International Journal of Science and Research Archive
eISSN: 2582-8185
Cross Ref DOI: 10.30574/ijsra
Journal homepage: https://ijsra.net/

(RESEARCH ARTICLE)

Check for updates

Credit Card Fraud Detection

Mayowa Timothy Adesina *and Luke Howe

Data Analytics Department, College of Business, Kansas State University, KS, USA.

International Journal of Science and Research Archive, 2024, 12(02), 2072–2080

Publication history: Received on 05 July 2024; revised on 12 August 2024; accepted on 14 August 2024

Article DOI: https://doi.org/10.30574/ijsra.2024.12.2.1430

Abstract

Credit card fraud poses a significant threat globally, impacting both individuals and businesses. Leveraging machine learning techniques for fraud detection is a powerful strategy, capitalizing on the ability of algorithms to analyze vast amounts of historical transaction data. These models learn intricate patterns and anomalies indicative of fraudulent behavior, enabling them to discern subtle deviations from normal spending patterns. By considering factors like transaction frequency, location, and user behavior, machine learning systems can dynamically adapt and evolve, staying ahead of sophisticated fraud tactics. This approach not only enhances the accuracy of fraud detection but also allows for real-time monitoring, enabling swift intervention to prevent unauthorized transactions and safeguard financial assets, ultimately fortifying the resilience of credit card systems in the face of evolving security threats.

In this paper, we introduce a machine learning credit card fraud detection model leveraging the Random Forest algorithm, a well-established ensemble learning technique that amalgamates multiple decision trees for enhanced predictive accuracy. The study specifically addresses the inherent challenge of class imbalance in credit card fraud datasets by comparing the performance of Random Forest under various fine-tuning methods. These methods include random oversampling, class weights adjustment, as well as the utilization of both smoke and Tomek links (a method for under sampling) and smoke oversampling.

We evaluate our model on a real-world credit card fraud dataset from Kaggle, and show that our model achieves high accuracy, precision, recall, and F1-score, while reducing the false positive rate and the false negative rate. We also discuss the advantages and limitations of our model, and suggest some possible directions for future work.

This study adds to the expanding realm of machine learning applications in credit card fraud detection, with implications reaching beyond the financial sector into diverse industries. The objective is to optimize the model's efficacy in detecting fraudulent transactions by mitigating the challenges posed by imbalanced data.

Keywords: Machine Learning; Artificial Intelligence; Credit Card; Fraud; Random Forest; Imbalanced Dataset; Financial Industry.

* Corresponding author: Luke Howe

Copyright © 2024 Author(s) retain the copyright of this article. This article is published under the terms of the Creative Commons Attribution Liscense 4.0.

1. Introduction

This project does not exist to endorse or disparage any person or organization.

1.1. Purpose

The purpose of this study is to develop a robust and adaptive credit card fraud detection system utilizing advanced machine learning algorithms to address the escalating challenges posed by sophisticated fraudulent activities, particularly in the domains of finance, e-commerce, and insurance. Automated, data-driven fraud detection has become imperative due to the evolving nature of fraudulent tactics, surpassing the capabilities of traditional methods. The study aims to contribute to the development of a comprehensive and swift fraud detection system that can keep pace with the velocity of data while ensuring adaptability to emerging threats. By leveraging machine learning techniques, the proposed system seeks to enhance security measures, protect legitimate users, and uphold the integrity of financial transactions.

1.2. Statement of Problem

In the face of increasing complexities in fraudulent activities across diverse sectors, traditional fraud detection methods are struggling to provide effective and timely solutions. The limitations of manual analysis and the rapid evolution of fraudulent tactics underscore the need for an advanced, automated fraud detection system. The problem at hand is the absence of a comprehensive and adaptive solution capable of efficiently identifying and mitigating fraudulent transactions in real-time. To address this issue, the study focuses on developing a machine learning-based fraud detection system, recognizing the urgency of an adaptive approach to stay ahead of evolving threats.

