IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Credit Card Fraud Detection Using Machine Learning-A Review.

[1]Esha Santosh Mohite, [2]Monica Dashrath Bhumarkar, [3]Dr. Harshali Patil, [4]Dr. Jyotshna Dongardive

[1]Student, [2]Student, [3]Professor, [4]Professor

Department of Computer Science,

University of Mumbai, Mumbai, India

Abstract: Nowadays, the usage of credit cards has dramatically increased. As credit cards become the most popular mode of payment for both online and regular purchases, cases of fraud associated with them are also rising. Criminals use fake identification and numerous technologies to attract customers and get cash out of them. Due to the increase in fraud, various techniques and methodologies have been developed for detecting it. In this paper, we discuss the different ways credit card frauds are committed and which machine learning algorithms are used by investigators to detect fraud.

Index Terms - Machine Learning, credit card, fraud detection, SVM, Decision Tree, Logistic regression.

I. INTRODUCTION

Credit card fraud Detection is the process of relating purchase attempts that are fraudulent and rejecting them rather than recycling the order. Credit card is the small plastic card furnish to customers as a system of payment. Credit card fraud detection requires analysis of transaction histories, user behaviour patterns, and external data sources.

Fraud detection is a set of processes and analyses that allow business. Credit card fraud happens when someone steals or uses without your credit card or credit card information, your information or your card is used to make a purchase in a store or make a transaction online.

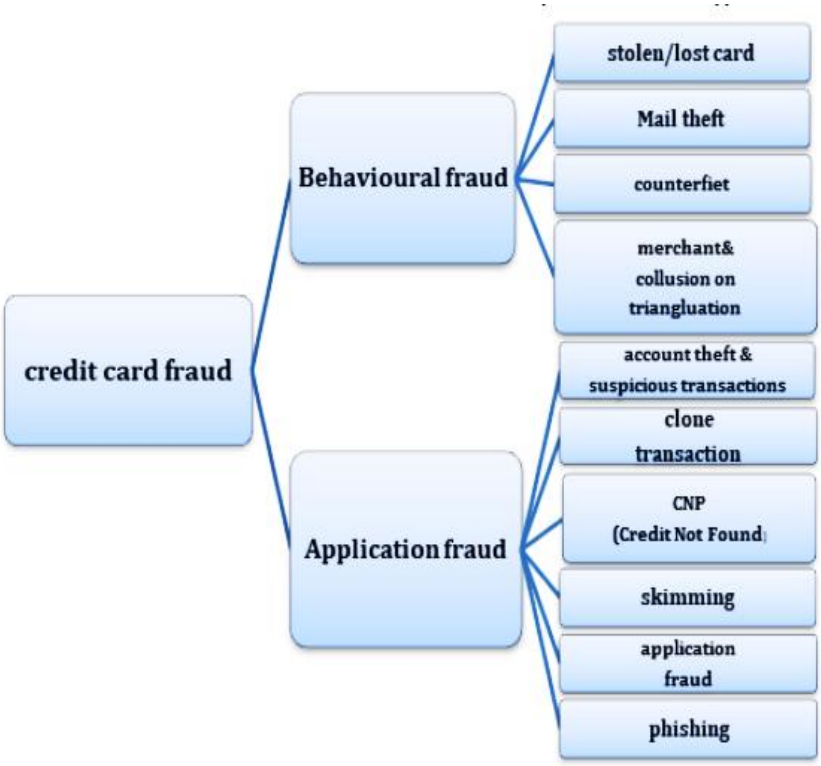
Credit card scam fall under identity theft and have become increasingly common. Nowadays scammers use your card details to perform unethical activities and imitate transaction from your account without your knowledge. Such events have led to the need for credit card security.

Credit card fraud detection with machine learning is a process of data investigation by a data science team and the development of a model that will provide the best results in revealing and preventing fraudulent transactions.

The models are trained as historical data to detect abnormal trends. machine learning algorithm can analyse massive volumes of data and detect patterns that humans may overlook. This can assist organizations in detecting fraud with greater accuracy and speed than traditional methods.

