Journal of Engineering Sciences
Vol 11, Issue 4 , April/2020
ISSN NO:0377-9254

CREDIT CARD FRAUD DETECTION USING LOGISTIC REGRESSION

K. Gowthami[1], K V L E Praneetha[2], G. Vinitha[3], Ch. Reddamma Kumari[4]
P. Sandhya Krishna[5]
[1,2,3,4]IV B.Tech, Department of Information Technology, Vignan's Nirula Institute of Technology & Science for Women, Peda Palakaluru, Guntur-522009, Andhra Pradesh, India.
[5]Asst.Professor, Department of Information Technology, Vignan's Nirula Institute of Technology & Science for Women, Peda Palakaluru, Guntur-522009, Andhra Pradesh, India.
sandhyamcaviit@gmail.com

ABSTRACT

Use of online transactions in day to day life has been increasing since last decade due to advancements in technology and network connectivity. Due to ease, simplicity and user friendliness of the online transaction system, new users are constantly joining the vast population benefitting from such system. Credit card fraud resulting from misuse of the system is defined as theft or misuse of one's credit card information which is used for personal gains without the permission of the card holder. To detect such frauds, it is important to check the usage patterns of a user over the past transactions. Credit card fraud refers to the physical loss of credit card or loss of sensitive credit card information. Many machine-learning algorithms can be used for detection.

Keywords— Credit card fraud, Machine learning algorithms, logistic regression.

1. INTRODUCTION

It is important that credit card companies are able to recognize fraudulent credit card transactions so that customers are not charged for items that they did not purchase [1]. Credit card fraud are often outlined because the bootleg use of any system or, criminal activity through the utilization of physical card or card info while not the information of the cardholder[2][3]. The MasterCard is also physical or virtual in an exceedingly physical-card, the cardholder presents his or her card physically to a business person for creating a payment. To hold out deceitful transactions during this quite purchase, a wrongdoer needs to steal the MasterCard.Machine learning algorithms are employed to analyse all the authorized transactions and report the suspicious ones[4][5]. These reports are investigated by professionals who contact the cardholders to confirm if the transaction was genuine or fraudulent. The investigators provide a feedback to the automated system which is used to train and update the algorithm to eventually improve the fraud-detection performance over time[6][7]. 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

2. LITERATURE REVIEW

Prepare Data for Logistic Regression

The assumptions made by logistic regression about the distribution and relationships in your data are much the same as the assumptions made in linear regression[8][9]. Much study has gone into defining these assumptions and precise probabilistic and statistical language is used. My advice is to use these as guidelines or rules of thumb and experiment with different data

preparation schemes[10][11]. Ultimately in predictive modelling machine learning projects you are laser focused on making accurate predictions rather than interpreting the results[12]. As such, you can break some assumptions as long as the model is robust and performs well.

Binary Output Variable: This might be obvious as we have already mentioned it, but logistic regression is intended for binary (two-class) classification problems[13]. It will predict the probability of an instance belonging to the default class, which can be snapped into a 0 or 1 classification[14][15].

Remove Noise: Logistic regression assumes no error in the output variable (y), consider removing outliers and possibly misclassified instances from your training data[16][17]. Gaussian distribution: Logistic regression is a linear algorithm (with a non-linear transform on output[18]). It does assume a linear relationship between the input variables with the output[41][42]. Data transforms of your input variables that better expose this linear relationship can result in a more accurate model[19][43]. For example, you can use log, root, Box-Cox and other univariate transforms to better expose this relationship[20][44].

Remove Correlated Inputs: Like linear regression, the model can over fit if you have multiple highly-correlated inputs[45][46]. Consider calculating the pair wise correlations between all inputs and removing highly correlated inputs[21][22]. Fail to Converge: It is possible for the expected likelihood estimation process that learns the coefficients to fail to converge[23]. This can happen if there are many highly correlated inputs in your data or the data is very sparse (e.g. lots of zeros in your input data)[24][25].

Further Reading
There is a lot of material available on logistic regression. It is a favorite in may disciplines such as life sciences and economics[46][47].

