International Journal of Engineering Research and Science & Technology

IJERST

ISSN : 2319-5991

www.ijerst.com

Email: editor@ijerst.com or editor.ijerst@gmail.com

International Journal of Engineering Research and Science & Technology
ISSN 2319-5991 www.ijerst.com
Vol. 20, Issue 4, 2024

CREDIT CARD FRAUD DETECTION SYSTEM USING MACHINE LEARNING

MERAVATH SUNIL, Mr. Thirupathi Rao
UG Student, Department of Electronics and Computer Engineering, JBIET, India.
Assistant professor, Department of Electronics and Computer Engineering, JBIET, India.

ABSTRACT

In recent years credit card became one of the essential parts of the people. Sudden increase in E-commerce, customer started using credit card for online purchasing therefore risk of fraud also increases. Instead of carrying a huge amount in hand it is easier to keep credit cards. But nowadays that too becomes unsafe. Now a days we are facing a big problem on credit card fraud which is increasing in a good percentage. The main purpose is the survey on the various methods appliedto detect credit card frauds. From the abnormalities, in the transaction, the fraudulent one is identified. We address this issue in order to implement some machine learning algorithm like random forest, logistic regression in order to detect this kind of fraud. In this paper we increase the efficiency in finding the fraud. However, we discussed and evaluated employee criteria. Currently, the issues of credit card fraud detection have become a big problem for new researchers. We implement an intelligent algorithm which will detect all kind offraud in a credit card transaction. We handled the problem by finding a pattern ofeach customer in between fraud and legal transaction. Isolation Forest Algorithmand Local Outlier Factor are used to predict the pattern of transaction for each customer and a decision is made according to them. In order to prevent data frommismatching, all attribute are marked equally.

INTRODUCTION

Nowadays as we can see that there is a huge increase online payment and the payment is mostlydone with the help of credit cards. It becomes a big problem for marketing company to overcome with the credit card fraudulent activities. Fraudulent can be done in many ways suchas tax return in any other account, taking loans with wrong information etc. Therefore, we needan efficient fraudulent detection model to minimize fraudulent activity and to minimize their losses. There are a huge number of new techniques which provide different algorithms which help in detecting number of credit card fraudulent activity. Basic understanding of these algorithms will help us in making a significant credit card fraudulent detection model. This paper helps us in finding doubtful credit card transaction by proposing a machine learning algorithms. Credit Card Fraudulent detection comes under machine learning, and the objective is to reduce such type of fraudulent activity. This type of fraud is happening from past, and tillnow not much research has done here in this particular area. The types of credit fraud in transactions are bankruptcy fraud, behavioral fraud, counterfeit fraud, application fraud [3].

.There are experiments done before on credit card fraudulent activity on basis of meta-learning.There is certain limit of meta-learning. There are two features which is introduced here in ourreport is True

International Journal of Engineering Research and Science & Technology
ISSN 2319-5991 www.ijerst.com
Vol. 20, Issue 4, 2024

Positive and False alarm. Both these features play an important role in catchingfraudulent because the rate of determining fraudulent behavior is quick. For the better performance of model, we need a better classifier. Different classifier can be combined togetherwith help of meta-learning.