1.3. Objectives and Scope of Study

The specific objectives include achieving Enhanced Security through advanced algorithms, the Utilization of Machine Learning for adaptive fraud detection, Protection of Legitimate Users from fraudulent activities, and the Application of the proposed system within the Financial Sector. The scope of this study encompasses the development, testing, and implementation of the fraud detection model, utilizing a dataset comprising over 1.2 million credit card transactions. The study seeks to contribute significantly to the reliability and security of financial transactions, ultimately benefiting both businesses and consumers.

2. Related Work

Machine learning methods can be used to detect and prevent fraudulent transactions by learning from historical data and finding patterns and anomalies. The first example of this is simple systems based on libraries such as Scikit-learn or Imbalanced-learn. These techniques extend into more advanced and complex examples of this are ensemble learning methods (ELMs) such as Random forest.

2.1. Background

The landscape of fraud detection has undergone a paradigm shift with the integration of machine learning, offering a transformative approach to automatically discern fraudulent patterns within vast datasets. Within this context, one of the pivotal challenges is posed by imbalanced datasets, wherein the instances of legitimate transactions significantly outnumber those of fraudulent ones. To address this issue, various sampling methods, such as oversampling and undersampling, are being employed. In this research, we focus on the application of the Random Forest algorithm, a powerful ensemble learning method, to tackle the intricacies of credit card fraud detection. The Random Forest model stands out for its ability to combine insights from multiple decision trees, providing a robust framework for discerning fraudulent activities. In order to further refine the model's performance, we explore different fine-tuning approaches, including random oversampling, class weights adjustment, smoke+tomek, smoke oversampling, and the conventional Random Forest. We will be explaining each of this models for better understanding.

2.1.1. Random forest

Forest is an ensemble learning method that combines multiple decision trees to create a more accurate and generalizable model. Each decision tree is trained on a random subset of the data and features, and the final prediction is obtained by averaging or voting the predictions of all the trees. Random Forest can handle high-dimensional and imbalanced data, and can also provide feature importance and variable selection. Random Forest is widely used for classification and regression tasks, and has been shown to perform well on credit card fraud detection.

2.1.2. Random oversampling

Random oversampling is a technique to deal with the class imbalance problem, which occurs when one class is significantly more frequent than the other in the data. Class imbalance can lead to poor performance and biased results, as the model tends to favor the majority class over the minority class. Random oversampling works by randomly duplicating the minority class examples until the class distribution is balanced. This can improve the recall and F1-score of the minority class, but it can also introduce overfitting and noise, as the same examples are repeated multiple times.

2.1.3. Class weights adjustment

Class weights adjustment is another technique to deal with the class imbalance problem, which assigns different weights to different classes according to their frequency in the data. The idea is to give more importance to the minority class and less importance to the majority class, so that the model can learn from both classes equally. Class weights adjustment can be applied to the loss function or the sampling strategy of the model, and can improve the accuracy and F1-score of the minority class, but it can also reduce the precision and increase the false positive rate of the majority class.

2.1.4. Smoke + tomek

This is a hybrid technique that combines undersampling and oversampling methods to deal with the class imbalance problem. Undersampling methods reduce the number of majority class examples, while oversampling methods increase the number of minority class examples. Smoke+tomek uses a combination of smoke and tomek links to achieve this. Smoke is a synthetic minority oversampling technique that generates new minority class examples by interpolating between existing ones. Tomek links are pairs of examples from different classes that are close to each other, and can be removed to create more clear boundaries between the classes. Smoke+tomek can improve the accuracy, precision, recall, and F1-score of both classes, but it can also lose some information and diversity from the original data.

