

# Implementation of Machine Learning Algorithm to Detect Credit Card Frauds

Nihar Ranjan, PhD  
Rajarshi Shahu College of  
Engineering, Tathawade, Pune

Sneha George  
Rajarshi Shahu College of  
Engineering, Tathawade, Pune

Pallavi Pathade  
Rajarshi Shahu College of  
Engineering, Tathawade, Pune

Rakshita Anikhindi  
Rajarshi Shahu College of Engineering,  
Tathawade, Pune

Sneha Kamble  
Rajarshi Shahu College of Engineering,  
Tathawade, Pune

## ABSTRACT

As the world is becoming more digitalized with every sector using the internet to flourish their businesses, online transactions have become an inevitable part of life. There has been a steady rise in the number of online transactions and this will continue to increase in the future as well. One of the major modes of online transactions is credit cards and along with its extensive use comes its major drawback, that is, credit card fraud. Machine learning plays a vital role in detecting credit card frauds as it is not possible for banks to monitor every transaction. This paper explores different machine learning algorithms used to detect credit card frauds.

## Keywords

Machine learning, Credit card, Fraud detection, Random forest, Logistic regression, Decision tree, Resampling

## 1. INTRODUCTION

In recent years, online transactions have taken up high importance in one's life. From small-scale traders to large businesses, all make sure to have their online presence to extend their products and services to a larger audience. Since distance is not a barrier, people prefer online transactions most of the time. This is also because online transactions come with more advantages like it is more convenient, the user can keep a track of his/her spending, and have a disciplined budget. According to statistics, 80% of people prefer cards over cash[1]. Banks also try to increase their customer engagement by providing various benefits on credit cards like the ease of credit, EMI facility, incentive and offers, flexible credit, purchase protection, and making foreign transactions without incurring a fee[2].

But one of the major disadvantages of credit card use is credit card fraud.

### 1.1 Credit Card Fraud

Credit Card fraud is a major problem of business institutes as well as financial institutes. With the advancement in technology, fraudsters are also taking up advanced ways to commit fraud. Credit card fraud can be stated as "using an individual's card without the knowledge of the owner of the card". A credit card fraud can happen with the physical loss of the card or by just knowing the necessary information required for a transaction.

The necessary information is:

- Name of cardholder

- Secure code(CVV code)
- Card number
- Expiration date

Most credit card frauds take place in one of the following ways:[3]

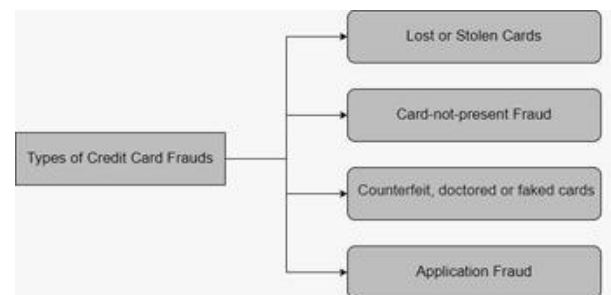


Figure 1. Types of Frauds

Lost or stolen cards: This is when the fraudsters make use of cards that were lost or stolen without the knowledge of the card owner. The thief may use this card to make online purchases.

Card-not-present fraud: In this type of fraud, only the availability of sensitive information like the account holder's name, card number, and the expiration date is needed to commit the crime.

Counterfeit, doctored, or faked cards: Skimmers are devices that are used to obtain credit card details illegally. The fraudsters install these at gas pumps or ATMs to collect card data[4]. This data is then encoded into a fake or doctored card which is hard to distinguish from real ones.

Application fraud: When the fraudsters use someone else's information to get a new credit card by luring a card owner to give away sensitive information or by stealing their details, it is called application fraud. They may also provide fake or stolen documents to vindicate their application.

But along with the world, the fraudsters have now moved to digital ways to carry out their frauds. Fraudsters can now replicate complete websites which look exactly like the genuine ones to deceive the user. Phishing techniques are also becoming popular ways of committing fraud. The hacking of sensitive information from online sites is another way for

fraudsters to carry out their crime.

Based on data revealed by the Reserve Bank of India (RBI) stated in [5], Fraudsters siphoned off ₹ 615.39 crores in more than 1.17 lakh cases of credit and debit card frauds over 10 years (April 2009 to September 2019). Another data revealed [6], in 2018, \$24.26 Billion was lost due to payment card fraud worldwide, and credit card fraud increased by 18.4 percent in 2018 and is still climbing. This is a major issue for the economy as well as the banks issuing credit cards. It causes uncertainty among the customers and makes the financial framework fragile.

