Dean&Francis

ISSN 2959-6130

Credit Card Fraud Detection Using Machine Learning: A Case Study on Imbalanced Data

Ziming Liu[1,]

[1]*College of Social Science, Michigan State University, East Lansing, United States*
corresponding author: Liuzimi1@msu.edu

Abstract:

Credit card fraud remains a serious problem for financial institutions and consumers worldwide. The rapid growth of online and cashless payments offers opportunities for fraudulent activity. Although fraudulent activity is not common, it can cause serious economic losses and damage clients' trust. Traditional rule-based systems often fail to capture complex and evolving fraud patterns, making machine learning (ML) methods a promising alternative. This paper takes the case study of the widely used Kaggle credit card fraud detection dataset as an example, which contains 284,807 transactions, of which only 492 cases of fraud (<0.2%). This paper explored data preprocessing, category imbalance processing, and a variety of machine learning models, including logical regression, decision trees, random forests and eXtreme Gradient Boosting (XGBoost). The results show that the integrated method (especially XGBoost) performs better than the baseline model, with the recipient operating feature curve (AUC) below AUC of 0.98, which achieves a better balance between recall and accuracy. In addition to technical results, the study highlights its commercial implications for reducing financial risks, improving efficiency and customer trust. The results show that integrating machine learning into fraud detection systems has both analytical and strategic advantages, although continuous updates and real-time adaptation are still essential.

Keywords: Credit card fraud; machine learning; Class imbalance; Fraud detection; XGBoost

1. Introduction

Credit card fraud has become one of the most pressing challenges in modern finance. With the continuous development of e-commerce, mobile payments and global financial integration, digital transaction volume is growing exponentially. In this growth, fraudulent activity is also increasing and exploits

vulnerabilities in payment systems and security protocols. While fraudulent transactions account for only a small fraction of the overall transaction, the economic losses caused by the causes are unusually huge. In addition to direct monetary losses, fraud also undermines consumer confidence, puts reputational risks on banks and increases the operating burden on merchants and regulators.

As Aparício et al. observe, "the rules performance degrades over time due to concept drift, especially of adversarial nature [1]." Traditional fraud detection systems rely heavily on manually designed rules, such as transaction limits or geographical limits. While these rules may be effective in some cases, they remain too rigid and often unable to adapt to the evolving strategies of fraudsters. Furthermore, a rule-based approach leads to a large amount of false positives, leading to frustration among customers when legal transactions are blocked. This leads to a dilemma for financial institutions: tighter detection rules can reduce fraud but also harm the user experience; more easier rules may improve convenience but put the system at greater risks.

Machine Learning (ML) provides us with flexible and adaptive solutions. By learning patterns from historical transaction data, ML models can identify nuances between fraud and legitimate behavior that static rules may ignore. More importantly, ML systems can be retrained using new data to identify emerging fraud trends and make them more resilient in a dynamic environment. Previous research has shown that algorithms such as random forests, gradient elevation, and neural networks perform strongly in fraud detection tasks, especially when combined with technologies that address data imbalances [2].

Despite these advantages, ML-based fraud detection still faces two major challenges. First, since fraud is extremely rare compared to normal transactions, there may be extreme category imbalances. In the Kaggle dataset used in this study, less than 0.2% of transactions are fraudulent. This imbalance means that a model that predicts all transactions as legitimate can still reach more than 99 percent accuracy, meaning fraud cannot be detected at all. Therefore, more robust metrics (such as accuracy, recall rates, F1 scores, and subjects' working feature curve (ROC) – area under the curve (AUC)) are needed to assess the model in this context. Second, fraud detection often takes place in real time, which requires efficient algorithms that can scale to millions of transactions per day.

The objectives of this article include three aspects. First, our goal is to structurally analyze data exploration, pre-processing and modeling of the Kaggle credit card fraud detection dataset. Second, this paper tested the challenges and pros and cons of applying machine learning methods in highly uneven data sets. Finally, this paper combines technical results with real business impacts to demonstrate how improved fraud detection can not only reduce financial losses, but also improve operational efficiency and customer trust.