LITRATURE SURVEY

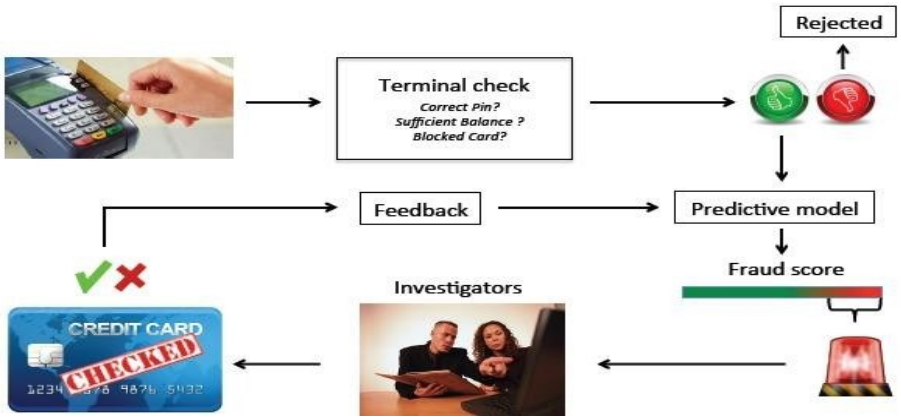
In our paper we referred to various papers for improving the performance of routing, reduce delay of information, reduce packet loss rate, reduce link failure, to improve packet delivery rate, to reduce energy consumption. There are a huge number of new techniques which providedifferent algorithms which help in detecting number of credit card fraudulent activity. Basic understanding of these algorithms will help us in making a significant credit card fraudulent detection model. This paper helps us in finding doubtful credit card transaction by proposing amachine learning algorithms. There are two features which is introduced here in our report is True Positive and False alarm. Both these features plays an important role in catchingfraudulent because the rate of determining fraudulent behavior is quick. As per today's Network plays an important role therefore it is mandatory for our models to be up to date to perform better detection capabilities. Whenever new fraudulent activity are detected then our model should be that much better to perform real time analysis. Other than traditional machine learning methods Fraudulent Detection System has been achieved through using Neural Networks [5]. To prevent personal information has become a huge task for financial company because there are a lot of attack on the system to steal someone personal information to performfraudulent. Our model has two essential feature which will help in finding abnormal behavior in form of charts for different column such as time, amount etc.

CREDIT CARD FRAUDULENT DETECTION SYSTEM

Unusual pattern which is known as outliers which not fulfill the expected behaviors is known as Anomaly detection. Many business applications are based upon this technique, unusual patterns in network are identified. Its helps in detecting credit card fraudulent as well as operating system fraudulent. Jupiter notebook we are going to take the credit card fraud detection as the case study so that we can understand the concept in detail. Outlier value is those value which shows an abnormal behavior from its neighbor or we can say that from standard data point. Generally, Outlier data termed as fraudulent transaction. Our experiment based upon catching fraudulent activity with the help of false alarm. Our model has focused onthe use of Isolated Random Forest and Local Outlier Factor, however previous works has alsobeen done using Bayesian Regularization and Gradient Descent Adaptive learningalgorithms[4]. There are many advantages of this system and one of the major advantage that we are recognizing the pattern and on the basis of pattern we made chart which will help to understand the fraudulent easily because it is easy to understand the data in the form of chart. We have plotted the chart for every features from V1 to V28. We should also keep the things in mind that other financial bank cannot read the other personal information. We can use this technique to find the scheme which will help in finding credit card fraudulent transaction. Theadvantages of this system is that it can work

BLOCK DIAGRAM

International Journal of Engineering Research and Science & Technology
ISSN 2319-5991 www.ijerst.com
Vol. 20, Issue 4, 2024



Block Diagram

METHODOLOGY

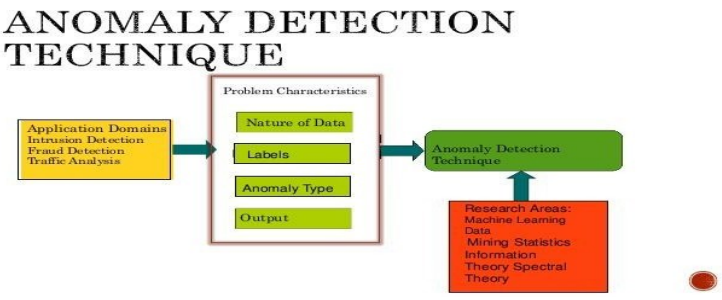
The task which is performed for the prediction of transaction and labelled as fraud is detectedon the basis of binary classification. We make two class for the prediction of fraud: class 0 andclass 1.

Class 0 if there is no fraud and class 1 to catch the fraud. This can be done with the help of binary classification.

