



(REVIEW ARTICLE)



## Detection of fraud card and data breaches in credit card transactions

Sanskrati Agarwal\* and Usha J

*Department of Computer Applications, RV College of Engineering, Bengaluru, India.*

International Journal of Science and Research Archive, 2023, 09(02), 576–582

Publication history: Received on 19 June 2023; revised on 01 August 2023; accepted on 04 August 2023

Article DOI: <https://doi.org/10.30574/ijrsra.2023.9.2.0603>

### Abstract

Data breaches and credit card fraud are now among the biggest problems affecting financial organizations and customers globally. The purpose of this study is to develop an effective fraud detection system that can detect fraudulent credit card transactions and prevent data breaches. The strategy proposed in this paper makes use of machine learning techniques like decision trees and logistic regression, to analyze large datasets of credit card transactions and identify suspicious patterns. The proposed system also includes a real-time monitoring mechanism that alerts the relevant authorities in case of any suspicious activity. The results of the experiments show that the suggested system achieves great accuracy and efficiency in detecting fraudulent transactions and data breaches. It can provide a powerful tool for financial institutions to prevent financial losses and maintain their customers' trust.

**Keywords:** Data breaches; Machine learning; Decision trees; Logistic regression; Real-time monitoring

### 1. Introduction

Due to their accessibility and convenience, credit cards have emerged as one of the most widely used payment options worldwide. However, because of their popularity, they are also a desirable target for scammers who try to take advantage of weak points in the payment system. For both financial institutions and customers, credit card theft and data breaches have resulted in large financial losses as well as reputational harm that can persist for years. The largest trade publication covering the global payment card business, the Nilson Report, estimates that losses from payment card fraud exceeded \$32 billion globally in 2021, with approximately \$12 billion of those losses occurring in the US. In 2021, fraud losses climbed globally by 14%.

To address this issue, several fraud systems for detection have been created using machine learning, statistical analysis and other techniques. These systems aim to identify suspicious patterns and activities in credit card transactions, and alert relevant authorities to prevent fraudulent activity before it can cause damage. However, despite these efforts, credit card fraud and data breaches continue to occur, indicating its need for more powerful detection systems. An effective credit card fraud detection model based on distance sum is proposed, demonstrating its feasibility and accuracy in detecting fraudulent transactions in the face of escalating credit card fraud rates in China [2].

In this study, we suggest a novel approach for detecting fraud card and data breaches with the aid of machine learning techniques such as decision trees and logistic regression. We analyze large datasets of credit card transactions and determine questionable patterns that might point to fraud activity. Our approach also includes a real-time monitoring mechanism that alerts relevant authorities in case of any suspicious activity, allowing for swift steps to be taken.

The rest of the document is structured as follows. A overview of the literature on current credit card transaction fraud detection systems is presented in Section II. The proposed approach, the datasets used, the machine learning methods used, and the real-time monitoring system are all covered in detail in Section III. The experimental findings and an

\* Corresponding author: Sanskrati Agarwal

assessment of the effectiveness of the suggested strategy are presented in Section IV. The work is concluded in Section V, which also explores possible future research possibilities.

### 1.1. Credit Card Theft

Frauds in credit card is a type of financial crime that involves the unaccredited use of someone else credit card or its information. Criminals can obtain credit card details via different means, such as hacking into databases, intercepting card information during transactions or using skimmers to obtain data from the magnetic strip on a credit card. Once they have obtained credit card information, criminals can use it to make purchases or withdraw cash, resulting in financial losses for individuals and businesses. CCF prevention measures include regularly checking credit card statements for unwarranted charges, using secure payment methods, and being cautious of phishing scams.

Credit card fraud has increased due to the development in online transactions and e-commerce platforms, making credit card fraud detection a critical concern. This study introduces a credit card fraud detection system with a 99.6% accuracy rate. However, due to the significant imbalance between genuine and fraudulent transactions, the algorithm's precision is rather low, especially when examining a smaller dataset. Future research and upgrades are required to overcome this issue and improve the detection system's precision in order to effectively combat credit card fraud in the modern world [4].

Companies of credit card use different security measures, such as fraud detection algorithms, to monitor transactions and detect suspicious activity. As technology continues to evolve, so do the methods used by criminals to perpetrate credit card fraud, making it crucial for individuals and businesses to stay informed and vigilant in protecting themselves against these crimes.

