Proceedings of ICFTBA 2024 Workshop: Finance in the Age of Environmental Risks and Sustainability
DOI: 10.54254/2754-1169/127/2024.OX18537

A Comparative Analysis of Machine Learning Models for Credit Card Fraud Detection

Wanru Zou[1,a,*]

[1]*International School, Beijing University of Posts and Telecommunications, Beijing, China*
a. Veronica_z@bupt.edu.cn
**corresponding author*

Abstract: The rapid growth of online transactions has greatly increased the autonomy of credit cards, creating serious challenges for financial institutions and consumers.Credit card fraud detection involves identifying unauthorized transactions conducted using stolen or fake credit card information. This study aims to explore the most influential features contributing to credit card fraud and evaluate the effectiveness of various machine learning models in predicting fraudulent transactions. Utilizing the "creditcard.csv" dataset, which contains real-world credit card transactions, we conducted feature selection and model comparison to enhance detection accuracy. The results demonstrate that Neural Networks and Support Vector Machines (SVM) are the most effective models, achieving high Matthews Correlation Coefficient (MCC) scores due to their ability to handle high-dimensional data and complex nonlinear relationships. In contrast, simpler models like Naive Bayes and Random Tree exhibited lower performance but can be improved through advanced techniques such as feature selection and data balancing. These findings highlight the importance of robust feature engineering and careful model selection in developing accurate and reliable fraud detection systems. Financial institutions can drastically improve their fraud detection capabilities by putting these innovative approaches into practice.

Keywords: Credit card fraud, Machine learning, Neural Networks, Support Vector Machines, Fraud detection systems.

1. Introduction

The consequences of such fraudulent activities are severe, leading to financial losses for both consumers and banks, decreased consumer trust, and increased costs for fraud prevention measures. Given the ever-evolving tactics of fraudsters, there is an urgent need for robust and accurate methods to detect and prevent fraudulent transactions in real-time. Traditional rule-based systems, while effective to some extent, often struggle to adapt to new fraud patterns and can generate a high rate of false positives. Machine learning and data mining techniques offer a promising alternative, as they can automatically learn and adapt to complex patterns in transaction data. By identifying key features and comparing model accuracies, this research seeks to provide insights into the best practices for developing robust fraud detection systems. Specifically, this study addresses the following questions: What are the key features in the transaction data that have a high impact on the likelihood of fraud? How do diverse models perform in predicting credit card fraud and which model is the most effective?

DOI: 10.54254/2754-1169/127/2024.OX18537

By providing answers to these queries, the research hopes to aid in the advancement of fraud detection mechanism that are more precise and useful.

2. Literature Review

The recognition of crimes by taking advantage of credit cards has become an important area of research, with many studies exploring different means, especially in machine learning, to increase awareness and prevent fraudulent activities. This section summarizes the findings of recent articles that provide insight into the use and consequences of various methods of credit card authentication.

2.1. Overview of Literatures

A thorough investigation into the development and application of various models for credit card default detection was given by Singh. [1]. Their study examined a number of models, emphasizing the value of feature selection and data preprocessing to get higher predictive accuracy.With a 99.87 percent accuracy rate, the Catboost algorithm is shown to be the most effective at detecting credit card fraud. The credit card fraud detection dataset was obtained via Kaggle.

Kaul et al. examined both machine learning (ML) and deep learning (DL) techniques in actual scenario [2]. The credit card dataset was analyzed using a variety of techniques, including 8 models. The project is run in Python, and the methods are applied to the revised dataset. Accuracy and f1-score are the cores to assess the methods. Along with other applied techniques, the result shows the highest level of precision for random forest classification (RFC), and the comparative analysis demonstrates that RFC outperforms other methods used.

To identify and evaluate fraud in online transactions, Kumar et al. started utilizing a variety of machine learning techniques [3]. The major objective is to create a novel approach to fraud detection for streaming transaction data by identifying patterns in historical transactions through analysis. Customers are divided into different groups according to their income. The window shows how you can coordinate activities in these groups to improve their behavior. Different classifiers are then trained for each group and the best performing classes are selected to predict fraud. The feedback method is used to solve the drift problem.