2.1.5. Smoke oversampling

Smoke oversampling is a variation of smoke that uses a different algorithm to generate synthetic minority class examples. Instead of interpolating between existing examples, smoke oversampling uses a clustering approach to identify the most informative examples and create new ones around them. This can reduce the noise and redundancy of the synthetic examples, and can also preserve the local structure and distribution of the minority class. Smoke oversampling can improve the accuracy, precision, recall, and F1-score of the minority class, but it can also introduce some distortion and overgeneralization.

2.1.6. Ground truth about the dataset

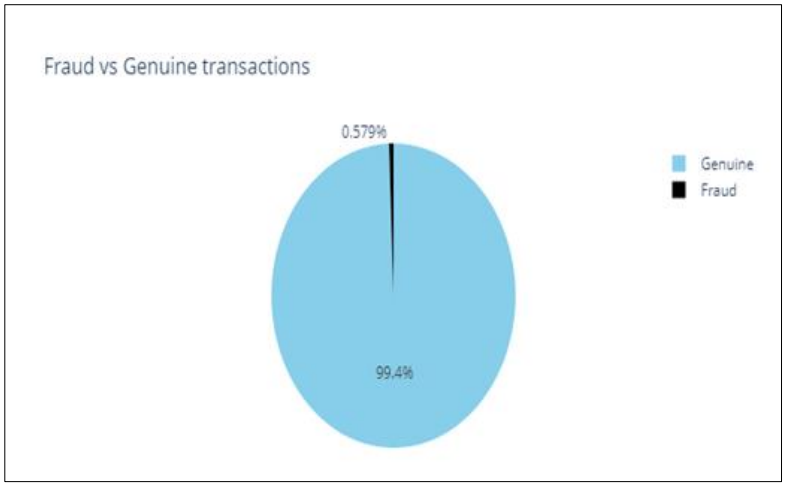

Figure 1 Ground truth about the transactions in the dataset

The dataset under examination comprises a substantial corpus of credit card transactions, surpassing 1.2 million entries. Notably, a mere 0.579% of these transactions are identified as fraudulent, underscoring the pronounced class imbalance within the dataset. This imbalance inherently poses a challenge for accurate model training, as the overwhelming majority of transactions are deemed legitimate. Comprising 23 columns, the dataset encapsulates diverse features, with temporal aspects such as transaction time and transaction amount being integral components.

The class label designates a binary classification, where a label of 1 signifies a fraudulent transaction, and 0 denotes a valid transaction. The inherent complexity of this dataset, characterized by its substantial size, imbalanced class distribution, and multidimensional feature set, makes it an ideal candidate for assessing and comparing the efficacy of various machine learning models for credit card fraud detection. It's pertinent to note that this dataset is publicly accessible on Kaggle, thereby facilitating its widespread use in the research community for benchmarking and model evaluation.

2.2. Related Work

Prior research has extensively explored the intersection of machine learning and fraud detection, particularly in the domain of credit card transactions. Imbalanced datasets have been a focal point, prompting the application of various sampling techniques to rebalance the class distribution. Notably, oversampling methods involve increasing the instances of the minority class, while undersampling aims to reduce the dominance of the majority class. Concurrently, the integration of pre-trained models, particularly those based on Bidirectional Encoder Representation using Transformers, has exhibited promising outcomes in enhancing fraud detection accuracy. However, the application of Random Forest in conjunction with different fine-tuning methods, such as random oversampling, class weights adjustment, smoke+tomek, and smoke oversampling, remains an area that demands nuanced exploration. This study aims to contribute to the existing body of knowledge by systematically evaluating the efficacy of these techniques in improving the Random Forest model's sensitivity to fraudulent patterns within a substantial dataset of over 1.2 million credit card transactions.

3. Methodology

3.1. Introduction

Random forest is the model of choice for this project. However different parameter tunning was adopted to enhance our model and to deal with the issue of class imbalance in the dataset. Let’s discuss each step used in the model development.

