

The best machine learning model for fraud detection on e-platforms: a systematic literature review

Alimatu – Saadia Yussiff¹, Lemdi Frank Prikutse¹, Georgina Asuah², Abdul – Lateef Yussiff¹, Emmanuel Dortey Tetteh¹, Norshahila Ibrahim³, Wan Fatimah Wan Ahmad⁴

¹Department of Computer Science and Information Technology, Faculty of Physical Sciences, University of Cape Coast, Cape Coast, Ghana

²Department of Data Science and Engineering, Faculty of Informatics, Eötvös Loránd University, Budapest, Hungary

³Department of Computer Science and Digital Technology, Faculty of Computing and Meta-Technology, Universiti Pendidikan Sultan Idris, Perak, Malaysia

⁴Department of Computer and Information Sciences, Universiti Teknologi PETRONAS, Seri Iskandar, Malaysia

Article Info

Article history:

Received Feb 2, 2024

Revised May 28, 2024

Accepted Jun 4, 2024

Keywords:

E-platforms

Financial transactions

Fraud detection

Machine learning algorithms

Online payment systems

ABSTRACT

The internet has been instrumental in the development and facilitation of online payment systems. However, its associated fraudulent activities on e-platforms cannot be overlooked. As a result, there has been a growing interest in the application of machine learning (ML) algorithms for fraud detection on financial e-platforms. The goal of this research is to identify common types of fraud on financial e-platform, highlight different machine learning algorithms employed in fraud detection, and derive the best machine learning algorithms for fraud detection on e-platforms. To achieve this goal, the research followed a nine steps systematic review approach to retrieve Journals and conference publications from science direct, Google Scholar and IEEE Xplore between 2018 and 2023. Out of 2,071 articles identified and screened, 44 publications (23 articles and 21 conference proceedings) satisfied the inclusion criteria for further analysis. The random forest algorithm turned out to be the best ML algorithm because it ranked first in the frequency of usage analysis and ranked first in the performance analysis with an average accuracy of 96.67%. Overall, this review has identified the kinds of fraud on financial e-platforms, and proclaimed the best and least ML algorithm for fraud detection on financial e-platform. This can help guide future research and inform the development of more effective fraud detection systems.

This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.



Corresponding Author:

Alimatu – Saadia Yussiff

School of Physical Sciences, Department of Computer Science and Information Technology

Faculty of Physical Sciences, University of Cape Coast

Cape Coast, Ghana

Email: asyussiff@ucc.edu.gh

1. INTRODUCTION

The advancement in technology and the evolution of the internet have paved the way for the establishment of modern services, including e-commerce and financial transactions. Traditional buyer-seller relationships and the shopping experience for many consumers have been significantly altered by e-commerce [1]. E-commerce and the widespread use of online banking have contributed to a recent uptick in the volume of monetary transactions. The Internet has been instrumental in the development and facilitation of online payment systems. However, this increased convenience presents several obstacles, the most

significant of which is the problem of fraudulent activities on these e-platforms [2]. Presently, online financial fraud has become a more intricate and difficult challenge to solve [3].

The most typical forms of fraud that can be committed on online systems that may cost consumers and e-commerce sites a lot of money include Payment fraud (criminals commit payment fraud by either stealing financial details or forging credentials to use for online purchases). Account takeover fraud (Fraudsters impersonate a real user in order to make illegal purchases or obtain sensitive information). Phishing scams (Scammers employ phishing emails, texts, and social media posts to fool consumers into divulging sensitive information by making it look to come from a trusted organization, such as a bank or online retailer). Identity theft (Theft of sensitive personal information, such as social security numbers or credit card data, can lead to the opening of fake accounts and the making of fraudulent purchases). Chargeback fraud (To commit chargeback fraud, a user must erroneously assert that they did not get the purchased products or services).

We must learn to recognize the red flags of online fraud and take preventative measures. The traditional methods of fraud detection on e-platforms, such as rule-based systems and manual review processes, are often unable to keep up with the rapidly evolving tactics of fraudsters. These methods are also resource-intensive and time-consuming, making them inefficient and often ineffective in detecting and preventing fraud [4]. Consequently, there has been a growing interest in the application of machine learning (ML) models for fraud detection on e-platforms. The goal of this research is to identify common types of fraud on financial e-platform, highlight which ML model are used for fraud detection on financial e-platforms and also ascertain the best ML for fraud detection on e-platforms in terms of performance metrics.

A machine learning algorithm is a type of algorithm that can learn from data and perform a task without being explicitly programmed to do so. These algorithms are "soft coded" in the sense that they improve the task at hand through iteratively changing or adapting their underlying structure [5]. By measuring prediction mistakes during the training phase, machine learning has an intelligent system that allows it to continuously learn [6]. Machine learning has proven to be a promising approach for fraud detection in e-commerce due to its ability to learn from large amounts of data and identify patterns that traditional methods may miss. By leveraging the power of ML, e-commerce platforms can build more effective fraud detection systems that are faster, more accurate, and less resource-intensive.

According to [5], depending on how the data is labeled, machine learning can be categorized as supervised, semi-supervised, or unsupervised. Supervised learning is the process of utilizing labeled datasets to train algorithms that effectively classify data or predict outcomes (e.g., classification and regression). Semi-supervised learning is a hybrid of supervised and unsupervised learning in which a portion of the data is partially labeled and the labeled portion is used to infer the unlabeled portion. In unsupervised learning, the learning system receives only input samples (e.g., clustering and estimation of probability density function).

However, the use of ML models for fraud detection on e-platforms is not without its own set of challenges. The performance of these models depends on the quality and size of the training data, the choice of appropriate features and algorithms, and the complexity of the fraud detection problem. To gain a better understanding of machine learning models for fraud detection on e-platforms, a systematic literature review is necessary. Such a review can provide an overview of the existing research on the topic, identify the different types of machine learning models that have been used for fraud detection, and highlight the best models. In this regard, the current study was conducted to perform a systematic literature review to analyze research studies to identify the common types of fraud on financial e-platforms and to determine the best machine learning algorithms in terms of performance metrics that have been used for fraud detection on e-platforms.

