International Journal of Research in Engineering, Science and Management
Volume-3, Issue-7, July-2020
journals.resaim.com/ijresm | ISSN (Online): 2581-5792

Credit Card Fraud Detection Using Bayesian Belief Network

M. Deekshith Kumar[1*], Sowmya[2], Abdul Mubarak[3], M. S. Dhanush[4]
[1,3,4]*Student, Dept. of Information Science and Engineering, Srinivas Institute of Technology, Valachil, India*
[2]*Assistant Professor, Dept. of Information Science and Engg., Srinivas Institute of Technology, Valachil, India*
*Corresponding author: kumardk8583@gmail.com

***Abstract*: The number of Credit card fraud cases are increased by day by day in online payment system. Therefore, it's essential to have a fraud detection in transaction system, which is implemented with the help of decision making algorithm. In the proposed system, we have applied two ML techniques suited for reasoning under indefiniteness: Artificial Neural Network(ANN) and Bayesian Belief Network. If the transaction is a fraudulent, it determined as examine the precious transaction and compare with a new current transaction. If its feature of the previous transaction and the current transaction vary considerably then the new current transaction may be a fraudulent or genuine transaction. These two machine-learning techniques approaches for the proper reason under indefiniteness. Bayesian network is also known as belief network and it is a types of artificial intelligence program that uses a variety of methods, and used for pattern identification and classification of data.**

***Keywords*: Credit cards, Fraud detection, Bayesian Networks, Naïve Bayes Classifier.**

1. Introduction

Credit card fraudulent scheme is an unauthorized account activity that is done through a credit card featuring system, with an intension of obtaining items without actually paying for it. Fraud can be defined as an illegal usage on any system. Technology is improved in now a days and this improvement in technology has significant effects on various fields including financial transaction. Now a day's credit card is the most widely used transaction method; this favour of a credit card is mainly due to the ease of transaction, the increasing popularity of network banking and mobile banking. Here technology will be a two sided coin which has a great improvement in the type of fraud or genuine that can occur.

In the paper we have discussed the problem of detecting the behaviour in a credit card transaction system if its fraudulent or not. Here we focus for the credit card transaction using Bayesian Belief Network (BBN). The system is to provide some computational learner with a set of training data consisting of some feature values that are built in to the system in which we want do the detection as a fraud and genuine. In this paper a research is done of using Bayesian Belief Network for the credit card fraud detection. We are focus on credit card fraud detection, but most of the similar properties also assign to other real problems such are cellular phone fraud, calling or text messages on respective credit card and computer network. Credit card fraud detection is defined as unauthorized card account activity for a person whom the account is not wilful. The person is using a credit card has not at all having a connection with the cardholder and does not intend to make the repayment for the purchase has they done.

2. Related Work

Many researchers have studied about Naïve Bayes Classifier and Bayesian Belief Network. In 2002 Sam Maes, Karl Tuv are proposed a system of detecting in a credit card fraudulent behaviour in a credit card transaction system. Here they focus on a two machine learning techniques to the credit card fraud detection problems are Naïve Bayes Classifier and Bayesian Belief Network. The process of learning this model is supposed to be able to correctly classify a transaction it has never known as before as fraudulent and not fraudulent. The biggest problem is assigned to a fraud detection is the lack of both literature providing experimental result and of a real world data values for academic researches to perform experiments.

Bayesian network represents the dependence between the variables and gives a dense details of the joint probability distribution. At every iteration of the algorithm, they represent the network with certain pattern taken to a training set. The features of the pattern are represented to the different Bayesian network in an input layer and produce an output layer for a pattern. Neural networks [1] are an easy, fast and reliable techniques to obtain better result in different area. In this operation it found that the great difficulty in applying of neural network resides in the choice of good set of pre-processing operations and better exchange in between the different parameter that has to be selected. Credit card fraud detection is the process of identifying the selected data transaction are to be fraudulent detected in two types of genuine and fraud transaction. Credit card fraud detection is based on analysis of a card's spending behaviour. Transaction of features are extracted and transform a data from the data file in raw selection while giving a data it to train model set [2], genetic algorithm, artificial neural network, support vector machine, decision tree algorithm will applied for the model system. Bayesian belief network and the neural network are evaluated in the trained data

on credit card fraud detection system, they have illustrated an auto encoder of neural network and the logistic regression for fraud detection.