## 1.2 Credit Card Fraud Detection

When a transaction is initiated, it goes through a terminal check. In this terminal check, parameters such as the correct pin, sufficient balance, and status of the card are checked. If any of these parameters fail in the check, the transaction is rejected. Else the transaction will proceed further into a predictive model. The predictive model is trained using machine learning algorithms to detect any suspicious behavior. If the model classifies it as suspicious, the transaction is rejected and the same is communicated to the user/owner of the card. Otherwise, the transaction is completed smoothly without any hassle.

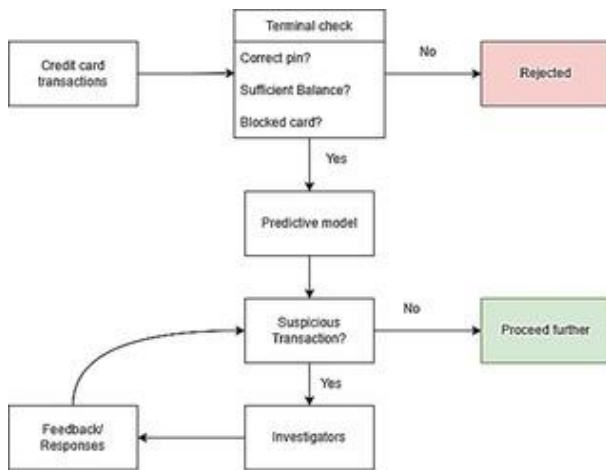


Figure 2. Credit Card Fraud Detection

## 2. LITERATURE SURVEY

In [7] machine learning algorithms like Logistic Regression, Decision Tree, Random Forest, Naïve Bayes, and neural network technique ANN have been used. They have concluded that the ANN model has the best accuracy and precision.

In [8], machine learning techniques like Logistic regression, Decision Tree, and Random forest have been used. By comparing all three methods, it is found that the random forest classifier is better than the logistic regression and decision tree.

In [9], the focus is mainly on credit card fraud detection in the real world. With the proposed scheme, using a random forest algorithm the accuracy of detecting fraud can be improved. The performance of the techniques is evaluated based on accuracy, sensitivity, specificity, and precision.

In [10], predictive models such as logistic regression, random forest, and XGBoost in combination with different resampling techniques have been applied. The experimental results

showed that random forest in combination with a hybrid resampling approach of SMOTE and TomekLinks removal performed better than other models.

In [11], they use SVM, NB, KNN, Logistic Regression, and Random forest algorithms to identify which algorithm works better for Credit card fraud detection. Based on accuracy or MCC's best value, they choose Random forest and KNN algorithm.

In [12], the aim is to overcome challenges with card frauds related datasets. They work on both imbalance and balanced datasets. Matthews Correlation Coefficient was the better parameter to deal with the imbalance dataset by applying the SMOTE for balancing the dataset.

In [13], different algorithms of machine learning that are used for detecting fraudulent transactions have been discussed. It also includes brief information on types of frauds and gives an idea about machine learning algorithms such as SVM, Logistic Regression, Random Forest, Neural network, Decision tree. They have also mentioned that random forest detects frauds with more accuracy.

In [14], they have proposed a system that is based on machine learning. They have focused on logistic regression algorithms and have given ideas about logistic function and also how data should be prepared for logistic regression.

## 3. DETECTION TECHNIQUES

### 3.1 Logistic Regression

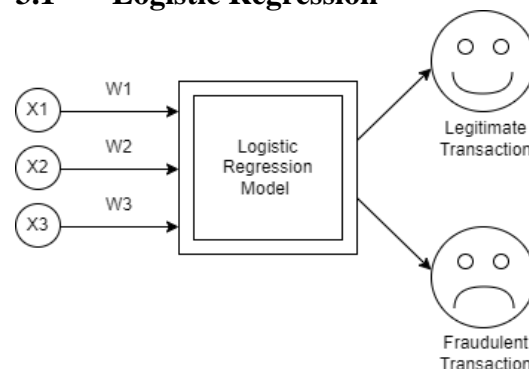


Figure 3. Logistic Regression

Logistic regression is a statistical method or a classification algorithm that predicts results in the form of binary values (1/0, Yes/No, True/False) in a given set of independent variables. It is used to understand the relationship between the dependent variable and the independent variables. Logistic Regression works with the sigmoid function. This algorithm is more suitable for less amount of data to provide good accuracy

