

Mobile Phone Transactions Fraud Detection

Chujun Chen

Brown Data Science Institute

Oct, 18 2023

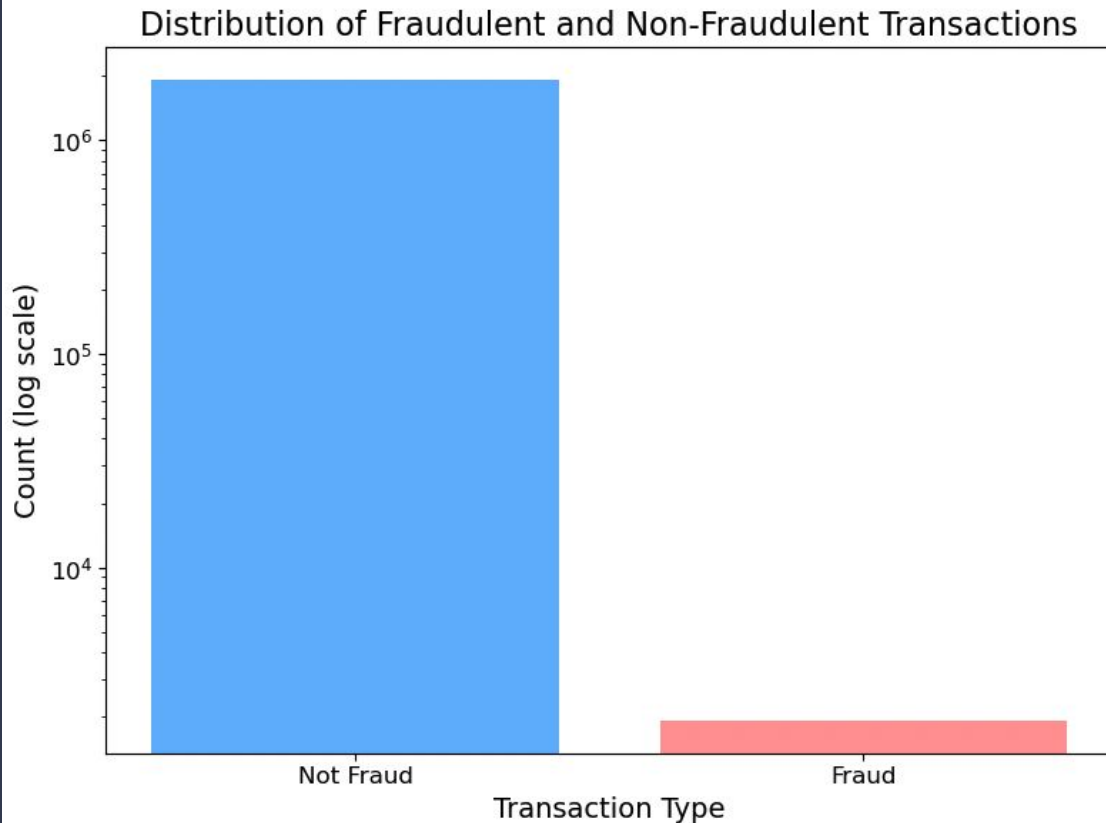
Github: <https://github.com/Christina-Chen01/DATA1030-MidtermProject-Fraud-Detection>

Classify a Transaction Based on Historical Transaction Patterns and Features

- Importance:
 - Ensure Financial Security
 - Prevent Financial Loss
 - Maintain trust in mobile platforms
- Characteristics:
 - 1.5M+ records
 - Non-iid (time series)
- Data Source: PaySim synthetic dataset on Kaggle
- Data Collection:
 - Simulates real transactions from a global mobile financial service provider.

EDA

Figure 1. Highlights the pronounced imbalance in the 'isFraud' target variable, where a mere 0.001% of transactions are labeled as fraudulent.



