



# Credit Card Fraud Detection Using Supervised Machine Learning Algorithms: - A Review

<sup>1</sup>K. Sai Srinivas

<sup>1</sup>Student, Department of Information Technology  
G M R Institute of Technology, Rajam, India

**Abstract:** Development of communication technologies and e-commerce has made the credit card as the most common technique for both online and regular purchases. The number of online exchanges has grown in huge amounts and online credit card exchanges holds a tremendous share of these exchanges. Because of lot of loop holes in this system and a lot of problems are being raised in this system in the method of credit card frauds. Due to this the industries and customers who are using credit cards are facing a huge loss. Therefore, it is essential to create mechanisms that ensure the security and integrity of credit card exchanges. Therefore, the banks and financial institutions offer credit card fraud detection applications much value and demand. This study proposes a scheme for detecting frauds in credit card using Supervised machine learning algorithms like, Support Vector Machine (SVM), K-Nearest Neighbour, Naïve bayes, Decision Trees, Random Forest and Logistic Regression technique to increase their quality of classification. This study using Supervised machine learning algorithms helps us to differentiate between genuine and fraudulent credit card transactions.

**Keywords:** - Credit Card, Fraud Detection, Machine Learning, Decision Trees, K-Nearest Neighbour (KNN), Support Vector Machine (SVM), Naïve Bayes, Random Forest, Logistic Regression, Machine Learning.

## I. INTRODUCTION

Nowadays, With the rise in the payment through credit card there are many people emerged who wanted to do some fraudulent activities and gain some profit using someone's credit card by performing some fraudulent transactions using somebody's credit card. With the expanding number of clients, credit card fakes are too expanding at a comparable pace. The credit card data of a specific person can be collected illegally and can be utilized for false transactions. Fraud prevention and fraud detection are two of the main mechanisms to avoid frauds and losses due to fraudulent activities. Fraud prevention is the upbeat mechanism with the goal of disabling the happening of fraud. Fraud detection systems come into play when the fraudsters go beyond to the fraud prevention systems and start a fraudulent transaction. With the purpose of automating the solution, this is a central problem that demands the involvement of fields such as machine learning. To detect whether the transaction done by using a credit card is fraudulent or not we need to develop some machine learning models which will predict whether the performed transaction in fraudulent or not. To develop a machine learning model firstly, we must take some algorithms and should pick an algorithm which gives high accuracy and we should develop a machine learning model with that algorithm, and we must train the machine learning model with a dataset so that the model can make predictions on the new transaction and let us know whether the transaction is fraud or not. So, in my research I will use supervised machine learning algorithms to evaluate the accuracy of fraud detection.

## II. LITERATURE SURVEY

In the paper [1] Rahul Powar, Rohan Dawkhar, Pratichi, "Credit Card Fraud Detection Using Machine Learning" proposes a machine learning model using K-Nearest Neighbour (KNN) algorithm for detecting the credit card frauds. In this, author uses the behavioural pattern of spending money which depends on past history of transactions and attributes such as location, daily expenses, transaction time of cardholder is compared with current transaction details to detect the credit card frauds. The author has developed a model which helps us to identify whether the new transaction is fraud or not.

In the paper [2] M. Amarender reddy, Dr. Pravin R Kshirsagar, D.Akshitha, G . Alekya, K.Divya rosy, "Implementation Of Credit Card Fraud Detection Using Support Vector Machine" proposes a machine learning model using Support Vector Machine(SVM) algorithm for detecting the credit card frauds. In this, author preprocesses the data to avoid malwares in the data and correct the errors in the data and balance the data in the given dataset. The author had implemented SVM method on a publicly available dataset, the author says that the results shows that the proposed SVM classification gives better performance compared to other.

In the paper [3] Heta Naik, "Credit card Fraud Detection based on Machine Learning Algorithms" proposes a machine learning model using Naïve Bayes algorithm for detecting the credit card frauds. In this, author says that the processing speed of the naïve bayes algorithm is high and is one of the most effective classification algorithms which helps in building the machine learning model faster and make the model to make quick predictions. Naïve Bayes is a probabilistic classifier, which means it make predictions on the basis of the probability of an object.

