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 1, January 2023

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 8.118

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: 8.118| A Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 1, January 2023 ||

| DOI:10.15680/IJIRSET.2023.1201073 |

Review paper on Credit Card Fraud Prediction using Machine Learning Algorithm

Kumar Ujjawal[1], Prof. Suresh. S. Gawande[2]

M. Tech. Scholar, Department of Electronics and Communication, Bhabha Engineering Research Institute, Bhopal, India[1]

Guide, Department of Electronics and Communication, Bhabha Engineering Research Institute, Bhopal, India[2]

ABSTRACT: As the internet is growing, small and big companies around the world have made their business online. Internet has become more important than computer television in today's time. Today, the customer is more dependent on the internet. Whether shopping or bank. All the work is being done through internet. For which payment the customer uses a credit card. Credit card is a tool that has enhanced our facilities. From which we buy items now and pay for them letter. Credit card payment has become common in the world of e- commerce.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

Now a day's virtual companies and the internet are changing the scenario of traditional commerce. As the internet provide a global market, more flexibility and more competition in market, e-commerce value is increased. E-commerce also provides easier and wide range of innovation in the field of banking and payment. It plays a main role in global and competitive market. Is global market increasing people is diverting toward digital market instead of traditional market. As it gives facilities to customer, it has number of limitations also. Online payment is the vital thing for Ecommerce or Digital market. As universal market increasing the demand for Financial and Banking sector for payment is become standard in present days. Online payment gives you benefits to perform transaction from anywhere, anytime and anyplace with the help of smart devices like laptop, mobile, desktop, PDA, etc. The main objective behind electronic payment growth is it removes the limitations of traditional commerce. A user does not have to stand in long queue and personally visits the Bank to settle transaction. It gives many benefits like transactions are become faster that are done in few minutes, no need to visits banks and stand in queue. There are two ways to performed Electronic payment that are either online or offline. Online payment can identify as virtual payment. In online payment, it requires account holder name, PIN, card number, expiry date, etc. sensitive information. Offline payment can identify as physical payment. For offline payment, presence of card holder and PIN are required [1, 2].

For online payment frauds first thing is required is credit card number of some card holder. There is a many way to accomplish these scams. Some popular methods for online credit card frauds are phishing, identity theft, skimming, lost or stolen card use, card cloning, etc. Apart from these methods some another mechanism that allow credit card scams such as, malware or key loggers who can hack credit card details while online transaction, scanning devices are used to read tour credit card details [3, 4].

While online payment does not require signature or PIN number of your card, it makes process easier. Most of the websites are stealing card details and selling them to third party, number of the fraudsters are available on dark web so difficulty to trap. For offline payment frauds two mechanisms are most important that are ATM system and POS system. Fraudsters are creating clone card, card trapping, forged ATM or POS devices, identity theft, lost or stolen card, suspicious system, network or device [5].

In online transaction that stops consumer scam due to copy of payment card numbers, this creation includes at least one reliable card host. Payment card processing steps are as follow:

- Customer choose host, and place online order. While placing online order payment card number is not sended.
- Merchant allocates an OrderID for the placed order.
- Customer verifies payment details via host using private key, merchant also request for payment verification via host. Customer and merchant both verify payment with the same orderID.
- Host compares orderID and improve private key. The host combine it with the set of private key to fetch

| |e-ISSN: 2319-8753, p-ISSN: 2347-6710| www.ijirset.com | Impact Factor: 8.118| A Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 1, January 2023 ||

| DOI:10.15680/IJIRSET.2023.1201073 |

payment card number.

- After that host sends request for payment verification of card holder through payment network.
- When card holder sends back the response of payment verification, response sends back to the merchant.
- Merchant completes the order and Send for payment via the host.
- During the online transaction, all information are transfer through the SSL in an encrypted form, and when recipient receives it in decrypted form.

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

F. Itoo et al. [3], 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. [4],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

| |e-ISSN: 2319-8753, p-ISSN: 2347-6710| www.ijirset.com | Impact Factor: 8.118| A Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 1, January 2023 ||

| DOI:10.15680/IJIRSET.2023.1201073 |

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. [5],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), 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.

D. Tanouz et al. [6],credit card frauds are straightforward and inviting targets. The number of online payment options has increased as a result of e-commerce and numerous other websites. As fraud rates rose, researchers began employing a variety of machine learning techniques to identify and investigate online transaction fraud. The primary objective of this paper is to develop and design a novel method for detecting fraud in Streaming Transaction Data by analyzing customer transaction details and identifying behavioural patterns. where cardholders are grouped according to the amount of their transactions. Then, utilizing the sliding window strategy [1], aggregating the transactions made by cardholders belonging to various groups permits the respective extraction of the groups' behavioral patterns. Later, distinct classifiers are individually trained over the groups. The classifier with the highest rating score can then be chosen as one of the best ways to anticipate fraud. Consequently, a feedback mechanism to address concept drift follows. We utilized a European credit card fraud dataset in this paper.