Logistic Function
Logistic regression is named for the function used at the core of the method, the logistic function[26]. The logistic function, also called the sigmoid function was developed by statisticians to describe properties of population growth in ecology, rising quickly and maxing out at the carrying capacity of the environment[27][28]. It's an S-shaped curve that can take any real-valued number and map it into a value between 0 and 1, but never exactly at those limits[48].

1 / (1 + e^-value)

Where e is the base of the natural logarithms (Euler's number or the EXP() function in your spreadsheet) and value is the actual numerical value that you want to transform[29]. Below is a plot of the numbers between -10 and 10 transformed into the range 0 and 1 using the logistic function[30][31].

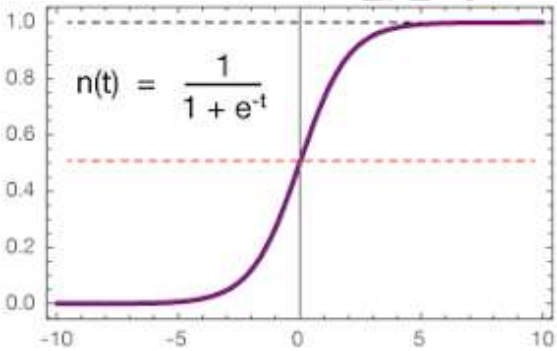

Fig 1 Logistic function representation

3. PROPOSED SYSTEM

In this proposed system we are doing the analysis of techniques of fraud detection which is user friendly and secure[32]. This system analyzes the feasibility of credit card fraud detection and proposes this detection procedures and its empirical process.

It contains only numerical input variables which are the result of a PCA transformation[33][34]. Unfortunately, due to confidentiality issues, we cannot provide the original features and more background information about the data[35][36]. Features V1, V2... V28 are the principal components obtained with PCA, the only features which have not been transformed with PCA are 'Time' and 'Amount'. Feature 'Time' contains the seconds elapsed between each transaction and the first transaction in the dataset[37]. The feature 'Amount' is the transaction Amount, this feature can be used for example-dependant cost-sensitive learning[38]. Feature 'Class' is the response variable and it takes negative value in case of fraud and positive value otherwise[39][40].

Implementation code

```
creditcard.drop(columns = 'time', inplace = True)
# Normalize the 'amount' column
scaler = StandardScaler()
creditcard['amount'] = scaler.fit_transform(creditcard['amount'].values.reshape(-1, 1))
# creditcard.drop(columns = 'amount', inplace = True)
X = creditcard.iloc[:, :-1]
y = creditcard.iloc[:, -1]
Xtrain, Xtest, ytrain, ytest = train_test_split(X, y, test_size = .33, stratify = y, random_state = seed)
logitreg_parameters = {'C': np.power(10.0, np.arange(-3, 3))}
logitreg = LogisticRegression(verbose = 3, warm_start = True)
logitreg_grid = GridSearchCV(logitreg, param_grid = logitreg_parameters, scoring = 'roc_auc', n_jobs = 70)
logitreg_grid.fit(Xtrain, ytrain) # logitreg_grid.best_params_ ; logitreg_grid.best_estimator_
# on OVER-Sampled TRAINing data
print("\n Thre recall score on Training data is:")
print recall_score(ytrain, logitreg_grid.predict(Xtrain)) # 0.58
print("\n Thre precision score on Training data is:")
print precision_score(ytrain, logitreg_grid.predict(Xtrain)) # 0.89
# on the separated TEST data
print("\n Thre recall score on Test data is:")
print recall_score(ytest, logitreg_grid.predict(Xtest)) # 0.58
print("\n Thre precision score on Test data is:")
print precision_score(ytest, logitreg_grid.predict(Xtest)) # 0.86
print("\n Thre Confusion Matrix on Test data is:")
print confusion_matrix(ytest, logitreg_grid.predict(Xtest))
```

4. RESULTS

The code prints out the number of fraud cases and no fraud cases detected and compares it with the actual values. This is used to calculate the accuracy of the different algorithms. The complete dataset is also used for testing and the results are printed. These results along with the classification report for algorithm is given in the output as follows, the positive values in the result gives no fraud occurred and the negative values in the result gives the fraud occurred.

