e-ISSN: 2582-5208

International Research Journal of Modernization in Engineering Technology and Science

(Peer-Reviewed, Open Access, Fully Refereed International Journal)

Volume:05/Issue:10/October-2023 Impact Factor- 7.868 www.irjmets.com

A SURVEY PAPER ON CREDIT CARD FRAUD DETECTION WITH THE HELP OF MACHINE LEARNING

Pooja Balu Wankhede[*1], **Dr. Dinesh D. Patil**[*2], **Dr. Priti Subramanium**[*3]

[*1]Student, Computer Science & Engineerig, Shri Sant Gadge Baba College Of Engineering And Technology, Bhusawal, Maharastra, India.

[*2]Head Of Dept. & Associate Professor, Computer Science & Engineerig, Shri Sant Gadge Baba College Of Engineering And Technology, Bhusawal, Maharastra, India.

[*3]Associate Professor, Computer Science & Engineerig, Shri Sant Gadge Baba College Of Engineering And Technology, Bhusawal, Maharastra, India.

DOI : https://www.doi.org/10.56726/IRJMETS45811

ABSTRACT

Credit Cards can be used in online transactions due to the convenience and ease of use. Credit card fraud is one of the leading causes of financial losses for credit card issuers and finance companies. Card fraud has cost credit card companies money. Currently, card fraud detection is the most common problem facing credit card companies. Credit card companies are searching for good systems and technologies to identify and reduce fraudulent transactions. There are a number of credit card detection techniques in machine learning. There are a number of credit card fraud detection techniques that have been examined and highlighted in this paper and have been compared in terms of their drawbacks and benefits. Credit cards are the most popular way to pay online because there are more and more people making electronic transactions susceptible to fraud. Credit cards have been a growing issue in recent years. It has caused a huge financial loss for individuals using credit cards as well as for books and merchants. Machine learning is one of the most effective techniques for detecting fraud. This paper surveys various fraud detection techniques and methods using machine learning and compares them using performance metrics, such as accuracy, precision and specificity.

Keywords: Credit Card Fraud, Random Forest, Deep Machine Learning, Online Transactions, Fraudulent Transactions.

I. INTRODUCTION

Credit fraud detection is a collection of techniques used to identify and prevent fraudulent transactions, whether online or in-store. It works by verifying that you're dealing with the correct cardholder and the transaction is legitimate. We can now pay for things online with credit cards. The number of people using credit cards for online things has increased significantly, and it's also caused a huge rise in credit card scams. Credit card scams involve using someone else's card or account info without permission. Scammers are adept enough to take advantage of any weak spots and always look for new ways to steal data, like skimming or phishing. When a website is set up to look like a legitimate site and people enter personal info like passwords, usernames, credit card info, etc., The scammers send out a ton of emails (tricks) that lead people to their fake websites. In today's highly competitive financial world, electronic payment methods play an essential role. They have made buying goods and services much easier. Not every customer that walks into the store will have cash on hand. These days, they value credit card and debit card payments highly. Consequently, businesses will have to modernize their setup in order to accept all forms of payment. It is anticipated that this scenario will worsen further during the coming years [1]. Machine learning helps you figure out which transactions are fake and which ones are legit. The emails appear to be from companies like PayPal banks, AOL and eBay they ask the victim to log their personal data in order to solve the problem. The scammer can make money by stealing the victim's identity and then stealing their money [2]. The consequences of credit card fraud have been devastating, resulting in a significant financial loss. Fraudulent credit card transactions are quickly identified using Machine Learning algorithms, which are equipped with powerful processing or computing capabilities and the capacity to process large data sets. This is one of the promising ways to reduce credit card fraud [3], [4].