2. METHOD

This section describes in detail how we carried out the methodological literature review. Based on the advantages of systematic literature review (SLR) such as comprehensiveness, reliability and unbiased coverage of relevant studies [7], this research adopted the SLR steps in [8] to conduct the SLR by following the key steps in Figure 1. The application of Figure 1 to our research, are further described in subsections 2.1-2.9.

2.1. Defining the research questions

According to [8], the first stage of SLR is to establish appropriate research questions. These questions should be focused, clear and guide what we wanted to find out from the research. In this regard, we derived the following three research questions (RQs) from our pre-planned topic:

RQ1: What are the common types of fraud on e-platforms?

RQ2: Which machine learning algorithms are used for Fraud detection on e-platforms?

RQ3: What are the best machine learning algorithms used on e-platforms for fraud detection?

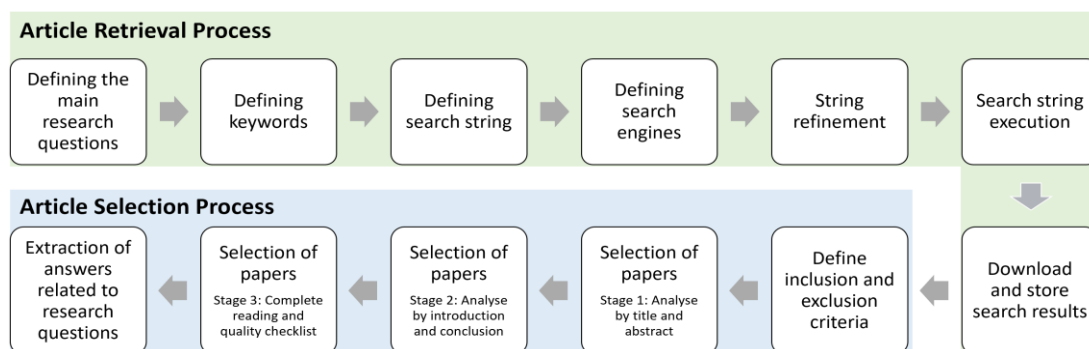


Figure 1. Systematic literature review process [8]

2.2. Defining concepts and keywords

In the second step, we then came out with the following three concepts in relation to the topic and research questions;

Concept 1: “Machine learning models”

Concept 2: “Fraud detection”

Concept 3: “e-platforms”

By considering the synonyms, other spelling variations and abbreviations of the concepts, we then derived the following keywords: “Machine Learning models”, “Machine Learning techniques”, “Machine learning algorithms”, “fraud detection”, “e-platforms”, and electronic platforms.

2.3. Defining search string

The above keywords were then merged using control vocabulary or Boolean operators. Thus, using the “population, intervention, comparison, outcome” (PICO) and the “MESH” approach, we came out with the following search string as shown in Figure 2.

“Machine Learning models” OR “Machine learning algorithms” OR “Machine Learning techniques” AND
 “fraud detection” AND “e-platforms” OR “electronic-platforms.”

Figure 2. Search string

2.4. Defining search engines

In this sub-section 4, search engines were defined. Our methodology used a suitable selection of databases to obtain comprehensive coverage of the literature and to raise the likelihood of discovering highly suitable articles. Consequently, the following electronic literature databases were included in our search; ScienceDirect, Google Scholar, and IEEEXplore.

2.5. String refinement

The string defined in sub-section 2.3 was tested in one of the search engines (Google Scholar). Once the string was applied, we verified if the returned papers were relevant. Known papers that are potential candidates for primary studies in this review (which exist on these engines) appeared in the search. In cases where there were no relevant results, the search string was parsed again to be calibrated. At the end, the refinement of the search criteria was done in each database.

2.6. Search string execution and search results

Once the search string was defined, it was adapted to each of the selected three search engines. As the search went on, the search results were documented, the number of articles each search returned, and the date of execution were also documented. Figure 3 demonstrated the resulting preferred reporting items for systematic reviews and meta-analyses (PRISMA) flow diagram derived from the research.

Thus, at the end of searching all the three databases, we found a total of 2,071 articles. These articles were imported into Rayaana to begin the systematic review process. First of all, 35 out of the 2,071 articles were de-duplicated in the “Rayaan” web tool. The remaining 2,036 articles were further screened on a title

and abstract basis in Rayyan. There was an initial conflict on 20 articles which was subsequently resolved. A total of 1,894 records were further excluded. The remaining 142 included articles were exported from “Rayyan” into Microsoft Excel for further screening based on their introductions, methodologies, findings, and conclusions. The eligibility criteria used is defined in the following section.

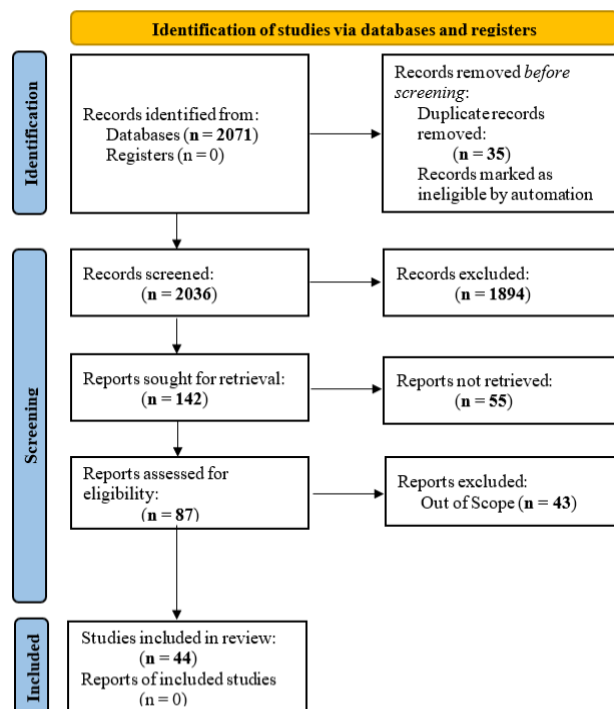


Figure 3. PRISMA flow diagram

2.7. Defining inclusion and exclusion criteria

In sub-section 7, we concentrated on examining several factors that affect how financial e-platforms operate. The chosen approach enabled us to list and categorize different machine learning models currently in use for detecting fraud on financial e-platforms. The review also included the evaluation of the effectiveness of different machine learning algorithms in detecting fraud on financial e-platforms. The approach also identified relevant data using the inclusion and exclusion criteria and in summarizing the findings.