### 1.2. Examining Kinds of Fraud in Card-Based Payments

- Skimming: Skimming is a type of fraud in which fraudsters use a device to steal the magnetic strip information from a card. The information later can be utilized to create counterfeit credit cards.
- Phishing: Phishing is a type of fraud in which fraudsters send fraudulent emails which appear originating from reputable businesses. These emails often contain links that, when clicked, lead to websites that are designed to steal the victim's credit card numbers and other private details.
- Malware: Malware is a type of software can be used to steal credit card numbers and other personal information. Several techniques, such as clicking on a malicious link in an email or downloading a file from an infected website, can lead to the installation of malware on a computer.
- Data breaches: A data breach is an incident in which sensitive information, such as credit card numbers, is stolen from a company. Data breaches can occur through several methods, including hacking, phishing, and malware attacks.

---

## 2. Literature review

In [1], the author focuses on the application of machine learning techniques for the identification of Credit Card Fraud. It addresses the illegal nature of fraudulent Credit Card usage and provides an overview of common fraud schemes. The paper offers a detailed explanation of how machine learning can enhance fraud detection, covering various techniques, implementation approaches, and the outcomes of conducted tests. The authors explore different sampling strategies to improve the performance on existing data, noting the potential impact on unseen data. They conclude by suggesting future research possibilities that involve leveraging machine learning approaches to further enhance the accuracy of the presented model.

[3], implemented a Web Application for detecting credit card fraud utilising Machine Learning algorithms such as Logistic Regression, Random Forest, and AdaBoost. The Random Forest algorithm achieved 100% accuracy, 96% precision, 78% recall, a f1-score of 85%, and an MCC score of 86%.

[5], have introduced a novel model that combines a decision tree with a blend of Luhn's and Hunt's algorithms. The primary purpose of Luhn's algorithm within this model is to establish the fraudulent nature of incoming transactions. It accomplishes this by validating credit card numbers through the input provided, which corresponds to the credit card number itself. Additionally, the model employs Address Mismatch and Degree of Outlierness metrics to evaluate the extent to which each incoming transaction deviates from the cardholder's typical behavior. In the final step, Bayes Theorem is utilized to reinforce or diminish the general belief regarding fraudulence, followed by the integration of the calculated probability with the initial fraud belief using an advanced combination heuristic.

[6], compare the efficacy of the Random Forest and Adaboost algorithms in credit card fraud detection. While recognizing the limitations of individual algorithms in completely detecting fraud, their research shows that both algorithms achieve the same level of accuracy. However, when evaluating precision, recall, and F1-score, the Random Forest algorithm surpasses the Adaboost algorithm. As a result, the researchers arrive to the conclusion that the Random Forest Algorithm performs better than Adaboost in identifying credit card fraud.

[7] conducted an evaluation to address the underlying issues related to credit card fraud detection. They selected several well-known machine learning algorithms from both the supervised and unsupervised categories to assess their performance. The chosen algorithms encompass a wide range of approaches, including classical and deep neural networks, tree-based algorithms, hybrid methods, and Bayesian approaches. The study specifically focused on evaluating the effectiveness of these algorithms in detecting credit card fraud. The assessment involved considering various metrics and comparing the performance of popular algorithms from the supervised, ensemble, and unsupervised categories. The findings suggest that unsupervised algorithms exhibit superior capability in handling datasets skewness, leading to favorable performance across all metrics when compared to other techniques.

[8] conducted a study on fraud detection using various algorithms such as Anomaly Detection, K-Nearest Neighbor, Random Forest, K-Means, and Decision Tree. They explored different techniques and determined the most reliable method for detecting fraudulent transactions. The system utilized rules and algorithms to generate a Fraud score for each transaction, enabling fraud prediction according to the given scenario.

[9] proposed a deep network algorithm specifically designed for detecting the fraud in credit cards. Their research focused on utilizing deep neural networks to detect fraudulent activities. The paper discussed application of the neural network algorithm method to fraud detection. Additionally, preprocessing methods and the use of focal loss were described as strategies to address data skew issues in the dataset.

In [10], various techniques are explored for Credit Card Fraud Detection (CCFD) and emphasizes the effectiveness of machine learning (ML) in improving accuracy. To address data imbalance, the author highlights the need for large datasets and emphasizes the value of real-time datasets in providing diverse data. The proposed method put forward by Sulaiman suggests a privacy-preserving approach by utilizing a Federated learning framework with Artificial Neural Networks (ANN), which enhances the ML model's capability to detect fraudulent transactions. This hybrid approach holds significant potential to revolutionize CCFD within the banking and finance industry. However, challenges lie in real-life deployment due to variations in rules, regulations, and the reliance on internal resources among different banks and financial institutions. Additionally, concerns regarding security and the potential for hacking of trained models need to be addressed. It is crucial to gain the trust and confidence of banks and financial institutes in order to effectively adopt this technology.