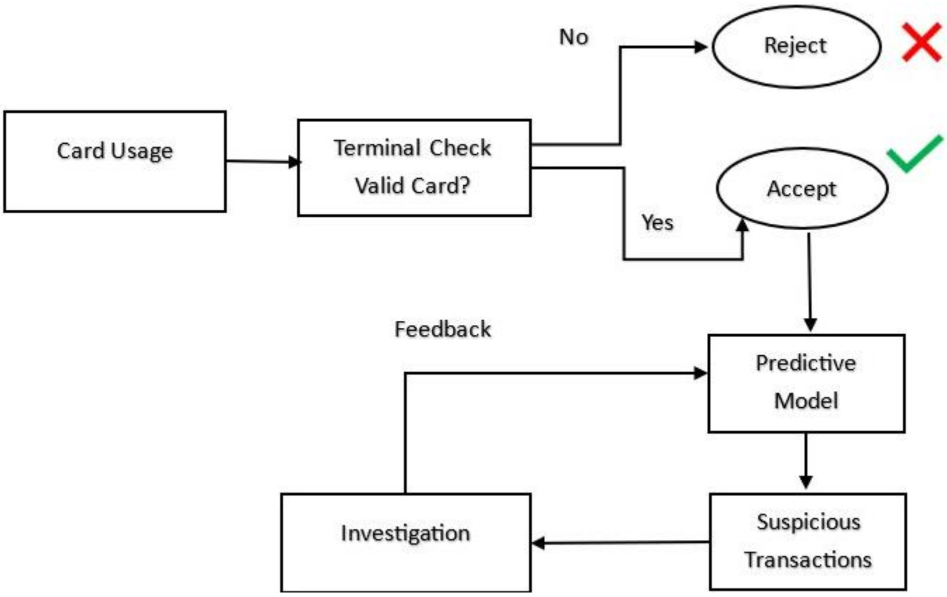

Figure 1: Fraud Detection Process

As shown in the figure1, Fraud detection process: Card usage, Terminal check, Valid Card, Reject, Accept, Predictive Model, Suspicious. At the terminal point, the transactions are first valid or not-valid. At the point of sale, there are some important things you need to check, like having enough money in your account and having a valid PIN. All valid transactions are verified and filtered accordingly. All valid transactions are then evaluated by the predictive model to determine whether they are actual or fake. The investigators follow up on each fraudulent alert and give feedback to the predictive model.

II. LITERATURE REVIEW

Xiaohan Yu et al. [5] Have proposed a Deep Neural Network (DNN) algorithm for the detection of credit card fraud. This paper outlines the neural network approach and the applications of the algorithm, as well as the preprocessing and focal loss techniques used to resolve data issues in the dataset. Fraudulent transactions are effectively identified with the aid of Machine Learning algorithms, which are capable of processing large datasets and have high computing power, making them a promising tool for reducing credit card fraud.

Y. Sahin et al. [6] Have proposed a decision tree method that cuts down on the total cost of misclassifying data and chooses the right splitting property at each node. They also compared their decision tree method for fraud identification to other models and showed it works well using performance metrics like accuracy and real positive rate.

Kibria & Sevkli [7] Utilize the grid search, build a deep learning (DML) model. The Support Vector Machine (SVM) model and Logistic Regression (LR) algorithm are two common machine learning algorithms whose performance is compared with the model's. The developed model is tested on the credit card dataset and the results are compared with Logistic regression and Support vector machine models.

Khatri et al., [8] a performance analysis of multilayer (ML) techniques for credit card fraud detection was conducted. This analysis focused on four ML approaches: DT, k-nearest neighbor (KNN), linear regression (LR), linear regression (R) and non-linear regression (NB). To evaluate the performance of each of these ML approaches, a dataset of highly imbalanced European cardholders was used. The experiments employed precision as a key performance metric to evaluate each classifier. The experimental results indicated that DT, KR, LR and RF achieved accuracies of 85.11, 91.11, 87.5, 89.77 and 6.52% respectively.

