A Review of Machine Learning Applications for Credit Card Fraud Detection with A Case study

Zahra Faraji

1Financial Service Analytics Institution, University of Delaware, United States
* Corresponding author: zfaraji@udel.edu

Article History
Received 2022-01-14
Revised 2022-02-10
Accepted 2022-02-03
Published 2022-02-15

Keywords
Machine learning,
Credit card fraud detection,
Artificial Neural Networks,
Logistic regression,
Random Forests,
XGBoost

How to cite?

Purpose - This paper aims to highlight the widely used supervised techniques applied for fraud detection. In addition, this paper aims to apply some techniques to evaluate their performance on real-world data and develop an ensemble model as a potential solution for this problem.

Design/Methodology- Different techniques applied in this study for fraud detection purposes are logistic regression, decision tree, random forest, KNN, and XGBoost. The confusion matrix gives information about the assignment of inputs to the different classes. This study uses precision and recall to evaluate the performance, calculated based on the confusion matrix.

Findings- XGBoost is the fastest and is expected to have the best performance; however, it is only outperforming the random forest in terms of accuracy, precision, recall, and f1-score. In general, the KNN and logistic regression have better performance, which means they better detect fraudulent transactions.

Practical Implications- The new model can be applied to new data instead of the previous techniques.

Copyright © 2022 The Author(s)
Introduction

Credit cardholders are encouraged to use their card over other ways of payment, such as debit cards. Indeed, there are some advantages to using credit cards compared with other payment methods; for example, they are so easy and fast to use, which support national or international transactions and withdrawing cash (Al Rubaie, 2021). However, some associated risks increase the attention for better credit card use management. Credit cards are an easy target for fraud, but detecting fraudulent transactions is usually an easy task (Al Rubaie, 2021; Dornadula, 2019). There are so many online businesses that only accept credit card transactions high usage of credit cards for online transactions led to the high number of fraudulent transactions.

Fraud detection is crucial, and the increasing trend of credit card online transactions makes strong prevention and detection techniques necessary. The total number of Visa and Master card users was about 2,183 million worldwide in 2020. Over 30% of the Visa card users and 24% of Master card users were in the United States (Alkhatib, 2021). The increasing number of cashless transactions resulted in an increasing trend of fraud rate for Card Not Present (CNP) transactions (Murli, 2015). In addition, the COVID pandemic led more people and merchants to use online transactions, highlighting the need to prevent and detect fraud. Consequently, Researchers and financial institutions are continuously seeking more efficient techniques to increase the safety of online transactions. Financial institutions are using prevention and detection mechanisms to check all the transactions, and Machine Learning (ML) techniques have been so helpful for this purpose.

There are different ML techniques to tackle credit card fraud detection, but we can classify them into main groups, including supervised, unsupervised, and reinforcement learning. The supervised learning techniques are applicable for classification and prediction problems, and data should be labeled for these techniques. This group contains techniques such as Support Vector Machine (SVM), Logistic Regression, Decision Tree, Naïve Bayes, K-Nearest Neighbor, Random Forest, Artificial Immune System, and Artificial Neural Network. On the other hand, the unsupervised learning techniques work with the unlabeled data and cluster the inputs based on their similarities. Some unsupervised ML techniques are K-means, Hidden Markov Model, Genetic Algorithm, Gradient Descendent, and DBSCAN (Zareapoor, 2015).

This paper is organized as follows: Section 2 gives information about the related credit card fraud detection techniques. Section 3 describes the experimental setup approach by explaining some data challenges, required pre-processing, and the different classifiers used to detect fraudulent credit card transactions. Section 4 presents the experimental results and discussion, and Section 5 concludes this work and provides some suggestions for future studies.

Literature Review

This section reviews some of the most important and widely used supervised classification techniques for credit fraud detection. Credit card fraud is a binary problem, as a transaction can be normal or fraudulent. The goal for the classification techniques is to classify a transaction based on the available features as fraud or legal.