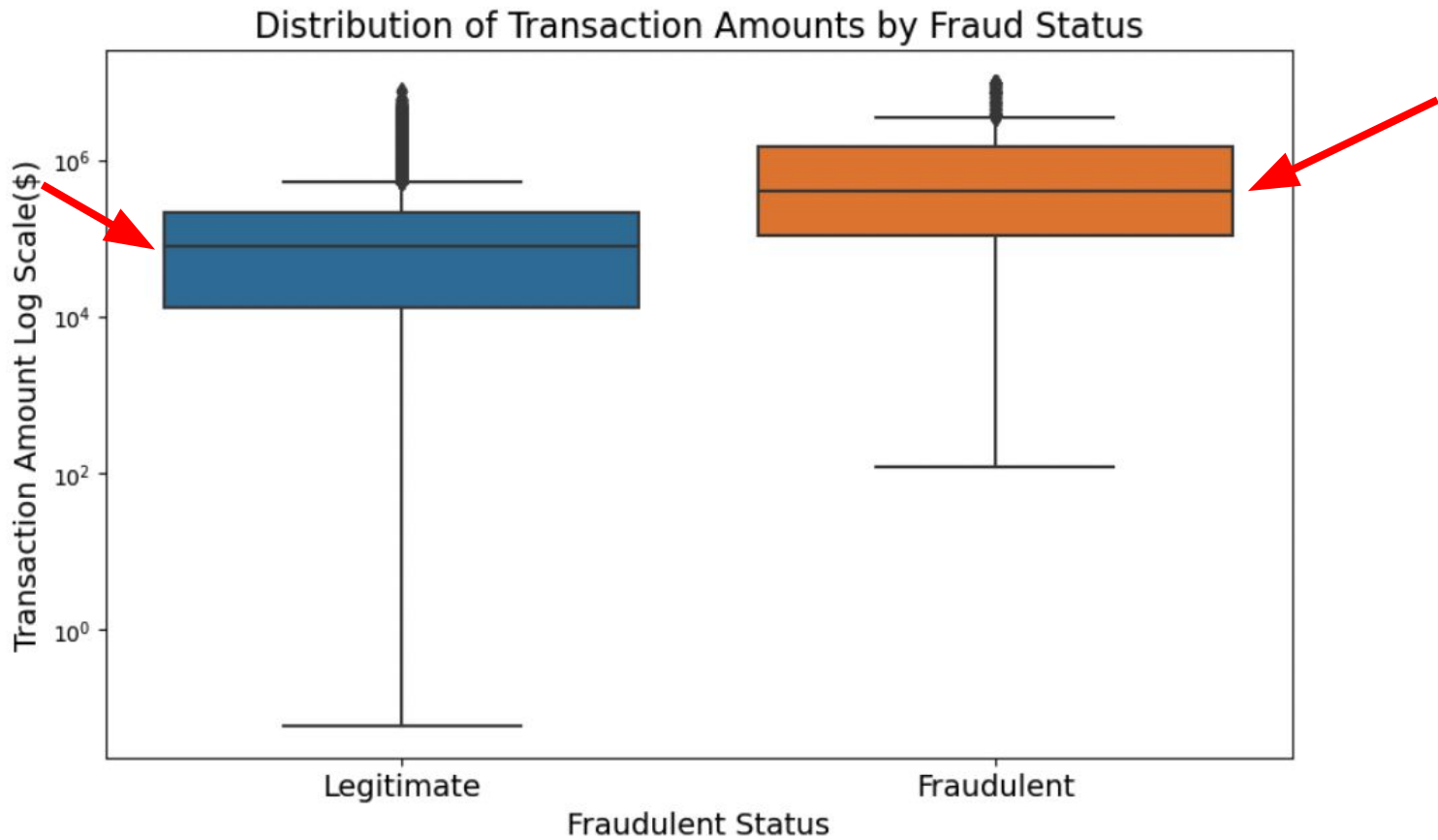


Figure 2. highlights a higher median of transaction amount for fraudulent transactions than its legitimate counterpart.

