**VISVESVARAYA TECHNOLOGICAL UNIVERSITY**

**BELAGAVI-590018**



**INTERNSHIP REPORT**

ON

***“CREDIT CARD FRAUD DETECTION”***

Submitted in partial fulfillment of the requirements for the award of

Degree of Bachelor of Engineering in computer science and technology

**Submitted by**

**Sumanth 4SU17CS104**

**Sai thejas C R 4SU17CS080**

**Prajwal M B 4SU17CS062**

**Vishwanath 4SU17CS121**

Under the Guidance of

**Sheik imam**

Carried out at

EBRAIN SOFT PVT.LTD

Bengaluru



**DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING**

**SDM INSTITUTE OF TECHNOLOGY**

**UJIRE-574240**

**2020-2021**

**S. D. M. INSTITUTE OF TECHNOLOGY**

(Affiliated To Visvesvaraya Technological University, Belagavi )

**UJIRE-574240**

**Department of Computer Science and Engineering**

**CERTIFICATE**

This is to Certify that the project work entitled “**CREDIT CARD FRAUD DETECTION**” is carried out by SUMANTH bearing USN **4SU17CS104**, SAI THEJAS C R bearing USN **4SU17CS080**, PRAJWAL M B bearing USN **4SU17CS062 and** VISHWANATH bearing USN **4SU17CS121** in partial fulfilment for requirements for **6TH SEMESTER** of the Visvesvaraya Technological University, Belagavi during the year 2020-21. It is certified that all corrections / suggestions indicated for Internship have been incorporated in the report deposited in the departmental library. The project report has been approved as it satisfies the academic requirements in respect of project work prescribed for the ebrain internship.

**Signature of the Guide Signature of the H.O.D**

(Guide NAME) **(Dr. Thayagaraju G. S)**

**ABSTRACT**

It is vital that credit card companies are able to identify fraudulent credit card transactions so that customers are not charged for items that they did not purchase. Such problems can be tackled with Data Science and its importance, along with Machine Learning, cannot be overstated. This project intends to illustrate the modelling of a data set using machine learning with Credit card Fraud Detection. The Credit card Fraud Detection Problem includes modelling past credit card transactions with the data of the ones that turned out to be fraud. This model is then used to recognize whether a new transaction is fraudulent or not. Our objective here is to detect 100% of the fraudulent transactions while minimizing the incorrect fraud classifications. Credit Card Fraud Detection is a typical sample of classification.

In this project we used Artificial Neural Networks to classify the transaction as fraud or normal.

**ACKNOWLEDGEMENT**

The internship opportunity we had with EBRAIN was a great chance for learning and professional development. Therefore, we consider our self as a lucky and we were provided to be a part of it.

We would like to thank SHEIK IMAM (ebrain Ltd) for providing us the opportunity for this training.

People working in the workspace made the training huge success and pleasant experience.

We extend our warm gratitude and regards to everyone who helped us during our internship.

**TABLE OF CONTENTS**

|  |  |  |
| --- | --- | --- |
| **ch.no** | **Particulars** | **Page no** |
| 1 | Company profile | 1-3 |
| 2 | Basic study | 4-7 |
| 3 | Project introduction | 8-9 |
| 4 | Problem definition | 10 |
| 5 | Problem solution | 11 |
| 6 | Screenshots | 12-18 |
| 7 | Results | 19-20 |
| 8 | Conclusion | 21 |
| 9 | References | 22 |

*Credit card Fraud Detection*

**E-BRAIN SOFTECH**

E-BRAIN Softech Pvt. Ltd. is a leading software development firm and training body, has been operational in Karnataka with a dedicated panel of experts from IT Industry. We provide services on Management, Education technology and guidance to anyone looking on any areas of interest. We are a team of qualified, experienced trainers & IT Professionals motivated to educate people by training & nurturing them to the best of their strengths.

**1.1 Services**

**Software Development**

Websites, apps, interfaces, AI, and more — we are a full-service product strategy, design, and development partner.

**User Experience**

Focusing on the customer, we design your customer’s journey, optimizing every touchpoint for convenience and delight.

**Training**

Machine Learning, Deep Learning, Python, Angular, Node JS, Core/Advance Java, Image Processing, Matlab, C/C++, Networking, Android

**1.2 products**

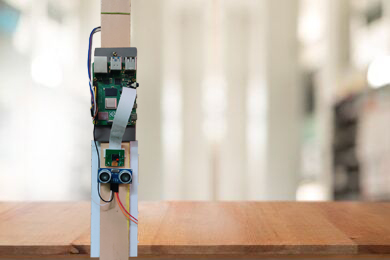
**Digital Eye**

According to the recent survey, India is now home to the largest number of blind people. Forty per cent of the world’s blind are in India and the official number is 12 million. The problems of blind and visually impaired people in India range from lack of basic necessities to global issues and prevailing social stigma attached to it. In India, the major issues faced by the blind and visually impaired persons are lack of disabled friendly infrastructure and transportation facilities as well. Most blind girls are not allowed to get out of the house for years together. Nevertheless, the blind deserve the same quality of life as that of a sighted person and have the right to participate equally in the society. As a step towards this; smart canes/sticks can prove to be extremely useful for independent and safe navigation of the blinds.

Dept cse, sdmit Page | 1

*Credit card Fraud Detection*

E-Brain employs a resourceful effort in initiating Digital Eye, which envisions easing the pain of people with visual impairment and blind to be self-relined. Digital Eye is a smart navigating-stick with ultrasonic proximity sensors to detect the presence of target objects and GPS module to guide these blind people to reach out their destinations independently. This stick comprises camera equipped with the object detection algorithm integrated with ultrasonic sensors to detect any upcoming obstacles and to sense the exact distance from that obstacle using voice based bot and a water-detection model integrated with the voicebot to discern between water and no-water regions.

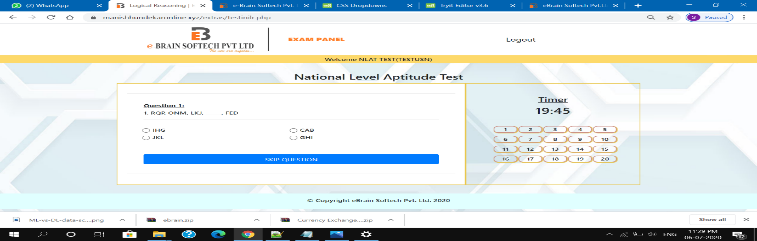


**Onlinex**

This software provides a best platform for MCQ Examinations and auto evaluation. This software also provides easy evaluation of scores automatically. After Successful Evaluation it will generate a score card for every student Individually. These Report cards will be sent to their Registered Email Id automatically without any man power.

