WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Fraud Detection of Credit Cards

Aswin Babu,Liyo Luko Sam,Ambily Merlin Kuruvila

Saintgits College of Applied Sciences, Pathamuttom, Kottayam,686532, India

ABSTRACT

Credit card is a thin rectangular metal card used for purchasing goods with credit limits or withdraws cash in advance. Due to cashless transaction every people use ATM card and credit card for transaction, so fraud can also be increase. Credit cards frauds are increasing day by day. Frauds can take money easily from our account. In order to avoid this situation, credit card fraud detection techniques should be implemented. Machine Learning techniques are used to detect the credit card frauds. There are different fraud detection algorithms and also more advanced algorithm that relies on machine learning which are used to reduce these losses.

Keywords: Credit card, transaction, frauds, algorithm, machine learning

1. Introduction

We know that there are a variety of e-commerce sites. Now people are more invested in purchasing items through such e-commerce sites. They are also demanding online payments for buying their products. Payments like these can now be done through various online payment methods. As the people who are using online payments day by day, the more they use online websites to make these kinds of payments, there are more chances of getting cheated by fraud sites. We should be aware to make source payments.

The advantage of having machine learning credit fraud detection from conventional frauds are by using the help of machine learning. Using this machine learning frauds can easily be detected, real-time streaming is possible, verification methods can be done within a short period of time and also can identify hidden correlations in data

On the other hand, assembly of the fraud detection requires a great deal of time and also multiple verification methods are needed and they only mostly get obvious fraud activities. Fraud detection is a set of activities which are taken to prevent money from being obtained through false pretences. In ML based credit card fraud detection, if provide the best results in revealing and preventing fraudulent transactions.

Here, machine learning is a set of methods and techniques that let computers to recognize pattern and trends and generate prediction based on it. The machine Learning has many uses in everyday life for email spam detection, image detection, health industry etc.

2. Fraud Detection

Commerce and banking channels are increasing and providing consumers more ways to transact than ever before. From physical channels like credit, debit and pre-paid cards; checks; ATMs and point-of-sale terminals; to digital channels like ACH, wire, internet, telephone, mobile devices and crypto-currencies. Consumers, businesses, merchants, financial institutions and many others all benefit from anytime and anywhere commerce. Sophisticated fraud threats are multiplying even faster: malware and Trojans; account takeover and identify theft; credit abuse and bust-out scams; ACH and wire fraud; data breaches; money laundering and employee fraud. In fact, a single data breach can compromise tens of millions of account holders in a matter of seconds.

Fraud detection are the activities that are used to prevent money or property from being obtained through false or wrong pretences. These threats put consumers, business, merchants and bank at risk for sudden and huge potential financial losses. They slow commerce and inhibit business growth and damaged reputation customer trust and loyalty. The responsibility for their protection falls onto the hands of the banks and merchants and is compounded by mounting national and international regulations and compliance mandates. We must act fast to keep with the onslaught of threats, this means getting better and faster in order to thwart fraudulent transaction in real time.

By analysing these threats, we can assess typical and non-typical behavioural profiles; identify and address new threads immediately and at a sustainable cost while minimizing customer impact. Some companies also try to build their own kind of fraud management tools. These can be expensive. So there are so many fraud detection algorithm and is managed by fraud detection analyst.

2.1 Fraud detection techniques

Fraud detection can be found either by using the statistical analysis of data kinds of techniques or artificial intelligence (AI). The following are included in the use of statistical data analysis techniques:

- Calculating statistical parameters
- Regression analysis
- Probability distribution and their models.
- Data matching

The techniques used by the AI in order to detect frauds are:

- Data mining – A process used to group large raw data into useful data to find patterns and detect fraud.
- Neural network - A series of algorithm which is used to identify patterns. Using these patterns, we detect the fraud.
- Machine learning - A part of artificial intelligences that teach systems to learn, identify patterns etc.
- Pattern recognition - An automated identifying of patterns and regularities in data.

2.2 Different types of Fraud

Many fraud activities are increasing day by day. The most common fraud is the Credit Card frauds. In order to avoid such fraud methods, fraud detection techniques should be developed and implemented. Many fraud activities take place through online mode.
Healthcare, insurance and banking are the most common industries whose fraud activities are taking place. These industries deal with large amount of money and hence financial frauds are the most common faced by these industries. Apart from the above-mentioned fraud type there are a wide number of frauds such as email frauds, consumer frauds etc. we should be much more careful so as to not be dragged into these kinds of frauds.