Artificial neural network(ANN) is one of the most robust classifier to search out a hidden patterns among the different attribute values. Artificial network is works same as human brain. ANN consists of a sorted layer throughout that initial layer is input data layer and last layer is output data layer. It should have included in form of hidden layer or no hidden layer. If Neural network embrace quiet one hidden layer, then it's deep learning in each layer has completely different neurons, and each overgrowth cell is connected with connected edges. Output of every overgrowth cell may even be a performance of its respective unit.

3. System Implementation

The methods are used in the proposed system are discussed in this section. Proposed system uses Bayesian Networks and Naïve Bayes classifier for classification of credit card fraud detection where it will train the data by using the collected data set value and when the user proves test data it will take the decision and shows the results.

In the training phase, a classifier is generated for fraud and genuine transactions are collected in feature. These are transmitted to the feature extractor, which extracts featured data value through the predefined class variable, time and amount will depends on class. In the testing phase, the classifier determines whether the transaction is a fraud or genuine. When a data request occurs were the file containing all the transaction data are to be imported and detect a feature of extraction, which extracts a feature is for fraud and genuine transactions. Those features values are used as an input data to be classifier. The transaction of specific classifier determines a fraud and genuine transactions in the imported file which contained all the transaction then it result as the transaction are fraud or not. The below figure 1 shows the block diagram of proposed system.

A. Bayesian Networks

Naïve Bayesian Classifier is sets for the credit card transaction classification in a research of Bayesian Network is because of its simplicity. The Naïve Bayesian Classifier is one of the forms of Bayesian Network where the conditional liberty of attributes is supported. Figure 1 shows the Block diagram for the credit card fraud detection system using Bayesian network.

B. Naive Bayes Classifier

Bayes classifier is in light of Bayes hypothesis with an impact of freedom between the predicted variable. Naïve Bayes classifier accepts that the closeness of a specific feature in a class is inconsequential to the closeness of some other features. Naïve Bayesian model is difficult to construct but the feature is valuable for the substantial datasets, a classifier can used to beat even deeply difficulty classification strategies. Here, P(c|x) is the posterior probability of class predictor. P(c) is prior probability of class, P(x|c) is the probability, the probability of predictor will known class. P(x) is the predictor for a prior probability.

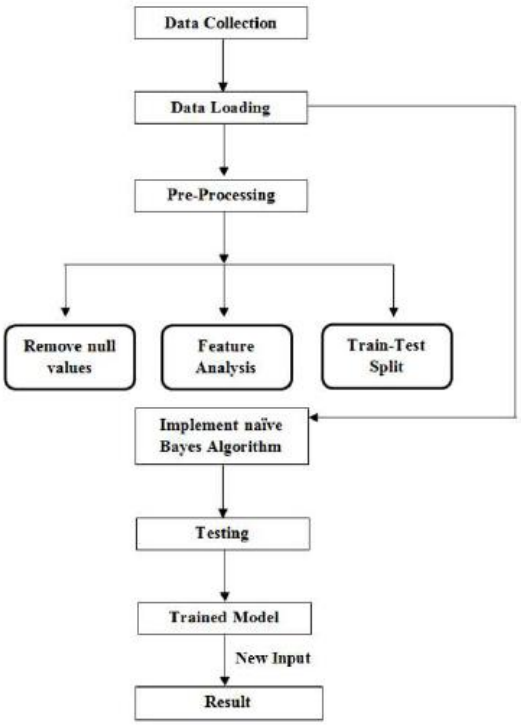

Fig. 1. Block diagram for credit card fraud detection using Bayesian network

- Naïve Bayes classifier applied to learning tasks in each instance x is described by conjunction of attributes values and the target function as f(x) can taken on any data value from other finite set V.
- A set of training data samples of the target function is provided and new instance is presents and also will described by a tuple of attributes values(a1,a2,…,am).
- The learner will ask to predict the target value, or classification for the new instance.

Bayesian approach is to classify the new data sample is assigned to the most probable target value, VMAP and given the attribute values(a1,a2,..,am) that describe the instance are,

$$v_{MAP} = \underset{v_j \in V}{\operatorname{argmax}} P(v_j | a_1, a_2 \ldots a_n)$$

Using Bayes theorem to rewrite an expression as

$$v_{MAP} = \underset{v_j \in V}{\operatorname{argmax}} \frac{P(a_1, a_2 \ldots a_n | v_j) P(v_j)}{P(a_1, a_2 \ldots a_n)}$$

$$= \underset{v_j \in V}{\operatorname{argmax}} P(a_1, a_2 \ldots a_n | v_j) P(v_j) \qquad \text{equ (1)}$$

The naïve Bayes classifier is depends on the assumption of an attribute values are conditionally independent to a given target values. Which means that the assumption is that given the target value of data value, the probability of observing the

conjunction (a1, a2,..am) is just the product of the probability for the individual attributes:

$$P(a_1, a_2 \ldots a_n|v_j) = \prod_i P(a_i|v_j)$$

Substituting this into Equation (1),
Naïve Bayes Classifier:

$$V_{NB} = \underset{v_j \in V}{\text{argmax}}\, P(v_j) \prod_i P(a_i|v_j) \qquad \text{equ (2)}$$

Where, VNB denotes the target value of output by the naïve Bayes classifier.