### 3.2 Decision Tree

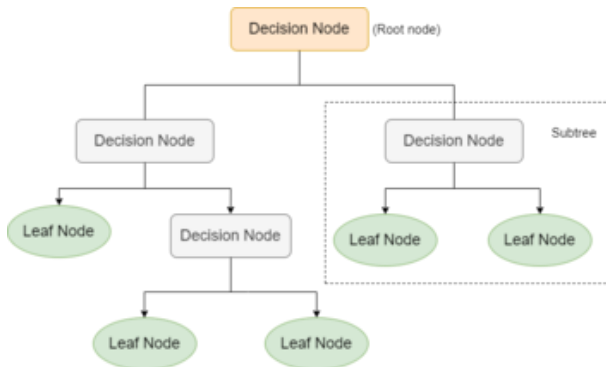


Figure 4. Decision Tree

Decision Tree is a supervised learning technique. It can be used for both classification and regression problems but is mostly used for classification problems. It is a tree-structured classifier, which consists of two nodes namely the decision node and the leaf node. The decision nodes represent the features of a dataset, branches represent the decision rules and each leaf node represents the outcome.

### 3.3 Random Forest

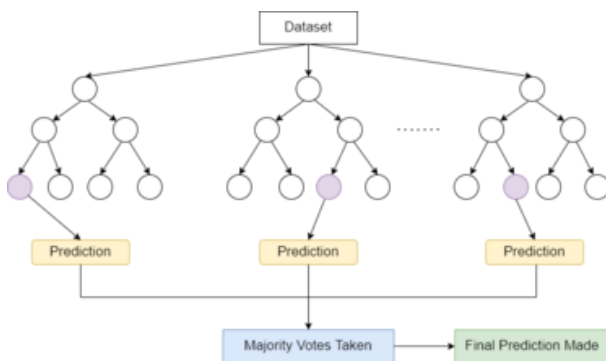


Figure 5. Random Forest

Random forest is an ensemble learning-based supervised machine learning algorithm which can be used for both classification and regression problems. In ensemble learning, one can join different types of algorithms or multiple algorithms of the same type to generate one ideal predictive model. This helps in the reduction of errors and improves the prediction. Random Forests are created from subsets of data and the final output is based on average, taking care of the overfitting problems. This algorithm works efficiently with a large amount of data.

### 3.4 Naive Bayes

Formula of Naive Bayes Theorem:-

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

where,



Figure 7. XGBoost

$P(A|B)$  is the posterior probability of the hypothesis that the evidence is true.

$P(B|A)$  is the likelihood of the evidence given that the hypothesis is true.

$P(A)$  is the prior probability of the hypothesis.

$P(B)$  is the prior probability that the evidence is true.

Naïve Bayes is a supervised learning algorithm based on Bayes Theorem which is used to solve classification problems. It uses Bayes Theorem for calculating probabilities and conditional probabilities. It works efficiently with large amounts of data. It can be implemented by giving independent features as input and dependent features as output. The probability of the dependent feature is calculated with respect to the independent features.

### 3.5 Support Vector Machine (SVM)

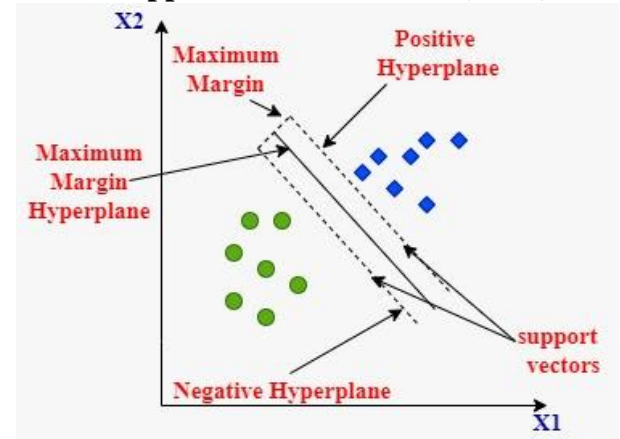


Figure 6. Support Vector Machine (SVM) Model Graph

SVM is the most popular supervised learning algorithm which is used for classification-based problems and regression problems. It can solve linear and non-linear problems. SVM can be used to analyze data for classification and regression using algorithms and kernels in SVM. It is a classifier defined by separating hyperplanes. SVM classifiers provide good accuracy. They use a subset of training points hence as a result use very less memory.

### 3.6 XGBoost

XGBoost or Extreme Gradient Boosting decision-tree-based ensemble Machine Learning algorithm. It uses bagging and boosting ensemble learners with decision trees. Then Gradient Boosting technique is applied which employs gradient descent algorithm to optimize the loss function. This algorithm is portable, supports all major programming languages and can be used to solve various types of problems [15].

Comparison of algorithms:

Table 1 shows the results of the algorithms discussed in this paper in terms of accuracy and precision. From the table, it can be seen that Logistic regression gives the best accuracy and precision. Though the accuracy of the Support Vector Machine algorithm is a little higher than that of Logistic Regression, the latter is considered the best among all of them as it gives the best precision.