II. Some common types of credit card frauds:

Credit card frauds can be online as well as offline. Online fraud like over the phone, by text and in person. Fake emails, information stolen in data or credit card stolen out of your mailbox.



- **Card not present fraud**:
  Fraudsters steal a cardholder's credit card and personal information and then use it to buy online or by phone. This fraud is hard to prevent because there is no physical card to examine, and the vendor cannot confirm the buyer's identity.[1]
- **Account takeover:**
  After stealing personal information, Scammers change passwords and PIN number so they can take over the account. This type of fraud will quickly be detected when the cardholder tries to use their card or log in to their account online.
- **Credit card Skimming:**
  Skimmers are devices that steal credit card information from magnetic strip on the back of the card. It is a type of cyber-crime in which stealers steal credit card or debit card data that is transmitted when people swipe their cards. Scammers attach them to credit card reader machines in ATM's, Retail stores, Gas stations, and other.
- **Lost or stolen cards:**
  The clearest way that your credit card could be compromised is through theft. [7]
  The culprit used stolen personal information to apply for a credit card. This type of fraud can go undetected until the victim applies for credit card themselves or checks their credit report.
- **Application fraud:**
  Application fraud is when someone applies for a credit card with false information. To detect application fraud, the solution is to implement a fraud system that allows identifying suspicious applications. To detect application fraud, two different situations have to be distinguished: when applications come from a same individual with the same details, the so-called duplicates, and when applications come from different individuals with similar details, the so-called identity fraudsters.[7]
- **Phishing:**
  Phishing is the act of personally asking for card information from you. numerous scammers pose as insurance agent or other marketers on the phone and convince you to give your card information under the excuse of making an agreement.
- **Clone Transaction:**
  Clone transactions are frequently a popular method of making transactions like an original one or duplicating a transaction. This can happen when an organization attempts to get payment from a partner repeatedly by sending the same statement to different departments.[2]

  The conventional method of rule-based fraud detection algorithm does not work well to distinguish a fraudulent transaction from irregular or incorrect transactions. For case, a user could click the submission button two times by accident or order the same product over again.

III. Machine Learning-based Fraud Detection:

a) Detecting fraud automatically
b) Real-time streaming
c) Less time needed for verification methods.
d) Identifying hidden correlations in data [2]

IV. Machine learning algorithms for credit cards fraud detection:

1. Logistic Regression (LR)
2. Support Vector Machine (SVM)
3. Decision Tree (DT)
4. Artificial Neural Network (ANN)

1. Logistic regression:

Logistic regression is a statistical model used to analyze the relationship between a dependent variable with two outcomes (binary or binomial logistic regression) or multiple categories (multinomial logistic regression) and one or more independent variables, which can be either qualitative or quantitative.

This method stands out as a direct approach for addressing both classification and regression problems. Its versatility extends to applicatic ns such as email categorization and spam detection, among other use cases. Logistic regression calculates the probability of a binomial or multinomial output, incorporating the sigmoid function to model data interactions between dependent and independent variables. This versatile method finds application in contemporary research, such as classifying transactions as fraudulent or legitimate. Despite its efficiency, caution is advised with high-dimensional datasets to prevent potential overfitting. t excels in delivering improved accuracy and provides insights into class distribution within the feature space, similar to certain alternative methods. However, a drawback lies in its reliance on the assumption of linearity between dependent and independent variables. While both classification and regression fall under supervised learning, their output characteristics diverge. In regression, the output variable represents a numeric value.

In the realm of regression tasks, the output resides on a consistent scale, meaning it is expressed as a whole number or decimal value. Alternatively, in a classification task, the algorithm's outcome assigns input variables to specific pre-defined categories, accurately placing them within their designated groups. Termed as a logic model, logistic regression operates as a binary classification, determining the conditional probability of one of the two possible outcomes of the response variable. This determination is based on a linear combination of two or more input variables, adjusted by the logistic function. In binary classification, the model's task is to predict the response variable as one of the two possible classes, typically 0 or 1. Logistic Regression can be elucidated through the logistic function, also known as the Sigmoid function, which transforms real input (x) into a probability value within the range of 0 to 1.