C. *Fraud detection*

Using Bayesian Network will require to have real genuine and fraud transaction in this system. Transaction is determined as genuine transaction if there has no complaint from the credit card holder about transaction during the defined period (for example, 3 or 4 months). If the detection predicted that the transaction is fraud, then it is determined as fraud or else its genuine transaction. Construction of static data is made separately for each registered a credit card in a bank system. Structure can made using all the genuine transaction in an credit card and the fraud transaction is been assigned for different cards. To prevent a mistakes for building a data values ratio of genuine transactions per fraud can be set by bank person. A transaction is realized as educts to fraud if the probability of transaction is genuine for the card is less than the defined value and if probability of transaction is fraud is greater than maximum fraud probability. Maximum value of legal probability and large number of fraud probability are settled by the bank member.

D. *Data Analysis*

The data values have been selected and used to holds the records of cardholders who made transactions using their credit card. This dataset values holds the record of transactions that were made within two days and total transactions are 284,807 transactions from which 358 transactions were found as fraudulent which makes the dataset highly variance, more predicted as the positive class i.e., fraud transactions are 0.2% out of total transactions. The dataset is in CSV format i.e., in a format where the data values are separated by commas of different attributes.

As PCA transformation of input values has been done in the dataset which makes the dataset contain only numerical data input values. Unfortunately, the source of dataset does not provide the more information, original features and information due to confidentiality issues. Principal components that were obtained by PCA transformations are nothing but the numeric data values under attributes V1 to V28 and the only feature that were not transformed using the transformation by Principal Component Analysis are the attributes or features ‘TIME’ and ‘AMOUNT’ in table.

‘TIME’ attribute or feature holds the data that denotes the elapsed a time in the transaction of the dataset and each transaction. And the attribute of ‘AMOUNT’ hold the data which represents nothing but the transaction amount and this feature can find its use for cost-sensitive and example-dependent machine learning. Finally, the attribute or feature of ‘CLASS’ is the response variable and it takes values ‘1’ in the case of fraudulent transaction as positive result in credit card and value ‘0’ in the case of the genuine transaction.

4. Results and Discussion

The testing has two sets of transaction is been generated. Each set consists of transaction detects for a training and testing. The first set is considered to estimate the output of naïve bayes classification using several approaches for the class probability calculation. Then the training data consists of transaction which variable in time transaction is corresponds to several amounts. There has many transactions from the set correspond to be a fraud and genuine with their respective time.

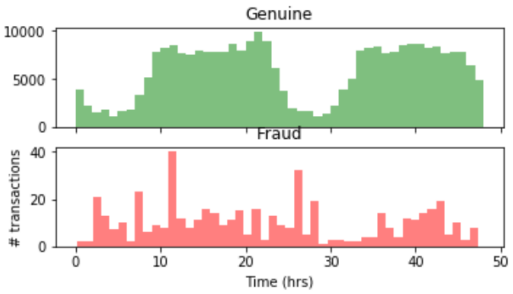

Fig. 2. Naive Bayes classification result using different approaches for the respective transaction

The second set of transaction is sufficient amount of data taken for a transaction in credit card will deals to a genuine transaction. If a transaction is varying in transferring an amount which causes a fraud transaction with their corresponding Time and Amount in classifier.

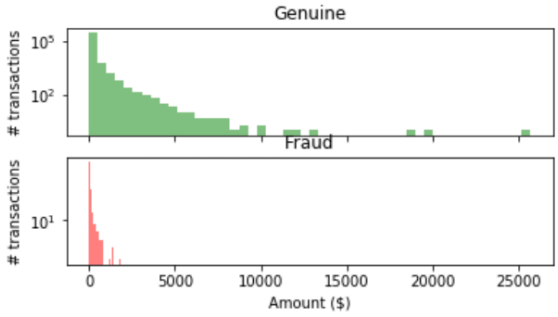

Fig. 3. Naive Bayes classification result using different transaction with the corresponding amount

Transaction can handle the value in table value in which the class attributes will give an 0 and 1 values. Figure 2 and 3 shows the Naive Bayes classification result using different transaction with corresponding amount. If the value in transaction is 0 which assign as Genuine Class and otherwise the transaction will goes to 1 class is shows as Fraud transaction in class. System is implemented according to the proposed method and

following outputs is being obtained as results.

