# Fraud Detection Project Overview

## Objective:

To detect fraudulent transactions using various machine learning techniques, ensuring high accuracy, precision, and recall to minimize financial losses and enhance security.

## Procedure:

### Data Cleaning & Preparation:

Ensured the dataset was accurate and ready for analysis by handling missing values and correcting data types.

### Feature Engineering:

Selected key features with strong correlations to fraud, including V4, V11, V12, V14, V17, and V18.

### Anomaly Detection:

Model Used: Isolation Forest

Why Chosen: Effective for detecting outliers and anomalies in high-dimensional datasets.

Results: Identified 5,669 anomalies out of 170,589 transactions.

### Pattern Recognition:

Model Used: Random Forest

Why Chosen: Robust, handles large datasets, and provides feature importance insights.

## Results:

#### Confusion Matrix:

[[85075 74]

[ 6 85434]]

Classification Report:

Class 0 (Not Fraud):

Precision: 1.00

Recall: 1.00

F1-Score: 1.00

Class 1 (Fraud):

Precision: 1.00

Recall: 1.00

F1-Score: 1.00

Overall Accuracy: 100%

### Predictive Analytics:

Model Used: Logistic Regression

Why Chosen: Simple, interpretable, and effective for binary classification problems.

#### Results:

#### Confusion Matrix:

[[83318 1831]

[ 5116 80324]]

**Classification Report:**

**Class 0 (Not Fraud):**

Precision: 0.94

Recall: 0.98

F1-Score: 0.96

**Class 1 (Fraud):**

Precision: 0.98

Recall: 0.94

F1-Score: 0.96

Overall Accuracy: 96%

## Cross-Validation:

Purpose: To ensure model robustness and prevent overfitting.

Method: 5-fold cross-validation for the Random Forest model.

Results:

Cross-validation scores: [0.9988, 0.9998, 0.9990, 0.9996, 0.9996]

Mean cross-validation score: 0.9994

### Combining predictions:

Improved Accuracy: Combining predictions can reduce the likelihood of errors that any single model might make. If one model misses fraud, the other might catch it, leading to better overall performance.

Robustness: Combining models can make your prediction system more robust and reliable

by mitigating the biases or limitations inherent in any one model.

Balanced Performance: Averaging helps balance precision and recall, providing a more

stable prediction system.

This ensemble approach enhances the robustness of your fraud detection system, ensuring

higher level of accuracy and reliability.

## Key Takeaways:

Accuracy: Achieved 96% accuracy with high precision and recall, indicating robust model performance.

Balanced Performance: Maintained high F1-scores for both fraudulent and non-fraudulent transactions.

Generalization: Cross-validation confirmed the model's ability to generalize well to new data, reducing the risk of overfitting.

Anomaly Detection: Effectively used Isolation Forest to flag potentially fraudulent transactions for further investigation.

Feature Importance: Identified key features contributing most to fraud detection, enhancing model interpretability.