S. Khatri et al. [7],in this paper, author presents statistics analysis of credit card growth in Indian market and different methods for credit card fraud detection, probability to produce FP (False Positive) and FN (False Negative) and legal and illegal transaction. Author proposed HMM method i.e. based on set of transaction probability. For that it does not require fraud signature only spending pattern of card holder is sufficient to detect fraud. Whenever credit card transaction occur system send it to fraud detection system to the bank for verify card's details. System does not bother about items it only checks amount and make cluster of data transaction. Later on, variance in amount is identifying for detect fraud. For implementing result author use MATLAB because it has in built HMM functionality. Using HMM method detection become easy and it removes complexity.

M. S. Kumar et al. [8], analyze HMM behavior with credit card transaction to find out transaction is genuine or not. If incoming credit card transaction is not accepted by HMM by high probability it means transaction is illegal, but with this system have to take care about genuine transaction. For maintain this spending pattern of card holder is analyzed. Pattern is sent to FDS of bank that identify transaction and generate alert automatically for bank. Author discussed problem with this technique detection and action occur in case of complain is registered by customer. Fraud is detected on basis of IP address. For this it requires cyber-crime experts and lots of man power. Author concludes that proposed system require certain updating of application in case of system requirements are change. For that some front end and back ends technologies are required for maintenance.

M. K. Dejan et al. [9], describe about need of credit card in today's payment system, different types of payment schemes of credit card and individual risk factor of card payment. For commit fraud key factor is credit cards details like its CVV number, expiry date, card type, etc. author used different types of method like HMM and data mining, etc. that analyze the spending pattern of user and their relationship between data to find out inconsistency in transaction. Here, proposed model identifies whether the current transaction is valid or invalid. If transaction is invalid then it terminates the current transaction otherwise continue with current transaction.

| |e-ISSN: 2319-8753, p-ISSN: 2347-6710| www.ijirset.com | Impact Factor: 8.118| A Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 1, January 2023 ||

| DOI:10.15680/IJIRSET.2023.1201073 |

VaishnaviNath et al. [10],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.
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

| |e-ISSN: 2319-8753, p-ISSN: 2347-6710| www.ijirset.com | Impact Factor: 8.118| A Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 1, January 2023 ||

| DOI:10.15680/IJIRSET.2023.1201073 |

(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

SVM are used in many applications like medical, military for classification purpose. SVM are employed for classification, regression or ranking function. SVM depends on statistical learning theory and structural risk minimization principal. SVM determines the location of decision boundaries called hyper plane for optimal separation of classes as described in figure 1. Margin maximization through creating largest distance between separating hyper plane and instances on either side are employed to minimize upper bound on expected generalization error. Classification accuracy of SVM not depends on dimension of classified entities. The data analysis in SVM is based on convex quadratic programming. It is expensive as quadratic programming methods need large matrix operations and time consuming numerical computations.

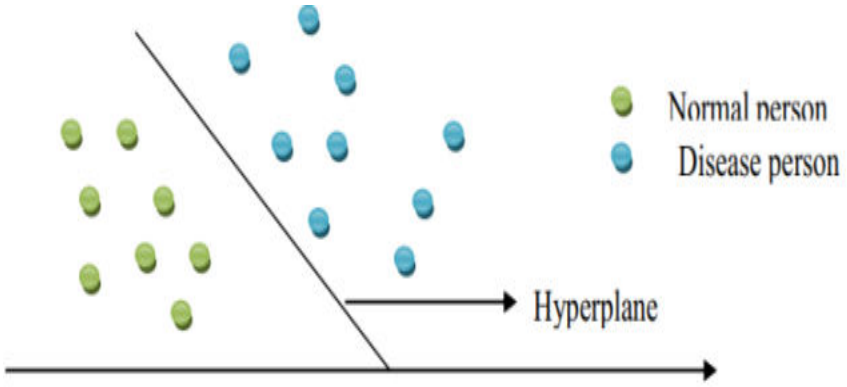

Fig. 1: Support Vector Classification

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 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

Today's many ways for credit card fraud detection. If we applied these algorithms for fraud detection in credit card system, then after the transaction of credit card we can be predicted easily the probability of fraud transactions and before anti-fraud transaction series. Reduce risks from bank transaction. Credit card fraud detection is very important

| |e-ISSN: 2319-8753, p-ISSN: 2347-6710| www.ijirset.com | Impact Factor: 8.118| A Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 1, January 2023 ||

| DOI:10.15680/IJIRSET.2023.1201073 |

topic for research. The percentage of fraud in this field is increasing. Fraudsters use new techniques to fraud. In our research, we have talked about fraud with customer, in addition, using the technique of data mining, how to remove the fraud is told. 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] 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.

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

[5] 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.

[6] D. Tanouz R. R. Subramanian and D. Eswar "Credit Card Fraud Detection Using Machine Learning" IEEE 2021.

[7] S. Khatri A. Arora and A. P. Agrawal "Supervised Machine Learning Algorithms for Credit Card Fraud Detection: A Comparison" IEEE 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] M. K. DejanVarmedja "Credit Card Fraud Detection - Machine Learning methods" IEEE Xplore 2019.

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

[11] 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.

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

[13] 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.

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

INNO SPACE
SJIF Scientific Journal Impact Factor
Impact Factor: 8.118

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