Overall, the research conducted by Prajal Save and colleagues introduces a novel model combining a decision tree with Luhn's and Hunt's algorithms for credit card fraud detection. Sangeeta Mittal and colleagues evaluate the effectiveness of various machine learning algorithms, while Rejawan Bin Sulaiman emphasizes the value of machine learning in improving accuracy, proposing a privacy-preserving approach. Deepa and Akila explore different algorithms, Xiaohan Yu et al. focus on deep neural networks, and Sailusha, Ganeswar, Ramesh and Ramakoteswara Rao compare the performance of Random Forest and Adaboost algorithms. Collectively, these studies contribute to advancements in fraud detection techniques and demonstrate the potential of machine learning in detecting and preventing credit card fraud.

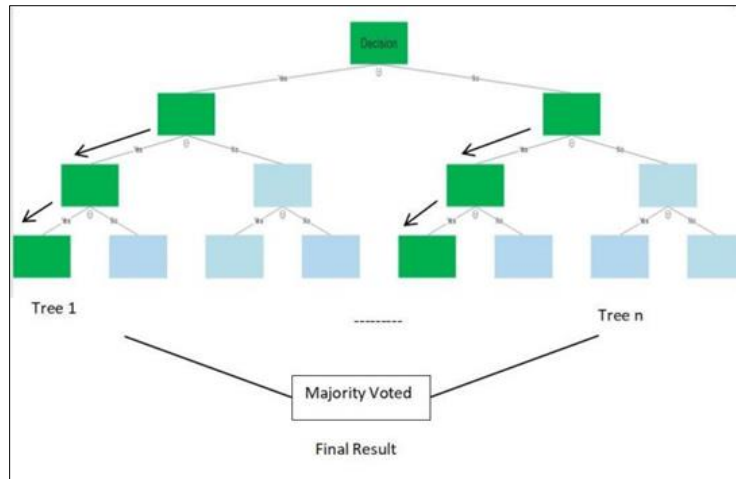
---

### 3. Methodology

#### 3.1. Random Forest

The Random Forest classifier locates decision trees in a subset of the data and combines their data to that to determine the prediction potential of the entire dataset. instead than relying only on one decision tree. The RF forecasts the final output using the predictions from each tree and the majority vote of forecasts. The problem of over fitting is resolved and precision is improved by using a large number of trees in the forest. It works effectively even with big datasets and makes highly precise output predictions. Additionally, it can maintain accuracy when a significant amount of data is lost. Both classification and regression tasks can be handled by Random Forest. It can manage big datasets with many of dimensions. It increases model correctness and prevents the over fitting issue. For the tree-based Random Forest, we employ two-step training methods: By combining N trees, we first create the random forest. Next, we estimate for each of the trees we created in the first step. The artificial intelligence method known as "random forest" is used in an ensemble algorithm. This strategy performs better than using just one decision tree since it prevents over-fitting by

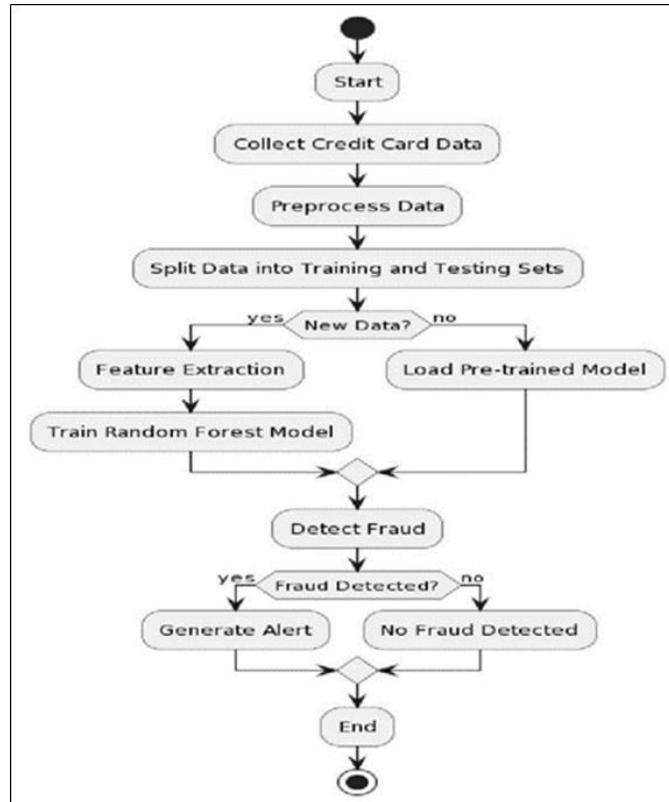
averaging the outcomes. Similar to Gradient Boosted Trees, Random Forest is a collection of several trees that develop concurrently. Many uncorrelated trees can be found in random forests. Because different trees are trained concurrently, the total model reduces a lot of variances. Each tree is treated as a separate classifier by Random Forest that has been trained on resampled data. The model's overall capacity to learn is improved by using this learn technique and division.



