www.ijsrnsc.org

Available online at www.ijsrnsc.org

IJSRNSC

Volume-11, Issue-2, April 2023
Research Paper

Int. J. Sc. Res. in Network Security and Communication

E-ISSN:2321-3256

A Review of Credit Card Fraud Detection Using Machine Learning

N.S. Shroff[1*], A.S. Vaishnav[2]

[1]Department of Computer Engineering, Government Engineering College, Gandhinagar, India
[2]Department of Computer Engineering, Government Polytechnic, Gandhinagar, India

**Corresponding Author: amitvaishnav1112@gmail.com*

Received: 21/Feb/2023, Accepted: 20/Mar/2023, Published: 30/Apr/2023

Abstract - Nowadays fraud has been increasing due to the establishment of online payment mode on different E-commerce platform.A credit card is a form of payment that lets you buy goods or services on credit from an issuer, usually a bank. You can make purchases up to a specified limit and then pay them off over time either in full or with minimum payments.There are several types of security features including fraud protection, verified by visa and master card secure code, address verification systems, and biometric authentication. Additionally, some cards offer the additional security feature of a chip and pin system which requires that the cardholder enter a secret code to make purchases.Still fraud has been executed using this card. In this fraud, banks, merchants, and organisations are losing billions of dollars. According to one survey, the prevalence of credit card fraud is rising by 12.5% a year. It is crucial to identify fraud using secure and effective methods.

Nowadays, hybrid algorithms and artificial neural networks are used to detect fraud since they perform better than other methods. We will use dataset variables like "duration," "amount of transaction," and "V1 to V28" as derived parameters for this. We will build a model that will separate out fraudulent transactions from other transactions using machine learning techniques or algorithms.

Keywords—Machine learning, Hybrid algorithms, Fraud, Fraudulent and Credit card

I. INTRODUCTION

Fraud is defined as stealing something that belongs to someone else without the user's knowledge. There are several types of fraud, including online fraud, offline fraud, and resource fraud. Among all this credit fraud that is categorised under "online frauds," it is nowadays becoming a major problem. Credit card fraud detection is a very tough and difficult process to detect since fraudsters always attempt to pass off every fraudulent transaction as legitimate.

Credit card fraud can be done by using the information of any person to perform different transactions.Thereare different ways to commit credit card fraud by stealing the customer's information.

- Directly from the customer
- Through a payment gateway, such as PayPal or Stripe.
- Through a third-party credit card processor, such as Square or Adyen.
- Through a merchant account provider, such as First Data or Worldpay.
- Through a mobile wallet, such as Apple Pay or Google Pay.
- Through a credit card reader, such as a Point of Sale (POS) system.

We can control this type of fraud by making people aware of it and reducing the financial loss that different organisations have to bear.

Global losses due to credit card fraud were around 1,68,260 crores of Indian rupees in 2017 and are expected to steadily rise by 2020, when they are projected to reach 2,28,775 crores of Indian rupees. Over 2.9 crore people in India currently use credit cards.But since the development of technology, Cybercrime is done from different places in the world, and in India, Jamtara has been the hub of cybercrime for the past five years. In 2019, 107 Jamtara citizens were detained on suspicion of cybercrime[4]. According to Reserve Bank of India (RBI) data, fraudsters stole 615.39 crore in more than 1.17 lakh cases of credit and debit card theft over a ten-year period (April 2009 to September 2019) [15].

There are many challenges faced while designing and implementing credit fraud detection techniques, such as the unavailability of datasets due to security reasons, imbalanced data, operational efficiency, and incorrect flagging. There are many algorithms to control fraud such as, random forest [1], optimisedlightGBM [4], k nearest neighbor, neural networks [5], logistic regression, decision trees [7], support vector machines [8], and naive bayes [9]. This study evaluates the performance of the various algorithms based on how well they were able to determine that the transaction was legal or illegal. Performance indicators like accuracy, specificity, and precision are used to make the comparison. In comparison to previous approaches, the K Nearest Neighbor algorithm exhibited great accuracy and precision.

Our objective behind this study is to develop better algorithms with respect to all existing algorithms to classify credit card transactions. Some current algorithms classify incorrect transactions as opposed to original transactions.

II. RELATED WORK

To find credit card fraud, numerous supervised and unsupervised machine learning methods are applied. But supervised algorithms are mainly used due to the highly imbalanced dataset in credit card fraud detection. We have studied different research papers and described below the work done by different authors.

In [1], for managing the imbalance dataset and classifying transactions as legitimate or fraudulent, logistic regression, random forest, and Naive Bayes algortithms are implemented. Among all above algorithms the random forest classifier gives 96.7741% accuracy, 100% precision and 91.1111% recall.