2.8. Selection of papers

“Rayyan” web tool which is designed to help researchers working on systematic reviews, scoping reviews, and other knowledge synthesis projects, was used in this review for screening and coding of studies. In order to select the relevant papers for the actual analysis we went through three stages of analyses. First, we conducted title and abstract screening of the relevant articles by marking them in Rayyan as either “included”, “excluded” or “doubtful”. Secondly, in order to refine the selection in the previous stage, the introductions and conclusions of the selected papers were analysed. Finally, the third stage of the analyses involved complete reading of selected articles, quality check and comparative analysis.

Thus, the 142 articles exported to Microsoft Excel were subjected to the following eligibility criteria: i) Reports not retrieved (unavailability of full text) and ii) Reports or articles that are out of scope.

The screening results indicated that out of the total of 142 papers, 55 full-text papers were unavailable and 43 papers were out of scope. The remaining papers included for further analyses were 44.

2.9. Extraction of answers to research questions

At this stage, the research questions were answered by analyzing the selected papers in the previous step with the aid of a spreadsheet. As we read each of the selected papers, the possible answers extracted was posted directly in the spreadsheet. Figure 4 shows a snapshot of the spreadsheet used for the final stage of screening while the results of the research are presented in the Findings and discussions section.

| No. | Title | authors | year | Objectives | Abstract | Methodology | Types of Fraud | Machine Learning Algorithms Used | Findings/Outcomes | Remarks |
|-----|---|--|------|--|---|--|---|--|--|---------|
| 1 | Credit Card Fraud Detection Using Conditional Tabular Generative Adversarial Networks (CT-GAN) and Supervised Machine Learning Techniques | Pali, Yushar | 2021 | To evaluate the effectiveness of using CT-GAN and machine learning techniques in improving the accuracy of credit card fraud detection, in comparison to other state-of-the-art techniques for handling class imbalance challenges | Credit card fraud has been a major concern for financial institutions and business stakeholders for a long time and due to its ever-increasing nature, it has been a global topic of interest for researchers. Machine learning has proved to be one of the promising approaches for the detection and prediction of frauds. Despite having various advantages, there is no ideal model to handle this task due to the various factors involved. On the other hand, the class imbalance is one of the major and frequently occurring challenges while dealing with fraud detection tasks which hamper the model performance. There are several previously explored studies combining machine learning algorithms with various data pre-processing techniques to handle class imbalance challenges. To take this research further we have used a novel approach of combining supervised machine learning algorithms like Logistic Regression, Random Forest, XGBoost with Conditional Tabular Generative Adversarial Networks (CT-GAN) for balancing skewed data by data augmentation. We have used the SelectBest feature selection method for selecting the most significant feature for our analysis. After testing the proposed technique on our machine learning algorithms which are trained on both unbalanced and balanced data, we have observed a significant increase in model performance in terms of F1-score, recall, AUC score and Gmean. The results show that the Random Forest model outperforms other models in all terms with 100% recall value followed by XGBoost having recall of 2%, after applying our proposed technique whereas Logistic Regression has shown the | We utilized a hybrid combination of a data modeling method CT-GAN and feature selection technique-SelectBest in this study to handle the class imbalance challenge while using machine learning models and developed 3 models using this approach. | credit card fraud detection | Logistic regression, Random Forest, XGBoost with Conditional Tabular Generative Adversarial Networks (CT-GAN) for balancing skewed data by data augmentation | The results show that the Random Forest model outperforms other models in all terms with a 100% recall value followed by XGBoost having a recall of 2% after applying our proposed technique whereas Logistic Regression has shown the most significant increase in performance from 78% recall to | |
| 2 | CREDIT CARD FRAUD DETECTION USING DEEP LEARNING | Dian, Kaoufar and Lahcen, Laouani and Mohamed, Azzou | 2022 | aims to use deep learning to detect fraud in online transactions | As the number of online businesses is increasing continuously, millions of individuals have switched their preferences toward the online experience as it requires less effort. With this rise comes fraud. Scammers always find techniques to use other people's data for unauthorized online purchases. Therefore, this capstone project aims to use deep learning to detect fraud in online transactions. While there exist several machine learning classification models to perform the fraud detection task, deep learning has demonstrated promising results in terms of accuracy, precision, and recall. For this project, the Long Short-Term Memory (LSTM) model will be employed to achieve maximum performance. However, before feeding the data into the model to train, it is important to balance it and normalize it. The dataset used in this project is retrieved from Kaggle.com, and it contains 284,807 online transactions made by European credit cardholders. The LSTM model reached a prediction accuracy of 99.94%, a precision of 99.89%, and a recall of 99.99%. | Data Analysis and modeling Long Short-Term Memory (LSTM) model will be employed to achieve maximum performance | Detection of fraud in online credit card transactions | deep learning | The LSTM model reached a prediction accuracy of 99.94%, a precision of 99.89%, and a recall of 99.99% | |

Figure 4. Snapshot of screening spreadsheet

3. RESULTS AND DISCUSSION

This section gives a detailed discussion of the results of the analysis. The initial database search returned a total of 2,071 results. After applying the selection and eligibility criteria, 44 publications were left for detailed analysis and the results are presented in the following sub-sections.

3.1. Common types of fraud on e-platforms

The first research question was to investigate the common types of fraud on e-platforms, after a thorough analysis of the 44 articles included in the studies, the result revealed that the common types of electronic frauds on financial e-platforms are: Credit card fraud, banking transaction fraud and E-commerce fraud with frequency of 36, 5 and 3 respectively. Therefore, credit card fraud is the most prevalent form of fraud on e-platforms. Figure 5 and Table 1 summarizes the results.

