International Journal for Modern Trends in Science and Technology, 6(12): 24-27, 2020

ISSN: 2455-3778 online
DOI: https://doi.org/10.46501/IJMTST061205
Available online at: http://www.ijmtst.com/vol6issue11.html

Credit Card Fraud Detection System

Shashank Singh[1] | Meenu Garg[1]

[1]Information Technology, Maharaja Agrasen Institute Of Technology

To Cite this Article
Shashank Singh and Meenu Garg, "Credit Card Fraud Detection System", *International Journal for Modern Trends in Science and Technology*, 6(12): 24-27, 2020.

Article Info
Received on 06-November-2020, Revised on 18-November-2020, Accepted on 25-November-2020, Published on 29-November-2020.

ABSTRACT

It is essential that Visa organizations can distinguish false Mastercard exchanges so clients are not charged for things that they didn't buy. Such issues can be handled with Data Science and its significance, alongside Machine Learning, couldn't be more important. This undertaking expects to outline the demonstrating of an informational collection utilizing AI with Credit Card Fraud Detection. The Credit Card Fraud Detection Problem incorporates demonstrating past Visa exchanges with the information of the ones that ended up being extortion. This model is then used to perceive if another exchange is fake. Our target here is to identify 100% of the fake exchanges while limiting the off base misrepresentation arrangements. Charge card Fraud Detection is an average example of arrangement. In this cycle, we have zeroed in on examining and pre- preparing informational indexes just as the sending of numerous irregularity discovery calculations, for example, Local Outlier Factor and Isolation Forest calculation on the PCA changed Credit Card Transaction

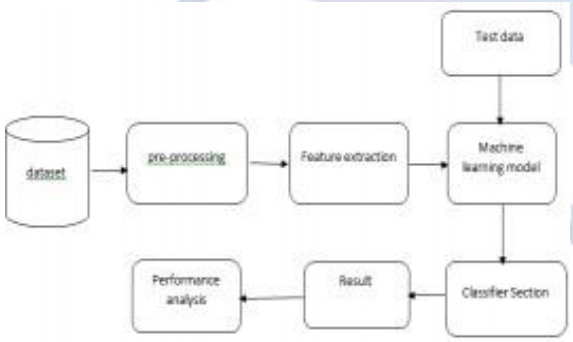
KEYWORDS: *Credit card· fraud detection ·applications of machine learning, data science, isolation forest algorithm, local outlier factor, automated fraud detection*

INTRODUCTION

'Fraud' in credit card transactions is unauthorized and unwanted usage of an account by someone other than the owner of that account. Necessary prevention measures can be taken to stop this abuse and the behaviour of such fraudulent practices can be studied to to minimize it and protect against similar occurrences in the future.In other words, Credit Card Fraud can be defined as a case where a person uses someone else's credit card for personal reasons while the owner and the card issuing authorities are unaware of the fact that the card is being used. Fraud detection involves monitoring the activities of populations of users in order to estimate, perceive or avoid objectionable behaviour, which consist of fraud, intrusion, and defaulting. This is a very relevant problem that demands the attention of communities such as machine learning and data science where the solution to this problem can be automated. This problem is particularly challenging from the perspective of learning, as it is characterized by various factors such as class imbalance. The number of valid transactions far outnumber fraudulent ones. Also, the transaction patterns often change their statistical properties over the course of time.

These are not the only challenges in the implementation of a real-world fraud detection system, however. In real world examples, the massive stream of payment requests is quickly scanned by automatic tools that determine which transactions to authorize. Machine learning algorithms are employed to analyse all the authorized transactions and report the suspicious ones. These reports are investigated by

professionals who contact the cardholders to confirm if the transaction was genuine or fraudulent. The investigators provide a feedback to the automated system which is used to train and update the algorithm to eventually improve the fraud-detection performance over time. Risk assessment is widely used at banks around the globe. Because credit risk assessment is very important, risk rates are evaluated using a variety of techniques. Banks group clients by profile. During assessment the financial history of clients and subjective consumer considerations are evaluated. Those figures are objective, which reflect the financial statements of the company. Detection of fraud involves monitoring and analysing the behaviour of different users in order to estimate detection unwanted behaviour. To effectively detect credit card fraud, we want to know the diverse technologies, algorithms and types involved in detecting credit card fraud. There are various algorithms to detect the credit card fraud and each have their own advantages and accuracy the algorithms are:-K-nearest neighbour, Linear regression, Ada Boost, Naive Bayes, J48, Logistic Regression, Random Forest algorithm etc. The null hypothesis is the credit card transaction is correct and not fraud. Hence false positive is whether it is a correct and genuine transaction and therefore the system model predicts it as fraud transaction and raises a warning .This means completely normal customers looking to form a sale would deter faraway from making purchases. False negative is a serious issue as the transaction is fraudulent and the system model predicts it as non-fraudulent. In our case, a false negative is far more serious than false positive as our system model would prove costly if it predicts fraudulent transactions as genuine



