Credit Card Fraud Detection Using Machine Learning – Project Documentation

Project by: Prajkta Chodankar  
Date: 22/04/2025

# Objective

Credit card fraud is a significant and growing global financial threat. The goal of this project is to build a machine learning model capable of accurately detecting fraudulent credit card transactions in order to mitigate financial loss and enhance security.

# Dataset Overview

• Source: Kaggle Dataset - Credit Card Fraud Detection(<https://www.kaggle.com/datasets/mlg-ulb/creditcardfraud>)  
• Size: 284,807 records and 31 features  
• Features:  
 - Time, Amount, Class  
 - V1 to V28 are anonymized numerical features (PCA-transformed to protect sensitive information)  
• Target Variable: Class  
 - 0: Legitimate  
 - 1: Fraudulent  
• Class Distribution: Highly imbalanced  
 - Fraud cases represent approximately 0.172% of the dataset.

# Data Preprocessing

1. Missing Values:  
 - No null values were found.  
2. Duplicate Records:  
 - Detected and removed 1,081 duplicate records.  
3. Data Split:  
 - 70% for training, 30% for testing.  
4. Imbalance Handling:  
 - Applied SMOTE (Synthetic Minority Over-sampling Technique).

# Exploratory Data Analysis (EDA)

• Distribution Plots:  
 - Visualized class imbalance (fraud vs. non-fraud).  
 - Analyzed transaction amount and time distributions.  
• Correlation Analysis:  
 - Created a heatmap to inspect correlations among features.

# Model Development & Selection

Multiple machine learning models were trained and evaluated:  
1. Logistic Regression  
2. K-Nearest Neighbors (KNN)  
3. Decision Tree  
4. Random Forest  
5. AdaBoost  
6. XGBoost

# Best Performing Model: K-Nearest Neighbors (KNN)

• Accuracy: 99.4%  
• F1-score: 81%  
• Why KNN?  
 - Robust to overfitting  
 - Balanced performance on both fraud and non-fraud classes  
 - High recall and precision for the minority (fraud) class

# Conclusion

• Successfully built a high-performance fraud detection model using machine learning.  
• Applied data balancing techniques to improve fraud detection.  
• Achieved high accuracy and robustness with the KNN model.  
• The model is ready to support real-world anti-fraud operations.