International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 07 Issue: 04 | Apr 2020 www.irjet.net p-ISSN: 2395-0072

CREDIT CARD FRAUD DETECTION

Nikhita K Raj[1], Dr.S.Kuzhalvai Mozhi[2]

[1]Dept. Of ISE, The National Institute of Engineering, Mysore
[2]Professor, Dept. of ISE, The National Institute of Engineering, Mysore

---

Abstract – In this paper a credit card fraud detection prototype is proposed. Credit card fraud has been a major issue in recent years. The number of frauds is increasing day by day and huge amount of money is lost due to false transactions. Machine learning algorithms are used to detect fraud in this prototype. Initially the model is trained with less data and once the model is ready, it is used to predict fraud for dynamic data. SMS and mail notification using SMTP protocol is used to alert the user regarding fraud.

The algorithm with highest accuracy predicts the fraud and voting method ensures good accuracy of prediction.

***Key Words*: Credit card fraud, Machine learning algorithms, Geographic location, Transactions, prediction**

1. INTRODUCTION

Machine learning is opted because there is a massive amount of transaction using credit cards and hence we have to narrow down the huge data by filtering and analyzing it. Frauds are less than one percent of the transactions and hence difficult to detect. Machine learning algorithm can be iteratively applied for a set of data and analyze the pattern for different situation.

The main criteria on which fraud is detected are geographic location, amount of transaction and frequency of transaction.

When there will be a sudden fluctuation in the amount of transaction or changing location, the user will be sent an SMS and email which alerts the user.

In this prototype, six algorithms have been used to detect fraud. And the algorithm with highest accuracy predicts the fraud dynamically.

1.1 Methodologies

Naives bayes is a group of algorithm based on bayes theorem where every pair of feature classified is not dependent on each other. The basic assumption is that each feature makes an independent and equal contribution to outcome. This assumption is not true in real world situation and hence probability of an event is found.

Decision tree is a tree like model of decisions and a way to display an algorithm that only contains conditional control statements. It breaks down a dataset into smaller subset and associated decision tree is developed. The result is the tree with decision nodes.

Random forest classifier is a supervised learning model that uses labeled data to learn how to classify unlabeled data. The output will be the mode of classes (Classification) or mean prediction(regression) of individual trees.

Ridge is a normalized regressor type which is linear in form. This algorithm will be used when an equation does not have an unique solution. It is applicable when more than two co-efficient are present. It is used to reduce standard errors in the system.

Support vector regression is proven to be an effective tool in real time estimation. It is used for working with continuous values. They perform regression while maintaining all important features. In this paper those features are frequency of transaction, location. Based on these criteria it is predicted as fraud or not.

Linear regression is used to depict the relationship between the scalar and explanatory vatriables. Linear predictor functions are used to model the relation. The features that are not known are estimated from the data, mostly mean is used for that purpose.

1.2 Data attributes

The dataset considered consists of nine attributes. The fraud is mainly detected on the basis of geographic location, amount of transaction and the frequency of transactions.

Consider a user purchases an item in one place and it is impossible to make a purchase at another country within a short period of time. Such a purchase would be predicted as fraud. Similarly there will be a pattern for amount of purchase of a user, and if it suddenly fluctuates it will be predicted as fraud and user will be sent an alert message.

Table-1 consists of the attributes used in fraud detection. After collecting the data, main step will data filtering because the quality of data determines how well a machine leaning algorithm will perform.

Table -1: Attributes

Sl no	Attribute	Description
1.	Name	Name of the credit card holder
2	Acc_num	Account number
3.	Date_time	Date and time of transaction
4.	Location	Location of transaction
5.	Trx_per_day	Number of transactions per day
6.	Amt	Amount of transaction
7.	Result	Fraud or not
8.	Email and phone_num	To send notification about the fraud

2. IMPLEMENTATION

The detection of credit card fraud using machine learning is one of the reliable approaches. Pandas library of python is imported to handle data manipulation because of its powerful and expressive data structure. Sci-kit learn library consists of all the machine learning algorithms considered. Matplotlib is a plotting library for python programming language and hence imported. Label encoder is used to convert strings to tokens.

Then the path of the excel sheet is provided to import the dataset. Data. Head ( ) function is used to consider the specific dataset and read them. Then the dataset is checked for null values and specific actions are taken to replace the values. The date and time attribute is converted to string.

Fit_transform function is used to predict the fraud based on Location and the result. Train_test_split is used to split the dataset into training and testing data. Test_size = 0.2 indicates 20% of the data is considered as testing data and remaining 80% of the data is taken as training dataset.

Numpy library is used to pass the data as array. It is a library for python language which supports large arrays and also consists of mathematical functions to operate on these arrays.
X variable consists of all the eight attribute columns except the result and the result attribute will be stored in variable Y.

```
clf = model
clf.fit (X_train, Y_train)
Accuracy = clf.score (X_test, Y_test)
```

Clf is used to store trained model values and further used to predict based on previously stored values. Pyqt package is used to build the User interface applications. Hence accuracy is calculated by all the six algorithms and will be displayed on the User interface. The algorithm with highest accuracy is considered for prediction.

3. EVALUATION

All the six algorithms are applied to the data and prediction about the fraud is done. Once the prediction of all the algorithms are done adaboost algorithm is applied to improve accuracy .The prediction is done on the basis of geographic location and amount. The algorithm with highest accuracy is dynamically utilized to predict the fraud.

The dataset is divided into training dataset and the testing dataset using train_test_split. The training data accounts for 80% and remaining 20% is considered as testing dataset. Accuracy is used to predict the result. Accuracy of all the six algorithms is displayed in the User interface and algorithm with highest accuracy is considered dynamically.

The main advantage of this prediction is it is applicable to dynamic and also real time data. User interface can be used to display about the fraud. When account details are given on the user interface all the transactions of the user and the details about fraud or not are displayed.

4. CONCLUSION

In this paper the main effort has been to detect the fraud based on the location and the frequency of transaction. All the algorithms are applied simultaneously and the algorithm with highest accuracy is used to estimate the fraud. SMS and email are used to alert the user regarding the fraud. When

the user is notified about the fraud, further actions can be taken like blocking of the card etc. Application of Adaboost algorithm and the voting technique helps to improve accuracy.

REFERENCES

[1] john o.awoyemi, Adebayo o. adetunmbi, Samuel A. Oluwadare "Credit card Fraud detection using machine learning techniques: a comparative analysis", 2017 International Conference on Computing Networking and informatics(ICCNI), **DOI:** 10.1109/ICCNI.2017.8123782

[2] Anuruddha Thennakoon:Chee bhagyani,Sasitha premadasa;"Real time credit card fraud detection usingmachinelearning" **DOI:** 10.1109/CONFLUENCE.2019.8776942

[3] Kuldeep randhawa;Chi kiong loo;manjeevan seera;Chee peng lim and asoke K.Nandi;"Credit card fraud detection using adaboost and majority voting", vol 6,2018

[4] Pawan kumar,Fahad iqbal ; "Credit card fraud identification using machine learning Approaches"