The experimental findings demonstrated that Logistic Regression achieved an impressive accuracy of 97.70%.

2. Support Vector Machine:

SVMs aim to find the optimal hyperplane that separates different classes in the feature space, and support vectors play a crucial role in defining this boundary. These support vectors are the data points that have the most significant influence on the placement of the hyperplane, making SVM effective in handling classification and regression tasks.

SVM handle non-linearly separable data using the kernel trick. By mapping the data into a higher-dimensional space, SVM can find a hyperplane that effectively separates classes, even when the original data is not linearly separable. The kernel trick is a powerful technique that allows SVM to handle complex decision boundaries, making it versatile for a wide range of classification and regression problems.

SVMs highlighting their effectiveness in solving regression problems by discovering functions that forecast continuous values. You've emphasized their strengths, such as robustness to noise, the ability to handle non-linear data, and the importance of support vectors in determining the optimum decision boundary.

The mention of SVM as an inequality classifier defined by a separating hyperplane accurately captures its role in supervised learning for categorizing new examples based on labeled training data. This enables the algorithm to find a decision boundary or hyperplane that can effectively classify the data.[4]
A SVM is a inequality classifier formally defined by a separating hyperplane. In other words, given labelled training data (supervised learning), the algorithm outputs an optimal hyperplane which categorizes new examples. [6]

1) Set up the training data for model creation.
2) Set up SVM's parameters.
3) SVM Trainer
4) SVM Predictor

The experimental outcomes showed that the SVM, obtained 97.50% accuracy.

3. Decision Tree:

While decision trees are indeed widely used in machine learning, it's important to note that they are not the only algorithm in the field. Decision trees are known for their simplicity, interpretability, and efficiency in handling large datasets. However, other machine learning algorithms, such as Support Vector Machines, Neural Networks, Random Forests, and Gradient Boosting, among others, also play significant roles in various applications.
Each algorithm has its strengths and weaknesses, and the choice often depends on the specific characteristics of the data and the goals of the task.
Decision trees are a type of supervised learning algorithm that makes decisions by recursively partitioning the data based on features. The process starts with a root node representing the entire dataset, and at each node, a decision is made based on a specific feature. This decision leads to branching, creating nodes for different outcomes.
the process continues until reaching endpoints or leaves, where the final decisions or predictions are made. Each path from the root to a leaf represents a unique decision-making process based on the features of the data. This structure makes decision trees easy to interpret and suitable for a variety of tasks.It consists of Nodes, Edges, and Leaf nodes.
Decision trees are non-parametric supervised learning techniques that can be working for classification. They create decision rules with a tree-like structure using actual data attributes. Decision Trees evolved from the way humans make decisions.[3]
The ID3 algorithm is used by training on a dataset to produce a decision tree which is stored in memory. At runtime, this decision tree is used to divide new hidden test cases by working down the decision tree using the values of this test case to come at a terminal node that tells you what class this test case belongs to. [6]
Determine the entropy for each attribute in the dataset. Partition the dataset into subsets based on the attribute with the minimum entropy (or equivalently, maximum information gain). Form a decision tree node incorporating that attribute. Recursively apply the process to subsets using the remaining attributes.

4. Artificial Neural Network:

Networks (ANN) is a machine learning algorithm inspired by the human brain. Typically, ANN relies on two main approaches: supervised and unsupervised methods.
The Unsupervised Neural Network is widely applied in fraud detection, boasting a sensitivity rate of 95%. It works by identifying patterns among current credit cardholders and those established in previous transactions, allowing it to recognize potential fraudulent activities.
If the information from current transactions corresponds to past ones, the likelihood of detecting a fraudulent case increases. Additionally, artificial neural network (ANN) methods exhibit significant fault tolerance.
For example, the production of output remains intact even in the presence of corruption in a single or multiple cells. Thanks to its high speed and efficient processing capabilities, artificial neural networks (ANN) can be deemed as a valuable solution for Credit Card Fraud Detection (CCFD).[8]

The ANN algorithm has two parts: Training part and testing part.