Vimala devi et al. [9] Used three machine learning algorithms to detect fake transactions. There are lots of ways to measure how well a classifier or predictor is doing, like the support vector machine, random forest, and decision tree. These metrics are either dependent on the prevalence of the classifier or independent of it. These techniques are used to detect credit card fraud, and they compare the results of the algorithms.

III. TECHNIQUES TO USE IN CREDIT CARD FRAUD DETECTION

1. RANDOM FOREST:

A random forest classifier finds the decision trees within a subset of the data and aggregates their data to get the full predictive power of the dataset. Using a ton of trees in a forest increases accuracy and eliminates the problem of overfitting. Random Forest predicts output with great accuracy and runs quickly even with big data. It also keeps accuracy even when a lot of data is gone. Random Forest is great for both classifying and regression tasks. This algorithm is capable of processing large datasets with a high degree of dimensionality. It enhances the model's accuracy and eliminates the issue of overfitting. The tree-based random forest is trained using two-stage techniques: first, the random forest is generated by combining N trees, and then an estimate is made for each tree that is generated in the first stage [10]. Since different trees are trained at the same time, the total model reduces a large number of variations. Random Forest learns each tree as if it were a separate classifier trained on the same data set. The overall learning ability of the random forest increases due to the use of this learning strategy and divide [11][12].

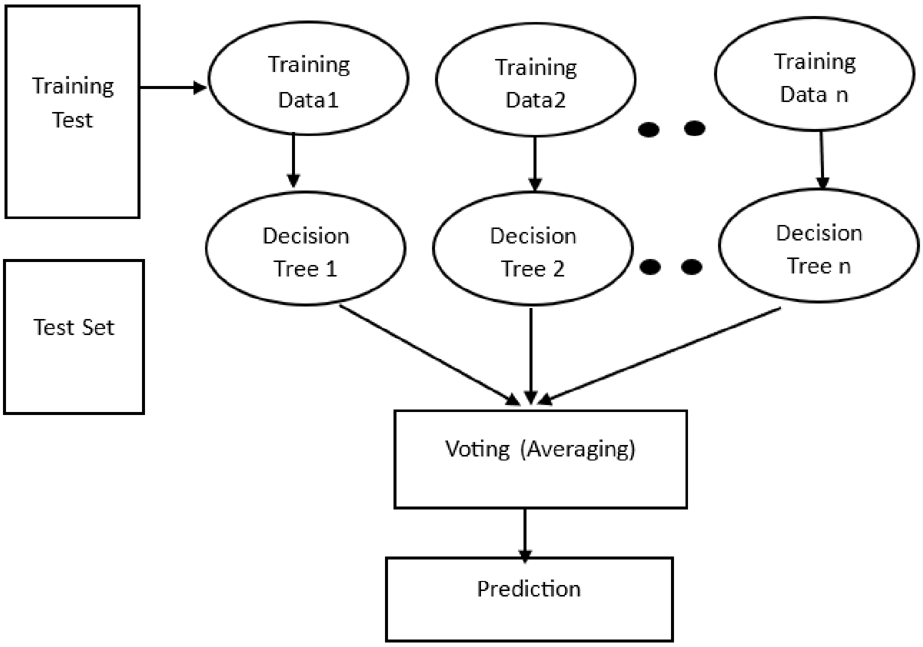

Figure 2: Random Forest

Steps:

These steps follow in Figure 2 above. The first step: select (K) the drill set as random data points. Second step, build the DT linked to the selected Subsets. Next, decide how many decision trees (N) to construct. Then, pass out steps 1 and 2 one more time. Finally, find the forecasts for each decision tree of the new data points and rank the new data points according to the category that gets the most votes. Let us examine the following example to comprehend the operation of RF: Let's say you have a dataset with various fruit image kinds. This dataset will be returned by the RF classifier.

