IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

CREDIT CARD FRAUD DETECTION SYSTEM

[1]Shashank Singh [2]Meeenu Grag

[1]Student, [2]Teacher
[1]Information Technology,
[1]Maharaja Agrasen Institute of Technology, Delhi, India

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

Index Terms - Credit card· fraud detection ·applications of machine learning, data science, isolation forest algorithm, local outlier factor, automated fraud detection

I. INTRODUCTION

I.

'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 implementa2on of a real-world fraud detec2on system, however. In real world examples, the massive stream of payment requests is quickly scanned by automa2c tools that determine which transac2ons to authorize. Machine learning algorithms are employed to analyse all the authorized transac2ons and report the suspicious ones.

II. LITERATURE REVIEW

Fraud act as the unlawful or criminal deception intended to result in financial or personal benefit. It isdeliberate 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 thatshed 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 classifica2on problem with variable misclassifica2on costs .

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. Banksgroup clients by profile. During assessment the financial history of clients and subjec2ve consumer considera2ons are evaluated. Those figures are objec2ve, which reflect the financial statements of the company. Detec2on of fraud involves monitoring and analysing the behaviour of different users in order to es2mate detec2on unwanted behaviour. To effec2vely detect credit card fraud, we want to know the diverse technologies, algorithms and types involved in detec2ng 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, Logis2c Regression, Random Forest algorithm etc. The null hypothesis is the credit card transac2on is correct and not fraud. Hence false posi2ve is whether it is a correct and genuine transac2on and therefore the system model predicts it as fraud transac2on and raises a warning .This means completely normal customersn looking to form a sale would deter faraway from making purchases. False nega2ve is a serious issue as the transac2on is fraudulent and the system model predicts it as non-fraudulent. In our case, a false negatve is far more serious than false posi2ve as our system model would prove costly if it predicts fraudulent transactions as genuine.So this concludes the litervature review of credit. Financial fraud is an ever growing menace with far reaching consequences in the finance industry, corporate organizations, and government. Fraud can be defined as criminal deception with intent of acquiring financial gain. High dependence on internet technology has enjoyed increased credit card transactions. As credit card transactions become the most prevailing mode of payment for both online and offline transaction, credit card fraud rate also accelerates. Credit card fraud can come in either inner card fraud or external card fraud. Inner card fraud occurs as a result of consent between cardholders and bank by using false identity to commit fraud while the external card fraud involves the use of stolen credit card to get cash through dubious means. A lot of researches have been devoted to detection of external card fraud which accounts for majority of credit card frauds. Detecting fraudulent transactions using traditional methods of manual detection is time consuming and inefficient, thus the advent of big data has made manual methods more impractical. However, financial institutions have focused attention to recent computational methodologies to handle credit card fraud problem.

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

III. DATASET

In this project, we are testing 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.

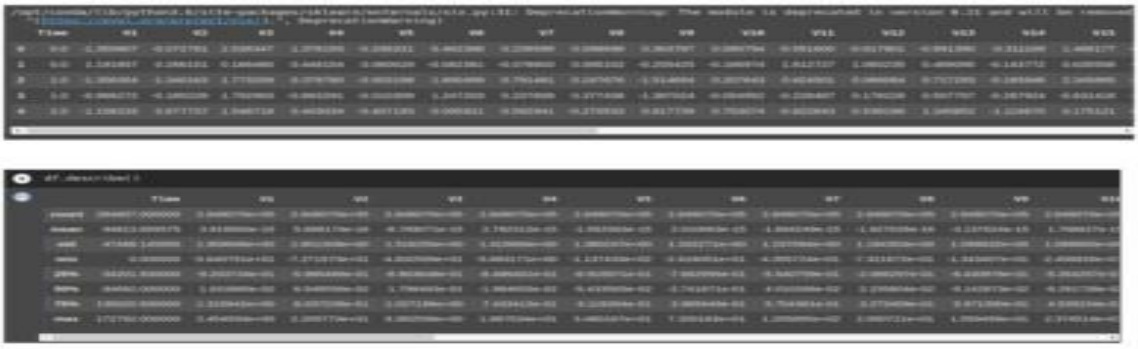


Figure 1: Dataset

IV. METHODS

LOGISTIC REGRESSIONS

In statistics, the **logistic model** (or **logit model**) is used to model the probability of a certain class or event existing such as pass/fail, win/lose, alive/dead or healthy/sick. This can be extended to model several classes of events such as determining whether an image contains a cat, dog, lion, etc. Each object being detected in the image would be assigned a probability between 0 and 1, with a sum of one.