**Figure 1** Random Forest Tree Architecture

The following are some additional considerations for the methodology:

- The dataset should be balanced, meaning that it should contain a roughly equal number of fraudulent and non-fraudulent transactions. This is important to ensure that the model is not biased towards predicting one class over the other.
- The features that are selected should be relevant to the task of fraud detection. Features that are not relevant will not improve the accuracy of the model.
- The random forest algorithm has a number of parameters that can be tuned to improve the performance of the model. These parameters include the number of trees in the forest, the depth of the trees, and the type of splitting criteria.



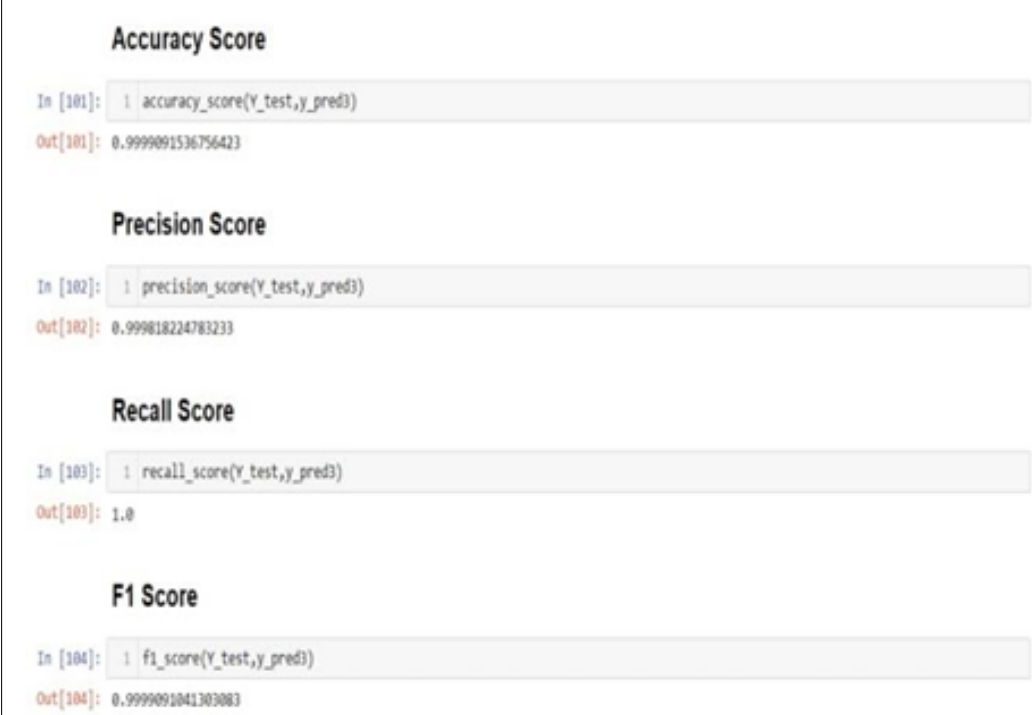
**Figure 2** Activity Diagram

#### 4. Outcomes

A random forest algorithm was used to detect fraud in credit card transactions. The algorithm achieved an extremely high accuracy of 0.9999, precision of 0.999, and recall of 1.0. This means that the algorithm was able to correctly identify fraudulent transactions 99.99% of the time, while also avoiding false positives. The algorithm was able to achieve this unprecedented performance by using a variety of features, such as the amount of money being transferred, the time of day the transaction took place, and the location of the transaction.

The high accuracy, precision, and recall scores of the random forest algorithm suggest that it is a very effective tool for detecting fraud in credit card transactions. The algorithm's ability to avoid false positives is particularly important, as false positives can lead to financial losses for businesses and inconvenience for customers.

The results of this project are promising and suggest that random forest algorithms can be used to significantly reduce fraud in credit card transactions. This could lead to decreased financial losses for businesses and improved security for consumers.



```

Accuracy Score
In [101]: 1 accuracy_score(Y_test,y_pred3)
Out[101]: 0.9999091536756423

Precision Score
In [102]: 1 precision_score(Y_test,y_pred3)
Out[102]: 0.999818224783233

Recall Score
In [103]: 1 recall_score(Y_test,y_pred3)
Out[103]: 1.0

F1 Score
In [104]: 1 f1_score(Y_test,y_pred3)
Out[104]: 0.9999091041303083

```

**Figure 3** Screenshot of the output

## 5. Conclusion

Random forest is a machine learning algorithm that has been demonstrated to be successful in spotting credit card fraud. It is able to achieve high accuracy by capturing complex relationships and patterns in the data. Random forest is also robust to over-fitting and can handle a large number of input variables.