**Table 1. Accuracy and Precision comparison table for different ML algorithms**

Model	Accuracy	Precision
Logistic Regression	94.84	97.58
Decision Tree	92.88	99.48
Random Forest	94.00	93.99
XGBoost	90.76	89.96
Naive Bayes	91.62	97.09
Support Vector Machine	94.99	95.98

#### 4. CONCLUSION

Credit card fraud is a very serious issue that has been affecting the victims as well as banks and credit card companies with the loss of money. The use of machine learning algorithms has proven to be beneficial to detect these frauds. In this paper, we studied different machine learning algorithms like Logistic regression, Decision tree, Random forest, XGBoost, Naive Bayes, and SVM. These come under supervised machine learning algorithms. Random forest and XGBoost are ensemble learning algorithms built on Decision trees.

As a future scope of this study, a credit card fraud detection system can be developed which can detect fraudulent transactions in real-time giving the fraudsters no time to escape. These systems can be implemented in the banking sector and online eCommerce sites. The system can be very helpful for tracking fraudulent transactions and thus preventing the card owner from paying for what he/she did not purchase.

#### 5. REFERENCES

[1] Jayprakash S Naidu. "₹615.39cr lost to debit, credit card frauds", Hindustan Times, Feb 11, 2020

[2] "Credit Card Fraud Statistics", Shift Processing, September 2021

[3] Devi Meenakshi, Janani, Gayathri, Mrs. Indira. Credit Card Fraud Detection using Random Forest. International Research Journal of Engineering and Technology (IRJET), Volume: 06 Issue: 03 | Mar 2019

[4] Shakya, Ronish, "Application of Machine Learning Techniques in Credit Card Fraud Detection" (2018). UNLV Theses, Dissertations, Professional Papers, and Capstones. 3454.

[5] Varun Kumar K S, Vijaya Kumar V G, Vijay Shankar A, Pratibha K, "Credit Card Fraud Detection using Machine Learning Algorithms", International Journal of Engineering Research & Technology (IJERT), ISSN: 2278-0181, Vol. 9 Issue 07, July-2020

[6] Lakshmi S V S, SelvaniDeepthiKavila. "Machine Learning For Credit Card Fraud Detection System", International Journal of Applied Engineering Research, ISSN 0973-4562, Volume 13, 2018

[7] NayanUchhana, Ravi Ranjan, Shashank Sharma, Deepak Agrawal, AnuragPunde. Literature Review of Different Machine Learning Algorithms for Credit Card Fraud Detection. International Journal of Innovative Technology and Exploring Engineering (IJITEE), 6 April 2021

[8] VaishnaviNathDornadulaa\*, GeethaSa."Credit Card Fraud Detection using Machine Learning Algorithms". International Conference on Recent Trends in Advanced Computing 2019.

[9] Harish Paruchuri, "Credit Card Fraud Detection Using Machine Learning: A Systematic Literature Survey". ABC Journal of Advanced Research, Volume 6, No 2(2017).

[10] K. Gowthami, K V L E Praneetha, G. Vinitha, Ch. ReddammaKumari, P. Sandhya Krishna. Credit Card Fraud Detection Using Logistic Regression. Journal of Engineering Sciences(JES) Vol 11, Issue 4, April/2020

[11] Vishal Morde, XGBoost Algorithm: Long May She Reign!, Towards Data Science, April 8, 2021

[12] NiharRanjan, Rajesh Prasad, "Automatic Text Classification using BPLion- Neural Network and Semantic Word Processing", Imaging Science Journal Print ISSN: 1368-2199, Online ISSN:1743-131X, July 2017, pp 1-15

[13] NiharRanjan, Rajesh Prasad," LFNN: Lion fuzzy neural network-based evolutionary model for text classification using contextand sense based features", Applied Soft Computing, ISSN: 1568-4946, July 2018, pp 994-2018

[14] NiharRanjan, MidhunChakkaravarthy,"Evolutionary andIncremental TextDocument Classifier usingDeep Learning"International Journal of Grid and Distributed Computing Vol. 14, No. 1, (2021), pp. 587-595

[15] NiharRanjan, MidhunChakkaravarthy "A Brief Survey of Machine Learning Algorithms for Text Document Classification onIncremental Database" TEST:Engineering & Management, ISSN: 0193-4120, pp 25246 – 25251

[16] NiharRanjan, Rajesh S. Prasad " Author Identification in Text Mining for used in Forensics" ISSN 2321- 9637, Volume 1, Issue 5, December 2013, pp. 568-571