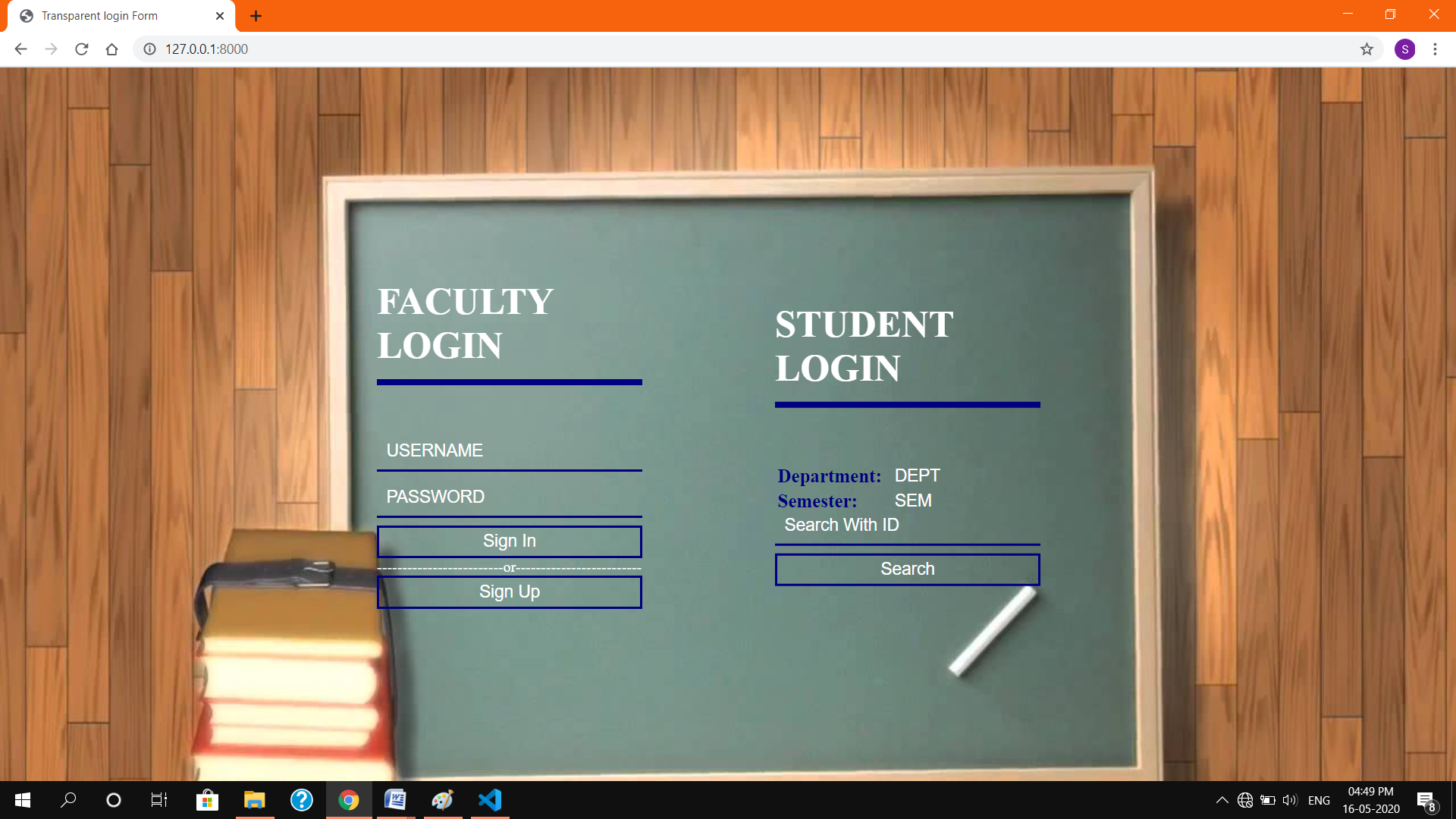
Dept cse, sdmit Page | 2

*Credit card Fraud Detection*



**AI Attendance**

The proposed software takes the participation of the student using facial identification technique. This participation is recorded by a high quality camera which is installed as a part of classroom and continuously capturing the images of the students, detects their faces, contrast distinguished appearances and mark the attendance. After Marking the attendance, it will also generate a report and will be stored for further use. With that the software also provides an acknowledgment message to the student if he is absent.



Dept cse, sdmit Page | 3

*Credit card Fraud Detection*

**BASIC STUDY**

**2.1 PYTHON**

Python is an interpreted, object-oriented, high level programming language with dynamic semantics. Its high-level built-in data structures, combined with dynamic typing and dynamic binding, make it very attractive for Rapid Application Development.

Python is dynamically typed and garbage-collected. It supports multiple programming paradigms, including procedural object-oriented, and functional programming. Python is often described as a "batteries included" language due to its comprehensive standard library

The Python 2 language, i.e. Python 2.7.x, was officially discontinued on 1 January 2020 (first planned for 2015) after which security patches and other improvements will not be released for it. With Python 2's end-of-life, only Python 3.5.x and later are supported.

Python uses dynamic typing and a combination of reference counting and a cycle-detecting garbage collector for memory management. It also features dynamic name resolution (late binding), which binds method and variable names during program execution.

**2.2 ANACONDA NAVIGATOR**

Anaconda Navigator is a desktop graphical user interface (GUI) included in Anaconda® distribution that allows you to launch applications and easily manage con packages, environments, and channels without using command-line commands.

In order to run, many scientific packages depend on specific versions of other packages. Data scientists often use multiple versions of many packages and use multiple environments to separate these different versions.

Navigator is an easy, point-and-click way to work with packages and environments without needing to type conda commands in a terminal window. You can use it to find the packages you want, install them in an environment, run the packages, and update them – all inside Navigator.

**Anaconda** is a free and open-source distribution of the Python and R programming languages for scientific computing (data science, machine learning applications, large-scale data processing, predictive analytics, etc.), that aims to simplify package management and deployment. Package versions are managed by the package management system conda. The Anaconda distribution includes data-science packages suitable for Windows, Linux, and MacOS.

Dept cse,sdmit Page | 4

*Credit card Fraud Detection*

**Anaconda distribution** comes with 1,500 packages selected from PyPI as well as the conda package and virtual environment manager. It also includes a GUI, **Anaconda Navigator**, as a graphical alternative to the command line interface (CLI).

The big difference between conda and the pip package manager is in how package dependencies are managed, which is a significant challenge for Python data science and the reason conda exists.

many scientific packages depend on specific versions of other packages. Data scientists often use multiple versions of many packages and use multiple environments to separate these different versions.

**2.3 JUPYTER NOTEBOOK**

The Jupyter Notebook is an open-source web application that allows you to create and share documents that contain live code, equations, visualizations and narrative text. Uses include: data cleaning and transformation, numerical simulation, statistical modelling, data visualization, machine learning, and much more.

A Jupyter Notebook can be converted to a number of open standard output formats

Jupyter Notebook can connect to many kernels to allow programming in many languages. By default, Jupyter Notebook ships with the Python kernel. As of the (October 2014), there are currently 49 Jupyter-compatible kernels for many programming languages, including Python, R, Julia and Haskell.

In addition to running your code, it stores code and output, together with markdown notes, in an editable document called a notebook. When you save it, this is sent from your browser to the notebook server, which saves it on disk as a JSON file with extension.

