Retraction

Retraction: Machine Learning based Fraud Analysis and Detection System (J. Phys.: Conf. Ser. 1916 012115)

Published 23 February 2022

This article (and all articles in the proceedings volume relating to the same conference) has been retracted by IOP Publishing following an extensive investigation in line with the COPE guidelines. This investigation has uncovered evidence of systematic manipulation of the publication process and considerable citation manipulation.

IOP Publishing respectfully requests that readers consider all work within this volume potentially unreliable, as the volume has not been through a credible peer review process.

IOP Publishing regrets that our usual quality checks did not identify these issues before publication, and have since put additional measures in place to try to prevent these issues from reoccurring. IOP Publishing wishes to credit anonymous whistleblowers and the Problematic Paper Screener [1] for bringing some of the above issues to our attention, prompting us to investigate further.

[1] Cabanac G, Labbé C and Magazinov A 2021 arXiv:2107.06751v1

Retraction published: 23 February 2022
Machine Learning based Fraud Analysis and Detection System

N Kousika¹, Vishali G¹, Sunandhana S¹, Vijay Arvind M¹
¹Department of Computer Science and Engineering, Sri Krishna College of Engineering and Technology, Coimbatore, India.
kousika@skcet.ac.in

Abstract. The spectacular surge in the proportion of credit card transactions, web based purchases, has led to a surge in fraudulent activities recently. For any business establishment, credit card security is a major concern. In this respect, credit card fraud is hard to identify. Thus it became imperative to implement effectual fraud detection systems for all credit card issuing banks to mitigate their losses. Betrayed transactions with real transactions in actuality are often dispersed and simple methods of matching are not enough to detect them accurately. The paper proposes an algorithm based on Machine Learning credit card fraud detection to solve the issue of a fraudulent transaction. This framework nominally increases the probability of card fraud by exponential activity. The results show that the accuracy of Random Forest, Support Vector Machine and KNN classifiers achieves respectively 94.84%, 89.46%. Random Forest could even predict new fraud cases very quickly.

Keywords: Credit Card Fraud, Machine Learning algorithm, Fraud detection.

1. Introduction

Using credit cards for a variety of purchases is quickly becoming a necessity. With this fantastic and quick method of performing transactions comes a large amount of risk. Based on rapid internet funding growth, digital transfers and rapid banking industry expansion, credit card card usages are becoming more common in daily living [1]. A credit card can be made in two forms. When it is displayed physically, the first time you use this card for a charge, cancellation or transfer. If a card is not available, e.g. for online transactions or payments (some information is required, such as the CVV number, cardholder name, PIN, security query, etc.).

Accurate, quick and effective methods for the detection of credit cards have become a hot issue in recent investigation. Currently, the Bayes network algorithm is the most commonly used data mining algorithm for credit card fraud [2]. In credit card data leak prevention, there seems to be no genuine solution for ensuring the safety of the card if it is safely put in the pocket of the owner, or when it is used by a third party. The investigating organisation has been interested in the frauds by credit card, and a number of approaches have been recommended to detect theft.

2. Literature Survey

For multiple technologies, ml algorithms have been used to secure credit card payments. Fraudulent credit card data set tested the impact of the naives and regression models Fraudulent credit card dataset was assessed for the effectiveness of naive bays, k-near neighbour and logistic regression. A large part of
the progress that has been made with cyber retailers is with regard to marketing and other tactics that have the effect of attracting clients to a data center. In order to solve the data problem of information asymmetry and aim to evolutionize embedding capacity by two important algorithm techniques, a data capture system for credit cards using whale optimization and SMOTE (synthetic minority optimization technique) has been implemented. In many ways, the existing solution is unfeasible. Analysis of the transaction model is nothing short of an invasion of privacy and more about a system full of risks and misjudgments.

3. Proposed Approach:
This study has a number of steps to resolve the problem. Beginning with the data set to be used, the data set must be pre-processed.

![Figure 1 Proposed Flow](image)

The test and train data must be set before the application of the machine learning classifier. The classifier can be incorporated after configuration of the data. The last step was to determine the algorithms for machine learning. The steps are as described. The following explains Figure 1 in more detail. In subchapters each of the steps is explained.

A. Dataset

The information comes from Kaggle, which gathers data on the history of prior transactions. The data is stable and the classifications are good. The dataset includes 30 transaction records features and 1 label. Class Label: 1- Fraud, 0- Standard.

B. Preprocessing

As discussed earlier in this thread, our dataset consists of 30 features and even the evaluation of those lowest effective dose in the dataset. This method searches key features and selects all 30 features for training the machine using Machine learning models. Then split the data into test and train. Train the machine with 80% training data. This training data contains feature and label. Remain 20% split-up was used to test the machine learning model.

