IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 9, June 2022
Impact Factor: 6.252

Machine Learning for Credit Card Fraud Detection System

Saurabh Rahangdale[1], Sayali Gedam[2], Dikshita Pusam[3], Pratik Harne[4]
Students, Department of Information Technology[1,2,3,4]
Government College of Engineering, Amravati, Maharashtra, India

Abstract: *In recent years, for banks has become terribly troublesome for police investigation the fraud in credit-card system. Machine learning plays a significant role for police investigation the credit-card fraud within the transactions. For predicting these transactions banks build use of assorted machine learning methodologies, past knowledge has been collected and new options square measure been used for enhancing the prophetical power. The performance of fraud police investigation in credit-card transactions is greatly full of the sampling approach on data-set, choice of variables and detection techniques used. This paper investigates the performance of supply regression for credit-card fraud detection. Dataset of credit-card transactions is collected from Kaggle and it contains a complete of two,84,808 credit-card transactions of a ecu bank knowledge set. It considers fraud transactions because the "positive class" and real ones because the "negative class". the info set is very unbalanced, it's concerning zero.172% of fraud transactions and also the rest square measure real transactions.*

Keywords: Fraud detection, Credit-card, Logistic regression Algorithm

I. INTRODUCTION

Credit card fraud could be a large move term for larceny and fraud committed mistreatment or involving at the time of payment by mistreatment this card. the aim is also to buy product while not paying or to transfer unauthorized funds from associate account. credit-card fraud is additionally associate add on to fraud. As per the data from the us Federal Trade Commission, the larceny rate of identity had been holding stable throughout the mid-2000s, however it absolutely was inflated by twenty-one % in 2008. albeit credit-card fraud, that crime that most of the people accompany ID larceny, bated as a proportion of all ID larceny complaints in 2000, out of thirteen billion transactions created annually, around ten million or one out of each 1300 transactions clad to be fallacious.

Also, 0.05% (5 out of each 10,000) of all monthly active accounts was dishonest. Today, fraud detection systems area unit introduced to manage one-twelfth of 1% of all transactions processed that still interprets into billions of greenbacks in losses. Credit-card Fraud is one among the most important threats to business institutions these days. However, to combat the fraud effectively, it's necessary to initial perceive the mechanisms of death penalty a fraud. credit card fraudsters use an oversized range of how to commit fraud. In straightforward terms, credit card Fraud is outlined as "when a personal uses another individuals' credit card for private reasons whereas the owner of the cardboard and also the card institution aren't conscious of the very fact that the card is being used". Card fraud begins either with the felony of the physical card or with the necessary knowledge related to the account, together with the card account range or alternative info that essentially be out there to a businessperson throughout a permissible dealing. Card numbers usually the Primary Account Number (PAN) area unit usually reprinted on the card, and a tape on the rear contains the info in machine-readable format. It contains the subsequent Fields:

- Name of card holder
- Card number
- Expiration date
- Verification/CVV code
- Type of card

There are additional ways to commit credit card fraud. Fraudsters are terribly gifted and fast-moving folks. within the ancient approach, to be known by this paper is Application Fraud, wherever someone can provide the incorrect info

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 9, June 2022
Impact Factor: 6.252

regarding himself to induce a credit card. there's additionally the unauthorized use of Lost and taken Cards, that makes up a big space of credit card fraud. There square measure additional enlightened credit card fraudsters, beginning with those that manufacture faux and Doctored Cards; there also are those that use Skimming to commit fraud. they're going to get this info survived either the magnetic strip on the rear of the credit card, or the info keep on the sensible chip is derived from one card to a different. website biological research and False bourgeois Sites on the web have gotten a well-liked methodology of fraud for several criminals with a talented ability for hacking. Such sites square measure developed to induce folks handy over their credit card details while not knowing they need been swindled.

II. RELATED WORK

A.Shen etal (2007) demonstrate the potency of classification models to credit card fraud detection downside and also the authors projected the 3 classification models i.e.., call tree, neural network and supplying regression. Among the 3 models' neural network and supplying regression outperforms than the choice tree. M.J.Islam et al (2007) projected the applied math frame work for creating call below uncertainty. once reviewing theorem theory, naïve Thomas Bayes classifier and k-nearest neighbour classifier is enforced and applied to the dataset for credit card system.Y. Sahin and E. Duman (2011) has cited the analysis for credit card fraud detection and used seven classification ways took a significant role. In this work they need enclosed call trees and SVMs to decrease the chance of the banks. they need recommended Artificial Neural networks and supplying Regression classification models ar a lot of useful to enhance the performance in sleuthing the frauds. Y. Sahin, E. Duman (2011) has cited the analysis, used Artificial Neural Network and applying Regression Classification and explained ANN classifiers outdo LR classifiers in determination the matter below investigation. Here the coaching knowledge sets distribution became a lot of biased and also the distribution of the coaching knowledge sets became a lot of biased and also the potency of all models minimized in catching the dishonest transactions.

