IJIRSET
International Journal of Innovative Research in
SCIENCE | ENGINEERING | TECHNOLOGY

e-ISSN: 2319-8753 | p-ISSN: 2347-6710


# INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

## IN SCIENCE | ENGINEERING | TECHNOLOGY

Volume 12, Issue 4, April 2023

ISSN INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 8.423

9940 572 462 | 6381 907 438 | ijirset@gmail.com | www.ijirset.com

International Journal of Innovative Research in Science, Engineering and Technology (IJIRSET)

|e-ISSN: 2319-8753, p-ISSN: 2347-6710| www.ijirset.com | Impact Factor: 8.423| A Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 4, April 2023 ||

| DOI:10.15680/IJIRSET.2023.1204106 |

# Review paper on Credit Card Fraud Prediction using Machine Learning Algorithm

**Vivek Kumar[1], Dr. Neelesh Jain[2], Dr. Ahtesham Farooqui[3]**

M. Tech. Scholar, Department of Computer Science and Engineering, SAM College of Engineering and Technology, Bhopal, India[1]

Professor, Department of Computer Science and Engineering, SAM College of Engineering and Technology, Bhopal, India[2]

Associate Professor, Department of Computer Science and Engineering, SAM College of Engineering and Technology, Bhopal, India[3]

**ABSTRACT:** The advent of e-commerce has resulted in many online transactions, which inturn has led to many security issues. Online payments or electronic payments are provided with maximum security, even though, like every invention by human, is not free from flaws. E-payments are still a most popular mode of transaction, hence attracts suppliers and customers alike. Hence effective means of combating the online threats is the need for the day. One of the major e-transactions that face these threats is the credit/debit cards. Of the security issues facing banks everywhere, prevention of credit cards frauds has always been a high priority credit card frauds come in several ways. In this research, we have been told the credit card fraud and the techniques used to reduce that fraud. A number of techniques have been created to prevent fraud from the credit card, which can reduce the fraud. But along with reducing the fraud there are some drawbacks, advantages, characteristics with these techniques. There are many techniques such as different machine learning algorithm by which fraud is detected and removed.

**KEYWORDS: -**Machine Learning, Credit Card, Fraud Prediction

## I. INTRODUCTION

The process of electronic payment has been used since 1970's. Many methods were designed to provide payments. After the internet came into being and after the incorporation of internet into many areas, the world has seen an explosive number of people using it. Payment using electronic means has started becoming a common mode of transaction during the late 90's. It was during this period when ideas came about for cashless transactions. These served as the basis for the electronic commerce or cashless transactions. This was also the period when electronic commerce came into full view. This got researchers working on the process and many academic studies conducted under this area. Since this process also involved a lot of commercial interest, many commercial agencies were also interested in this area and hence this area of research started bustling with activity. Many ideas were proposed for carrying out these cashless transactions, some of them were even launched into the market. But many failed to reach the targeted audience and hence failed. These systems achieved quite extensive deployment but failed to generate an economic return. At the same time many companies started up new methods of payments for business-to-consumer sector [1].

The Electronic Payment (E-payment) is a method of value exchange in electronic commerce, where the value is transferred via the Internet andcommunication technologies. The electronic payment systems have evolved from traditional payment systems and consequently the two types of systems have much in common. An electronic payment system denotes any kind of network service that includes the exchange of money for goods or services. E-payment is conducted in different e-commerce categories such as Business-to Business (B2B), Business-toConsumer (B2C), Consumer-to- Business (C2B) and Consumer-to-Consumer (C2C) [2, 3].

The advent of e-commerce has initiated a new issue. It deals with the security while performing these transactions. One of the main problems that persuades in the issue of online transactions is the credit card frauds. These frauds have always been the highest of priority to the banks. These issues are of highest importance and are set to grow more. This problem is also set to strike the customer base as one of the highest priority issues. Over the recent years, these types of frauds have seen a huge raise, due to the increase in the usage of online transactions. The usage of payment cards has triggered a raise in the corresponding frauds and hence has become an issue of importance. The increase in usage

|e-ISSN: 2319-8753, p-ISSN: 2347-6710| www.ijirset.com | Impact Factor: 8.423| A Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 4, April 2023 ||

| DOI:10.15680/IJIRSET.2023.1204106 |

of cards and crime related to it has converted these types of frauds into an organized crime activity. The use of credit cards has become one of the common tasks in the life of an average person [4].
The usage of credit cards is one of the easier modes of transactions; hence it is also the most flexible and the easy way of payment. Detection of frauds performed using credit cards is a tedious as well as a crucial process. Since it also involves the customer's interests and flexibility, this process should be performed carefully. This has recently become a study of great importance due to the increase in the usage of credit cards and increase in the frauds detected. Even though this processis of high priority, handling this in an efficient way is very important, since a large customer base is involved [5, 6]. A simple flaw in the detection process might prove to be a heavy loss. Fraud is defined as "willful misrepresentation intended to deprive another of some right" and is a major source of concern in a number of applications such as ecommerce, telecommunication industry, computer intrusion, etc. Credit card fraud is a growing problem in the credit card industry [7].