3. Technologies Used

3.1. Machine Learning

Machine learning brings together statistics and computing to enable computers to be able to find out the way, to do a given task without being programmed to try to do so. Just as your brain uses experience to enhance a task so can computers. For example, we can have a computer that can tell the difference between a picture of dog and a cat. We can begin by feeding it images and telling it this one a dog that one is a cat. A computer programmed to learn will seek statistical pattern within the data that will enable it to recognize a cat or dog. It's the computer not the programmer that identifies those patterns and establishes the algorithm by which future data will be sorted. One example of a simple yet highly effective algorithm is to find the optimal line separating cats from dogs. When the computer sees a new picture, it checks which side of the line it falls on and they say either cat or dog. But of course, there can be mistakes. The more data the computer receives the more finely tune its algorithm becomes and the more accurate its predictions turn out to be. Machine learning is already widely applying. It's the technology behind facial recognition, text to speech recognition, spam filters on your inbox, online shopping or viewing recommendations and so much more. At the University of Oxford machine learning researches are combining statistics and computer science to build algorithms that can solve more complex programs more efficiently using less computing power.
Machine learning algorithms build mathematical model based on sample data, known as training data in order to make predictions and decisions. In order to be able, perform any task effectively without using instructions, patterns and interface methods, computer uses algorithms and static model. This study is known as machine learning.

3.1.1. Types of Machine Learning

- Supervised Learning

  Supervised learning uses labelled data to train machine learning models. Labelled data indicates that the output is already known to us. The model just needs to map inputs to outputs. An example for supervised learning is to train a machine that identifies the image of an animal

- Unsupervised Learning

  Unsupervised learning uses unlabelled data to train machine. Unlabelled data means there is no fixed output variable. The model learns from all the data discovers patterns and features that it has got from the data and then returns an output. An example to unsupervised learning technique that uses the images of vehicles to classify what type of a vehicle that is, either a bus or truck. So the model learn by identifying out the parts of that vehicle, like its length and width, the front and rear end covers roof hoods, the types of wheels used etc. Based on these the features the model classifies whether the vehicle is bus or a truck.

- Reinforcement Learning

  RL drains a machine in order to take suitable accents and maximize reward in particular situations. Here it uses an agent and also an environment so as to produce some actions and their rewards. The agent has both a start and an end state. But there might be different parts for reaching the end state. In this learning technique there is no predefined target variable. An example for the RL technique to train a machine to identify the shape of an object is to give a list of different objects such as square triangle, rectangle, triangle or a circle. With this the model tries to predict the form of the thing.

4. Applications Used

4.1 Logistic Regression

Logistic regression may be considered as a technique which is used for the traditional statistics also as machine learning. It is also quite similar to linear regression except that the logistic regression predicts whether something is True or False, instead of predicting something continuous like the size of an object. Mathematically speaking, a binary logistic model features a variable with mainly these 2 values: - Pass or Fail which has been represented by an indicator variable. The logistic function also called sigmoid function is used to predict population growth and carrying capacity of environment. Also, rather than fitting a line to the info logistic regression fits an s-shaped logistic function.

1 / (1 + e^-value)

Where e is the Euler's number or the EXP () function in our spread sheet and value is actual numerical value.

4.1.1 Sigmoid Function

To map values to probabilities that are predicted, we rather use sigmoid function. The value must lie between the range 0 and 1. We use sigmoid function in machine learning for mapping predicted values to probabilities.

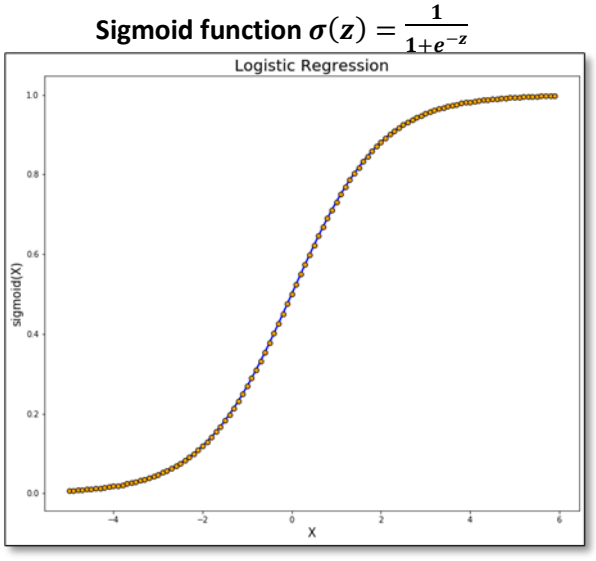


Fig 1 – Logistic Regression

4.1.2 Hypothesis Representation

When using logistics regression, we use a formula for hypothesis that is,

$$\sigma(Z) = \sigma(\beta_0 + \beta_1 X)$$

The expected values of our hypothesis lie between 0 and 1.

$$Z = \beta_0 + \beta_1 X \quad h\Theta(x) = \text{sigmoid}(Z)$$

i.e. $h\Theta(x) = 1/(1 + e^{-(\beta_0 + \beta_1 X)})$

$$h\theta(X) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 X)}}$$

Fig 2 – The hypothesis of logistic regression

4.1.3 Module

Module 1: Frame the problem.

