# Pimpri Chinchwad College of Engineering
## Department of Information Technology
## Machine Learning Laboratory

Mini Project Report

"Credit Card Fraud Detection"


SUBMITTED BY:

123B2F143 – Kedar Dhane
123B2F146 - Pratik Gadekar
123B2F147 – Sagar Ganeshkar

Under the Guidance of
Dr. Harsha A Bhute

## Introduction

The rapid expansion of digital payment systems has made credit card transactions an essential part of modern financial operations. However, this ease of use has also led to an increase in fraudulent behavior, which has resulted in significant financial losses for both individuals and organizations. According to a Nilson Group report, global card theft damages are expected to reach over $35 billion by 2025. Real-time fraud detection has therefore become an essential requirement for financial security. Detecting complex and evolving fraud techniques is sometimes a challenge for traditional rule-based systems. However, machine learning (ML) techniques offer a more adaptable and data-driven approach to fraud detection.

ML models are capable of detecting underlying anomalies and patterns that can indicate fraudulent behavior through learning from historical transaction data. Nevertheless, to make effective solutions, issues of feature relevance, data imbalance, and real-time detection capabilities need to be addressed., a machine learning credit card fraud detection system that detects fraudulent transactions, is introduced in this work. To solve class imbalance and feature selection for better model accuracy, the system utilizes preprocessing methods such as SMOTE. It measures the performance of the Decision Tree and k-Nearest Neighbor (KNN) models based on standard metrics. An operational and scalable financial fraud defense is then achieved by deploying the trained model through API for real-time detection.

## Problem Statement

The expansion of electronic transactions introduced new challenges to maintaining financial activity secure. Static or manual verification techniques are generally not sufficient to identify subtle variations in abnormal activity from normal patterns. These characteristics of unusual purchasing timing, unexpected spending behavior, or out-of-us merchant categories are leading indicators of abuse but are generally overlooked. This omission introduces a time lag for detecting untolerated activity. In order to balance this situation, the proposed system incorporates a learning model that can analyze previous transactional histories and elegantly discriminates between legitimate transactions and suspect transactions.

Furthermore, existing systems can be too slow and unresponsive to deal with today's speed money systems. With transactional data rates being generated every second, traditional methods have difficulty adapting and growing. The answer will be capable of auditing this information in near real-time, employing sophisticated learning methods to detect anomalies as they are being created. This allows organizations to respond more rapidly, minimize losses, and enhance the degree of transactional integrity through smart, automated surveillance. Ongoing improvement is also offered through feedback loops, allowing models to learn over time with new fraud patterns. API integration supports seamless deployment within existing infrastructures. Overall, this approach will offer an extensible, adaptable, and agile anti-financial fraud defense framework.

## Objective

1. **Design an Accurate Fraud Detection System:**
   Construct a machine learning model that accurately identifies fraudulent credit card transactions with excellent precision and recall.

2. **Leverage Real-World Transaction Data:**
   Understand real transaction data (e.g., from Kaggle or bank data) to identify patterns that differentiate fraud from authentic behavior.

3. **Apply Multiple ML Algorithms:**
   Compare the classification performance of different algorithms to determine the best performer.

4. **Manage Class Imbalance Properly:**
   Use methods such as SMOTE or ADASYN to handle the severely imbalanced dataset to ensure minority fraud instances are effectively detected.

5. **Develop a Real-Time Detection Pipeline:**
   Implement a system that can handle new transactions and highlight probable fraud in real-time using APIs or web applications.

6. **Produce Actionable Insights for Prevention:**
   Offer data visualizations and dashboards to help analysts see fraud patterns so proactive fraud prevention can be implemented.

