Research on Detecting Credit Card Fraud Through Machine Learning Methods

Shunning Dai[1, *]

[1] SWJTU-Leeds Joint School, Southwest Jiaotong University, Chengdu, 610097, China

*Corresponding author: dsn@my.swjtu.edu.cn

Abstract. The heavy use of credit cards inevitably leads to the escalation of fraud technology and a surge in fraudulent behavior. Machine learning, a multi-interdisciplinary discipline with numerous algorithms, can effectively detect and prevent financial fraud. This study focuses on several common machine learning methods applied to fraud detection and then evaluates how they perform on real data, including Bagging, Random Forest, Decision Tree, and AdaBoost. However, the proportion of fraudulent transactions in real transaction data is extremely unbalanced. SMOTE can determine the data imbalance problem, while confusion matrices visualize the classification results of different classes. The experiment results reveal that Random Forest performs best for both unbalanced and balanced data. It indicates that random forest is better for detecting fraudulent transactions.

Keywords: Machine Learning, Fraud Detection, SMOTE, Confusion Matrix

1 Introduction

Due to technological advances and the introduction of new e-financing options, digital payment has become the most popular payment method in recent years. As online payment platforms become increasingly important in everyday life, credit card fraud has proliferated and caused significant losses. Amanze and Onukwugha (2018) found that credit card fraud in Nigeria continued to increase between 2014 and 2016 [1]. Moreover, in accordance with a 2020 report by "UK Finance", the number of reported card fraud incidents in the UK had reached GBP 574.2 million. The growing number of credit card frauds has undoubtedly reduced public confidence in the financial sector and affected daily life. While financial institutions are already trying to address this problem with various fraud detection models, they can still not stem their rapid growth.

To effectively prevent and detect financial fraud, relevant financial institutions and scientists are committed to developing professional analytical techniques. There are abundant previous works of literature on fraud detection, which is mainly divided into three aspects. In terms of Support vector machine (SVM), Xu et al. (2015) suggested a model for detecting credit card fraud online utilizing the optimized support vector machine model for banking data. It was found that SVM is superior to the ID3+BP hybrid

V. Escudero et al. (Eds.): BADS 2022, AHCS 11, pp. 1030–1037, 2023.
https://doi.org/10.2991/978-94-6463-102-9_107

model [2]. In terms of Hidden Markov Model (HMM), Agrawal and Kumar et al. (2015) suggested an identification model of credit card fraud after examining cases combining genetic algorithms, behavior-based and HMM. The results showed that the model benefited from credit card fraud [3]. In terms of Artificial Neural Network (ANN), Sahin and Duman (2011) combined ANN and Logistic Regression (LR) to detect the entire process of credit card transactions in an automated and efficient way to ensure their security and efficiency. The results showed that ANN is superior to LR [4].

Machine learning has already had a significant impact on many tasks and jobs performed by humans. Previous research on machine learning has been fruitful and mainly includes three aspects. In the medical field, Vaishya et al. (2020) showed that machine learning methods were identified to aid diagnosis and help researchers develop treatments for COVID-19 [5]. In the financial sector, the Wall Street Journal reported in 2010 that Rebellion Research has employed machine learning to forecast financial crises. In the field of digital media, Dey et al.(2020) found that machine learning methods can optimize smartphone performance and improve user experiences based on user interactions with the phone [6].

Therefore, this article wants to investigate the feasibility of machine learning methods in fraud detection. In the process of fraud detection, Machine learning can perform multi-processing data analysis using rich data and surveillance models, build anti-fraud models in real-time, and identify fraudulent behaviour in real-time based on current user characteristics. This study intends to train four supervised learning algorithms on the same credit card transaction dataset and select the best-performing algorithm by comparing its Accuracy, Precision, Recall, and F1-Score.

2 Methods

This section reviews some supervised classification methods widely used for credit card fraud detection. Based on the available characteristics of the transaction data, classifiers can classify transactions as fraud or legality according to the available characteristics of transaction data.

