Credit Fraud Detection using Machine Learning

1st Sachin Kumar Singh
Computer Science and Engineering
MIT ADT University
Pune, India

2nd Amritansh Srivastava
Computer Science and Engineering
MIT ADT University
Pune, India

3rd Shivendu Kumar
Computer Science and Engineering
MIT ADT University
Pune, India

4th Saurav Kumar
Computer Science and Engineering
MIT ADT University
Pune, India

5th Prof Kiran Bidua
Computer Science and Engineering
MIT ADT University
Pune,India

Abstract—Scam using a credit card is a simple and convenient case. E-commerce and many other paid internet sites have increased their online modes of operation, which increases the risk of online fraud. As the rate of fraud has increased, researchers have begun to use a variety of techniques used to train fraud detection and analysis in online transactions. The main goal is to project, develop and develop a new fraud detection method Streaming data about operations that are aimed at detailed analysis of the most recent customer operations and extracting data about people's behavior. This work was carried out by a dataset of credit card transactions entering the EU, a card containing 284,807 operations. In this project, we used an algorithm in an isolated forest for the purpose of classifying an observation. As a result, the model allows sample training to classify test select. Therefore, once the model is trained, we had to draw in the accuracy of the results obtained

Index Terms-Credit card fraud , isolation forest algorithm , local outlier , automated fraud detection, data science

1 INTRODUCTION

For a while there has been a strong interest in the ethics of banking, but also the spiritual difficulties of fraudulent behavior. Fraud for the purpose of obtaining services/goods and / or money is widespread in practice and is a growing problem in everyone in the modern world. Fraudsters dealing with cases that involve a criminal justice target are generally difficult to detect. Credit cards-one of the most well-known cases of fraud, but far from simple; fraud can occur in any type of line of credit, like: personal debts, mortgages, loans, and retail sales. In addition, the situation with fraud has changed over the past few decades, at least as technologies have changed and evolved. The most important task is to help companies and financial institutions like banks take fraud prevention and control measures in order to be effective and prompt when this happens . This is a very urgent problem that requires the attention of the learning and data science communities. The solution to this problem can be automated. This problem is particularly challenging from a learning perspective, which is characterized by a number of factors, such as: class imbalance. The total number of valid trades is even more misleading. In addition, model operations often change their statistical properties over a certain period of time. This is the only problem when using real fraud detection systems. In the real world, a great use of force is to meet quickly by scanned automated means to determine which operations will have made mistakes. Machine learning algorithms to use, analyze, all transaction confirmation and reporting in suspicious cases. These reports are reviewed by experts, contact the owner to check whether the transaction is original or deceptive. You give feedback, automated system both for training and updating the algorithm, so that we can ultimately improve the performance of the fraud detection system over time

II. RELATED WORK

A fraud that is illegal, or a criminal that aims to defraud a person from buying, financial, or personal gain. This is an intentional act that is contrary to law, regulation, or policy in order to obtain unauthorized financial gain. Many texts related to rain encountered or cheat discoveries in this area have already been published and -where to be discovered. A broad inquiry by most Clifton Phua and his colleagues showed that the domain employees 'methods include" program data mining, " automatic fraud detection, and adversarial testing. Others in the article by Suman, GJUS&T scientists, researchers at HCE present training methods such as supervised and unsupervised fraud, credit cards. Even though these methods and algorithms have had unexpected success, and in some areas -provide reliable and consistent solutions to fraud detection problems.

The same research areas are presented by Wen-Yu Fan N. A. Wang, and wherever they are used for mining acceptance, emission, mining and distance-from-sum algorithms to accurately predict fraudulent emulation operations and collect operational credit card data for certain commercial bank level. Outlier mining is the field of data mining, which, in principle,,is to use money both online and in this area. All that found is an object that is independent of the underlying system, i.e. these operations are wrong. They take attributes of customer behavior and values based on the feature attributes calculated that distance between the object's cost control and the feature, and then when a certain value is set. Unconventional methods, like a hybrid of data mining and sophisticated network classification algorithms for perceiving illegal copies of the actual map dataset, operations based on networks, an algorithm that allows you to create a copy of the sample deviation from a group has been demonstrated effectively, usually in a large-scale online operation environment. Attempts have been made to completely move away from the new aspect in this. Attempts to improve the registration process and interaction contacts, in case of a fraudulent transaction. In this case, a fraudulent transaction that is bəyənilmiş in the system will be alerted to the alert and the connection must be for failure of the current operation. An artificial genetic algorithm, from a strategy that sheds new light in this area, takes a different direction of cheating. It turned out that it is necessary to find fake operations correctly and minimize the number of false positives. Even if it was a problem related to the variable classification value of the correct classification.

