**Fraud Detection in Rent Payments: Methodology and Model Selection**

### **Abstract**

Fraud detection in financial transactions is a critical challenge in today's digital economy. This study aims to develop a machine learning-based approach for identifying fraudulent rental payments. By leveraging supervised and unsupervised learning models, we analyze transaction patterns to distinguish fraudulent activities from genuine transactions. The study explores various techniques for handling imbalanced datasets and evaluates model performance using industry-standard metrics. Future enhancements focus on advanced feature engineering and real-time deployment for improved fraud detection accuracy.

### **Introduction**

Detecting fraudulent transactions in rental payments is essential to prevent financial losses and enhance security in digital payments. Traditional rule-based fraud detection systems are insufficient against evolving fraudulent techniques. Machine learning provides a more adaptive and intelligent approach to identifying anomalies. This study investigates multiple models, including XGBoost and Isolation Forest, to detect fraudulent patterns effectively. The research emphasizes handling data imbalance, feature selection, and evaluation metrics to ensure accurate predictions and minimal false positives.

### **Methodology**

The goal of this project is to detect fraudulent rental transactions using machine learning techniques. The dataset used for this task includes financial transactions labeled as fraudulent or genuine. The key steps followed in this study are:

1. **Data Collection & Preprocessing**
   1. Datasets from Kaggle (PaySim, IEEE-CIS, Credit Card Fraud) were considered.
   2. Features such as transaction amount, balance changes, and transaction type were extracted.
   3. Missing values in the dataset were handled by removing rows with NaN values.
   4. Standardization was applied to numerical features for better model performance.
2. **Handling Imbalanced Data**
   1. Fraudulent transactions were underrepresented in the dataset.
   2. SMOTE (Synthetic Minority Over-sampling Technique) was used to balance the dataset.
   3. Class weighting in model loss functions was also explored.
3. **Model Selection**
   1. **Supervised Models:**
      1. **Logistic Regression & XGBoost:** XGBoost was chosen for its high performance in fraud detection tasks.
   2. **Unsupervised Models:**
      1. **Isolation Forest & One-Class SVM:** Used for anomaly detection in cases where labeled fraud data was scarce.
4. **Evaluation Metrics**
   1. **Precision, Recall, and F1-score** were used to handle the precision-recall tradeoff.
   2. **ROC-AUC Score** was used to assess classification performance.

### **Improvements & Future Optimization**

* **Feature Engineering Enhancements:**
  + Introduce time-series analysis to detect abnormal transaction patterns over time.
  + Use geolocation and device-based anomalies for better fraud detection.
* **Model Optimization:**
  + Hyperparameter tuning for XGBoost and Isolation Forest.
  + Use ensemble learning to combine multiple models for better accuracy.
* **Deployment Considerations:**
  + Deploy the trained model on a cloud-based platform for real-time fraud detection.
  + Implement automated model retraining with fresh transaction data to improve performance continuously.

This study demonstrates the feasibility of machine learning models in fraud detection and highlights future directions to enhance accuracy and reliability.