Through this case study, this paper has contributed to the academic and practical discussion of fraud detection. It demonstrates the potential of integrated machine learning approaches (especially eXtreme Gradient Boosting (XGBoost)) to address rare event characteristics of fraud detection, while also underlining the importance of assessing metrics, interpretability, and business relevance.

Furthermore, the study aims to highlight the interdisciplinary nature of fraud detection, which links computer science, statistics, economics and behavioural research closely. This interdisciplinary perspective is one of the core innovations of the article, as existing studies often focus only on technical modelling and overlook the economic and behavioural foundations of fraud. Fraud is not only a technical problem, but also a human problem, as fraudsters constantly exploit various incentives and vulnerabilities. Understanding the motivations and strategies of criminals helps us design better models, and understanding customer behavior is just as important to reducing false positives. Incorporating behavioural motivation into model design represents another innovative aspect of this research, extending fraud detection beyond pure data patterns to include psychological and strategic thinking. By applying machine learning technology in a broader social and economic context, this article emphasizes that effective fraud detection requires powerful algorithms and an overall perspective on the financial ecosystem. This systemic viewpoint forms a third innovation of the study, as it positions fraud detection within the dynamics of markets, incentives and human decision-making rather than treating it as an isolated computational task.

2. Methodology

2.1 . Data Source

The data set for this study is the credit card fraud detection dataset on Kaggle. The dataset contains anonymous credit card transactions for European cardholders within two days of September 2013. Of the 284,807 transactions, only 492 were fraudulent transactions, or about 0.17% of the dataset. These features consist of 30 variables: 28 anonymous features generated by the main component analysis (V1–V28), and two non-primary components analyze features: time and amount. The target variable "category" indicates whether the transaction is fraudulent (1) or legal (0).

The dataset is widely used in fraud detection research because it represents real-world challenges: high dimensions, anonymous features, and severe category imbalances.

The imbalance rate is about 1:579, making the dataset a

ISSN 2959-6130

typical case for rare event predictions.
In addition, the anonymization of variables adds to a layer of complexity, as researchers cannot directly link features to real-world behavior. This forced assessments to rely solely on statistical signals, not on specific areas of interpretation. Since the model must detect fraud purely from hidden patterns, such conditions make the dataset a great benchmark for testing the generalization capabilities of the algorithm.

2.2 . Model Introduction

As stated by Niu et al. in their study, "this paper train … Logistic Regression, Decision Tree, Random Forest, Extreme Gradient Boosting …" to compare their performance on the fraud detection task [3].
This study selected several machine learning algorithms for comparison. Each model is selected based on its balance between interpretability, computational efficiency, and classification performance.
Logic Regression: A linear model for standardized data. It is simple to understand, easy to explain, and is often used as a benchmark.
Decision Tree: A nonlinear model that divides data based on characteristic thresholds. It can highlight important variables, but it is easy to overfit.
Random Forests: A collection of decision trees that reduce overfitting and improve robustness. It is perfect for handling imbalanced datasets with category weights.
XGBoost: A gradient lifting method that combines many weak learners into a strong classifier. It is known for its high predictive ability and effective handling imbalances through parameter adjustments.
Neural Network (optional): a feedforward model with hidden layers. While it captures complex patterns, it needs to be carefully adjusted and spent longer training time.
The study focused on XGBoost, which has been proven to be effective in fraud detection competitions and previous studies.
Still, simpler models like logical regression and decision trees are also important for comparison, as they provide baselines and highlight the relative benefits that are achieved by more complex algorithms.

2.3 . Data Preprocessing

Before training, this paper took several pre-processing steps: Data Cleaning: No missing values or duplicate lines are found, simplifying the pre-processing process.
Feature scaling: StandardScaler standardizes time and amount variables to match the characteristic scale of Principal Component Analysis (PCA) conversions. Training-Test Split: Break the dataset into 70% training sets and 30% test sets. This paper used a hierarchical sample to align fraud in both groups with the normal sample ratio. Category imbalance: During training, this paper apply category weights to higher penalties for fraud cases. This paper also tested other methods such as Synthetic Minority Over-sampling Technique (SMOTE) oversampling and undersampling, but the class weighting method provides more stable results and less overfitting.
These pre-processing steps ensure that the dataset is ready for robust model assessment. It is worth noting that in the actual banking system we can see preprocessing often involves feature engineering, such as aggregating customer history records or detecting spending speed patterns. While there are no such context features in the dataset, the standardized process here still reflects the core logic of preparing unbalanced transaction data for machine learning analytics.

