Credit Card Fraud Detection Based on Machine Learning Prediction

Ge Yang

School of Software, Jiangxi University of Finance and Economics, NanChang JiangXi 330000, China.
2202100330@stu.jxufe.edu.cn

Abstracts. In recent years, credit card fraud has become increasingly rampant, posing a major threat to financial security. To effectively detect and prevent credit card fraud, this study combines three machine learning algorithms, namely Random Forest (RS), Support Vector Machine (SVM), and Logistic Regression (LR), to deeply analyze credit card transaction data through cross-validation with different multiplicity. The study results show that Random Forest performs best in terms of precision and F1 scores, SVM performs well in terms of recall, and logistic regression has a high Area Under Curve (AUC) value in distinguishing between fraudulent and non-fraudulent transactions. Through meticulous data preprocessing, feature engineering, and model optimization, this study significantly improves the performance and stability of each model.The research results of this paper provide an important reference for building an efficient and reliable credit card fraud detection system, which has important practical application value and theoretical significance across different sectors and industries.

Keywords: Credit Card Fraud Detection, Machine Learning, Random Forest, Support Vector Machine, Cross Validation.

1 Introduction

With the rapid development of digitalization and Internet finance, credit cards have become one of the most important tools for global payment transactions. However, the widespread use of credit cards has led to an increase in credit card fraud, resulting in significant economic losses and a crisis of confidence for consumers and financial institutions. According to statistics, credit card fraud costs the global economy billions annually. Therefore, it is necessary to conduct pre-purchase credit card testing [1].

Traditional fraud detection methods, such as rule-based systems, have become less effective due to their fixed logic and limited adaptability, leading to high false positives and underreporting rates. With the advancement of machine learning technology, its application in credit card fraud detection shows significant advantages. Machine learning can learn and predict unknown fraudulent behaviors from large amounts of data, significantly improving the accuracy and efficiency of fraud detection [2].

Y. Wang (ed.), *Proceedings of the 2024 2nd International Conference on Image, Algorithms and Artificial Intelligence (ICIAAI 2024)*, Advances in Computer Science Research 115,
https://doi.org/10.2991/978-94-6463-540-9_5

Research has shown that using complex algorithms can reduce the false alarm rate while improving the identification of fraudulent transactions..

This paper aims to improve credit card fraud detection using algorithms like random forest and logistic regression. The goal is to compare the performance of different machine learning models on real credit card transaction data and their effectiveness in detecting fraud. This research will preprocess and feature engineer the data, apply multiple machine learning algorithms for training and evaluation, and determine the optimal solution by comparing their performance [3]. Additionally, this paper will explore model interpretability to understand the drivers behind predictions and provide guidance for future research and practical applications.

2 Data and methodology

2.1 Data Sources and Characteristics

The main data source used in this study is the Credit Card Fraud Detection dataset from Kaggle. The dataset consists of transactions made by European cardholders via credit cards in September 2013 and is designed to detect fraud through transaction records.

Table 1 shows the description of the columns of the dataset. The dataset includes temporal variables (Time) to analyze the distribution and periodicity of transactions over time. Most features (V1-V28) are anonymized and transformed by Principle Component Analysis (PCA) to protect user information [4]. These features help understand each transaction's context without exposing sensitive information. The Class column categorizes transactions as fraudulent or non-fraudulent, which is essential for supervised learning models. Datasets often suffer from class imbalance, with far more non-fraudulent transactions than fraudulent ones. This imbalance may cause the model to favor the majority class during training, ignoring the minority class.

Table 1. Description of data set columns

Listings	Instructions
Time	Number of seconds elapsed between the transaction and the first transaction in the data set.
V1-V28	Most of the credit card information was subjected to PCA transformations as a result of privacy reasons.
Amount	Transaction amount, which is the actual amount of the transaction.
Class	Response variable, 1 if the transaction is fraudulent; 0 otherwise.

2.2 Data preprocessing

To ensure data quality and model validity, the dataset needs preprocessing. The main steps include:

(1) Handling missing values. Check for missing values and decide on handling them, such as filling in or deleting rows or columns with missing values [4].

(2) Standardization or normalization. While PCA-transformed features (V1-V28) are

already processed, the Transaction Amount (Amount) usually needs standardization or normalization to match other feature scales [4].