LITERATURE REVIEW

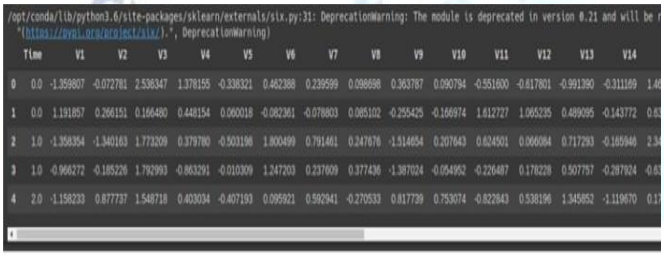
Fraud act as the unlawful or criminal deception intended to result in financial or personal benefit. It is a deliberate act that is against the law, rule or policy with an aim to attain unauthorized financial benefit. Numerous literatures pertaining to anomaly or fraud detection in this domain have been published already and are available for public usage. A comprehensive survey conducted by Clifton Phua and his associates have revealed that techniques employed in this domain include data mining applications, automated fraud detection, adversarial detection. In another paper, Suman, Research Scholar, GJUS&T at Hisar HCE presented techniques like Supervised and Unsupervised Learning for credit card fraud detection. Even though these methods and algorithms fetched an unexpected success in some areas, they failed to provide a permanent and consistent solution to fraud detection. A similar research domain was presented by Wen-Fang YU and Na Wang where they used Outlier mining, Outlier detection mining and Distance sum algorithms to accurately predict fraudulent transaction in an emulation experiment of credit card transaction data set of one certain commercial bank. Outlier mining is a field of data mining which is basically used in monetary and internet fields. It deals with detecting objects that are detached from the main system i.e. the transactions that aren't genuine. They have taken attributes of customer's behaviour and based on the value of those attributes they've calculated that distance between the observed value of that attribute and its predetermined value. Unconventional techniques such as hybrid data mining/complex network classification algorithm is able to perceive illegal instances in an actual card transaction data set, based on network reconstruction algorithm that allows creating representations of the deviation of one instance from a reference group have proved efficient typically on medium sized online transaction. There have also been efforts to progress from a completely new aspect. Attempts have been made to improve the alertfeedback interaction in case of fraudulent transaction. In case of fraudulent transaction, the authorised system would be alerted and a feedback would be sent to deny the ongoing transaction.

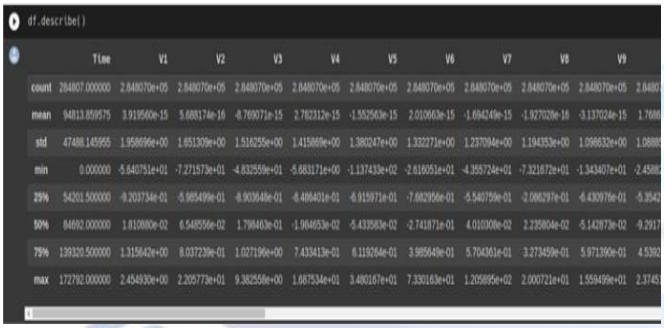
Artificial Genetic Algorithm, one of the approaches that shed new light in this domain, countered fraud from a different direction. It proved accurate in finding out the fraudulent transactions and minimizing the number of false alerts. Even though, it was accompanied by

classification problem with variable misclassification costs.

METHOD

In this project we will use various predictive models to see how accurate they are in detecting whether a transaction is a normal payment or a fraud. As described in the dataset, the features are scaled and the names of the features are not shown due to privacy reasons. Nevertheless, we can still analyze some important aspects of the dataset. The datasets contain transactions made by credit cards in September 2013 by european cardholders. This dataset presents transactions that occurred in two days, where we have 492 frauds out of 284,807 transactions. The dataset is highly unbalanced, the positive class (frauds) account for 0.172% of all transactions.





It contains only numeric input variables which are the result of a PCA transformation. Unfortunately, due to confidentiality issues, we cannot provide the original features and more background information about the data. Features V1, V2, ... V28 are the principal components obtained with PCA, the only features which have not been transformed with PCA are 'Time' and 'Amount'. Feature 'Time' contains the seconds elapsed between each transaction and the first transaction in the dataset. The feature 'Amount' is the transaction Amount, this feature can be used for example-dependant cost- sensitive learning. Feature 'Class' is the response variable and it takes value 1 in case of fraud and 0 otherwise.