**2.4 PANDAS**

Pandas is a library written for the Python programming language for data manipulation and analysis. In particular, it offers data structures and operations for manipulating numerical tables and time series, which is a Panel Data.

In computer programming,pandas is a software library written for the Python programming language for data manipulation and analysis. In particular, it offers data structures and operations for manipulating numerical tables and time series. It is free software released under the three-clause BSD license. The name is derived from the term "panel data", an econometrics term for data sets that include observations over multiple time periods for the same individuals.

Dept cse,sdmit Page | 5

*Credit card Fraud Detection*

Pandas is a fast, powerful, flexible and easy to use open source data analysis and manipulation tool, built on top of the python programming language.

**2.5 NUMPY**

NumPy is a package in Python used for Scientific Computing. NumPy package is used to perform different operations. The ndarray (NumPy Array) is a multidimensional array used to store values of same datatype. In computer programming, pandas is a software library written for the Python programming language for data manipulation and analysis. In particular, it offers data structures and operations for manipulating numerical tables and time series. It is free software released under the three-clause BSD license. The name is derived from the term "panel data", an econometrics term for data sets that include observations over multiple time periods for the same individuals.

NumPy targets the CPython reference implementation of Python, which is a nonoptimizing bytecode interpreter. Mathematical algorithms written for this version of Python often run much slower than compiled equivalents. NumPy addresses the slowness problem partly by providing multidimensional arrays and functions and operators that operate efficiently on arrays, requiring rewriting some code, mostly inner loops using NumPy.

**2.7 MATPLOTLIB**

Several toolkits are available which extend Matplotlib functionality. Some are separate downloads,others ship with the Matplotlib source code but have external dependencies.

 Basemap: map plotting with various map projections, coastlines, and political boundaries.

 Cartopy: a mapping library featuring object-oriented map projection definitions, and arbitrary point, line, polygon and image transformation capabilities. (Matplotlib v1.2 and above)

 Excel tools: utilities for exchanging data with Microsoft Excel

 GTK tools: interface to the GTK+ library

 Qt interface

 Mplot3d: 3-D plots

 Natgrid: interface to the natgrid library for gridding irregularly spaced data.

Dept cse, sdmit Page | 6

*Credit card Fraud Detection*

**2.8 MACHINE LEARNING**

Machine learning is an application of artificial intelligence (AI) that provides systems the ability to automatically learn and improve from experience without being explicitly programmed. Machine learning focuses on the development of computer programs that can access data and use it learn for themselves.

Machine learning is closely related to computational statistics, which focuses on making predictions using computers. The study of mathematical optimization delivers methods, theory and application domains to the field of machine learning. Data mining is a field of study within machine learning, and focuses on exploratory data analysis through unsupervised learning. In its application across business problems, machine learning is also referred to as predictive analytics.

There are also some types of machine learning algorithm that are used in very specific case, but three main methods are used today.

**i. supervised learning:** supervised learning is one of the most basic types of machine learning. in this type, the machine learning algorithm is trained on labelled data. Even though the data needs to be labelled accurately for this method to work, supervised learning is extremely power full when used in the right circumstances.

**ii. Unsupervised learning:** unsupervised machine learning holds the advantages of being able to work with unlabelled data. This means that human labour is not required to make the dataset machine-readable, allowing much larger datasets to be worked on by the program.

**iii. Reinforcement learning:** this learning directly takes inspiration from how human beings learn from data in their lives. it features an algorithm that improves upon itself and learns fromnew situation using a trail-and-error method. Favourable outputs area encouraged or ‘reinforced’ and non-favourable outputs are discouraged.

Dept cse, sdmit Page | 7

*Credit card Fraud Detection*

**PROJECT 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 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 feedback to the automated system which is used to train and update the algorithm to eventually improve the fraud-detection performance over time.

Dept cse, sdmit Page | 8

Credit card Fraud Detection

Fraud detection methods are continuously developed to defend criminals in adapting to their fraudulent strategies. These frauds are classified as:

● Credit Card Frauds: Online and Offline

● Card Theft

● Account Bankruptcy

● Device Intrusion

● Application Fraud

● Counterfeit Card

● Telecommunication Fraud

Some of the currently used approaches to detection of such fraud are:

● Artificial Neural Network

● Fuzzy Logic

● Genetic Algorithm

● Logistic Regression

● Decision tree

● Support Vector Machines

● Bayesian Networks

● Hidden Markov Model

● K-Nearest Neighbour

Dept cse,sdmit Page | 9

*Credit card Fraud Detection*

**PROJECT DEFINATION**

Using data from a credit card issuer, a neural network based fraud detection system was trained on a large sample of labelled credit card account transactions and tested on a holdout data set.

The neural network was trained based on anonymized features .

The main aim of the project is to find whether the given transaction is fraud or genuine .

Benefits of Credit card fraud detection:

● Banks can save literally millions of dollars a year. If they halt suspicious

transactions early and confirm fraud.

● Building good relationships with customers .

● Time will be reduced to find the fraud transactions.

● could detect credit card behavior and usage patterns in a large volume of customer and

transactional data

Dept cse,sdmit Page | 10

*Credit card Fraud Detection*

**PROJECT SOLUTION**

We are going get the solution for the project by following steps below:

1. In this project we are using pandas,numpy,tensorflow,matplotlib,sklearn and seaborn modules . We import all the packages required .

2. Using the pandas library , we are going to read and analyze the dataset .

3. Then we are going to select the anonymized features and plot each of the features to find similar distributions to drop it from the dataset .

4. Based on the plots, new features are created to identify values where fraudulent transactions are more common and creating new features for the normal transactions .

5. Create dataframes of only Fraud and Normal transactions.

6. Set X\_train equal to 80% of the fraudulent transactions . Then add 80% of the normal transactions to X\_train.X\_test contains all the transactions which are not in X\_train.

7. Shuffle the dataframes so that the training is done in a random order and add our target features to y\_train and y\_test.Drop target features from X\_train and X\_test.

8. Transform each feature in features so that it has a mean of 0 and standard deviation of 1. This helps with training the neural network.

9. Then create the Layers for the neural network.

10. Train the model by setting training epoch, batch size and learning rate.

11. Find the predicted values, then use them to build a confusion matrix.

12. A confusion matrix is a table that is often used to describe the performance of a classification model (or "classifier") on a set of test data for which the true values are known.

