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

Online Credit Card Fraud Detection System

Miss. Jadhav Shradha M. [1], Miss. Chalwa Prabhavati V [2], Miss. Bakshi Asmita S.[3]

Prof. Prof. Santosh waghmode (Assistant Professor) [4]

[1,2,3]Student, Computer Engineering, JSPM's Imperial College of Engineering and Research, Maharashtra, India
[4]Assistant Professor, Dept. of computer Engineering, JSPM's Imperial College of Engineering and Research, Maharashtra, India

---

Abstract - *Online transactions are rapidly increasing and it will mostly be done by the credit cards. Loss of physical credit cards or loss of credit card information will make a person pay much more amount. The hackers are there to do fraud with people. So, there was a need to detect fraudulent transaction and to secure online credit card transaction. For analysis of this problem and to overcome this fraud related to the credit cards, we have developed a new system namely "Online credit card fraud detection and prevention system" using machine learning. We have experimented with multiple algorithms to check the accuracy of the system. This system chooses random forest algorithm to check whether the transaction is normal or fraud. Apart from random forest algorithm there are also other algorithms like unbalanced classification, logistic regression, multilayer perception, Naive Bayes. This system uses multiple decision trees at the backend of the random forest algorithm.*

Keywords: - Random forest, decision tree, logistic regression, classification, credit card fraud.

1.INTRODUCTION

In present scenario, many companies all over the world are growing rapidly. The companies are trying to provide best services to the customers. For that companies are processing huge amount of data on daily basis. This data also contains personal and financial details of the customers. So, the companies need to store the data to process it and the security of data is important. If this data is not secured then it can be used by other companies or in worst case scenario it can be stolen. In some cases, the financial information is stolen which can be used for fraud transaction and which can harm respective individuals.

Today, online shopping has become a popular trend of daily purchases. The criminals are doing malicious activities such as Trojan and pseudo based-station. Increase in fraud events has become a serious issue now-a-days when cardholder's information is stolen by the criminals.

Credit card fraud detection is a relevant problem that draws the attention of machine learning and computational intelligence communities, where the large number of automation solutions has been proposed.

All around the world data is available, from small to big organizations are loading the information that has high volume, variety, speed and worth. This information comes from various sources like social media followers, likes and comments, user's purchase behaviors. All this information is used for analysis and visualization of the hidden data patterns. For frauds, the credit card is an easy and friendly target because without any risk, a significant amount of money is obtained within a short period.

In this paper we have used online banking transaction repository data set, we have trained data set for checking or analyzing and classify if the transaction is fraud transaction or normal transaction.

2. RELATED WORK

Because of credit card fraud events there is huge financial losses. Trojan & phishing technologies used by the criminals to hack the information of other people's credit card. Therefore, the fraud detection method is important. Because of fraud detection method we can identify a fraud in time when criminal uses fake card to consumer. In this paper two kinds of random forest algorithm are used to train the behavior feature of normal & fraud transactions.

The goal of data analytics is to define hidden patterns &use them support informed conclusions in a variety of situations. Credit card fraud has highly mismatch publicly available datasets. We identify the most important variables that may lead to higher accuracy in credit card fake transaction detection.

Credit card fraud refers to the loss of sensitive credit card information or physical loss of credit card. For the credit card detection purpose there are many machine learning algorithms. In this paper they use several algorithms for classifying the transaction is fraud or the genuine one. In this research Credit card fraud detection dataset was used. For the purpose of Oversampling "Synthetic Minority Over Sampling Technique (SMOTE)" was used. Dataset was split into two parts, training data & test data. The algorithms used in the research were Logistic Regression, Random Forest, Naïve Bayes & Multilayer Perceptron. Result show that each algorithm can be used for credit card fraud detection with high accuracy.

The number of transactions by credit cards are increasing rapidly with the rapid development of electronic commerce. The most popular transaction mode is online shopping, cases of transaction fraud is also increasing. In this research, we propose a novel fraud detection method that comprises of four stages. First apply the cardholder's historical transaction data to divide all cardholders into different groups. Next, we abstract a collection of specific behavioral patterns for each cardholder based on the combined transactions & the cardholder's historical transactions. Then they train a set of classifiers for each group on the base of all behavioral patterns. Finally, to detect online fraud they use the classifier set.

In credit card transactions detecting fraud is maybe one of the best testbeds for computational intelligence algorithm. This problem involves: Concept drift, class imbalance, verification latency. This lack of practicality concerns two main aspects: 1) the way & timing with which supervised information is provided.2) the measures used to asses fraud-detection performance. This paper has three major contributions. First, they propose, with the help of their industrial partner, this system also demonstrate the most appropriate performance measures to be used for fraud-detection purposes. Second, we design an assess a novel learning strategy that effectively addresses class imbalance, concept drift & verification latency. Third, in this system demonstrate the impact of class unbalance & concept drift in a real-world data stream.