3.2. Data Collection & Preparation

3.2.1. Data collection

The data utilized in this project was sourced from Kaggle, a well-known platform for data science competitions and datasets. Specifically, the dataset was provided by Shenoy in 2020 and generated using Sparkov, a tool for simulating credit card transaction data. This simulated dataset serves as a valuable resource for developing and testing credit card fraud detection methodologies.

Sparkov is a tool designed for the generation of synthetic transaction data, offering a controlled environment to simulate various scenarios, including both normal and fraudulent activities. The dataset created by Sparkov on Kaggle is structured to mimic real-world credit card transactions, providing a diverse set of features and patterns that are commonly encountered in actual financial transactions.

3.2.2. Data preprocessing

Data preprocessing is a critical step in the machine learning pipeline, encompassing the cleaning and transformation of raw data to create a structured and usable dataset for training models. Raw data often comes with various imperfections, such as missing values, outliers, or inconsistencies. Cleaning involves addressing these issues to ensure the quality and integrity of the dataset. Techniques like imputation, where missing values are filled in using statistical methods, and outlier detection and removal help enhance the dataset's reliability. This process was completed by Mayowa Adesina. Furthermore, inconsistencies in data formats or representations are harmonized during cleaning to facilitate seamless integration into machine learning algorithms.

Once the data is cleaned, the next step is transforming it into a format suitable for model training. This involves encoding categorical variables, standardizing numerical features, and creating new features through techniques like feature scaling or dimensionality reduction. Encoding categorical variables converts non-numeric data into a numerical format that models can comprehend, while standardization ensures that numerical features are on a consistent scale, preventing certain features from dominating others. Feature engineering plays a crucial role in transforming the dataset by creating new informative features or capturing relevant patterns that can enhance the model's predictive capabilities. In essence, data preprocessing is akin to preparing the raw material for a machine learning model's training, ensuring that it can effectively learn from the data and make accurate predictions in real-world scenarios.

3.2.3. Model construction(random forest)

Scikit-learn library to implement the Random Forest algorithm. The ensemble nature of Random Forests, consisting of multiple decision trees, allows the model to capture intricate patterns and relationships within the data, making it well-suited for the complexities inherent in credit card fraud detection. Hyperparameter tuning was performed to optimize the model's performance, including the adjustment of parameters such as the number of trees, maximum depth, and minimum samples per leaf. This iterative process aimed to strike a balance between model complexity and generalization.

The Random Forest model's training leveraged the binary nature of the dataset, with a focus on accurately classifying transactions as either fraudulent or genuine. The scikit-learn implementation facilitates the straightforward integration of the Random Forest algorithm into the machine learning workflow, providing a powerful tool for constructing robust and accurate fraud detection models. In the subsequent sections of the methodology, I will delve into the metrics used to evaluate the model's performance, shedding light on its precision, recall, accuracy, F1 score, and the nuanced insights provided by the confusion matrix. These metrics collectively offer a comprehensive assessment of the Random Forest model's efficacy in identifying credit card fraud while balancing the costs associated with false positives and false negatives.

3.2.4. Hyperparameter tuning

The hyperparameter tuning process played a pivotal role in optimizing the performance of the Random Forest model for credit card fraud detection. Leveraging the scikit-learn library, I systematically explored and adjusted hyperparameters such as the number of trees, maximum depth of each tree, and the minimum samples required for a leaf node. This iterative optimization aimed to strike a delicate balance between model complexity and generalization, ensuring that the Random Forest effectively captured patterns in the binary training dataset while avoiding overfitting. The tuning process involved evaluating the model's performance across different hyperparameter configurations, utilizing techniques like grid search and cross-validation to identify the parameter set that maximized predictive accuracy. The resulting fine-tuned Random Forest model is well-positioned to discern subtle nuances in the data indicative of fraudulent transactions, contributing to a robust and efficient credit card fraud detection system.

