



## SURVEY ON CREDIT CARD FRAUD DETECTION USING MACHINE LEARNING TECHNIQUES

NAGARAJU T A

*Assistant Professor*

Dept.of Electronics and Communication Engineering  
Government Engineering College  
Ramanagara, India

MANJUNATH H R

*Assistant Professor*

Dept.of Information Science and Engineering  
Alva's Institute Of Engineering and Technology  
Mangaluru,India

**Abstract**— Today, the Internet is an important factor in our lives. Due to the widespread use of the Internet, online shopping varies from day to day. Credit Card is an easy way to shop online and pay off debts. Due to this sudden pandemic situation and developed technology people prefer the online payment for tiny transactions. Therefore, Credit Card becomes a popular and convenient way to trade money online and is growing very fast. People in the developed countries like USA about 7 in 10 are having at least one credit card and 9 out of 10 transactions are done in online mode. Which gave a lot of possibility to the credit card fraudsters, and to stop/detect those frauds, now it is very essential to develop the model to prevent such unwanted frauds. In this regard a machine learning model is proposed a to detect the credit card frauds.

**Keywords**— *Credit Card, Fraud detection, Deep leaning, Machine learning*

### Introduction

Recently, there is a growing usage of the credit cards due to the present pandemic situation in the world even for the tiny payments. People often choose e-commerce and internet services to improve productivity and reduce time-consuming jobs. Online transactions are becoming increasingly popular. As a result, fraudulent activities have become quite common in various industries around the world, particularly in the financial sector. In financial institutions, credit card fraud is considered to be the most problematic form of fraud and there is a need to develop prevention tools as soon as possible. Figure 1 shows a summary of the total credit card fraud value and cents lost by \$ 100 from 2013 to 2027 worldwide [2]. The total amount of credit card fraud has grown significantly, from about \$ 13.7 billion in 2013 to about \$ 38.5 billion by 2027 worldwide. By 2021, credit card fraud as a whole is projected to lose 32.4 billion, up 7.7 cents to \$ 100 in total. In order to dramatically reduce the effects of credit card fraud, fraud detection methods need to be investigated for strict management.

There are two possibilities to stop this fraud that is prevention and detection mechanism. In the prevention there certain mechanisms to prevent the fraud, sometimes the prevention mechanisms won't work. In that case we will go for the detection mechanism, which is actually used after the fraud is happed. Using the detection mechanisms we will try to detect the fraud. There are several approaches for the detection of the fraud transactions among which, machine learning models are believed to be more accurate. One of the most successful techniques for credit card fraud detection is Machine Learning. ML mainly uses classification and regression algorithms for fraud detection.



**Figure 1. Card fraud Worldwide [2].**

Credit card fraud detection identifies suspicious transactions, events, and behaviours for further investigation. Each and every operation generates hundreds of data points that are evaluated for signs of fraud derived from past data. Modern machine learning powered fraud detection systems consider the tiniest changes in a customer's behaviour patterns in milliseconds with a high degree of precision.



**Source: Nilson Report [2]**

These machine learning techniques are basically divided into two types, supervised and unsupervised learning algorithms. The Supervised learning algorithms are trained in prior with the training data set and later tested on the test data set. Whereas the unsupervised algorithms are untrained algorithms which are tested on the dataset directly, the model will be learned with the experience. For the credit card fraud detection the supervised models are trained with the labelled transactions. And the unsupervised algorithms use the peer group analysis that groups customers according to their profile, identifies fraud based on customers spending habit. Some more approaches which are currently used are neural networks, decision tree, Naive Bayes, Support Vector Machines and K-Nearest Neighbours.

An important subfield of machine learning is Deep learning, which functions similar to human brain commonly used for data processing and other activities like decision making. It includes various models like NN and convolution neural networks. The neural network model is used for transaction and data prediction.

### Credit Card Fraud Detection Challenges

In this Section the challenges for the credit card fraud detection are identified

**Lack of Data** Lack of data can be considered in two contexts: lack of literature on the topic and lack of training/test data (public credit card transactions databases). The latter is a problem for the scientist and not so much for the industry, as credit card processing houses have vast amounts of data



**Scalability** It is a technical problem often ignored in the literature. One must strive to design robust and scalable systems to sustain a continual, large stream of transactions.

**Feature Engineering** It is a classic topic in data mining and is particularly important in credit card fraud detection. There is a limited set of features at disposal and little information from which to create a card model.

**Concept Drift:** Credit card fraud patterns change over time as the market and technology change, and both fraudsters and card processors adapt to the changes. This changes the underlying patterns and data and is referred to as “concept drift” [2].

**Learning Model Selection:** There are many different Machine learning models can address this problem. Each of them presents an optimization problem with many hyper parameters to tune.