Artificial Neural Network

Artificial Neural Network (ANN) imitates the human brain based on interconnected nodes or neurons. Each ANN has input, hidden, and output layers. Using ANN, the computer can learn from data and humanly make decisions. (Mehndiratta S. &., 2019). The neurons receive and process the signals and send them to the subsequent neurons for further process. There is a weighted connection between the active nodes in the associated layers. ANN has been widely used because of its ability to extract patterns in complex data. (Lim, 2021). Many researchers applied this algorithm for credit card fraud detection purposes (Murli, 2015; Paruchuri,
The Convolution Neural Network (CNN) and Recurrent Neural Network (RNN) are two types of neural networks implemented by the researcher to detect fraudulent activities. (Tiwari, 2021).

**Genetic Algorithm (GA)**

The genetic algorithm can optimize different problems such as fraud detection and prevention for online transactions (Paruchuri, 2017). This method is a subset of the natural evolution idea to reproduce the fitter or stronger individuals for the next generation (Tiwari, 2021).

**Logistic Regression**

The logistic regression model is a generalized linear technique used when the predictive variable is binary. This technique uses the logistic function to model the probability for different classes (Tiwari, 2021). Logistic regression has been widely used for fraud detection (Ngai, 2011). Some studies showed that logistic regression could outperform other credit card fraud data (Paruchuri, 2017).

**Decision Tree**

A decision tree is a predictive model for classification and regression problems that map the inputs to the possible classes. This supervised technique has a tree-like structure that contains root nodes, other nodes that split the data based on the features, and leaves. At each node, the classes in the dataset are separated based on some conditions on the features. The splitting process achieves the fullest purity (Paruchuri, 2017). This technique is widely used for credit card fraud detections (Ngai, 2011; Paruchuri, 2017).

**Random Forest**

Random forest is another supervised technique that uses the bagging idea to improve the results by combining multiple single trees. This approach uses a random subset of each tree's features and a training dataset to overcome the disadvantages of a single decision tree (Tiwari, 2021). This method has been used for credit card fraud detection for online and offline transactions (Paruchuri, 2017).

**Support Vector Machines (SVM)**

SVM is another supervised method that is applicable for classification and regression problems. This technique can handle high-dimensional data by transforming a non-linear input into a linear task. SVM has been applied to detect different types of fraud in banking industries (Paruchuri, 2017). SVM creates the decision boundary to segregate the classes.

**K-Nearest Neighbors (KNN)**

KNN is a supervised method used for classification and regression challenges. This method classifies each input based on the nearest feature space in the training dataset without generalization (Lim, 2021; Mehndiratta S. &., 2019). This technique is fast but, as mentioned, doesn’t make any generalizations. Indeed, if the K-nearest neighbors are fraudulent transactions, it is highly possible that a new instance is classified as fraudulent.

**Naïve Bayesian Classifier**

This method is only applicable for classification problems that leverage the Bayesian rules to calculate the conditional probabilities of the classes (Tiwari, 2021). The class with the highest conditional probabilities will be assigned to the instances. An important assumption for this model is that the effect of a feature on a given label is unrelated or independent from the value of the other features (Mehndiratta S. &., 2019).

**Gradient Boosted Trees:**

This is a supervised technique for classification and regression problems using an ensemble of decision trees. It uses the gradient boosting framework and builds sequences of weak models or trees, and at each iteration, it
tries to predict the error of the previous model (Mishra, 2018). XGBoost is based on gradient boosted decision trees, which are also applicable for categorical and continuous target variables. It is more regularized than gradient boosted trees for more generalization, resulting in better speed and performance (Meng, 2020).

**Hidden Markov Model (HMM)**

This stochastic model contains unobserved or hidden states that can be predicted by the underlying Markov Chain. Indeed, the only available information is the observational data, and there is no information related to the states. This model's main assumption is that future event are only related to the current state but not to the previous states (Tiwari, 2021).