3.2.5. Cross-validation

Stratified K-fold cross-validation, a key component in evaluating the Random Forest model's performance, was employed with n_splits=5 to ensure a robust and unbiased assessment of its predictive capabilities for credit card fraud detection. This method involves dividing the binary training dataset into five folds while preserving the class distribution, thus mitigating the risk of skewed representations in any given fold. Each iteration of the cross-validation process involves training the Random Forest model on four folds and validating it on the fifth, with this process repeated five times to yield a comprehensive evaluation. By stratifying the dataset in this manner, I aimed to provide a more reliable estimation of the model's ability to generalize across diverse scenarios, ensuring that it can effectively discern fraudulent transactions while maintaining consistency and reliability across different subsets of the data. This approach enhances the robustness of the model evaluation and contributes to the overall trustworthiness of the credit card fraud detection methodology.

3.3. Evaluation metrics

Five performance metrics were employed in this research work to help us understand the classification model better and to select the best model for deployment.

3.3.1. Recall

Recall, also known as sensitivity or true positive rate, is a critical metric in credit card fraud detection. It measures the proportion of actual fraudulent transactions that the model correctly identifies. A high recall is paramount in this context as it ensures that the model minimizes false negatives, thereby reducing the instances where fraudulent activities go undetected. For credit card fraud detection, where the consequences of missing a fraudulent transaction are severe, achieving a high recall is a priority to enhance the model's sensitivity to identifying potential fraud.

3.3.2. Precision

This is the ratio of correctly predicted positive instances to the total predicted positive instances. In the context of credit card fraud detection, precision signifies the accuracy of the model in identifying transactions as fraudulent. A high precision is crucial for minimizing false positives, ensuring that legitimate transactions are not incorrectly flagged as

fraudulent. Given the potential inconvenience and disruption caused by false alarms in financial transactions, achieving a balance between precision and recall is vital in designing an effective and reliable fraud detection model.

3.3.3. Accuracy

The ratio of correctly predicted instances to the total instances, provides a holistic measure of the model's overall correctness. While accuracy is a widely used metric, it may not be the sole determinant of model efficacy, especially in imbalanced datasets where the majority class dominates. In credit card fraud detection, where fraudulent transactions are typically rare, accuracy alone may be misleading. Therefore, a comprehensive evaluation considering other metrics is essential to ensure the model's effectiveness.

3.3.4. F1 score

The F1 score, or harmonic mean of precision and recall, provides a balanced assessment of a model's performance. This metric is particularly valuable in credit card fraud detection, where both false positives and false negatives have significant implications. The F1 score considers both the precision and recall, offering a single metric that reflects the trade-off between these two crucial aspects of model performance. Striking the right balance is essential to ensure that the model effectively identifies fraudulent transactions while minimizing the impact of false alarms on legitimate transactions.

3.3.5. Confusion matrix

The confusion matrix is a fundamental tool for understanding the performance of a credit card fraud detection model. It provides a detailed breakdown of true positives, true negatives, false positives, and false negatives. This matrix is instrumental in calculating other metrics such as precision, recall, and accuracy. Examining the confusion matrix allows for a nuanced analysis of the model's strengths and weaknesses, offering insights into the types of errors it may make. Understanding the distribution of predictions helps refine the model and fine-tune its parameters, contributing to an optimized credit card fraud detection system.

4. Results

4.1. Introduction

The dataset comprises 7,506 instances of fraudulent transactions, equivalent to an approximate sum of 4 million US dollars. These transactions involve dealings with over 600 distinct merchants. Notably, the months of March and May account for the highest percentages of fraudulent transactions, representing 12.50% and 12.46% of the total fraudulent set in the dataset, respectively. Predominant transaction categories include Grocery_POS, Shopping_net, and misc_net. Bar chart below represents different categories in the dataset.