In the paper [4] Emmanuel Ileberi, Yanxia Sun, Zenghui Wang, - "Performance Evaluation of Machine Learning Methods for Credit Card Fraud Detection Using SMOTE and AdaBoost" proposes a machine learning model using Decision Trees Algorithm for detecting the credit card frauds. In this, author uses Decision Trees and develops a machine learning model which makes predictions on classifying the transactions whether they are fraud or not, and boosts the model using AdaBoost and gets a better accuracy.

In the paper [5] M. Suresh Kumar, V. Soundarya, S. Kavitha, E. S. Keerthika, E. Aswini- "Credit Card Fraud Detection Using Random Forest Algorithm" proposes a machine learning model using Random Forest Algorithm (RFA) for detecting the credit card frauds. In this, author takes a data set and performs cleaning on the dataset and divides the dataset into training dataset and testing dataset. After dividing the author applies Random Forest Algorithm. After this analysis the accuracy of credit card fraud transactions can be obtained which will be finally represented in the form of graphical representation.

In the paper [6] Hala Z Alenzi, Nojood O Aljehane- "Fraud Detection in Credit Cards using Logistic Regression" proposes a machine learning model using Logistic Regression classifier algorithm for detecting the credit card frauds. In this, author says that these credit card services are much more efficient and economical to use, but with rapid increasing of the credit cards, it has become easy for the attackers to prey the people by sending fraud messages. So here the author uses Logistic Regression classifier algorithm for detecting the frauds transaction in the credit card transactions.

### III. METHODOLOGY

In this paper, the problem mainly depends on classification of credit card transactions. So, machine learning predictive models will decide the credit card transaction is fraudulent or not fraudulent, after predictive models are formulated using the training data. Classification should be done for two categories namely fraud transaction or legal transaction. By using supervised machine learning algorithms K-Nearest Neighbour (KNN), Support Vector Machine (SVM), Naïve Bayes, Random Forest, Logistic Regression to train the model with the data and find the credit card transaction is legal or fraud.

#### K-NEAREST NEIGHBOR(KNN):

The k-nearest neighbor (KNN) [1] is a simple, easy to implement supervised machine learning algorithm that can be used to solve both classification and regression problems. The concept of K-nearest neighbor analysis has been used in several anomaly detection techniques. The performance of KNN algorithm is impacted by three main factors:

- The distance rule is used to find the nearest neighbors.
- The distance rule is used to derive a classification from the k-nearest neighbor.
- The number of neighbors is used to classify the new sample.

In process of KNN, we classify any incoming transaction by calculating of closest point to new incoming transaction. Then if the nearest neighbor be legal, then the transaction indicates as a legal and vice versa.

#### SUPPORT VECTOR MACHINE(SVM):

SVM algorithm [2] is used for classification and pattern analysis. It is a classification method to classify or predict patterns into two classes; fraud or legal. This technique is used for binary classifications. SVM works primarily on the Histogram-based data, and at the same time it retains relatively accurate results. Usually, SVM is faster than other gradient boosting algorithms. The main ideology behind this algorithm is to prevent some issues like overfitting, fights against Gradient bias, categorical features support, and provide good results with default parameters.

- Import libraries and load the csv file.
- Summarize and pre-process the data.
- The categorical information will be changed into numerical data by one-hot encoding.
- Use the SVM model to find whether the incoming transactions are fraud or legal.

#### NAÏVE BAYES:

Naïve Bayes [3] is a classification algorithm and is a collection of classification algorithms based on Bayes' Theorem. It is not a single algorithm but a family of algorithms, and all of them share a common principle, i.e., each and every pair of features that are being classified is independent of each other.

- Bayes theorem: Bayes' Theorem states that the conditional probability of an event, based on the occurrence of another event, is equal to the likelihood of the one event given by another event being multiplied with the probability of another event.

$$\text{Formula: } - P(B/A) = (P(A/B) * P(B)) / P(A)$$

Were

$$P(A) = \text{Probability of A}$$

$$P(B) = \text{Probability of B}$$

$$P(A/B) = \text{Posteriori probability of B}$$

- Naïve Bayes classifier algorithm is very easy and fast. This algorithm needs very much less training data and is highly scalable.