Challenges with Artificial Neural Network (ANN) While solving an image classification problem using ANN, the first step is to convert a 2-dimensional image into a 1-dimensional vector prior to training the model. This has two Basic work of theme finalise Software Setup

Little Brief about tools

Task 1: Basic of Data Science Why use, when use, where use in detail.

Understanding our data - Gather Sense of our data

Task 1: Learn about How Visualize Data, Libraries used for Visualization (Matplotlib, Numpy, Pandas)

Task 2: Learn about Linear Algebra, Statistics and Probability (Basics)

Task: Learn about getting data(Web Scraping, Reading files), working with data(Cleaning, munging, manipulating data, rescaling, dimensionality reduction)

Preprocessing

Scaling and Distributing Splitting the Data

Random UnderSampling and Oversampling Distributing and Correlating

Anomaly Detection

Dimensionality Reduction and Clustering (t-SNE)

METHODOLOGY

In credit card transactions, various fraudulent activity detection techniques have been implemented in the minds of researchers to methods for developing models based on artificial intelligence, data mining, fuzzy logic, and machine learning. Apps are installed within fraudulent sample data sets. These data points, include customers name, customers age and value of the customer account, and origin of the credit card. So, with regard to card fraud, if the usage of cards to commit fraud are proven to be more, the fraud of a transaction using a credit card will be equally so, but if this were to decrease, the level of contribution would be equal. Detection and prevention of credit card fraud using Machine Learning is accomplished by using the classification and regression algorithms. We use supervised machine learning algorithms such as Random Forest Algorithms to detect online or offline fraud card

The Confusion Matrix:

Here is again, how the confusion matrix works:

Upper LeW Square: The amount of correctly classified by our model of no fraud transactions.

Upper Right Square: The amount of incorrectly classified transactions as fraud cases, but the actual label is no fraud .

Lower LeW Square: The amount of incorrectly classified transactions as no fraud cases, but the actual label is fraud .

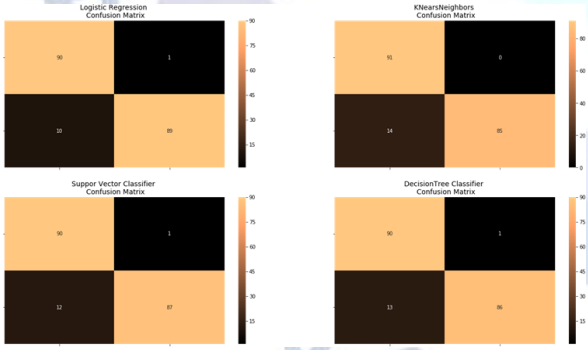
Lower Right Square: The amount of correctly classified by our model of fraud transactions. Summary (Keras || Random UnderSampling)

Dataset: In this final phase of testing we will fit this model in both the random undersampled subset and oversampled dataset (SMOTE) in order to predict the final result using the original dataframe testing data.

Neural Network Structure: As stated previously, this will be a simple model composed of one input layer (where the number of nodes equals the number of features) plus bias.

Other characteristics: The learning rate will be 0.001, the optimizer we will use is the AdamOptimizer, the activation function that is used in this scenario is "Relu" and for the final outputs we will use sparse categorical cross entropy, which gives the probability whether an instance case is no fraud or fraud (The prediction will pick the highest probability between the two.

EXPERIMENTAL RESULTS



CONCLUSION

Cedit card fraud is without a doubt an act of criminal dishonesty. This article has listed out the most common methods of fraud along with their detection methods and reviewed recent findings in this field. This paper has also explained in detail, how machine learning can be applied to get better results in fraud detection along with the algorithm, pseudocode, explanation its implementation and experimentation results. While the algorithm does reach over 99.6% accuracy, its precision remains only at 28% when a tenth of the data set is taken into consideration. However, when the entire dataset is

fed into the algorithm, the precision rises to 33%. This high percentage of accuracy is to be expected due to the huge imbalance between the number of valid and number of genuine transactions.Food Monitoring plays a leading role in health-related problems, it is becoming more essential in our day-to-day lives. Since people are dependent on smart technologies, provision of an application to automatically monitor the individuals diet, helps in many aspects.

REFERENCES

[1] "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

[2] CLIFTON PHUA1, VINCENT LEE1, KATE SMITH1 & ROSS GAYLER2 " A Comprehensive Survey of Data Mining-based Fraud Detection Research" published by School of Business Systems, Faculty of Information Technology, Monash University, Wellington Road,

[3] "Survey Paper on Credit Card Fraud Detection by Suman" , Research Scholar, GJUS&T Hisar HCE, Sonepat published by International Journal of Advanced Research in Computer Engineering & Technology (IJARCET) Volume 3 Issue 3, March 2014

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

[5] 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

[6] 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

[7] 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