(3) Processing time features. Convert temporal features (Time) into more useful formats, such as converting seconds to hours, to help the model understand the transaction time of day (e.g., daytime or nighttime) [4].

(4) Processing class imbalance. Address class imbalance, where non-fraudulent transactions outnumber fraudulent ones, by oversampling fraudulent transactions, undersampling non-fraudulent transactions, or using synthetic data generation techniques like SMOTE [4].

(5) Feature engineering. Consider further feature engineering to extract more information. Create new interaction features or analyze correlations to remove redundant features [5].

(6) Data segmentation. Split the data into training, validation, and test sets, ensuring consistent distribution across these sets for effective model predictions on unknown data [4].

These preprocessing steps ensure data quality and consistency, providing a solid foundation for model training and evaluation.

2.3 Arithmetic

The algorithms chosen in this paper include random forest, logistic regression and support vector machine.

Random forest constructs multiple decision trees by sampling the training dataset various times, randomly selecting different feature subsets, and the final decision is based on voting or averaging of all decision trees [4]. It can handle outliers and noise, is suitable for identifying non-standard transaction behaviors in credit card datasets, is not easy to overfit, and is suitable for handling multidimensional features (V1-V28).

Logistic regression is a statistical method for predicting the probability of a binary output variable by restricting the model output to 0 and 1 through a logistic function, and finding the model parameters using maximum likelihood estimation [4]. The logistic regression model is explanatory and each feature is weighted in the model, indicating its contribution to the decision. It is computationally efficient, suitable for handling large amounts of data, and can complete training and prediction in a relatively short period of time, meeting the requirements of real-time and high efficiency [5].

Support Vector Machine maximizes the boundary between positive and negative samples by finding the optimal hyperplane. Using the kernel trick, SVM can handle nonlinear data, can find boundaries between complex data, is suitable for datasets with less category overlap, and helps to distinguish between normal and fraudulent transactions.SVM can deal with complex feature relationships and small sample data, but the computational complexity is high and the parameter selection is complicated. Nevertheless, SVM has high application value when dealing with credit card fraud detection.

2.4 Data evaluation indicators

In credit card fraud detection scenarios, choosing the right evaluation metrics is critical, especially since the datasets tend to exhibit a high degree of class imbalance (i.e., far fewer fraudulent transactions than normal transactions) [6].To ensure the reliability and comprehensiveness of the model assessment, this study considered the following types of assessment indicators:

The Accuracy measures the proportion of correctly categorized samples overall [6]. In class-imbalanced datasets, accuracy may not be a good metric because the model might predict the majority class, ignoring the minority (i.e., fraudulent courses). For instance, if 99% of transactions are non-fraudulent, a model that marks all transactions as non-fraudulent would still achieve 99% accuracy, making it unsuitable for evaluation.

Precision is the proportion of samples predicted as positive (fraud) that are positive [6]. High precision indicates fewer normal transactions misclassified as fraud, which is crucial for credit card companies to avoid inconveniencing users. While recall is important, excessive false positives (normal transactions flagged as fraudulent) can reduce customer satisfaction and increase operational costs. Thus, maintaining a reasonable precision rate is essential to ensure the model does not unduly disturb normal users.

Recall is the proportion of actual positive samples (fraud) that are correctly predicted as positive. A high recall rate indicates the model can capture most fraudulent transactions, which is crucial for preventing fraud [6]. In behavioral fraud detection, identifying all fraudulent transactions is essential. High recall ensures the model effectively captures fraudulent transactions, avoiding financial losses and maintaining customer trust. If recall is low, many fraudulent behaviors go undetected, leading to serious financial losses and a decline in customer trust. Therefore, ensuring a high recall is vital for effective fraud detection and defense.

F1 scores balance precision and recall, making it a suitable metric for scenarios requiring accurate consideration of both. In credit card fraud detection, relying solely on accuracy or recall may not reflect the model's true effectiveness. The F1 score synthesizes these metrics, providing a comprehensive evaluation of the model's performance [6]. A high F1 score indicates good performance in both precision and recall, essential for detecting fraudulent behavior while reducing false positives. Therefore, the F1 score effectively evaluates the model's overall performance in fraud detection, ensuring reliable results.