## Literature SURVEY

R.Patidar and L.Sharma et al., [3] proposed ANN with genetic algorithm to detect credit card fraud. Original and fraudulent transactions are compared using NN, if there is large difference then it is considered as illegal transaction and if there is a small difference then it is considered as legal transaction. For improving the efficiency and selecting a parameter for network the feed forward back propagation was used. Hence with genetic algorithm and ANN A credit card fraud detection can be accomplished.

M. Zareapoor and P. Shamsolmoali et al.,[4] proposed bagging ensemble for fraud detection. UCSD-FICO dataset has been used for this experiment which has both legal and fraud transactions. Bagging ensemble based on decision tree algorithm provides the good system performance(fraud catching rate, false alarm rate, balanced classification rate and Matthews correlation coefficient).

H.Tran and K.P.Tran et al.,[5], used anomaly detection techniques for credit card fraud detection based on the reasons that the detection of fraud must be very flexible in order to track the continuous evolution of fraud over time and the occurrence of unknown anomalies. And proposed anomaly detection technique and One class support vector machine and T2 control chart methods. A real life data set of e-commerce transaction from Europe is used in this experiment. In which 400 transactions were considered for experiment in which equal number of original and fraud transactions were present. Then the experiment is done for both OCVM and T2 control chart and found that OCVM works better than T2 .

K.Seeja and M.Zareapoor in [5] used a matching algorithm and Apriori algorithm. Matching algorithm helps in comparing each customer patterns and forming fraudulent and original patterns. And then Apriori algorithm is applied for obtaining frequent item set for original and fraudulent patterns. Hence original patterns and fraudulent patterns for each customer is obtained.it makes easier for fraud detection by comparing the new transaction to both the patterns and know whether it is fraudulent transaction or original transaction.

A.Roy, J.Sun and R.Mahoney in [6], used six classifiers with a dataset before and after the pre-processing phase; The results show a significant improvement when they use the Under sampling technique with the dataset. The dataset used with this paper consist of 284,807 transactions; only 492 of them are fraud transactions. After using the Random Under sampling technique, they changed the ratio to 1:1, which means the number of fraud transactions is the same as the number of legal transactions. They used precision and recall evaluating the classifiers with both datasets; they found out that using the Under sampled dataset increase the precision for all the classifiers significantly.

In [7], Shiyang Xuan and Guanjun Liu et al., the performance of random forest models has been examined. A real-life B2C dataset on credit card transactions is used in the experiment.



In their research they have considered 3 months online transactions. Over that data set they have conducted an experiment using RF model, found the legal transactions and fraud transactions successfully. With this experiment they concluded that the RF model works nicely on the small data sets and it has a problem with an imbalanced data.

Dejan Varmedja et.al.,[8] proposed the various machine learning algorithms and analysed them concerning to credit card fraud detection methods. The various methods of machine learning are Logistic Regression, Naïve Bayes, Random Forest, Multilayer Perceptron. Here for multilayer perceptron is used which consist of 4 hidden layers and relu activation functioned is used that is to avoid negative values and optimizer used is Adam for its best performance As a result for Logistic regression the accuracy score is 97.46% with the data set containing 56962 samples in which 98 fraud transactions. For the same dataset Naïve Bayes and Random Forest, accuracy score is 99.23% and 99.96% respectively. At last for ANN it was 99.93% of accuracy as we can observe that random forest gives the best result in case of credit card fraud detection.

Rimpal R. Popat and Jayesh Chaudhary et al.,[9] they made a survey on credit card fraud detection, considering the major areas of credit card fraud detection that are bank fraud, corporate fraud, Insurance fraud. They had focused on the techniques which are Regression, classification, Logistic regression, Support vector machine, Neural network, Artificial Immune system, K-nearest Neighbor, Naïve Bayes, Genetic Algorithm, Data mining, Decision Tree, Fuzzy logic-based system, etc. In which, they have explained six data mining approaches as theoretical background that are classification, clustering, prediction, outlier detection.

Kuldeep Randhawa et.al, [10] they used twelve machine learning algorithms for credit card fraud detection in which their range standard from a neural network to deep learning. They are tracing the performance of benchmark and real-world datasets. As there related study explains about single and hybrid models. For both the parameters (Benchmark and real-world datasets) they had given the results using there twelve selected algorithms that are Naïve Bayes, Random Forest, Decision Tree, Gradient Boosted Tree, Decision Stump, Random Tree, Neural Network, Linear Regression, Deep Learning, Logistic Regression, SVM, Multilayer Perceptron. When experimented with real-world data the accuracy rate was above 90% even with 30% noise in the dataset.