In [2], the first approach, fraud is detected by building a tree based on a user's activity. In the second approach, to identify a victim, a forest is built based on the user's activity. The result shows that there are common techniques to detect credit card fraud with respect to their degree of precision.

In [3], the author used different methods, such as the homogeneous and heterogeneous poissonprocesses,for credit card fraud detection to calculate the probability of fraud based on the use of different intensity parametric functions. He also used ensemble methods for classifications, and then both were compared.

In [4] Optimized Light Gradient Boosting Machine (OLightGBM) and a Bayesian-based hyperparameter optimization algorithm that is intelligently integrated with LightGBM. Experiments were performed on the credit card dataset, and it was found that the new approach gives approximately 97% accuracy.

In [5] for the classification of fraudulent transactions from the dataset, many machine learning and deep learning techniques are implemented, such as multiple linear regression, logistic regression, K nearest neighbor, naive bayes, random forest, and neural network, among all, K nearest neighbour performs best with an f1-score of 0.75 and precision of 0.78.

In [6], two techniques, random forest and adaboost algorithms, are implemented to extract fraudulent transactions. Random forest will overtake the adaboost algorithm and become a better performer as compared to the adaboost algorithm.
In [7], the algorithm will evaluate historical customer transactions and extract their behavioural pattern; then, unique groups of transaction information are formed based on this behavioural pattern using the sliding window protocol.

In [8], for converting a balanced dataset from an unbalanced dataset, the SMOTE technique was used for oversampling. Moreover, training and test data were collected as part of the feature selection process. Various algorithms are used, such as logistic regression, random forest, naive bayes, ANN, and multilayer perceptrons; among all of them, random forest gives the best performance. Different parametric measures such as recall, accuracy, and precision were used to measure the performance of all algorithms.

In [9], the study is based on four major fraud incidents that occur in the real world. Different algorithms are implemented to solve each fraud incident, and an optimised approach among all algorithms is selected as the final result. Additionally, predictive analysis is also used to identify fraudulent transactions among all transactions.
In [10], different supervised and unsupervised learning algorithms were used for the classification of fraudulent transactions, and supervised learning algorithms generated better results as compared to unsupervised algorithms.

In [11] implemented two phases for fraud detection classification. In the first phase, the local outlier factor (LoF), which specifies the numerous characteristics that must be used, In the second phase, isolation forest algorithms isolate transactions with a high incidence of anomaly detection.

In [12], a variety of supervised learning algorithms are used, and an ensemble method is used to implement a stacking classifier.The stacking classifier was compared with various algorithms, and it gave the highest accuracy of 95.27% among all algorithms.

In [13], there are two different types of random forest algorithms that were implemented based on different classifiers, among which the first was a normal random forest algorithm and the second was CART-based random forest II, which gives better results with the highest accuracy of 96.77%.

III. DATASET

In credit card fraud detection, one of the major challenges is a highly imbalanced dataset, which is balanced by applying undersampling, oversampling, or both methods.
In [1,5,6,7,8,10,12], European dataset of credit card holders from September 2013 for two days were used as a dataset, which is available on Kaggle. In [2,4] public sample credit card transaction dataset was used. In [3], the dataset contains 95 662 transactions from November 15, 2018 to February 13, 2019. [9] used a Brazilian bank dataset with 345,735 instances and an imbalance ratio of 25.7. In [11], a dataset of 280,000 transactions, of which 28,490 are fraudulent transactions, was used for classification.