13. Then we can see the prediction of the model by creating a web app using the flask framework .

*Dept cse,sdmit Page|11*

*Credit card Fraud Detection*

**SCREENSHOTS**

fig 1.1 represent the importing of packages that comes under libraries

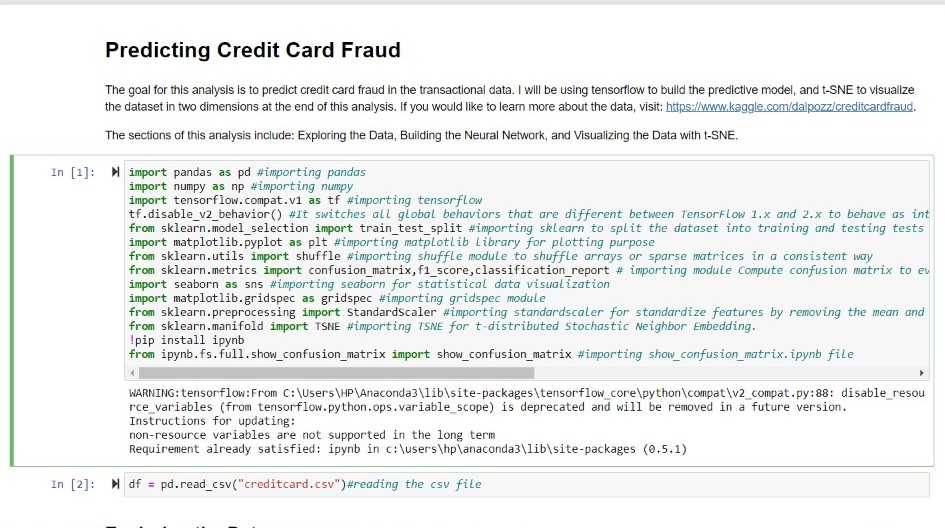
 fig 1.1

fig 1.2 represents reading data from file and copied into another file

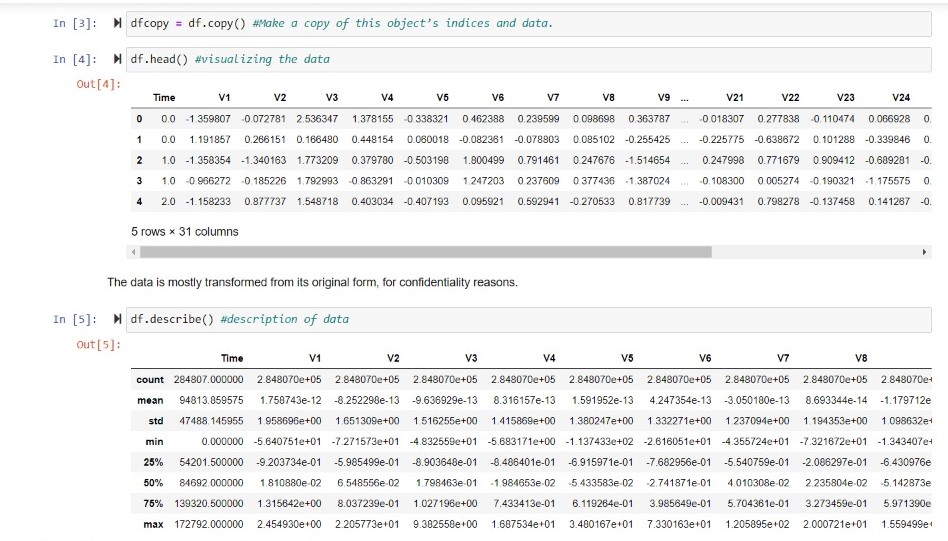
****

fig 1.2

Dept cse,sdmit Page | 12

*Credit card Fraud Detection*

Fig 1.3 represents describing the fraud and normal data

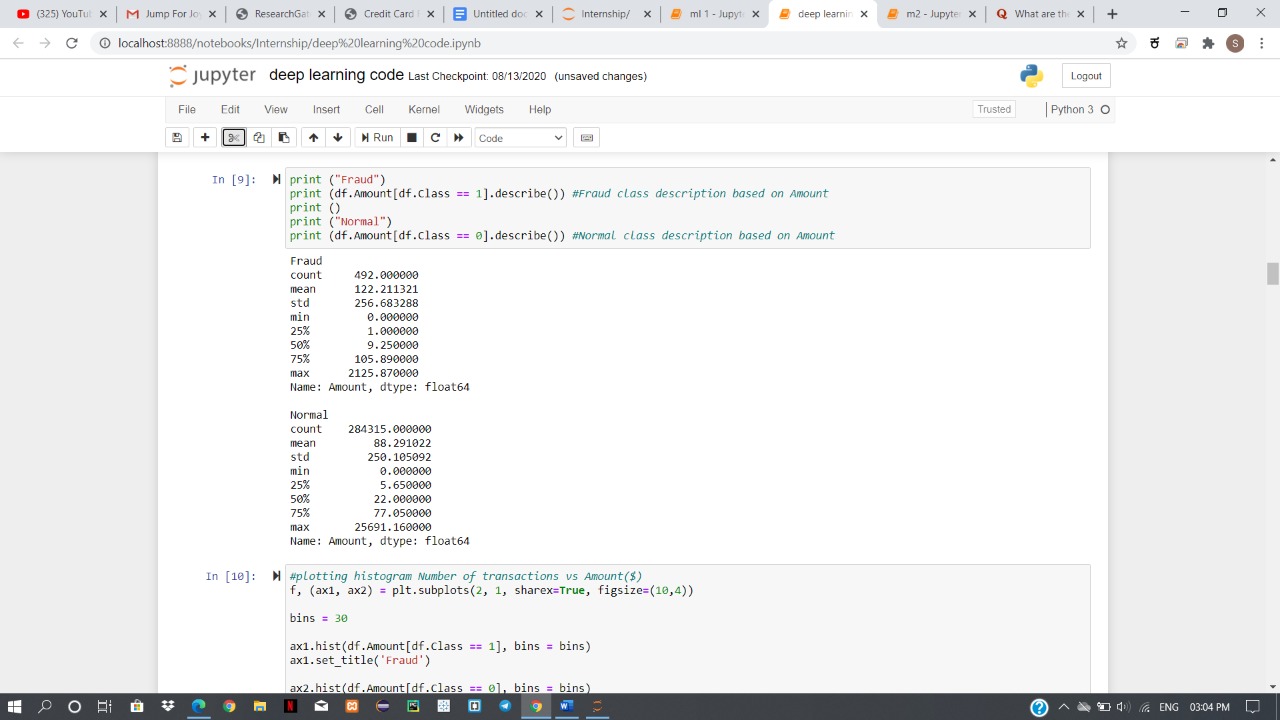


Fig 1.3

Fig 1.4 shows the plotting of the fraud and normal data

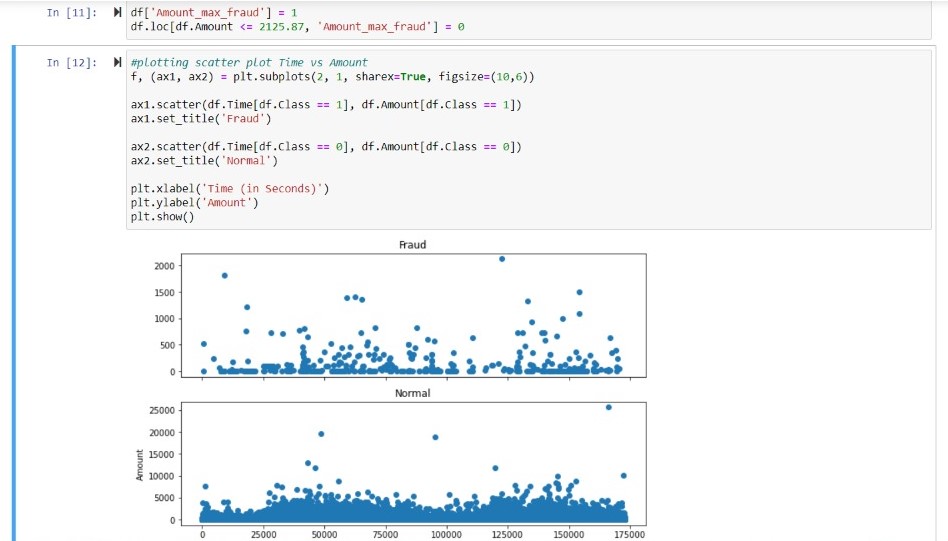


Fig 1.4

Dept cse,sdmit Page | 13

*Credit card Fraud Detection*

Fig 1.5 shows usage of lambda function to identify where fraudulent transactions are common