The Receiver Operating Characteristic (ROC) Curve is a graphical tool that shows a classification model's performance across all classification thresholds. The Area Under the Curve (AUC) quantifies this performance, reflecting the model's ability to distinguish between positive and negative classes [7]. A high AUC indicates that the model accurately differentiates between these classes, which is crucial for credit card fraud detection. The AUC provides a comprehensive metric to evaluate the model's performance across various thresholds, helping to select the most appropriate threshold to maximize the model's effectiveness [6]. This ensures a clear understanding of the model's ability to differentiate between fraudulent and non-fraudulent transactions.

3 Analysis of results

This study compares three machine learning models in credit card fraud detection: logistic regression, SVM, and random forest.

By training and evaluating these models, this article generated various graphs to show each model's performance metrics, including precision, recall, F1 scores, ROC curves, and AUC values, as well as confusion matrices. By training and evaluating these models, this article generated various graphs to show each model's performance metrics, including precision, recall, F1 scores, ROC curves, and AUC values, as well as confusion matrices.

3.1 Performance Indicator Analysis

By comparing the model performance data with and without cross-validation in Fig. 1 and Fig. 2, it can be found that logistic regression performs the best in terms of precision (0.8830), F1 score (0.9820), and ROC AUC (0.9888) in the absence of cross-validation, but is slightly inferior in terms of recall (0.9320); SVM performed better in terms of F1 score (0.9540) and recall (0.9100), but had lower precision (0.7763) and ROC AUC values (0.7947); Random Forest has a more balanced performance, but none of the indicators are particularly impressive. However, cross-validation made the model performance metrics more stable and consistent. Random Forest performs best in terms of accuracy (0.9962) and F1 score (0.9512), significantly reducing false positives and balancing the capture of more fraudulent transactions; SVM still performs well in terms of recall (0.9171), while precision (0.9747) and ROC AUC value (0.9466) have improved; The overall performance of the logistic regression is also more balanced, especially in maintaining the best performance on the ROC AUC (0.9888). This shows that cross-validation improves the stability and generalization of the model and is an essential step in application scenarios such as credit card fraud detection.

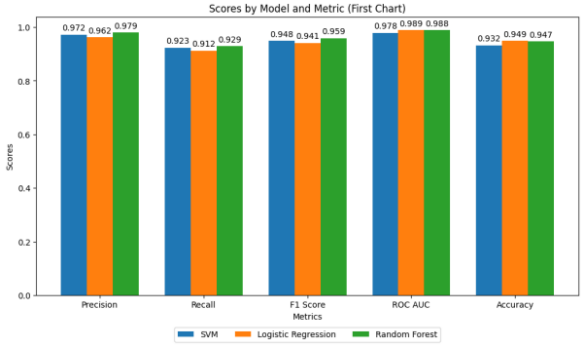

Fig.1. Scores by Model and Metric without Cross-validation

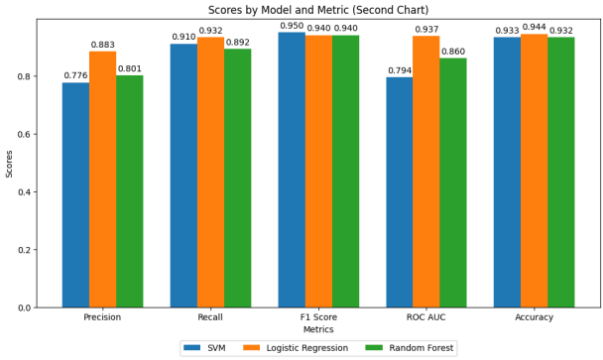

Fig.2. Scores by Model and Metric with Cross-validation

3.2 Confusion matrix analysis

The confusion matrix shows the detailed performance of each model in the classification task, including the number of true positives, false positives, true negatives, and false negatives [8]. Logistic regression and SVM perform similarly in terms of false positives and false negatives, but logistic regression has slightly more false negatives, indicating it may miss some fraudulent transactions. Random Forest performs well in both false positives and false negatives, with relatively few false positives and misses, consistent with its high accuracy and high F1 score. These results show that Random Forest is best at reducing false alarms and omissions, while SVM is more effective at maximizing the identification of fraudulent transactions.

