SHS Web of Conferences **218**, 02017 (2025) https://doi.org/10.1051/shsconf/202521802017
ICDDE 2025

A Machine Learning Approach to Credit Card Transaction Fraud Prediction

Zixuan Liu*

College of Arts and Sciences, Boston University, Boston, MA, 02215, United States

Abstract. Credit card fraud has a significant impact on the financial industry and is now a growing concern. Machine learning can minimize bias against legitimate transactions and enable accurate identification of fraud. This study explores machine learning techniques to address category imbalances in credit card fraud detection datasets to mitigate economic losses while improving model performance. The results show that logistic regression outperforms other classifiers, including support vector classifiers (SVC), K-nearest neighbor classifiers (KNN), and decision trees, achieving an optimal balance between fraud detection and minimizing false positives. By conducting data processing techniques such as feature scaling and dataset balancing, the model shows an effective identification of fraudulent transactions that rarely exist in a vast number of legitimate transactions. In addition, simple neural networks trained on oversampled data reveal higher recall rates but at the cost of higher false positives, highlighting the tradeoff between accuracy and fraud detection sensitivity. These findings underscore the importance of choosing models that can both effectively detect fraud and minimize disruption to legitimate transactions, which also provide valuable insights for financial institutions seeking to enhance their fraud detection systems.

1 Introduction

Today, with the increasing volume of transactions, leading fraudsters continue to develop advanced methods for making further fraud, which has become an alarming issue. Cruz's research found that 62 million Americans had their credit or debit cards fraudulently used last year alone, with more than $6.2 billion in unauthorized spending each year. More surprisingly, around 63% of U.S. credit card users have become victims of credit card fraud, and more than half of them have experienced repeated incidents, which highlights the significance of dealing with this troublesome problem [1]. This also further raises the concern of the credit card fraud issue and illustrates the importance of improving the accuracy and efficiency of fraud detection for financial companies and society as a whole.

With the increasing threat of credit card fraud, extensive research has also been applied on detection methods in academia. Earlier research by Bolton and Hand pioneered the use of statistical outlier detection and rule-based systems to flag abnormal transactions [2]. However, as fraudsters' strategies have also become more sophisticated, traditional methods

---

* Corresponding author: Zixuan05@bu.edu

have proven inefficient due to their only focusing on static analysis, a limitation that has prompted people to turn to machine learning (ML) methods. Based on this, Chauhan and his colleagues studied the effectiveness of supervised learning models such as random forests and logistic regression (LR), in processing unbalanced data sets, which is a common challenge in fraud detection [3]. Recently, some deep learning techniques like convolutional neural networks (CNNs) and recurrent neural networks (RNNs) have been explored by Ibomoiye and Nobert for their ability to solve the computational cost and interpretability but still hide potential challenges [4]. Whitrow and his team highlighted the risk of overfitting ML models due to extreme class imbalances, while Bhattacharyya calls for the need for real-time processing on minimizing the potential factors of fraud alerts [5, 6].

Collectively, these studies emphasize the need for balancing the accuracy of the model, computational efficiency, and ethics. This research aims to explore anti-fraud detection to minimize the misclassification of legitimate transactions and reduce customer dissatisfaction caused by financial losses.

2 Dataset description

The dataset employed in this paper originates from Kaggle and was produced by Janio Martinez Bachmann [7]. The whole dataset consists of 284,807 credit card transactions recorded with 492 labeled as fraud, over two days in September 2013, 30 features, and 1 target variable providing a comprehensive representation of real-world credit card activity. Among these, Time and Amount are the two main untransformed features. Time indicates how many seconds have passed since the first transaction, and the Amount shows the transaction value in euros. The remaining 28 features (V1 to V28) are anonymized numerical values that are transformed to protect sensitive details. The Class variable acts as a label on showing the fraudulence where 1 represents a fraudulent transaction and 0 represents a non-fraudulent transaction. The dataset faces some significant challenges, and severe class imbalance becomes the most remarkable problem. Across the entire dataset, fraudulent transactions accounted for only 0.17% (about 492 cases) of the dataset, while 99.83% (284,315 cases) were legitimate, making it difficult for ML models to detect fraud under such an extremely unbalanced premise.