2.1 Random Forest

Random forest can be considered as an algorithm that integrates several decision trees across ensemble thinking. For the classification task, random forest chooses to output the classes chosen by the majority of trees, while for the regression task, the average value of each tree is selected to be returned [7]. Random forest can reasonably speculate on a large amount of data without configuration. This method was used when detecting credit card fraud both online and offline.

2.2 Decision Tree

Decision tree is by far the most easily understood concept by humans because it can visually and unambiguously represent decisions and decision-making. Decision tree can be thought of as a predictive model that represents a correspondence between the object attribute and the object value. Each node in the decision tree represents the criteria for determining the attributes of the object, its branch represents the objects that meet the conditions of the node, and the leaves of the tree denote the prediction results of the object. This method is widely used for credit card fraud detection [8].

2.3 Bagging

Bagging is an essential integrated learning method commonly used in classification and regression, which can enhance its precision and stability by decreasing the square deviation of the results while avoiding overfitting. Its working mechanism can be boiled down to the construction of multiple classifiers or regressions by means of multiple rounds of sampling substitution. The final prediction is the average performance of the sample on these learners. Bagging is also currently being used in the financial industry for deep learning models, including fraud detection, credit risk assessment and option pricing issues [9].

2.4 AdaBoost

Adaboost is a binary classification model that belongs to supervised learning in machine learning, whose basic idea is to train various weak classifiers in the same training set, then combine them reasonably to form a strong classifier [10]. This method is also a classifier with high accuracy and is simple to operate without the need for feature screening. In addition, Adaboost has proven to be suitable for fraud detection in the banking system.

3 Class Imbalance Problem

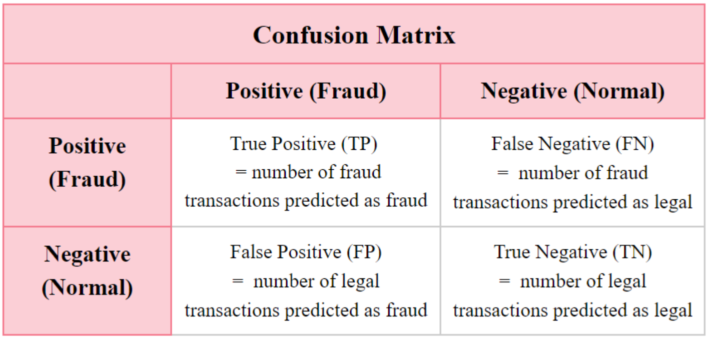
In effect, the number of fraudulent credit card transactions is well below the number of legitimate transactions. As a result, the misclassification of a few class instances is high when training machine learning algorithms. In this section, two basic approaches to class imbalance handling are presented.

3.1 Synthetic Minority Oversampling (SMOTE)

SMOTE can be regarded as a synthetic sampling method that starts from the minority class samples, finds neighbouring samples, and synthesizes new minority class samples so that the number of positive samples is roughly the same as the number of negative samples. The actual operation of this technique involves oversampling fraudulent transactions and subsampling normal transactions.

3.2 Confusion Matrix

Confusion matrix is an array layout that makes it possible to visualize the performance of an algorithm. Each line of the confusion matrix denotes the instance of an actual class, and each column denotes the instance of an expected class. Figure 1 presents the confusion matrix.



Confusion Matrix		
	Positive (Fraud)	**Negative (Normal)**
Positive (Fraud)	True Positive (TP) = number of fraud transactions predicted as fraud	False Negative (FN) = number of fraud transactions predicted as legal
Negative (Normal)	False Positive (FP) = number of legal transactions predicted as fraud	True Negative (TN) = number of legal transactions predicted as legal

Fig. 1. Confusion Matrix

More advanced classification indicators can be obtained from the confusion matrix: as shown below:

3.2.1. Accuracy.

Accuracy is the most original measure of classification issues, reflecting the percentage of the total sample in which the predictions are correct, defined as formula (1):

$$Accuracy\ = \frac{TP+TN}{TP+TN+FP+FN} \tag{1}$$

In the case of an imbalance in the sample, Accuracy is not suitable for the assessment of model performance.