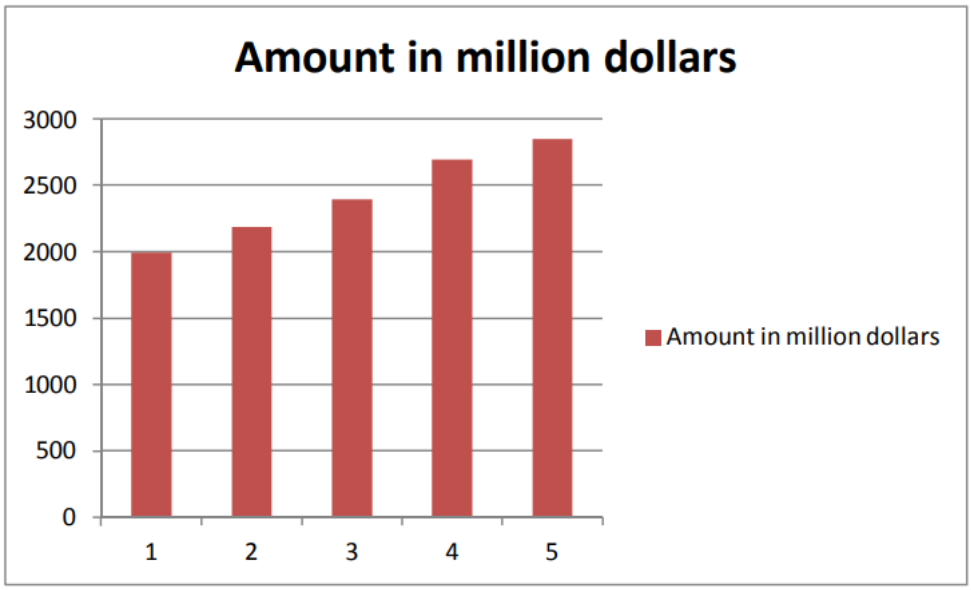


Figure 1: Yearly Credit Card Frauds

## II. RELATED WORK

**Vipul Jain et al. [1],** one of the most significant issues affecting the financial sector is credit card fraud. The covid-19 pandemic and advancements in technology are contributing to an increase in users and credit card usage. Cases of fraud also continue to rise on a daily basis as more people use credit cards. It is difficult to develop an efficient method for the detection of credit card fraud despite the efforts of the research community to investigate numerous credit card fraud detection techniques and the shifting nature of credit card fraud. A dataset based on actual credit cards was used in this study. Random Forest, logistic regression, and AdaBoost are the three machine learning algorithms used to find the fraud transaction in this dataset. Based on their accuracy and Mathews Correlation Coefficient (MCC) Score, the machine learning algorithms are compared. The Random Forest Algorithm had the highest MCC score and accuracy of the three. The machine learning web application is created with the help of the Streamlit framework.

**Yathartha Singh et al. [2],** in order to prevent customers from being charged for goods they have not yet purchased, Credit Card Management Organizations are crucial in selecting fraudulent transactions for your credit rating. With the assistance of data and technical data and machine knowledge, these issues can be resolved. This novice demonstrates credit card fraud identification and the knowledge gained through the use of gadgets. Modeling past credit card transactions with the facts of those who have been found to be fraudulent is part of the problem of detecting a credit score. The version is then used to delay determining whether or not new transactions are fraudulent. Our objective is to minimize fraudulent misclassification and detect all fraudulent transactions. A common class model is for detecting credit card fraud. We focused on the analysis and preprocessing of several anomaly detection algorithms and record

|e-ISSN: 2319-8753, p-ISSN: 2347-6710| www.ijirset.com | Impact Factor: 8.423| A Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 4, April 2023 ||

| DOI:10.15680/IJIRSET.2023.1204106 |

sets in PCA-converted credit card transaction statistics, such as "neighbor outliers" and "forest zone isolation" algorithms.

