

# CREDIT CARD FRAUD DETECTION USING PREDICTIVE MODELLING

V. Rajasekhar<sup>1</sup>, Dr. M. Saravanamuthu<sup>2</sup>

<sup>1</sup>Student, Department of Computer Applications, Madanapalle institute of technology and science, India

<sup>2</sup>Asst. Professor, Department of Computer Applications, Madanapalle institute of technology and science, India

\*\*\*

## ABSTRACT

these days absolutely everyone exploitation their credit score playing cards for diverse capabilities Billions of coins' loss way to flawed credit score Cards. fraud detection may be a critical disadvantage shifting huge cash corporations that have magnified way to the enlargement in credit card transactions. This paper provides detection of frauds in credit card transactions, exploitation information processing strategies of prophetic modeling, provision Regression, and speak to Tree. provision regression is hired only for binary type of information. because it gives reasonably-priced end result to the user. credit card fraud dataset is amassed for the education checking out the system. the information set consists of credit card transactions in Sep 2013 with the aid of using European cardholders. This know-how set present transactions that passed off in 2 days, anywhere we've 490 frauds out of 284,800 transactions. By exploitation absolutely distinctive libraries in system the output are going to be written. Finally, confusion matrix is hired to devise verity and foretold know-how. This system gives the pretty 90th accuracy to the user.

**Key Words:** provision Regression, Confusion Matrix, Fraud and Non-Fraud, name Tree, predictive modelling.

## 1. INTRODUCTION

Online shopping is growing day by day. In modern society, the use of credit cards is increasing. Credit cards are used to purchase and service many things that humans need with the help of virtual and physical cards, but virtual cards are used for online transactions and physical cards are used for offline transactions. In physical card-based purchases, the credit card holder (a human) physically presents the card to pay for the resource. An attacker would need to steal a credit card to make a fraudulent transaction for this type of purchase. If the cardholder is unaware of the loss of the card, it can result in significant financial loss to the banking sector that provides access to the credit card.

Credit card fraud detection based on analysis of purchase data from previous cardholders promises a promising way to reduce credit card fraud success rates. Each cardholder can be represented by different patterns, as people tend to exhibit specific behavioral profiles. Therefore, it contains information about typical purchase categories, time spent on

last purchase, amount of money spent, etc. Deviations from such patterns are potential threats to the system.

## 2. LITERATURE REVIEW

The writer Sushmito Ghosh [1] has completed the studies on credit score card fraud detection. the writer makes use of the neural community idea for detection of frauds in credit score card. The synthetic neural community became skilled with specific frauds like lack of credit score playing cards, neglected playing cards, taking fraud playing cards from bank, counterfeit fraud, mail-order fraud and NRI (non-acquired issue) fraud, etc. The synthetic neural networks detected substantially extra fraud money owed with substantially fewer fake positives (decreased through a component of 30) over rule primarily based totally fraud detection procedures, this stuff completed through the usage of confusion matrix.

In 2014 the writer Gaurav Mhatre [2] completed the studies on credit score card fraud detection the usage of Hidden Markov Model (HMM). The writer fashions the series of operations in credit score card transaction processing the usage of a HMM and display how it could be used for the detection of frauds. An HMM is skilled with ordinary conduct of cardholder. If an incoming credit score card transaction isn't frequent through the HMM with sufficiently excessive probability, it's far taken into consideration to be fraudulent.

The writer John O. Awoyemi [3] completed the studies on credit score card fraud detection the usage of k-manner Algorithm, In this most effective the 2 features (referred to as fraud and non-fraud) with the maximum variance have been used to educate the Support vector gadget set of rules. The set of rules may have 2 clusters namely, zero for non-fraud and 1 for fraud. The writer additionally experimented with specific values for the hyper aircraft parameters, however all of them produced comparable outputs. through Changing the dimensionality of the data (lowering it to extra dimensions than 2) additionally made the extra distinction at the very last values.

## 3. PROPOSED ALGORITHM

In proposed approach Logistic regression, predictive modelling and choice tree is used for frauds detection. Logistic regression set of rules is used for binary type of data. So that it's going to provide the higher accuracy to

stumble on the frauds. Logistic Regression gadget gaining knowledge of set of rules for fraud detection and print predicting rating in line with the set of rules. Finally, Confusion matrix turned into plotted on genuine and predicted. If the range of transactions passed off on one card at one time in order to be taken into consideration as a fraud. at that point choice tree is used to take the choice on every card. it really works primarily based totally at the range of transactions are occurring on every credit score card. choice tree saves the masses of time and additionally offers the higher accuracy.

