JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

ARA – Credit Card Fraud Detection System

[1]Amey Salaskar, [2]Abhijeet Potale, [3]Rushikesh Gawde, [4] Dr. Savita Sangam

[1,2,3,4]Department of Information Technology, SSJCOE, India

Abstract: Credit card fraud is a growing global concern. It involves use of some other persons credit/debit card for purchasing purposes or withdrawing funds. To address this issue, various fraud detection methods, including rule-based systems, machine learning algorithms, and deep learning models, have been developed. The work presented reviews the existing credit card fraud detection techniques, focusing on their strengths, weaknesses, and performance metrics. Additionally, we propose a fraud detection approach that combines with various machine learning algorithms to enhance fraud detection accuracy. Our experimental results determine the fraudulent results based on user input data. The system detects the fraud based on the data that is fed to the system by individuals. The detection is carried out by using various machine learning algorithms.

Index Terms - **Credit card fraud, Fraud detection, fraudulent, Machine learning, Fraud detection rate**

I. INTRODUCTION

'Fraud' in credit card transactions is unauthorized and unwanted usage of an account by someone other than the owner of that account. Credit card fraud is a serious problem affecting financial institutions and cardholders globally, leading to significant financial losses and privacy breaches. Traditional rule-based fraud detection systems are limited in their ability to adapt to evolving fraud patterns and may generate high false positive rates, inconveniencing legitimate cardholders.

Machine learning (ML) has emerged as a powerful tool in combating credit card fraud due to its ability to detect complex patterns in large datasets. ML algorithms can learn from historical transaction data to identify fraudulent activities and adapt to new fraud patterns in real-time.

The work presented provides an overview of the use of machine learning in credit card fraud detection. We discuss various ML techniques, including supervised learning, unsupervised learning, and deep learning, and how they can be applied to detect fraudulent transactions. We also highlight the challenges and limitations of ML-based fraud detection and discuss future research directions in this field.

II. LITERATURE REVIEW

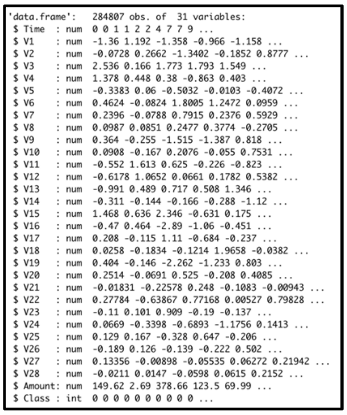
In today’s world the credit card fraud has become the major issue all over the world. The attackers try to steal sensitive information from the user and try to make unauthorized transactions.

To deal with such problems some of the experts have created the system to minimize the risk of credit card fraud. This system detects whether the transaction made is fraud or not. Such type of systems has really helped the society and there are very low chances of frauds.

From the ref. [1], the authors K.Ratna Sree Valli, P.Jyothi, G.Varun Sai, R.Rohith Sai Subash implemented different machine learning algorithms on an imbalanced dataset such as logistic regression, naïvebayes, random forest with ensemble classifiers using boosting technique. The credit card has become the most popular mode of payment. Each user uses the credit card for purchasing an item or any transaction purposes. The Dataset used by the authors was made available from Kaggle. This dataset contains the transactions, which was made in two days, in September 2013 by European cardholders. The dataset was divided into trained data set and test data set. 70% of the data set was under training and the remaining 30% was under testing.

From the ref. [2] according to (Credit card statistics 2021) the number of people using credit cards around the world was 2.8 billion in 2019, in addition 70% of those users own a single card at least. Reports of Credit card fraud in the US rose by 44.7% from 271,927 in 2019 to 393,207 reports in 2020. The purpose of the work presented by the author “Meera AlEmad” is to stop fraudsters from the unauthorized usage of customers’ accounts and to identify the credit card fraud using various machine learning algorithms. After collecting the necessary datasets four machine learning was created - KNN, SVM, Logistic Regression and Naïve Bayes. The dataset is sectioned into a ratio of 70:30, the training set will be the 70% and remaining set will be the testing set which is the 30%. As mentioned below we have added the system outputs which have been created by some of the experts which gives the brief overview of how Data Visualization tools and techniques and various machine algorithms can be used to detect the fraudulent transactions and gives the best result.

