# Credit Card Fraud Detection - Project 11

## 1. Introduction

This report analyzes a credit card transactions dataset to detect fraudulent activities. Machine learning models including Logistic Regression, Random Forest, and Isolation Forest were used. Handling of class imbalance was critical due to the rarity of fraud cases.

## 2. Key Questions

- What percentage of transactions are fraudulent?  
- Which features are most indicative of fraud?  
- How do different ML models perform in detecting fraud?  
- Can we reduce false positives without missing fraud cases?

## 3. Tasks Performed

- Loaded and explored the dataset.  
- Handled class imbalance using undersampling.  
- Trained Logistic Regression, Random Forest, and Isolation Forest models.  
- Evaluated models using precision, recall, F1-score, and ROC curves.  
- Visualized confusion matrices for performance comparison.

## 4. Analysis Approach

The dataset was highly imbalanced, with less than 0.2% of the transactions being fraudulent. We balanced the dataset using random undersampling of the majority class. Features were standardized, and models were trained on the balanced data. Model performance was evaluated using classification metrics and visualizations.

## 5. Findings

- Fraudulent transactions constituted less than 0.2% of all transactions.  
- Random Forest outperformed Logistic Regression in terms of precision and recall.  
- Isolation Forest detected anomalies but with lower precision compared to supervised models.  
- ROC-AUC scores indicated strong separation between fraudulent and non-fraudulent transactions.

## 6. Conclusion

Credit card fraud detection requires careful handling of imbalanced data. Supervised learning models like Random Forest are effective for detecting fraudulent transactions, while anomaly detection methods can provide additional support. Ongoing model evaluation and tuning are essential to maintain high detection rates and minimize false positives.