Fig 1.5

Fig 1.6 represents the normal and fraud data with counts

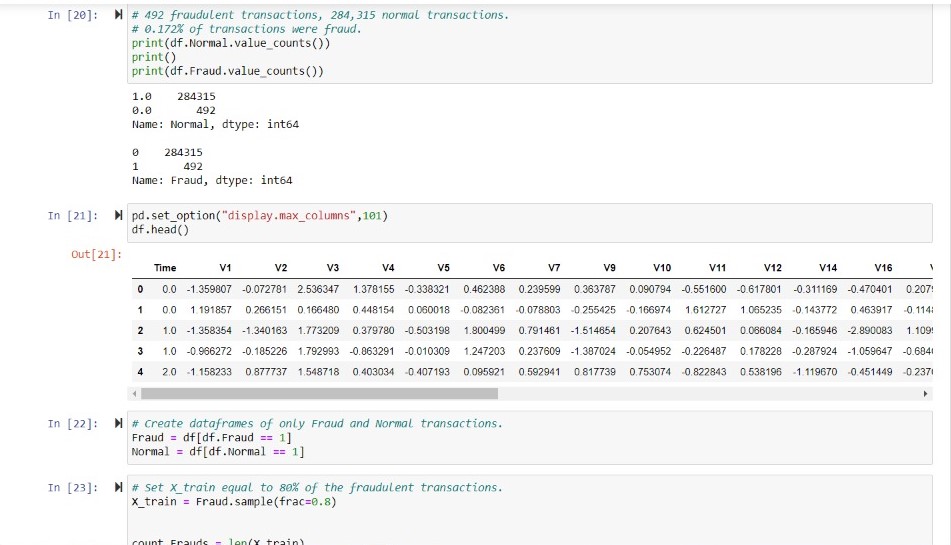


Fig 1.6

Dept cse,sdmit Page | 14

*Credit card Fraud Detection*

Fig 1.7 represents displaying the shape of matrix



Fig 1.7

Fig 1.8 shows the ratio with the fraud data

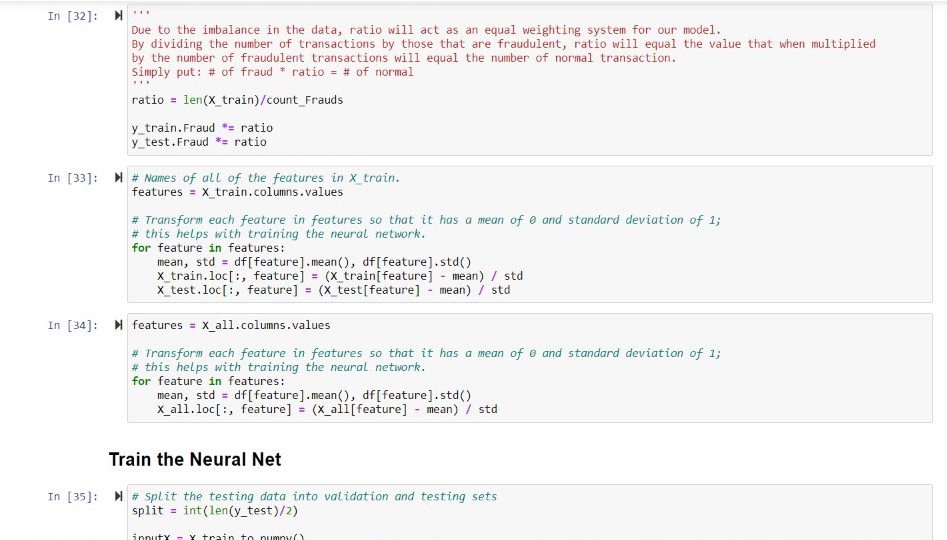


Fig 1.8

*Dept cse,sdmit Page|15*

*Credit card Fraud Detection*

Fig 1.9 shows that actual code to find the accuracy for different epochs

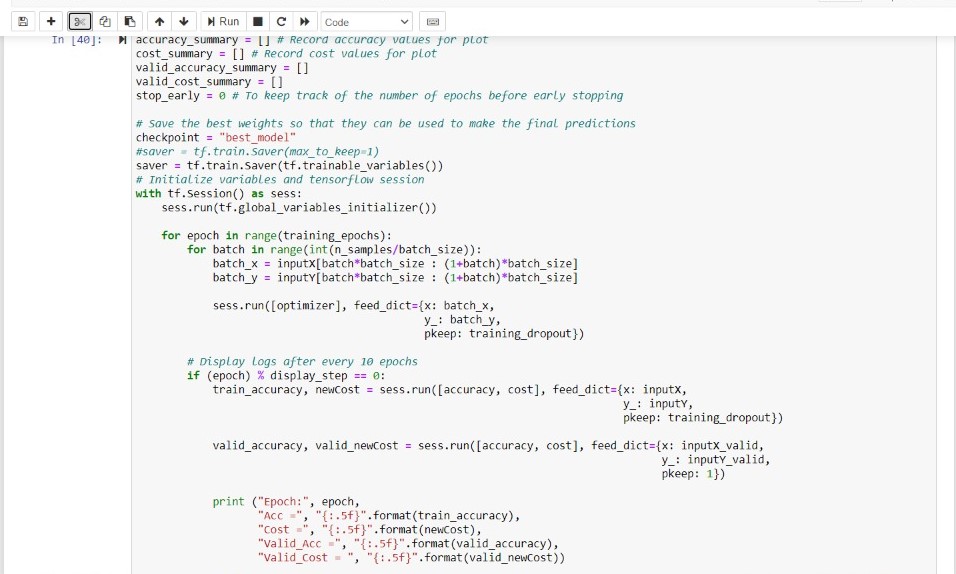


Fig 1.9