WHAT ARE ANOMALIES?

Anomalies can be categorized as following:

- Point Anomalies: Point anomaly is a single instance of data. The credit card fraudulent detection technique is based on “amount spend”.
- Contextual Anomalies: The best example of contextual anomaly is time-series data.
- Collective Anomalies: Here, Detection of anomaly is based on a set of data instances collectively. Therefore, a set of data will help in detecting fraudulent anomaly. If someone try to theft personal data from server it will come under collective anomaly and named as cyber-attack.



Anomaly Detection Technique

LEARNING INTRODUCTION OF MACHINE

International Journal of Engineering Research and Science & Technology
ISSN 2319-5991 www.ijerst.com
Vol. 20, Issue 4, 2024

AI is a mechanism which features algorithms and calculations based on a normal human intelligence to address a problem. The AI behaves and approaches a problem in a similar waythat a normal human brain would. Its working mechanism is influenced by human thinking. Acollection of expectation and result is achieved by AI by portraying information in a form termed as 'test information' without making use of any predetermined models or being trainedin that particular domain. Problems catering to non-related dimensions such as email sifting, PC vision, location of system gate crashers are addressed. Thus it is assertive that it is not possible to train an AI to address a particular domain, instead an AI trained with general problem solving abilities, builds up its own algorithms for a set of problems.

An AI engine is allocated with responsibility of prediction or analysis using a PC framework and set of data. For this an AI engine is allocated with packages of scientific methods, logisticcalculations, data sets and knowledge about the field of the problems for performing. At the initial stage AI makes use of various algorithm to perform exploratory analysis for marking outvarious features of the given problem. Information mining is one of the necessary tool used byvarious AI models for this purpose. Moreover, the entire operation of AI is carried based on unsupervised learning model which leaves a very less room for training a robust AI for only aproblem specific solution. However, for business purposes modifications are performed beforeits application.

SYSTEM SPECIFICATION

The necessity for the most part dependent on two classes: they are practical portray every singlerequired usefulness for framework administrations which are given by the customers. Non useful necessities characterize the framework properties and compels. The equipment prerequisites indicate the equipment functionalities and required speed and limit of the fringe. The product prerequisites incorporate programming expected to create and run the framework.

HARDWARE SPECIFICATION

- System - Core i5
- Mobile - Android
- Monitor - RGB Colour
- Hard Disk - 2 TB
- Mouse - Microsoft
- Ram - 8GB

SPECIFICATION OF THE SOFTWARE

- Operating system - Win 10
- Dataset - csv
- Language - Python

SOFTWARES USED

- Python 3.5

International Journal of Engineering Research and Science & Technology

ISSN 2319-5991 www.ijerst.com

Vol. 20, Issue 4, 2024

GENERAL

- NumPy 1.11.3
- Matplotlib 1.5.3
- Pandas 0.19.1
- Seaborn 0.7.1
- SciPy
- Scikit-learn 0.18.1

DESIGN ENGINEERING

The UML is used for business and production based works. The task of using UML is to provide a solution or working of a product or model using visual representation. UML involvesusage of lock diagrams and flow chart to depict the interrelation and workflow of a model. Sometimes it is also used for planning purposes or analysis as a reference for further development of a project

- Provides direction with regards to the requests of the group exercise.
- Software ancient rarities create.
- Directs of errand to individual designers and group.
- Offer the criteria to check & estimate the task's item & exercise.

IMPLEMENTAION

Implementation phase brings out the design tweaked out into a operational system. Hence thiscan be deliberated to be most precarious juncture in accomplishing the efficacious system andin convincing the user faith that system will operate and be effective. This phase encompassesvigilant planning & design, examination of prevailing system and constraints on execution, design & scheming of methods to change over.

PROCEDURE FOLLOWED DURING IMPLEMENTATION

The application – Credit Card Fraud Detection which is in itself the complete & full-fledged GUI enabled application to envisage/foresee the authenticity & legitimacy of a transaction has been implemented, as per the following steps:

- Install Anaconda from an reliable source.