## Logistic Regression

Logistic regression is much like linear regression however interpret curve is the use of herbal logarithm of the “odd” of the goal variable that is evolved with the aid of using statistician David cox in 1958.To expect the possibility of an final results which has values both 0 or one, sure or no and fake or true. The prediction is primarily based totally on the usage of one or numerous predictors; logistic regression produces logistic curves, which can be values of 0 and one.

Linear regression version is used to expect binary goal variables. Binary goals variables both zero or one. The linear regression equation

$$Y = \beta_0 + \beta_1 + \sum_i \dots \dots \dots (1)$$

In equation (2) the real price of Y is binary variable, then the anticipated Y may be much less than 0 or extra than one. Logistic Regression or log it version is a regression version wherein the established variable is express and analyzes the connection among a couple of unbiased variables. Binary Logistic Regression version is used to estimate the possibility of a binary reaction primarily based totally on one or greater predictors.

## The Logistic curve

The technique of logistic regression is becoming a regression curve,  $y = f(x)$  whilst y includes binary code (0, 1—failure, success) data. The reaction variable is a binary variable and x is numerical. In equation (3), the connection among x and y is decided via way of means of logistic curve, to match curve the usage of logistic regression. The form of logistic curve is an s\_ fashioned or sigmoid curve. A logistic curve startswith sluggish, linear growth, accompanied via way of means of exponential growth, which then sluggish once more to a strong state. A easy logistic characteristic is described via way of means of the formula

$$Y = \frac{e^x}{1 + e^x} = \frac{1}{1 + e^{-x}} \dots \dots \dots (2)$$

Logistic Regression is a type technique that go back the chance of binary structured variable can be anticipated from the unbiased variables.

## Predictive Modelling

Once we create a version, we are able to use many times, to decide the opportunity of outcomes. So are expecting version is reusable. Historical records is used to teach an algorithm. The predictive modeling procedure is an iterative procedure and regularly includes education the version, the usage of more than one fashions at the equal dataset.

### The Process of Predictive Modeling

- Creating the version:
- To create a version to run one or greater algorithms at the records set.
- Testing a version: The trying out is performed on beyond records to look how the great version predicts.
- Validating a version: Using visualization gear to validate the version.
- Evaluating a version: Evaluating the great suit version from the fashions used and selecting the version proper outfitted for the records.

## Decision Trees

Decision timber are used to pick among numerous guides of action. It gives powerful shape to analyze the feasible outcomes. Decision timber use tree shape to construct type or regression model. A selection tree is a flowchart like tree shape, in which non leaf node denotes a check on attribute.

Which corresponds to the first-rate predictor with inside the statistics? Decision timber may be used to analyses the explicit statistics and numerical statistics. One of the set of rules is used to construct a de is selection tree is ID3 that's evolved through J. Ross Quinlan. This set of rules makes use of pinnacle down method and grasping search. The pinnacle down method is recursive divide-and-triumph over method. Backtracking isn't always used on this set of rules. The studying of selection timber from education tuples the use of ID3 and CART (Classification and Regression Trees) algorithms have been invented independently of each other across the identical time. The ID3 and CART algorithms are used to generate selection tree induction. These algorithms additionally observe pinnacle down method in recursive manner. Decision tree is constructed primarily based totally on education tuples are recursively partitioned into smaller subsets.

### System Architecture

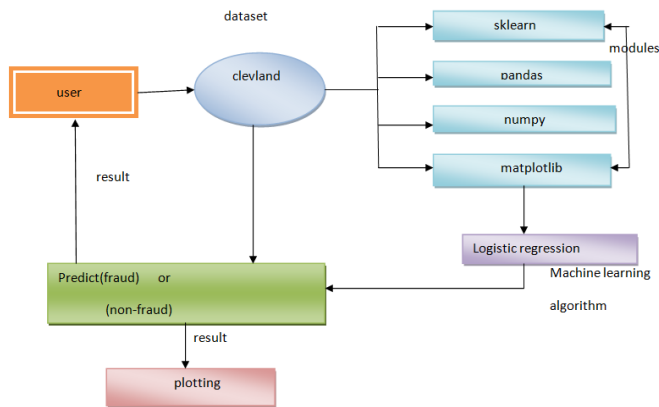


Fig.1. system Architecture

Fig.1. represents the how the frauds are detected is described the use of grade by grade process. Initially the customer’s credit score card is taken as enter and it's far in comparison with dataset. With the same homes if any dataset is observed in dataset then the associated output can be exceeded to person. Otherwise via way of means of the use of one-of-a-kind python libraries, one-of-a-kind techniques and algorithms the output can be expected. For the expected output the quantity of cash loss the use of credit score card is plotted via way of means of the use of confusion matrix and additionally the output is shared with the person as fraud or non-fraud.