#### DECISION TREES:

Decision Trees (DTs) [4] are a non-parametric supervised learning method used for classification and regression. The goal is to create a model that make predictions on the value of a target variable by learning simple decision rules inferred from the data features. Decision Trees algorithm is first used and then the Adaptive Boosting algorithm is used to boost the accuracy of the model which we get using Decision Trees algorithm. This approach uses a tree-like construct to make the predictions. Some of the advantages of using DT include the fact that is simple to interpret and it does not require an extensive data preparation. AdaBoost is a machine learning algorithm. Mainly developed for binary classification. AdaBoost approach is used to improve the performance of some classifiers such as decision tree with regard to performance metrics such as the accuracy.

**RANDOM FOREST:**

Random Forest [5] is also called as Random Decision Forest (RFA) which is used for Classification, Regression and other tasks that are performed by constructing multiple decision trees. Random Forest Algorithm is based on supervised learning and the main advantage of this algorithm is, it can be used for both Classification and Regression.

- Pre-process the dataset and divide the data set into training and testing datasets.
- The pre-processed dataset will be analyzed again and then a confusion matrix will be obtained.
- In confusion matrix the dataset will be partitioned into four blocks as True Positive (TP), True Negative (TN), False Positive (FP) and False Negative (FN).
- The dataset will be partitioned continuously until all the data is validated.
- Now all these partitioned data will be evaluated and finally it will be represented as separate graphs.

**LOGISTIC REGRESSION:**

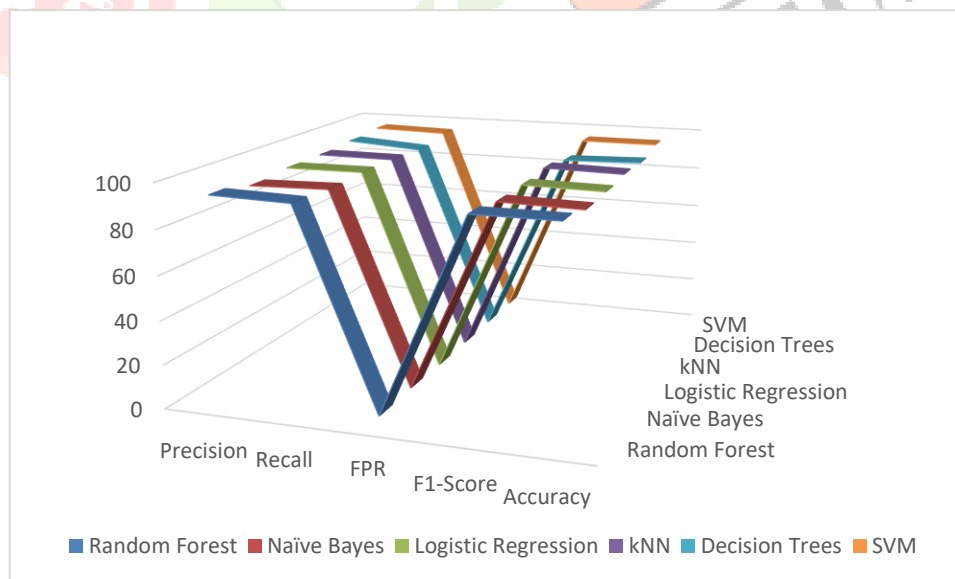
Logistic Regression [6] is a Machine Learning algorithm-based classification method. It's a technique for forecasting a categorized outcome variable from a set of individual factors.

- The logistic function's curve reflects the probability of items such as how the cells are carcinogenic or not, whether a rodent is obese or not based on its bodyweight, and so on.
- The logistic function's curve reflects the probability of items such as how the cells are carcinogenic or not, whether a rodent is obese or not based on its bodyweight, and so on.
- A mathematical formula for translating expected values into probabilities is the sigmoid function. Every real value between 0 and 1 is converted to a distinct number.
- The value of the logistic regression should be within 0 and 1, and this must not surpass this limitation, else a "S" curve would result. The S-form curve is the Sigmoid function, commonly referred as logistic function.