- Import packages: pandas, Scipy, Matplotlib, Seaborn

International Journal of Engineering Research and Science & Technology

ISSN 2319-5991 www.ijerst.com

Vol. 20, Issue 4, 2024

- Load the dataset, a dataset is the pool of data for analytical/critical purpose, a (.CSV)file.
- Reconnoiter and get through the dataset through data. shape, data. describe.
- Split the dataset into training dataset and testing dataset.
- Plot histogram of the dataset to epitomize/depict numerical data.
- Determine the count of fraud cases by checking if class is 0 or 1.
- In the similar procedure, get the correlation matrix.
- Next, there is a need to determine the local outlier factor.
- This is followed by use of random forest algorithm to find accurate results.
- The GUI is developed using PyQt library.
- The PyQt library, provides tools to achieve a complete GUI enabled application,similar to swings in java environment.
- Define the constructor in the file.
- Write down the entire implementation inside, thus encapsulating everything inside aGUI-enabled python file.

DATASET DESIGN

The dataset holds information about credit card transactions which has been made in a span oftwo days. The number of frauds have been calculated as 492 out of 284,807 transactions. The details have been given in form of positive and non-positive numerical values. The dataset contains 31 features which has been labelled as V1-V28 due to confidential reasons. The feature which has been reveled are Time and Amount of transaction. Here time denotes the number of seconds elapsed from the first transaction of Day 1. Amount of transaction consistsof positive value denoting deposit and non-positive value denoting