### 4. RESULTS

Below Fig.2 Represents the Fraud and Non-fraud cards and for each card the amount money loss. For every 100 normal credit cards there must be fraud occurred for minimum 30 cards.

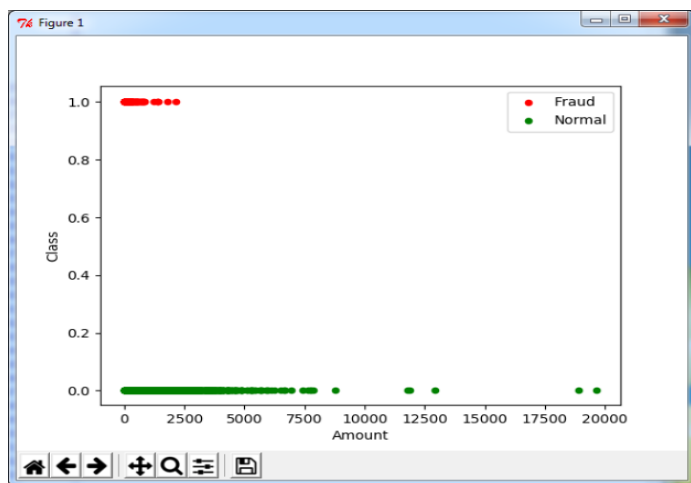


Fig.2: Fraud detection

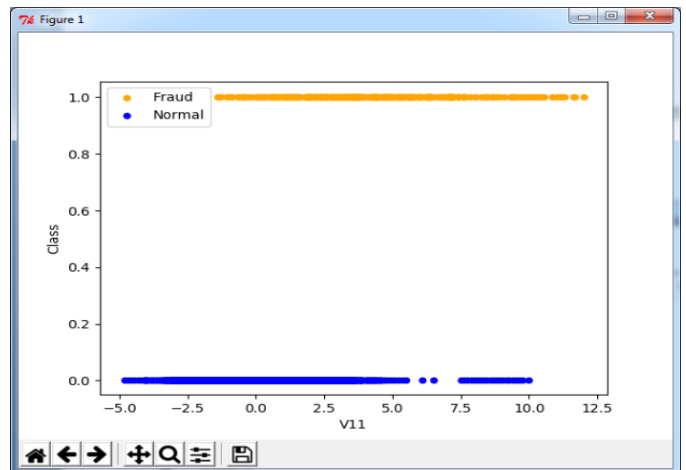


Fig.3.Fraud and normal cards

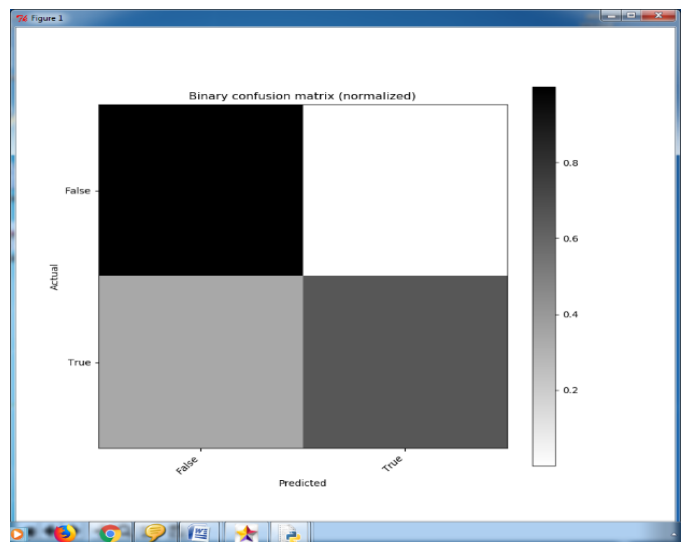


Fig.4. Confusion Matrix for Predicted value

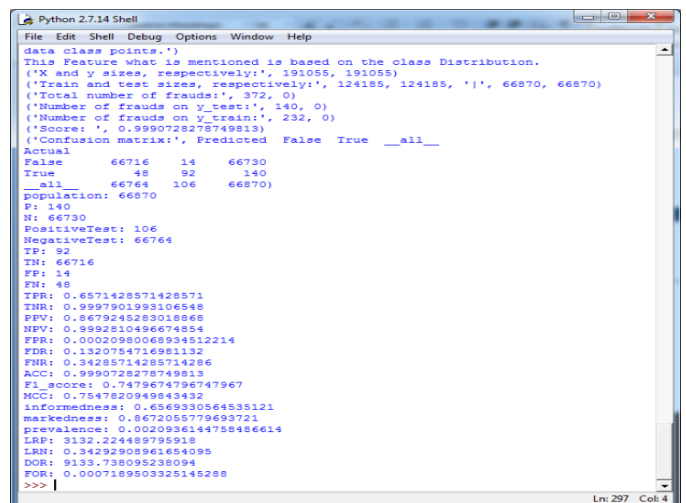


Fig.5. No. of frauds non-frauds

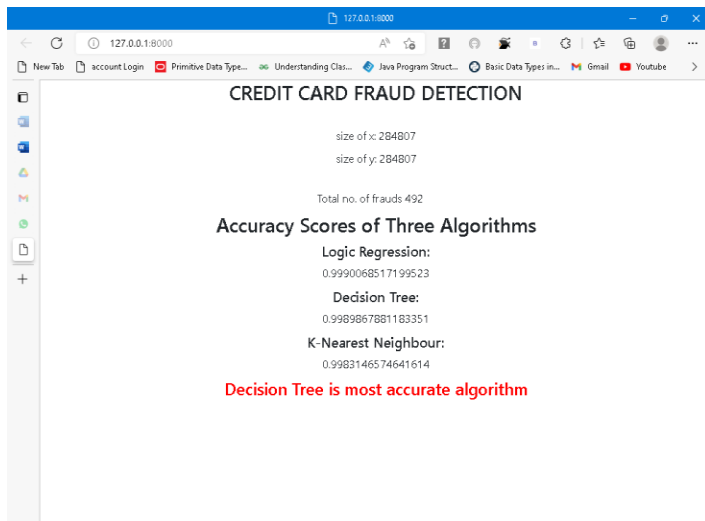


Fig.6. accuracy of models on data

Nonyelum. "Data mining application in credit card fraud detection system." *Journal of Engineering Science and Technology* 6.3 (2011): 311-322.

- [4] 4. Quinlan, J. Ross. "Induction of decision trees." *Machine learning* 1.1 (1986): 81-106.
- [5] 5. Quinlan, J. R. (1987). "Simplifying decision trees". *International Journal of Man-Machine Studies*. 27 (3): 221. doi:10.1016/S0020-7373(87)80053-6.
- [6] 6. Salazar, Addisson, Gonzalo Safont, and Luis Vergara. "Surrogate techniques for testing fraud detection algorithms in credit card operations." *Security Technology (ICCST), 2014 International Carnahan Conference on. IEEE, 2014.*