The figure 2.1 bellow shows the structure of the dataset where all attributes are shown with their types.



```
'data.frame':   284807 obs. of  31 variables:
 $ Time  : num  0 0 1 1 2 2 4 7 7 9 ...
 $ V1    : num  -1.36 1.192 -1.358 -0.966 -1.158 ...
 $ V2    : num  -0.0728 0.2662 -1.3402 -0.1852 0.8777 ...
 $ V3    : num  2.536 0.166 1.773 1.793 1.549 ...
 $ V4    : num  1.378 0.448 0.38 -0.863 0.403 ...
 $ V5    : num  -0.3383 0.06 -0.5032 -0.0103 -0.4072 ...
 $ V6    : num  0.4624 -0.0824 1.8005 1.2472 0.0959 ...
 $ V7    : num  0.2396 -0.0788 0.7915 0.2376 0.5929 ...
 $ V8    : num  0.0987 0.0851 0.2477 0.3774 -0.2705 ...
 $ V9    : num  0.364 -0.255 -1.515 -1.387 0.818 ...
 $ V10   : num  0.0908 -0.167 0.2076 -0.055 0.7531 ...
 $ V11   : num  -0.552 1.613 0.625 -0.226 -0.823 ...
 $ V12   : num  -0.6178 1.0652 0.0661 0.1782 0.5382 ...
 $ V13   : num  -0.991 0.489 0.717 0.508 1.346 ...
 $ V14   : num  -0.311 -0.144 -0.166 -0.288 -1.12 ...
 $ V15   : num  1.468 0.636 2.346 -0.631 0.175 ...
 $ V16   : num  -0.47 0.464 -2.89 -1.06 -0.451 ...
 $ V17   : num  0.208 -0.115 1.11 -0.684 -0.237 ...
 $ V18   : num  0.0258 -0.1834 -0.1214 1.9658 -0.0382 ...
 $ V19   : num  0.404 -0.146 -2.262 -1.233 0.803 ...
 $ V20   : num  0.2514 -0.0691 0.525 -0.208 0.4085 ...
 $ V21   : num  -0.01831 -0.22578 0.248 -0.1083 -0.00943 ...
 $ V22   : num  0.27784 -0.63867 0.77168 0.00527 0.79828 ...
 $ V23   : num  -0.11 0.101 0.909 -0.19 -0.137 ...
 $ V24   : num  0.0669 -0.3398 -0.6893 -1.1756 0.1413 ...
 $ V25   : num  0.129 0.167 -0.328 0.647 -0.206 ...
 $ V26   : num  -0.189 0.126 -0.139 -0.222 0.502 ...
 $ V27   : num  0.13356 -0.00898 -0.05535 0.06272 0.21942 ...
 $ V28   : num  -0.0211 0.0147 -0.0598 0.0615 0.2152 ...
 $ Amount: num  149.62 2.69 378.66 123.5 69.99 ...
 $ Class : int  0 0 0 0 0 0 0 0 0 0 ...
```

figure.2.1 The Structure of Dataset

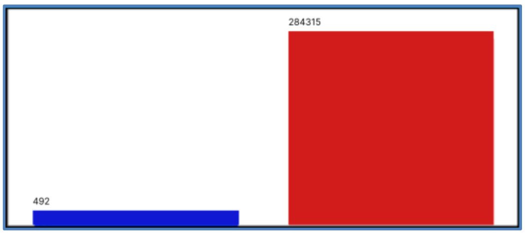
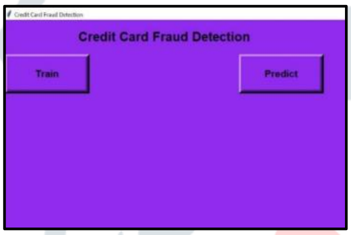


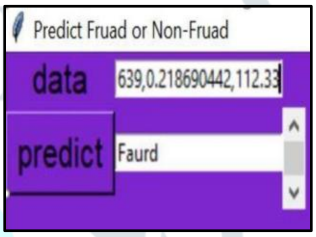
figure.2.2 Distribution of Class

The figure 2.2 shows the class distribution, the red bar shows non-fraudulent transactions which contains 284,315 variables and the blue bar shows the fraudulent transactions which contains 492 variables.

The authors from the ref [3] build the credit card fraud detection system using algorithms like – Logistic Regression, Naïve bayes, Decision Tree and ANN.

The figure 2.3 shows the user interface of the system. The window consists of two buttons – train and predict. When the user clicks in “train” button the model will be trained and clicking on “predict” button it will redirect the user to next window.