International Journal of Engineering Research and Science & Technology

ISSN 2319-5991 www.ijerst.com

Vol. 20, Issue 4, 2024

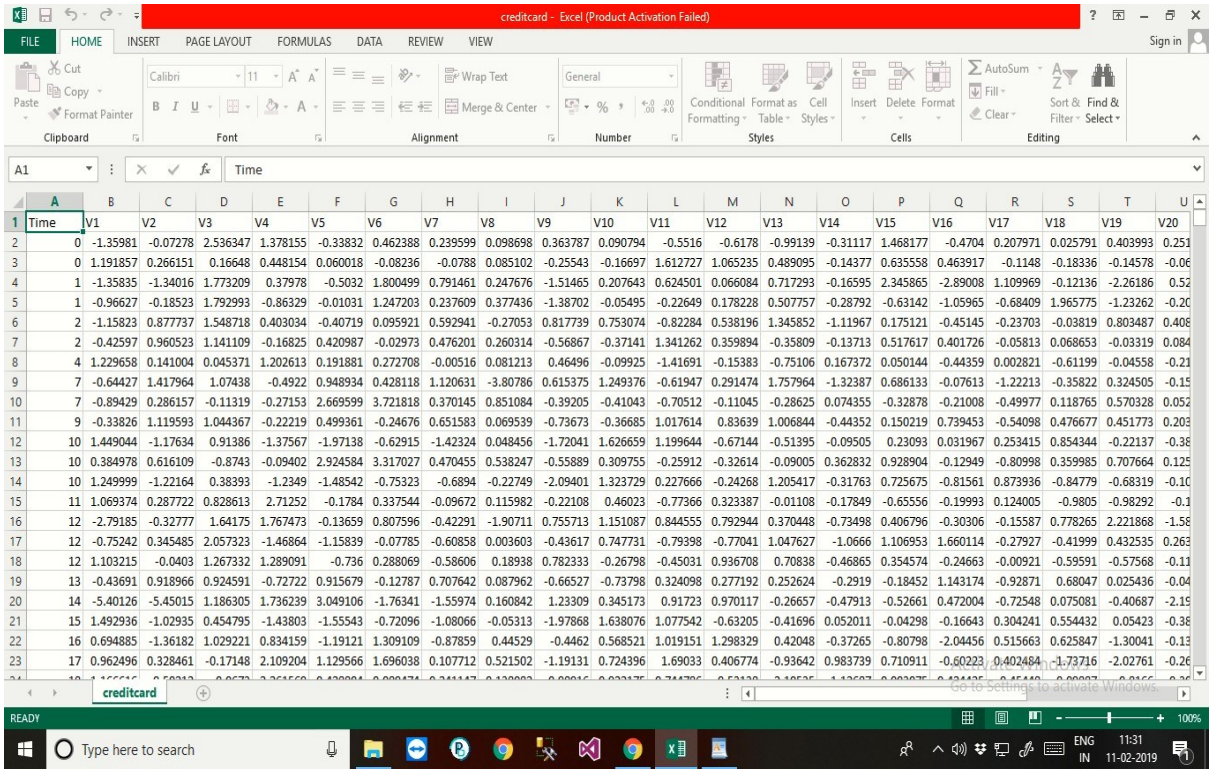


Figure: Dataset Design

DATA DESCRBE

The shape and characteristics of the data values belonging to each column has been describedin the following step. The data describe stage belongs to starting stage of exploratory analysisstage.

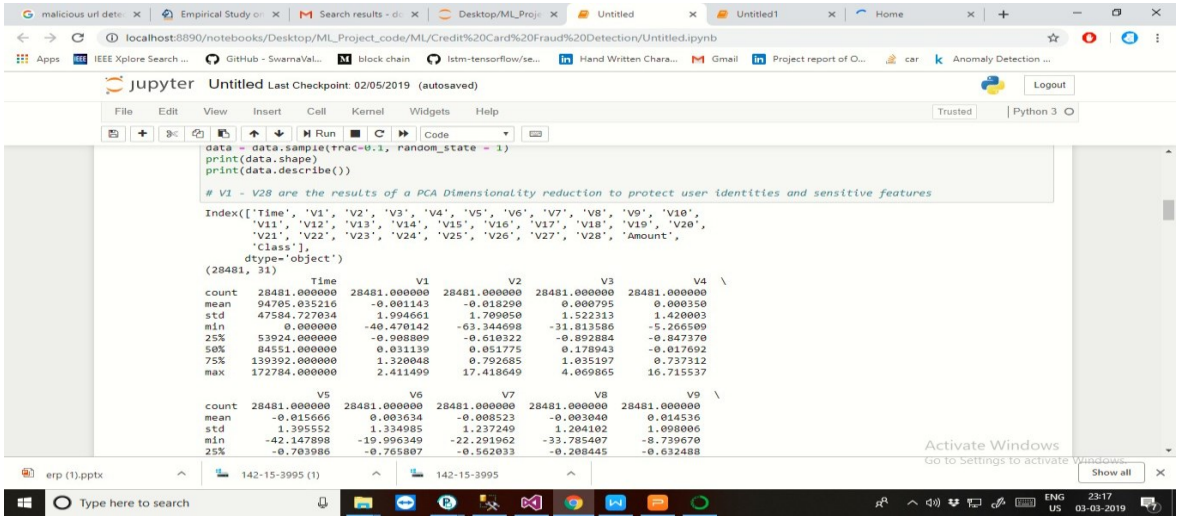


Figure : Data Describe

PREPROCESSING

International Journal of Engineering Research and Science & Technology

ISSN 2319-5991 www.ijerst.com

Vol. 20, Issue 4, 2024

The data values has been plotted using histogram describing the numerical distribution of thedata values.

