Credit Card Fraud Detection in R

Renuka Shirke[1], Manal Surve[1], Pradnesh Shitap[1]

U.G. Student, Department of Information Technology, Finolex academy of Management and Technology, Ratnagiri,

Maharashtra, India[1]

ABSTRACT: A phenomenal growth in e-commerce industry has led to an exponential increase in the use of credit cards. The number of credit card transactions especially for online purchases has led to substantial rise in fraudulent activities. In recent years, for the credit card issuing banks implementation of efficient fraud detection systems has thus become imperative. It is vital that credit card companies are able to identify fraudulent credit card transactions so that customers are not charged for items that they did not purchase. Such problems can be tackled with Data Science and its importance, along with Machine Learning, cannot be overstated. Various different classification methods, i.e. decision tree, neural networks and logistic regression, etc. are tested for their applicability in fraud detections in this paper. This study investigates the efficacy of applying classification models to credit card fraud detection problems. It provides a useful framework to choose the best model to recognize the credit card fraud risk.

KEYWORDS: Credit card fraud detection, Classification models, Logistic regression, Decision tree, Random forest, Data science, Automated fraud detection.

I.INTRODUCTION

Credit card frauds can happen by physical theft of card/details or virtually via imposters/hackers. Occurrences of credit card frauds are rare, nonetheless involving good amount of loss. Credit card frauds can be detected in-vivo through suitable Machine Learning models. It's an active research topic, unfortunately with very few data available. Size, balance and dynamic nature of credit card fraud data are real challenges. We need a fast, efficient and possibly transparent Machine Learning model to detect fraud.

This paper proposes a credit card fraud detection model that is mainly devoted to detection of external card fraud. External card fraud is mainly embodied at using the stolen, fake or counterfeit credit card to consume, or using cards to get cash in disguised forms, such as buying the expensive, small volume commodities or the commodities that can easily be changed into cash. External card frauds accounts for a majority of credit card frauds.

The main objective behind our system is to analyze different techniques and find the most appropriate technique which not only helps us to minimize the fraud but also help in future enhancement in the system.

The remainder of this paper is organized as follows. Section II reviews related work about the credit card system. Section III provides a brief description of the proposed system architecture and methodology. Section IV shows conclusion

II.LITERATURE SURVEY

"The efficiency of classification models and the authors proposed the three classification models i.e., decision tree, neural network and logistic regression. among the three models neural network and logistic regression outperforms than the decision tree." by a. shen, r. tong, y. deng, "application of classification models on credit card fraud detection", service systems and service management 2007 international conference, pp. 1-4, 2007. "machine learning technique like logistic regression, decision tree and random forest were used. sensitivity, specificity, accuracy and error rate are used to evaluate the performance for the proposed system. the accuracy for logistic regression, decision tree and random forest classifier are 90.0, 94.3, and 95.5 respectively. by comparing all the three method, found that random forest classifier is better than the logistic regression and decision tree." by lakshmi s v s s, selvanideepthikavila (2018), "credit card fraud detection using machine learning and data science" - this paper explained about how machine learning can be applied to get better results in fraud detection. while the algorithm does reach over 99.6% accuracy, its precision remains only at 28% when a tenth of the data set is taken into consideration. however, when the entire dataset is fed into the algorithm, the precision rises to 33%. being based on machine learning algorithms, the program will only increase its efficiency over time as more data is put into it" by). s p maniraj, adityasaini, swarna deep sarkar, shadabahmed's.
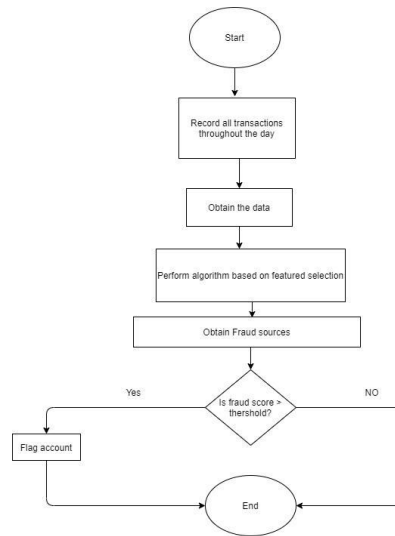
III. PROPOSED SYSTEM