Logistic regression is a statistical model that in its basic form uses a logistic function to model a binary dependent variable, although many more complex extensions exist. In regression analysis, **logistic regression**[1] (or **logit regression**) is estimating the parameters of a logistic model (a form of binary regression). Mathematically, a binary logistic model has a dependent variable with two possible values, such as pass/fail which is represented by an indicator variable, where the two values are labeled "0" and "1". In the logistic model, the log-odds (the logarithm of the odds) for the value labeled "1" is a linear combination of one or more independent variables ("predictors"); the independent variables can each be a binary variable (two classes, coded by an indicator variable) or a continuous variable (any real value). The corresponding probability of the value labeled "1" can vary between 0 (certainly the value "0") and 1 (certainly the value "1"), hence the labeling; the function that converts log-odds to probability is the logistic function, hence the name. The unit of measurement for the log-odds scale is called a *logit*, from ***log****istic un****it***, hence the alternative names. Analogous models with a different sigmoid function instead of the logistic function can also be used, such as the probit model; the defining characteristic of the logistic model is that increasing one of the independent variables multiplicatively scales the odds of the given outcome at a *constant* rate, with each independent variable having its own parameter; for a binary dependent variable this generalizes the odds ratio.

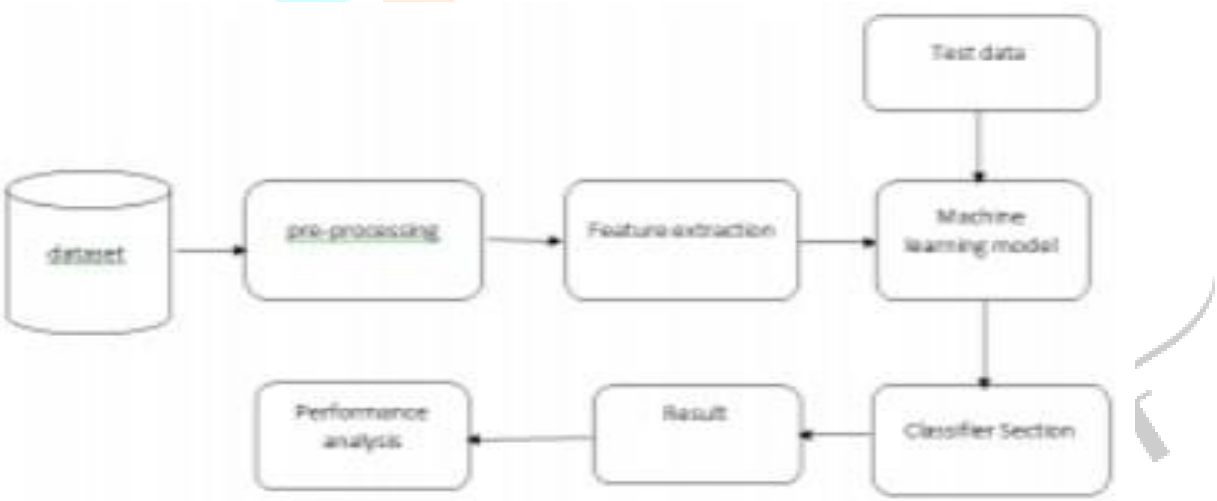
XGBOOST ALGORIHM

XGBoost or extreme gradient boosting is one of the well-known gradient boosting techniques(ensemble) having enhanced performance and speed in tree-based (sequential decision trees) machine learning algorithms. XGBoost was created by Tianqi Chen and initially maintained by the Distributed (Deep) Machine Learning Community (DMLC) group. It is the most common algorithm used for applied machine learning in competitions and has gained popularity through winning solutions in structured and tabular data. It is open-source software. Earlier only python and R packages were built for XGBoost but