Khan et al. emphasizes the importance for credit card organizations to identify fraudulent transactions to prevent customers from being charged for purchases they didn't make [4]. This problem can be solved with data science and machine learning which are important for this. The project aims to demonstrate the use of statistics in examining credit cards. It involves analyzing previous credit card transactions to identify patterns of fraud. The model is then used to determine whether new activity is fraudulent. The goal is to 100% detect fraud while limiting benefits. The work focuses on analyzing and pre-processing data and using various anomaly detection algorithms such as Random Forest and KNN.

2.2. Discussion on Outcomes

The reviewed literature highlights several critical aspects of credit card fraud detection. Singh et al. demonstrated the effectiveness of the Catboost algorithm with an accuracy of 99.87%, while Kaul et al. found that Random Forest Classification (RFC) performed best, showing the importance of selecting robust algorithms. Both studies underscored the significance of feature selection and data preprocessing in enhancing model accuracy. The third research introduced a new method of collecting cardholders based on the quantity of activities and using sliding window strategies to analyze the behaviors and work out the kinks through the feedback system. Khan et al. focused on achieving a balance between detecting fraudulent transactions and minimizing false positives, using anomaly detection algorithms like Random Forest and KNN. In summary, robust feature engineering, careful

DOI: 10.54254/2754-1169/127/2024.OX18537

algorithm selection, and the use of advanced techniques are essential for developing accurate and reliable credit card fraud detection systems.

3. Methodology

3.1. Dataset Description

The dataset applied for testing is named of "creditcard.csv" dataset, which contains real-world credit card transactions collected. It comprises 284,807 transactions with 492 instances of fraud, making it highly imbalanced. Each transaction is represented by 30 features, including the transaction amount and 28 anonymized features from a Principal Component Analysis (PCA) transformation.

Time: The time elapsed between the first transaction in the dataset and each subsequent transaction (in seconds).

V1-V28: Principal components obtained from the PCA transformation which ensure the confidentiality of the original feature values.

Amount: The transaction amount which can be used as an input for predictive models.

Class: The sign where '0' indicates a legitimate transaction and '1' stands for a fraudulent one.

The highly imbalanced nature of the dataset poses a significant challenge for model training as the minority class (fraudulent transactions) represents only 0.172% of the total transactions. Effective handling of this imbalance is crucial for developing accurate and reliable fraud detection models.

3.2. Feature Distribution

3.2.1. Amount Distribution

The distribution of amounts varies significantly between the two data sets. This feature cannot be deleted. The amounts involved in credit card fraud are generally smaller compared to those of normal credit card users. This indicates that fraudsters tend to choose smaller amounts to avoid drawing the attention of the credit card holder (see Figure 1).

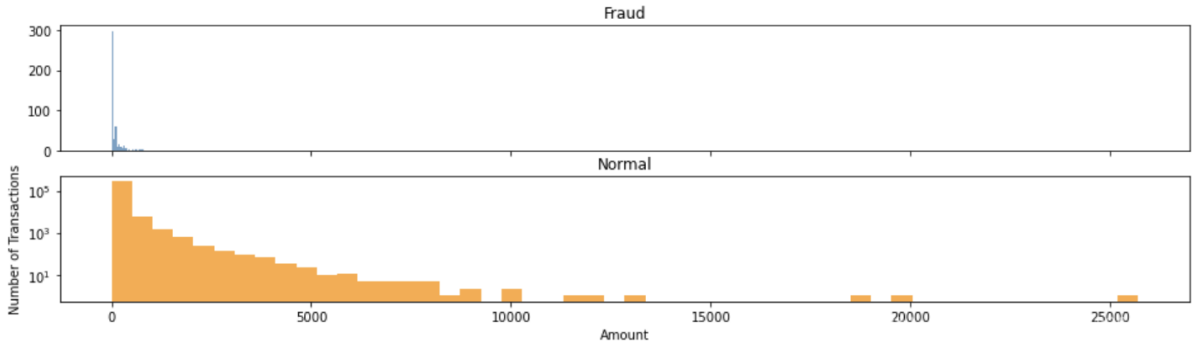

Figure 1: Distribution of Transaction Amounts for Fraudulent and Non-Fraudulent Transactions

3.2.2. Time Distribution

It can be observed that the distribution of fraudulent transactions is more dispersed, showing two peak indications. In contrast, the distribution of normal transactions is more concentrated, with higher transaction frequency observed during two specific time periods (see Figure 2).

