Multiple Machine Learning Algorithms-based Credit Card Fraud Detection

Manho Yeung *

Statistics and Data Science, University of California Santa Barbara, Santa Barbara, USA

* Corresponding Author Email: myeung@ucsb.edu

Abstract. In recent decades, the rise of E-commerce has made credit cards a popular choice for transactions due to their convenience. However, this has also led to an increase in credit card fraud, posing a significant financial threat. To tackle this issue, this study investigates the use of machine learning models for credit card fraud detection. The research explores three models: Gradient Boosting, Decision Trees, and K-Nearest Neighbors (KNN), assessing their performance using metrics like Area under the Precision-Recall Curve (AUPRC), precision, recall, and F1-score. The study employs the Synthetic Minority Oversampling Technique (SMOTE) algorithm to address the dataset's class imbalance, creating a balanced training dataset. Among the models, Gradient Boosting achieves the highest AUPRC of 0.81, indicating its effectiveness in general scenarios. However, its precision in detecting fraudulent transactions remains a challenge. The Decision Tree model provides a balanced performance with an AUPRC of 0.59, while KNN stands out with a higher precision for fraudulent transactions, achieving an AUPRC of 0.77. These findings highlight the need for continuous improvement in fraud detection techniques, suggesting that hybrid models or ensemble approaches could enhance detection capabilities by leveraging the strengths of different models.

Keywords: Machine learning, Artificial intelligence, Credit card fraud detection.

1. Introduction

In recent decades, E-commerce has become extremely prevalent and widespread in the world. Among all methods of E-commerce, credit card is one of the most used approaches due to its obvious convenience. There is a trend that people are used to purchasing goods by using credit cards instead of paying with cash. With this trend, credit card frauds are becoming increasingly common in people's daily lives, which is a type of financial fraud [1, 2], referring to utilizing unauthorized credit card for the aim of illegally financial possession, such as goods, services, or money [3]. As e-commerce and online payment keeps expanding, credit card frauds will be more prevalent, which is a critical financial problem [4, 5]. Accordingly, with credit card fraud detection based on machine learning, it is feasible to determine whether a transaction is fraudulent or not.

In the previous decades, there were various researches that endeavor to propose methods that can do credit card fraud detection. For instance, Jain et al. Have experimented with several techniques to detect credit card fraud such as Support Vector Machines (SVM), Artificial Neural Networks (ANN), Bayesian Networks, logistic regression, fuzzy logic-based systems and more [6]. In their research, they have found that the worst performance is provided by SVM and fuzzy logic, while ANN, bayesian networks give the highest accuracy. Moreover, SVM, fuzzy logic and logistic regression also provide lowest detection rate, while bayesian networks give the highest ones. These methods do an excellent job, but challenges remain, especially because of more advanced tactics of fraudsters. These models often fail to accommodate the new fraudulent methods. Therefore, new methods and continuous improvement in fraud detection is required.

This research set out to elevate the efficiency of credit card fraud detection mechanisms by deploying and scrutinizing three machine learning models: Gradient Boosting, K-Nearest Neighbors (KNN), and Decision Trees. Each model's efficacy is assessed through metrics including Area under the Precision-Recall Curve (AUPRC), precision, recall, and F1-score.

The study embarks with meticulous data preprocessing, tackling missing values and normalizing features. To combat the prevalent issue of dataset imbalance, this study employs the Synthetic

Minority Over-sampling Technique (SMOTE) algorithm, which concocts synthetic instances of the minority class, thereby ensuring an evenly distributed dataset. Subsequently, each model undergoes training on this resampled dataset, with hyperparameter tuning conducted to finetune their performance. The models are then appraised based on their prowess in accurately flagging fraudulent transactions.

This paper delves deeper into the work carried out from two perspectives: (i) the readily available methods for fraud detection, and (ii) the techniques available to manage imbalanced data. As mentioned above, SMOTE is used in this research to avoid imbalance in data. Furthermore, some of the prominent Machine Learning algorithms are used for credit fraud detection, including decision trees, gradient boosting, and K-nearest neighbor.