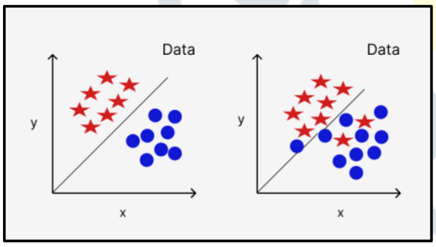
data: 639,0.218690442,112.33

Faurd

figure.2.3 The User Interface of System figure.2.4 Detection of fraud data

The figure 2.4 shows the next interface when the user clicks on predict button. Here the user will see two features’ “data” and “predict button”. In data field the user enters the credit card data and then clicks on predict button. The system will show whether the entered data is fraud or not.

From the ref [4] the below figure shows the graphical representation of linear regression. The figure shows that how the data can classified into equal classes and sometimes can overlap over each other.



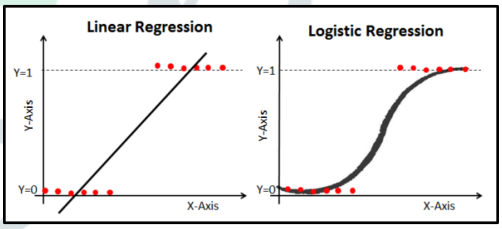


figure.2.5 Limitations of linear regression figure.2.6 Comparison of Linear & Logistic Regression

As shown in figure 2.5 The “star” represents the fraud transactions and “the disc” represents non-fraudulent transaction. The left graph represents linear regression that can classify data and the line that is passing through divides it into two equal categories or classes.[5]

The right side of the figure 2.5 represents the limitations of linear regression. When the data overlap over each other the line that is passing through cannot classify the data into equal categories or classes.

To overcome this problem referring to paper [6] logistic regression is being introduced. The below figure 2.6 shows the visual comparison between linear and logistic regression. As shown in the figure 2.6 the left side shows the graphical representation of linear regression. The logistic regression only deals with the continuous variable. After plotting the data, you will see a logistic regression hyperplane (straight line) which is not much understandable whether the data is being classified into equal classes or not.

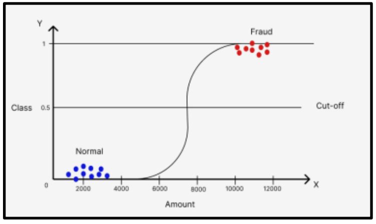


figure.2.7 Logistic Regression on Credit Card Data

$$Y = \frac{1}{1 + e^{-x}}$$

equation.2.1

The above figure 2.7 illustrates the graphical representation of logistic regression.

The logistic regression is related to classification of data using which we can classify the data into two equal classes.

In logistic regression there is an important function called as “Sigmoid Function”. The below equation.2.1 shows the formula of sigmoid function.

From the above equation.2.1, Y represents the output, e represents to euros constant (2.718), and -x represents independent variable which we need to transform.

So basically, sigmoid function in logistic regression is simply trying to convert the independent variable into a expression of probability that ranges between 0 & 1 with respect to the dependent variable.[7]

The logistic regression maps the real values in the interval of 0 & 1.

0 – means there is no possibility or probability of a certain occurrence.

1 – means there is probability of certain occurrence.

Logistic Regression can be helpful in Fraud detections, disease diagnosis, emergency detection, span or no-span.

The below figure 2.8 shows the fraud status based on gender using pie chart.

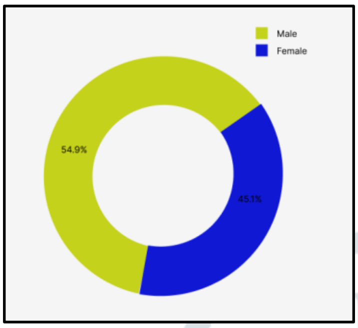


figure.2.8 Fraud Status based on Gender

From reference [8] the authors made the visualization using pie chart that based upon the gender category it is classifying that how much percentage of male and female has made the credit card fraud. The “yellow” represents fraud done by male and “blue” represents fraud done by female.

The above figure illustrates that 54.9% fraud are done by male and 45.1% fraud are done by female.

The above figure.2.9 shows that how many percentages of frauds are made from the different categories. The graph represents five categories – shopping, grocery, misc., transport, and homecare. Among which the shopping category has highest number of frauds (1.13%) whereas homecare has least frauds (0.16%).