3. Results

Use metrics that are more suitable for imbalance data to evaluate the model, including accuracy, recall rates, F1 values, ROC-AUC, and confusion matrices. As stated by Hajek et al. in Fraud Detection in Mobile Payment Systems: "… detection performance is negatively affected by the extreme class imbalance in financial fraud data [4]." Since all transactions are forecast to be legal, accuracy will exceed 99%, but fraud detection rate is zero, so simple accuracy is not emphasized. 3.1 Model performance logic regression: ROC-AUC is 0.94, the recall rate is relatively high, but the accuracy is moderate. This is a good benchmark, but not enough to deploy. Decision Tree: A higher recall rate is shown on training data, but overfitting, resulting in weak generalization capacity. Random forests: balanced recall rates and precision rates, with robust results in all dimensions. Its ROC-AUC is about 0.97. XGBoost: Best Performance, ROC-AUC | 0.98. It strikes a good balance between recall rates (detecting more fraud) and accuracy (avoiding false positives). Neural Networks: Results are close to random forests, but require longer training and tuning. This suggests that, while the deep model is competitive, its computational cost and complexity may exceed the advantages of such table datasets unless there are more abundant features available. 3.2 The main findings of category imbalances are the main challenges. If there is no category weighted or redraw, models tend to predict all transactions as legitimate transactions. XGBoost performs better than other models, especially in rare event detection, showing its advantages in handling tilt distribution. The feature importance analysis of tree-based models highlights that certain PCA features (e.g. V14, V17) are powerful indicators of fraud. While anonymization hinders direct interpretation, these variables are always the most important predictor, suggesting that fraudulent transactions have subtle statistical structures that machine learning models can take advantage of.

This reinforces the idea that even without a clear domain knowledge, powerful algorithms can identify hidden relationships. Evaluation metrics are crucial: While all models are accurate at more than 99%, only accuracy, recall rates, and ROC-AUC reveal real differences in fraud detection. All models have an accuracy of more than 99%, with only accuracy, recall rates and ROC-AUC revealing the real difference in fraud detection [5]. Specifically, the accuracy-recall curve suggests that a similar model of the ROC-AUC may behave very differently depending on the threshold settings, so careful calibration must be performed before deployment. 3.3 Commercial significance of results These findings are of practical significance in addition to technical accuracy: financial institutions can reduce losses by identifying risk transactions in real time. False positives can be minimized, thereby increasing customer trust and reducing complaints about payment blocks. Fraud detection systems that are compatible with machine learning have a competitive advantage in both compliance and customer experience.

In addition, providing investigators with sorting alerts (the most suspicious case-first) helps to allocate limited human resources more effectively. This means that even if the models are not perfect, they can significantly improve the efficiency of the workflow by prioritizing cases with the highest probability of fraud.

Overall, the results confirm that machine learning, especially integration methods like XGBoost, can provide a powerful way to credit card fraud detection when combined with careful treatment of category imbalances [6-8].

4. Method

This study explores the application of machine learning methods in credit card fraud detection based on Kaggle dataset, which is characterized by extreme imbalances in categories. Through data exploration, pre-processing and model comparison, the analysis shows that machine learning can effectively identify fraudulent transactions as long as the category imbalance is properly addressed and appropriate evaluation indicators are adopted. In the tested model, integrated methods such as random forests and XGBoost performed significantly better than simple models. Among them, XGBoost performed best, with ROC-AUC of about 0.98, achieving a reliable balance between recall and accuracy. The findings highlight the importance of not only accuracy, but also accuracy, recall, F1 and ROC-AUC indicators when making rare event predictions.