DOI: 10.54254/2754-1169/127/2024.OX18537

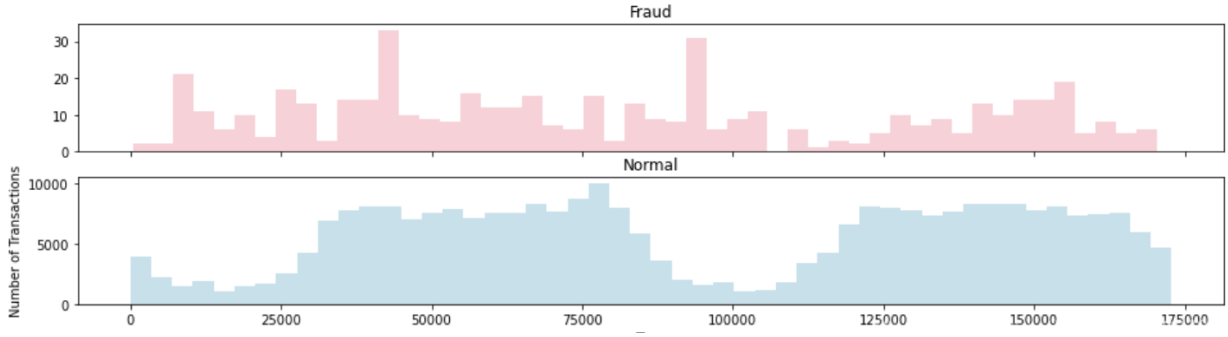

Figure 2: Time Distribution of Transactions for Fraudulent and Non-Fraudulent Transactions

3.3. Feature Importance Analysis

3.3.1. Correlation Analysis

This is kind of analysis is in order to investigate the relavance between numerical features and the classes [5]. Features with high correlation with the target were selected for modeling. In cases of credit card fraud, a more pronounced correlation is observed between certain variables. Changes in variables V1-V19 exhibit certain patterns in the samples of fraudulent credit card transactions. These features are derived from a Principal Component Analysis (PCA) transformation, and their specific meanings are anonymized to maintain data confidentiality (see Figure 3).