3 Methodology

3.1 Data processing

Data preprocessing techniques were applied for solving the severe class imbalance problem where only 0.17% of transactions in the dataset were fraudulent. The two unconverted features, Time and Amount, are scaled using RobustScaler to minimize the impact of outliers and ensure consistency between features. Random undersampling creates a balanced dataset that matches 492 non-fraudulent transactions with 492 fraudulent transactions, resulting in a 50:50 class distribution, which helps reduce the bias but still has a risk of losing valuable information from most classes that potentially affect model generalization. To mitigate this problem, the dataset is segmented before resampling, ensuring that model evaluation is performed on the original, unaltered data.

3.2 Correlation analysis and feature selection

Feature selection is a crucial step in improving the performance of fraud detection by identifying the most relevant variables in the data set. The dataset consists of 30 anonymous principal components (V1- V28) and two unconverted features, Time and Amount, which represent the time elapsed since the first transaction and the value of the transaction expressed in euros, respectively. To analyze the relationship between these characteristics and fraudulent activity, a correlation matrix was created using the subsampled data set, as shown in Fig. 1. This matrix reveals patterns that were previously masked by a large number of non-fraudulent transactions. Certain features, such as V10, V12, V14, and V17, showed a strong negative correlation with fraudulent transactions, with others having a weak or negligible correlation. These findings are further supported by the boxplot of the distribution in Fig. 2, which compares fraudulent (red) and legitimate (blue) transactions. The boxplot shows that these characteristic values for fraudulent transactions are consistently low, with the sharp drop in V10 being particularly significant, suggesting that it can catch anomalies in transaction behavior. Similarly, declines in V12 and V14 reinforce their predictive importance, suggesting that low values for these features are strong indicators of fraudulent activity. To further refine the dataset, the interquartile interval (IQR) method was used to detect and remove extreme outliers. Through the filtration of outside values from the Q1 to Q3 range, the dataset becomes more structured, reducing potential distortions, and the model's ability to generalize also improves. The combination of correlation analysis and outlier removal optimizes the data set for fraud detection, with particular emphasis on features that illustrate strong negative correlations. This refined data set provides a stronger basis for training classification models and ensures greater accuracy and reliability in distinguishing between fraudulent and legitimate transactions.

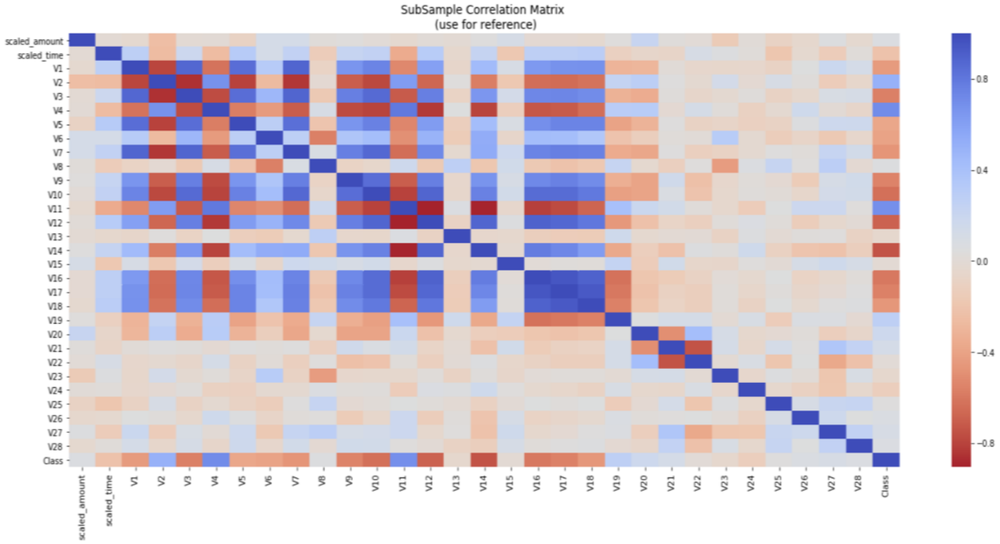


Fig. 1. Correlation matrix for subsampled dataset (Photo/Picture credit: Original).