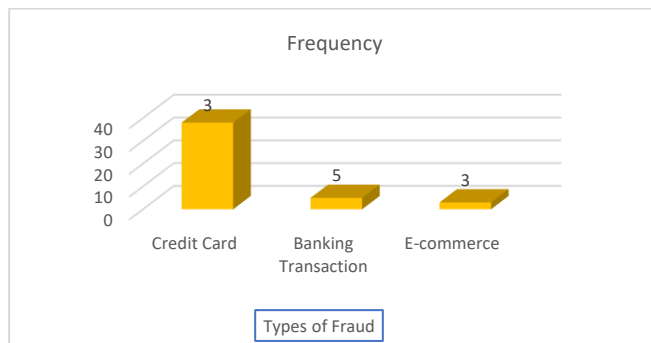


Figure 5. Types of electronic fraud

Table 1. Types of electronic fraud

| No. | Type of Fraud | Frequency | Percentage (%) |
|-----|---------------------|-----------|----------------|
| 1 | Credit Card | 36 | 81.82 |
| 2 | Banking Transaction | 5 | 11.36 |
| | E-commerce | 3 | 6.82 |

3.2. Machine learning algorithms used for fraud detection on e-platforms

The second research question was to find out which machine learning algorithms are used for fraud detection on e-platforms. Table 2 is the summary of results obtained from the 44 articles we worked with. It

is evident from Table 2 and Figure 6 that 23 different machine learning models were used in the 44 different publications we worked with.

Table 2. Different machine learning algorithms for fraud detection

| No. | Machine Learning (ML) Algorithm | Reference | Frequency of Usage |
|-----|--|--|--------------------|
| 1 | Long Short-Term Memory (LSTM) | [9]-[11] | 3 |
| 2 | Logistic Regression (LR) | [10], [12], [13], [14]-[29] | 19 |
| 3 | Random Forest (RF) | [10], [12], [15], [17]-[24], [26]-[40] | 27 |
| 4 | XGBoost | [14], [22], [21], [35], [40], [41] | 6 |
| 5 | AdaBoost | [9], [30], [21], [44], [45], [40], [42], [43] | 9 |
| 6 | Support Vector Machine (SVM) | [13], [15]-[17], [19], [20], [26], [27], [29], [31], [32], [46]-[49] | 15 |
| 7 | Naïve Bayes (NB) | [10], [15], [25]-[27], [31], [49], [50] | 8 |
| 8 | K-Nearest Neighbour (KNN) | [10], [11], [13], [17], [19], [21], [25-27], [31], [48], [50] | 12 |
| 9 | Multi-Layer Perceptron (MLP) | [10], [21], [32], [50] | 4 |
| 10 | Complement Naïve Bayes | [10] | 1 |
| 11 | Gaussian Naïve Bayes | [10], [19], [21], [50] | 4 |
| 12 | Bernouli Naïve Bayes | [10] | 1 |
| 13 | Light Gradient Boosting Machine (LGBM) | [10], [14], [16] | 3 |
| 14 | Decision Tree | [10], [13], [15], [17], [18], [21]-[24], [28], [35], [39] | 12 |
| 15 | Gradient Boost | [15], [21], [23], [30] | 4 |
| 16 | Convolutional Neural Network (CNN) | [9], [11], [19], [45] | 4 |
| 17 | Isolation Forest | [32], [51], | 2 |
| 18 | Artificial Neural Network (ANN) | [15], [20], [23], [25], [48], [49], [52] | 7 |
| 19 | k-means | [19], [45], [52], | 3 |
| 20 | Linear Regression | [35] | 1 |
| 21 | J48 Algorithm | [38] | 1 |
| 22 | Perceptron | [10], [25], [26] | 3 |
| 23 | CatBoost | [40] | 1 |

Figure 6 also illustrates that the least used models are complement naïve Bayes, Bernoulli naïve Bayes, isolation forest, linear regression, J48, and CatBoost algorithms. Other algorithms like K-nearest neighbor (KNN), decision tree, support vector machine (SVM), logistic regression, and random forest fall within the top five with random forest being the most used machine learning algorithm for fraud detection.

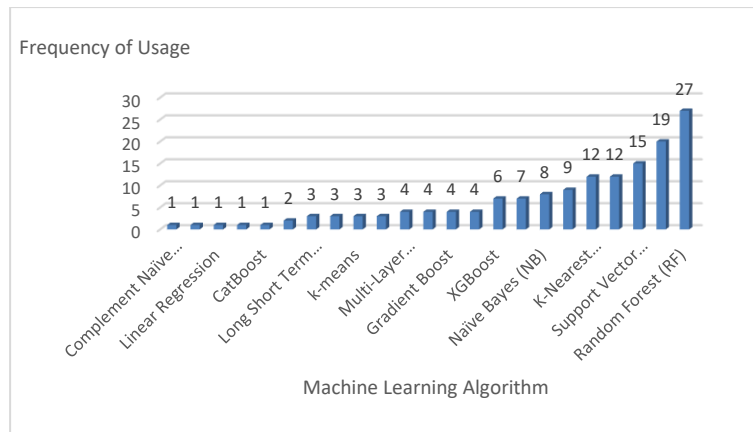


Figure 6. Different machine learning algorithms for fraud detection

3.3. The best machine learning algorithms used on e-platforms for fraud detection

Research question three was to determine the best machine learning algorithms used on e-platforms for fraud detection. In order to answer this question, we compared and analyzed the performance of the top five machine learning algorithms based on their frequency of usage. The performance metric used in this review is accuracy. Since most of the authors used a common dataset (transactions made by credit cards in September 2013 by European cardholders.), it will be prudent to analyze and compare their performances. The results from the analysis are presented in Figure 7.