Result screens python

Fig 2 open the file in python GUI

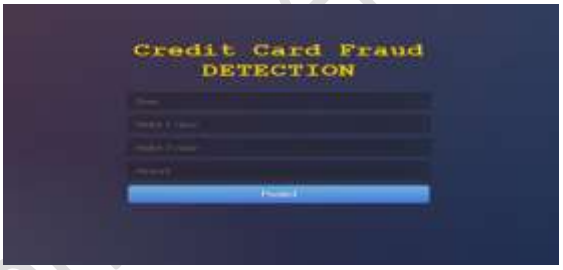

Fig 3 enter value1 and value2

Fig 4 Negative value results the fraud occurred

Fig 5 Positive values in results no fraud occurred

Output screens of jupyter

Fig 6 Open Dataset

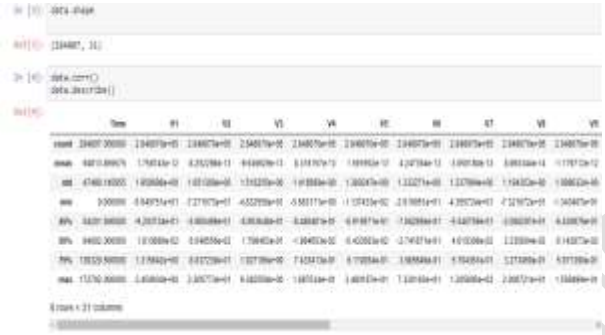

Fig 7 Check the shape and correlation

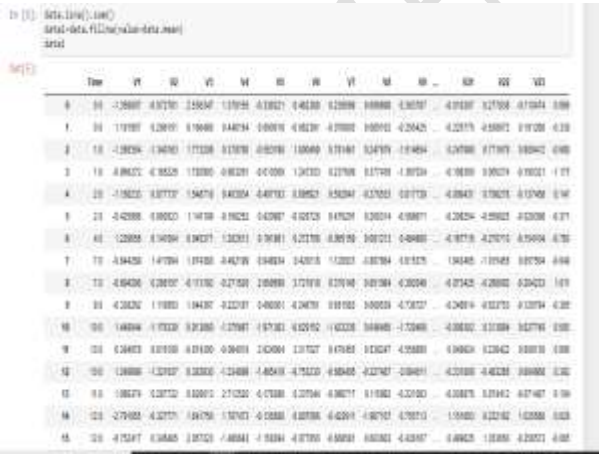

Fig 8 Value of Correlation between Variables

Fig 9 Accuracy of different algorithms

5. CONCLUSION

Fraud detection is a complicated issue, requires more planning and practice to implement with machine learning algorithms. This makes sure that the customer's cash is secure and no longer without difficulty tampered with. Usage of credit cards is increased in each field of day to day life. Credit card Frauds also be increased. To enhance security of the transaction systems an effective way efficient credit card detection system is established.

REFERENCES

[1]. Lakshman Narayana Vejendla and A Peda Gopi, (2019)," Avoiding Interoperability and Delay in Healthcare Monitoring System Using Block Chain Technology", Revue d'Intelligence Artificielle , Vol. 33, No. 1, 2019,pp.45-48.

[2]. Gopi, A.P., Jyothi, R.N.S., Narayana, V.L. et al. (2020), "Classification of tweets data based on polarity using improved RBF kernel of SVM" . Int. j. inf. tecnol. (2020). https://doi.org/10.1007/s41870-019-00409-4.

[3]. A Peda Gopi and Lakshman Narayana Vejendla, (2019)," Certified Node Frequency in Social Network Using Parallel Diffusion Methods", Ingénierie des Systèmes d' Information, Vol. 24, No. 1, 2019,pp.113-117.. DOI: 10.18280/isi.240117

[4]. Lakshman Narayana Vejendla and Bharathi C R ,(2018),"Multi-mode Routing

Algorithm with Cryptographic Techniques and Reduction of Packet Drop using 2ACK scheme in MANETs", Smart Intelligent Computing and Applications, Vo1.1, pp.649-658. DOI: 10.1007/978-981-13-1921-1_63 DOI: 10.1007/978-981-13-1921-1_63