III. THE PROPOSED METHOD CONSISTS IN

Credit card transactions are not known compared to previous transactions and customer service. This is the beginning of a line of many complex problems in the real world, if the concept of an operational problem [1]. This is the concept of a transaction, you can say that a variable, for example, changes over time and in an unpredictable way. These variables we need to provide a high level of information imbalance. The main goal of this study is to solve the problem of eliminating the concept of operations performed in the real world, according to a scenario. Table 1 1 shows that the basic functions are accepted when performing all operations

Attribute name	Description
Transaction Id	Identification number of a transaction
Cardholder Id	Unique Identification number of given to the cardholder
Amount	Amount transferred or credited in a particular transaction by the customer
Time	Details like time and date , to identify when the transaction was made
Label	To specify whether the transaction is genuine or fraudulent

IV.Description of the data set

Data set [11], which contains transactions made by cardholders, duration 2 days, two days, month In September 2013. There are only 284,807 operations out of which 492 want to say that 0.172% of operations are fraudulent transactions. This data set is very unbalanced. After surgery, giving the client information is the answer, in connection with the issue of confidentiality, therefore, the most features of the datasets must be transformed by using the key, component analysis (PCA). V1, V2, V3,..., V28-the agreement on PARTNERSHIP and cooperation between the EU and Russia (PCA) applied

S NO	Feature	Description
1	Time	Time in seconds to specify the elapses between the curent transaction and frist transaction
2	Amount	Transaction amount
3	Class	0-not fraud 1-fraud

RESULTS

The pin, which returns the number of false-positive detection and it is compared with the actual values. It used to be how to calculate the accuracy score and the accuracy of the algorithm. It is part of the data that we have used for a more rapid tests for 10% off the entire data set. The complete data set is also used to indicate the end of the the results can be printed out. These results, together with the classification of report the algorithm is not given, the output is as follows, with a grade of 0 this means that the transaction has been determined to be valid, the property is 1 it means that there was a fraud in the transaction. These results will be compared with the class of items, check it out

the false-positive rate. This will result in about 10% of the data that is being used for the purpose of

```
Isolation Forest
Number of Errors: 71
Accuracy Score: 0.99750711000316

              precision    recall  f1-score   support

           0       1.00      1.00      1.00     28432
           1       0.28      0.29      0.28        49

    accuracy                           1.00     28481
   macro avg       0.64      0.64      0.64     28481
weighted avg       1.00      1.00      1.00     28481


Local Outlier Factor
Number of Errors: 97
Accuracy Score: 0.9965942207085425

              precision    recall  f1-score   support

           0       1.00      1.00      1.00     28432
           1       0.02      0.02      0.02        49

    accuracy                           1.00     28481
   macro avg       0.51      0.51      0.51     28481
weighted avg       1.00      1.00      1.00     28481
```

Results with the complete dataset is used:

```
Isolation Forest
Number of Errors: 659
Accuracy Score: 0.9976861523768727

              precision    recall  f1-score   support

           0       1.00      1.00      1.00    284315
           1       0.33      0.33      0.33       492

    accuracy                           1.00    284807
   macro avg       0.66      0.67      0.66    284807
weighted avg       1.00      1.00      1.00    284807


Local Outlier Factor
Number of Errors: 935
Accuracy Score: 0.9967170750718908

              precision    recall  f1-score   support

           0       1.00      1.00      1.00    284315
           1       0.05      0.05      0.05       492

    accuracy                           1.00    284807
   macro avg       0.52      0.52      0.52    284807
weighted avg       1.00      1.00      1.00    284807
```

FUTURE ENHACEMENT:

At the same time, we were not able to reach the goal of 100% accuracy in deception the discovery is, in the end, we decided to create a system that can a lot of time, and is very close to the target. That is all in a project such as this, there is still much room for improvement here. The in this project, makes it possible for several of the algorithms an integrated together into modules, and the result can be this, combined with the precision and accuracy of the final result. This is a the model can also be improved by the addition of more algorithms. However, the output of these algorithms this has to be one of the same size as the others. As soon as the the condition is not met, the modules can easily be added, as is done in the of the code. There is a large body of work, and the versatility of the project. More room for improvement, which can be found in the data set. About shown as to the accuracy of the algorithm improves when the size of the data is also increased. Therefore, there will be no more information sure, the model will be more accurate in the detection of fraud and, in order to reduce the number of

false-positive results. This requires, however, that the with the official support of the bank.