Module 2: Collect the raw data.

Module 3: Import the libraries.

- pandas, numpy, sklearn etc.

Module 4: Process the data for analysis

- Perform the logistic regression on data
- Explore the data

Module 5: Perform in-depth analysis.

Module 6: Communicate results of the analysis.

4.2 Decision Tree

Decision tree is a tree structure that contains a root node, branches and leaf node. Each internal node in a tree denote test on an attribute, branch denotes the outcome of a test, and each leaf node in the tree denote the class labels. It does not require any domain knowledge and is easy to comprehend. Decision tree consists of two phases:

Tree construction in which all the training examples are at the root and all the partition examples recursively based on selected attributes.

Tree pruning which identify and remove outlier holding branches.

A machine learning researcher named J. Ross Quinlan developed a decision tree algorithm known as Iterative Dichotomiser (ID3). It adopt greedy algorithm and there is no backtracking. The attribute selection measure can be done in one of the following ways: information gain, gini index and gain ratio

4.2.1 Algortihm

- Tree is constructed in a top-down recursive divide-and-conquer manner.
- At start, all the training examples are at the root.
- Attributes are categorical.
- Examples are partitioned recursively based on selected attributes.
- Test attributes are selected on the basis of a heuristic or statistical measure.

4. Progress of Application

We import all the libraries and browse the information from kaggle. Then we discover the top of the information that has five rows and thirty one columns and print the outline of the information.

After commercialism the information we tend to explore the information and check whether or not there's any null worth gift or not. Dealings category distribution is summarized with regarding ninety nine of dealings marked as traditional and regarding zero.173% marked as deceitful.

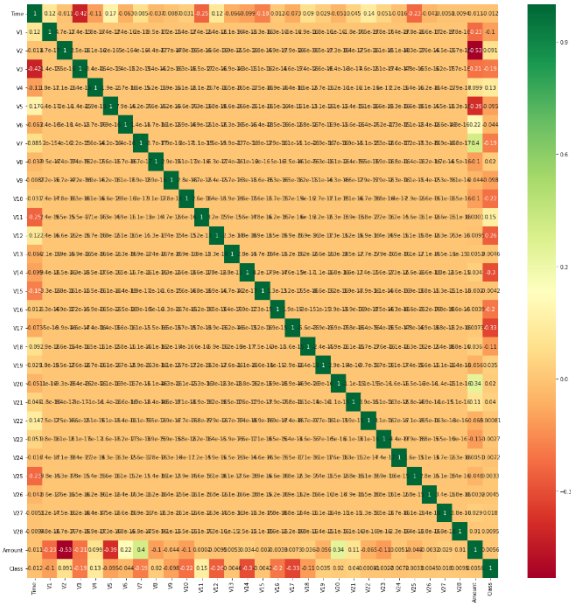
```
In [2]: import numpy as np
        import pandas as pd
        import sklearn
        import scipy
        import matplotlib.pyplot as plt
        import seaborn as sns
        from sklearn.metrics import classification_report,accuracy_score
        from sklearn.ensemble import IsolationForest
        from sklearn.neighbors import LocalOutlierFactor
        from sklearn.svm import OneClassSVM
        from pylab import rcParams
        rcParams['figure.figsize'] = 14, 8
        RANDOM_SEED = 42
        LABELS = ["Normal", "Fraud"]
```

```
In [3]: data = pd.read_csv('creditcard.csv',sep=',')
        data.head()
```