2. DECISION TREE:

This is a supervised learning methodology, which is a graphical representation of possible solutions to a decision based on certain situations [13], as shown in Figure 3 and is a tree-based classifier. Starting from a root node, nodes within nodes symbolize the features of a dataset, decision rules are represented by branches, and the result is represented by each leaf node. In a decision tree, they have the purpose of making decisions and communicating. A decision tree simply poses a question and then breaks it down into sub trees according to the answer. DT can be used to solve classification and regression problems; its most common use cases are classification problems. The algorithm looks at the top of a tree to find the dataset's class. It compares the root trait to the record attribute and tracks the offshoot on its way to the nearest node, which is calculated according to the relation [14].

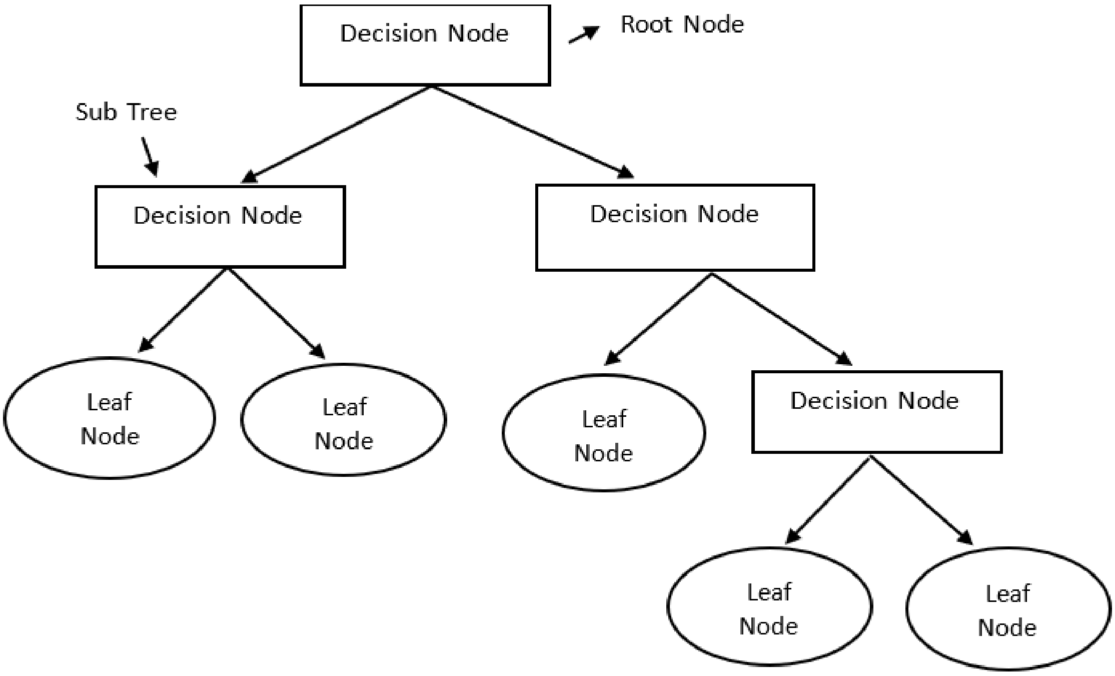

Figure 3: Decision Tree

Steps:

As shown in figure 3 above, the first step begins with S as the root node that contains the complete set of data. Using the attribute selection measure, identify the best trait in the dataset in the second step. If the nodes cannot be classified, at that point, the last node is called a foliate node. The decision node is the location where the decision to split the tree is made. The final choice is represented by a leaf node [15]. A decision node and a single leaf node are further divided from the root node. At the end, the node is split into two parts (accepted and rejected offers).

3. LOGISTIC REGRESSION:

LR does not assume that the independent variables are linearly related, nor does it assume that the variance within each group is the same, making LR a less rigorous statistical analysis. A logistic regression analysis was conducted to estimate the likelihood of credit card fraud [16]. Categorical variables are predicted using logistic regression, which uses dependent variables. Think about the following situation: Right now, there are two classrooms in use. Regression demonstrates the connections between a number of independent and dependent variables [17]. An algorithm that supports both regression and classification, but is mainly used for classification. Probabilities 0 and 1 are then calculated by the algorithms.