Changjun Jiang et.al, [11] proposed a novel fraud detection method that has four stages they first utilize the historical transaction data to divide them into groups to form clusters of transactions having the same behaviour then thus they came up with a sliding window strategy to aggregate transactions. This algorithm is used to characterize the behavioural pattern of a cardholder then after aggregation, we use the new window formed the feature extraction is done. At last, the classification takes place and classifies behavioural patterns and assignments. As a result, their method of Logistic Regression with raw data (RawLR), Random Forest with aggregation data (AggRF), and Random Forest and feedback technique with aggregation data (AggRF +FB) are the best method with 80% accuracy as compared to other methods.

Hassan Najadat, Ayah Abu Aqouleh, et at.,[12] the Dataset used is IEEE-CIS Fraud Detection resulting in 433 features in the dataset and 590540 instances. However, there are 378 features contain a lot of null values and thus they were ignored. The dataset is highly imbalanced such that there are 569875 transactions that are legitimate transactions, while 20663 fraud transactions which means only 3.626% of the transactions were fraud transactions and this leads the models to predict only legitimate transactions while acting



poorly when trying to predict fraud transactions. Thus it is important for dataset classes to be balanced when training the models. To solve this problem Random Under Sampling, Random Over Sampling, SMOTE techniques applied to the dataset.

Machine Learning models are widely used in detecting credit card fraud and have proved their efficiency in getting high scores. Deep learning is a subset of machine learning technique that teaches computers to perform tasks which are natural to the human. A computer model learns to perform classification tasks directly from image, text or sound where it builds features automatically based on training data. Deep Learning models including neural networks, convolutional neural networks, Long short-term memory(LSTM) and Gated Recurrent Unit (GRU) provide state of the art results in various classification and prediction tasks

### Conclusion

Credit card fraud detection is one of the potential and interesting research areas for the young researchers as its usage and exploitation is increasing day by day especially in this pandemic situation. There are different approaches in use for the credit card fraud detection. Among which we found machine learning model as the most effective. With the present survey it is came to know that the machine learning models promise high accuracy and detection rate which is deficient with other approaches. In the present research a serious attempt is made using Machine learning and Deep learning models to detect the credit card fraud.

### References

- [1] Kubat, M.;Widmer, G. Adapting to drift in continuous domains. In Machine Learning: ECML-95, Proceedings of the 8th European Conference on Machine Learning Heraclion, Crete, Greece, 25–27 April 1995; Springer: Berlin/Heidelberg, Germany, 1995; pp. 307–310.
- [2] NILSON REPORT News and Statistics for Card and Moblie payemnts 2020.
- [3] R. Patidar, L. Sharma et al., “Credit card fraud detection using neural network.” International Journal of Soft Computing and Engineering (IJSCE), vol. 1, no. 32-38, 2011.
- [4] M. Zareapoor, P. Shamsolmoali et al., “Application of credit card fraud detection: Based on bagging ensemble classifier,” Procedia computer science, vol. 48, no. 2015, pp. 679–685, 2015.
- [5] P. H. Tran, K. P. Tran, T. T. Huong, C. Heuchenne, P. HienTran, and T. M. H. Le, “Real time data-driven approaches for credit card fraud detection,” in Proceedings of the 2018 International Conference on EBusiness and Applications. ACM, 2018, pp. 6–9.
- [6] A. Roy, J. Sun, R. Mahoney, L. Alonzi, S. Adams, and P. Beling, “Deep learning detecting fraud in credit card transactions,” in 2018 Systems and Information Engineering Design Symposium (SIEDS). IEEE, 2018, pp. 129–134.
- [7] Shiyang Xuan, Guanjun Liu, “Random Forest for Credit Card Fraud Detection” in 2018 IEEE
- [8] Dejan Varmedja, Mirjana Karanovic, Srdjan Sladojevic, Marko Arsenovic, and Andras Anderla, Credit Card Fraud Detection - Machine Learning methods, Publish in:18th International Symposium INFOTEH-JAHORINA, 20-22 March 2019 IEEE.
- [9] Rimpal R. Popat and Jayesh Chaudhary, A Survey on Credit Card Fraud Detection using Machine Learning, 2018 (IEEE), pp. 1120 -1125
- [10] Kuldeep Randhawa, Chu Kiong Loo, Manjeevan Seera, Chee Peng Lim, Ashoke K. Nandi, Credit Card Fraud Detection Using AdaBoost and Majority Voting, Published in: IEEE Access on 15 February 2018, vol. no.6, pp. 14277 – 14283.
- [11] Changjun Jiang, Jiahui Song, Guanjun Liu, Member, IEEE, Lutao Zheng, and Wenjing Luan, Credit Card Fraud Detection: A Novel Approach Using Aggregation Strategy and Feedback Mechanism, 2018 IEEE, pp. 2327-4662.
- [12] Hassan Najadat, Ayah Abu Aqouleh ,Ola AlItiti, MutazYounes,”Credit Card Fraud Detection Based on Machine and Deep Learning”, in 2020.