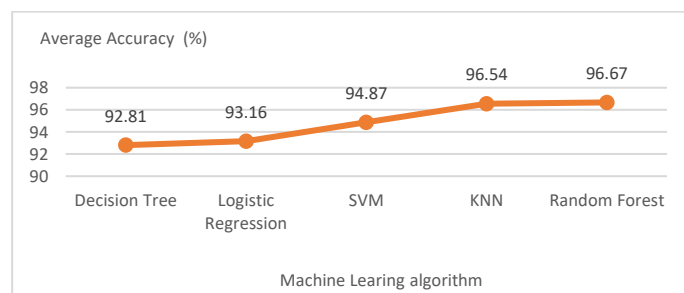


Figure 7. Average accuracies of top 5 (frequency of usage) ML algorithms

Results from Figure 7 demonstrated that the Decision Tree algorithm performs the least with an average accuracy of 92.81%. This is followed by logistic regression, SVM, KNN, and random forest algorithm with an average accuracy of 93.16, 94.87, 96.54, and 96.67 respectively. The result also indicated that, even though the KNN ranked 4th in terms of frequency of usage as shown in Table 2, it ranked 2nd in terms of performance with an accuracy of 96.54% as illustrated in Figure 7. The logistic regression algorithm ranked 2nd in terms of frequency of usage but it is the 4th best in terms of performance with an average accuracy of 93.16%. Also, The SVM algorithm ranked 3rd in terms of frequency of usage and ranked 3rd in terms of performance with an average accuracy of 94.87%. More importantly, the random forest algorithm ranked 1st in terms of frequency of usage and ranked 1st in terms of performance with an average accuracy of 96.67%.

4. CONCLUSION

This research revealed that credit card fraud is more common in the area of financial fraud on e-platforms. Based on our systematic literature review credit card fraud accounted for 82.61% of the financial fraud cases on e-platforms. Secondly, this review identified 23 different machine-learning algorithms that were used to detect fraud on e-platforms. Notable amongst them is random forest which had the highest frequency of usage followed by logistic regression and support vector machine, whereas algorithms like linear regression, Bernouli naïve Bayes, J48 algorithm, and CatBoost had the lowest frequency of usage. The research also revealed that the predominant usage of a model does not guarantee its superior performance, as exemplified by the cases of KNN and logistic regression. The logistic regression algorithm ranked 2nd in terms of frequency of usage but ranked 4th best in terms of performance with an average accuracy of 93.16%. On the contrary, the KNN ranked 4th in terms of frequency of usage but ranked 2nd in terms of performance with an accuracy of 96.54%. The random forest algorithm turned out to be the best machine learning algorithm as it is ranked first in the frequency of usage analysis and first in the performance analysis with an average accuracy of 96.67%. Hence, the random forest algorithm should be considered ahead of other machine learning algorithms for such fraud detection analysis on e-platforms. By identifying the strengths and weaknesses of different approaches, this review can help guide future research and inform the development of more effective fraud detection systems. Overall, this review has identified the kinds of fraud on financial e-platforms, and proclaimed the best and least ML algorithm for fraud detection on financial e-platform. This can help guide future research and inform the development of more effective fraud detection systems. This review was limited to financial fraud detection and accuracy was the only performance metric used for the comparative analysis. Future work could therefore consider performance metrics like recall and precision in the comparative analysis.

REFERENCES




- [1] A. Srivastava, P. K. Bala, and B. Kumar, "New perspectives on gray sheep behavior in E-commerce recommendations," *Journal of Retailing and Consumer Services*, vol. 53, p. 101764, Mar. 2020, doi: 10.1016/j.jretconser.2019.02.018.
- [2] A. M. Alnasrawi, A. M. N. Alzubaidi, and A. A. Al-Moadhen, "Improving sentiment analysis using text network features within different machine learning algorithms," *Bulletin of Electrical Engineering and Informatics*, vol. 13, no. 1, pp. 405–412, Feb. 2024, doi: 10.11591/eei.v13i1.5576.
- [3] D. Choi and K. Lee, "An artificial intelligence approach to financial fraud detection under IoT environment: a survey and implementation," *Security and Communication Networks*, vol. 2018, pp. 1–15, Sep. 2018, doi: 10.1155/2018/5483472.
- [4] K. G. Al-Hashedi and P. Magalingam, "Financial fraud detection applying data mining techniques: A comprehensive review from 2009 to 2019," *Computer Science Review*, vol. 40, p. 100402, May 2021, doi: 10.1016/j.cosrev.2021.100402.
- [5] I. El Naqa, R. Li, and M. J. Murphy, Eds., *Machine Learning in Radiation Oncology*. Cham: Springer International Publishing, 2015. doi: 10.1007/978-3-319-18305-3.
- [6] S. Masrom, N. H. Abdul Samad, R. Septiyanti, N. Roslan, and R. A. Rahman, "Machine learning prediction for academic