3.2.2. Precision.

Precision refers to the likelihood of a positive sample being accurately predicted, defined as formula (2):

$$Precision\ = \frac{TP}{TP+FP} \tag{2}$$

Accuracy and Precision are in no way similar to each other. Precision is a concept only for positive samples, whereas Accuracy is the overall accuracy of the prediction, including both positive and negative samples.

3.2.3. Recall.

Recall represents the model's actual ability to identify positive samples. In other words, it is equivalent to the percentage of fraudulent transactions forecast by models over actual fraudulent transactions, which is defined as formula (3):

$$Recall = \frac{TP}{TP+FN} \tag{3}$$

3.2.4. F1-Score

F1-Score accounts for *Precision* and *Recall* in the classification model and can be considered as a weighted average of them, with values between 0-1. It is defined as formula (4):

$$F1 - Score = \frac{TP}{TP + \frac{1}{2}(FP + FN)} \tag{4}$$

4 Experiment

This study applies classification methods to fraud detection in Bagging, Decision Tree, Random Forest, and AdaBoost. This experiment aims to find an appropriate algorithm to process large quantities of data into a fraud detection model. Therefore, two actual comparisons of these machine learning methods will be made before and after balancing credit card data.

4.1 Data description

The dataset used in this experiment is derived from research collection and analysis by Worldline and Université Libre de Bruxelles during a research collaboration on financial fraud detection. The data shows 284807 transactions that occurred in two days, including 492 scams, with a fraud rate of only 0.173%.

4.2 Experimental process

This experiment consists of five steps, which are as follows:

(1) Import the required modules and datasets, print out the basics of the dataset and visualize their transaction distribution.

(2) Detect the distribution of features, normalize the 'Amount' and 'Time' features, and then replace the original data with the normalized fields and data.

(3) Import the models and make a function to print out the classification report for four models trained on this unbalanced dataset.

(4) Use SMOTE method to balance the data.

(5) Use the same function but add a simple part to print out the figure "Precision-Recall Curve", and apply these models to balanced data.

(6) Compare and analyze the performance of these four classifiers on unbalanced and balanced data.

4.3 Result

The running tool used in this experiment is Jupyter Notebook which is a web application for interactive computation. In this experiment, Accuracy, Precision, Recall and F1-score are employed as evaluation criteria for model comparison.

Figure 2 represents the model performance before balancing the data. Overall, Bagging and Decision Tree are the two best-performing models for unbalanced data.

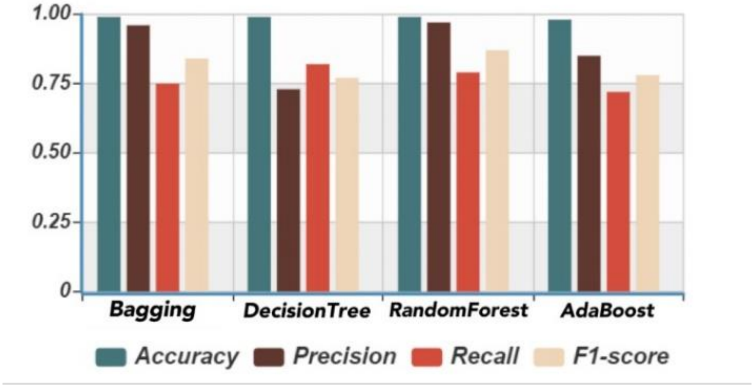


Fig. 2. Model Performance for Unbalanced Data

Figure 3 gives the performance measurements for all the applied models after SMOTE. Decision Tree performs better than Bagging, while Bagging is far superior to the other two models.