Fig 2.0 shows the result or accuracy for different epochs

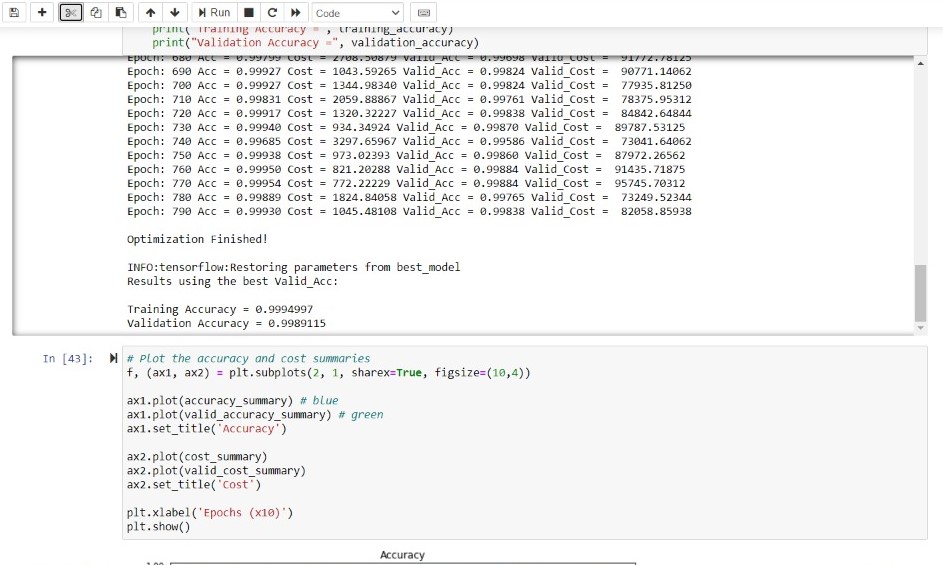


Fig 2.0

*Dept cse,sdmit Page|16*

*Credit card Fraud Detection*

Fig 2.1 represents the plot for predicted and true label

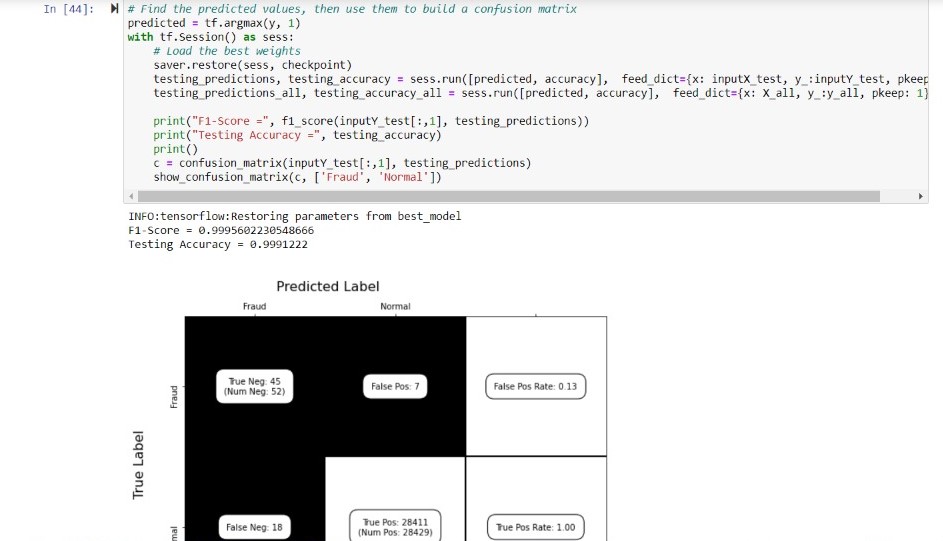


Fig 2.1

Fig 2.2 shows that displaying accuracy and writing a file

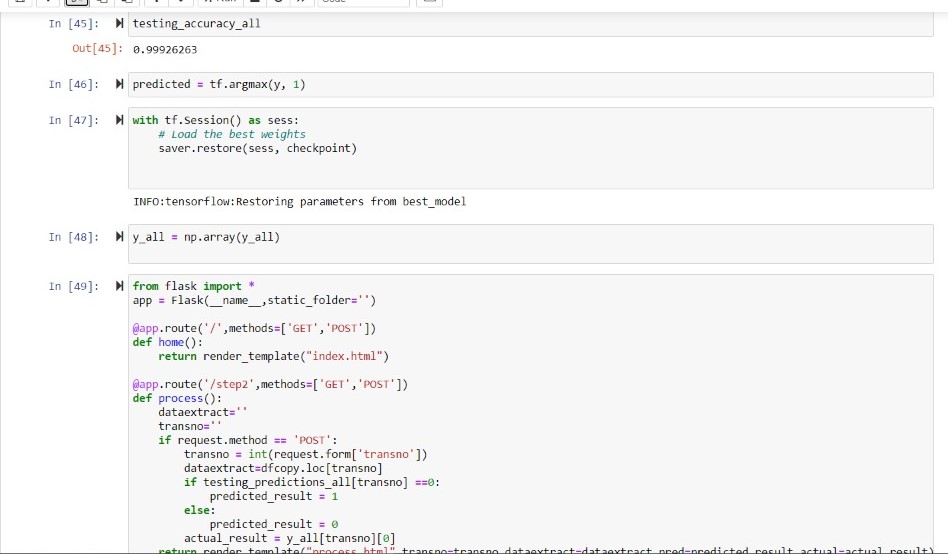


Fig 2.2

*Dept cse,sdmit Page|17*

*Credit card Fraud Detection*

Fig 2.3 represents plotting of histograms

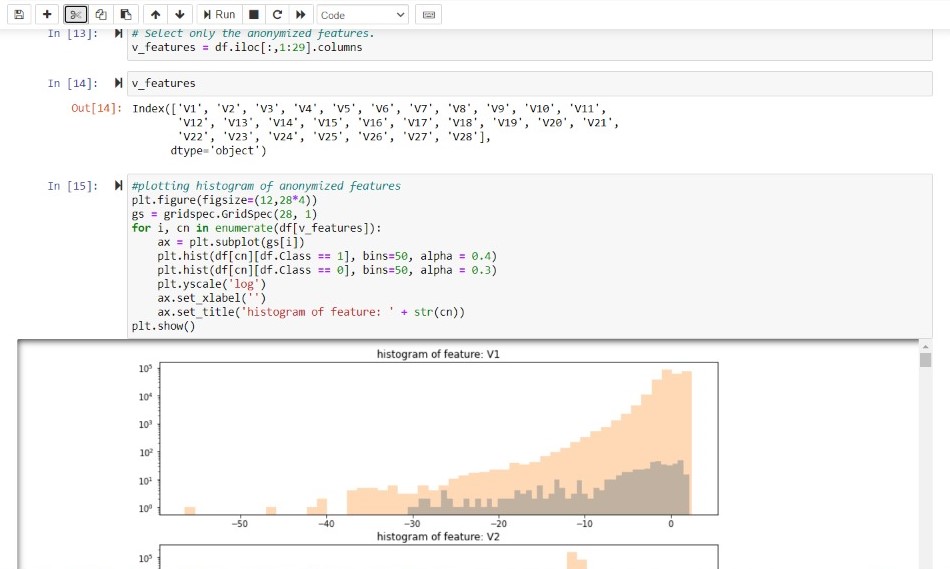


Fig 2.3

