INTERNATIONAL JOURNAL FOR INNOVATIVE RESEARCH IN MULTIDISCIPLINARY FIELD ISSN: 2455-0620 Volume - 7, Issue - 9, Sept – 2021
Monthly, Peer-Reviewed, Refereed, Indexed Journal with IC Value: 86.87 Impact Factor: 6.719
Received Date: 13/09/2021 Acceptance Date: 26/09/2021 Publication Date: 30/09/2021

Credit Card Fraud Detection using Machine Learning Techniques

[1] Preet shah [2] Nivedita Sinha [3] Sejal Thakkar

[1, 2]Student, [3]Assistant Professor,
1, 2, 3Computer Engineering Department, IITE Indus University, Ahmedabad, India
Email – [1]preetshah986@gmail.com [2] nivedita.sinha2068@gmail.com [3]seju101ster@gmail.com

Abstract: *In the realm of money, as innovation developed, a new system of business making came into the picture. The credit card system is one of them. Banking and online shopping has now become the most normal activities among the majority. As innovation propels so does the danger partner with these activities. The convenience of this online exchange has now gotten more mainstream across the world. So, it is fundamental that we should be exceptionally cautious of the expanded Fraud exercises. It is essential that credit card organizations can recognize fraud credit card exchanges so clients are not charged for things that they didn't buy. Such issues can be handled with machine learning techniques. The sort of fraud doesn't continue as before for each situation, so this turns out to be exceptionally urgent in concocting the best calculation for the fraudulent exchange.*

Key Words: *credit card fraud, random forest, K-nearest neighbour, outlier decision method.*

1. INTRODUCTION:

Credit card fraud finding is while an exchange receipts steps to block whipped money, stocks, or conveniences achieved through an unlawful credit card business. Credit card fraud can happen together by the client or by another person. To stay away from happening such frauds, there are numerous strategies designed. On the off chance that such frauds occur, then, how to follow the abused exchanges is also explained. Several remarkable calculations are proposed to give security to the computerized information exchanges from unapproved access. Yet at the same time, there are a few disadvantages in either way. This paper deals with few techniques in recognition of credit card fraud. This paper presents the survey of techniques such as Random Forest and K-Nearest Neighbour, furthermore, predicts the best calculation to distinguish the fraudulent exchange dependent on a given situation**.**

2. LITERATURE REVIEW: Any bank that issues credit cards to its clients ought to have powerful measures set up to find fraudulent credit card exchanges. It is being seen that an assortment of approaches can be utilized to find fraudulent credit card exchanges. The distinctive existing methodologies used to find the fraudulent credit card exchanges are recognized and delegated as follows:

3. RANDOM FOREST:

This algorithm is a further developed form of the decision tree algorithm it utilizes a blend of decision trees to give the better outcome. Each sole decision tree draft for the different conditions will work on self-assertive informational collections and the decision trees. Each tree gives the chance of the fraud business and non-fraudulent also. Random decision forests and Random forests are the gathering learning methods for arrangement, forecast, and extra positions that capacity by building a monstrous volume of decision trees at practice time and yielding the class that is the method of the modules (categorization) or mean forecast (regression) of the different trees. Random decision forests exactly for decision trees nature of over fitting to their activity set. Use Case: Consider a situation where an exchange is made. Presently, an outline is made in transit the random forest in AI is utilized in fraud discovering algorithms is as displayed in the figure below:

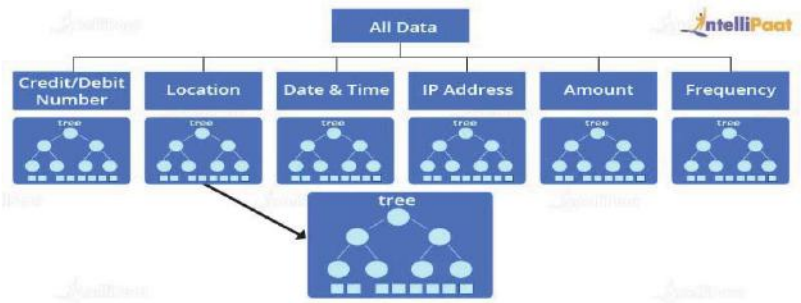

Figure 1: *Random Forest*

Two random forests that are RTRF and CRF which are diverse in their base classifiers are utilized. RTRF is utilized as a learning technique for grouping and relapse while CRF makes numerous CART trees and joins their anticipated outcome. The Dataset contains 30,000,000 person transactions from which just 82,000 settlements were set apart as false so the proportion of misrepresentation is 0.27% that may bring about the data imbalance issue. So, the RU technique is finished on valid settlements to manage information awkwardness issues. The CART-based random forest has an exactness of 96.77%, which is greatly improved when contrasted with the random tree-based random forest [7]. Technique for highlight expansion in the time measurement and container network is embraced to search some profound provisions on the foundation of the extended include. A bunch of the classifiers like Support Vector Machine, Random Forest, Neural Network, Capsule network, and CNN. The outcome shows that a case organization has a most noteworthy exactness of 99.21%. Impediment of container network is that utilization times are more [8]. Since the number of genuine exchanges is consistently higher than deceitful exchanges, the SMOTE procedure is utilized to balance the Dataset. The stacking classifier has higher precision than RF [19]. Algorithms like random forest, support vector machine and logistic regression are utilized. Assessment of the result is done on two performance measures: a region under the ROC curve (AUC) and normal accuracy (AP) and with the static and steady learning approach. To deal with the class unevenness issues SMOTE technique is utilized, which adjusts the dataset. At the point when measure with AUC scores random forest gives a superior outcome that is 91.48% in static learning additionally logistic regression gives 91.07% in the gradual learning approach. In the AP scores measure, the random forest gives 84.83% in static learning, and in gradual learning logistic regression gives 84.13% which is better compared to contrasted with others [9]. [10] Analyses some machine learning algorithms and announces that the random forest gives very great outcomes on the boundaries of precision, review, and exactness. Scikit-learn is utilized for the arrangement of data. Scikit-learn is an open-source library in python, it contains highlights like arrangement, regression, Clustering algorithms. The disarray matrix is shaped when the RF is applied to the pre-processed data. The performance is investigated dependent on the disarray matrix and gives accuracy between 90%-95% [11].

Table 1: Comparison of Random forest with different datasets

Reference	Methodology & Tools	Dataset	Merits	Demerits	Accuracy
[8]	NB, Random Forest	UCSD –FICO data mining contest 2009 dataset	It is independent of attribute values	Inconsistent behavior over a larger time range	The accuracy produced by RF is 83%
[11]	Thresholding Bayes minimum risk classifier	European card processing technology	RF-T performs better LR-MR performs better	LR-T performs worst RF-MR perform worst	RF-MR-A is consistently the best method
[15]	RF1: - RTRF,RF2: -CRF.	Dataset from a Chinese E-commerce company	RF gives better results when compared to decision tree	On large dataset problem such as data imbalanced may occur	RF II gives higher accuracy of 96.77% and precision of 89.46%
[18]	SVM, RF, NN, Capsule network, CNN.	Financial company of China	Capsule network achieve good performance	Capsule network consumes much time	Capsule network has the highest accuracy: 99.21% recall: 95.20%
[19]	Random forest, logistic regression, SVM, decision tree, KNN, NB, SMOTE	dataset by European cardholders	Supervised learning makes better prediction.	Supervised learning requires prior classification to anomalies	Accuracy by stacking classifier is more than that of Random Forest
[20]	RF, SVM, LR. SMOTE Python, standard scikit	European cardholder Dataset	Static learning approach gives better solution than incremental learning approach	Static learning approach is not a long-term solution	RF is better in static learning and LR is better in incremental learning
[21]	LR, RF, NB, multilayer perception.	European cardholders' data	Achieves high precision and recall	Requires balancing data	Good accuracy

