# Detecting Payment Fraud Using Machine Learning

Project ID:24



| SRN: | Name: |
|---|---|
| PES2UG23CS361 | Naman Nagar |
| PES2UG23CS336 | Megha Dhanya |



## Problem Statement

Payment fraud poses a major financial threat, costing institutions billions each year. Detecting fraud in real time is challenging because fraudulent transactions form only a tiny fraction of total activity, creating a class imbalance problem. Traditional methods struggle to identify these rare events accurately. This project develops machine learning models—such as Logistic Regression, Random Forests, and Neural Networks—to distinguish legitimate and fraudulent transactions effectively using data balancing techniques like oversampling (SMOTE).

## Objective

To build a robust machine learning system capable of detecting fraudulent credit card transactions by applying classification models and using oversampling (SMOTE) to address class imbalance.

## High-Level Architecture

1. Data Preprocessing: Handle missing data, scale features, and separate labels.
2. Balancing: Apply SMOTE oversampling to handle class imbalance.
3. Model Training: Train multiple ML models like Logistic Regression, Random Forest, SVM, ANN, etc.
4. Evaluation: Compare models using Precision, Recall, F1-Score, and AUC.
5. Explainability: Use SHAP values to identify key contributing features.
6. Visualization: Display Confusion Matrices and ROC Curves for each model.

## Models Implemented

• Logistic Regression
• K-Nearest Neighbors
• Decision Tree
• Random Forest
• Gradient Boosting
• Naive Bayes

• Support Vector Machine
• Artificial Neural Network

## Results Summary

| **Model** | **Precision** | **Recall** | **F1-Score** | **AUC** |
|---|---|---|---|---|
| Random Forest | 1.0000 | 0.8750 | 0.9333 | 1.0000 |
| Gradient Boosting | 1.0000 | 1.0000 | 1.0000 | 1.0000 |
| SVM | 1.0000 | 0.7500 | 0.8571 | 0.9997 |
| ANN | 1.0000 | 0.7500 | 0.8571 | 0.9828 |

## Conclusion

Among all models, Random Forest demonstrated the most balanced and reliable performance with an AUC of 1.0000, combining high recall and precision. It effectively captured complex nonlinear patterns in transactions and minimized false alarms. SMOTE oversampling successfully mitigated class imbalance, enabling the model to learn from rare fraud cases.

## References

We used csv file from Kaggle.

**Credit Card Fraud Detection Dataset**
https://www.kaggle.com/mlg-ulb/creditcardfraud