

---

## CREDIT CARD FRAUD DETECTION USING MACHINE LEARNING (KNN)

Tejas Jadhav\*<sup>1</sup>

\*<sup>1</sup>UG Student, Department Of Information Technology, B.K Birla College Kalyan,  
(Empowered Autonomous Status) Kalyan, Maharashtra, India.

---

### ABSTRACT

In this paper, we study the credit card fraud detection using KNN. we are in a world where cash/money transaction is on the verge of extension digital transaction is taking rapid growth, As there is a huge amount of online transaction in Day-to-Day life. Since credit cards are a common way to pay for things, fraud related to them is also increasing. Detecting a fraud is not easy like catching it but doing it quickly too. The rapid growth of fraudulent activities related to credit cards caused significant financial institutions. More people are trying to cheat the system with fake or unauthorized credit card transactions which is leading to financial. In this research, the dataset helps to know fraud transactions by using algorithms & models for development.

**Keywords:** Fraud Detection.

---

### I. INTRODUCTION

In today's rapidly evolving digital landscape, the convenience and ubiquity of credit have revolutionized the way we conduct financial transactions. While this convenience has undeniably enhanced our daily lives, it has also given rise to a significant challenge in the techniques to exploit vulnerabilities in the payment system, making it imperative for financial institutions, merchants, and individuals to employ advanced fraud detection methods.

Credit card fraud poses financial risks and the trust and security of the financial system. Detecting fraudulent transactions swiftly and accurately is crucial for safeguarding the interests of financial institutions, merchants and cardholders'-Nearest Neighbors (KNN) is a smart computer program that can help catch credit card fraud. It works by comparing each credit card transaction to others and finding the ones that look similar. This helps us find strange or suspicious transactions that might be fraudulent. In this study, putting a flash on KNN works for catching credit card fraud. Here is an explanation for basic ideas of what is and how it works for finding fraud. Here is an explanation of how we can use the KNN to get alerts from the fraud detection System. It covers what this means for real life, including the benefits and challenges. Credit card fraud detection is a critical application in the field of finance, where the objective is to identify and prevent fraudulent credit and transactions.

#### Types of fraud

- 1. card testing:** The one who often tests stolen credit card details by making small transactions. Machine learning can identify the transaction and anomaly detection.
- 2. First-party fraud:** This type of fraud involves individuals who intentionally apply for credit cards with the intent to commit fraud. Machine learning models can evaluate application data and behavior to identify potential first-party frauds.
- 3. card-present fraud:** Card-present fraud relates to in-person transactions such as fake credit cards or stolen. Location changes, multiple attempts of pins.
- 4. Amount takeover (ATO):** Here, attackers get access to the cardholder account & make it an unapproved transaction. It can monetarize the account login and password to detect innumerous behavior.

### II. METHODOLOGY

In research, the very first step includes collecting and preprocessing the dataset of a credit card. The dataset includes fraud as well as normal transactions which conclude the name of the customer, amount, location & time, whereas the dataset must be clean and standardized. In machine learning various libraries are included such as NumPy which is a Python library used for numerical & interchangeably. Pandas for different tasks, Scikit-Learn used for functionalities, Matplotlib for data visualization, seaborn visualization, and Pickle module for saving trained machine learning. Including these others standard libraries are also used for handling the dataset processing the data. Relevant features are selected to improve the model and performance and reduce

computational complexity Features such as transaction-specific attributes are considered The dataset is divided into training and testing sets. cross-validation techniques may be applied to ensure the models are general. The KNN algorithm is chosen for its suitability in detecting patterns in credit card transactions. The selected KNN model is a trained dataset, which is used to learn patterns and relationships within the data.

### Machine Learning

Machine learning is a branch of artificial intelligence (AI) & their model is used for evolving fraud pins. They can continuously learn from new data and adjust their prediction making them useful for detecting previously unseen types of fraud.

### Logistic regression

Logistic regression is a supervised machine learning algorithm used to predict the probability of a binary outcome or event such as whether a credit card transition is fraud or legitimate. The LR model uses the logistic function also known as the sigmoid function to transfer the linear combination of input features into a probability score.

### K-Nearest Neighbors (KNN)

k-Nearest Neighbors is a specific machine learning algorithm used to classify transactions as potentially based on their similarity to their known transactions. KNN is a type of instance-based or lazy learning algorithm that makes predictions by finding the k nearest data point in the training dataset to a given data point & classifying.