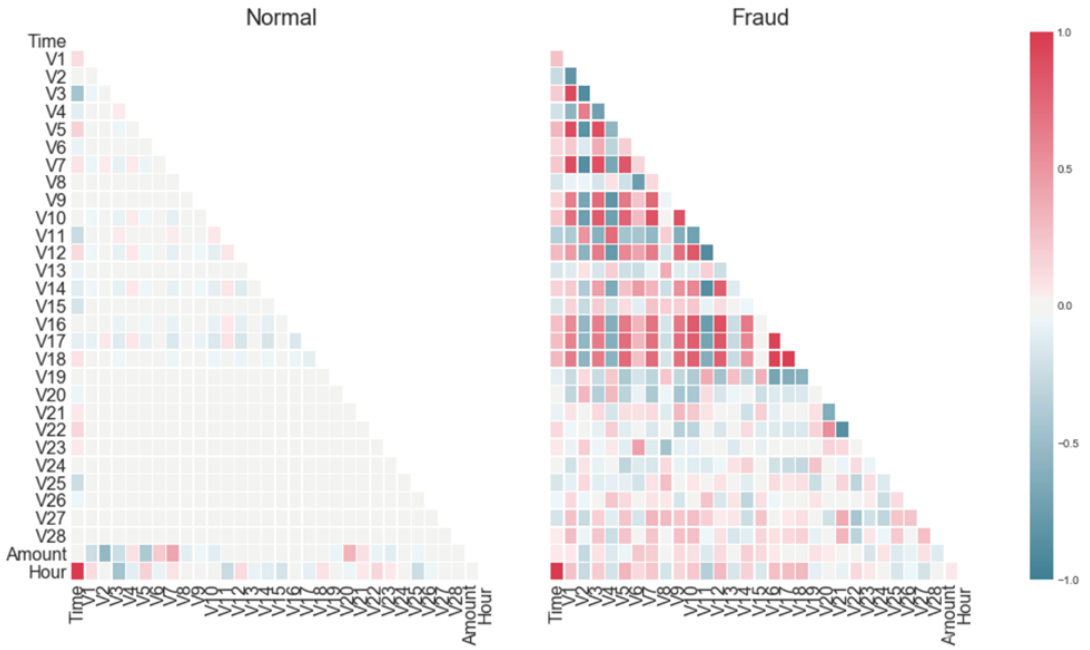

Figure 3: Correlation Matrix of Features and Target Variable

3.3.2. Importance Analysis

Feature importance scores from a Random Forest model were used to identify the most impactful features. The analysis reveals that certain PCA-derived components have significant influence on fraud detection, emphasizing the importance of these features in the predictive modeling process (see Figure 4).

DOI: 10.54254/2754-1169/127/2024.OX18537

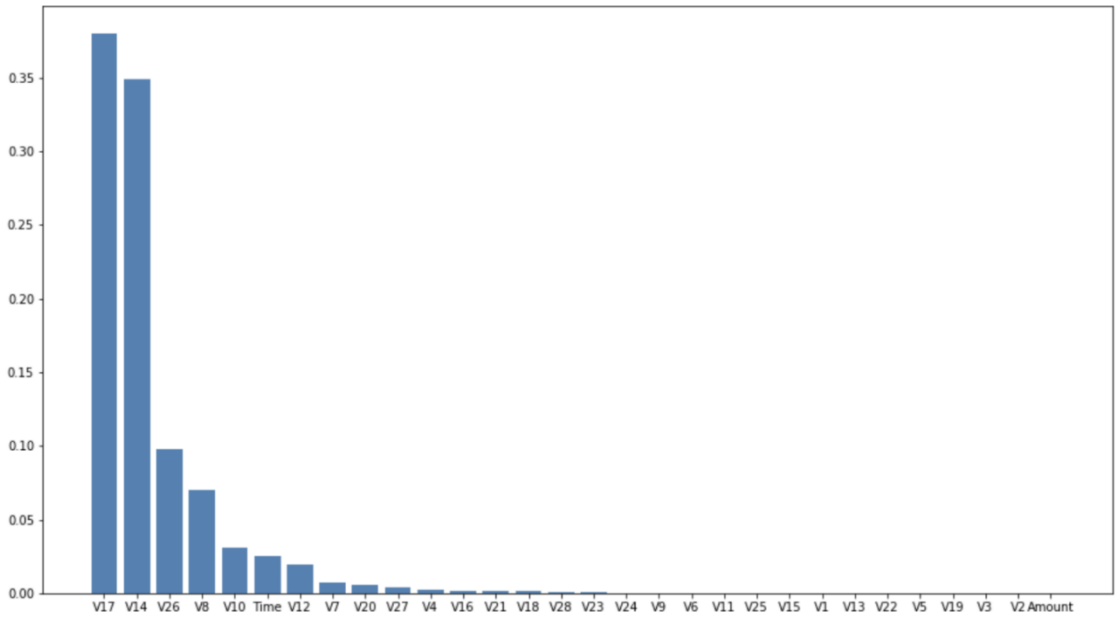

Figure 4: Importance Scores from Random Forest Model

3.4. Model Training

After the basic data cleaning and feature processing steps, this paper focuses on training 12 models, all of which are commonly used methods in machine learning. Additionally, these 12 models will be ranked and classified based on performance according to certain accuracy metrics. The comparative analysis of the models is conducted from multiple dimensions, and their performance may vary due to the different characteristics of the dataset [5].

4. Model Comparison

In this section, several machine learning models for the selected dataset are compared. The goal is to identify the most effective models by evaluating their performance on a highly imbalanced dataset. Multiple algorithms are analyzed, their strengths and weaknesses are assessed, and the best-performing models are selected based on specific metrics.