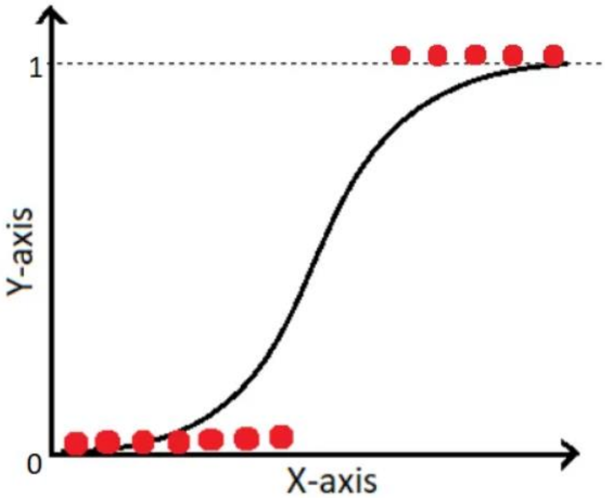

Figure 4: Decision Tree

IV. ADVANTAGES AND DISADVANTAGES OF MACHINE LEARNING TECHNIQUES

Techniques	Advantage	Disadvantage
Random Forest	• RF can perform both Classification and Regression tasks. • Can also handle large data sets increase dimension. • This is to promote a thorough model and avoid over-compatibility problems.	• Although RF can be asked for both classification and regression function, this function not any more suitable for regression tasks.
Decision Tree	• It's simple to understand and to put into action. • It can be very helpful in problem solving for decision making and action planning. • High flexibility, which helps you to think of all possible solutions to a problem. • Minimal data cleaning required.	• This technique's numerous layers contribute to its complexity. • It might have an overfitting problem, which the RF algorithm can fix. The DR Arithmetic's level of difficulty could rise.
Logistic Regression	• It is easier to use, understand, and very effective to train. Regarding the distribution of classes in the feature space, it makes no assumptions.	• The nonlinear problem cannot be solved by logistic regression because it has a linear decision surface.

V. RESULTS AND DISCUSSION

We note that while every technology has benefits, there are drawbacks as well that reduce its efficiency and impair its capacity to recognize and identify fraudulent transactions.

VI. CONCLUSION

Credit card fraud has become a worldwide concern. Fraud causes huge financial losses around the world. The primary goal of this research is to create algorithms that give credit card issuers the information they need to more quickly and affordably detect fraudulent transactions. Different machine learning algorithms compare, including Logistic Regression, Decision Trees and Random Forests. Each of the fraud detection techniques discussed in this survey article has benefits and drawbacks. The researchers forecast and highlight transactions that are fraudulent by using various performance metrics and algorithms. These kinds of surveys will enable researchers to develop the most accurate hybrid approach for detecting fraudulent credit card transactions.

ACKNOWLEDGEMENTS

The authors wish to thank Shri Sant Gadge Baba College of Engineering and Technology, Bhusawal, India for supporting this research work.

VII. REFERENCES

[1] Y. Abakarim, M. Lahby, and A. Attioui, “An efficient real time model for credit card fraud detection based on deep learning,” in Proc. 12th Int. Conf. Intell. Systems: Theories Appl., Oct. 2018, pp. 1–7, doi: 10.1145/3289402.3289530.

[2] K. J. Barker, J. D ‘Amato, and P. Sheridon, “Credit card fraud: awareness and prevention”, J. Finance Crime, vol. 15, no. 2, pp 398–410, 2008.

[3] C. Reviews, “a Comparative Study: Credit Card Fraud,” Vol. 7, issue 19, pp. 998-1011, 2020.