**Dileep M R et al. [3],** in the world of finance, as the technology grown,new systems of business making came into picture. Credit cardsystem is one among them. But because of lot of loop holes inthis system, lot of problems are aroused in this system in the method of credit card scams. Due to this the industry and customers who are using credit cards are facing a huge loss.

There is a deficiency of investigation lessons on examiningpractical credit card figures in arrears to privacy issues. In the manuscript an attempt has been made for finding the frauds in the credit card business by using the algorithms which adopted machine learning techniques. In this regard, two algorithms are used viz Fraud Detection in credit card using Decision Tree and Fraud Detection using Random Forest. The efficiency of the
model can be decided by using some public data as sample. Then, an actual world credit card facts group from a financial institution is examined. Along with this, some clatter is supplemented to the data samples to auxiliary check the sturdiness of the systems. The significance of the methods used in the paper is the first method constructs a tree against the activities performed by the user and using this tree scams will be suspected. In the second method a user activity based forest will have constructed and using this forest an attempt will be made in identifying the suspect. The investigational outcomes absolutely show that the mainstream elective technique attains decent precision degrees in sensing scam circumstances in credit cards.

**F. Itoo et al. [4],** the threat of financial fraud is growing at a faster rate and has a devastating effect on the economy, collaborative institutions, and administration. Because of the development of internet technology, credit card transactions are growing at a faster rate, leading to a high level of internet dependence. With advancements in technology and an increase in credit card use, fraud rates pose a threat to the economy. Fraudsters are also developing new patterns or loopholes to pursue credit card transactions as a result of the inclusion of new security features. because of which frauds' and regular transactions' patterns of behavior constantly shift. Additionally, the credit card data is highly skewed, making it difficult to accurately predict fraudulent transactions. The re-sampling (over-sampling or under-sampling) technique is used to pre-process skewed or imbalanced data in order to get better results. This study used three different proportions of datasets, and random under-sampling was used for skewed datasets. The following three machine learning algorithms are utilized in this work: Naive Bayes, logistic regression, and K-nearest neighbor The comparative analysis of these algorithms records their performance. The work is done in Python, and the accuracy, sensitivity, specificity, precision, F-measure, and area under the curve are used to measure how well the algorithms work. Based on these measurements, a logistic regression-based model for predicting fraud was found to be superior to other prediction models like K-nearest neighbor and naive bayes. Applying techniques for under sampling to the data prior to developing the prediction model also yields better results.

**E. S. C. R. S K Saddam Hussain et al. [5],**the primary goal of this project is to find real-world credit card fraud. The number of credit card transactions has significantly increased as a result of recent growth. The goal is to get items from an account without having to pay for them or use money that hasn't been approved. All credit card issuer banks must reduce the cost of implementing an efficient fraud detection system. The fact that neither the card nor the cardholder are required to complete a credit card transaction is one of the most difficult obstacles. As a result, the seller is unable to determine whether the customer making the purchase is an actual cardholder. This system, which makes use of the random forest algorithm, the decision tree algorithm, and the support vector machine algorithm, improves fraud detection accuracy. A classification procedure for observing the data set and improving the accuracy of the resulting data is known as a random forest algorithm. The precision, sensitivity, and accuracy of the techniques are used to evaluate their performance. Some of the provided data are processed to identify fraud detection and provide the graphic model with visualization.

**E. Ileberi et al. [6],**the number of daily online card transactions has increased as a result of technological advancements like e-commerce and financial technology (FinTech) applications. Consequently, there has been an increase in credit card fraud that affects merchants, banks, and card issuers. Therefore, it is absolutely necessary to develop mechanisms that guarantee the integrity and security of credit card transactions. Using real-world imbalanced datasets generated from European credit cardholders, we implement a machine learning (ML)-based framework for credit card fraud detection in this study. Using the Synthetic Minority over-sampling Technique (SMOTE), we re-sampled the dataset to eliminate class imbalance. The following ML techniques were used to evaluate this framework: Decision Tree (DT), Extra Tree (ET), Extreme Gradient Boosting (XGBoost), Support Vector Machine (SVM),

|e-ISSN: 2319-8753, p-ISSN: 2347-6710| www.ijirset.com | Impact Factor: 8.423| A Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 4, April 2023 ||

| DOI:10.15680/IJIRSET.2023.1204106 |

Logistic Regression (LR), Random Forest (RF), and The Adaptive Boosting (AdaBoost) method was used in conjunction with these machine learning algorithms to improve their classification quality. The accuracy, recall, precision, Matthews Correlation Coefficient (MCC), and Area Under the Curve (AUC) were used to evaluate the models. To further validate the findings of this study, the proposed framework was applied to a highly skewed synthetic credit card fraud dataset. The results of the experiments showed that using the AdaBoost improves the proposed methods' performance. In addition, the boosted models produced outcomes that were superior to those produced by existing methods.