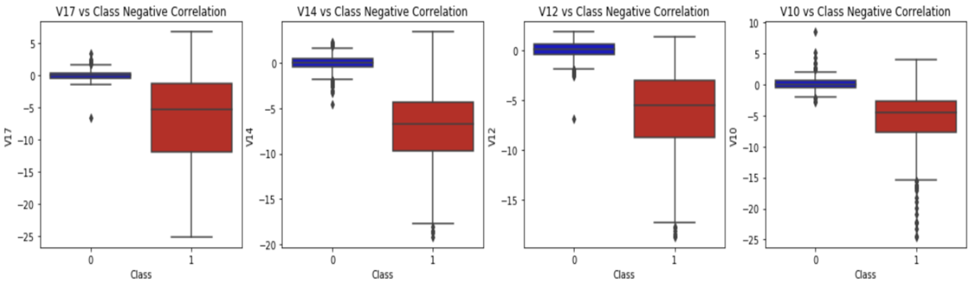
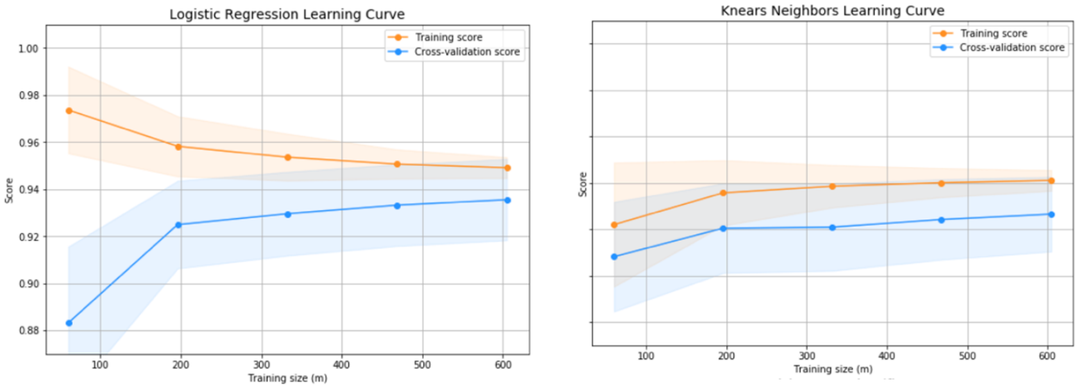


Fig. 2. Boxplot of negatively correlated features (V10, V12, V14, V17) with fraudulent transactions (Photo/Picture credit: Original).

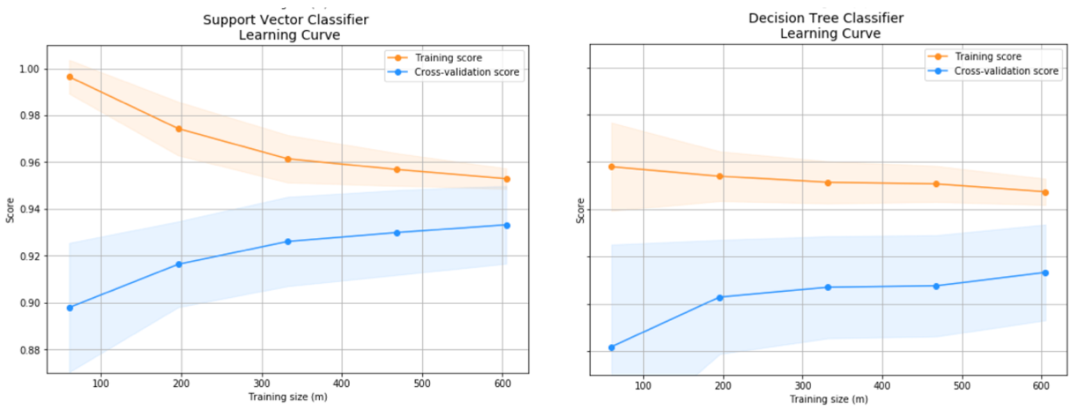
3.3 Model selection and implementation

Selecting the appropriate machine learning model is critical in identifying fraudulent transactions effectively. This study evaluated four classifiers - logistic regression (LR), support vector classifier (SVC), K-nearest neighbor (KNN), and decision tree (DT) applying ROC-AUC to evaluate their performance. Fig. 3 shows the learning curves of the four classifiers, with the horizontal axis representing the training size and the vertical axis representing the model accuracy. Each plot contains two curves: the training score (orange) that represents the model's accuracy on the training data and the cross-validation score (blue) that measures generalization on unseen data. LR (Fig. 3A) shows stable convergence between the training scores, and the validation scores emerge strong generalization without overfitting. In contrast, other models demonstrate a tendency to overfit, performing well on training data but struggling on unseen transactions, leading to higher cases of false positives or missed reports of fraud. The ROC curve in Fig. 4 further reveals the tradeoff between true and false positive rates. The higher the curve, the better the detection capability. LR had the highest ROC-AUC score (0.9799), followed by SVC (0.9747). DT and KNN were less effective, with ROC-AUC of 0.9174 and 0.9246, respectively. These results highlight LR's excellent balance between accuracy and generalization, making it the most appropriate classifier for this dataset.