- misconduct prediction: an analysis of binary classification metrics,” *Bulletin of Electrical Engineering and Informatics*, vol. 13, no. 1, pp. 388–395, Feb. 2024, doi: 10.11591/eei.v13i1.5629.
- [7] A. M. Al-Sabaei, H. Alhussian, S. J. Abdulkadir, and A. Jagadeesh, “Prediction of oil and gas pipeline failures through machine learning approaches: A systematic review,” *Energy Reports*, vol. 10, pp. 1313–1338, Nov. 2023, doi: 10.1016/j.egy.2023.08.009.
- [8] M. Thilakarathne, K. Falkner, and T. Atapattu, “A systematic review on literature-based discovery workflow,” *PeerJ Computer Science*, vol. 5, p. e235, Nov. 2019, doi: 10.7717/peerj-cs.235.
- [9] F. K. Alarfaj, I. Malik, H. U. Khan, N. Almusallam, M. Ramzan, and M. Ahmed, “Credit card fraud detection using state-of-the-art machine learning and deep learning algorithms,” *IEEE Access*, vol. 10, pp. 39700–39715, 2022, doi: 10.1109/ACCESS.2022.3166891.
- [10] M. M. M. Megdad, B. S. Abu-Nasser, and S. S. Abu-Naser, “Fraudulent financial transactions detection using machine learning,” *International Journal of Academic Information Systems Research (IJAIRS)*, vol. 6, no. 3, pp. 30–39, 2022.
- [11] S. Khatri, A. Arora, and A. P. Agrawal, “Supervised machine learning algorithms for credit card fraud detection: A comparison,” in *Proceedings of the Confluence 2020 - 10th International Conference on Cloud Computing, Data Science and Engineering*, IEEE, Jan. 2020, pp. 680–683. doi: 10.1109/Confluence47617.2020.9057851.
- [12] V. Jain, H. Kavitha, and S. Mohana Kumar, “Credit Card Fraud Detection Web Application using Streamlit and Machine Learning,” in *2022 IEEE International Conference on Data Science and Information System (ICDSIS)*, IEEE, Jul. 2022, pp. 1–5. doi: 10.1109/ICDSIS55133.2022.9915901.
- [13] S. Khatri, A. Arora, and A. P. Agrawal, “Supervised Machine Learning Algorithms for Credit Card Fraud Detection: A Comparison,” in *2020 10th International Conference on Cloud Computing, Data Science & Engineering (Confluence)*, IEEE, Jan. 2020, pp. 680–683. doi: 10.1109/Confluence47617.2020.9057851.
- [14] B. Keswani *et al.*, “Adapting Machine Learning Techniques for Credit Card Fraud Detection,” in *Advances in Intelligent Systems and Computing*, vol. 1087, 2020, pp. 443–455. doi: 10.1007/978-981-15-1286-5_38.
- [15] M. B. Islam, C. Avornu, P. K. Shukla, and P. K. Shukla, “Cost Reduce: Credit Card Fraud Identification Using Machine Learning,” in *2022 7th International Conference on Communication and Electronics Systems (ICCES)*, IEEE, Jun. 2022, pp. 1192–1198. doi: 10.1109/ICCES54183.2022.9835811.
- [16] N. K. Trivedi, S. Simaiya, U. K. Lilhore, and S. K. Sharma, “An efficient credit card fraud detection model based on machine learning methods,” *International Journal of Advanced Science and Technology*, vol. 29, no. 5, pp. 3414–3424, 2020.
- [17] K. Huang, “An Optimized LightGBM Model for Fraud Detection,” *Journal of Physics: Conference Series*, vol. 1651, no. 1, p. 012111, Nov. 2020, doi: 10.1088/1742-6596/1651/1/012111.
- [18] K. Abhirami, A. K. Pani, M. Manohar, and P. Kumar, “An Approach for Detecting Frauds in E-Commerce Transactions using Machine Learning Techniques,” in *Proceedings - 2nd International Conference on Smart Electronics and Communication, ICOSCE 2021*, IEEE, Oct. 2021, pp. 826–831. doi: 10.1109/ICOSCE51865.2021.9591720.
- [19] A. Singh, A. Singh, A. Aggarwal, and A. Chauhan, “Design and Implementation of Different Machine Learning Algorithms for Credit Card Fraud Detection,” in *2022 International Conference on Electrical, Computer, Communications and Mechatronics Engineering (ICECCME)*, IEEE, Nov. 2022, pp. 1–6. doi: 10.1109/ICECCME55909.2022.9988588.
- [20] G. Sandhya, M. Abishek, S. Gunal Kumar, and R. S. Jisenthira Kumar, “Credit Card Fraud Detection using Machine Learning Algorithms,” in *Lecture Notes in Networks and Systems*, vol. 516, 2023, pp. 313–320. doi: 10.1007/978-981-19-5221-0_30.
- [21] P. Sharma, S. Banerjee, D. Tiwari, and J. C. Patni, “Machine learning model for credit card fraud detection-A comparative analysis,” *International Arab Journal of Information Technology*, vol. 18, no. 6, pp. 789–796, 2021, doi: 10.34028/iajit/18/6/6.
- [22] A. Menshchikov, V. Perfilev, D. Roenko, M. Zykin, and M. Fedosenko, “Comparative Analysis of Machine Learning Methods Application for Financial Fraud Detection,” in *2022 32nd Conference of Open Innovations Association (FRUCT)*, IEEE, Nov. 2022, pp. 178–186. doi: 10.23919/FRUCT56874.2022.9953872.
- [23] B. P. Verma, V. Verma, and A. Badholia, “Hyper-Tuned Ensemble Machine Learning Model for Credit Card Fraud Detection,” in *2022 International Conference on Inventive Computation Technologies (ICICT)*, IEEE, Jul. 2022, pp. 320–327. doi: 10.1109/ICICT54344.2022.9850940.
- [24] J. Domashova and O. Zabelina, “Detection of fraudulent transactions using SAS Viya machine learning algorithms,” *Procedia Computer Science*, vol. 190, pp. 204–209, 2021, doi: 10.1016/j.procs.2021.06.025.
- [25] J. K. Afriyie *et al.*, “A supervised machine learning algorithm for detecting and predicting fraud in credit card transactions,” *Decision Analytics Journal*, vol. 6, p. 100163, Mar. 2023, doi: 10.1016/j.dajour.2023.100163.
- [26] M. Â. L. Moreira *et al.*, “Exploratory analysis and implementation of machine learning techniques for predictive assessment of fraud in banking systems,” *Procedia Computer Science*, vol. 214, no. C, pp. 117–124, 2022, doi: 10.1016/j.procs.2022.11.156.
- [27] E. Strelcenia and S. Prakoornwit, “Comparative Analysis of Machine Learning Algorithms using GANs through Credit Card Fraud Detection,” in *2022 International Conference on Computing, Networking, Telecommunications & Engineering Sciences Applications (CoNTESA)*, IEEE, Dec. 2022, pp. 1–5. doi: 10.1109/CoNTESA57046.2022.10011268.
- [28] P. K. Sadineni, “Detection of Fraudulent Transactions in Credit Card using Machine Learning Algorithms,” in *2020 Fourth International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC)*, IEEE, Oct. 2020, pp. 659–660. doi: 10.1109/I-SMAC49090.2020.9243545.
- [29] S. Patil, V. Nemade, and P. K. Soni, “Predictive Modelling for Credit Card Fraud Detection Using Data Analytics,” *Procedia Computer Science*, vol. 132, pp. 385–395, 2018, doi: 10.1016/j.procs.2018.05.199.
- [30] R. Jhangiani, D. Bein, and A. Verma, “Machine Learning Pipeline for Fraud Detection and Prevention in E-Commerce Transactions,” in *2019 IEEE 10th Annual Ubiquitous Computing, Electronics and Mobile Communication Conference, UEMCON 2019*, IEEE, Oct. 2019, pp. 0135–0140. doi: 10.1109/UEMCON47517.2019.8992993.
- [31] S. Saraf and A. Phakatkar, “Detection of Credit Card Fraud using a Hybrid Ensemble Model,” *International Journal of Advanced Computer Science and Applications*, vol. 13, no. 9, 2022, doi: 10.14569/IJACSA.2022.0130953.
- [32] A. Joyson, A. G. R, and B. Tejashwini, “Credit Card Fraud Identification Using Machine Learning Algorithm,” *Journal of Contemporary Issues in Business and Government*, vol. 27, no. 3, p. 2021, 2021, doi: 10.47750/cibg.2021.27.03.220.
- [33] A. Kumar, D. Prusti, I. S. Purusottam, and S. K. Rath, “Real time SOA based credit card fraud detection system using machine learning techniques,” in *2021 12th International Conference on Computing Communication and Networking Technologies (ICCCNT)*, IEEE, Jul. 2021, pp. 1–6. doi: 10.1109/ICCCNT51525.2021.9579598.
- [34] M. S. Kumar, V. Soundarya, S. Kavitha, E. S. Keerthika, and E. Aswini, “Credit Card Fraud Detection Using Random Forest Algorithm,” in *2019 3rd International Conference on Computing and Communications Technologies (ICCCT)*, IEEE, Feb. 2019, pp. 149–153. doi: 10.1109/ICCCT2.2019.8824930.
- [35] K. D. Singh, P. Singh, and S. S. Kang, “Ensembled-based Credit Card Fraud Detection in Online Transactions,” in *AIP Conference Proceedings*, 2022, p. 050009. doi: 10.1063/5.0108873.