**Yakub K. Saheed et al. [7],** credit Card Fraud (CCF) is a serious challenge facing credit card holder and the credit card delivering companies in the past decades. There are two levels CCF are performed, the transaction level frauds and application level frauds. This paper focuses on the application level of CCF detection using Genetic Algorithm (GA) as a feature selection technique. The GA feature selection technique is in two phases, the first phase is designated as the first priority features where eight (8) attributes were selected as the fittest attributes. At second stage which is referred to as the second priority features where another set of eight (8) attributes were considered and selected. The Naïve Bayes (NB), Random Forest (RF) and Support Vector Machine (SVM) supervised machine learning techniques were used for the detection of CCF on German credit card dataset which is an imbalance dataset. The experimental findings of the proposed model revealed that the first priority features are the most important features. Also, the obtained results showed that the RF algorithm outperformed NB and SVM in terms of accuracy, fraud detection rate and precision.

**VaishnaviNath et al. [9],**presents the impact of credit card in e-commerce, different features provided by credit card and how transaction fraud is occurring in today's e-payment system. To overcome with this problem author proposed data mining technique to secure credit card payment system. Data mining techniques used different data analysis tools to find out unknown, illegal and different pattern from large number of payment transaction data set. Author discussed about different types of fraud detection techniques available for credit card transaction, impact of credit card fraud on card holder, merchant and bank. Paper describes different types of prevention techniques available for credit card fraud system that helps to prevent credit card fraud.

## III. MACHINE LEARNING ALGORITHM

Uproarious information is available in the heap of substance that will be identified through the anomaly strategies. The information can be spatial or can be a transient method spatial connected with the geological conditions and worldly connected with the time perspectives [11]. The principle point of exception identification is to deal with the loud information that is introduced in the heap of text. Different methods for recognizing abnormalities in Text are specified in below:

### Learning

The main property of an ML is its capability to learn. Learning or preparing is a procedure by methods for which a neural system adjusts to a boost by making legitimate parameter modifications, bringing about the generation of wanted reaction. Learning in an ML is chiefly ordered into two classes as [12].

- Supervised learning
- Unsupervised learning

### Supervised Learning

Regulated learning is two stage forms, in the initial step: a model is fabricated depicting a foreordained arrangement of information classes or ideas. The model developed by investigating database tuples portrayed by traits. Each tuple is expected to have a place with a predefined class, as dictated by one of the qualities, called to have a place with a reclassified class, as controlled by one of the traits called the class name characteristic. The information tuple are dissected to fabricate the model all things considered from the preparation dataset [13].

### Unsupervised learning

It is the kind of learning in which the class mark of each preparation test isn't knows, and the number or set of classes to be scholarly may not be known ahead of time. The prerequisite for having a named reaction variable in preparing information from the administered learning system may not be fulfilled in a few circumstances.

|e-ISSN: 2319-8753, p-ISSN: 2347-6710| www.ijirset.com | Impact Factor: 8.423| A Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 4, April 2023 ||

| DOI:10.15680/IJIRSET.2023.1204106 |

Data mining field is a highly efficient techniques like association rule learning. Data mining performs the interesting machine-learning algorithms like inductive-rule learning with the construction of decision trees to development of large databases process. Data mining techniques are employed in large interesting organizations and data investigations. Many data mining approaches use classification related methods for identification of useful information from continuous data streams.

**Nearest Neighbors Algorithm**

The Nearest Neighbor (NN) rule differentiates the classification of unknown data point because of closest neighbor whose class is known. The nearest neighbor is calculated based on estimation of k that represents how many nearest neighbors are taken to characterize the data point class. It utilizes more than one closest neighbor to find out the class where the given data point belong termed as KNN. The data samples are required in memory at run time called as memory-based technique. The training points are allocated weights based on their distances from the sample data point. However, the computational complexity and memory requirements remained key issue. For addressing the memory utilization problem, size of data gets minimized. The repeated patterns without additional data are removed from the training data set [14].