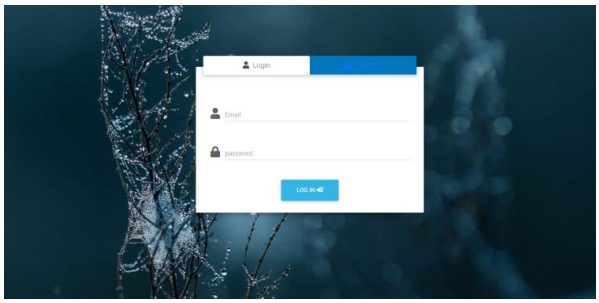

Fig. 4. Main page with login form

Figure 4 shows main page with login form of the proposed system where user has to enter the login details.

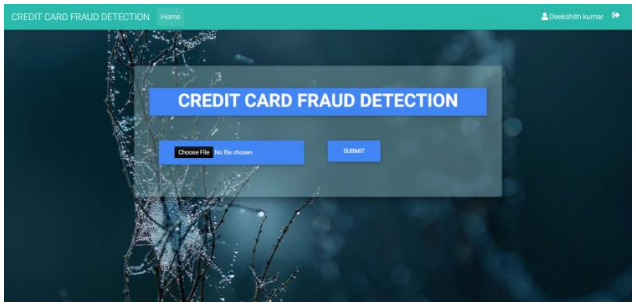

Fig. 5. Index page to predict sample data

Figure 5 shows the index page of the proposed system where the test data is uploaded.

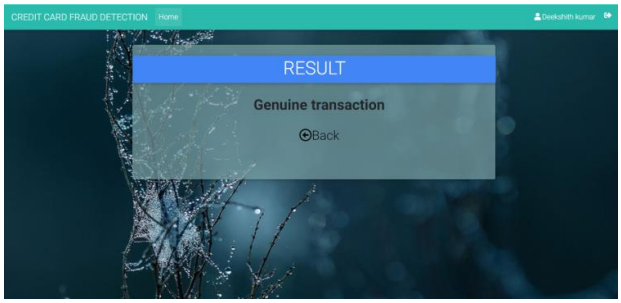

Fig. 6. Predicted result of genuine transaction

The above figure 6 show the predicted result of genuine transaction and it belongs to class 0 in the classification.

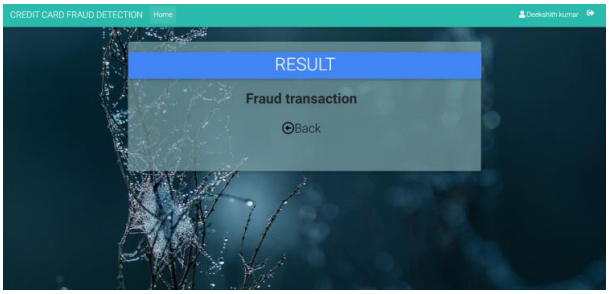

Fig. 7. Predicted result of fraud transaction

The above figure 7 show the predicted result of fraud transaction and it belongs to class 1 in the classification.

5. Conclusion

Testing of result will approaches for the class probability prediction in a Naïve Bayesian Classifier to prove that the developed input data represents a method is better than the other assigned approach values. The Naïve Bayesian Classifier is depends on the input data representation method is exact what is tested by calculated probabilities for transaction with attribute values is not been observed in training phase.

The transaction set are closer to a real area history of transaction from the credit cards. The output of result will show a Bayesian Network is more accurate than the Naïve Bayesian Classifier. This is disturbed with using the fact of conditional dependence between the attributes in Bayesian Network, but it requires more difficult to calculation and as training process. The transaction of data value available in dataset which is trained with their results as fraud or genuine transaction which is predicted by a testing data value for individual transaction. The developed input data representation method manage a implementation of Bayesian Network like one of the base system in a real credit card fraud detection and predicting a system.

References

[1] Lev Mukhanov, “Using Bayesian Belief Networks for credit card fraud detection”, International Conference on Artificial Intelligence and Applications, February 2008.

[2] Sai Kiran and Jyoti Guru Rishabh Kumar, Naveen Kumar, Deepak Katariya, Maheshwar Sharma, “Credit card fraud detection using Naïve Bayes model based and KNN classifier”, International Journal of Advance Research, Ideas and Innovations in Technology, Volume 4, Issue 3, May 2018.