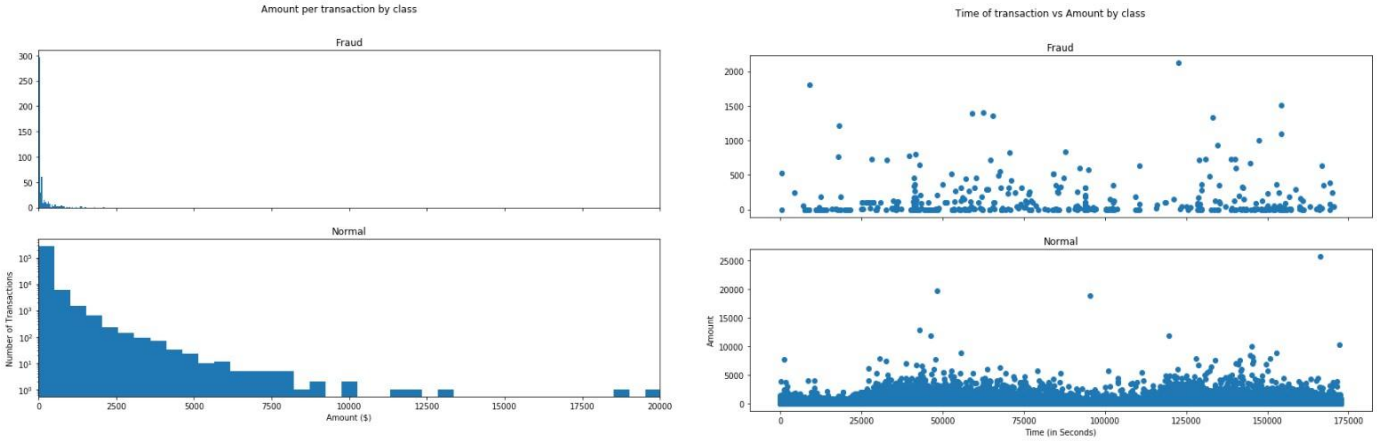
Out[3]:

	Time	V1	V2	V3	V4	V5	V6	V7	V8	V9	...	V21	V22	V23	V
0	0.0	-1.359807	-0.072781	2.536347	1.378155	-0.338321	0.462388	0.239599	0.098698	0.363787	...	-0.018307	0.277838	-0.110474	0.066
1	0.0	1.191857	0.266151	0.166480	0.448154	0.060018	-0.082361	-0.078803	0.085102	-0.255425	...	-0.225775	-0.638672	0.101288	-0.339
2	1.0	-1.358354	-1.340163	1.773209	0.379780	-0.503198	1.800499	0.791461	0.247676	-1.514654	...	0.247998	0.771679	0.909412	-0.689
3	1.0	-0.966272	-0.185226	1.792993	-0.863291	-0.010309	1.247203	0.237609	0.377436	-1.387024	...	-0.108300	0.005274	-0.190321	-1.175
4	2.0	-1.158233	0.877737	1.548718	0.403034	-0.407193	0.095921	0.592941	-0.270533	0.817739	...	-0.009431	0.798278	-0.137458	0.141

5 rows × 31 columns



Later the we tend to came to seek out the time dealings vs. quantity dealings and exploitation supplying regression we discover the correlation. The graphs are given below:



6. Advantages

1. Higher accuracy of fraud detection: Machine learning technologies take into account multiple information albeit embodies the littlest details of the information. Thus, they need higher exactness and come the foremost relevant values.

2. Less manual work required for extra verification: Machine Learning offers the analyst the relevant values so it'll enhance the accuracy of the information. this could reduce the burden of analyst.

3. Fewer false declines: once a system founds a legitimate dealings and wrong cancel it, during this state of affairs false declines or false positives happen.

4. Ability to spot new patterns and adapt to changes: Machine learning algorithms change analysts to spot new suspicious patterns and build new rules to forestall new forms of scams.

7. Conclusion

This paper reviewed about credit card fraud detection using machine leaning. In today's world fraud detection system has become an important essential in banks, institution especially financial institutions to minimize their loses. Credit card fraud detection system is capable for providing the most essential features to detect fraudulent and legitimate transaction. We have to update the dataset because in this day to day processing it becomes updated.

REFERENCES

https://searchsecurity.techtarget.com/definition/frauddetection#:~:text=Fraud%20detection%20is%20a%20set,or%20using%20stolen%20credit%20cards., https://www.investopedia.com/terms/d/datamining.asp#:~:text=Data%20mining%20is%20a%20process,increase%20sales%20and%20decrease%20costs.

https://www.investopedia.com/terms/n/neuralnetwork.asp#:~:text=A%20neural%20network%20is%20a,organic%20or%20artificial%20in%20na ture.

https://www.sas.com/en_in/insights/analytics/machinelearning.html#:~:text=Machine%20learning%20is%20a%20method,decisions%20with%20minimal%20human%20intervention.

https://en.wikipedia.org/wiki/Pattern_recognition#:~:text=Pattern%20recognition%20is%20the%20automated,computer%20graphics%20and%20machine%20learning. https://www.geeksforgeeks.org/machine-learning/