**Naive Bayes Classifier**

Naive Bayes Classifier technique is functioned based on Bayesian theorem. The designed technique is used when dimensionality of input is high. Bayesian Classifier is used for computing the possible output depending on the input. It is feasible to add new raw data at runtime. A Naive Bayes classifier represents presence (or absence) of a feature (attribute) of class that is unrelated to presence (or absence) of any other feature when class variable is known. Naïve Bayesian Classification Algorithm was introduced by Shinde S.B and AmritPriyadarshi (2015) that denotes statistical method and supervised learning method for classification. Naive Bayesian Algorithm is used to predict the heart disease. Raw hospital dataset is employed. After that, the data gets preprocessed and transformed. Finally by using the designed data mining algorithm, heart disease was predicted and accuracy was computed.

**Support Vector Machine**

Support Vector Machine (SVM) is a method used in pattern recognition and classification. It is a classifier to predict or classify patterns into two categories; fraudulent or non fraudulent. It is well suited for binary classifications. As any artificial intelligence tool, it has to be trained to obtain a learned model. SVM has been used in many classification pattern recognition problems such as text categorization, and face detection. SVM is correlated to and having the basics of nonparametric applied statistics, neural networks and machine learning [10] [11][ 12]. SVM weight implements cost sensitive learning. Similar to SVM, the weighted SVM is used to maximize the margin of separation and minimize the classification error. The margin boundary is used to separate the classes. In Cost Based Support Vector Machine (CS-SVM) different weights are assigned to the classes. Effective decision boundary is learned by adjusting the weights of the different classes. It improves the accuracy of the prediction rate. To handle the imbalanced dataset, the algorithm using different error cost for the positive (C+) and the negative (C-) classes is used [13]. The training of a SVM classifier involves finding a hyperplane, as its decision surface, that separates the positive training examples from the negative ones with the largest margin.

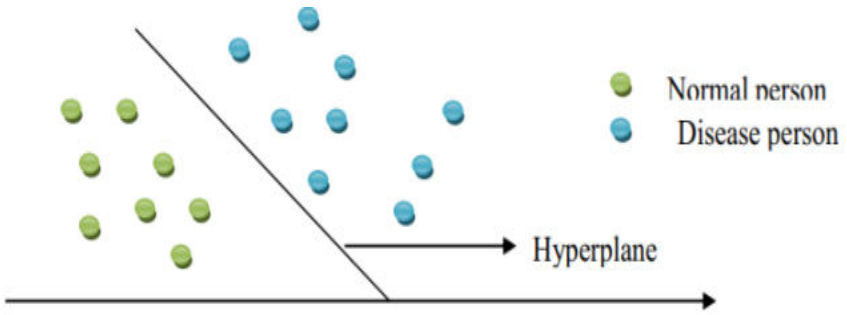


Figure 2: Support Vector Classification

## IV. SIMULATION PARAMETER

Dataset was separated into two datasets (70%/30%, preparing/testing) to keep away from any predisposition in preparing and testing. Of the information, 70% was utilized to prepare the ML model, and the excess 30% was utilized

|e-ISSN: 2319-8753, p-ISSN: 2347-6710| www.ijirset.com | Impact Factor: 8.423| A Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 4, April 2023 ||

| DOI:10.15680/IJIRSET.2023.1204106 |

for testing the presentation of the proposed movement arrangement framework. The articulations to compute accuracy and review are given in Equations (2) and (3).

Accuracy gives a proportion of how precise your model is in anticipating the real up-sides out of the absolute up-sides anticipated by your framework. Review gives the quantity of real up-sides caught by our model by grouping these as obvious positive. F-measure can give a harmony among accuracy and review, and it is liked over precision where information is uneven.

Accordingly, F-measure was used in this review as a presentation metric to give a decent and fair measure utilizing the equation.

$$\text{Precision} = TP\ TP + FP \times 100$$
$$\text{Recall} = TP\ TP + FN \times 100$$

$$F - \text{measure} = 2 * \text{Precision} * \text{Recall}\ \text{Precision} + \text{Recall} \times 100$$

Where,

TP—True Positive, FP—False Positive, FN—False Negative

## V. CONCLUSION

This research provides a complete composition of structures for efficient fraud detection. It initially starts with the usage of clustering and outlier detection techniques. These techniques are considered to be the basis for finding data that does not belong to the current data pattern. These are further made accurate by the usage of SVM and behavior bases SVM. SVM, being a binary classifier helps in providing the user with a result of whether the current transaction is legitimate or fraudulent. There are many techniques of data mining such as Decision tree, Neural network, Genetic algorithm, HMM etc. The main goal of the Fraud Detection System is to eliminate the fraud completely from the system. In this use of these, credit card transactions have been largely pursued to remove fraud. Each method has contributed a lot on its own behalf. With these methods, fraud can be reduced to a great extent. And our system also does.

