International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 06 Issue: 04 | April - 2022 Impact Factor: 7.185 ISSN: 2582-3930

Credit Card Fault Detection using Machine Learning

Anjali Sharma[1] , Ananya Gupta[2] , Gargi Agarwal[3] , Garima Bisht[4]

[1,2,3,4]*B.Tech. Students, KIET GROUP OF INSTITUTIONS, DELHI-NCR, GHAZIABAD-201206, U.P.., INDIA*

ABSTRACT:

Credit card fraud event occurs often times and then, at that point, bring about enormous monetary misfortune. Law breakers can utilize a few advancements, for example, phishing or trojans to take others credit card data. Subsequently, effective fraud detection strategy will be significant as it can distinguish extortion when a criminal uses card to devour. A strategy is utilized to authentic exchange information including ordinary exchanges and misrepresentation once to acquire ordinary / extortion conduct highlights depends on machine learning methods.

Each year, a loss of 5% of income by an organisation is towards fraud. In fact, a RTI query revealed that a total of 2480 cases of fraudulent transactions that summed up to rupees 32,000 crores rattled 18 public sector Indian banks. As such we can say that fraud poses a serious concern for all the financial as well as trading organizations. In this paper, we have tried to explore what fraud is in financial terms and its different forms, rule-based approach to fraud detection, data science techniques used to detect these frauds and challenges faced during building a fraud detection system for transactions involving credit cards

INTRODUCTION

The following are the paper's main commitments:

1. To combat fraud detection issues, we put our dataset through a variety of machine learning algorithms, included K-nearest neighbour (KNN), logistic regression, support vector machine (SVM), and random forest.
2. On testing our dataset on above mentioned algorithm we have
   a. K- Nearest neighbour (KNN) : KNN being a supervised machine learning technique that is used to address problems like regression and classification.
   b. Random forest: We use N decision tree models using the random forest algorithm. The whole value is predicted by all models. The final goal value when will be forecasted using the majority vote method. The random forest approach generates a sample data set at random for creating the individual decision tree. The bootstrap samples are a type of sample data set. If we wish to

produce the forest with an decision trees the technique first makes and movers trap samples. Decision tree model will be will afterwards for each bootstrap sample.

c. Logistic regression: The distribution of classes in the features space are not taken into account by logistic regression. It's simple to adapt to a variety of classes (multinomial regression). It's great for categorising records that aren't known.

d. Decision tree: The simple and the most often used classification algorithm is the decision tree. The decision tree method evaluates all of the data's offered features when constructing the model and comes up with the most important ones.

Because of this benefit, decision tree algorithms are also employed to determine the significance of feature matrices which was employed in feature selection process.

One of the important features have been discovered, the model is trained using training data to provide a set of rules.

e. Support vector machine (SVM: SVM is a method for classifying and detecting patterns. Being a classification system it divides the pattern into two categories:
1: Fraudulent
2: Non-Fraudulent

Fraud by definition is an illegal act of deceiving a person, organization or entity of money, property, or legal rights. The most common type of frauds are financial frauds which cause an organization 5% of yearly losses in their revenue on an average. According to the reports of the Reserve Bank of India (RBI) a total of 911 frauds involving credit card transactions amounted to rupees 65.25 crore for the financial year 2019-2020.

Common types of frauds one may encounter in his daily life are online fraud, where all his money may be illegally transferred to different accounts, credit card frauds where details of the credit card, credit card number, pin etc. are used for illegal transfer of money. Other types of frauds include threat, stealing or theft of inventory, where a defaulter may break into a bank and undertake illegal activities causing harm to assets. Thus, it becomes really important to detect such frauds and safeguard people from losing their money.

Rule based Approach

Detecting fraud in transactions through credit card using traditional methods involves setting up an algorithm written by fraud detection analysts which is based on stringent rules. In case of change from the normal pattern while detecting a new fraud, all the alterations are carried out manually, either by modifying an existing algorithm or by constructing a new one.. In the traditional approach, whenever there is an increase in customers and their credit card data, the human effort to compile the algorithm also increases. Thus, in conclusion it can be said that the rule-based approach is costly and time consuming.

RELATED WORK:

We have used various machine learning algorithms through which we have seen that random forest algorithm is the most effective method for spotting a new scamthe credit card using machine learning algorithms. We have tested the dataset on these machine learning algorithms.