K-Nearest Neighbour:

K-Nearest Neighbours algorithm utilizes Machine Learning alongside some supervised learning. We need to give prior data knows as the training data for this model to work [6]. This preparation information will arrange the directions into various gatherings dependent on the trait. This algorithm utilizes information mining, design acknowledgment, and interruption identification procedures. The critical advantage in utilizing this algorithm is this doesn't have any basic suppositions on the information distribution [7]. The Gaussian Mixture Model (GMM) on the other hand vigorously utilizes suppositions on the given information, in view of this the K-Nearest Neighbours algorithm is generally utilized, in real situations.

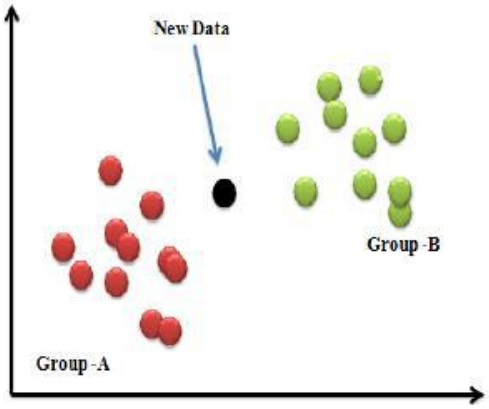

Figure 2: *K-Nearest Neighbour*

KNN builds the fraud recognition rate and lessening bogus alert rates. KNN works well in frameworks utilizing supervised learning techniques. In view of comparability measures, KNN saves every single open occurrence and orchestrates new cases. SVM is a strategy dependent on measurable learning that is reasonable for binary classification techniques where just two classes are required for example real and fraud information. Begging ensemble classifier dependent on a decision tree which is created to work on the robustness and accuracy of algorithms of machine learning utilized in regression and classification. KNN and Outlier detection methods increase the fraud detection rate and bogus alert rates. KNN works well in frameworks utilizing supervised learning techniques. KNN based fraud detection techniques need two gauges for example similitude or distance count between two events of information. In K-NN any of the settlements is figured for its nearest detail. On the off chance that it demonstrates to be fraudulent then the algorithm shows cheating. This strategy is fast with fewer off-base cautions and for that, both fraudulent and authentic models are taken care of to prepare the informational collections. KNN can identify fraud during the season of exchange. KNN distinguishes fraud from memory restriction. Outlier detection is ordered into two kinds for example supervised also, unsupervised. In supervised, specialists train the design to get knowledge and characterize huge contrasts. In unsupervised it doesn't need to be prepared with the lawful and illicit exchange.
K nearest neighbour is a sort of supervised learning utilized for classification and regression issues. It is utilized in tracking down a comparable example in a prior exchange of the cardholder. Some machine learning algorithms are utilized like NB, LR, and KNN. KNN has the best result of 97.69% precision for discovering fraud exchanges.

Table 2: Comparison of K-nearest neighbour with different datasets.

Methodology & Tools	Dataset	Merits	Demerits	Accuracy
SVM NB KNN Bagging ensemble classifier.	UCSD-FICO data corpus	A small change in data does not affect the hyper plane easy to implement	1)SVM takes very long training time 2) Sometime unable to make prediction 3) KNN is sensitive to noise dataset	SVM gives better accuracy than others. SVM -20% NB-15% KNN-10% BEC-3%
KNN, OD. Distance metrics	Downloaded from UCI website	1) There is no requirement of predictive model before classification. 2) UL is preferred	1)KNN cannot detect fraud at the time of transaction 2)Difficult to tutor the system	KNN gives better accuracy (72%) than outlier detection
Naïve Bayes (NB), k-nearest neighbor (KNN) and logistic regression. Hybrid technique	European cardholder dataset	If Credit card fraud detection datasets are available and balanced than it is easy to find fraud	Credit card fraud detection datasets are rarely available and highly imbalanced	NB – 97.92%, KNN – 97.69% LR – 54.86% (w.r.t accuracy, sensitivity, specificity, MCC)

