

# Report

## Credit Card Fraud Detection Using Anomaly Detection

### PROJECT TEAM :

#### Group Members:

Jallepalli Dwaraka - Data Collection

Kushal Mangal– Data preprocessing

Nishant Prajapati - Exploratory Data Analysis

Kunal Yadav - Model Development & Tuning

Mahi Uddhavrao Parate - Results Analysis & Documentation

### EXECUTIVE SUMMARY :

This project implements Isolation Forest algorithm for credit card fraud detection on a highly imbalanced dataset of 284,807 transactions. Our optimized model achieves 57% fraud recall with ROC AUC of 0.947, demonstrating strong anomaly detection capabilities. The analysis reveals critical temporal patterns where fraud rates peak at 1.4% during early morning hours (1-4 AM), providing actionable insights for real-time fraud prevention systems.

### 1. DATASET OVERVIEW:

**Source:** European credit card transactions (September 2013)

**Size:** 284,807 transactions with 31 features

**Time Period:** 48 hours of transaction data

**Class Distribution:**

- Normal: 284,315 (99.83%)

- Fraud: 492 (0.17%)

#### Features:

- V1-V28: PCA-transformed features (privacy protection)

- Time: Seconds elapsed between transactions

- Amount: Transaction value in dollars

- Class: Binary target (0=Normal, 1=Fraud)

**Data Quality:** Zero missing values, clean dataset ready for modeling.

## **2. KEY FINDINGS FROM EXPLORATORY ANALYSIS :**

### **2.1 Transaction Amount Patterns**

- Average transaction: \$88.47
- Median transaction: \$22.00
- Distribution: Highly right-skewed with 75% of transactions under \$77.51
- Fraud vs Normal: Fraudulent transactions show more variability in amount distribution

### **2.2 Temporal Fraud Patterns (Critical Discovery)**

- High-Risk Period: Hours 0-5 (Midnight to 5 AM)
- Peak fraud rate: 1.4% at 1 AM
- Why it matters: 8x higher than daytime average (0.17%)
- Business impact: Targeted monitoring during off-peak hours
- Low-Risk Period: Hours 9-23 (Day and Evening)
- Fraud rate: Stable at 0.05-0.2%
- Insight: Normal business hours show baseline fraud activity

### **2.3 Class Imbalance Challenge**

- Extreme imbalance (492 frauds in 284,807 transactions) requires anomaly detection techniques

## **3. ANOMALY DETECTION MODEL: ISOLATION FOREST :**

Why Isolation Forest?

- Unsupervised approach: Doesn't require balanced training data
- Anomaly-focused: Specifically designed to isolate outliers
- Scalability: Efficient on large datasets (300 trees trained)
- Interpretability: Provides anomaly scores for threshold tuning

Model Configuration:

- Algorithm: Isolation Forest
- Estimators: 300 trees
- Contamination: Tuned from 0.00167 to 0.01
- Features: 30 (V1-V28 + scaled Amount + scaled Time)

## **4. PERFORMANCE RESULTS :**

Baseline Model (Contamination = 0.00167)

- Precision: 0.239
- Recall: 0.239
- F1-Score: 0.239

- ROC AUC: 0.947

Confusion Matrix:

- TN: 282,893, FP: 360, FN: 360, TP: 113

Optimized Model (Contamination = 0.01)

- Precision: 0.096

- Recall: 0.575

- F1-Score: 0.164

- ROC AUC: 0.947

Business Impact:

- 272 frauds detected (vs 113 baseline)

- 159 additional frauds caught = significant loss prevention

- Trade-off: More false alarms require investigation resources

Threshold Analysis:

- 40% Recall → 14.5% Precision

- 50% Recall → 10.4% Precision (Recommended)

- 60% Recall → 10.0% Precision

## **5. VISUAL INSIGHTS :**

- PCA Visualization: Fraudulent transactions separate from normal clusters.

- Precision-Recall Curve: Demonstrates recall-precision trade-off.

## **6. BUSINESS RECOMMENDATIONS :**

Implementation:

- Deploy optimized model (contamination=0.01)

- Focus monitoring during 0-5 AM high-risk hours

- Two-tier detection: automatic blocks + manual review

ROI:

- Baseline detection: \$28,250 saved

- Optimized detection: \$68,000 saved

- Additional prevention: \$39,750 per 48 hours

Operational Considerations:

- False positive rate: ~1%

- Review workload: ~2,800 flagged transactions / 48 hrs

## **7. TECHNICAL METHODOLOGY :**

- Preprocessing: Scaling, duplicate removal
- Model Training: Full dataset (unsupervised)
- Evaluation Metrics: Recall, Precision, ROC AUC

## **8. LIMITATIONS :**

Current Limitations:

- Recall ceiling: 57%
- Precision trade-off: Higher detection = more false positives
- Model trained on 48-hour dataset

## **9. CONCLUSION :**

Isolation Forest achieves 0.947 ROC AUC and 57% fraud recall.

Key discoveries: Temporal fraud patterns, tunable system, actionable recommendations.

Potential savings: \$39,750 per 48 hours of fraud prevention.