[5]. Lakshman Narayana Vejendla and Bharathi C R, (2018), "Effective multi-mode routing mechanism with master-slave technique and reduction of packet droppings using 2-ACK scheme in MANETS", Modelling, Measurement and Control A, Vol.91, Issue.2, pp.73-76. DOI: 10.18280/mmc_a.910207

[6]. Lakshman Narayana Vejendla , A Peda Gopi and N.Ashok Kumar,(2018)," Different techniques for hiding the text information using text steganography techniques: A survey", Ingénierie des Systèmes d'Information, Vol.23, Issue.6,pp.115-125.DOI: 10.3166/ISI.23.6.115-125

[7]. A Peda Gopi and Lakshman Narayana Vejendla (2018), "Dynamic load balancing for client server assignment in distributed system using genetic algorithm", Ingénierie des Systèmes d'Information, Vol.23, Issue.6, pp. 87-98. DOI: 10.3166/ISI.23.6.87-98

[8]. Lakshman Narayana Vejendla and Bharathi C R,(2017),"Using customized Active Resource Routing and Tenable Association using Licentious Method Algorithm for secured mobile ad hoc network Management", Advances in Modeling and Analysis B, Vol.60, Issue.1, pp.270-282. DOI: 10.18280/ama_b.600117

[9]. Lakshman Narayana Vejendla and Bharathi C R,(2017),"Identity Based Cryptography for Mobile ad hoc Networks", Journal of Theoretical and Applied Information Technology, Vol.95, Issue.5, pp.1173-1181. EID: 2-s2.0-85015373447

[10]. Lakshman Narayana Vejendla and A Peda Gopi, (2017)," Visual cryptography for gray scale images with enhanced security mechanisms", Traitement du Signal,Vol.35, No.3-4,pp.197-208. DOI: 10.3166/ts.34.197-208

[11]. A Peda Gopi and Lakshman Narayana Vejendla, (2017)," Protected strength approach for image steganography", Traitement du Signal, Vol.35, No.3-4,pp.175-181. DOI: 10.3166/TS.34.175-181

[12]. Lakshman Narayana Vejendla and A Peda Gopi, (2020)," Design and Analysis of CMOS LNA with Extended Bandwidth For RF Applications", Journal of Xi'an University of Architecture & Technology, Vol. 12, Issue. 3,pp.3759-3765. https://doi.org/10.37896/JXAT12.03/319.

[13]. Chaitanya, K., and S. Venkateswarlu,(2016),"DETECTION OF BLACKHOLE & GREYHOLE ATTACKS IN MANETs BASED ON ACKNOWLEDGEMENT BASED APPROACH." Journal of Theoretical and Applied Information Technology 89.1: 228.

[14]. Patibandla R.S.M.L., Kurra S.S., Mundukur N.B. (2012), "A Study on Scalability of Services and Privacy Issues in Cloud Computing". In: Ramanujam R., Ramaswamy S. (eds) Distributed Computing and Internet Technology. ICDCIT 2012. Lecture Notes in Computer Science, vol 7154. Springer, Berlin, Heidelberg

[15]. Patibandla R.S.M.L., Veeranjaneyulu N. (2018), "Survey on Clustering Algorithms for Unstructured Data". In: Bhateja V., Coello Coello C., Satapathy S., Pattnaik P. (eds) Intelligent Engineering Informatics. Advances in Intelligent Systems and Computing, vol 695. Springer, Singapore

[16]. Patibandla, R.S.M.L., Veeranjaneyulu, N. (2018), "Performance Analysis of Partition and Evolutionary Clustering Methods on Various Cluster Validation Criteria", Arab J Sci Eng ,Vol.43, pp.4379–4390.

[17]. R S M Lakshmi Patibandla, Santhi Sri Kurra and N.Veeranjaneyulu, (2015), "A Study on Real-Time Business Intelligence and Big Data",Information Engineering, Vol.4,pp.1-6.

[18]. K. Santhisri and P.R.S.M. Lakshmi,(2015), " Comparative Study on Various Security Algorithms in Cloud Computing", Recent Trends in Programming Languages ,Vol.2,No.1,pp.1-6.

