algorithm for Credit Card Fraud Detection System," Sensors, vol. 22, no. 23, p. 9321, 2022.

[25] S. K. Kamaruddin and V. Ravi, "Credit card fraud detection using big data analytics: use of PSOAANN based one-class classification," in Proceedings of the international conference on informatics and analytics, 2016, pp. 1–8.

[26] B. Chandra and M. Gupta, "An efficient statistical feature selection approach for classification of gene expression data," J. Biomed. Inform. vol. 44, no. 4, pp. 529–535, 2011.

[27] A. G. Karegowda, A. S. Manjunath, and M. A. Jayaram, "Comparative study of attribute selection using gain ratio and correlation based feature selection," Int. J. Inf. Technol. Knowl. Manag. vol. 2, no. 2, pp. 271–277, 2010.

[28] G. Vaughan, "Efficient big data model selection with applications to fraud detection," Int. J. Forecast., vol. 36, no. 3, pp. 1116–1127, 2020.

[29] B. Stojanović et al., "Follow the trail: Machine learning for fraud detection in Fintech applications," Sensors, vol. 21, no. 5, p. 1594, 2021.

DOI: https://doi.org/10.54216/JCIM.140224
Received: January 28, 2024 Revised: April 06, 2024 Accepted: July 12, 2024