This gives a brief representation to the user. The user can able to see in which categories most of the frauds are made.

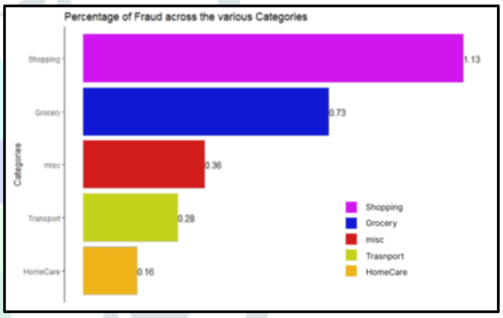


figure.2.9 Fraudulent transactions based on various categories

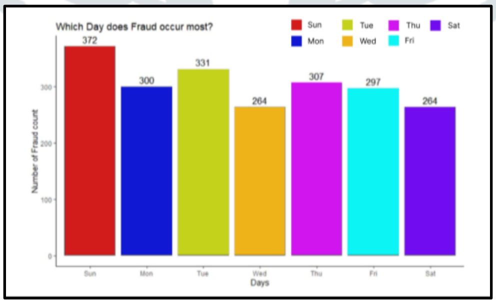


figure.2.10 Occurrence of Fraud transactions on weekdays

The above figure.2.10 from paper [9] shows the occurrence of fraud transactions on weekdays. The x-axis represents the days and y-axis represents number of fraud count. The figure illustrates that in weekdays which day has the highest number of fraud count. The color bars represent days of week “red” represents Sunday, “blue” represents Monday, “yellow” represents Tuesday, “orange” and so on. The bar graph shows that Sunday is the day on which most fraud occurs (372) and Wednesday & Saturday are days on which least fraud occurs (264).

There are many researches about credit card fraud detection where the authors usually imply the machine algorithms such as random forest, logistic regression, data visualization techniques. So, from the research papers we can make a brief overview that how credit card fraud detection system works as.

Here is a simplified system flow for a credit card fraud detection system using machine learning:

The below figure.2.11 shows the block diagram of the previous credit card fraud detection system.

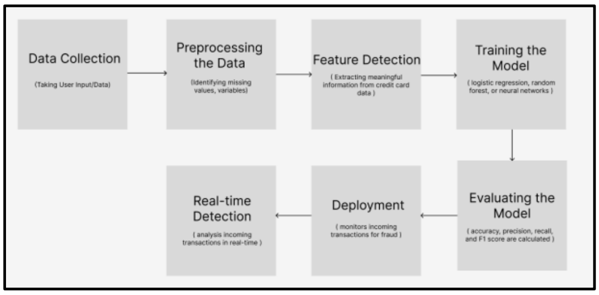

Data Collection
(Taking User Input/Data)
Preprocessing the Data
(Identifying missing values, variables)
Feature Detection
(Extracting meaningful information from credit card data)
Training the Model
(logistic regression, random forest, or neural networks)
Evaluating the Model
(accuracy, precision, recall, and F1 score are calculated)
Deployment
(monitors incoming transactions for fraud)
Real-time Detection
(analysis incoming transactions in real-time)

figure.2.11 Block Diagram of Credit Card Fraud System

From [10] the authors have made the detailed description about the flow of credit card fraud detection system. The steps are as follows-

1. **Data Collection**: Transaction data is collected from credit card transactions, including transaction amount, timestamp, merchant information, and other relevant features.
2. **Data Preprocessing**: The collected data is pre-processed to clean and prepare it for analysis. This may include handling missing values, encoding categorical variables, and scaling numerical features.
3. **Feature Engineering**: Relevant features are selected or engineered from the transaction data. This step aims to extract meaningful information that can help differentiate between legitimate and fraudulent transactions.
4. **Model Training**: A machine learning model is trained on the pre-processed data using a labelled dataset. Various ML algorithms such as logistic regression, random forest, or neural networks can be used for this purpose.
5. **Model Evaluation**: The trained model is evaluated using a separate test dataset to assess its performance in detecting fraudulent transactions. Performance metrics such as accuracy, precision, recall, and F1 score are calculated to evaluate the model's effectiveness.
6. **Deployment**: Once the model is trained and evaluated, it can be deployed into a production environment where it can continuously monitor incoming transactions for fraud.
7. **Real-time Detection**: In the production environment, the deployed model analysis incoming transactions in real-time. If a transaction is flagged as potentially fraudulent, it can be further investigated or blocked to prevent financial losses.

