# Credit Card Fraud Detection Report

## 1. Introduction

This project aims to detect fraudulent credit card transactions using machine learning techniques. The dataset consists of anonymized transaction data, and we evaluate multiple models to determine the best fraud detection approach.

## 2. Dataset Overview

The dataset contains transaction details with a balanced distribution of fraudulent and non-fraudulent transactions. The key steps include data exploration, preprocessing, model training, evaluation, and insights.

## 3. Data Preprocessing

- Feature scaling was applied to normalize transaction amounts.  
- The dataset was split into training (80%) and testing (20%) sets.  
- No SMOTE (Synthetic Minority Over-sampling Technique) was used as the dataset was already balanced.

## 4. Model Training & Performance

### 4.1 Logistic Regression

Logistic Regression is a simple and efficient model for binary classification. We applied 5-fold cross-validation and achieved the following results:  
- ROC-AUC Score (CV): 0.9935  
- Accuracy: 96%  
- F1-Score: 96%  
While efficient, this model may not capture complex fraud patterns.

### 4.2 Random Forest

Random Forest is an ensemble method that uses multiple decision trees. To prevent overfitting, we reduced the model complexity (max depth = 10, n\_estimators = 50). The results showed:  
- ROC-AUC Score (CV): 0.9999  
- Accuracy: 100%  
- F1-Score: 100%  
While achieving high performance, this model may still be overfitting.

### 4.3 XGBoost

XGBoost is a boosting algorithm known for its efficiency and accuracy. We fine-tuned the model using cross-validation and reduced complexity:  
- ROC-AUC Score (CV): 0.9999  
- Accuracy: 100%  
- F1-Score: 100%  
This model provided the best performance but also showed signs of overfitting.

### 4.4 Model Comparison

| Model | Accuracy | ROC-AUC Score | F1-Score |  
|---------------------|----------|--------------|----------|  
| Logistic Regression | 96% | 0.9935 | 96% |  
| Random Forest | 100% | 0.9999 | 100% |  
| XGBoost | 100% | 0.9999 | 100% |  
Both Random Forest and XGBoost show perfect accuracy, indicating possible overfitting.

## 5. Insights & Recommendations

- \*\*Fraud Detection is Highly Effective\*\*: Machine learning models provide high accuracy for fraud detection.  
- \*\*Random Forest & XGBoost Might Be Overfitting\*\*: The models achieve 100% accuracy, suggesting they may not generalize well to unseen data.  
- \*\*Logistic Regression is More Practical\*\*: Given its speed and decent performance, Logistic Regression is suitable for real-time fraud detection.  
- \*\*Business Strategy\*\*: Use Logistic Regression for low-risk transactions and XGBoost for high-value transactions requiring deeper analysis.

## 6. Deployment Strategy

For real-world implementation, the fraud detection model should be deployed as an API integrated into the payment processing system. The steps include:  
1. \*\*Train & Save the Model\*\*: Convert the trained model into a `.pkl` file.  
2. \*\*Build a Fraud Detection API\*\*: Use Flask or FastAPI.  
3. \*\*Real-Time Monitoring\*\*: Deploy on AWS or Azure to analyze transactions live.

## 7. Conclusion

This project successfully demonstrated that machine learning can effectively detect fraudulent transactions. While XGBoost and Random Forest provided perfect accuracy, Logistic Regression remains a practical choice for real-time fraud detection. Further work includes hyperparameter tuning and real-world deployment.