IV. COMPARISON

Table 1 – Comparative study of credit card fraud detection methodologies

Ref. No.	Methodology	Findings	Results
[1]	➢ Decision tree ➢ Random Forest ➢ Logistic Regression ➢ Naive Bayes	Combining all supervised algorithm to get better performance	Random forest classifier performs Accuracy - 96.7741%
[2]	➢ Clustering technique ➢ Gaussian Mixture Technique ➢ Bayesian Network Technique	Credit card number,location,datetime,IPaddress,Amount and frequency are considered as parameters	Three Technique are used to find fraudulent transactions but fraud occurrences can be possible through other intermediate channels
[3]	➢ Poisson process ➢ various intensity functions ➢ Machine learning. ➢ Gradient boosters ➢ LightGBM ➢ XGBoost ➢ CatBoost	➢ Poisson Process ➢ Ensembles ➢ Computation Process	Poisson process models performs better compare to gradient boosting models shown through ROC AUC graph
[4]	➢ Optimized LightGBM ➢ LightGBM ➢ Bayesian-based hyperparameter	Precision Recall and ROC AUC curve are shown as proposed approach	Optimized LightGBM give best performance Accuracy - 97%
[5]	➢ Multiple Linear Regression ➢ Logistics Regression ➢ K Nearest neighbor ➢ Naive Bayes ➢ Random forest ➢ Neural network	Use of different machine learning and deep learning algorithms	K Nearest neighbor F1-Score : 0.75 Precision : 0.78
[6]	➢ Random Forest algorithm ➢ Adaboost algorithm	Different performance metrics are used to check efficiency of algorithm	Random Forest algorithm give better performance
[7]	➢ Local outlier factor ➢ Isolation forest ➢ Logistics Regression ➢ Decision tree ➢ Random forest	MCC and SMOTE used to handle Imbalance dataset	Logistics Regression has highest accuracy
[8]	➢ Logistic Regression ➢ Naive Bayes ➢ Random Forest ➢ Multilayer Perceptron ➢ ANN	SMOTE and Over sampling used to handle Imbalance dataset	Random forest classifier performs Best Accuracy - 99.93%
[9]	➢ Support Vector	Models that considers various parameters	Logistics Regression has better

	Machine ➢ Naive Bayes ➢ K-Nearest Neighbor ➢ Logistic ➢ Regression.	such response code,web address and transaction amount	accuracy Accuracy - 74.00%
[10]	➢ Supervised ➢ Unsupervised ➢ Ensemble	Difficult to handle dataset using Unsupervised algorithms	Ensemble classifier performs Best Accuracy - 99.00%
[11]	➢ Genetic algorithm ➢ Decision tree ➢ Algorithm for Anomaly Detection ➢ Local Outlier Factor (LOF) ➢ Isolation Forest Algorithm (IFA)	Analyze Parameters such as Time, Transaction Limit,Country, Class,Merchantvendor,Amount Avg transaction, Issuing bank and Location	Reduce the number of parameters by applying various dimensionality reduction techniques.
[12]	➢ Stacking classifier ➢ Linear Regression ➢ Decision tree ➢ K Nearest neighbor ➢ Naive Bayes ➢ Random forest	Difficult to handle dataset using Unsupervised algorithms	Stacking classifier performs best Accuracy - 95.27%
[13]	➢ Random forest I Random tree-based ➢ Random forest II CART-based	Three Experiments are carried out to compare both algorithms	Random forest II CART-based has highest accuracy Accuracy - 96.77%

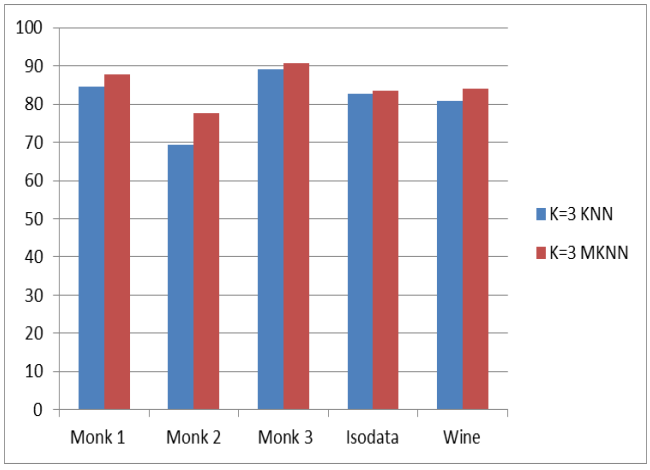
V. GAP IDENTIFIED

We studied all research papers and found that the credit card dataset is highly imbalanced, as well as that there are different algorithms implemented on this dataset. The table below shows the accuracy of all algorithms.

Table 2 - Comparison of different algorithms with their accuracy on credit card dataset (Kaggle)[5]

Sr. No.	Algorithms	Precision	Recall	F1-score
1	Random Forest	0.69	0.75	0.71
2	Multiple Linear Regression	0.75	0.79	0.77
3	**K Nearest Neighbor (KNN)**	**0.75**	**0.81**	**0.78**
4	Neural Network	0.75	0.81	0.78
5	Logistics Regression	0.66	0.76	0.70
6	Naive Bayes	0.02	0.86	0.05

From Table 1, it is clearly identified that KNN is the best classifier for credit card fraud detection, but as per [14], it is also found that Modified KNN (MKNN) gives better accuracy compared to KNN when K = 3, K = 5, and K = 7. Here, MKNN and KNN perform classification on different datasets shown in the bar chart in Figure 1.