A: LR

B: KNN



C: SVC

D: DT Classifer

Fig. 3. Learning curves for classifiers (Photo/Picture credit: Original).

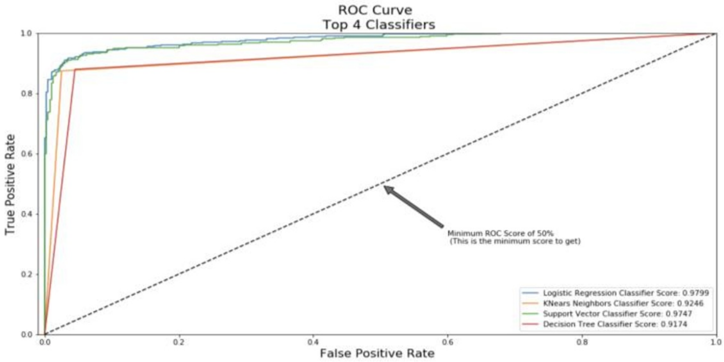


Fig. 4 Receiver Operating Characteristic (ROC) curves for classifiers (Photo/Picture credit: Original).

3.4 Neural network implementation

Neural networks can capture complex patterns through multiple layers of computation that are explored to enhance fraud detection. The model was trained on both undersampled and oversampled datasets, with SMOTE used to balance the data in general. The results exhibit that it has an advantage in capturing complex transaction patterns but still lacks precision, which leads to a higher false positive rate but still lacks precision, which leads to a higher false positive rate. In contrast, Fig. 5 demonstrates the ROC curve for Logistic Regression (LR), where the curve rises sharply toward the upper left corner, indicating strong performance in distinguishing fraudulent from legitimate transactions. The curve remains well above the diagonal red dashed line, representing random classification, confirming LR's balance between sensitivity and specificity. These results underscore LR's stability and interoperability for fraud detection and makes it and a better choice compared to the neural network. The findings further suggest the future exploration of hybrid models or cost-sensitive learning to improve deep learning approaches for fraud prevention.

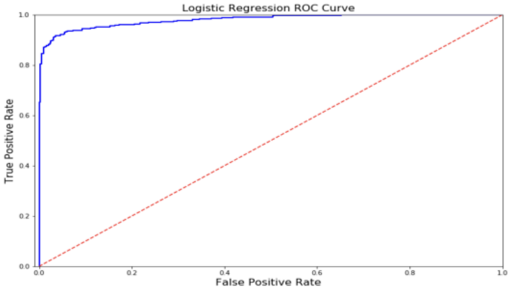


Fig. 5. LR ROC curve (Photo/Picture credit: Original).

3.5 Cross-validation and testing

To better ensure the robustness of the trained models, the dataset was divided into two parts: training and testing sets; thus, the cross-validation method was applied to evaluate more accurately the performance of the model by iterative training and validating on different subsets of the dataset. The process of cross-validation is shown in Fig. 6. The X-axis represents multiple iterations of cross-validation, while the Y-axis represents the dataset composition of the training set and the verification set. Oversampling techniques are applied to address class imbalances, allowing models to learn more efficiently from fraudulent transactions. The classification results demonstrate the confusion matrix of LR (Fig. 7A), KNN(Fig. 7B), SVC(Fig. 7C), and DT model (Fig. 7D), respectively. These matrices illustrate the relationship between the actual classification and the predicted classification of transactions, where correctly classified cases appear along the diagonal, and misclassifications are reflected in the off-diagonal values. The results show that the LR model achieves the best balance between fraud detection and minimizing false positives. Most transactions with minimal error classification are classified correctly. Since models with higher recall rates typically introduce more false positives, this result highlights the tradeoff between maximizing fraud detection and avoiding false positives. When some models struggle with misclassification, LR shows the most stable and interpretable performance.