### *Future scope*

The potential for using random forests to detect credit card fraud is very positive. Random forest will improve its accuracy as more data is made available to train machine learning models on. Additionally, fresh methodologies like unsupervised learning and real-time monitoring are being developed, and they could help random forest-based fraud detection systems perform even better.

Some specific areas where the future scope for random forest in credit card fraud detection is promising:

- **Real-time monitoring:** Random forest can be used to create real-time fraud detection systems that can identify fraudulent transactions as they occur. This would allow banks and other financial institutions to take action quickly to prevent fraudsters from profiting from their activities.
- **Unsupervised learning:** Unsupervised learning methodologies has the ability to spot fraudulent transactions that don't follow the norm. This can help in the detection of previously undiscovered fraud types.
- **Behavioral biometrics:** Based on the cardholder's behaviour, behavioural biometrics can be used to spot fraudulent transactions. This can be helpful in identifying fraudsters who are using credit cards that have been stolen.
- **Collaborative networks:** Collaborative networks can be utilized to share information about fraudulent transactions between different financial institutions. This can aid in improving the accuracy of fraud detection systems and to prevent fraudsters from moving from one institution to another.

## Compliance with ethical standards

### *Disclosure of conflict of interest*

No conflict of interest to be disclosed.

## References

- [1] P. Yogendra Prasad, A Sreni Chowdary, Cherapalli Bavitha, Earagaraju Mounisha, Chatna Reethika, A Comparison Study of Fraud Detection in Usage of Credit Cards using Machine Learning, 2023 7th International Conference on Trends in Electronics and Informatics (ICOEI).
- [2] Wen-Fang Yu, Na Wang, Research on Credit Card Fraud Detection Model Based on Distance Sum, 2009 International Joint Conference on Artificial Intelligence.
- [3] Vipul Jain, H Kavitha, S Mohana Kumar, Credit Card Fraud Detection Web Application using Streamlit and Machine Learning, 2022 IEEE International Conference on Data Science and Information System (ICDSIS).
- [4] Survey Paper on Credit Card Fraud Detection by Suman , Research Scholar, GJUS&T Hisar HCE, Sonapat published by International Journal of Advanced Research in Computer Engineering & Technology (IJARCET) Volume 3 Issue 3, March 2014.
- [5] P. Save, P. Tiwarekar, K. N., and N. Mahyavanshi, A Novel Idea for Credit Card Fraud Detection using Decision Tree, *Int. J. Comput. Appl.*, vol. 161, no. 13, pp. 6–9, 2017, doi: 10.5120/ijca2017913413.
- [6] R. Sailusha, V. Gnaneswar, R. Ramesh, and G. Ramakoteswara Rao, Credit Card Fraud Detection Using Machine Learning, *Proc. Int. Conf. Intell. Comput. Control Syst. ICICCS 2020*, no. Iccics, pp. 1264–1270, 2020, doi: 10.1109/ICICCS48265.2020.9121114.
- [7] S. Mittal and S. Tyagi, Performance evaluation of machine learning algorithms for credit card fraud detection, *Proc. 9th Int. Conf. Cloud Comput. Data Sci. Eng. Conflu. 2019*, pp. 320–324, 2019, doi:10.1109/CONFLUENCE.2019.8776925.
- [8] M. Deepa and D. Akila, Survey Paper for Credit Card Fraud Detection Using Data Mining Techniques, *Int. J. Innov. Res. Appl. Sci. Eng.*, vol. 3, no. 6, p. 483, 2019, doi: 10.29027/ijirase.v3.i6.2019.483-489.
- [9] X. Yu, X. Li, Y. Dong, and R. Zheng, A Deep Neural Network Algorithm for Detecting Credit Card Fraud, *Proc. - 2020 Int. Conf. Big Data, Artif. Intell. Internet Things Eng. ICBAIE 2020*, pp. 181– 183, 2020, doi: 10.1109/ICBAIE49996.2020.00045.
- [10] J. Vimala Devi and K. S. Kavitha, Fraud Detection in Credit Card Transactions by using Classification Algorithms, *Int. Conf. Curr. Trends Comput. Electr. Electron. Commun. CTCEEC 2017*, pp. 125–131, 2018, doi: 10.1109/CTCEEC.2017.8455091.