Implementation of Q-learning algorithm for credit card fraud detection.

Bhavna Gopchandani ,Pooja Sharma, Pooja Lanjewar, Sakshi Kodgule ,Shreya Chavan

Department of Computer Science and Engineering Jhulelal Institute of Technology, Lonara Nagpur

Abstract:

Financial fraud is an ever growing threat with way consequences within the economic business. Now-adays the use of electronic transactions are increasing because they are easy and time saving for customers. Individual cards has its unique identification information which give services to the user and will be paid according to their use. German Credit Data is one of the most famous datasets in the realm of fraud detection and it is available in two formats that is categorical attribute and numerical attribute. Increased in number of credit card transactions opens the door for thieves to steal credit card details and commit fraud. Thieves steals the account of users which includes the card account number or other information that would be necessarily available to a merchant during legitimate transactions. To overcome this situation we are using Q- learning algorithm which is an off policy reinforcement learning algorithm. It's considered off-policy because the Q- learning function learns from actions that are outside the current policy like taking random actions and therefore policy is not needed. The 'Q' in Q-learning algorithm stands for the quality that is applied on the raw and pre- processed knowledge ofdata. This paper describes the existing application related to credit card fraud detection.

Keywords:

Fraudulent transaction, Reinforcement learning algorithm, Q learning, Fraud detection, Fraud prevention.

I. Introduction:

In today's world, online shopping has become an integral part of our life .Now a days the use of electronic transaction are increasing because they are easy and time saving. As online shopping becomes the most prevailing mode of payment for both online as well as regular purchase, frauds related with it are also accelerating. A credit card is payment card issued to users to enable the cardholder to pay a merchant for goods and services based on the cardholder's promise to the card issuer to pay them for the amounts and another agreed charges. Individual cards has its unique identification information which give services to the user and will be paid according to use [6].

As increased in number of credit card transactions open the door for thieves to steal credit card details and commit fraud. Credit card begins with the theft of the physical card or with the compromise of data associated with the account including the card account number or other information that would necessarily be available to a merchant during legitimate transactions.

There are many techniques which are used for detecting the frauds during credit card transactions such as Decision Tree, Genetic Algorithm, Neural Network, Outlier Analysis and logistic regression. In this techniques work detection is done based on time analysis and pattern analysis, which had a lot of drawbacks. The time analysis is sequence of well-defined data points measured at the consistent time interval over the period of time. The pattern analysis is used to transform the given data set into particular pattern.[12]The various existing applications related to credit card fraud detection has been discussed in section II.

ISSN: 2233-7857 IJFGCN Copyright ©2020 SERSC

II. Related Work:

[1]Domingos Bruno, et.al Describes Decision Tree technique that sorts the information into branch-like fragments that manufacture an altered tree with a root hub, inward hubs, and

leaf hubs. Decision tree is easy to understand and implemented. Decision tree is a calculation that uses a tree like diagram or model of choices what's more, their potential results to foresee a ultimate conclusion this calculation employments restrictive Explanation. Decision tree is also called as choice tree which uses the graph or model to predict the final output. For inductive Decision learning these sorts of calculations are well- known and have been effectively connected to abroad scope of assignments. The general rule used in decision tree is 'If condition 1 and condition 2 but not condition 3 then outcome'. The main disadvantage of decision tree is that they are unstable, meaning that a small change in the data can leads to large change in the structure of optimal decision tree.

Carneiro E.M., et al.describes Neural system[3] based representation identification depends absolutely on the human mind working head. Neural system innovation has made a computer fit for the think. As human mind learn through past involvement and utilize its information or involvement in settling on the choice in day by day life issue the same procedure is connected with the credit card misrepresentation recognition innovation called as neural network. The Neural Network technique can work for organized, unstructured, and envisioned datasets. The greater computational burden,

proneness to over fitting, and the empirical nature of model development is the main drawbacks of neural network. Its concealed layers do fitting component changes of envisioned datasets. Thus, the Neural Network method has a distinct advantage over the Linear Regression method in cases of pictured datasets.