2. Method

2.1. Dataset preparation

The Credit Card Fraud Detection dataset is a compelling compilation of anonymized credit card transactions conducted in September 2013 by European cardholders, which can be downloaded from Kaggle [7]. It encapsulates transactions spanning 2 days, with a notable yet sparse occurrence of 492 fraudulent transactions out of a colossal 284,807 total transactions, thus presenting a staggering imbalance where fraudulent instances constitute a mere 0.172% of the dataset. This dataset is intricately detailed with 30 variables, including 'Time', which meticulously logs the seconds elapsed between each transaction and the inaugural transaction within the dataset, and 'Amount', delineating the monetary value of each transaction. The other 28 variables, V1 through V28, are principal components meticulously extracted via a PCA transformation, ensuring the confidentiality and privacy of the cardholders' sensitive information. The target variable, 'Class', is a binary indicator of transaction legitimacy, flagging fraudulent transactions with a 1 and genuine transactions with a 0. The dataset's profound class imbalance necessitates the use of the AUPRC for accuracy measurement, rather than the conventional confusion matrix accuracy. This dataset, amassed through a synergistic research collaboration between Worldline and the Machine Learning Group of Université Libre de Bruxelles (ULB), stands as a paramount resource for the advancement and rigorous evaluation of credit card fraud detection algorithms, offering deep insights into the complexities of fraudulent transaction patterns.

Given the profound class imbalance inherent in the credit card fraud detection dataset, conventional accuracy metrics fall short in providing meaningful evaluations. Consequently, this study utilized AUPRC as a more robust measure of performance. To counteract this imbalance, this study employed SMOTE during preprocessing, a method that generates synthetic samples to bolster the minority class, thereby enhancing the dataset's balance and the robustness of subsequent machine learning models.

2.2. Machine learning-based prediction

In order to construct efficacious models for detecting fraudulent credit card transactions, this study meticulously deployed three sophisticated machine learning algorithms: Gradient Boosting, Decision Tree, and K-Nearest Neighbors (KNN) [8-10]. Each model was carefully crafted and trained to unearth fraudulent activities within the heavily skewed dataset, preprocessed using SMOTE to optimize performance and reliability. Gradient Boosting stands as a paragon of ensemble learning techniques, adept at forging a formidable predictive model through the amalgamation of numerous weak learners, typically decision trees. This algorithm operates sequentially, with each new model endeavoring to rectify the residual errors of the combined ensemble, thereby systematically diminishing bias and variance. The implementation leveraged the GradientBoostingClassifier from the sklearn. Ensemble module. Hyperparameters, such as the number of estimators and learning rate, were meticulously fine-tuned to achieve optimal performance.

The process began with initializing and training the Gradient Boosting model on the resampled training set. Post-training, predictions were made on the original test set, and the model's performance was evaluated using AUPRC. The results revealed the model's adeptness at identifying fraudulent transactions, reflected in the AUPRC score. The confusion matrix and classification report provided further insights into the model's precision and recall, which were subsequently visualized through heatmaps and precision-recall curves, underscoring the model's efficacy.

The Decision Tree model, a stalwart in machine learning, constructs a hierarchical, flowchart-like structure where internal nodes denote feature comparisons, branches symbolize decision rules, and leaf nodes represent outcomes. Despite its simplicity, this model is potent, capable of capturing intricate decision boundaries. The implementation of the study employed the DecisionTreeClassifier from the sklearn. Tree module. This study meticulously experimented with parameters such as maximum depth, minimum samples split, and minimum samples leaf to strike a balance between overfitting and generalization.

Training the Decision Tree model followed a similar trajectory: the model was trained on the SMOTE-augmented dataset, predictions were made on the original test set, and the model's performance was scrutinized using AUPRC. The resulting confusion matrix and classification report were analyzed and visualized, revealing the model's strengths and areas for improvement. The precision-recall curve further highlighted the model's proficiency in distinguishing between fraudulent and non-fraudulent transactions.

K-Nearest Neighbors epitomizes simplicity and power in non-parametric classification methods. The algorithm identifies the 'k' training samples nearest to a test sample and predicts the class label that is most prevalent among these 'k' neighbors. For the implementation, this paper utilized the KNeighborsClassifier from the sklearn. Neighbors module, setting the number of neighbors to five. This model was trained on the resampled dataset to enhance its capability in detecting fraudulent transactions.

The KNN model followed the established training and evaluation protocol: it was trained on the SMOTE-adjusted dataset, predictions were made on the original test set, and performance was evaluated using AUPRC. The confusion matrix and classification report provided a detailed account of the model's predictive accuracy and error rates. Visualization through heatmaps and precision-recall curves underscored the model's effectiveness and pinpointed potential areas for refinement.

3. Results and Discussions

The performance of different models was provided in Table 1. They are evaluated by four different metrics including AUPRC, precision, recall and F1-score.

Table 1. Performance of different models.