When making a credit card fraud detection model, it is very important to abstract the right features from transactional data. This is usually done by combining the transactions in order to observe the outgoings behavioral patterns of the customers. In this system propose to create a new set of features based on studying the periodic behavior of the time of a transaction using the von Mises distribution. By including the proposed periodic features into the methods, the result shows an average rise in savings of 13%.

3. PROPOSED METHOD

After referring some papers on Credit card fraud detection, we observed that the previous systems for credit card fraud detection that are detects only frauds that are happens during transaction. but we are going to develop such system that can detect frauds as well as prevent the frauds during transaction.

our future work is to focus on preventing the frauds which are rapidly increased now a days. in existing system there are number of algorithms are used but after comparing all those algorithms we find the Algorithm which has the best accuracy i.e. Random Forest Algorithm.

In our system we are using the Random forest algorithm which is a classification algorithm consisting of many decision trees.

The random forest algorithm takes the decision depends on the transaction.

The proposed system will work on following phases:

- While doing transaction user will enter all his card details and credentials then on the basis of trained random forest algorithm the system will classify if the transaction is fraud or normal.
- But someone want to hack or access money and make transaction illegally then they will not able to do transactions when system has detected the transaction as fraud and then the action will be taken on it.
- In our system we working on large data sets and some parameters depends on that parameter our proposed system can detect the transactions are fraud or normal.

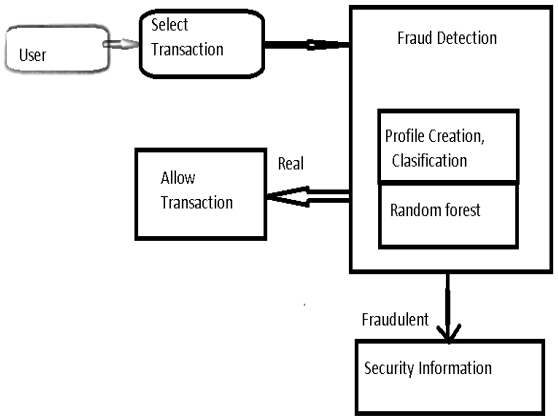


Fig.1 System diagram

The Most Important thing of machine learning is classification. we want to know what class. And in this proposed system following classification Algorithms are used to predict the output.

3.1 The Random Forest Algorithm

Random forest classifier, it is the most powerful and popularly used algorithm in machine learning. It consists number of individual decision trees and make decisions. It has simplicity

and diversity and because of that it can performs both classification and regression tasks.

• How Random Forest Algorithm Works

Random forest Algorithm is a supervised learning algorithm. As the name suggest "forest" it is a

group of decision trees.

• Steps of random forest algorithm:

1. Pick N random records from the dataset.

2. Build a decision tree based on these n records.

3. Choose the number of trees you want in your

algorithm and repeat steps 1 and 2.

4. In case of a regression problem, for a new record,

Each and every tree in the forest forecasts a value for Y (output).

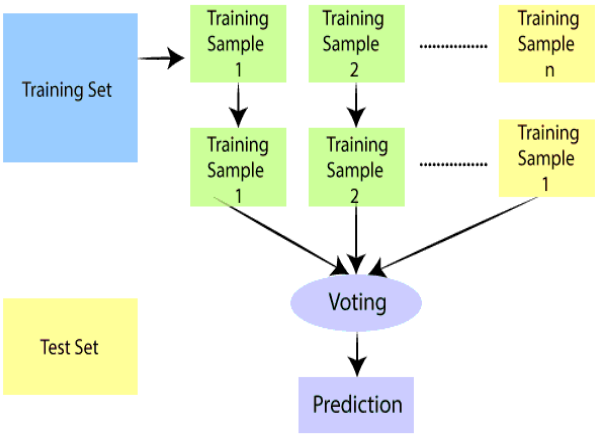


Fig.2 Random forest Algorithm example

3.2 Decision Tree Algorithm

A decision tree is a decision support tool. It is a non-parametric supervised learning method used for classification and regression. it works on training data and produces a decision tree. Decision trees are very excellent tool for helping to choose between number of courses of action. The main aim od decision tree is to create training model for predict class.

Decision Tree Algorithm Steps

1. set the best attribute of the dataset at the root of the tree.

2. divide the training set into subsets. Subsets should be made in the degree of a way that each subset contains data with the same value for an attribute.