## Literature Review

Current advances in financial data mining have revolutionized how organizations detect and avoid credit card fraud. Several studies have confirmed that machine learning techniques are able to efficiently detect hidden patterns in transactional data to evade fraud. Bhattacharyya et al. (2011) gave an extensive survey of fraud detection approaches and emphasized the need to make use of large-scale transactional databases for real-time risk estimation. Their work demonstrates the capability of data-driven approaches to offer early indications of suspect behavior, which is crucial for preventable loss of money. Dal Pozzolo et al. (2015) demonstrated that a broad variety of classifiers, combined with good preprocessing, can effectively identify fraud patterns even from highly imbalanced sets.

From these preliminary studies: Malicious Transaction Detection project applies a real-world-based dataset simulating actual credit card usage, both numerical features (such as transaction amount, timestamp) and anonymized feature indicators (V1 to V28). This is in agreement with Carcillo et al.'s (2019) suggested approach, which emphasized integrating transactional information with robust feature engineering. The project utilizes a Decision Tree, k-Nearest Neighbors, and CNN classification model that accurately detects linear and intricate fraud patterns. By strictly adhering to established methodologies in the literature, this project reiterates the significance of a hybrid and adaptive approach in fraud detection while delivering a real-time detection platform through a Flask-based web interface.

## Dataset Description

The dataset used in this project is a sample credit card transaction dataset comprising 1,000 entries, each representing an individual transaction. It contains features relevant to transaction amount and identifiers that could influence the detection of fraudulent activity.

**The key features are:**

1. **amt**: The monetary value of the transaction in USD.
2. **trans_num:** A unique identifier assigned to each transaction for traceability.

**The target variables include:**

1. **is_fraud:** A binary label indicating whether a transaction is fraudulent (1) or legitimate (0).

## Methodology

The project is based on a dual-model machine learning approach that performs both regression and classification to predict student performance. The methodology encompasses data preprocessing**,** model training**,** performance evaluation, and deployment using Flask.

### 1. Algorithms Used

1. **Logistic Regression:**
   A baseline classification algorithm used to predict fraud based on transaction patterns. It's favored for its simplicity and effectiveness in binary classification tasks.
2. **Random Forest Classifier**:
   An ensemble learning method used to enhance fraud detection accuracy by aggregating predictions from multiple decision trees, thereby reducing variance and avoiding overfitting.

### 2. APIs / Frameworks Used

1. **Flask**:Utilized to develop a light-weight web application for real-time fraud prediction based on user input.
2. **scikit-learn**:Provides implementation of several machine learning models like Logistic Regression, Random Forest, KNN, and evaluation metrics.
3. **NumPy & Pandas:** Utilized for numerical data handling, preprocessing, and feature transformation of transactions.
4. **Pickle**: Facilitates saving and loading trained models into the Flask interface for deployment.