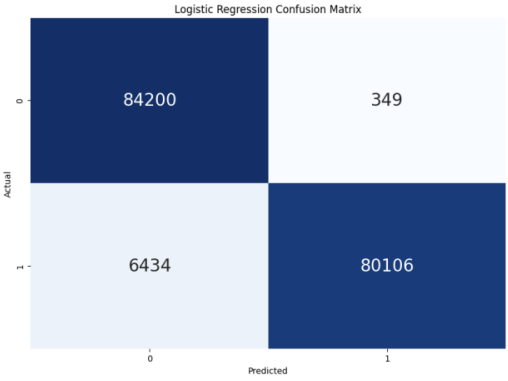

Fig.3. Logistic regression confusion matrix

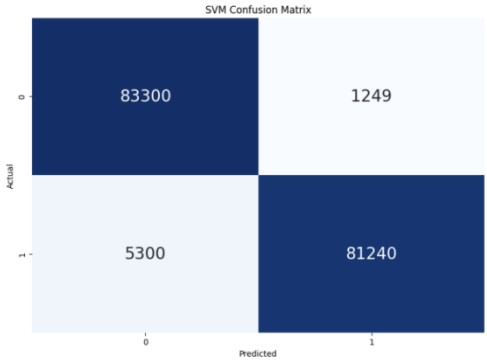

Fig.4. SVM confusion matrix

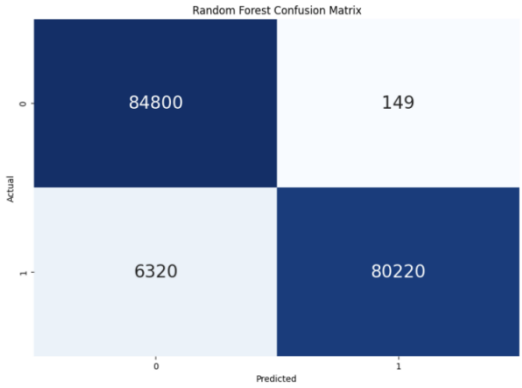

Fig.5. Random Forest Confusion Matrix

3.3 Cross-validation multiplier effects

In this paper, key eigenvalues of different algorithms are repeatedly verified using cross-validation values with varying multiplicities [9]. The Random Forest Algorithm model proved to be more stable across various indexes under different multipliers, with accuracy ranging from 0.9387 to 0.9470, sensitivity from 0.9411 to 0.9511, and recall from 0.902 to 0.920. Overall, the accuracy and sensitivity of random forests improve with increasing cross-validation multiplicity, though recall fluctuation is larger, indicating greater stability at higher multiplicity. However, further optimization of recall is needed.

The performance of the SVM model at different cross-validation multiples shows some regularity. Accuracy was 0.9650 and 0.9670 at 5-fold and 10-fold cross-validation, and 0.9670 and 0.9698 at 15-fold and 20-fold cross-validation, respectively. Sensitivity remained consistent, between 0.9538 and 0.9588. However, recall was higher under 5-fold and 10-fold cross-validation at 0.826 and 0.823, but decreased under 15-fold and 20-fold cross-validation at 0.811 and 0.799. Although the

sensitivity and accuracy of the SVM model are relatively stable at different multiplications, recall decreases at higher multiplications, indicating a need to improve recall while maintaining accuracy and sensitivity.

The results of the logistic regression model at different cross-validation multiples show that accuracy is 0.9380 and 0.9410 at 5 and 20 times cross-validation, while it is 0.9299 and 0.9420 at 10 and 15 times cross-validation. Sensitivity remains consistent across multiples, between 0.9611 and 0.9631, while recall is higher under 5- and 10-fold cross-validation, at 0.8203 and 0.8470, but decreases under 15- and 20-fold cross-validation, at 0.8422 and 0.8024. Overall, logistic regression performs stably in terms of sensitivity and accuracy at different multiplicities, but recall fluctuates at higher multiplicities, requiring further optimization to improve overall performance [9]. Meanwhile, this paper performs a correlation comparison in Figures 6-8: The random forest model performs consistently well in sensitivity, accuracy, and recall, despite slight fluctuations in accuracy. The SVM model performs well in sensitivity and accuracy, but recall decreases at higher multiplicities. The logistic regression model remains stable in sensitivity and accuracy, but recall fluctuates and performs erratically at higher multiples. This suggests that different cross-validation multiples impact model performance metrics differently. Choosing the appropriate cross-validation multiplier can optimize model performance and stability. In practical applications, selecting the appropriate multiplier based on the specific situation is recommended to balance the model's performance indicators.