From a practical point of view, fraud detection using machine learning models can bring tangible benefits to financial institutions and online businesses. These models not only reduce direct financial losses, but also improve operational efficiency by reducing false alarm rates and enabling investigative teams to focus on real threats. Furthermore, fewer false positives can enhance customer trust and satisfaction, which is crucial in the digital economy. Another important point is that these systems help regulate compliance, indicating that agencies are taking proactive steps to prevent fraud and protect consumers. In a competitive financial market, having advanced fraud detection technology can also be a differentiated advantage, and a safer platform will attract more users. Looking ahead, as fraud methods evolve, fraud detection systems must be updated regularly. Future prospects include real-time fraud detection processes, capturing complex patterns using deep learning, and finding organized fraud gangs through web-based analytics. In addition, interpretability approaches such as SHapley Additive exPlanations (SHAP) values or (Local Interpretable Model-agnostic Explanations) LIME can be integrated to make model predictions more transparent to regulators and investigators. Ethical issues such as data privacy and fairness should also be carefully considered, as too radical fraud filters may have disproportionate effects on certain client groups. In short, incorporating machine learning into fraud detection is not only a technological improvement, but also a strategic business decision that enhances security, compliance and customer confidence. While the results are encouraging, this study has some limitations. The dataset used is from Kaggle and represents only a short time window, lacking contextual variables such as user behavior, geographic information, and device-level features. Furthermore, the model was evaluated in an offline environment, which may not fully reflect the challenges faced in real-world deployments. Incorporating behavioral, geographic, and device-level features, and evaluating models in real-time deployment environments. Furthermore, integrating advanced technologies such as deep learning for complex pattern recognition, along with interpretable tools like SHAP or LIME, could improve the effectiveness and transparency of fraud detection systems.

5. Conclusion

The research is far more meaningful than credit card transactions. Similar categories of imbalances, real-time decision-making and confrontational adaptation challenges are also found in areas such as insurance fraud detection, money-laundering prevention and even cybersecurity. Thus, lessons learned from credit card fraud detection can be used to inform the wider application of fintech and risk management. At the same time, researchers must recognize the limitations of benchmark datasets like Kaggle, which lack context variables and represent only a narrow window of time. Future research should explore a richer dataset, including behavioral, geographical and device-level features, and should evaluate the model

ISSN 2959-6130

in real-time deployment environments, rather than offline simulations. By combining technological innovation with responsible governance, institutions can ensure that machine learning systems can not only effectively detect fraud, but also maintain fairness, transparency and long-term trust in the digital financial ecosystem. In short, the future of fraud detection will depend on the collaboration of data scientists, financial experts and policymakers. Only through interdisciplinary efforts will machine learning reach its full potential and provide safe, adaptive and customer-friendly solutions.

References

[1] Aparício, D., Barata, R., Bravo, J., Ascensão, J. T. & Bizarro, P. (2020) ARMS: Automated Rules Management System for Fraud Detection. Proceedings of KDD, 20.

[2] Bhuiyan, M. & Farabi, S. F. (2024) Enhancing Credit Card Fraud Detection: A Comprehensive Study of Machine Learning Algorithms and Performance Evaluation.

[3] Kaggle (2020) Credit Card Fraud Detection. Available at: https://www.kaggle.com/datasets/mlg-ulb/creditcardfraud.

[4] Analytics Vidhya (2022) Exploratory Data Analysis (EDA): Credit Card Fraud Detection Case Study.

[5] Mastercard (2025) At Mastercard, AI Is Helping to Power Fraud-Detection Systems. Business Insider.

[6] Niu, X. T., Wang, L. & Yang, X. L. (2019) A Comparison Study of Credit Card Fraud Detection: Supervised Versus Unsupervised. arXiv preprint, arXiv:1904.10604.

[7] Hajek, P., Abedin, M. Z. & Sivarajah, U. (2022) Fraud Detection in Mobile Payment Systems Using an XGBoost-Based Framework. PMC/Biomedical Journal.

[8] Gao, J., Zhou, Z., Ai, J., Xia, B. & Coggeshall, S. (2019) Predicting Credit Card Transaction Fraud Using Machine Learning Algorithms. Journal of Intelligent Learning Systems and Applications, 11(3), 33-63.