## REFERENCES

[1] Vipul Jain, H Kavitha and S Mohana Kumar, “Credit Card Fraud Detection Web Application using Streamlit and Machine Learning”, International Conference on Data Science and Information System (ICDSIS), IEEE 2022.

[2] Yathartha Singh, Kiran Singh and Vivek Singh Chauhan, “Fraud Detection Techniques for Credit Card Transactions”, 3rd International Conference on Intelligent Engineering and Management (ICIEM), IEEE 2022

[3] Dileep M R, Navaneeth A V and Abhishek M, “Novel Approach for Credit Card Fraud Detectionusing Decision Tree and Random Forest Algorithms”, Third International Conference on Intelligent Communication Technologies and Virtual Mobile Networks (ICICV 2021).

[4] F. Itoo and S. Singh "Comparison and analysis of logistic regression Naïve Bayes and KNN machine learning algorithms for credit card fraud detection" International Journal of Information Technology vol. 13 no. 4 pp. 1503-1511 2021.

[5] E. S. C. R. S K Saddam Hussain "Fraud Detection in Credit Card Transactions Using SVM and Random Forest Algorithms" IEEE Xplore 2021.

[6] E. Ileberi Y. Sun and Z. Wang "Performance Evaluation of Machine Learning Methods for Credit Card Fraud Detection Using SMOTE and AdaBoost" IEEE vol. 9 no. 5 pp. 165286-165294 2021.

[7] Yakub K. Saheed, Moshood A. Hambali, Micheal O. Arowolo, YinusaA. Olasupo, "Applicat ion of GA Feature Select ion on Naive BayesRandom Forest and SVM for Credit Card Fraud Detection", DecisionAid Sciences and Application (DASA), International Conference on, pp. 1091-1097, 2020.

[8] M. S. Kumar V. Soundarya S. Kavitha E. Keerthika and E. Aswini "Credit Card Fraud Detection Using Random Forest Algorithm" IEEE 2019.

[9] VaishnaviNathDornadula and S Geetha "Credit Card Fraud Detection using Machine Learning Algorithms" Procedia Computer Science vol. 165 pp. 631-641 2019.

[10] I. Sadgali N. Sael and F. Benabbou "Fraud detection in credit card transaction using neural networks" Proceedings of the 4th International Conference on Smart City Applications (SCA ’19) 2019.

[11] Xuan S, Liu G, Li Z, Zheng L, Wang S, Jiang C (2018) Random forest for credit card fraud detection. In: ICNSC 2018—15th IEEE International conference on networking, sensing and control, pp 1–6

|e-ISSN: 2319-8753, p-ISSN: 2347-6710| www.ijirset.com | Impact Factor: 8.423| A Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 4, April 2023 ||

| DOI:10.15680/IJIRSET.2023.1204106 |

[12] Mishra A, Ghorpade C (2018) Credit card fraud detection on the skewed data using various classification and ensemble techniques. In: 2018 IEEE International students' conference on electrical, electronics and computer science, SCEECS 2018.

[13] G. L. S. Xuan "Random forest for credit card fraud detection" IEEE 15th International Conference on Networking Sensing and Control (ICNSC) 2018.

[14] I. Sohony R. Pratap and U. Nambiar "Ensemble learning for credit card fraud detection" Proceedings of the ACM India Joint International Conference on Data Science and Management of Data Association for Computing Machinery 2018.

[15] J. O. Awoyemi and A. O "Credit card fraud detection using machine learning techniques: A comparative analysis" IEEE 2017.

[16] Malini N, Pushpa M., "Analysis on credit card fraud identification techniques based on KNN and outlier detection", In: Proceedings of the 3rd IEEE international conference on advances in electrical and electronics, information, communication and bio-informatics, AEEICB 2017, pp 255–258, 2017.

INNO SPACE
SJIF Scientific Journal Impact Factor

Impact Factor: 8.423

doi cross ref

ISSN INTERNATIONAL STANDARD SERIAL NUMBER INDIA

निस्केयर NISCAIR

# INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

## IN SCIENCE | ENGINEERING | TECHNOLOGY

9940 572 462 6381 907 438 ijirset@gmail.com

www.ijirset.com

Scan to save the contact details