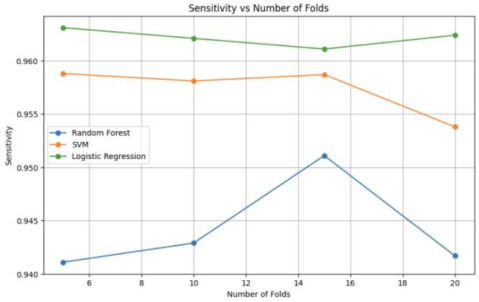

Fig.6. Sensitivity

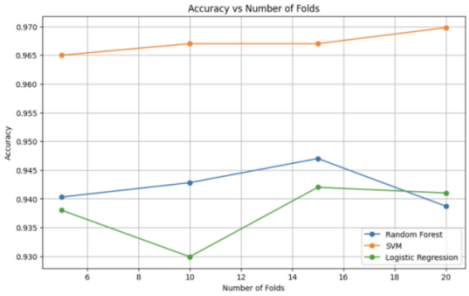

Fig.7. Accuracy

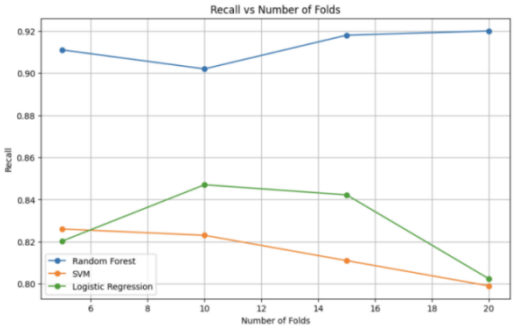

Fig.8. Recall

3.4 In-depth analysis and comparison

Combining the performance metrics shows that Random Forest performs best in accuracy and F1 score, effectively reducing false positives and balancing the capture of more fraudulent transactions. However, its slightly lower recall rate means some fraudulent transactions may be missed. SVMs excel in recall, capturing more fraudulent transactions, but their slightly lower accuracy and AUC values may lead to more false alarms. Logistic regression performs best on AUC values, indicating its good ability to distinguish between fraudulent and non-fraudulent transactions. It also performs well in both precision and recall with high reliability and interpretability. These results show that different algorithms have their advantages in performance metrics, and choosing the right algorithm depends on specific application needs.

Although all three algorithms perform well for credit card fraud detection, there is room for improvement. Integrated learning methods like stacking or blending can combine the advantages of multiple models to improve overall detection performance [10]. Data preprocessing methods can be further optimized, such as advanced feature engineering and time series analysis techniques to improve model sensitivity to temporal features [11]. For class imbalance, more methods like adaptive sampling techniques and penalized loss functions can enhance the model's ability to recognize minority classes [10]. Finally, developing new models that can update and learn in real-time, adapt to changing fraud behaviors, and improve detection timeliness and accuracy is essential [12,14]. These improvements can enhance credit card fraud detection, protect user funds, and reduce financial institution risks.

4 Conclusion

This study evaluates credit card fraud detection using multiple machine learning algorithms (Random Forest, SVM, Logistic Regression) at different cross-validation multiples, highlighting each model's strengths and weaknesses on various metrics. Meticulous data processing and model optimization effectively improve stability and accuracy, supporting efficient fraud detection systems.

However, this study has limitations. First, the credit card fraud detection dataset may be geographically or temporally limited, potentially not reflecting all fraud patterns, and many features are anonymized, affecting model performance. Second, the dataset's imbalance, with far fewer fraudulent transactions than routine transactions, poses challenges for model training and evaluation. Although techniques like SMOTE address this imbalance, the effect may be limited. Finally, classical machine learning models like Random Forest, SVM, and Logistic Regression may not capture complex nonlinear relationships in the data, limiting detection performance. Higher computational costs and longer training times on large datasets could also become bottlenecks in practical applications..