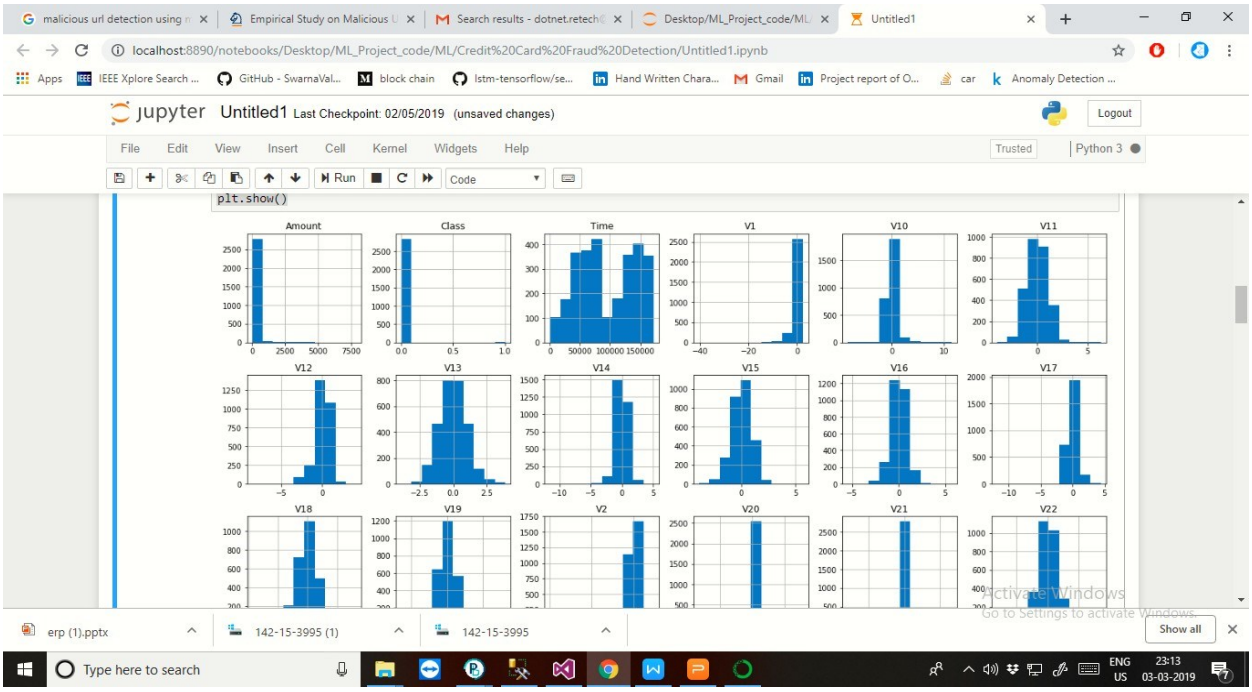


Figure : Histogram

FIND FRAUD

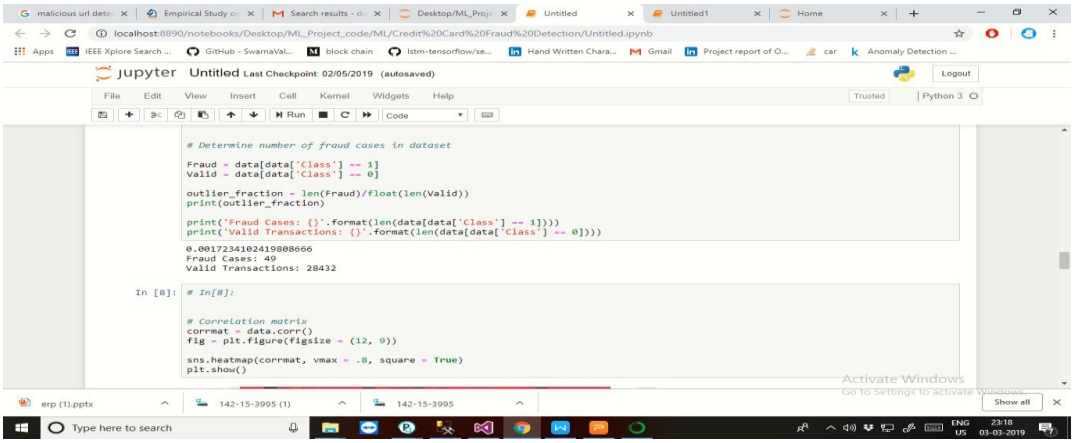


Figure : Fraud Diagram

The above picture describes the number of fraud that has been detected by the model. Thenumber of fraud detected has varied for two different algorithms.