Fig 2.4 shows that final link to check for normal or fraudulent transactions

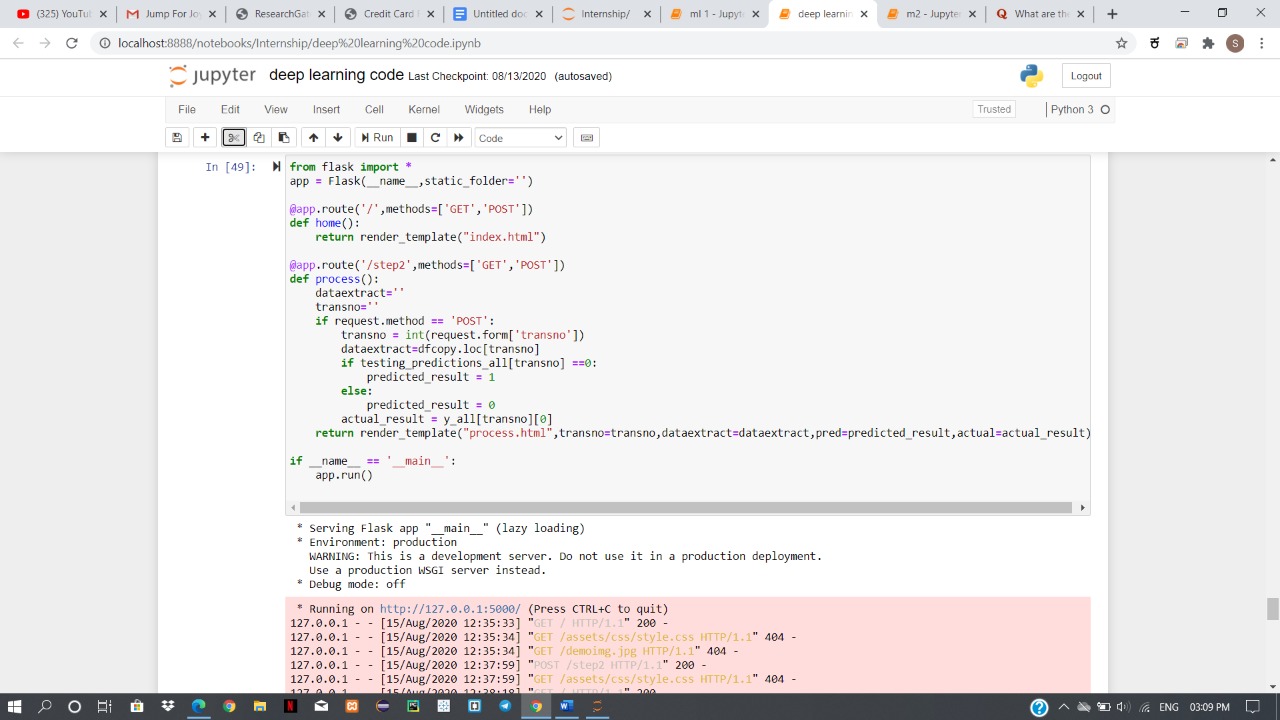


Fig 2.4

*Dept cse,sdmit Page|18*

*Credit card Fraud Detection*

**RESULTS**

Fig 3.1 shows that the picture asking for the transaction number

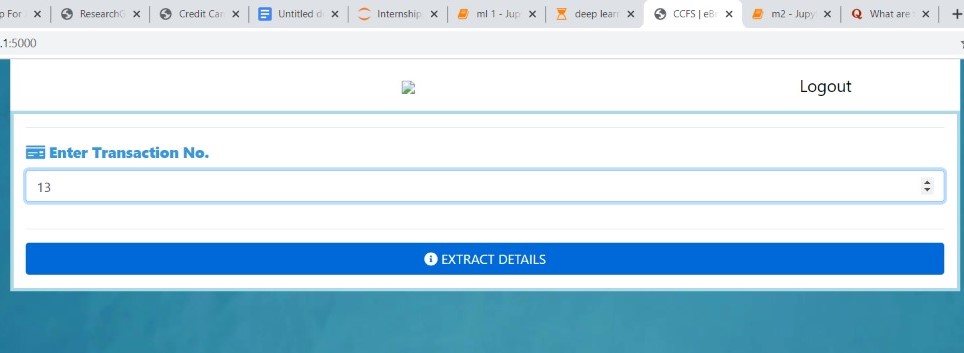


Fig 3.1

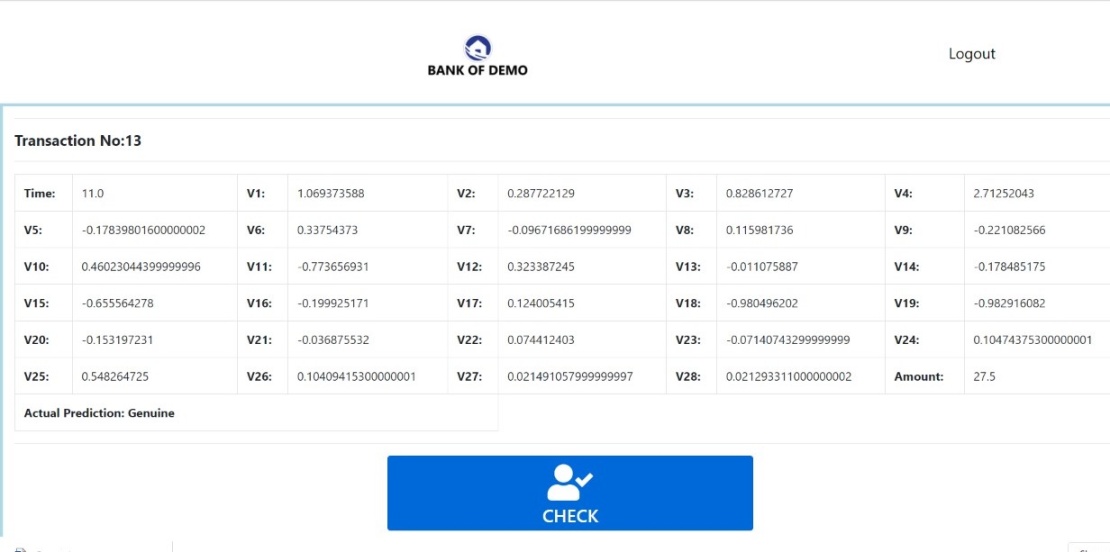
fig 3.2 shows that checking for the transactions whether fraudulent or normal .

Fig 3.2

Dept cse,sdmit *Page|19*

*Credit card Fraud Detection*

Fig 3.3 shows that displaying the transactions with genuine

.