III. PROPOSED METHODOLOGY

The below figure.3.1 shows the block diagram of ARA-Credit Card Fraud Detection System.

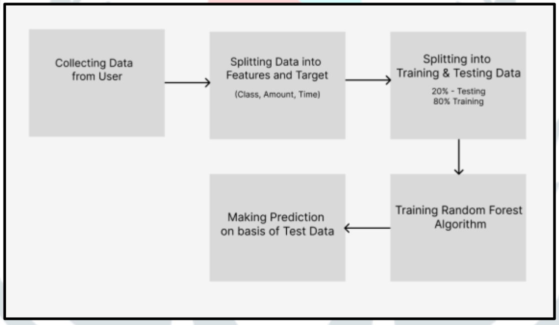

Collecting Data from User
Splitting Data into Features and Target
(Class, Amount, Time)
Splitting into Training & Testing Data
20% - Testing
80% Training
Training Random Forest Algorithm
Making Prediction on basis of Test Data

Figure.3.1 ARA system block diagram

The previous credit card fraud detection system relied on data visualization techniques and basic algorithms to identify the fraudulent transactions. Most of the systems implemented the Data Visualization techniques to predict the credit card data These techniques include - histogram, pie chart, bar chart, heat map, scatter plots and many more [11].

The proposed new credit card fraud detection system (ARA – Credit Card Fraud Detection System) integrates machine learning algorithms to enhance fraud detection capabilities. By leveraging historical transaction data, the new system can learn and adapt to new fraud patterns in real-time.

In our ARA – Credit Card Fraud Detection System to enhance detection, we are introducing several new features to further improve the system's accuracy and efficiency. These additional features are designed to address limitations identified in the current system and to incorporate advanced techniques to combat emerging fraud patterns. By integrating these new features, we aim to strengthen the fraud detection capabilities of our system and provide better protection for cardholders and financial institutions.

Comparing to other systems our “ARA Credit Card Fraud Detection System” has the better performance as it detects the fraudulent transaction within fraction of seconds. In our ARA system we are providing options such as- user can upload files, can view files report, can view account details, and change the password which are not available in any other systems. And most importantly our system provides the best security as the login can be done through one organization at a time, it will not be available to all users in organization.

Following are some features that we are providing:

1. **Login**: The login page provides the user authentication. Before doing any analysis or prediction on your data you must login to the dashboard for further access. For security purposes we are providing only organization level login feature i.e not every user can login. So, the data remains safe with only one super user of each organization.
2. **User Dashboard**: After Successful login the user enters the dashboard page. We are providing options such as- user can upload files, can view files report, can view account details, and change the password.
3. **Upload Files**: User can upload their files for prediction and analysis. User can upload either single data file or multi data file. The user must specify the name & select the file and then click on submit button

4. **Files Report**: user can see the uploaded data. We are providing 4 features inside files report – view data, delete, prediction and analysis.
5. **My Account:** user can see the details of the account.
6. **Change Password**: Our system also provides the feature of changing the password. If the user forgets the password, then he/she can easily access this feature.

IV. RESULTS:

Below figure shows the implementation of our ARA – Credit Card Fraud Detection System.

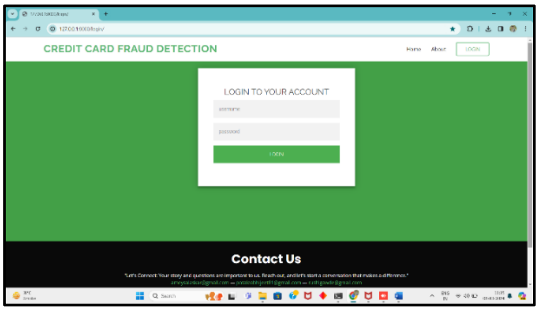


figure.4.1. Login Page

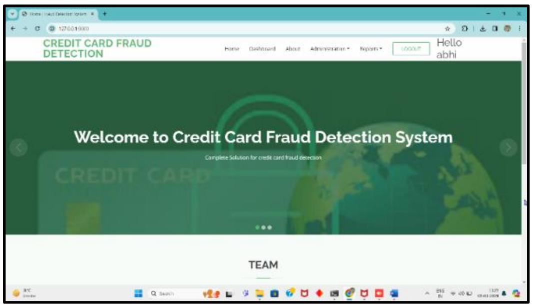

CREDIT CARD FRAUD DETECTION
Hello abhi
Welcome to Credit Card Fraud Detection System
TEAM

figure.4.2. Home Page

figure.4.1 and figure.4.2 represents the login page & home page of ARA – Credit Card Fraud Detection System..