## III. MODELING AND ANALYSIS

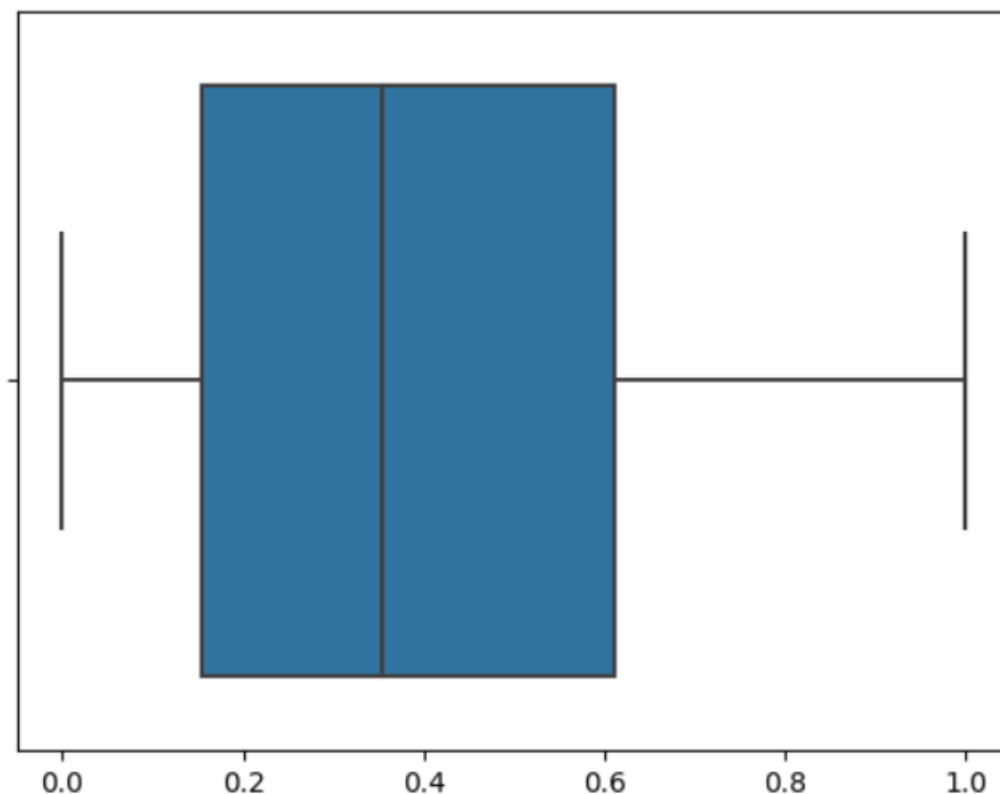


Figure 1: Plot Graph Of Credit Card Detection

## IV. RESULTS AND DISCUSSION

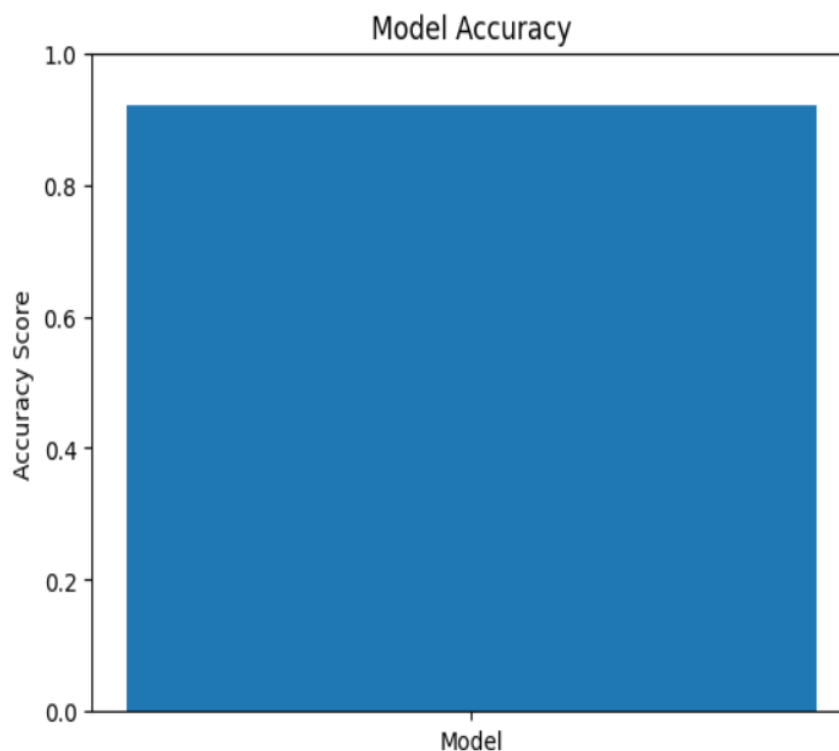
The model is to build a credit card fraud detection using the k-nearest neighbors (KNN) classification. The code begins by importing necessary, libraries, including NumPy, scikit-learn, Matplotlib, seaborn, and pickle. The data is loaded by naming "creditcard\_2023.csv" using pandas and store it in 'creditcard\_data' dataframe. In a dataset, the missing values are checked by the code. It extracts the feature from the dataset features stored in x & y variables which are normalized using min-max scaling to scale range between 0&1. Principal component analysis is used to reduce the dimensionality of the data. The dataset is splatted into training and testing sets

