Fraud Detection In Rent Payments

Task: Develop an ML model to detect fraudulent rental transactions using historical tenant behavior and transaction data.¶

Fraud Detection in Rent Payments: A Supervised & Unsupervised Learning Approach

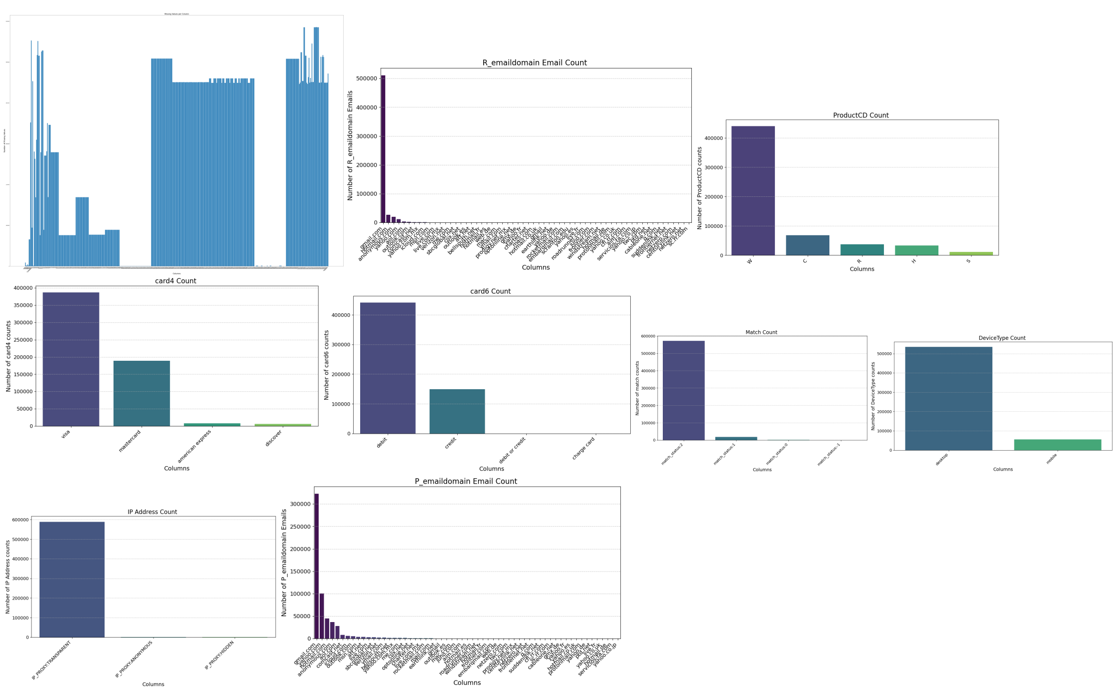
1. Introduction

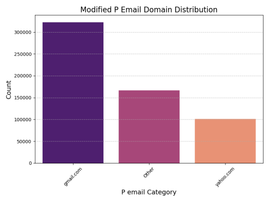
To detecting fraudulent rental transactions is critical for ensuring secure financial transactions. This study explores a **hybrid approach** that combines both **supervised** and **unsupervised** machine learning models to identify fraudulent activities.

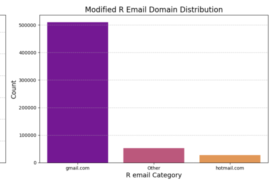
2. Feature Engineering

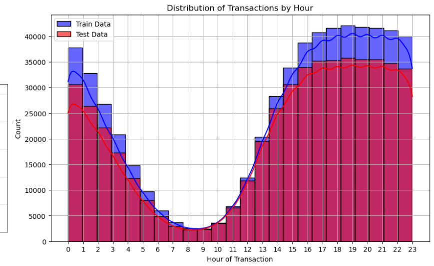
Feature selection and engineering play a crucial role in identifying fraudulent patterns. The key extracted features include:

- **Time-based spending patterns**: Identifying irregular spending behaviors at unusual hours
- **Transaction frequency analysis**: Detecting abnormally high transaction volumes over short periods.
- **Transaction location anomalies**: Comparing transaction locations to known user behaviors and detecting mismatches.

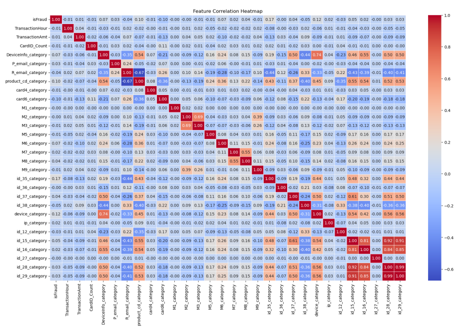








Correlation



3. Handling Imbalanced Data

Fraud detection datasets are highly **imbalanced**, with fraudulent transactions being rare. To counter this, we apply:

- **SMOTE (Synthetic Minority Over-sampling Technique)**: Generated synthetic fraud cases to balance the dataset.
- **Class Weighting in Loss Function**: Adjusted model training to prioritize minority class detection.

4. Supervised + Unsupervised Learning

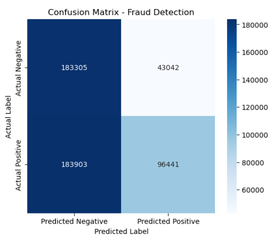
Since we did not have the ground truth, to achieve effective fraud detection, we used Isolation Forest for creating anomaly of 2 groups (0-non-fraud, 1-fraud). The anomaly is considered as ground truth for Supervised Learning of the architecture and then have applied XGBoost to measure the effectiveness

- **Unsupervised Learning Models:**
  - **Isolation Forest**: Detects anomalies by isolating fraudulent transactions.
- **Supervised Learning Models: Performed Supervised learning on the**
  - **XGBoost**: A robust, tree-based model that effectively handles structured data and imbalanced datasets.

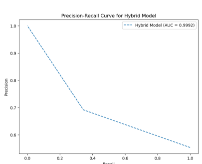
5. Evaluation Metrics and Precision-Recall Tradeoff for XGBoost:

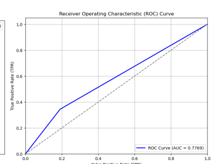
Given the imbalanced nature of fraud detection, traditional accuracy is insufficient. Instead, we evaluate models using:

Confusion Matrix



- **F1-Score**: Balances Precision and Recall to measure overall effectiveness.
- **Precision-Recall Curve & AUC-ROC**:
    - **Precision**: Measures the percentage of predicted fraud cases that are actual fraud.
    - **Recall**: Measures the ability to detect all fraudulent cases.
    - **AUC-ROC Score**: Assesses the model's ability to distinguish between fraudulent and non-fraudulent transactions.





Model Accuracy and Loss-

Model Accuracy: 0.9121
Model Log Loss: 0.7438