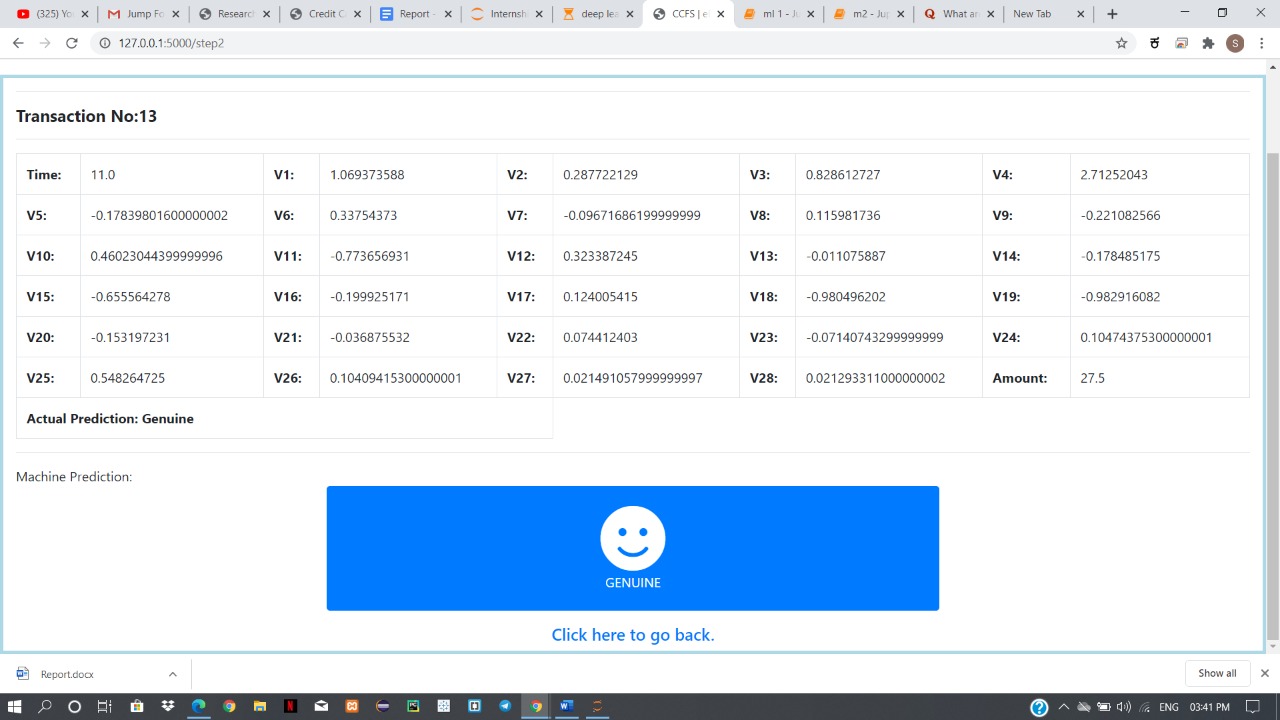


Fig 3.3

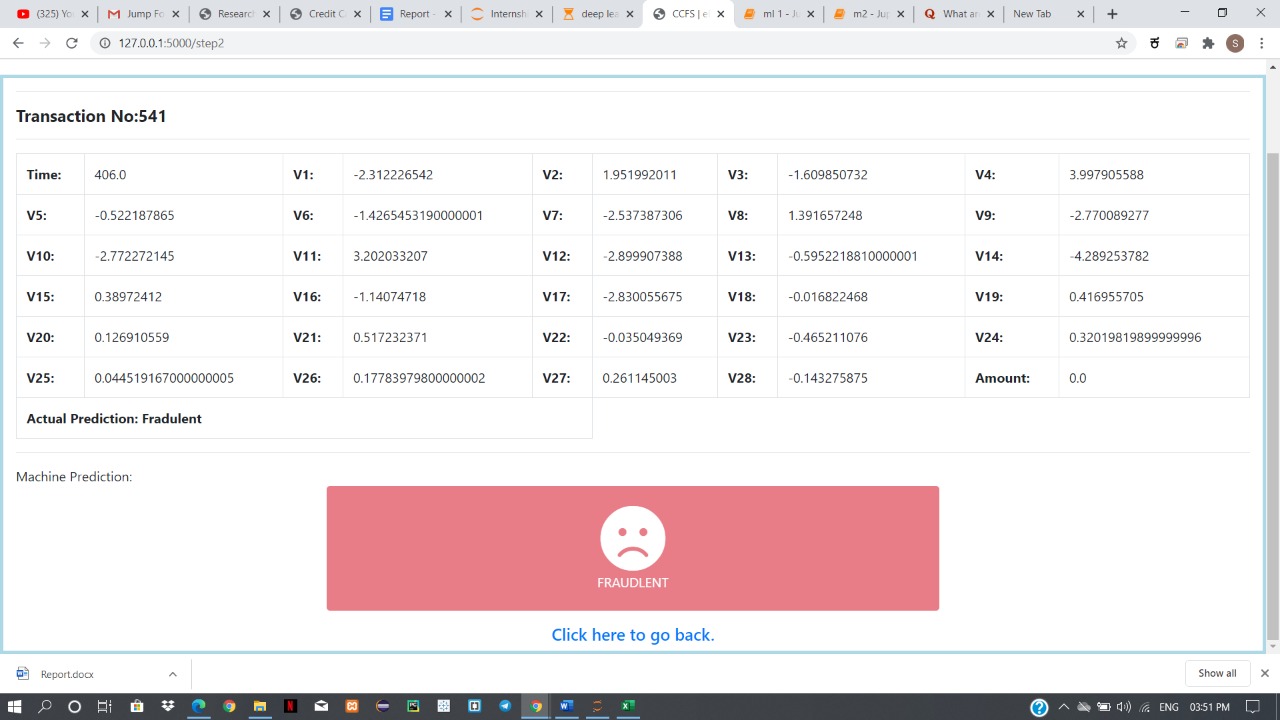
fig 3.4 shows that displaying the transactions are fraudulent

Fig 1.9

Fig 3.4

Dept cse,sdmit Page | 20

*Credit card Fraud Detection*

**CONCLUSION**

Credit card fraud detection is done in this project . Also the scheme through which the project is achieved is Artificial Neural Network . Credit card fraud is without a doubt an act of criminal dishonesty . This project is helpful to detect the fraud at the accuracy of around 99% . It helps banks to take measures on credit card fraud . This method is really very effective . also other techniques can be used to do this project.

Dept cse,sdmit Page | 21

Credit card Fraud Detection

**REFERENCES**

[1] “Credit Card Fraud Detection using Machine Learning and Data Science” by S P Maniraj,Aditya Saini and Swarna Deep Sarkar Shadab Ahmed ,International Journal of Engineering Research & Technology,Vol. 8 Issue 09, September-2019

[2] “Using Neural Network for Credit Card Fraud Detection” by Sevdalina Georgieva, Maya Markova and Velizar Pavlov,Conference Paper in AIP Conference Proceedings,October 2019

Dept cse,sdmit Page | 22