- [36] T.-H. Lin and J.-R. Jiang, "Credit Card Fraud Detection with Autoencoder and Probabilistic Random Forest," *Mathematics*, vol. 9, no. 21, p. 2683, Oct. 2021, doi: 10.3390/math9212683.
- [37] N. T. N. Anh, T. Q. Khanh, N. Q. Dat, E. Amouroux, and V. K. Solanki, "Fraud detection via deep neural variational autoencoder oblique random forest," in *2020 IEEE-HYDCON*, IEEE, Sep. 2020, pp. 1–6. doi: 10.1109/HYDCON48903.2020.9242753.
- [38] D. Shaohui, G. Qiu, H. Mai, and H. Yu, "Customer Transaction Fraud Detection Using Random Forest," in *2021 IEEE International Conference on Consumer Electronics and Computer Engineering (ICCECE)*, IEEE, Jan. 2021, pp. 144–147. doi: 10.1109/ICCECE51280.2021.9342259.
- [39] T. Mauritsius, S. Alatas, F. Binsar, R. Jayadi, and N. Legowo, "Promo Abuse Modeling in E-Commerce Using Machine Learning Approach," in *2020 8th International Conference on Orange Technology (ICOT)*, IEEE, Dec. 2020, pp. 1–6. doi: 10.1109/ICOT51877.2020.9468744.
- [40] B. Kilic, A. Sen, and C. Ozturan, "Fraud Detection in Blockchains using Machine Learning," in *2022 Fourth International Conference on Blockchain Computing and Applications (BCCA)*, IEEE, Sep. 2022, pp. 214–218. doi: 10.1109/BCCA55292.2022.9922045.
- [41] K. N. Brahmajirao D, "Recognizing Credit Card Fraud Using Machine Learning Methods," *Turkish Journal of Computer and Mathematics Education*, vol. 12, no. 12, pp. 3271–3278, 2021, doi: 10.17762/turcomat.v12i12.8005.
- [42] "Credit Card Fraud Detection using XGBoost Classifier with a Threshold Value", doi: 10.21203/rs.3.rs-1722294/v1.
- [43] K. Randhawa, C. K. Loo, M. Seera, C. P. Lim, and A. K. Nandi, "Credit Card Fraud Detection Using AdaBoost and Majority Voting," *IEEE Access*, vol. 6, pp. 14277–14284, 2018, doi: 10.1109/ACCESS.2018.2806420.
- [44] M. Y. Turaba, M. Hasan, N. I. Khan, and H. A. Rahman, "Fraud Detection During Financial Transactions Using Machine Learning and Deep Learning Techniques," *Proceedings of the 2022 IEEE International Conference on Communications, Computing, Cybersecurity and Informatics, CCCI 2022*, 2022, doi: 10.1109/CCCI55352.2022.9926503.
- [45] A. Mishra, "Fraud Detection: A Study of AdaBoost Classifier and K-Means Clustering," *SSRN Electronic Journal*, 2021, doi: 10.2139/ssrn.3789879.
- [46] S. Kumar, V. K. Gunjan, M. D. Ansari, and R. Pathak, "Credit Card Fraud Detection Using Support Vector Machine," *Lecture Notes in Networks and Systems*, vol. 237, pp. 27–37, 2022, doi: 10.1007/978-981-16-6407-6_3.
- [47] K. Poongodi and D. Kumar, "Support vector machine with information gain based classification for credit card fraud detection system," *International Arab Journal of Information Technology*, vol. 18, no. 2, pp. 199–207, 2021, doi: 10.34028/IAJIT/18/2/8.
- [48] A. RB and S. K. KR, "Credit card fraud detection using artificial neural network," *Global Transitions Proceedings*, vol. 2, no. 1, pp. 35–41, Jun. 2021, doi: 10.1016/j.gltp.2021.01.006.
- [49] C. . Sumanth, P. P. Kalyan, B. Ravi, and S. Balasubramani, "Analysis of Credit Card Fraud Detection using Machine Learning Techniques," in *2022 7th International Conference on Communication and Electronics Systems (ICES)*, IEEE, Jun. 2022, pp. 1140–1144. doi: 10.1109/ICES54183.2022.9835751.
- [50] H. M. M. H. Vidanelage, T. Tasnavijitvong, P. Suwimonsatein, and P. Meesad, "Study on Machine Learning Techniques with Conventional Tools for Payment Fraud Detection," in *2019 11th International Conference on Information Technology and Electrical Engineering (ICITEE)*, IEEE, Oct. 2019, pp. 1–5. doi: 10.1109/ICITEE.2019.8929952.
- [51] S. Gupta, Meenu, S. Patel, S. Kumar, and G. Chauhan, "Anomaly Detection in Credit Card Transactions Using Machine Learning," *International Journal of Innovative Research in Computer Science & Technology*, vol. 8, no. 3, pp. 67–71, May 2020, doi: 10.21276/ijrcst.2020.8.3.5.
- [52] B. A. Abdulsalami, A. A. Kolawole, M. A. Ogunrinde, M. Lawal, R. A. Azeez, and A. Z. Afolabi, "Comparative Analysis of Back-propagation Neural Network and K-Means Clustering Algorithm in Fraud Detection in Online Credit Card Transactions," *Fountain Journal of Natural and Applied Sciences*, vol. 8, no. 1, Oct. 2019, doi: 10.53704/fujnas.v8i1.315.