using 'train\_test\_split. A loop iterates over different numbers of neighbors to build KNN models with various numbers of neighbors, difference weighting schemes, and distance metrics. A KNN model makes predictions on the test data, and the accuracy is computed and printed.

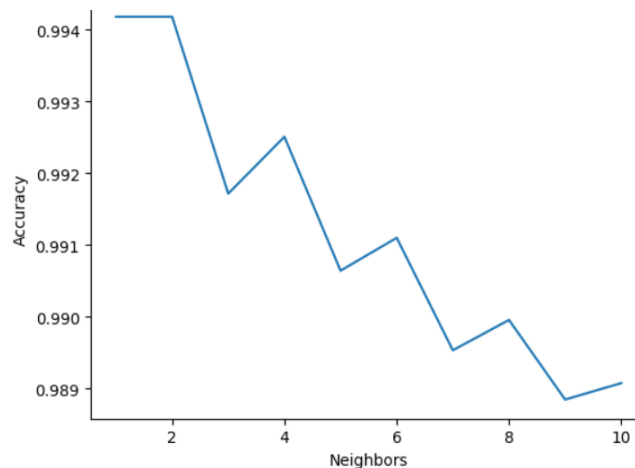
This code adds Gaussian noise to your training data and then fits the KNN model using the noisy data. It prints the accuracy with noisy data, which should be different from the accuracy without noise, it is likely to be lower.

**Table**

MODEL	ACCURACY
LOGISTIC REGRESSION	0.95
KNN	0.99



**Fig 2: Logistic Regression**



**Fig 3: KNN Model**

**V. CONCLUSION**

The code demonstrates the process of building a KNN-based credit card fraud detection model and logistic regression model with a focus on data preprocessing, dimensionality reduction, model training, and evaluation.

The accuracy of the KNN model on the test data is printed as a percentage, representing the model's ability to correctly classify fraud and non-fraud transactions. The code allows you to experiment with different model configurations, but it's important to note that intentionally reducing accuracy is generally not a goal in practical applications. In real-world scenarios, the aim is to build accurate fraud detection models to minimize financial losses and protect consumers.

## VI. REFERENCES

- [1] K. Chaudhary, J. Yadav, and B. Mallick, 'A review of Fraud Detection Techniques: Credit Card', Int. J. Comput. Appl., vol. 45.
- [2] A. Shen, R. Tong, and Y. Deng, 'Application of Classification Models on Credit Card Fraud Detection', in 2007 International Conference on Service Systems and Service Management, Changdu, China: IEEE, Jun. 2007, pp. 1-4. doi: 10.1109/ICSSSM.2007.4280163.
- [3] P. K. Chan and S. J. Stolfo, 'Toward Scalable Learning with Non-Uniform Class and Cost Distributions: A Case Study in Credit Card Fraud Detection'.
- [4] D. Varmedja, M. Karanovic, S. Sladojevic, M. Arsenovic, and A. Anderla, 'Credit Card Fraud Detection - Machine Learning methods', in 2019 18th International Symposium INFOTEH-JAHORINA (INFOTEH), East Sarajevo, Bosnia, and Herzegovina: IEEE, Mar. 2019, pp. 1-5. doi: 10.1109/INFOTEH.2019.8717766.
- [5] T. C. Tran and T. K. Dang, 'Machine Learning for Prediction of Imbalanced Data: Credit Fraud Detection', in 2021 15th International Conference on Ubiquitous Information Management and Communication (IMCOM), Seoul, Korea (South): IEEE, Jan. 2021, pp. 1-7. doi: 10.1109/IMCOM51814.2021.9377352.