III. PROPOSED TECHNIQUE AND EXPERIMENTAL RESULTS

The proposed techniques are used in this paper, for detecting the frauds in credit card system. The comparison is made for different machine learning algorithms such as Logistic Regression, Decision Trees, Random Forest, to determine which algorithm gives suits best and can be adapted by credit card merchants for identifying fraud transactions.

3.1 Processing Steps

The entire document should be in Times New Roman or Times font. Type 3 fonts must not be used. Other font types may be used if needed for special purposes.

Algorithm steps:

- Step 1: Read the dataset.
- Step 2: Random Sampling is done on the data set to make it balanced.
- Step 3: Divide the dataset into two parts i.e., Train dataset and Test dataset.
- Step 4: Feature selection are applied for the proposed models.
- Step 5: Accuracy and performance metrics has been calculated to know the efficiency for different algorithms.
- Step 6: Then retrieve the best algorithm based on efficiency for the given dataset.

3.2 Logistic Regression

Logistic Regression is one amongst the classification algorithmic rule, wont to predict a binary values during a given set of independent variables (1 / 0, Yes / No, True / False). To represent binary / categorical values, dummy variables area unit used. For the aim of special case within the logistic regression may be a linear regression, once the ensuing variable is categorical then the log of odds area unit used for variable quantity and additionally it predicts the chance of incidence of an incident by fitting information to a logistic function. Such as

$$O = e^{(I0 + I1*x)} / (1 + e^{(I0 + I1*x)})$$

Whereas, O is the predicted output

I0 is the bias or intercept term

I1 is the coefficient for the single input value (x).

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 9, June 2022
Impact Factor: 6.252

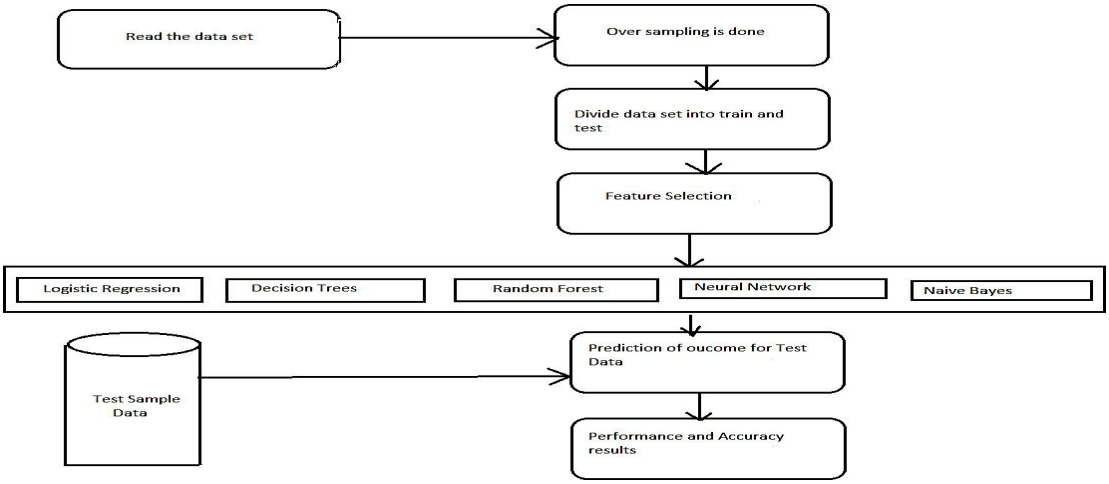

Figure 1: System Architecture

Each column within the input {data|input file|computer file} has an associated I constant (a constant real value) that has to be learned from the training data.

$y = e^{(b0 + b1*x)} / (1 + e^{(b0 + b1*x)})$

Logistic regression is started with the simple linear regression equation within which variable quantity may be enveloped in an exceedingly link perform i.e., to start with supply regression, I'll 1st write the simple linear regression equation with variable quantity enveloped in an exceedingly link function:

$A(O) = \beta 0 + \beta(x)$

Whereas, A (): link function

O: outcome variable

x: dependent variable

A function is established using two things:

1) Probability of Success(pr) and

2) Probability of Failure(1-pr).

pr should meet following criteria:

a) probability must always be positive (since p >= 0)

b) probability must always be less than equals to 1 (since pr <= 1). By applying exponential in the first criteria and the value is always greater than equals to 1.

$pr = \exp(\beta_o + \beta(x)) = e^{(\beta_o + \beta(x))}$

For the second condition, same exponential is divided by adding 1 to it so that the value will be less than equals to 1

$pr = e^{(\beta o + \beta(x))} / e^{(\beta o + \beta(x))} + 1$

Logistic function is employed within the supplying regression during which cost function quantifies the error, because it models response is compared with truth value.