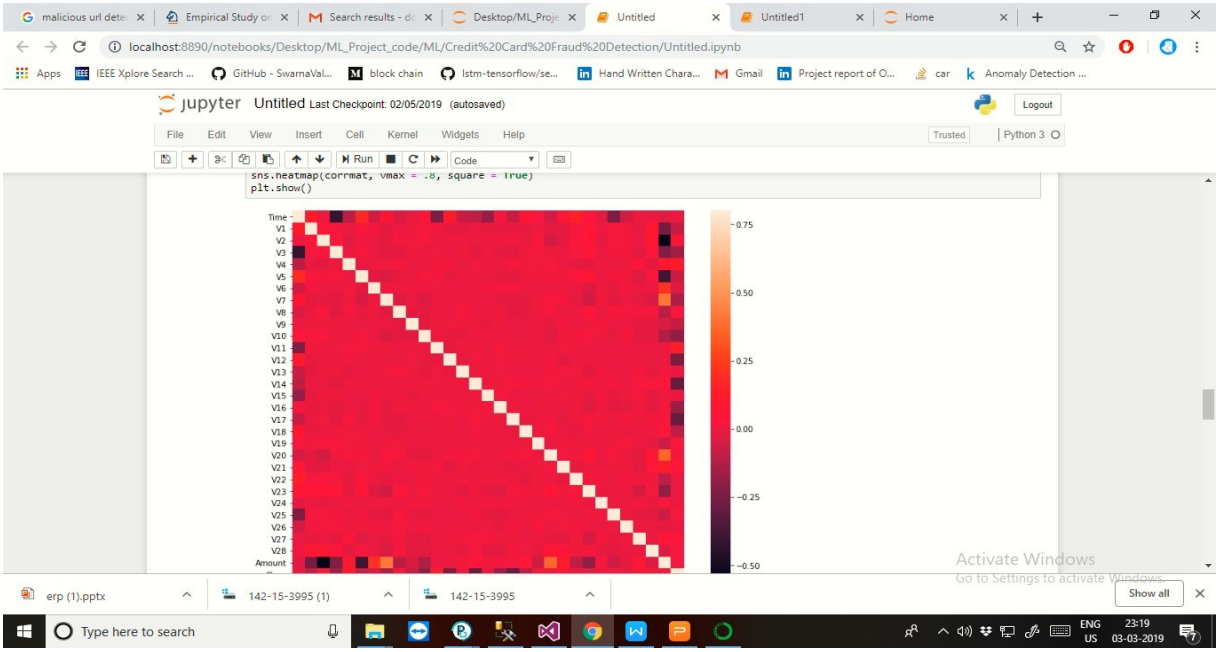
HEATMAP

International Journal of Engineering Research and Science & Technology

ISSN 2319-5991 www.ijerst.com

Vol. 20, Issue 4, 2024

The heat map has been plotted based on correlation matrix of the features. A correlation matrixdescribes the relation of features with each other. The level of correlation has been ranged from0.0-1.0 with 1(white



shade) denoting the features to be equilateral and 0(black shade) denoting the features with no interrelation.

Figure: Heat Map Diagr

PREDICTION

The prediction that has been achieved using the Isolation Forest Algorithm and Local OutlierFactor Algorithm has been shown below