Another way to tackle the fraud detection problem is combining or using multiple machine learning techniques to perform better. Different studies compared multiple methods together and even made a hybrid technique to boost the model (Faraji Z., 2020; Faraji Z., 2020; Asgharizadeh, 2014; Faraji, Z., Fleischhacker, A., 2020). Research tackled the credit card fraud detection problem with three supervised techniques, support vector machines, logistic regression, and random forests. One study compared the naïve Bayes model and logistic regression for fraud detection. The findings revealed that the naïve Bayes model could converge faster, but the discriminative logistic regression has the lower asymptotic error (Ng, 2002). Another research studied the three models: logistic regression, decision tree, and the neural network for fraud detection. The findings showed the decision tree underperformed the other models (Shen, 2077). However, in another study, the decision tree outperformed the SVM and had better performance detecting fraudulent instances (Sahin, 2011). The findings suggested that the logistic regression had better performance. Still, SVM considered more fraudulent instances in the training data (Bhattacharyya, 2011). Another study applied 10 ML techniques and compared their performance using the accuracy and confusion matrix. They applied Decision Tree, KNN, SVM, Logistic Regression, Random Forest, Naïve Bayes, Gradient Boosting, XGB Classifier, and Stacking Classifier to detect the fraudulent instances logistic regression showed better performance (Dhankhad, 2019).

Different standard and hybrid machine learning techniques were used to detect credit card fraud activities in one study. The hybrid model applied the AdaBoost with majority voting technique for a real-world dataset in which noises were added to test the model robustness. Random Forest showed the best performance compared to other models in this study (Randhawa, 2018). Based on the literature, different machine learning techniques can be applied to detect fraud activities for the credit card. Each technique has its benefits and disadvantages (Faraji, Z., Fleischhacker, A., 2020; Faraji Z., in press); however, a researcher should choose a technique based on the available data and associated feasibilities. For example, SVM requires lots of memory and considerable time to train the data. All the attributes should be independent for naïve Bayes; KNN needs scaled data, which is not applicable for high-dimensional data (Uchhana, 2021). Indeed, it is necessary to consider the data and computation limitations to choose the appropriate technique (Faraji, Z., Fleischhacker, A., 2020; Faraji Z., 2020).

**Research Method**

**Data Challenges**

Usually, the dataset for fraud detection is unbalanced as the number of fraudulent transactions is considerably less than the normal transactions. All the previously mentioned methods can predict the majority class in the data, but the fraudulent class is considered noises. Therefore, misclassification for minority class instances is high while using the standard techniques for fraud data (Mishra, 2018). However. The goal is to classify the minority class with good performance. One solution is to apply the Synthetic Minority Oversampling Technique (SMOTE) for raw data before the modeling. SMOT oversamples the minority class and under samples the
majority class. Indeed, the new dataset is balanced. This method creates "synthetic" from the minority class using the KNN technique (Mishra, 2018).

Even if the model classifies all the instances as normal transactions with highly imbalanced fraud data, the accuracy would be high but misleading. Therefore, accuracy is a weak performance metric for fraud data. This study uses precision and recall to evaluate the performance, calculated based on the confusion matrix. The confusion matrix gives information about the assignment of inputs to the different classes. The following table shows the confusion matrix (Al Rubaie, 2021).