3. Repeat steps 1 and 2 on each subset until you got leaf nodes in all the branches of the tree.

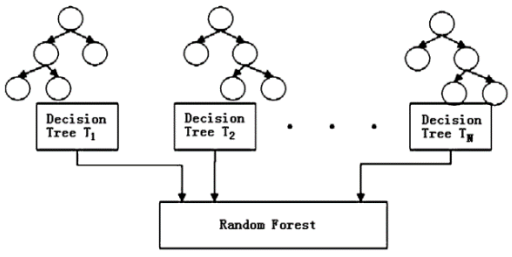


Fig.3 Decision Tree Example

- **Disadvantage of Decision Tree Algorithm**

1. In decision tree Algorithm high chances of overfitting

2. Training time is relatively expensive

3. There may be difficult to handle non-numeric data.

3.3 Logistic Regression

Logistic Regression is a traditional statistics technique that is very important widely used as a machine learning tool. Logistic regression is a linear algorithm. Logistic regression is a suitable regression analysis to conduct when the dependent variable is binary.

Like all regression analysis, the logistic regression is a predictive analysis. Logistic regression is used to relate data and to explain the relationship between one dependent binary variable and one or more nominal, ordinal, interval or ratio-level independent variables.

Logistic Regression Algorithm

```
x=DF.drop(columns=["Outcome"])

y=DF["Outcome"]

xtrain,xtest,ytrain,ytest=train_test_split(x,y,test_size=0.25)

lm2=LogisticRegression()

lm2.fit(xtrain,ytrain)

y_pred=lm2.predict(xtest)

print(y_pred)

print(metrics.accuracy_score(ytest,y_pred)*100)

cr=classification_report(ytest,y_pred)

print(cr)
```

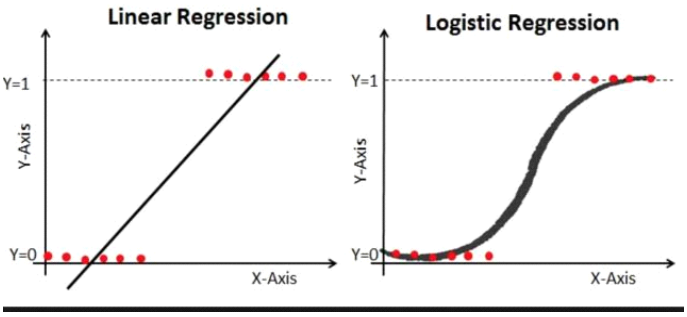
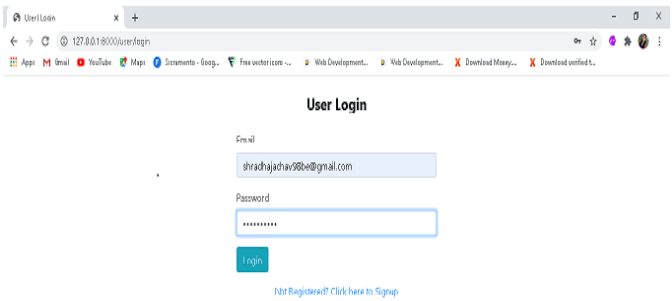
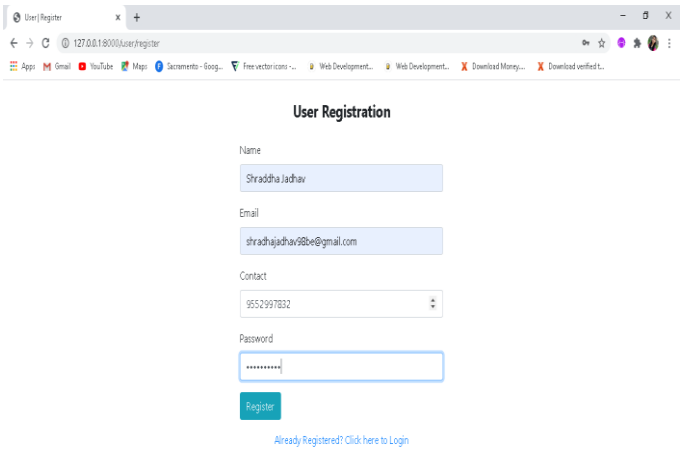
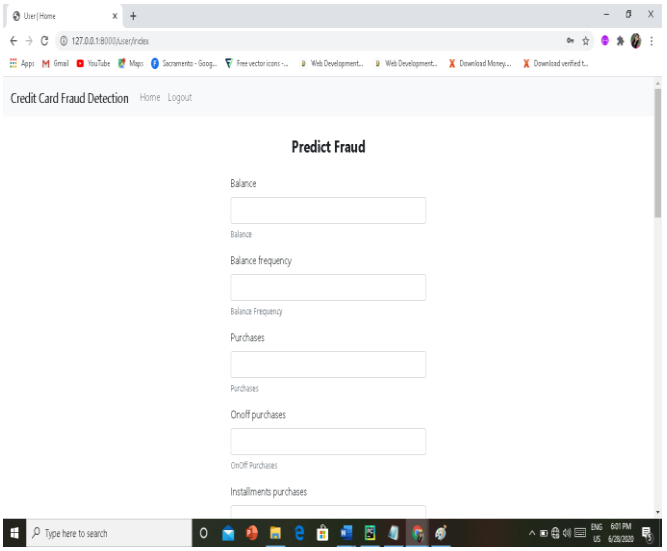