Daniel Garner, et al.[4]describes Genetic Algorithm which is an enhancement strategy that attempt to recreate common development forms. genetic algorithm is also called as Hereditary Algorithm. The Genetic pool of a particular population for a given issue possibly contains the arrangement, or a superior arrangement that means it will give a solution or a better solution to a given problem. This is the essential thought behind the genetic calculation. The main goal of Artificial Genetic Algorithm is to improved the solution to a problem. This improvement is completed by keeping the best mix of information factors. It streamlines the issue definition and furthermore creates a goal work that is the method for figuring out which person produces the best result. There are Three Different operations which are carried out by the genetic algorithm which are as follow:

A. Selection: It is the survival of the fittest and the inclination is constantly given to better results.

B. Mutation: It depends on attempting arbitrary mixes what's more, assessing the outcome (achievement or disappointment) of the result.

C. It is finished by joining parts of good results in the desire for making a surprisingly better result.

Aishwarya Kaneri, et al.[5] describes Outlier analysis systems which are connected to recognize false Master card utilization. Distinguishing fake charge card utilization is like recognize an anomaly. The information comprises of a few measurements, for example, client id, sum spent, and time between continuous card utilizations, etc. whenever we are dealing with large data set the outlier analysis is efficiently used. The fraud are reflected in value-based records.

Xueton Xuetong Niu, et al.[6] describes the dataset and preprocessing approach which is used the dataset contains numerical input variables which are from a PCA transformation due to confidentiality issue. For the non-numerical features of "Time" and "Amount", we normalize them by using Robust Scalar which scales the data according to the quantile range. Specifically for the supervised learning models, to tackle

the heavily unbalanced problem, random down sampling is used to avoid the bias results toward the non-fraudulent class.

KaithekuzhicalLeenaKurien1 and et al. describes the two techniques Logistic Regression and Random Forest classifier [7] are used. Logistic Regression classifier was selected as it is used for predictive analysis and it is used to describe data independent variables. Random Forest classifier is an ensemble of random decision trees. Random Forest achieved better results than decision trees. Its default hyper parameter setting produces good prediction results. The main trees in forest the algorithm is slow.

Nator junior carvahllo da costa and et al. describes A naïve baye's classifier [8] which is basic probabilistic classifier dependent on the baye's hypothesis. Bayesian system can process the restrictive likelihood of a hub dependent on given qualities allotted to different hubs. There are a few points of interest of Bayesian Network, for example, the capacity to deal with fragmented information sources, the learning of causal relationship, etc. The explanations for the making of such model is the capacity of kid hubs to work autonomously without hindering other kid hubs and especially impact the likelihood of root hub.

Hamzath Ali Shukur, et al. describes two techniques Information mining and machine learning [9] which are famous techniques to study and battle the Visa misrepresentation cases. There are countless investigations that misused the quality of information mining and AI to avert the Visa false exercises. In view of Self-Organizing Map and Neural Network, the investigation of got Receiver Operating Curve (ROC) over 95.00% of extortion cases without false cautions rate. There are Different modern procedures dependent on Machine learning, Artificial Intelligence, and this has been developed is as yet advancing to recognize fraudulent in credit card.

III. Experimental setup:

In this process a system and method for detecting the real-time testing of compromised financial transaction device such as credit cards and debit cards has been discussed. The system uses Q learning method and the implementation of specialized models to enhance accuracy of credit card fraud detection. The proposed system mainly focused on credit card transactions and their

use for the detection of fraudulent activities. A dataset of past transaction is required for

computer. The dataset formally contains two factors in credit card fraud detection which are time and amount of transactions.

The features are extracted from database which consist of normal transaction and fraudulent transaction by defining the class label as 0 and 1.By using this two classes of dataset we can easily identify a fraud in credit card transaction. Then by applying pattern analysis which are mainly concentrating on behavioral pattern of credit card transaction and identify whether the transaction belongs to fraudulent or normal transactions. The flow diagram indicates a user profile that has been build based on its past transactions. Whenever a new entry of same user will come then our module will check the behavior of new transaction by applying pattern analysis and compare it with the database to recognize whether the transaction is fraudulent transaction or normal one.

The standard database comprises of 28880 transactions. Features like timestamp, volume, total amount, class labels are extracted from the database to differentiate between the fraudulent and normal transactions. Features are extracted by reading and processing the information for evaluating the results of credit card. For fraud transactions the class label is 1 and for normal transactions the class label is 0.

