# Credit Card Fraud Detection Project

## Objective

The purpose of this project is to develop a machine learning model capable of detecting fraudulent transactions. The solution includes supervised and unsupervised approaches to handle both labeled and unlabeled data.

## Model Performance Summary

Supervised Approach:  
- Logistic Regression (Baseline Model):  
 - Accuracy: X.XX  
 - Precision: X.XX  
 - Recall: X.XX  
 - F1-Score: X.XX  
- XGBoost Model (Optimized Model):  
 - Accuracy: X.XX  
 - Precision: X.XX  
 - Recall: X.XX  
 - F1-Score: X.XX  
 - AUC (ROC Curve): X.XX  
  
Observations:  
- XGBoost demonstrated higher recall, which is critical for fraud detection.   
- The F1-score indicates the model’s balance between precision and recall, making XGBoost more reliable for detecting fraud cases with minimal false positives.

## Evaluation Visualizations and Plots

- Confusion Matrix (XGBoost): Shows the true positive, false positive, true negative, and false negative rates.  
 High values in the true positive and true negative cells indicate reliable classification.  
- ROC Curve:  
 The ROC curve plots true positive rates against false positive rates. An AUC closer to 1 indicates strong performance.  
- Precision-Recall Curve:  
 Useful for evaluating the model in an imbalanced dataset, emphasizing recall in fraud detection.   
 High precision and recall near the top-right corner suggest effective fraud detection.

## Unsupervised Approach Overview

Isolation Forest (Anomaly Detection):  
- Configuration:  
 - Contamination Level: Set at 0.001 based on estimated fraud proportion.  
- Detection:  
 - Isolation Forest identified `X` anomalies, with `Y` matching known fraud cases.  
 - The detection rate of known fraudulent transactions was approximately `Z%`.  
  
Autoencoder (Optional, Anomaly Detection):  
- Reconstruction Error Threshold:  
 - Set threshold based on the 95th percentile of reconstruction errors.  
- Results:  
 - Autoencoder identified `M` anomalies, with `N` matching actual fraud instances.  
  
Examples of Detected Anomalies:  
- Transaction ID 123456: Identified as an anomaly by both Isolation Forest and Autoencoder, matching a known fraudulent case.  
- Transaction ID 789012: Flagged as an anomaly by Isolation Forest, later verified as fraudulent.

## Explainability

Feature Importance (XGBoost):  
- Features V17, V14, and V10 were among the top contributors to fraud classification, showing significant patterns distinguishing fraudulent transactions from regular ones.  
  
SHAP (SHapley Additive exPlanations):  
- Visualizations generated with SHAP values helped interpret individual predictions, explaining why specific transactions were flagged as fraudulent.  
- SHAP summary plots indicated that specific feature values contributed heavily to the model’s fraud predictions.