Metrics	Gradient Boosting	Decision Tree	K-Neareast Neighbour
AUPRC	0.8089	0.5941	0.7697
Precision	1.00(0) / 0.12(1)	1.00(0) / 0.43(1)	1.00(0) / 0.48(1)
Recall	0.99(0) / 0.91(1)	1.00(0) / 0.76(1)	1.00(0) / 0.87(1)
F1-Score	0.99(0) / 0.22(1)	1.00(0) / 0.55(1)	1.00(0) / 0.62(1)

3.1. Gradient boosting

The Gradient Boosting model showcased formidable prowess in detecting fraudulent transactions, boasting an AUPRC of 0.81. Delving into the confusion matrix, it revealed 56,231 true negatives, concerning 633 false positives, 9 false negatives, and 89 true positives. This paints a picture of a model that excels at pinpointing non-fraudulent transactions with near-perfection but stumbles when tasked with identifying fraud. For non-fraudulent cases, the precision hit a flawless 1.00, yet it plummeted to a dismal 0.12 for fraudulent cases. Recall stood at 0.99 for non-fraudulent and a notable

0.91 for fraudulent transactions. The F1-score mirrored these results: 0.99 for non-fraudulent and a meager 0.22 for fraudulent transactions.

Visual aids, such as the confusion matrix shown in Figure 1 and the precision-recall curve shown in Figure 2, further illuminate these findings. The confusion matrix underscores the overwhelming number of true negatives and the relatively scant true positives, showcasing the model's strength in verifying legitimate transactions but its struggle with fraudulent ones. The precision-recall curve, crowned with an AUPRC of 0.81, underscores the model's general effectiveness but also signals the pressing need for refinement, especially in bolstering precision for fraud detection.

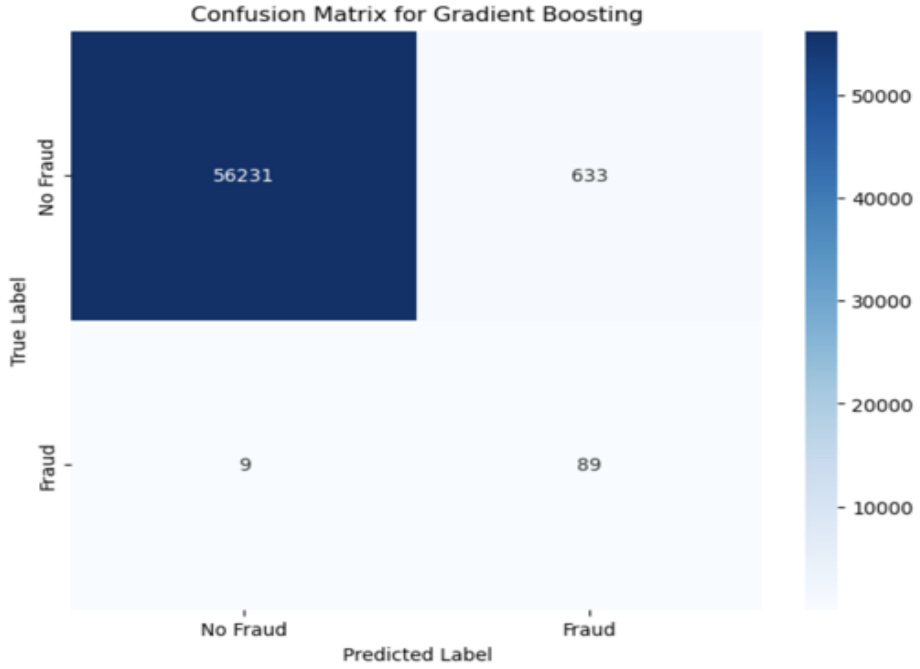

Figure 1. The confusion matrix of the gradient boosting (Photo/Picture credit: Original).