now it has extended to Java, Scala, Julia and other languages as well. XGBoost algorithm was developed as a research project at the University of Washington. Tianqi Chen and Carlos Guestrin presented their paper at SIGKDD Conference in 2016 and caught the Machine Learning world by fire. Since its introduction, this algorithm has not only been credited with winning numerous Kaggle competitions but also for being the driving force under the hood for several cutting-edge industry applications. As a result, there is a strong community of data scientists contributing to the XGBoost open source projects with ~350 contributors and ~3,600 commits on GitHub. The algorithm differentiates itself in the following ways:

1. A wide range of applications: Can be used to solve regression, classification, ranking, and user-defined prediction problems.
2. Portability: Runs smoothly on Windows, Linux, and OS X.
3. Languages: Supports all major programming languages including C++, Python, R, Java, Scala, and Julia.
4. Cloud Integration: Supports AWS, Azure, and Yarn clusters and works well with Flink, Spark, and other ecosystems.

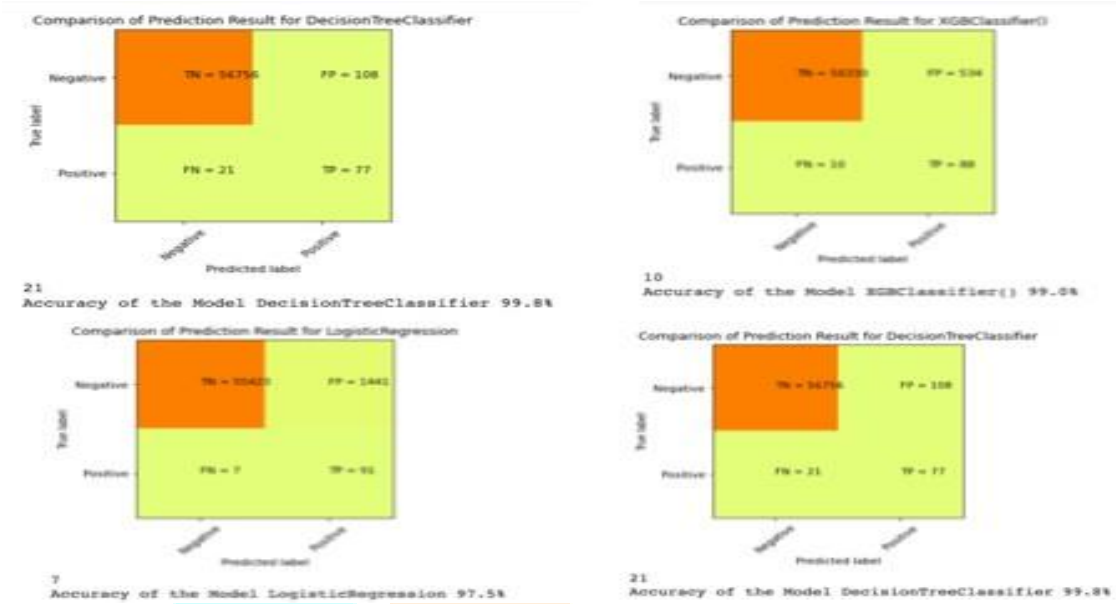
V. METHODOLOGY

In credit card transac2ons, various fraudulent ac2vity detec2on techniques have been implemented in the minds of researchers to methods for developing models based on ar2ficial 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 transac2on using a credit card will be equally so, but if this were to decrease, the level of contribu2on would be equal. Detec2on and preven2on of credit card fraud using Machine Learning is accomplished by using the classifica2on and regression algorithms. We use supervised machine learning algorithms . 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

VI. EXPERIMENTAL RESULT



VII. 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

II. REFERENCES

1. "Credit Card Fraud Detection Based on Transaction Behavior –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 Scolar, GJUS&T Hisar HCE, Sonepat published by International Journal of Advanced Research in Computer Engineering & Technology (IJARCET) Volume 3, March 2014

4. David J.Wetson, DavidJ.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 Analysis By 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