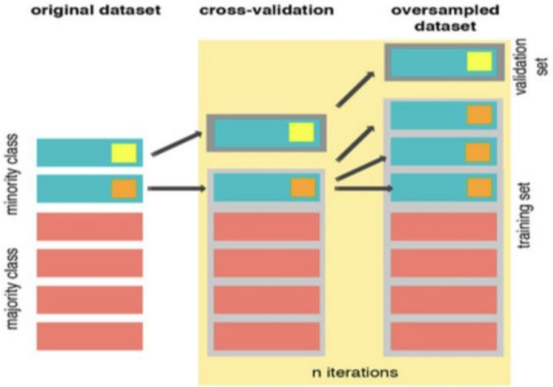
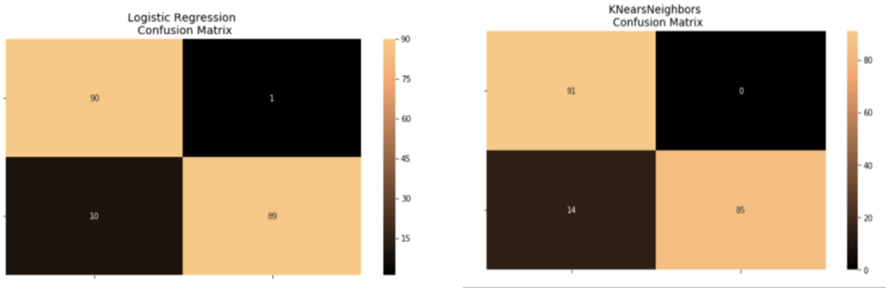
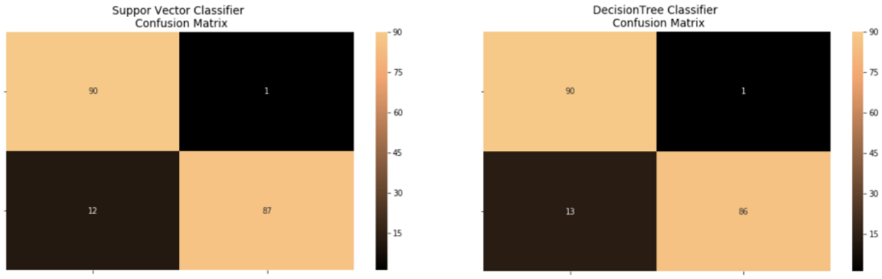


Fig. 6. Cross-validation training and validation process (Photo/Picture credit: Original).



A: LR B: KNN



C: SVC D: DT Classifier

Fig. 7. Confusion matrices for different classification models (Photo/Picture credit: Original).

3.6 Business impact

Through the implementation of such models, banks and payment service providers could have a chance to develop more efficient fraud detection systems that minimize financial losses. Fraudulent transactions pose substantial risks to financial institutions, including revenue losses and potential liabilities. According to the Nilson Report, in 2022, global credit card fraud losses reached $33 billion, which emphasizes the need for advanced fraud detection mechanisms [8]. Applying ML models helps to reduce the fraudulent transactions cases, so that saves millions in fraudulent chargebacks and unauthorized withdrawals for institutions.

Another critical aspect of the business impact is building customer trust by minimizing false positives. Many existing fraud detection systems flag legitimate transactions as suspicious, annoying and frustrating customers when payments are declined, or other additional verification steps are required. A study conducted by Castillo discovered that false positives are a leading cause of customer dissatisfaction in digital payment systems [9]. Implementing ML models helps retain customer loyalty and satisfaction so that financial institutions can ensure seamless transactions. When customers trust their financial service providers to protect them from fraud without unnecessary disruptions, they are more likely to continue using digital payment services, ultimately benefiting institutions through higher retention rates [10]

4 Challenges and future works

Despite some achievements in fraud detection, there are still challenges remaining. The most significant issue is class imbalance, with fraudulent transactions representing only 0.17% of the dataset. For dealing with the problem, techniques like undersampling and SMOTE helped balance the data, but they have limitations. Undersampling runs the risk of losing valuable information from most classes, which also means that while SMOTE introduces a synthetic pattern, it does not fully reflect real-world fraud [11]. Future research could explore hybrid models that combine traditional ML with deep learning approaches, as well as cost-sensitive learning, to better address class imbalance and improve fraud detection accuracy [12].

Another challenge is the interpretability of anonymous features (V1- V28), making it difficult to design targeted fraud detection rules successfully without specific insights. Applying explainable AI techniques like SHapley Additive exPlanations (SHAP) could help identify the most influential features and improve the transparency of the model, which would allow financial institutions to have a better understanding on the model's decisions [13]. In this study, the Neural Network model showed signs of promise, but it still could have large improvements on fraud recall on the oversampled dataset, and implicates a requirement of more advanced architectures. Based on such scenario, Long Short-Term Memory networks (LSTMs) or transformer-based models could offer a better capture of sequential dependencies in transaction patterns and help to improve the accuracy of fraud detection [14]. Additionally, online learning techniques could also be integrated for adapting to develop fraud patterns in real time [15]. Collaboration with financial institutions to access real-time transaction data would further enhance model robustness and practicality.