C. Model Selection

Machine Learning is known a part of Artificial Intelligence, involving the progress of approaches and practices for learning the computer. The analysis of the supervised learning algorithms in this study is performed. Here the supervised classifying algorithm is used to relate the model assessment [3].

1. Support Vector Machine

Support vector machines (SVM) are a series of supervised classification and regression learning methods. Support vectors are the nearest data points to the surface of decision (or hyperplane). Optimal hyperplane is derived from the lowest independent enhancements function class. Set hyperplanes H to the following in figure 2.
w•xi +b +1 for yi = +1 for w•xi +b = -1 for yi = -1

The planes H1 and H2 are:
H1: xi w• +b = +1
W•xi +b =-1 H2:

The support tips are the points of planes H1 and H2.

II. K-NN Algorithm

The nearest neighbouring algorithms belong to the "simplest" machines supervised and have been well-studied over the previous decade in the area of information processing [4]. Although neural networks are gaining wide acceptance of computer vision and machine learning, the cross-section between computer vision, pattern classification and biometrics is one of the most frequently used neighbouring models. KNN is a supervised learning algorithm that simply stores labelled workouts that show the training phase [5]. That is why kNN is also known as a lazy algorithm in learning in figure 3.

Random Forest Classifier

Random forests are an integration of decision trees, in which each tree is independently and equally distributed to all forest trees by the values of a random vector. The result of the growth of a group of trees and the vote for the most popular class was a considerable improvement in prediction performance.

1. In case b = B: 1.
(a) Draw samples from the training data for the bootstrap Z size N settings.
(b) Create a random-forest tree Tb to the data bootstrapped, repeating recursively the next steps of every tree terminal node until the minimum node size nmin is reached.

i. Choose m variable from the p variables randomly.
ii. Please select the best split/variable between m.

iii. Divide the node into two nodes of the daughter.

2. Output trees \( \{T_b\}_B \) ensemble.

For a new point \( x \) to predict:

Back: by \( B \text{rf}(x) = \prod_{b=1}^B T_b(x) \).

Classification: Let \( C\text{ing}(x) \) be the \( b \)th random-forest tree class prediction.

\( \text{CEB rf}(x) = \text{vote by majority } \{C\text{EB}(x)\}_B \)

4. Experiments And Result Analysis

In Python IDE, this work was carried out with Python. We have chosen to divide our primary data based on the exact results of various data proportions: 80 percent of total for learning and 20 percent of the total for the test phase in figure 4-6.

This study will examine the exact comparison of three classification algorithms. Upon completion of the classification, results of a comparison of precision show that Random Forest algorithm achieved high accuracy 94%. Besides, the results view will be displayed. In this study, to analyze the estimation accuracy, confusion matrix and ROC curve was used.

![Figure.4 Accuracy Comparison](image-url)
5. Conclusion
Fraud detection by credit card is a serious problem. Companies are therefore making significant investments in new algorithms to help identify and avoid suspicious charges. This article is intended to pinpoint business transactions financial fraud using machine algorithms. The results have shown that Random Forest is more effective than SVM and KNN using various methods such as the matrix of accuracy, recall, reliability, true positive rate and false positive rates. Our future research involves the construction of this model as a prototype. The prototype includes a highlight on how the exit solution is restricted and how this e-business framework has enough card holder authentication.

References
[1] A. O. Adewumi and A. A. Akinyelu, A survey of machine-learning and nature-inspired based credit card fraud detection techniques, Int. J. Syst. Assur. Eng. Manag., vol. 8, pp. 937–953, Nov. 2017.
[2] B. Lebichot, Y. A. Le Borgne, L. He-Guelton, F. Oblé, and G. Bontempi, Deep-learning domain adaptation techniques for credit cards fraud detection, in Proc. INNS Big Data Deep Learn. Conference, Genoa, Italy, 2019, pp. 78–88.
[3] J. O. Awoyemi, A. O. Adetunmbi, and S. A. Oluwadare, Credit card fraud detection using machine learning techniques: A comparative analysis, Proc. IEEE Int. Conf. Comput. Netw. Informatics, ICCNI 2017, vol. 2017-Janua, pp. 1–9, 2017.

[4] Haldorai, A. Ramu, and S. Murugan, Social Aware Cognitive Radio Networks, Social Network Analytics for Contemporary Business Organizations, pp. 188–202. doi:10.4018/978-1-5225-5097-6.ch010

[5] R. Arulmurugan and H. Anandakumar, Region-based seed point cell segmentation and detection for biomedical image analysis, International Journal of Biomedical Engineering and Technology, vol. 27, no. 4, p. 273, 2018.