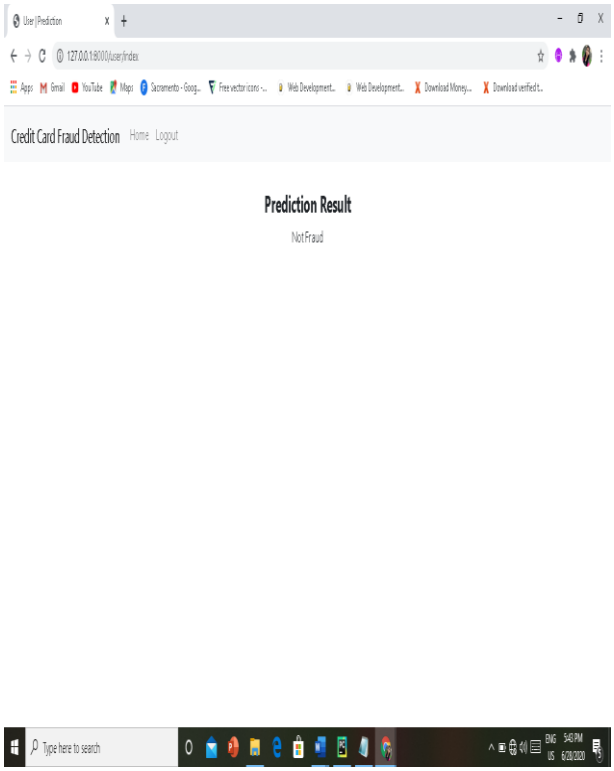
Fig.4 Logistic regression graph

4. RESULT









5. CONCLUSIONS

Credit card fraud signifies a very serious commercial problem. These frauds can lead to huge losses. online paying money by the credit card is increased and also credit card frauds so there is need to detect this happened fraud transactions and provide security to the users about credit card.

So, the main purpose of this paper is for detecting as well as preventing the frauds during transactions. There are still some problems in previous system such as accuracy. This paper has surveyed the performance of Random Forest Algorithm. Random forest obtains good results on small data set as well as large data set.

6. REFERENCES

[1] Global Facts (2019). Topic: Startups worldwide. [online] Available at: https://www.statista.com/topics/4733/startups-worldwide/ [Accessed 10 Jan. 2020].

[2] Legal Dictionary (2019). Fraud - Definition, Meaning, Types, Examples of fraudulent activity. [online] Available at: https://legaldictionary.net/fraud/ [Accessed 15 Jan. 2020].

[3] A. Mishra, C. Ghorpade, "Credit Card Fraud Detection on the Skewed Data Using Various Classification and Ensemble Techniques" 2018 IEEE International Students' Conference on Electrical, Electronics and Computer Science (SCEECS) pp. 1-5. IEEE

[4] S. Bachmayer, "Artificial Immune Systems," Artificial Immune Systems, vol. 5132, pp. 119–131, 2008.

[5] M. Krivko, "A hybrid model for plastic card fraud detection systems," Expert Systems with Applications, vol. 37, no. 8, pp. 6070–6076, Aug. 2010.

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

[7] D. J. Weston, D. J. Hand, N. M. Adams, C. Whitrow, and P. Juszczak, "Plastic card fraud detection using peer group analysis," Advances in Data Analysis and Classification, vol. 2, no. 1, pp. 45–62, Mar. 2008.

[8] O. S. Yee, S. Sagadevan, N. Hashimah, and A. Hassain, "Credit Card Fraud Detection Using Machine Learning As Data Mining Technique," vol. 10, no. 1, pp. 23–27.

[9] K. Randhawa, C. K. Loo, M. Seera, C. P. Lim, and A. K. Nandi, "Credit Card Fraud Detection Using AdaBoost and Majority Voting," IEEE Access, vol. 6, pp. 14277– 14284, 2018.

[10] N. Mahmoudi and E. Duman, "Detecting credit card fraud by Modified Fisher Discriminant Analysis," Expert Syst. Appl., vol. 42, no. 5, pp. 2510–2516, 2015.