4.1. Performance Metrics

The distinction of each model is assessed according to two values including accuracy for detecting fraudulent (Frauds) and genuine (Genuines) transactions, and the Matthews Correlation Coefficient (MCC), which is an evaluation metric used to assess the performance of classification models. It ican provide comprehensive information about the quality of the results [6].

Table 1: This caption has one line so it is centered.

Model	Accuracy (Frauds)	Accuracy (Genuines)	MCC
Navie Bayes	83.130	97.730	0.219
Decision Tree	81.098	99.951	0.775
Random Forest	42.683	99.988	0.604
Gradient Boosted Tree	81.098	99.936	0.746
Decision Stump	66.870	99.963	0.711

DOI: 10.54254/2754-1169/127/2024.OX18537

Table 1: (continued).

Random Tree	32.520	99.982	0.497
Deep Learning	81.504	99.956	0.787
Neural Network	82.317	99.966	0.812
Multi Layer Perceptron	80.894	99.966	0.806
Linear Regression	54.065	99.985	0.683
Logistic Regression	79.065	99.962	0.786
Support Vector Machine	79.878	99.972	0.813

4.2. Method Classification and Characteristics

Based on Table 1 above, from the perspective of comparing MCC values, there are two groups of models identified: those with MCC values greater than 0.8 and those less than 0.6. These correspond to groups that exhibit superior and inferior performance in detecting this dataset, respectively.

4.2.1. Superior Models

Neural network is a computational method that represents a natural, organic network that processes complex information through multiple layers of neurons (input, hidden layers, output).
Support Vector Machine is a classification algorithm that seeks for the optimal separating hyperplane in a high-dimensional space to achieve classification, suitable for both linear and non-linear separable data.

4.2.2. Inferior Models

The assumption of Naive Bayes that attributes are independent of each other is often not valid in practical applications, which has a certain impact [7].

Random Tree constructs multiple independent decision trees and performs majority voting for classification.

4.3. Performance on the Dataset

4.3.1. Dataset Characteristics

The dataset is extremely imbalanced with fraudulent transactions accounting for a very small proportion, while normal transactions are the majority. This imbalance poses challenges for model training and performance evaluation. The dataset contains nearly 30 features with complex nonlinear relationships and over 28,000 transaction records, requiring handling of large amounts of data.

4.3.2. Performance Differences

Neural Network and Support Vector Machine: Effective in handling high-dimensional features and imbalanced data by adjusting loss functions and class weights. Capable of processing complex nonlinear relationships, improving classification accuracy and generalization ability.

Naive Bayes and Random Tree: Suitable for quick preliminary analysis but perform poorly on high-dimensional and imbalanced datasets. Naive Bayes assumes feature independence, leading to poor performance on high-dimensional datasets. Random Tree is prone to overfitting, even with ensemble methods, resulting in poor generalization ability.

DOI: 10.54254/2754-1169/127/2024.OX18537

4.4. Optimization of Poorly Performing Methods

4.4.1. Naive Bayes

(1) Feature Selection and Engineering

Remove irrelevant features: Eliminate irrelevant or noisy features to reduce dimensionality.

Feature combination: Create new combined features to capture interactions between features.

(2) Balancing the Dataset

Upsampling the minority class: Use techniques such as SMOTE to increase fraudulent transaction samples [8].

Downsampling the majority class: Reduce the number of normal transaction samples to balance the dataset.

(3) Adjusting Decision Threshold

Adjust the classification decision threshold based on the model output probabilities to improve minority class detection.

(4) Ensemble Methods

Bagging: Train multiple Naïve Bayes models by random sampling of the dataset, then vote on the results [9].

Boosting: Use techniques such as AdaBoost to assign higher weights to misclassified samples and iteratively train multiple models.

4.4.2. Random Tree

(1) Feature Selection and Engineering

Feature importance: Select the most influential features based on their importance.

Feature standardization: Standardize or normalize features to reduce differences in feature scales.

(2) Balancing the Dataset

Generating synthetic samples: Use algorithms such as ADASYN (Adaptive Synthetic Sampling) to generate more minority class samples [10].