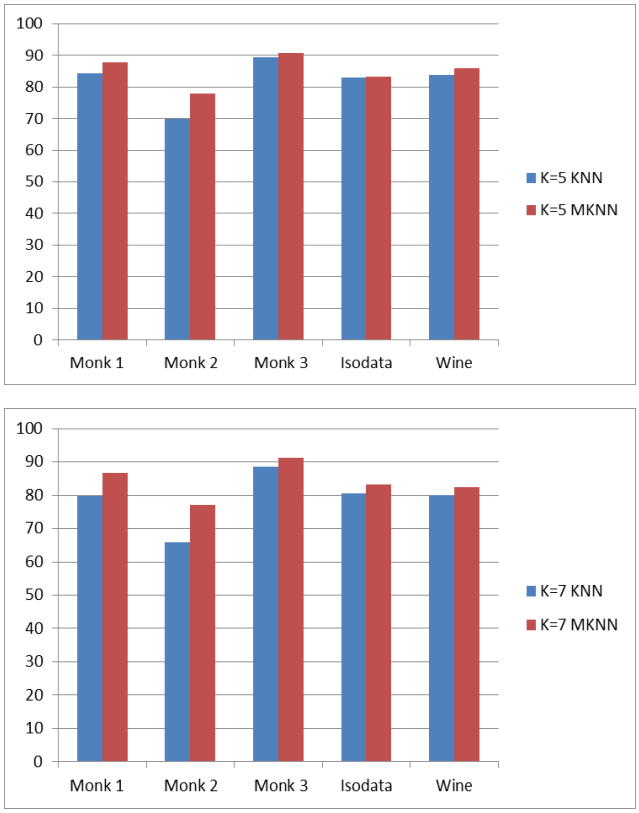


Figure 1 - Comparison of KNN and MKNN on different datasets[14]

VI. PROPOSED SYSTEM

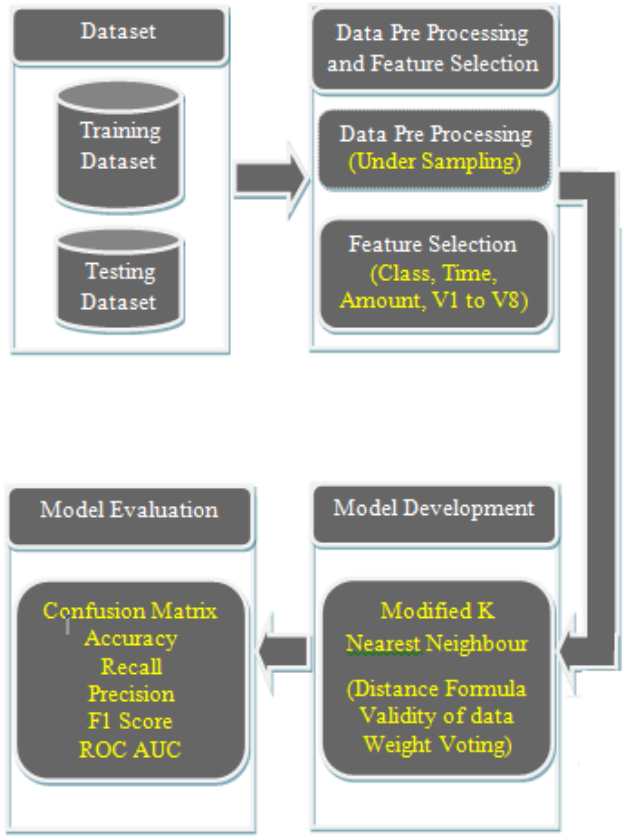


Figure 2 - Proposed system using MKNN algorithm

Based on the gap identified, we conclude that MKNN should be implemented on credit card fraud detection for better classification of fraudulent transactions across all transactions. We designed a proposed system, as shown in Figure 2, in which the following steps are used to implement a new approach algorithm.

In the first step, the dataset is divided into training(70%) and test(30%) data.

In the second step, data preprocessing is carried out using an undersampling technique for equality among different types of transactions in the dataset.

In the third step, the model will be developed using our new approach and modified KNN algorithm.

In the fourth step of the model, evaluation will be done using different performance metrics.

In MKNN, two extra calculations—validity of data and weight voting—are added to the existing KNN algorithms, which in turn gives a better classification of transactions.

VII. CONCLUSION

Our study revealed that while there are numerous algorithms for detecting credit card fraud none of them can accurately distinguish between fake and genuine transactions in real life scenario, but KNN algorithms perform best among all supervised algorithms in most of cases. MKNN which modified version of KNN gives higher accuracy on other datasets. KNN algorithm gives an average accuracy of 82.042% and the MKNN algorithm gives an average accuracy of 85.112% when K is set to 5 [14]. So we try to implement a modified KNN algorithm on a credit card dataset to classify between fraudulent and legitimate transactions by considering different values of K, and we try to improve performance by measuring their efficiency using different performance metrics parameters.