## 5. CONCLUSION

Detection of Frauds in credit score playing cards may be very tough process. Detecting every credit score card for frauds, takes the masses of time. But on this set of rules the version simplest predicts for the frauds for extra transactions which are occurring the usage of simplest one credit score card. Most of the time this method is used to expect the correct outcomes.

This gadget gaining knowledge of fraud detection educational confirmed the way to address the hassle of credit score card fraud detection the usage of gadget gaining knowledge of. It within reason clean to give you a easy version, enforce it in Python and get incredible outcomes for the Credit Card Fraud Detection challenge on Kaggle.

## REFERENCES

- [1] Ghosh [1], Sushmito, and Douglas L. Reilly. "Credit card fraud detection with a neural-network." *System Sciences, 1995. Proceedings of the Twenty-eighth Hawaii International Conference on. Vol. 3. IEEE, 1994.*
- [2] Srivastava [2], Abhinav, et al. "Credit card fraud detection using hidden Markov model." *IEEE Transaction on dependent and secure computing* 5.1 (2008): 40-48.
- [3] J. O. Awoyemi [3], A. O. Adetunmbi and S. A. Oluwadare, "Credit card fraud detection using machine learning techniques: A comparative analysis," *2017 International Conference on Fraud detection on creeit card using Computing Networking and Informatics (ICCNI), 2017, pp. 1-9, doi: 10.1109/ICCNI.2017.8123782. Ogwueleka, Francisca*