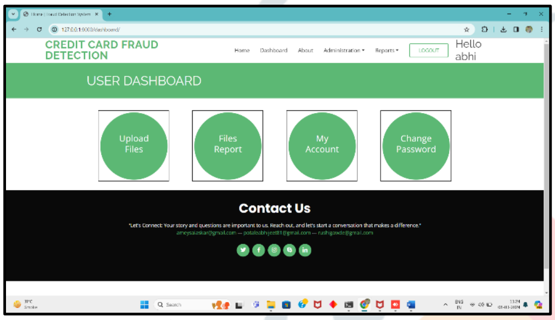


figure.4.3. Dashboard

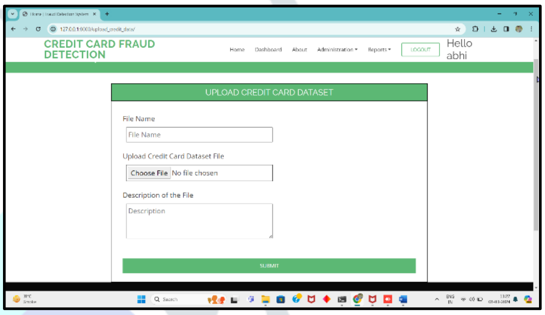

CREDIT CARD FRAUD DETECTION
Hello abhi
UPLOAD CREDIT CARD DATASET
File Name
Upload Credit Card Dataset File
Choose File No file chosen
Description of the File
SUBMIT

figure.4.4. uploading Data

Above figure.4.3 represents the dashboard of ARA – Credit Card Fraud Detection System.
Above figure.4.4 represents the user uploading data of ARA – Credit Card Fraud Detection System

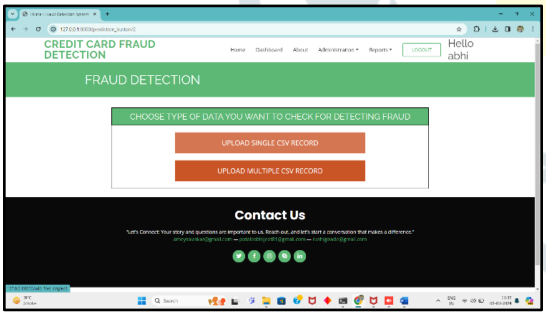

CREDIT CARD FRAUD DETECTION
Hello abhi
FRAUD DETECTION
CHOOSE TYPE OF DATA YOU WANT TO CHECK FOR DETECTING FRAUD
UPLOAD SINGLE CSV RECORD
UPLOAD MULTIPLE CSV RECORD
Contact Us

figure.4.5 uploading single/Multiple Data

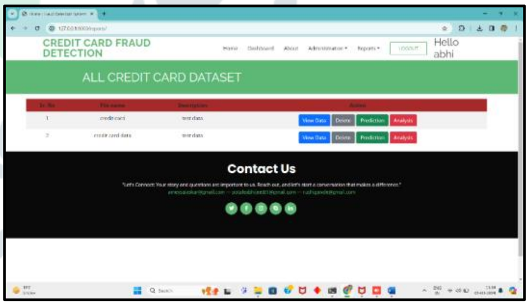


figure.4.6. Access the features for dataset

Above figure.4.5 & figure.4.6 represents uploading data and accessing features of ARA – Credit Card Fraud Detection System