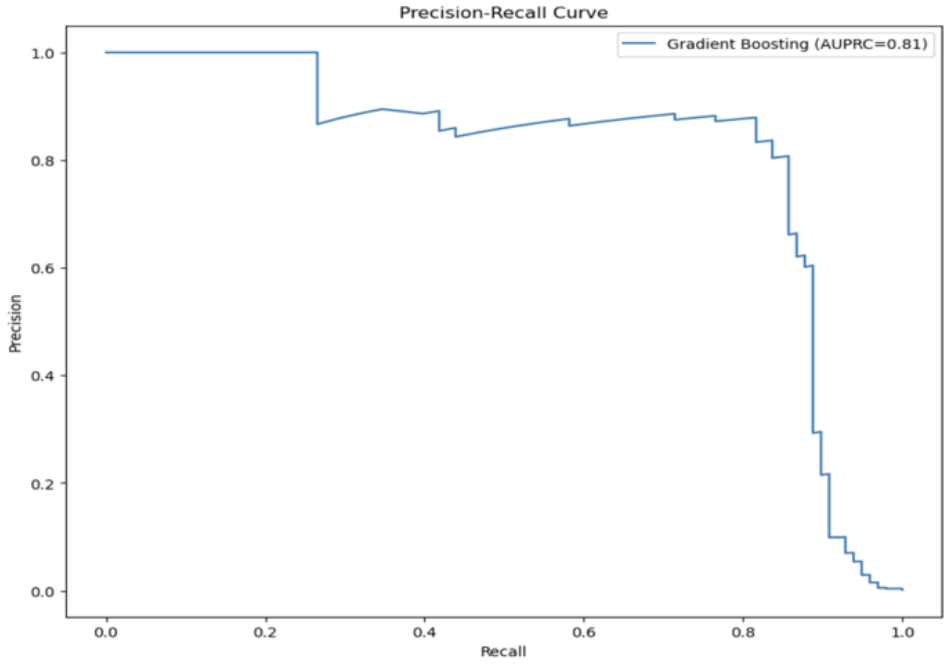

Figure 2. The precision-recall curve of the gradient boosting (Photo/Picture credit: Original).

3.2. Decision tree

In stark contrast, the Decision Tree model, despite its simplicity, offered intriguing insights. Garnering an AUPRC of 0.59, it indicated a moderate proficiency in differentiating between

fraudulent and non-fraudulent transactions. Its confusion matrix showed 56,767 true negatives, 97 false positives, 24 false negatives, and 74 true positives. This model demonstrated exceptional accuracy in recognizing non-fraudulent transactions, with a perfect precision of 1.00. Yet, its precision for fraudulent transactions, at 0.43, although better than Gradient Boosting, still left much to be desired. Recall was impeccable for non-fraudulent transactions at 1.00 and a decent 0.76 for fraudulent ones. The F1-score reflected this dichotomy, with a perfect 1.00 for non-fraudulent and 0.55 for fraudulent transactions.

The confusion matrix shown in Figure 3 visualization sheds light on the model's adeptness in accurately identifying the bulk of non-fraudulent transactions but also exposes a higher incidence of false positives and false negatives when juxtaposed with the Gradient Boosting model. The precision-recall curve shown in Figure 4, with an AUPRC of 0.59, encapsulates the model's balanced yet moderate performance, hinting at potential benefits from further tuning and refinement to enhance its fraud detection capabilities.

Figure 3. The confusion matrix of the decision tree (Photo/Picture credit: Original).

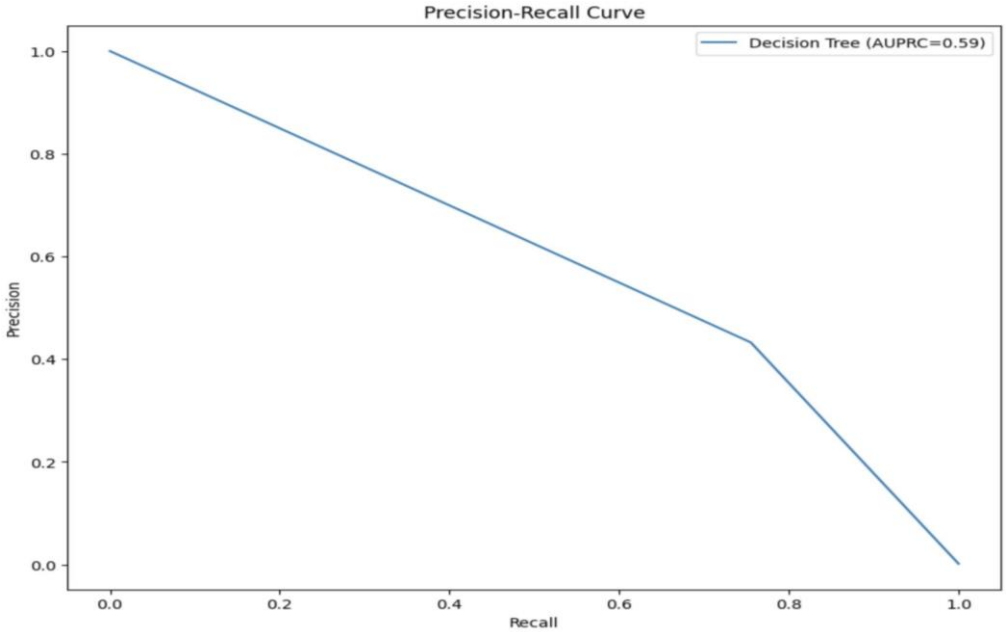

Figure 4. The precision-recall curve of the decision tree (Photo/Picture credit: Original).