$X(\theta) = -1/m * (\sum y_i \log(h\theta(x_i)) + (1 - y_i) \log(1 - h\theta(x_i)))$

Whereas, hθ(xi): logistic function

yi: outcome variable Gradient descent is a learning algorithm

3.3 Performance Metrics

The basic performance measures derived from the confusion matrix. The confusion matrix may be a two-by-two matrix table contains four outcomes created by the binary classifier. numerous measures like sensitivity, specificity, accuracy and error rate are derived from the confusion matrix.

Accuracy is calculated as the total number of 2 correct predictions(A+B) divided by the overall number of the dataset(C+D). It is calculated as (1-error rate).

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 9, June 2022
Impact Factor: 6.252

Accuracy=A+B/C+D

Whereas, A=True Positive
B=True Negative
C=Positive
D=Negative

Error rate is calculated as total number of two incorrect predictions (F+E) divided by total number of the dataset (C+D).

Error rate=F+E/C+D

Whereas, E=False Positive
F=False Negative
C=Positive
D=Negative

Sensitivity is calculated as number of correct positive predictions(A) divided by total number of positives(C).

Sensitivity=A/C

Specificity is calculated as the number of correct negative predictions(B) divided by the total number of negatives(D).

Specificity=B/D

Accuracy, Error-rate, Sensitivity and Specificity area unit wont to report the performance of the system to discover the fraud within the credit card.

In this paper, 3 machine learning algorithms are developed to find fraud in credit card system. to judge the algorithms, 60 minutes of the dataset is employed for coaching and four-hundredth is employed for testing and validation. Accuracy, error rate, sensitivity and specificity are used to judge for various variables for three} algorithms as shown in Table 1. The accuracy result's shown for logistic regression is 92.7. The comparative results show that the Random Forest performs higher than the logistic regression and decision tree techniques.

Table I: Comparison between Three Different Algorithms

Feature Selection	Logistic regression	Decision tree	Random Forest
For 5 variables	87.2	89	90.1
For 10 variables	88.6	92.1	93.6
For all Variables	90.0	94.3	95.5

IV. CONCLUSION

In this paper, Machine learning technique supply regression is employed to notice the fraud in credit card system. Sensitivity, Specificity, accuracy and error rate are accustomed evaluate the performance for the proposed system. The accuracy for supply regression is 90.0%.

ACKNOWLEDGMENT

The heading of the Acknowledgment section and the References section must not be numbered. Causal Productions wishes to acknowledge Michael Shell and other contributors for developing and maintaining the IEEE LaTeX style files which have been used in the preparation of this template.

REFERENCES

[1]. S. Bhattacharyya, S. Jha, K. Tharakunnel, J. C. Westland, "Data mining for credit card fraud: A comparative study", Decision Support Systems, vol. 50, no. 3, pp. 602-613, 2011.

[2]. S. Maes, K. Tuyls, B. Vanschoenwinkel, B. Manderick, "Credit card fraud detection using Bayesian and neural networks", Proceedings of the 1st international naiso congress on neuro fuzzy technologies, pp. 261-270, 2002.

[3]. M. Wegmuller, J. P. von der Weid, P. Oberson, and N. Gisin, “High resolution fiber distributed measurements with coherent OFDR,” in Proc. ECOC’00, 2000, paper 11.3.4, p. 109.

[4]. S. Maes, K. Tuyls, B. Vanschoenwinkel, B. Manderick, "Credit card fraud detection using Bayesian and neural networks", Proceedings of the 1st international naiso congress on neuro fuzzy technologies, pp. 261-270, 2002

[5]. S. Patil, H. Somavanshi, J. Gaikwad, A. Deshmane, R. Badgujar, "Credit Card Fraud Detection Using Decision Tree Induction Algorithm", International Journal of Computer Science and Mobile Computing (IJCSMC), vol.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 9, June 2022
Impact Factor: 6.252

4, no. 4, pp. 92-95, 2015, ISSN ISSN: 2320-088X.

[6]. K. Chaudhary, B. Mallick, "Credit Card Fraud: The study of its impact and detection techniques", International Journal of Computer Science and Network (IJCSN), vol. 1, no. 4, pp. 31-35, 2012, ISSN ISSN: 2277-5420.

[7]. G. Singh, R. Gupta, A. Rastogi, M. D. S. Chandel, A. Riyaz, "A Machine Learning Approach for Detection of Fraud based on SVM", International Journal of Scientific Engineering and Technology, vol. 1, no. 3, pp. 194-198, 2012, ISSN ISSN: 2277-1581.

[8]. B.Meena, I.S.L.Sarwani, S.V.S.S.Lakshmi," Web Service mining and its techniques in Web Mining" IJAEGT, Volume 2,Issue 1 , Page No.385-389.

[9]. A. Shen, R. Tong, Y. Deng, "Application of classification models on credit card fraud detection", Service Systems and Service Management 2007 International Conference, pp. 1-4, 2007.