[4] R. Sailusha, V. Gnaneswar, R. Ramesh, and G. Ramakoteswara Rao, “Credit Card Fraud Detection Using Machine Learning”, Proc. Int. Conf. Intell. Comput. Control Syst. ICICCS 2020, no. Iciccs, pp. 1264–1270, 2020.

[5] X. Yu, X. Li, Y. Dong, and R. Zheng, “A Deep Neural Network Algorithm for Detecting Credit Card Fraud,” Proc. -2020 Int. Conf. Big Data, Artif. Intell. Internet Things Eng. ICBAIE 2020, pp. 181-183, 2020.

[6] Y. Sahin and E. Duman, "Detecting credit card fraud by ANN and logistic regression," 2011 International Symposium on Innovations in Intelligent Systems and Applications, Istanbul, Turkey, 2011, pp. 315-319,

IEEE doi: 10.1109/INISTA.2011.5946108.

[7] G. Kibria and M. Sevkli, "Application of Deep Learning for Credit Card Approval: A Comparison with Application of Deep Learning for Credit Card Approval: A Comparison with Two Machine Learning Techniques", no. January, 0-5,2021.

[8] S. Khatri, A. Arora and A. P. Agrawal, "Supervised Machine Learning Algorithms for Credit Card Fraud Detection: A Comparison," 2020 10th International Conference on Cloud Computing, Data Science & Engineering (Confluence), Noida, India, 2020, pp. 680-683.

[9] J. Vimala Devi and K. S. Kavitha, "Fraud Detection in Credit Card Transactions by using Classification Algorithms," 2017 International Conference on Current Trends in Computer, Electrical, Electronics and Communication (CTCEEC), Mysore, India, 2017, pp. 125-131, doi: 10.1109/CTCEEC.2017.8455091.

[10] "Machine Learning Random Forest Algorithm – Javatpoint", https://www.javatpoint.com/machine-learning-random-forest-algorithm(accessed Apr, 2023).

[11] R. Sailusha, V. Gnaneswar, R. Ramesh, and G. Ramakoteswara Rao," Credit Card Fraud Detection Using Machine Learning," Proc. Int. Conf. Intell. Comput. Control Syst. ICICCS 2020, no. Iciccs, pp. 1264–1270, 2020, doi: 10.1109/ICICCS48265.2020.9121114.

[12] I. Sadgali, N. Sael, and F. Benabbou," Detection and prevention of credit card fraud: State of art," MCCSIS 2018 - Multi Conf. Comput. Sci. Inf. Syst. Proc. Int. Conf. Big Data Anal. Data Min. Comput. Intell. 2018, Theory Pract. Mod. Comput. 2018 Connect. Sma, no. March 2019, pp. 129–136, 2018.

[13] V. Patil and U. Kumar Lilhore," A Survey on Different Data Mining & Machine Learning Methods for Credit Card Fraud Detection," Int. J. Sci. Res. Comput. Sci. Eng. Inf. Technol. © 2018 IJSRCSEIT, vol. 5, no. 10, pp. 320–325, 2018, doi:10.13140/RG.2.2.22116.73608.

[14] "Machine Learning Random Forest Algorithm – Javatpoint," https://www.javatpoint.com/machine-learning-random-forest-algorithm(accessed Apr, 2023).

[15] Liang J, Qin Z, Xiao S, Ou L, Lin X," Efficient and secure decision tree classification for cloud-assisted online diagnosis services," IEEE Trans Dependable Secure Comput. 2019, vol 18, no.4, pp. 1632-1644, 2019.

[16] S. Venkata Suryanarayana, G. N. Balaji, and G. Venkateswara Rao," Machine learning approaches for credit card fraud detection," Int. J. Eng. Technol., vol. 7, no. 2, pp. 917–920, 2018, doi: 10.14419/ ijet. v7i2.9356.

[17] J. Han, M. Kamber, "Data Mining: Concepts and Techniques", Second ed, Morgan Kaufmann Publishers, 2006, pp. 285–464.