Training part:

Step 1: START
Step2: Loading and observing the dataset.

- pd.read.csv(.csv) # reads the dataset.
- resampling of data
- StandardScaler () #scaling and normalization of data

Step 3: Data pre-processing

- Train_test_split () #Splitting of data

Step 4: Training the model.

- Dense () #Adding data to activation function.

Step 5: Analysing the model.

- Prediction of fraud is made, and this trained data is stored .it can used to test (training the model takes longer time so it is stored)
- **Step 6**: STOP
- **Testing part:**

Def ANN
It is carried out similar way only difference is that the stored trained model is used to test the data and classify it.[5]
By using an artificial neural network (ANN) which gives accuracy approximately equal to 100% is best suited for credit card fraud detection It gives accuracy more than that of the unsupervised learning algorithms.

V. Evaluation measure

The final outcome undergoes evaluation using a confusion matrix, where precision, recall, and accuracy are computed. The confusion matrix revolves around two classes: the actual class and the predicted class. In the assessment of the end result, precision, recall, and accuracy are calculated based on a confusion matrix, which involves the actual class and predicted class within its two-class framework. [5]

	Actually Positive (1)	Actually Negative (0)
Predicted Positive (1)	True Positives (TPs)	False Positives (FPs)
Predicted Negative (0)	False Negatives (FNs)	True Negatives (TNs)

True Positive occurs when both values are positive (1), while True Negative is the scenario where both values are negative (0). False Positive refers to a situation where the true class is 0, but the predicted class is 1. False Negative is when the actual class is 1, but the predicted class is 0. In instances of True Positive, both values are positive (1); in True Negative, both values are negative (0). False Positive involves the true class being 0 with a predicted class of 1. False Negative occurs when the actual class is 1, but the predicted class is 0.

- **Precision defined as follows:**
  Precision = true positive / Actual result
  Precision = true positive/(true positive + false positive)
- **Recall defined as follows:**
  Recall = true positive / predicted result
  Recall = true positive/(true positive + false negative)
- **Accuracy defined as:**
  Accuracy = (true positive + true negative)/ total

VI. Conclusion:

This review paper explores various techniques employed in Credit Card Fraud Detection, highlighting Machine Learning as a valuable means to enhance accuracy. With fraud posing a significant challenge to the credit card industry, exacerbated by the rise of electronic transactions, implementing advanced Fraud Prevention and Detection methods becomes crucial. Machine Learning methods, informed by cardholder behavior, offer continuous improvement in fraud prevention accuracy. For those seeking web development services for financial applications vulnerable to fraud, it's advisable to engage a provider with substantial expertise in both web development and Machine Learning, particularly in the context of Fraud Detection.

VII. What to expect next?

Regrettably, even with the surge in technological layoffs at the close of 2022, criminals can exploit Machine Learning, as a pool of skilled talent remains jobless, potentially turning to fraudulent activities. Anticipate more sophisticated automated ML-powered attacks in 2023, surpassing basic velocity checks and leading to increased instances of multi-accounting fraud. The use of ML is also expected to drive customized phishing attacks, raising the prevalence of card fraud and identity theft. Disturbingly, reports indicate cybercriminals experimenting with ChatGPT for developing malware and fraud-enabling tools, posing threats to accounts, system access, and organizational infrastructure. At SPD Technology, over four years, we've provided AI solution development, empowering businesses to confront both physical and digital risks swiftly with advanced Machine Learning algorithms. Our vigilant experts monitor emerging threats, continuously integrating up-to-date security measures into the solutions they deliver.[2]

VIII. References:

[1] Bhatla, T. P., Prabhu, V., & Dua, A. (2003). Understanding credit card frauds. Cards business review, 1(6), 1-15. https://popcenter.asu.edu/sites/default/files/problems/credit_card_fraud/PDFs/Bhatla.pdf