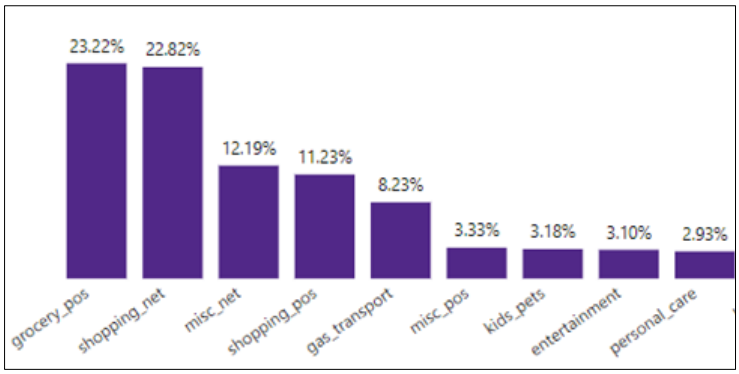

Figure 2 Different categories in the Fraudulent transactions in the dataset

Within the dataset, New York, Pennsylvania, Texas, and California emerged as the states with the highest incidence of fraudulent transactions.

4.2. Results

4.2.1. Performance Comparison

This report presents the performance metrics of a Random Forest model applied to credit card fraud detection under various sampling and weighting strategies. The metrics include Recall, Precision, F1 Score, and Accuracy.

Performance comparison

```
predictions = pd.concat([rf_score, over_rf_score, smote_rf_score, SMOTETomek_rf_s
predictions.sort_values(by=['Recall'], ascending=False)
```

	Random Forest with	Recall	Precision	F1 Score	Accuracy
1	Random Oversampling	0.868	0.091	0.164	0.947
4	Class weights	0.847	0.111	0.197	0.959
3	SMOTE + Tomek	0.818	0.698	0.753	0.996
2	SMOTE Oversampling	0.798	0.131	0.226	0.968
0	No Under/Oversampling	0.287	0.951	0.441	0.995

Figure 3 Results of different techniques used with Random Forest

The assessment reveals the effectiveness of the SMOTE + Tomek method, showcasing its robustness in achieving elevated precision, recall, and accuracy. Similarly, the Class Weights technique demonstrates a harmonious equilibrium between precision and recall. The selection of the optimal strategy for the credit card fraud detection system should align with specific goals and priorities, considering the trade-offs between precision and recall. Further investigation into these strategies, coupled with a detailed analysis of the confusion matrix, will play a pivotal role in enhancing and fine-tuning the model for seamless real-world deployment. Confusion matrix using Random Oversampling provides the best results in the binary classification task.

- True Positives (TP): 1955 instances - These are the transactions that the model correctly identified as fraudulent.
- True Negatives (TN): 367,062 instances - These are the transactions correctly classified as non-fraudulent by the model.
- False Positives (FP): 19,689 instances - These are non-fraudulent transactions that the model incorrectly predicted as fraudulent.
- False Negatives (FN): 297 instances - These are fraudulent transactions that the model failed to identify correctly.

This confusion matrix provides a detailed insight into the model's performance, allowing for a nuanced evaluation of its strengths and weaknesses. In particular, it sheds light on the balance between correctly identifying fraudulent transactions (True Positives) and avoiding misclassifying legitimate transactions as fraudulent (False Positives).

The Receiver Operating Characteristic Curve is a plot of the true positive rate against the false positive rate for different threshold values. The ROC curve provides insights into the model's performance across different classification thresholds. AUC-ROC Curve for Random Forest with Class weights: 0.9039 was achieved. The Precision-Recall (P/R) curve is a graphical representation of the performance of a classification model at different classification thresholds. We have an average precision of 0.66 which is equivalent to 66%.

5. Discussion

The credit card fraud detection project was a thrilling adventure of building and testing a Random Forest model using a mysterious dataset from Kaggle, created by Sparkov. The data was carefully cleaned, trimmed, and transformed to make it ready for the modeling process. The Random Forest algorithm was selected for its ability to combine multiple decision trees and produce a powerful prediction through the scikit-learn library. The experiment involved different ways of handling the imbalanced data, such as Random Oversampling, Class Weights, SMOTE + Tomek, SMOTE Oversampling, and No Under/Oversampling. The model was fine-tuned to achieve the best performance, and the results were measured using various metrics such as Recall, Precision, F1 Score, and Accuracy through stratified K-fold cross-validation.