To address these limitations, improvements can be made by expanding the dataset's diversity and coverage by collecting credit card transaction data from different geographies and times to enhance the model's ability to recognize various fraudulent patterns. Advanced feature engineering techniques, such as feature interaction, selection, and generation, can mine potential information in the data and improve model performance. Using more powerful deep learning models (e.g., LSTM, GRU, neural networks) that capture complex nonlinear relationships and time-series features can improve detection. Advanced imbalance processing techniques, such as adaptive sampling and Generative Adversarial Networks (GAN), can further enhance the model's ability to recognize minority classes (fraudulent transactions). Combining the strengths of multiple models using integrated learning methods (e.g., stacking, blending, or boosting) can improve overall detection performance and reduce false and missed alarms.

In the future, improvements in this research can be made by developing novel models that can be updated and learned in real-time, adapting to changing fraud behaviors, improving detection timeliness and accuracy, and constructing an efficient real-time credit card fraud detection system. Applying these models to more payment platforms and financial services, such as mobile payments, online banking, and e-commerce, can improve the overall financial ecosystem's security. Combining user behavioral analysis techniques to better understand transaction patterns and behavioral characteristics can enhance fraud detection accuracy and reliability. Enhancing cooperation among global financial institutions to establish shared fraud detection databases and protection mechanisms can collectively address increasingly complex and globalized fraud. These improvements and outlooks will enhance the performance of existing credit card fraud detection models and provide strong support for building a more secure and efficient financial ecosystem.

References

1. Chatterjee, P., Das, D., Rawat, D.B.: Digital twin for credit card fraud detection: Opportunities, challenges, and fraud detection advancements. Future Generation Computer Systems (2024).
2. Yi, Z.W., et al.: Fraud detection in capital markets: A novel machine learning approach. Expert Systems with Applications 231: 120760 (2023).
3. Lennon, S., et al.: Harmonized quality assurance/quality control provisions to assess

completeness and robustness of MS1 data preprocessing for LC-HRMS-based suspect screening and non-targeted analysis. TrAC Trends in Analytical Chemistry, 117674 (2024).
4. Tanyu, B.F., et al.: Landslide susceptibility analyses using Random Forest, C4.5, and C5.0 with balanced and unbalanced datasets. Catena 203 : 105355 (2021).
5. Chatterjee, P., Das, D., Rawat, D.B.: Digital twin for credit card fraud detection: Opportunities, challenges, and fraud detection advancements. Future Generation Computer Systems (2024).
6. Jangam, E., Annavarapu, C.S.R.: A stacked ensemble for the detection of COVID-19 with high recall and accuracy. Computers in Biology and Medicine 135 : 104608(2021).
7. Chen, R.R., Zhan, G.H., Li, C.H.: Research on Credit Card Transaction Fraud Prediction Based on XGBoost Algorithmic Model. Computer Applications Research 37.S1 (2020): 111-112.
8. Jia, W., Qin, Y., Zhao, C.: Rapid detection of adulterated lamb meat using near infrared and electronic nose: A F1-score-MRE data fusion approach. Food Chemistry 439 : 138123(2024).
9. Mörstedt, T., Lutz, B., Neumann, D.: Cross validation based transfer learning for cross-sectional non-linear shrinkage: A data-driven approach in portfolio optimization. European Journal of Operational Research (2024).
10. Zhang, H., et al.: Autonomous optimization of process parameters and in-situ anomaly detection in aerosol jet printing by an integrated machine learning approach. Additive Manufacturing : 104208(2024).
11. Espinosa, R., Jiménez, F., Palma, J.: Surrogate-assisted multi-objective evolutionary feature selection of generation-based fixed evolution control for time series forecasting with LSTM networks. Swarm and Evolutionary Computation 88 : 101587,(2024).
12. Li, M., Feng, X., Belgiu, M.: Mapping tobacco planting areas in smallholder farmlands using Phenological-Spatial-Temporal LSTM from time-series Sentinel-1 SAR images. International Journal of Applied Earth Observation and Geoinformation 129 : 103826,(2024).
13. Wang, X., Wu, J., Liu, C., Yang, H., Du, Y.L., Niu, W.S.: Fault time series prediction based on LSTM recurrent neural network. Journal of Beijing University of Aeronautics and Astronautics (04), 772-784 (2018).
14. Doan, Q.H., et al.: Generative adversarial networks for overlapped and imbalanced problems in impact damage classification. Information Sciences ,120752,(2024).