3.3. K-Nearest neighbors

The KNN model presented a balance between the two models discussed above, achieving an AUPRC of 0.77. The confusion matrix for the KNN model revealed 56,773 true negatives, 91 false positives, 13 false negatives, and 85 true positives. This model showed a commendable ability to detect fraudulent transactions with a precision of 0.48 for fraudulent cases, significantly better than Gradient Boosting. Recall for fraudulent transactions was also strong at 0.87, while non-fraudulent transactions had a precision and recall of 1.00.

The confusion matrix visualization for KNN highlights a robust identification of fraudulent transactions compared to the other models, with fewer false negatives and a reasonable number of false positives. The precision-recall curve, showing an AUPRC of 0.77, indicates the model's solid performance and potential as a reliable method for fraud detection.

3.4. Discussion

The results of these models highlight the varying strengths and weaknesses of different machine learning approaches in the context of credit card fraud detection. The Gradient Boosting model, with its high AUPRC, showcases its effectiveness in general scenarios but reveals challenges in precision for fraudulent transactions. The Decision Tree model, while simpler, provides a balanced approach with moderate performance, highlighting the trade-off between model complexity and detection capabilities. The KNN model stands out with its higher precision for fraudulent transactions, indicating its potential as a more focused fraud detection tool.

These findings suggest that while advanced models like Gradient Boosting offer high overall accuracy, simpler models like Decision Trees and KNN can provide better precision for specific tasks, such as fraud detection. Future research could explore hybrid models or ensemble techniques to further enhance detection capabilities, combining the strengths of various approaches to achieve even more robust and reliable fraud detection systems.

4. Conclusion

In conclusion, this study explored the efficacy of the above machine learning models—Gradient Boosting, Decision Trees, and KNN. SMOTE was used, so that each model was meticulously trained and evaluated based on metrics such as precision, recall, AUPRC, etc.

The outcomes indicated that while the Gradient Boosting model exhibited a high AUPRC of 0.81 and excelled in identifying non-fraudulent transactions, it struggled with precision for fraudulent cases. The Decision Tree model, although simpler, offered a balanced performance with an AUPRC of 0.59, highlighting its potential when model simplicity and interpretability are crucial. The KNN model emerged as a robust contender, achieving an AUPRC of 0.77 and demonstrating higher precision for fraudulent transactions compared to Gradient Boosting, indicating its effectiveness as a focused fraud detection tool.

While advanced models like Gradient Boosting provide high overall accuracy, simpler models like Decision Trees and KNN can deliver better precision for detecting fraud. Future research should delve into the realm of hybrid models or ensemble techniques, aiming to amplify detection capabilities by harnessing the strengths of diverse approaches.

References

[1] Yusuf Sahin, Serol Bulkan, Ekrem Duman. 2013. A cost-sensitive decision tree approach for fraud detection. Expert Systems with Applications, 40(15): 5916-5923.

[2] Aderemi O. Adewumi, Andronicus A. Akinyelu. 2017. A survey of machine-learning and nature-inspired based credit card fraud detection techniques. International Journal of System Assurance Engineering and Management, 8: 937-953.

[3] Tej Paul Bhatla, Vikram Prabhu, Amit Dua. 2003. Understanding credit card frauds. Cards Business Review, 1(6): 1-15.

[4] Sahil Dhankhad, Emad Mohammed, Behrouz Far. 2018. Supervised machine learning algorithms for credit card fraudulent transaction detection: a comparative study. In 2018 IEEE International Conference on Information Reuse and Integration (IRI), IEEE, 122-125.

[5] S. Patil, V. Nemade, P.K. Soni. 2018. Predictive modelling for credit card fraud detection using data analytics. Procedia Computer Science, 132: 385-395.

[6] Yashvi Jain, Namrata Tiwari, Shripriya Dubey, Sarika Jain. 2019. A comparative analysis of various credit card fraud detection techniques. International Journal of Recent Technology and Engineering, 7(5): 402-407.

[7] Credit Card Fraud Detection. 2019. Available: https://www.kaggle.com/mlg-ulb/creditcardfraud.

[8] Alexey Natekin, Alois Knoll. 2013. Gradient boosting machines, a tutorial. Frontiers in Neurorobotics, 7: 21.

[9] Yan-Yan Song, L.U. Ying. 2015. Decision tree methods: applications for classification and prediction. Shanghai Archives of Psychiatry, 27(2): 130.

[10] Leif E. Peterson. 2009. K-nearest neighbor. Scholarpedia, 4(2): 1883.