Our Findings show that SMOTE + Tomek performs well, while the Class Weights technique demonstrates more stability between precision and recall. Random Oversampling had the highest recall, and classified each transaction best as discussed in the confusion matrix.

The findings provided a solid foundation for further improvement and optimization of the model, considering the real-world challenges and the specific needs of fraud detection systems.

5.1. Future works

Dynamic Risk Assessment: For future works, incorporating dynamic risk assessment mechanisms into the credit card fraud detection model could enhance its adaptability to evolving patterns of fraudulent activities. Traditional risk assessment often relies on static parameters, but dynamic risk assessment considers changing factors in real-time, providing a more responsive and proactive approach to identifying potential threats. Integrating features that assess transaction risk dynamically, considering factors such as transaction frequency, geographic location, and recent user behavior, could contribute to a more resilient and adaptive fraud detection system. Behavioral Biometrics: In future research, the integration of behavioral biometrics could add an extra layer of security to the credit card fraud detection model. Analyzing user behavior patterns, such as typing dynamics, mouse movements, or transaction interaction sequences, can establish a unique behavioral profile for each user. Leveraging this behavioral biometric data alongside traditional transaction features may enhance the model's ability to distinguish between legitimate and fraudulent transactions, particularly in cases where stolen credentials are used for unauthorized transactions. Privacy-Preserving Techniques: As a direction for future exploration, a focus on privacy-preserving techniques is crucial in the realm of credit card fraud detection, especially given the sensitivity of personal information. Incorporating advanced privacy-preserving methods, such as Homomorphic Encryption, Secure Multi-Party Computation, Federated Learning, and Differential Privacy, can ensure the protection of personally identifiable information (PII) while still allowing the model to extract valuable insights. Strategies like aggregation, tokenization, data masking, and anonymization or pseudonymization can be further investigated to strike a balance between privacy preservation and model performance. This emphasis on privacy is vital in complying with regulations and building trust with users concerned about the security of their financial information.

These future works aim to augment the existing credit card fraud detection model by introducing dynamic risk assessment, behavioral biometrics, and advanced privacy-preserving techniques. By incorporating these elements, the model can evolve to address emerging challenges and adhere to increasingly stringent privacy standards, ensuring a more resilient and secure approach to credit card fraud detection in the ever-evolving landscape of financial transactions.

Compliance with ethical standards

Disclosure of conflict of interest

No conflict of interest to be disclosed.

References

[1] Blog, D. C. (2022, 05 23). Deepcheck. Retrieved from https://deepchecks.com/top-techniques-for-cross-validation-in-machine-learning/

[2] Google. (2023). Bard. Retrieved from bard.google.com: https://bard.google.com/chat/fa495c83077a5222

[3] IBM. (2023). Retrieved from IBM.com: https://www.ibm.com/cloud/learn/exploratory-data-analysis

[4] Igareta, A. (2021, 7 21). Towards Data Science. Retrieved from Towardsdatascience.com: https://towardsdatascience.com/stratified-sampling-you-may-have-been-splitting-your-dataset-all-wrong-8cfdd0d32502

[5] Inscribe. (2023). Retrieved from Inscribe.ai: https://www.inscribe.ai/fraud-detection/credit-fraud-detection

[6] N. Prabha, S. M. (2022). IEEE.org. Retrieved from ieeexplore.ieee.org: https://ieeexplore.ieee.org/document/9742878

[7] Open AI. (2023, 09). Chat GPT. Retrieved from https://chat.openai.com/

[8] Shenoy, K. (2020). Credit Card Transactions Fraud Detection Dataset. Retrieved from Kaggle: https://www.kaggle.com/datasets/kartik2112/fraud-detection