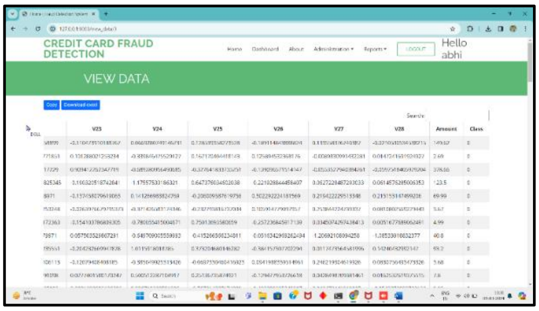


figure.4.7. View uploaded data

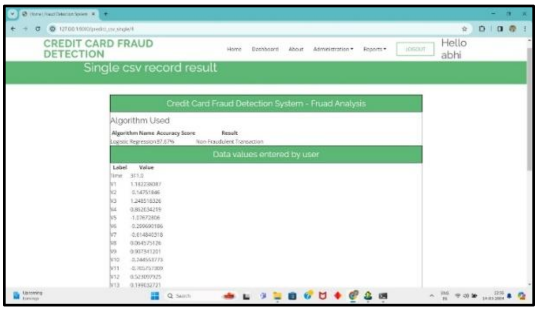

CREDIT CARD FRAUD DETECTION
Hello abhi
Single csv record result
Credit Card Fraud Detection System - Fruad Analysis
Algorithm Used
Data values entered by user

figure.4.8. Prediction of data

Above figure.4.7 & figure.4.8 represents uploaded data & prediction of data.

V. CONCLUSION AND FUTURE SCOPE:

Our study demonstrates the importance and effectiveness of using machine learning techniques for credit card fraud detection. Through the analysis of transaction data and the application of various machine learning algorithms, we can able to identify fraudulent transactions [12]. This not only helps financial institutions to save millions of dollars annually but also enhances the security and trust of cardholders.

Looking ahead, future research could focus on further improving the performance of fraud detection models by incorporating more advanced machine learning algorithms, such as deep learning and ensemble methods. Additionally, the integration of real-time data streams and the use of anomaly detection techniques could enhance the system's ability to detect emerging fraud patterns.

The future scope of credit card fraud detection is vast and evolving, with the enormous increase of fraudsters. We have now discovered fraudulent activity but we have not prevented it. Preventing known and unknown fraud in real time is not easy but it is feasible.

VI. REFERENCES:

[1] K.Ratna Sree Valli, P.Jyothi, G.Varun Sai, R.Rohith Sai Subash.(2020). “Credit card fraud detection using Machine learning algorithms”

[2] Meera AlEmad (2022). "Credit Card Fraud Detection Using Machine Learning”

[3] Varun Kumar K S, Vijaya Kumar V G, Vijay Shankar A, Pratibha K (2020) “Credit Card Fraud Detection using Machine Learning Algorithms”

[4] Hala Z Alenzi, Nojood O Aljehane (2020) “Fraud Detection in Credit Cards using Logistic Regression” Tabuk University, Tabuk City Kingdom Saudi Arabia

[5] Jonathan Kwaku Afriyie, Kassim Tawiah,Wilhemina Adoma Pels, Sandra Addai-Henne, Harriet Achiaa Dwamena, Emmanuel Odame Owiredu, Samuel Amening Ayeh, John Eshun (2023) “A supervised machine learning algorithm for detecting and predicting fraud in credit card transactions”

[6] Swarna B P, Asst Prof. Shivaleela. S (2021) “Credit card fraud detection system using machine learning” Dr Ambedkar institute of technology,Bangalore.

[7] Parmar, Jasmin & Patel, Achyut & Savsani, Mayur. (2020). Credit Card Fraud Detection Framework - A Machine Learning Perspective. International Journal of Scientific Research in Science and Technology. 431-435. 10.32628/IJSRST207671.

[8] Priya, G. & Saradha, S. (2021). “Fraud Detection and Prevention Using Machine Learning Algorithms: A Review.” 564-568. 10.1109/ICEES51510.2021.9383631.

[9] Dr. Savita Sangam (2020) “A Novel Machine Learning & NLP based approach for Analysing Phishing Attack”. International Journal for Research in Applied Science & Engineering Technology (IJRASET)

[10] S P, Maniraj & Saini, Aditya & Ahmed, Shadab & Sarkar, Swarna. (2019). “Credit Card Fraud Detection using Machine Learning and Data Science. International Journal of Engineering Research” and. 08. 10.17577/IJERTV8IS090031.

[11] Sadineni, Praveen Kumar. (2020). Detection of Fraudulent Transactions in Credit Card using Machine Learning Algorithms. 659-660. 10.1109/ISMAC49090.2020.9243545.

[12] Shirgave, Suresh & Awati, Chetan & More, Rashmi & Patil, Sonam. (2019). “A Review on Credit Card Fraud Detection Using Machine Learning. International Journal of Scientific & Technology Research.” 8. 1217-1220.