BIOGRAPHIES OF AUTHORS






Alimatu-Saadia Yussiff (PhD)    is a senior lecturer in the Department of Computer Science and Information Technology at the University of Cape Coast (UCC), Ghana, where she has taught a quiet numbers of computing courses. She obtained her PhD in Information Technology from Universiti Teknologi PETRONAS, Malaysia and MSc. in Information Technology from Sothern Polytechnic State University, Marieta – Georgia, USA. She has published many papers in refereed journals and international conference proceedings. Her research areas include e-learning, intelligent tutoring systems, User Experience design and Evaluation of User Interfaces for mobile and computers, human computer interaction, mobile and ubiquitous systems, and the role of women in technology and digital transformation. She is actively involved in STEM initiatives in Ghana to encourage young women to pursue STEM courses, soft skills and to advocate for secure internet use, internet safety and privacy online. She has a strong interest in bridging gender gap in computing and to make the cyber/internet safe for all. She is also a member of Ghana Science Association (GSA). She also serves as reviewer for some journals. She can be contacted at email: asyussiff@ucc.edu.gh






Lemdi Frank Prikutse    is a seasoned professional with over 20 years of experience in the Telecommunications and ICT industry. He is currently pursuing a PhD in Computer Science at the University of Cape Coast, Ghana. He is specializing in cryptographic techniques for data security in cloud computing. He also holds a Master's degree in Telecommunications Engineering from the Kwame Nkrumah University of Science and Technology (KNUST), Ghana. He also has a Bachelor of Engineering degree in Electronics from the Multimedia University, Malaysia. He is a recognized Telecoms Engineer and a member of the Ghana Institution of Engineering. His research interests are cybersecurity and data science. He can be contacted via email at: frankprikutse@gmail.com.






Georgina Asuah    is currently pursuing a Ph.D. in Computer Science at Eotvos Lorand University, Hungary. Researching on the Integration of Machine Learning and Artificial Intelligence in SAP HANA Optimization. She also holds a Master of Computer Applications (MCA) degree from Lovely Professional University, India with specialization in Data Science. She obtained her first degree in Mathematics (B. Ed Mathematics) from University of Cape Coast, Ghana. Her research interests cover machine learning, artificial intelligence, data mining, big data quality, enterprise resource planning (ERP), information systems. She can be contacted via email at: asuahgeorgina@inf.elte.hu.






Abdul-Lateef Yussiff (PhD)    is a senior lecturer in the Department of Computer Science and Information Technology at the University of Cape Coast (UCC), Ghana. He obtained PhD in Information Technology from Universiti Teknologi PETRONAS, Malaysia and obtained both MPhil Computer Science and MSc. Information Technology from Southern Polytechnic State University, Marietta – Georgia, USA. He has published many papers in refereed journals and international conference proceedings. He specializes in Artificial Intelligence, programming, problem analysis and image processing where he has published a quite number of articles. His main research interests include artificial intelligence, image processing, computer security and human computer interaction. He serves as reviewer for some journals. He can be contacted at email: ayussiff@ucc.edu.gh.






Emmanuel Dortey Tetteh    is a Lecturer at the Department of Computer Science and Information Technology, University of Cape Coast, Ghana. He obtained PhD in Information and Communication Engineering from the University of Electronic Science and Technology of China. His research interest includes software engineering, computer networking and information systems. He can be contacted at email: etetteh@ucc.edu.gh.



Norshahila Ibrahim (PHD)    is a senior lecturer at Faculty of Computing and Meta-Technology, Universiti Pendidikan Sultan Idris, Malaysia. She obtained her PhD in Information Technology at Universiti Teknologi PETRONAS. She has published many papers in refereed journals and international conference proceedings. She is a proactive consultant for assessing and validating research instruments in creative multimedia and digital games. In addition, she is a member of the editorial board in the International Journal of Heritage, Art and Multimedia (IJHAM), the Advances in Science, Technology and Engineering Systems Journal (ASTESJ), and Asian Social Science Journal (ASSJ). Her research interest includes digital games, mobile applications, multimedia applications, human computer and interaction and multimedia education. She can be contacted at email: shahila@meta.upsi.edu.my.



Wan Fatimah Wan Ahmad    Wan Fatimah Wan Ahmad has experienced of over 27 years as an academia. She obtained her PhD from the Universiti Kebangsaan Malaysia. She was formerly an Associate Professor with the Department of Computer and Information Sciences, Universiti Teknologi PETRONAS in Malaysia. She is currently working as a contract staff with CIS Department. She has been appointed as an adjunct Associate Professor at the School of Science and Technology, Asia e University (AeU), Subang Jaya 47500, Malaysia. Her research interests include topics on multimedia, human-computer interaction, mathematics education, e-learning and mobile learning. She led several research grants from Ministry of Science Technology and Innovation and Ministry of Higher Education in Malaysia. She won several awards in national and international levels of exhibition and commercialised few products. She can be contacted at email: fatimhd@utp.edu.my.