4. CONCLUSION:

Credit card fraud detection has been a keen space of research for specialists for quite a long time and will be a charming space of exploration in the coming future. This happens significantly because of nonstop change of techniques in frauds. Credit card fraud these days has become extremely infamous influencing endless lives. Not just that, it is likewise very hard to perceive such cheating progressively. Therefore, it is important to assemble people's

confidence in utilizing cards for online repayments. Identifying fraudulent credit card exchanges is a significant prerequisite to control different credits card frauds. Organizing a powerful framework to test and distinguish fraudulent card transactions is one of the most fundamental capacities for doing cash transactions. The advantage provided by Random Forest and KNN techniques is the high accuracy both techniques provide. The random forest provides an accuracy of 90-95% while KNN has an accuracy of around 97%. As both are machine learning techniques the effectiveness will increase over time as more dataset is put into it.

REFERENCES:

1. N. Shirodkar, P. Mandrekar, R. S. Mandrekar, R. Sakhalkar, K. M. Chaman Kumar and S. Aswale, "Credit Card Fraud Detection Techniques – A Survey," 2020 International Conference on Emerging Trends in Information Technology and Engineering (ic-ETITE), 2020, pp. 1-7, doi: 10.1109/ic-ETITE47903.2020.112.
2. S P Maniraj , Aditya Saini , Shadab Ahmed , Swarna Deep Sarkar, 2019, Credit Card Fraud Detection using Machine Learning and Data Science, INTERNATIONAL JOURNAL OF ENGINEERING RESEARCH & TECHNOLOGY (IJERT) Volume 08, Issue 09 (September 2019),
3. M. R. Dileep, A. V. Navaneeth and M. Abhishek, "A Novel Approach for Credit Card Fraud Detection using Decision Tree and Random Forest Algorithms," 2021 Third International Conference on Intelligent Communication Technologies and Virtual Mobile Networks (ICICV), 2021, pp. 1025-1028, doi: 10.1109/ICICV50876.2021.9388431.
4. *Hisar, T. and Hce Sonepat. "Survey Paper on Credit Card Fraud Detection." (2014).*
5. P. Jayant , Vaishali, 2014, Survey on Credit Card Fraud Detection Techniques, INTERNATIONAL JOURNAL OF ENGINEERING RESEARCH & TECHNOLOGY (IJERT) Volume 03, Issue 03 (March 2014)
6. N. Malini and M. Pushpa, "Analysis on credit card fraud identification techniques based on KNN and outlier detection," 2017 Third International Conference on Advances in Electrical, Electronics, Information, Communication and Bio-Informatics (AEEICB), 2017, pp. 255-258, doi: 10.1109/AEEICB.2017.7972424.
7. *Xuan, Shiyang et al. "Random forest for credit card fraud detection." 2018 IEEE 15th International Conference on Networking, Sensing and Control (ICNSC) (2018): 1-6.*
8. *Fu, Kang et al. "Credit Card Fraud Detection Using Convolutional Neural Networks." ICONIP (2016).*
9. M. Puh and L. Brkić, "Detecting Credit Card Fraud Using Selected Machine Learning Algorithms," 2019 42nd International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO), 2019, pp. 1250-1255, doi: 10.23919/MIPRO.2019.8757212.
10. Machine Learning, Image Processing, Network Security and Data Sciences: Second International Conference, MIND 2020, Silchar, India, July 30 - 31, 2020, Proceedings, Part II [1st ed.] 9789811563171, 9789811563188
11. Akanksha Bansal and Hitendra Garg 2021 *IOP Conf. Ser.: Mater. Sci. Eng.* 1116 012181