[2] https://spd.tech/machine-learning/credit-card-fraud-detection/

[3] Ayorinde, K. (2021). A methodology for detecting credit card fraud [Master's thesis, Minnesota State University, Mankato]. Cornerstone: A Collection of Scholarly and Creative Works for Minnesota State University, Mankato. https://cornerstone.lib.mnsu.edu/etds/1168

[4] Sushant Kumbhar , Ashish Lade , Abhishek Patil , Jaykishan Pandey, A. B. Ghandat, 2023, Support Vector Machine based Credit Card Fraud Detection, INTERNATIONAL JOURNAL OF ENGINEERING RESEARCH & TECHNOLOGY (IJERT) Volume 12, Issue 03 (March 2023), https://www.ijert.org/support-vector-machine-based-credit-card-fraud-detection

[5] Asha RB, Suresh Kumar KR, Credit card fraud detection using artificial neural network, Global Transitions Proceedings, Volume 2, Issue 1, 2021, ISSN 2666-285X, https://doi.org/10.1016/j.gltp.2021.01.006.

[6] 1Vijayshree B. Nipane, 2Poonam S. Kalinge, 3Dipali Vidhate, 4Kunal War, 5Bhagyashree P. Deshpande, Fraudulent Detection in Credit Card System Using SVM & Decision Tree , Volume 1, Issue5,(2016), https://www.ijsdr.org/papers/IJSDR1605113.pdf

[7] Delamaire, L., Abdou, H., & Pointon, J. (2009). Credit card fraud and detection techniques: a review. Banks and Bank systems, 4(2), 57-68. https://eprints.hud.ac.uk/id/eprint/19069/1/AbdouCredit.pdf

[8] Bin Sulaiman, R., Schetinin, V. & Sant, P. Review of Machine Learning Approach on Credit Card Fraud Detection. Hum-Cent Intell Syst **2**, 55–68 (2022). https://doi.org/10.1007/s44230-022-00004-0

[9] Ileberi, E., Sun, Y. & Wang, Z. A machine learning based credit card fraud detection using the GA algorithm for feature selection. J Big Data **9**, 24 (2022). https://doi.org/10.1186/s40537-022-00573-8

[10] Razaque, A., Frej, M. B. H., Bektemyssova, G., Amsaad, F., Almiani, M., Alotaibi, A., ... & Alshammari, M. (2022). Credit Card-Not-Present Fraud Detection and Prevention Using Big Data Analytics Algorithms. Applied Sciences, 13(1), 57.

[11] I. D. Mienye and Y. Sun, "A Deep Learning Ensemble With Data Resampling for Credit Card Fraud Detection," in IEEE Access, vol. 11, pp. 30628-30638, 2023, doi: 10.1109/ACCESS.2023.3262020.

[12]S P, Maniraj & Saini, Aditya & Ahmed, Shadab & Sarkar, Swarna. (2019). Credit Card Fraud Detection using Machine Learning and Data Science. International Journal of Engineering Research and. 08. 10.17577/IJERTV8IS090031., https://www.researchgate.net/publication/336800562_Credit_Card_Fraud_Detection_using_Machine_Learning_and_Data_Science

[13]Madhurya, M. J., Gururaj, H. L., Soundarya, B. C., Vidyashree, K. P., & Rajendra, A. B. (2022). Exploratory analysis of credit card fraud detection using machine learning techniques. Global Transitions Proceedings, 3(1), 31-37.

[14]Mohsen, O. R., Nassreddine, G., & Massoud, M. (2023). Credit Card Fraud Detector Based on Machine Learning Techniques. Journal of Computer Science and Technology Studies, 5(2), 16-30.

[15]Kumar, S., Gunjan, V. K., Ansari, M. D., & Pathak, R. (2022). Credit Card Fraud Detection Using Support Vector Machine. In Proceedings of the 2nd International Conference on Recent Trends in Machine Learning, IoT, Smart Cities and Applications: ICMISC 2021 (pp. 27-37). Springer Singapore.