<table>
<thead>
<tr>
<th></th>
<th>Positive (Fraud)</th>
<th>Negative (Normal)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Positive (Fraud)</td>
<td>True Positive (TP)</td>
<td>False Negative (FN)</td>
</tr>
<tr>
<td>Negative (Normal)</td>
<td>False Positive (FP)</td>
<td>True Negative (TN)</td>
</tr>
</tbody>
</table>

The performance metrics can be calculated based on the confusion matrix as follows:

True Positive (TP) = number of fraud transactions predicted as fraud
True Negative (TN) = number of legal transactions predicted as legal
False Positive (FP) = number of legal transactions predicted as fraud
False Negatives (FN) = number of fraud transactions predicted as legal

These values are applied to compute the following (Zareapoor, 2015):

\[
\text{Accuracy} = \frac{TP + TN}{TP + FP + TN + FN}
\]

\[
\text{Recall} = \frac{TP}{TP + FN}
\]

\[
\text{Precision} = \frac{TP}{TP + FP}
\]

\[
F_1 = \frac{TP}{TP + \frac{1}{2}(FP + FN)}
\]

Data Description
The data for this study is public data available on the Kaggle website. The data contains 284,807 transactions in a two-day duration that only a tiny fraction of that is labeled as fraudulent transactions. There are 28 features in the dataset transformed by principal component analysis (PCA). The transaction amount is another feature that needs to be scaled before training the model. The next table summarizes the transaction distribution, and Figure 1 visualizes the imbalance data.

<table>
<thead>
<tr>
<th>Total Transaction</th>
<th>Legal</th>
<th>Fraud</th>
<th>Fraud rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>284,807</td>
<td>284315</td>
<td>492</td>
<td>0.173%</td>
</tr>
</tbody>
</table>
Our Approach

The different classification techniques are applied in this study for fraud detection purposes are logistic regression, decision tree, random forest, KNN, and XGBoost. Their performances are compared to see which model can better extract the relationship between the features and detect fraudulent transactions. The logistic regression estimates the probabilities by using a logistic regression equation to model the relationship between the dependent and independent variables. The decision tree is another classification model that is applied to the data. This model has some advantages, such as being easy to interpret the results and visualize the tree, making it easier to communicate with non-technical persons. The random forest is applied to the dataset as an ensemble model that includes more trees. The final results are based on the most frequent predictions by individual trees. The gradient boosting method is a tree-based learning algorithm with a sequence of decision trees built on the previous trees' errors. This model is fast and can be an excellent choice for fraud data. In addition, cross-validation and hyperparameters tuning is applied for all the models in this study to find the best fit and avoid overfitting.

After training all the classifiers, a new ensemble model will be applied as a voting classifier to combine all the other classification techniques. The objective is to reduce the errors of single models, which helps the ensemble model make better predictions compared with the individual classifiers. If all the classifiers are considered as $C_1, C_2, C_3, C_4,$ and $C_5$, then the final classifier will take the votes as the majority of votes as the final prediction or $C_t$.

$$C_t = \text{Majority}\{C_1, C_2, C_3, C_4, C_5\}$$

Results

The next table shows the model's performance before balancing the data. According to Table 3, the uneven instances in the data resulted in poor performance for most of the models. Figure 2 shows the performance of the XGBoost model, which is overperforming the other models. Most accuracy scores are high as most normal transactions are classified correctly in imbalanced data. However, by looking at the other metrics, it is clear that the precision, recall, and f1-score are lower than the accuracy. Also, the XGBoost model is overperforming the other techniques based on all the metrics.
In the next step, the data is balanced using the SMOT method to oversample the fraud transactions and undersample the normal transactions. Figure 3 shows the new distribution for the dataset.

<table>
<thead>
<tr>
<th>Methods</th>
<th>Accuracy</th>
<th>Precision</th>
<th>Recall</th>
<th>F1-score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Logistic regression</td>
<td>0.81</td>
<td>0.90</td>
<td>0.63</td>
<td>0.74</td>
</tr>
<tr>
<td>Decision tree</td>
<td>0.99</td>
<td>0.77</td>
<td>0.76</td>
<td>0.77</td>
</tr>
<tr>
<td>Random forest</td>
<td>0.99</td>
<td>0.94</td>
<td>0.78</td>
<td>0.85</td>
</tr>
<tr>
<td>KNN</td>
<td>0.86</td>
<td>0.91</td>
<td>0.72</td>
<td>0.81</td>
</tr>
<tr>
<td>XGBoost</td>
<td>0.99</td>
<td>0.92</td>
<td>0.82</td>
<td>0.87</td>
</tr>
</tbody>
</table>