(3) Adjusting Model Parameters

Tree depth: Limit the maximum depth of the tree to prevent overfitting.

Minimum samples for splitting: Increase the minimum number of samples required to split a node to reduce overfitting.

(4) Ensemble Methods

Random Forest: Improve model performance by constructing multiple random trees and voting on the results.

Boosting: Use techniques such as Gradient Boosting to iteratively train multiple tree models, focusing on the errors of the previous model each time.

(5) Cross-Validation

Use k-fold cross-validation to select the best parameters and evaluate model performance [11].

5. Conclusion

This study successfully identified the key features influencing credit card fraud and evaluated the effectiveness of various machine learning models. The results demonstrate that Neural Networks and Support Vector Machines (SVM) are the most effective models, achieving high MCC scores thanks to their natural advantage in dealing with high-dimensional data and complex nonlinear relationships. In contrast, simpler models like Naive Bayes and Random Tree exhibited lower performance but can be improved through advanced techniques such as feature selection and data balancing. These findings highlight the importance of robust feature engineering and careful model selection in

DOI: 10.54254/2754-1169/127/2024.OX18537

developing accurate and reliable fraud detection systems. With this modern system implemented, it provides more possibilities for revelant institusions to detect frauds, which does help to solve problems in payment machinery.

References

[1] *Singh, A., Singh, A., Aggarwal, A., & Chauhan, A. (2022, November). Design and Implementation of Different Machine Learning Algorithms for Credit Card Fraud Detection. In 2022 International Conference on Electrical, Computer, Communications and Mechatronics Engineering (ICECCME). IEEE, 1-6.*

[2] *Kaul, A., Chahabra, M., Sachdeva, P., Jain, R., & Nagrath, P. (2021). Credit Card Fraud Detection Using Different ML and DL Techniques. In Proceedings of the International Conference on Innovative Computing & Communication (ICICC). Available at SSRN: https://ssrn.com/abstract=3747486 or http://dx.doi.org/10.2139/ssrn.3747486*

[3] *Kumar, V., Kumar, V., Vijayshankar, A., & Kumar, P. (2020). Credit Card Fraud Detection using Machine Learning Algorithms. Inter-national Journal of Engineering Research & Technology, 9. 7.*

[4] *Khan, S., Sanovar, S., Kumar, S., and Kumar, H. (2021) Credit Card Fraud Detection Using Machine Learning. International Journal of Scien-tific and Research Publications (IJSRP), 11(6), 2250-3153.*

[5] *Li, Z. (2024). A Feature Engineering Strategy for Data-Driven Predictive Models of Educational Building's Electricity Consumption. In 2024 7th International Conference on Advanced Algorithms and Control Engineering (ICAACE). IEEE, 542-546.*

[6] *Boughorbel S, Jarray F, El-Anbari M (2017) Optimal classifier for imbalanced data using Matthews Correlation Coefficient metric. PLOS ONE 12(6): e0177678.*

[7] *Wickramasinghe, I., Kalutarage, H. (2021) Naive Bayes: applications, variations and vulnerabilities: a review of literature with code snippets for implementation. Soft Comput, 25, 2277–2293.*

[8] *Awoyemi, J. O., Adetunmbi, A. O., & Oluwadare, S. A. (2017). Credit card fraud detection using machine learning techniques: A comparative analysis. In 2017 international conference on computing networking and informatics (ICCNI). IEEE, 1-9.*

[9] *Jafarzadeh, H., Mahdianpari, M., Gill, E., Mohammadimanesh, F., & Homayouni, S. (2021). Bagging and boosting ensemble classifiers for classification of multispectral, hyperspectral and PolSAR data: a comparative evaluation. Remote Sensing, 13(21), 4405.*

[10] *Munshi, R. M. (2024). Novel ensemble learning approach with SVM-imputed ADASYN features for enhanced cervical cancer prediction. PLoS One, 19(1), e0296107.*

[11] *Montesinos López, O. A., Montesinos López, A., & Crossa, J. (2022). Overfitting, model tuning, and evaluation of prediction performance. In Multivariate statistical machine learning methods for genomic prediction, 109-139.*