[19]. K.Santhi Sri and PRSM Lakshmi,(2017), "DDoS Attacks, Detection Parameters and Mitigation in Cloud Environment", IJMTST,Vol.3,No.1,pp.79-82.

[20]. P.R.S.M.Lakshmi,K.Santhi Sri and Dr.N. Veeranjaneyulu,(2017), "A Study on Deployment of Web Applications Require Strong Consistency using Multiple Clouds", IJMTST,Vol.3,No.1,pp.14-17.

[21]. P.R.S.M.Lakshmi,K.Santhi Sri and M.V.Bhujanga Ra0,(2017), "Workload Management through Load Balancing Algorithm in Scalable Cloud", IJASTEMS,Vol.3,No.1,pp.239-242.

[22]. K.Santhi Sri, P.R.S.M.Lakshmi, and M.V.Bhujanga Ra0,(2017), "A Study of Security and Privacy Attacks in Cloud Computing Environment", IJASTEMS,Vol.3,No.1,pp. 235-238.

[23]. R S M Lakshmi Patibandla and N. Veeranjaneyulu, (2018), "Explanatory & Complex Analysis of Structured Data to Enrich Data in Analytical Appliance", International Journal for Modern Trends in Science and Technology, Vol. 04, Special Issue 01, pp. 147-151.

[24]. R S M Lakshmi Patibandla, Santhi Sri Kurra, Ande Prasad and N.Veeranjaneyulu, (2015), "Unstructured Data: Qualitative Analysis", J. of Computation In Biosciences And Engineering, Vol. 2,No.3,pp.1-4.

[25]. R S M Lakshmi Patibandla, Santhi Sri Kurra and H.-J. Kim,(2014), "Electronic resource management using cloud computing for libraries", International Journal of Applied Engineering Research, Vol.9,pp. 18141-18147.

[26]. Ms.R.S.M.Lakshmi Patibandla Dr.Ande Prasad and Mr.Y.R.P.Shankar,(2013), "SECURE ZONE IN CLOUD", International Journal of Advances in Computer Networks and its Security, Vol.3,No.2,pp.153-157.

[27]. Patibandla, R. S. M. Lakshmi et al., (2016), "Significance of Embedded Systems to IoT.", International Journal of Computer Science and Business Informatics, Vol.16,No.2,pp.15-23.

[28]. AnveshiniDumala and S. PallamSetty. (2020),"LANMAR routing protocol to support real-time communications in MANETs using Soft computing technique", 3rd International Conference on Data Engineering and Communication Technology (ICDECT-2019), Springer, Vol. 1079, pp. 231-243.

[29]. AnveshiniDumala and S. PallamSetty. (2019),"Investigating the Impact of Network Size on LANMAR Routing Protocol in a Multi-Hop Ad hoc Network", i-manager's Journal on Wireless Communication Networks (JWCN), Volume 7, No. 4, pp.19-26.

[30]. AnveshiniDumala and S. PallamSetty. (2019),"Performance analysis of LANMAR routing protocol in SANET and MANET", International Journal of Computer Science and Engineering (IJCSE) – Vol. 7,No. 5, pp.1237-1242.

[31]. AnveshiniDumala and S. PallamSetty. (2018), "A Comparative Study of Various Mobility Speeds of Nodes on the Performance of LANMAR in Mobile Ad hoc Network", International Journal of Computer Science and Engineering (IJCSE) – Vol. 6, No. 9, pp. 192-198.

[32]. AnveshiniDumala and S. PallamSetty. (2018),"Investigating the Impact of IEEE 802.11 Power Saving Mode on the Performance of LANMAR Routing Protocol in MANETs", International Journal of Scientific Research in Computer Science and Management Studies (IJSRCSMS) – Vol.7, No. 4.

[33]. AnveshiniDumala and S. PallamSetty. (2016),"Analyzing the steady state behavior of RIP and OSPF routing protocols in the context of link failure and link recovery in Wide Area Network", International Journal of Computer Science Organization Trends (IJCOT) – Vol. 34 No 2, pp.19-22.