Credit-Card-Fraud-Detection-Based Transactions The behavior of the John-Richard, D., Ho, He, A., Ale, edited by Proc. Of the 2017 IEEE Region 10 Conference (TENCON), In Malaysia, From 5 To 8 November, 2017

[2] THE NEW JERSEY PHUA1 HAN LEE1, KATE SMITH1 & ROSS GAYLER2, " A Comprehensive survey of Data Mining-based Fraud Detection Research", published by the school of business System, Department Of Information Technology, Monash The University, Wellington Road, Clayton, Victoria 3800, Australia

[3] a Survey of the Paper in a Credit card Fraud Detection Suman" as a scientist, GJUS&T, Hisar HCE, Sonepat published by the International Journal of Advanced Research in Computer Engineering & Technology (IJARCET) volume 3 issue 3, March-2014

[4] "Research on Credit card Fraud Detection Model Based on the Distance from the To sum it up, Wen-Fang-YU-Na Wang", which was published in april 2009 In international Joint Conference on Artificial Intelligence

[5] for Credit card Fraud Detection, with the help of a Parenclitic The Network AnalysisBy, Massimiliano Zanin, Miguel Romance, Regino Criado, and SantiagoMoral," which was published by Hindawi The complexity Volume 2018, Article ID 5764370, 9 pages

[6] For Credit Card Fraud Detection: A Realistic Modeling a book-to-Learn Approach, which is published by the IEEE. TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS, VOL. 29, FEB. AUGUST 8, 2018

[7] for Credit card Fraud Detection Ishu Trivedi, a Wife, a Mrigya, Mridushi', which was published in the International Journal of Advanced Research Computer and communication technology, Vol. 5, Issue 1, January 2016

[8] David J. Wetson,David J. Hand, " M, Adams,Whitrow, and After Jusczak plastic card Fraud Detection using Peer Group An Analysis Of The Running Of The 2008 Edition

CONCLUSION

Of course, the credit card fraud is the act of a crime of dishonesty. This article reviewed the recent findings in the credit card field. In this paper, it has been established that the various forms of fraud, such as: bankruptcy, fraud, forgery, fraud, identity theft, fraud, enforcement of the the fraud, and all fraud, and to take steps to encourage the they may be located. These measures included, in pairs, to match the decision trees, clustering techniques, neural networks, and in the genetic algorithm. From an ethical point of view, it can be said that the banks and the your credit card company to try to locate all of the fraudulent in some cases,. However, this is unprofessional He's not likely to work, at the level of the a professional swindler, and therefore the cost to the bank, and their the discovery is a not-for-profit. The bank would then be confronted with the this is an ethical dilemma. They should try to detect such a the fraudulent cases or, as they need to in order to act in the best interest of the shareholders and, in order to avoid a not-for-profit cost? As a next step in this research will be the focus of with the implementation of a "suspicious," the scorecard is correct the

data for the evaluation. The main responsibilities will be to build it lots of models for the prediction of the fraudulent conduct, with respect to the one area of behavior, which are related to the different types of credit card fraud is identified, in this article, and for the evaluation of the the associated ethical implications. It is intended to be one of the most In some european countries, and probably in Germany, and then to extend it research in other european countries. Credit card fraud is, without a doubt, an act of criminal sin. This article is a list of the most frequently the methods of fraud, as well as their methods of inquiry, and the reviewing the most recent advances in this field. In this paper, it has also in detail, it is explained how the learning can be applied in order to obtain better results in the discovery of the fraud, and, in conjunction with the algorithm, the the pseudo-code, implementation, and interpretation of the experiments and results of operations. As for the algorithm, which is accessible via a 99.6% accuracy in the the level of precision that only 28 per cent of the one-tenth the size of the data to be taken into account. However, if the complete set of data is to be fed into the algorithm, and the accuracy will increase to 33%. This is a high level of the percentage of the level of accuracy that can be expected as a result of the great the imbalance between the number to be valid, and some the real conclusion. For the entire data set consists of only two days." the transaction records are only a small fraction of the data that can be since this is a new project that could be used in a commercial scale. Be based on machine-learning algorithms, the program will start to increase in effectiveness over time, for more information it is to be put into it.

REFERENCES

Credit Card Fraud Detection Based on Transaction Behaviour -by John Richard D. Kho, Larry A. Vea" published by Proc. Of the 2017 IEEE Region 10 Conference (TENCON), Malaysia, November 5-8, 2017

[1] "Research on Credit Card Fraud Detection Model Based on Distance Sum – by Wen-Fang YU and Na Wang" published by 2009 International Joint Conference on Artificial Intelligence

[2] "Credit Card Fraud Detection through Parenclitic Network AnalysisBy Massimiliano Zanin, Miguel Romance, Regino Criado, and SantiagoMoral" published by Hindawi Complexity Volume 2018, Article ID 5764370, 9 pages

[3]   "Credit Card Fraud Detection: A Realistic Modeling and a Novel Learning Strategy" published by IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS, VOL. 29, NO. 8, AUGUST 2018 [7] "Credit Card Fraud Detection-by Ishu Trivedi, Monika, Mrigya, Mridushi" published by International Journal of Advanced Research in Computer and Communication Engineering Vol. 5, Issue 1, January 2016

[4]   David J.Wetson,David J.Hand,M Adams,Whitrow and Piotr Jusczak "Plastic Card Fraud Detection using Peer Group Analysis" Springer, Issue 2008.