VIII. FUTURE SCOPE

One of our major challenges is the dataset, which is highly imbalanced. So we can use better techniques to convert it into a balanced dataset that can affect the accuracy of any algorithm. Moreover, we will implement MKNN for the classification of transactions with better accuracy.

REFERENCES

[1] D. Tanouz, R. R. Subramanian, D. Eswar, G. V. P. Reddy, A. R. Kumar and C. V. N. M. Praneeth, "Credit Card Fraud Detection Using Machine Learning," 5th International Conference on Intelligent Computing and Control Systems (ICICCS), pp. 967-972, 2021.

[2] M. R. Dileep, A. V. Navaneeth and M. Abhishek, "A Novel Approach for Credit Card Fraud Detection using Decision Tree and Random Forest Algorithms," Third International Conference on Intelligent Communication Technologies and Virtual Mobile Networks (ICICV),pp. 1025-1028, 2021.

[3] AnastasiiaIzotova, Adel Valiullin, “Comparison of Poisson

process and machine learning algorithms approach for credit card fraud detection" Procedia Computer Science,Volume 186,Pages 721-726,2021

[4] A. A. Taha and S. J. Malebary, "An Intelligent Approach to Credit Card Fraud Detection Using an Optimized Light Gradient Boosting Machine," in IEEE Access, vol. 8, pp. 25579-25587, 2020.

[5] M. Azhan and S. Meraj, "Credit Card Fraud Detection using Machine Learning and Deep Learning Techniques," 3rd International Conference on Intelligent Sustainable Systems (ICISS), pp. 514-518, 2020.

[6] R. Sailusha, V. Gnaneswar, R. Ramesh and G. R. Rao, "Credit Card Fraud Detection Using Machine Learning," 4th International Conference on Intelligent Computing and Control Systems (ICICCS), pp. 1264-1270, 2020.

[7] VaishnaviNathDornadula, S Geetha,"Credit Card Fraud Detection using Machine Learning Algorithms,Procedia Computer Science",Volume 165,Pages 631-641,2019.

[8] D. Varmedja, M. Karanovic, S. Sladojevic, M. Arsenovic and A. Anderla, "Credit Card Fraud Detection - Machine Learning methods," 18th International Symposium INFOTEH-JAHORINA (INFOTEH),pp. 1-5, 2018.

[9] A. Thennakoon, C. Bhagyani, S. Premadasa, S. Mihiranga and N. Kuruwitaarachchi, "Real-time Credit Card Fraud Detection Using Machine Learning,"9th International Conference on Cloud Computing, Data Science & Engineering (Confluence), pp. 488-493, 2019.

[10] S. Mittal and S. Tyagi, "Performance Evaluation of Machine Learning Algorithms for Credit Card Fraud Detection," 9th International Conference on Cloud Computing, Data Science &Engineering (Confluence), pp. 320-324, 2019.

[11] V. CeronmaniSharmila, K. K. R., S. R., S. D. and H. R., "Credit Card Fraud Detection Using Anomaly Techniques," 1st International Conference on Innovations in Information and Communication Technology (ICIICT),pp. 1-6, 2019.

[12] S. Dhankhad, E. Mohammed and B. Far, "Supervised Machine Learning Algorithms for Credit Card Fraudulent Transaction Detection: A Comparative Study," IEEE International Conference on Information Reuse and Integration (IRI), pp. 122-125, 2018.

[13] S. Xuan, G. Liu, Z. Li, L. Zheng, S. Wang and C. Jiang, "Random forest for credit card fraud detection," IEEE 15th International Conference on Networking, Sensing and Control (ICNSC),pp. 1-6, 2018.

[14] Okfalisa, &Gazalba, Ikbal&Mustakim, Mustakim& Reza, Nurul.Comparative analysis of k-nearest neighbor and modified k-nearest neighbor algorithm for data classification. pp. 294-298, 2017.

AUTHORS PROFILE

Namrata Shroff, received the B.E. degree in Computer Engineering from North Maharashtra University, India, and the M.E. degree in Computer Engineering from Gujarat Technological University, India. She is currently pursuing the Ph.D. degree with the Gujarat Technological University, India. Her research Interests include Data Analysis, Text Mining, Natural Language Processing and Pattern Recognition.

Amit S. Vaishnav, received the B.E. degree in Computer Engineering from Gujarat Technological University, India. He is currently pursuing the M. E. degree with the Gujarat Technological University, India. His research interests include Data Mining, Data Analysis, Machine Learning, Natural Language Processing and Deep Learning.