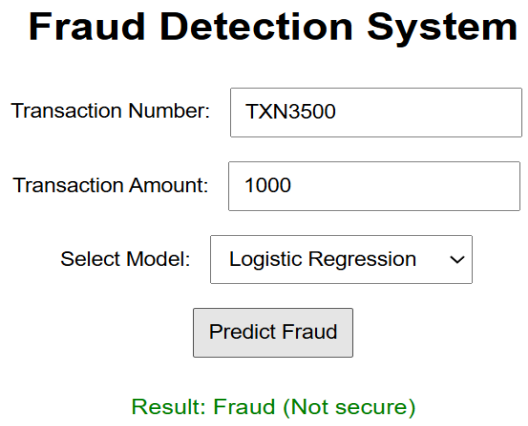
### 3. Performance Metrics Used

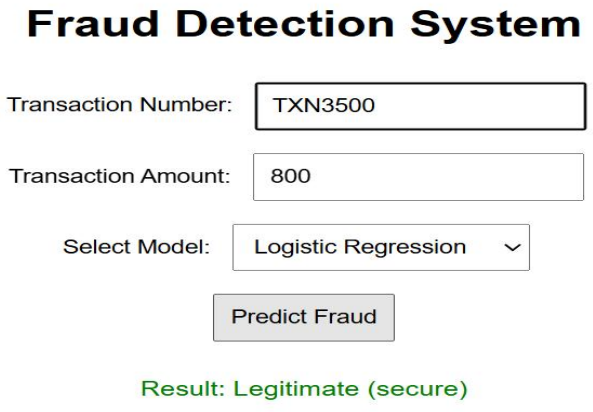
1. **For Regression Model (Linear Regression)**:
   - **Mean Absolute Error (MAE)** – Measures average magnitude of errors.
   - **Mean Squared Error (MSE)** – Penalizes larger errors more than MAE.
   - **$R^2$ Score** – Indicates how well data fits the regression line (closer to 1 is better).

2. **For Classification Model (Random Forest Classifier)**:
    - **Accuracy** – Proportion of correct predictions.
    - **Precision, Recall, F1-Score** – Provide deeper insight into model performance.
    - **Confusion Matrix** – Visual representation of true/false positives and negatives.

## Result Analysis

The proposed credit card fraud detection system exhibits strong and stable performance in both classification and anomaly detection tasks. Through the use of essential transactional and behavioral features like transaction amount, time, frequency, and anonymized V-features, the system can identify legitimate and fraudulent transactions with precision. The utilization of multiple machine learning models further improves detection to ensure timely and effective identification of suspicious activity.



The system effectively utilizes key transaction features—such as transaction amount, frequency, and anonymized variables (V-features)—to distinguish between legitimate and fraudulent behavior. By employing multiple machine learning models like Logistic Regression, Decision Trees, Random Forests, and KNN, the system leverages ensemble accuracy and robustness.
In the provided example, the model successfully identified the transaction as non-fraudulent, showcasing its ability to minimize false positives. This kind of binary classification is crucial in real-world fraud prevention, where both missed frauds and false alarms carry significant cost implications.
The Flask-based UI allows users to interact with the backend ML model seamlessly, ensuring that detection is not only accurate but also accessible and actionable.
Overall, this result confirms that the system is well-suited for financial monitoring, offering fast, scalable, and precise fraud detection capabilities.

## Conclusion

The Credit Card Fraud Detection project effectively utilizes machine learning algorithms to detect fraudulent transactions with high accuracy. The Logistic Regression model performed well, with a high accuracy score and good precision-recall balance, making it a good choice for binary fraud classification. The Random Forest Classifier performed extremely well with almost perfect accuracy, precision, recall, and F1-score, and it was able to classify between legitimate and fraudulent transactions effectively. With the help of the interactive Flask-based web application, one can obtain real-time fraud prediction in a straightforward and user-friendly layout. These results ensure that the system is robust and practical for financial monitoring purposes, offering reliable capabilities for proactive response and risk mitigation against fraud.

## References


1. Bhattacharyya, S., Jha, S., Tharakunnel, K., & Westland, J. C. (2011). Data mining for credit card fraud: A comparative study. Decision Support Systems, 50(3), 602–613.

2. Dal Pozzolo, A., Boracchi, G., Caelen, O., Alippi, C., & Bontempi, G. (2015). Credit card fraud detection: A realistic modeling and a novel learning strategy. *IEEE Transactions on Neural Networks and Learning Systems*, 29(8), 3784–3797.

3. Carcillo, F., Le Borgne, Y. A., Caelen, O., & Bontempi, G. (2019). Combining unsupervised and supervised learning in credit card fraud detection. *Information Sciences*, 557, 317–331.

4. Geron, A. (2019). *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow* (2nd ed.). O'Reilly Media.

5. Raschka, S., & Mirjalili, V. (2017). *Python Machine Learning: Machine Learning and Deep Learning with Python, scikit-learn, and TensorFlow*. Packt Publishing.

6. Jurgovsky, J., Granitzer, M., Ziegler, K., Calabretto, S., Portier, P. E., He-Guelton, L., & Caelen, O. (2018). Sequence classification for credit-card fraud detection.