Anomaly Detection Report on Credit Card Fraud Dataset

# 🧾 Overview

This report summarizes the results of anomaly detection using Isolation Forest and Local Outlier Factor (LOF) on the Credit Card Fraud Detection dataset. We compared their performance with and without Principal Component Analysis (PCA), evaluating how dimensionality reduction affects fraud detection.

# 📊 Dataset Summary

The dataset contains 284,807 transactions, with only 492 labeled as fraudulent (~0.17%). The features V1 to V28 are already PCA-transformed anonymized variables, with additional features like 'Amount' and 'Time'.

# 🔍 Model Results & Analysis

## Isolation Forest

### Without PCA

Fraud Recall: ~0.25  
Output:  
 precision recall f1-score support  
 0 1.00 1.00 1.00 284315  
 1 0.26 0.25 0.26 492  
Overall: Detected ~25% of frauds while maintaining high precision for normal transactions.

### With PCA (2 Components)

Fraud Recall: 0.00  
Output:  
 1 0.00 0.00 0.00 492  
All frauds were missed. Reducing to 2D PCA removed key fraud signal — not recommended.

### With PCA (95% Variance Retained)

Number of Components Retained: 27  
Fraud Recall: ~0.28  
Output:  
 1 0.28 0.28 0.28 492  
Best result: Balanced dimensionality reduction while preserving fraud-detection capability.

## Local Outlier Factor (LOF)

### Without PCA

Fraud Recall: 0.00  
Output:  
 1 0.00 0.00 0.00 492  
LOF failed to detect any frauds in the original feature space.

### With PCA (95% Variance Retained)

Fraud Recall: 0.00  
Output:  
 1 0.00 0.00 0.00 492  
Even with more informative features retained, LOF still could not detect any frauds.

# ✅ Key Takeaways

- Isolation Forest performed well, especially when using PCA with 95% variance.  
- Using only 2 PCA components severely hurt performance.  
- LOF was not effective at all for this dataset — likely due to the nature of fraud not aligning with local density-based outliers.  
- PCA is useful if done carefully: retain enough variance to keep rare signals intact.