[34]. AnveshiniDumala and S. PallamSetty. (2016),"Investigating the Impact of Simulation Time on Convergence Activity & Duration of EIGRP, OSPF Routing Protocols under Link Failure and Link Recovery in WAN Using OPNET Modeler", International Journal of Computer Science Trends and Technology (IJCST) – Vol. 4 No. 5, pp. 38-42.

[35]. VellalacheruvuPavani and I. Ramesh Babu (2019) ,"Three Level Cloud Storage Scheme for Providing Privacy Preserving using Edge Computing",International Journal of Advanced Science and Technology Vol. 28, No. 16, pp. 1929 – 1940.

[36]. VellalacheruvuPavani and I. Ramesh Babu,"A Novel Method to Optimize the

Computation Overhead in Cloud Computing by Using Linear Programming" ,International Journal of Research and Analytical Reviews May 2019, Volume 6, Issue 2,PP.820-830..

[37]. Anusha Papasani and Nagaraju Devarakonda,(2016),"Improvement of Aomdv Routing Protocol in Manet and Performance Analysis of Security Attacks", International Journal Of Research in Computer Science & Engineering ,Vol.6,No.5, pp.4674-4685.

[38]. Sk.Reshmi Khadherbhi,K.Suresh Babu , Big Data Search Space Reduction Based On User Perspective Using Map Reduce ,International Journal of Advanced Technology and Innovative Research Volume.07, IssueNo.18, December-2015, Pages: 3642-3647

[39]. B.V.Suresh kumar,Sk.Reshmi Khadherbhi ,BIG-IOT Framework Applications and Challenges: A Survey Volume 7, Issue VII, JULY/2018 pg.no 1257-1264

[40]. P.Sandhya Krishna,Sk.Reshmi Khadherbhi,V.Pavani, Unsupervised or Supervised Feature Finding For Study of Products Sentiment ,International Journal of Advanced Science and Technology, Vol 28 No 16 (2019).

[41]. K.Santhi Sri, Dr.Ande Prasad (2013), "A Review of Cloud Computing and Security Issues at Different Levels in Cloud Computing" , International Journal on Advanced Computer Theory and Engineering Vol. 2,pp 67-73.

[42]. K.Santhi Sri, N.Veeranjaneyulu(2018), "A Novel Key Management Using Elliptic and Diffie-Hellman for Managing users in Cloud Environment", Advances in Modelling and Analysis B,Vol.61,No.2,pp 106-112.

[43]. K.Santhi Sri, N.Veeranjaneyulu(2019), "Decentralized Key Management Using Alternating Multilinear Forms for Cloud Data Sharing with Dynamic Multiprivileged Groups", Mathematical Modelling of Engineering Problems,Vol.6,No.4,pp511-518.

[44]. S.Sasikala, P.Sudhakar, "interpolation of CFA color Images with Hybrid image denoising", 2014 Sixth International Conference on Computational Intelligence and Communication Networks, DOI 10.1109/.53 193 DOI 10.1109/CICN.2014.53, pp. 193-197.

[45]. Me. Jakeera Begum and M.Venkata Rao, (2015), "Collaborative Tagging Using CAPTCHA" International Journal of Innovative Technology And Research, Volume No.3, Issue No.5,pp,2436 – 2439.

[46]. L.Jagajeevan Rao, M. Venkata Rao, T.Vijaya Saradhi (2016), "How The Smartcard Makes the Certification Verification Easy" Journal of Theoretical and Applied Information Technology, Vol.83. No.2, pp. 180-186.

[47]. Venkata Rao Maddumala, R. Arunkumar, and S. Arivalagan (2018)"An Empirical Review on Data Feature Selection and Big Data Clustering" Asian Journal of Computer Science and Technology Vol.7 No.S1, pp. 96-100.

[48]. Singamaneni Kranthi Kumar, Pallela Dileep Kumar Reddy, Gajula Ramesh, Venkata Rao Maddumala, (2019), "Image Transformation Technique Using Steganography Methods Using LWT Technique" ,Traitement du Signalvol 36, No 3, pp. 233-237.