Figure 1 Design of proposed algorithm

The processing steps to detect the best algorithm for the given dataset can be described as follows – In the first step we will read the dataset. Next, Random Sampling is done on the data set to make it balanced. Further we will divide the dataset into two parts i.e., Train dataset and Test dataset. Feature selection are applied for the proposed models. Next, Accuracy and performance metrics has been calculated to know the efficiency for different algorithms. And lastly, retrieve the best algorithm based on efficiency for the given dataset.
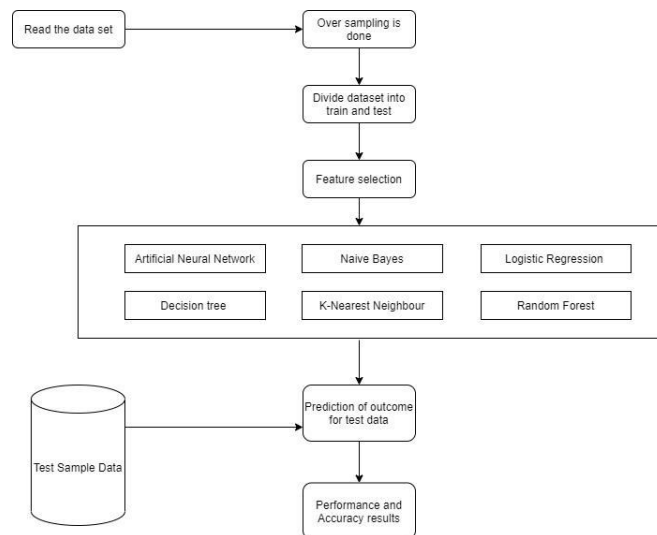

Figure 2 System Architecture

Implementation steps –
- Raw Data: The collected input data is in the form of csv files.
- Prepared Data: A process to gather context to the input data. Understanding the data for pre-processing and cleaning of datasets.
- Handling Class Imbalance: There are quite a few ways to curb this imbalance-   a. Under sampling, b. Oversampling, c. Combining the two

- Training and Testing Subset: If the dataset is imbalanced, many classifiers show bias for majority classes. The features of minority class are treated as noise and are ignored. Hence it is proposed to select a sample dataset.
- Applying algorithm: Following are the classification algorithms used to test the sub-sample dataset- a. Decision Tree b. Random Forest
- Predicting results: The test subset is applied on the trained model .The metrics used are precision and Recall score. The ROC Curve is plotted and the desirable results are achieved.

IV.CONCLUSION

In this paper, Machine learning techniques like Logistic regression, Decision Tree and Random forest, etc. are used to detect the fraud in credit card system. Sensitivity, Specificity, accuracy and error rate are used to evaluate the performance for the proposed system. We are trying to obtain best algorithm as a solution by comparing existing algorithms which will minimize the frauds. We present our work and demonstrate the advantages of the data mining techniques including neural networks, logistic regression and decision tree to the credit card fraud detection, for the purpose of reducing credit card frauds.

REFERENCES

1. A. Shen, R. Tong, Y. Deng, "Application of classification models on credit card fraud detection", Service Systems and Service Management 2007 International Conference, pp. 1-4, 2007.
https://www.researchgate.net/publication/4265092_Application_of_Classification_Models_on_Credit_Card_Fraud_Detection
2. Machine Learning For Credit Card Fraud Detection System Lakshmi S V S S1 ,SelvaniDeepthi Kavila2 1,2Department of CSE, Anil Neerukonda Institute Of Technology And Sciences(A), Visakhapatnam-531162,India
https://www.ripublication.com/ijaer18/ijaerv13n24_18.pdf
3. Credit Card Fraud Detection using Machine Learning and Data Science International Journal of Engineering Research & Technology (IJERT) ISSN: 2278-0181http://www.ijert.org IJERTV8IS090031 (This work is licensed under a Creative Commons Attribution 4.0 International License.) by S P Maniraj Assistant Professor (O.G.) &AdityaSaini, Swarna Deep SarkarShadab Ahmed
https://www.researchgate.net/publication/336800562_Credit_Card_Fraud_Detection_using_Machine_Learning_and_Data_Science