After balancing the data, all the previous machine learning techniques are applied. The next figure shows the confusion matrix for the XGBoost model, which illustrates the performance improvement after leveraging the SMOT method.
Table 4 gives detailed information about the performance measurements for all the applied models. A new ensemble model is also applied as a voting classifier to combine all the other classification techniques. The idea is that the ensemble model is stronger than the single model. The results are based on voting on the predicted classes, aiming to reduce the error. According to Table 4, the voting classifier model outperforms the other models.

<table>
<thead>
<tr>
<th>Methods</th>
<th>Accuracy</th>
<th>Precision</th>
<th>Recall</th>
<th>F1-score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Logistic regression</td>
<td>0.98</td>
<td>0.98</td>
<td>0.93</td>
<td>0.96</td>
</tr>
<tr>
<td>Decision tree</td>
<td>0.99</td>
<td>0.77</td>
<td>0.76</td>
<td>0.77</td>
</tr>
<tr>
<td>Random forest</td>
<td>0.99</td>
<td>0.93</td>
<td>0.78</td>
<td>0.85</td>
</tr>
<tr>
<td>XGBoost</td>
<td>0.99</td>
<td>0.93</td>
<td>0.83</td>
<td>0.88</td>
</tr>
<tr>
<td>KNN</td>
<td>0.98</td>
<td>0.98</td>
<td>0.94</td>
<td>0.96</td>
</tr>
<tr>
<td>Ensemble</td>
<td>0.99</td>
<td>0.99</td>
<td>0.99</td>
<td>0.99</td>
</tr>
</tbody>
</table>

The logistic regression performance is better than the decision tree, and the XGBoost outperforms the random forest in terms of accuracy, precision, recall, and f1-score. The KNN and logistic regression have better performance than other models. Their low false-negative rates mean better at capturing the fraudulent transactions and extracting the data pattern. The ensemble model gains better results by reducing the errors of the other models. Figure 5 visualizes the performance measurements.
Discussion
The ensemble model gains better results by reducing the errors of the other models. The logistic regression is the most straightforward technique in this paper, but its performance is better than the decision tree. The XGBoost is the fastest and expected to have the best performance; however, it is only outperforming the random forest in terms of accuracy, precision, recall, and f1-score. In general, the KNN and logistic regression have better performance, which means they better detect fraudulent transactions. These models are more straightforward than XGBoost, highlighting that the model complexity doesn’t guarantee good performance. Each dataset has its features and requires investigation to find the appropriate model.

Conclusion
This paper reviewed the machine learning techniques for credit card fraud detection problems. Fraud is an important problem for financial institutions; therefore, applying a model that can handle the data fast and efficiently is critical. Balancing the data is so crucial to achieving a stable and generalized approach. Being familiar with the different algorithms can be helpful to make a better decision to choose the right technique. In this paper, five different supervised techniques were applied for public Kaggle data and their performance was investigated with imbalance data and the balanced data. In addition, a new ensemble model was applied to improve the performance of the individual classifiers.

Limitations and Future Research
There are some limitations to this study; one limitation is the data. The data for this study was limited to one financial institution, which indicates that the results cannot be generalized for all the banks or financial institutions. Future studies can investigate the ML techniques with larger datasets. Also, another limitation is that the unsupervised techniques are not used in this study. In future studies, the new ensemble model can be applied for a group of supervised and unsupervised techniques to see the possibility of performance improvement. In future works, the unsupervised machine learning algorithms can be reviewed and compared with supervised techniques. In addition, the data for this study has only numerical features. However, other types of the data, such as textual data, can be beneficial to improve the fraud detection process.

Funding: This research received no external funding

Conflicts of Interest: The authors declare no conflict of interest

References