The classification of data is done for evaluating fraudulent and normal transaction. In this proposed system SVM (Support Vector Machine) is used for the classification. SVM are supervised learning with

associated learning algorithms that analyze data used for classification. In addition to performing linear classification, SVMs can efficiently perform a non-linear classification. Because, support vector machines employing the kernel trick that do not scale well to large numbers of training samples or large numbers of features in the input space, several approximations to the RBF kernel have been introduced. RBF the radial basis function kernel, is a popular kernel function used in various kernelized learning algorithms used in support vector machine classification.

In this proposed system, classification report is used for calculating the values of f1 score(score of fraudulent activities).

f1(score) = P + 2PR/(P + R)where, P= number of transaction which are correctly match with database of past transaction.

R= number of transaction which are not match with the database of past transaction.

IV. Results:

In this proposed system Q-learning algorithm which is machine learning algorithm used for improving the accuracy of credit card fraud detection.



Fig (A): Graph of analyzing fraud and non fraud transaction

The above figure (A) shows the fraud and non fraud transaction of single user in the form of graph which evaluate the result from the database of past transaction. As discussed above class label 1 defines fraudulent transaction and class label 0 defines non fraudulent transaction.



Fig (B): Time of transaction Vs Amount by class

The above figure (B) shows the amount of transaction with respect to time stamp for normal transaction and fraud transaction.

Graphical representation of normal transaction with respect to time and amount and the fraud transaction with respect to time and amount shown in figure (B).



Fig (C): Correlation Matrix with respect to amount and time

ISSN: 2233-7857 IJFGCN Copyright ©2020 SERSC The above figure (C) is a the correlation matrix which shows that none of the volumes from V1 to V28 PCA(Principle Component analysis) component have any correlation to each other however if class has some positive correlation and negative correlation with respect to V component but has no correlation with respect to time and amount.

Support Ve Accuracy S 0.05281866 Classifica	ector Score 514829 Ation	Machine: : 921414 Report :	51162		
	1	precision	recall	f1-score	support
	0 1	1.00 0.00	0.05	0.10 0.01	53860 155
micro a macro a weighted a Fig (D): Classific	avg avg avg cation Re	0.05 0.50 1.00	0.05 0.52 0.05	0.05 0.05 0.10	54015 54015 54015

The above figure (D) shows the classification report which was calculated the values of f1-score and support based on the values of precision and recall.

Conclusion:

This paper discusses the Credit fraud detection using Q-learning Algorithm. The result section depicts the various results showing the credit fraud detection graphically.

References:

- 1. XuetongNiu, Li Wang, Xulei Yang "A Comparison Study of Credit Card Fraud Detection-Supervised versus Unsupervised"(April 2019).
- 2. XuetongNiu, Li Wang, Xulei Yang transarXiv:1904.10604v1 [cs.LG] 24 (Apr 2019).
- 3. KaithekuzhicalLeenaKurien1 & Dr. Ajeet Chikkamannur "Detection and prediction of credit card using machine learning"(march 2019)
- 4. Hamzah Ali Shukur, SeferKurnaz" Credit Card Fraud Detection using Machine Learning Methodology "(March 2019)
- 5. Nator junior carvahllo da costa, N.M fonsecaferrira ,Aratasaraiva " credit card fraud detection by Naïve bayes" (Dec 2018).
- Domingos Bruno, Gustavo Rodrigues W., Nator Junior C. C Arata A. Saraiva1, Jos´e V. Moura Sousa1, Nuno Fonseca Ferreira "Learning model for fraud detection in credit cards "(Dec 2018).
- 7. AishwaryaKaneri,Anugrah S, Isha Bharti, SamruddhiJadhav, MitaliKadu " Fraud detection in online card transaction."(march 2018).
- 8. Navanshu Khare and Saad Yunus Sait"Credit Card Fraud Detection Using Machine Learning Models and Collating Machine Learning Models(2018).

ISSN: 2233-7857 IJFGCN Copyright ©2020 SERSC

- 9. Daniel Garner ,Ritika Wadhwa "Genetic algorithms for credit card fraud detection" ,IEEE Transactions May 2017.
- 10. Carneiro E.M., Dias L.A.V., Da Cunha A.M., MialaretL.F.S.,"Cluster analysis and artificial neural networks" (march 2017).
- 11. https://www.incompleteideas.net/book/ RLbook2018trimmed.pdf?source=post_page
- 12. https://budgeting.thenest.com/characteri stics-credit-cards-30499.html