**IV. CONCLUSION**

	Random Forest	Naïve Bayes	Logistic Regression	kNN	Decision Trees	SVM
<b>Precision</b>	93.998	91.201	92.8956	93.228	94.5891	95.9887
<b>Recall</b>	93.001	91.989	93.112	93.005	92.008	95.1234
<b>FPR</b>	4.665	4.7789	3.9785	3.889	3.998	3.9875
<b>F1-Score</b>	93.998	91.7748	92.112	93.479	91.003	95.1102
<b>Accuracy</b>	94.001	91.0021	91.5456	92.789	91.998	95.1023

**Table-1: - Different ML Algorithms with their Precision, Recall, False Positive Rate, F1-Score and Accuracy values**



**Fig-1: - Analysis of different ML Algorithms with their Precision, Recall, False Positive Rate, F1-Score and Accuracy values**

Fig-1 represents the visualization and comparison between the Precision, Recall, False Positive Rate, F1-Score and Accuracy values of different supervised machine learning algorithms.

## V. CONCLUSION

From the analysis conducted in this paper, Different supervised machine learning algorithms are applied on the data to detect the credit card fraud transactions. On the basis of comparative study performed among the applied various techniques K-Nearest Neighbours, Support Vector Machine, Naïve Bayes, Decision Trees, Random Forest, Logistic Regression. After analyzing the approaches towards the credit card fraud detection, Support Vector Machine gives good results when compared with the remaining algorithms. Hence the machine learning model which is developed using Support Vector Machine Algorithm Can be used very effectively in detecting the credit card scams which are being done.

## VI. REFERENCES

- [1] Rahul Powar, Rohan Dawkhar, Pratichi, "Credit Card Fraud Detection Using Machine Learning" - International Journal of Advance Scientific Research and Engineering Trends, Volume 5, Issue 9, September 2020, ISSN (Online) 2456-0774
- [2] M. Amarender reddy, Dr. Pravin R Kshirsagar, D.Akshitha, G . Alekya, K.Divya rosy, "Implementation Of Credit Card Fraud Detection Using Support Vector Machine" Vol 12, Issue 06, June /2021 ISSN NO: 0377-9254
- [3] Heta Naik, "Credit card Fraud Detection based on Machine Learning Algorithms" - International Journal of Computer Applications (0975 – 8887) Volume 182 – No. 44, March 2019
- [4] Emmanuel Ileberi, Yanxia Sun, Zenghui Wang, "Performance Evaluation of Machine Learning Methods for Credit Card Fraud Detection Using SMOTE and AdaBoost", December 22,2021-IEEE
- [5] M. Suresh Kumar, V. Soundarya, S. Kavitha, E. S. Keerthika, E. Aswini- "Credit Card Fraud Detection Using Random Forest Algorithm", 3rd International Conference on Computing and Communication Technologies (ICCCCT), 2019
- [6] Hala Z Alenzi, Nojood O Aljehane- "Fraud Detection in Credit Cards using Logistic Regression" - International Journal of Advanced Computer Science and Applications (IJACSA), Vol 11, Nov 12, 2020.
- [7] Dileep M R, Navaneeth A V, Abhishek M – "A Novel Approach for Credit Card Fraud Detection using Decision Tree and Random Forest Algorithms" - Proceedings of the Third International Conference on Intelligent Communication Technologies and Virtual Mobile Networks (ICICV 2021). IEEE Xplore Part
- [8] Dhankhad, S., Mohammed, E. and Far, B., 2018, July "Supervised machine learning algorithms for credit card fraudulent transaction detection: a comparative study" In 2018 IEEE International Conference on Information Reuse and Integration (IRI) (pp. 122-125). IEEE.
- [9] Yee, O.S., Sagadevan, S. and Malim, N.H.A.H., 2018. "Credit card fraud detection using machine learning as data mining technique", Journal of Telecommunication, Electronic and Computer Engineering (JTEC), 10(1-4), pp.23-27.
- [10] Awoyemi, J.O., Adetunmbi, A.O. and Oluwadare, S.A., 2017, October. Credit card fraud detection using machine learning techniques: A co parative analysis. In 2017 International Conference on Computing Networking and Informatics (ICCNi) (pp. 1-9). IEEE.