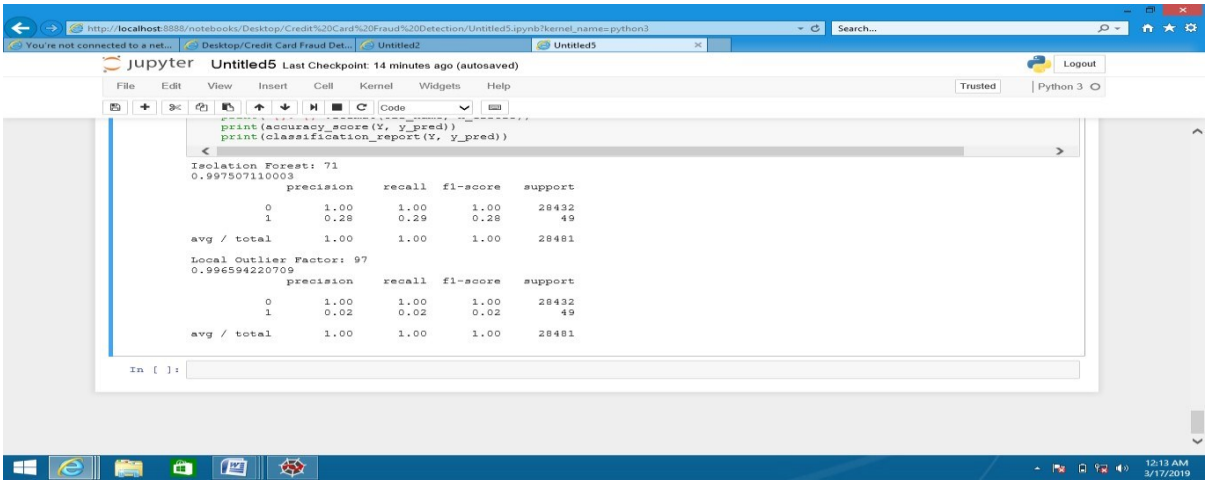


Figure: Accuracy Diagram

ISSN 2319-5991 www.ijerst.com

Vol. 20, Issue 4, 2024

CONCLUSION

In this model, we discussed about credit card fraud detection using machine learning. The proposed model has been extensively tested on different types of transactions. The results werepromising, almost all the fraudulent transactions could be detected successfully, and the proposed methodology has been compared with existing method and the results shows that proposed method performs superior than existing methods.

In this model, we detected the fraudulent transactions and recognized which illustrates the robustness of the proposed system. This proposed model took the trained dataset and performedclassification on basis of them, if the transaction was legal then it moved to class 0 and if the transaction was fraud then it moved to class 1, and significantly improve the detection accuracy.

The proposed method works efficiently in various platform, vivid environment and is a full- fledged cross platform application. The system has depicted robust, scalable and accurate performance to the degree that efficiency is taken into consideration in the Credit Card Fraud Detection System.

The system takes into consideration various factors and has been fulfilling or meeting all the project specifications documented.

REFERENCES

[1] V. Bhusari S. Patil, “Study of Hidden Markov Model in Credit Card Fraudulent Detection”, International Journal of Computer Applications (0975 – 8887) Volume 20– No.5, April 2011

[2] Priya Ravindra Shimpi, Prof. Vijayalaxmi Kadroli Angrish, “Survey on Credit Card Fraud Detection Techniques”, International Journal Of Engineering And Computer Science ISSN: 2319-7242 [3] Salvatore J. Stolfo, Wei Fan, WenkeLee, “Cost-based Modeling for Fraud and Intrusion Detection Results from the JAM Project”, In Proceedings of the ACM SIGMOD Conference on Management of Data, pages 207– 216, 2014.

[3] Delamaire. L. Abdou, HAH and Pointon. J,”Credit card f raud and detection techniques”, Banks and Bank Systems, Volume 4, Issue 2, 2009

[4] Suman, Nutan, “Review Paper on Credit Card Fraud Detection”, International Journal of Computer Trends and Technology (IJCTT) – volume 4 Issue 7–July 2013.

[5] Renu, Suman, “Analysis on Credit Card Fraud Detection Methods”, InternationalJournal of

Computer Trends and Technology (IJCTT) – volume 8 number 1 – Feb 2014.

[7] Sushmito Ghosh and Douglas L. Reilly, "Credit Card Fraud Detection with a Neural-Network" Proc. IEEE First Int. Conf. on Neural Networks, 2014.

[6] Deepak Pawar, SwapnilRabse, Sameer Paradkar, NainaKaushi, "Detection of Fraud in Online Credit Card Transactions", International Journal of Technical Research and Applications e-ISSN: 2320-8163.