5 Conclusion

In conclusion, credit card fraud has an increasing impact on society which brings a critical issue on the security of transactions. The whole study highlights the importance of ML in detecting fraudulent credit card transactions by addressing significant challenges, including class imbalance, feature scaling, and model selection. The analysis demonstrates the role of fraud detection techniques in enhancing financial security, it applies clear and precise data processing methods like undersampling and SMOTE the dataset was refined so that the model training could be improved and bias could be reduced. Feature correlation analysis further helped to identify critical variables influencing fraud detection and the process of the removal of outliers ensured more accurate predictions. After a series testing on models, the LR was identified as the most effective classifier, striking a balance between fraud recall and minimizing false positives. Among other models being tested, it is capable for the Neural Network Model to catch complex data patterns, but there are still remarkable challenges in real-world applicability. On the other hand, performance metrics like ROC-AUC were also being considered in the comparison between models which results in a confirmation that the traditional classifiers have better effectivity than the deep learning methods in this scenario.

Overall, the findings of this study have revealed important ML model implications for financial institutions, which provide meaningful insights in terms of the ways that help optimize the fraud detection models while maintaining customer trust. Besides, it also demonstrates that implementing robust ML solutions like minimizing false positives could help reduce potential financial losses and reinforce the transaction security of the customers. With the increasing trend of digital transactions, fraud detection systems must evolve with adaptive learning techniques to apply effective methods to combat emerging threats. In the future, efforts should focus on the development of more complex models, integrating real-time detection methods to make the models more interpretable and to ensure that they can be effectively deployed in financial institutions.

References

1. B. Cruz, Credit card fraud 2021 annual report: prevalence, awareness, and prevention. Securityorg, 2023. Available at: https://www.security.org/digital-safety/credit-card-fraud-report/
2. R. J. Bolton, Hand, D. J., Provost, F., Breiman, L. Statistical fraud detection: a review. Statistical Science **17**, 235–255 (2002)
3. S. Chauhan, L. Vig, Anomaly detection in ECG time signals via deep long short-term memory networks. In: 2015 IEEE International Conference on Data Science and Advanced Analytics (DSAA), October 2015
4. I. D. Mienye, N. Jere, Deep learning for credit card fraud detection: a review of algorithms, challenges, and solutions. IEEE Access **12**, 96893–96910 (2024)
5. C. Whitrow, D. J. Hand, P. Juszczak, D. Weston, N. M. Adams, Transaction aggregation as a strategy for credit card fraud detection. Data Min. Knowl. Discov. **18**, 30–55 (2008)
6. S. Bhattacharyya, S. Jha, K. Tharakunnel, J. C. Westland, Data mining for credit card fraud: a comparative study. Decis. Support Syst. **50**, 602–613 (2011)
7. J. M. Bachmann, Credit fraud || dealing with imbalanced datasets. Kagglecom, 2019. Available at: https://www.kaggle.com/code/janiobachmann/credit-fraud-dealing-with-imbalanced-datasets.
8. Nilson Report, Card fraud losses worldwide. Nilson Report, 2022. Available at: https://nilsonreport.com/articles/card-fraud-losses-worldwide-2/.
9. M. Castillo, Why credit card fraud alerts are rising, and how worried you should be about them. CNBC, September 12, 2024. Available at: https://www.cnbc.com/2024/09/12/why-credit-card-fraud-alerts-are-rising.html.
10. B. Lundgren, How software developers can fix part of GDPR's problem of click-through consents. AI Soc. (2020)
11. N. V. Chawla, K. W. Bowyer, L. O. Hall, W. P. Kegelmeyer, SMOTE: synthetic minority over-sampling technique. J. Artif. Intell. Res. **16**, 321–357 (2002)
12. H. He, E. A. Garcia, Learning from imbalanced data. IEEE Trans. Knowl. Data Eng. **21**, 1263–1284 (2009)
13. S. Lundberg, S.-I. Lee, A unified approach to interpreting model predictions. arXiv preprint arXiv:1705.07874 (2017)
14. A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, I. Polosukhin, Attention is all you need. arXiv preprint arXiv:1706.03762 (2017)
15. J. Gama, I. Žliobaitė, A. Bifet, M. Pechenizkiy, A. Bouchachia, A survey on concept drift adaptation. ACM Comput. Surv. **46**, 1–37 (2014)