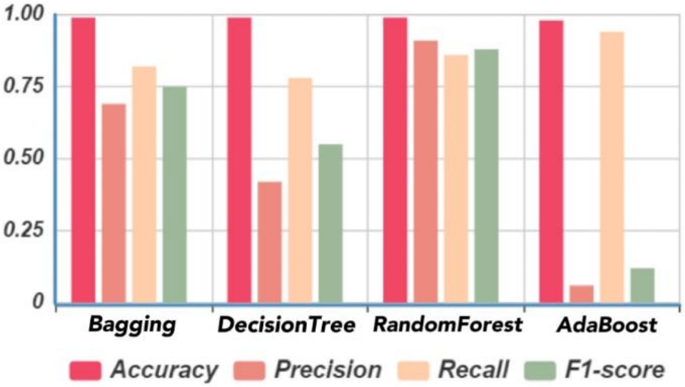


Fig. 3. Model Performance for Balanced Data

4.4 Discussion

For the unbalanced data, Random Forest and Bagging perform well on all performance criteria, while Decision Tree and AdaBoost have low F1-score. For the balanced data, Random Forest still maintains good performance, but the performance indicators of the other three models have decreased significantly except for Recall and Accuracy. In summary, Random Forest is the best-performing model for this dataset. Each dataset has its characteristics that require investigation and analysis to find the right model.

5 Conclusion

This study explores several machine learning methods for credit card fraud detection. Financial fraud has seriously hindered the development of the financial industry, so it's critical to find a model that can process data quickly and efficiently. This article conducts a series of experiments on four common detection methods to find the most effective algorithms among them. There are three main conclusions in this paper. Firstly, an imbalance in sample classes can result in a small sample classification containing too few features to extract rules from. Even if a classification model is obtained, it is particularly prone to overfitting problems. Secondly, SMOTE method and confusion matrix are proven to be able to balance the data effectively. Finally, the experiment compares and analyzes the classifier's four performance indicators, concluding that the random forest can better detect fraudulent transactions.

The implications of this research can be divided into two aspects. In terms of theory, this article enriches the literature in related fields. On the practical side, the research results of this experiment can open up a new field of vision for the banking and financial industry and help to develop a system that can effectively prevent financial fraud.

However, this study has two limitations. On the one hand, the dataset used in this study comes from only one financial institution, which means that the results may not apply to all financial institutions. Future research should be trained on complex and massive datasets. On the other hand, this lab does not use any unsupervised learning methods and integrated models, and future research can look at different types of machine learning techniques and try different sets of classifiers.

References

1. Amanze, B.C., Onukwugha, C.G. Credit Card Fraud Detection System in Nigeria Banks Using Adaptive Data Mining and Intelligent Agents: A Review. Journal 7(7), 2018.
2. Zareapoor, M., Shamsolmoali, P.: Application of credit card fraud detection: Based on bagging ensemble classifier. Journal 48, 679–685(2015).
3. Agrawal, A., Kumar, S., Mishra, A.K.: Credit card fraud detection: A case study. IEEE, pp. 5–7, 2015.
4. Sahin, Y., Duman, E.: Detecting credit card fraud by ANN and logistic regression. IEEE, pp. 315–319, 2011.
5. Vaishya, R., Javaid, M., Khan, I. H.et al: Artificial Intelligence (AI) applications for COVID-19 pandemic. Journal 14(4), 337–339(2020).
6. Vincent, J: The first AI-generated textbook shows what robot writers are actually good at. 2019.
7. Ho T.: The Random Subspace Method for Constructing Decision Forests. Journal 20(8), 832–844(1998).

8. Paruchuri, H.: Credit Card Fraud Detection using Machine Learning: A Systematic Literature Review. ABC Journal of Advanced Research, Journal 6(2), 113-120(2017).
9. Chawan, S.: Study of Data Mining Techniques used for Financial Data. International Journal of Engineering Science and Innovative Technology (IJESIT), 2013.
10. Hastie, T., Rosset, S., Zhu, J. et al: Multi-class AdaBoost. Statistics and Its Interface. Journal 2(3), pp. 349–360(2009).