ALGORITHM:

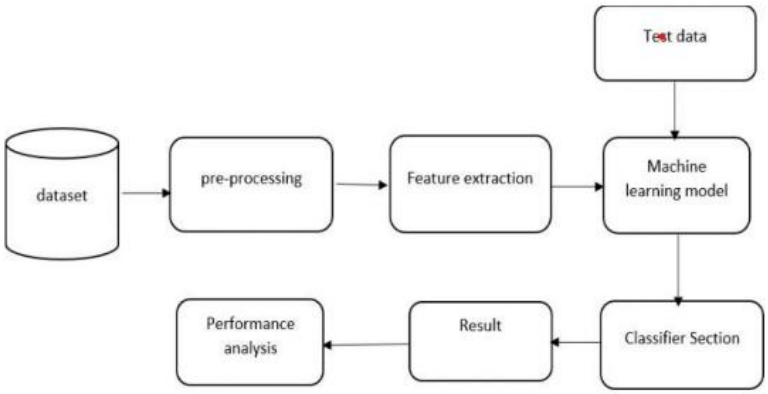
1. Taking dataset
2. The dataset used is real bank transactions made by the card holder.
3. Import necessary libraries, the necessary libraries are NumPy, pandas, seaborn, matplotlib
4. Importing Data: to import dataset pandas module is used in python, the data import command.
   Data= pd.read_csv("creditcard.csv")
5. Data understanding and processing: Vast majority of transaction in our dataset are legitimate, with only small percentage of being them fraudulent.
6. Now the data which we have is properly scaled with no duplicate or missing data.
7. Training data and testing data: Before we split our data we train and test our data - we now define dependent and independent variable 'a' dependent variable , 'b' independent variable
   a = data.drop('Class', axis = 1).values
   b = data['Class'].values
   a_train, x_test, a_train, b_test = train_test_split(a, b, test_size = 0.25, random_state = 1)
   The two sets of data: data that is used to train the model and data that will be used to test the model
   8. Building Model: one by one, testing of various machine learning models
   a. Decision tree: The decision tree model's accuracy score is
   DT = DecisionTreeClassifier(max_depth = 4, criterion = 'entropy')
   DT.fit(atrain, b_train)
   dt_bhat = DT.predict(a_test)
   for accuracy:
   print('Accuracy score of the Decision Tree model is {}'.format(accuracy_score(b_test, tree_bhat)))
   The decision tree model's F1 score is
   print('F1 score of the Decision Tree model is {}'.format(f1_score(b_test, tree_bhat)))
   Similarly, for other algorithms

TABLE 1 : CALCULATION OF ACCURACY AND F1 SCORE

S,NO.	Algorithm	Accuracy	F1 score
1.	DECISION TREE	0.99818	0.77514
2.	KNN	0.99415	0.83542
3.	LOGISTIC REGRESSION	0.99810	0.69235
4.	SVM	0.99252	0.77677
5.	RANDOM FOREST	0.99936	0.85322

From the above table we have examined that the random forest algorithm is better than other algorithms in credit card fraud detection.

FLOW-CHART: Fig1



CONCLUSION

In this paper we have examined different types of machine learning algorithms. We have evaluated highest accuracy and F1 score in random forest algorithm.

Accuracy is 99.936 and F1score is 0.85322.

REFERENCES

1. Seyedhossein, Leila, and M. R. Hashemi. ”Mining information from credit card time series for timelier fraud detection.” International Symposium on Telecommunications IEEE, 2011:619-624.

2. Shi, E., Niu, Y., Jakobsson, M., and Chow, R. (2010). Implicit Authentication through Learning User Behavior. International Conference on Information Security (Vol.6531, pp.99-113). Springer-Verlag.

3. Chan, P. K., Fan, W., Prodromidis, A. L., and Stolfo, S. J. (2002). Distributed data mining in credit card fraud detection. IEEE Intelligent Systems and Their Applications, 14(6), 67-74.

4. Gupta, Shalini, and R. Johari. ”A New Framework for Credit Card Transactions Involving Mutual Authentication between Cardholder and Merchant.” International Conference on Communication Systems and Network Technologies IEEE, 2011:22-26.

5. Bhattacharyya, S., Jha, S., Tharakunnel, K., and Westland, J. C. (2011). Data mining for credit card fraud: a comparative study. Decision Support Systems, 50(3), 602-613.