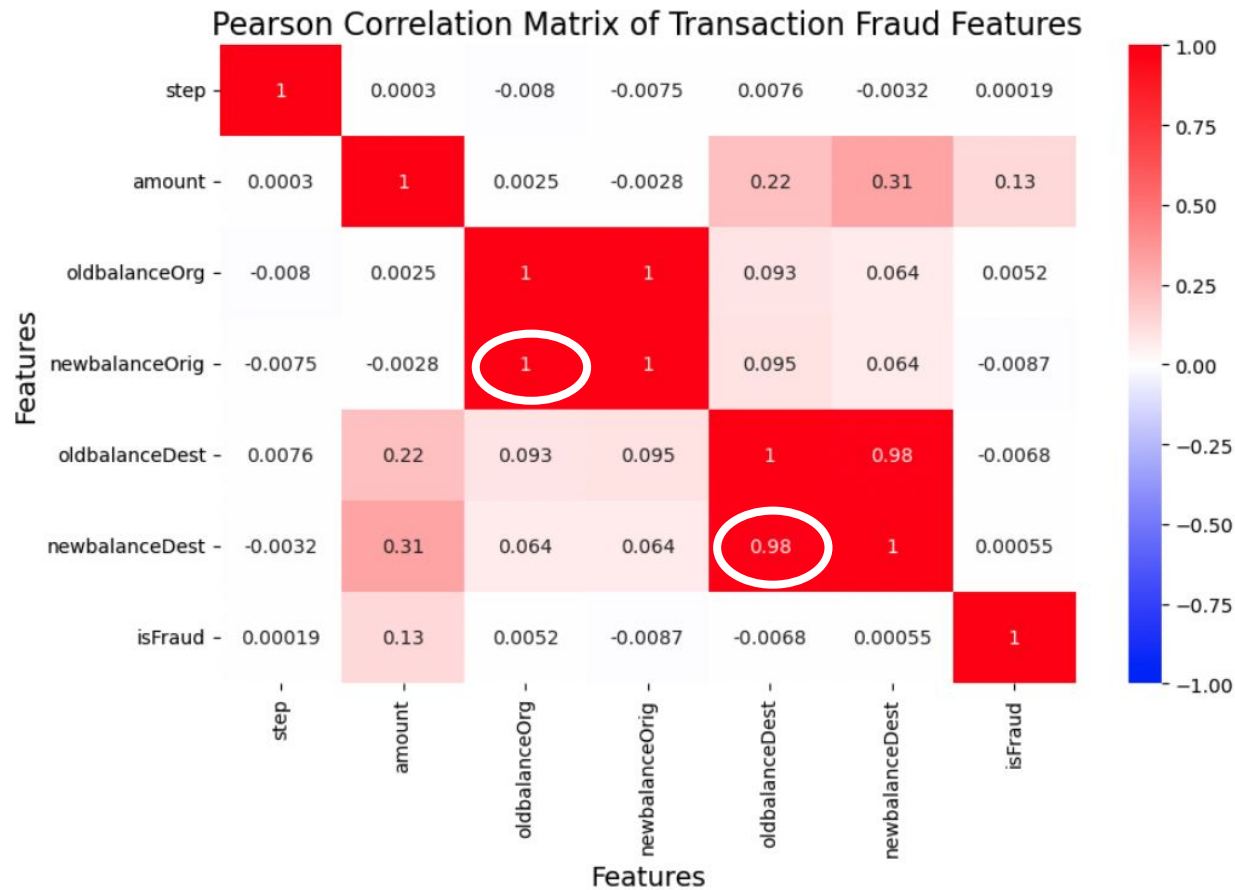
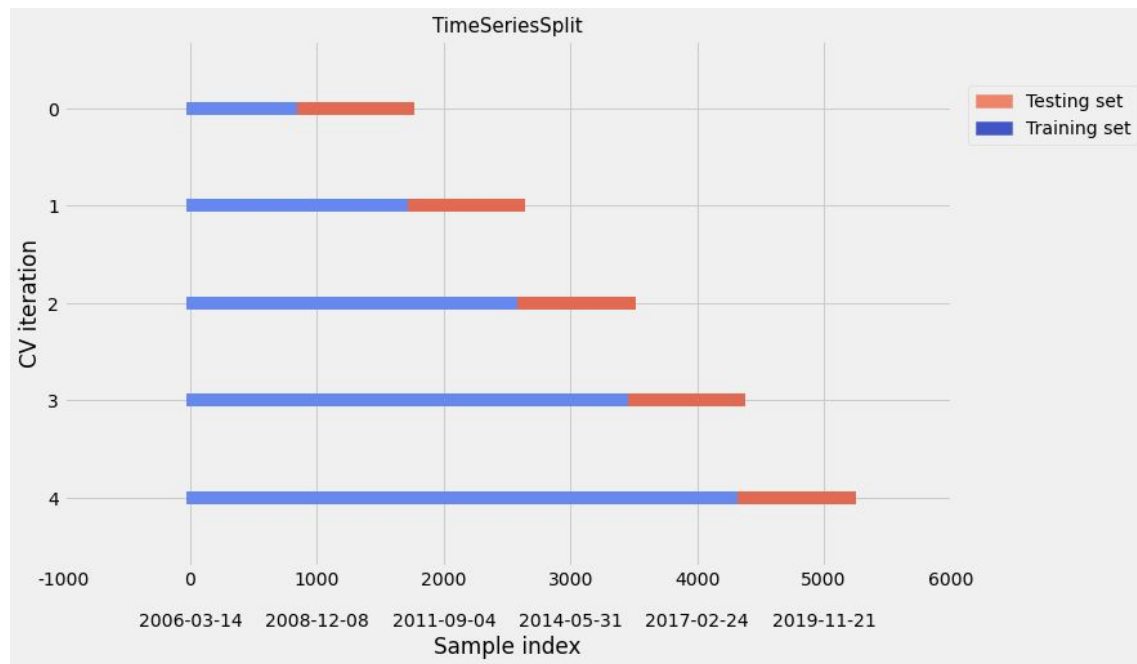


Figure 3: Pearson Correlation Coefficient matrix for features, highlighting potential multicollinearity that may necessitate features removal.

Data Splitting

- Target Variable Y:
 - Highly imbalanced
- TimeSeriesSplit:
 - Preserves chronological order
 - Prevents future data leakage



Preprocessor

- Variable ‘type’:
 - OneHotEncoder of five classes
 - Not ordered
- Variable ‘amount’
 - Standard Scaler

| | Train | Val | Test |
|---------------|---------|--------|--------|
| # Data Points | 1158108 | 386036 | 386036 |

| | Train/Val/Test |
|-------------------|----------------|
| # Features Before | 6 |
| # Feature After | 10 |

The background is a solid dark blue. Overlaid on this are three large, overlapping circles. The leftmost circle is a dark navy blue. The middle circle is a medium blue-grey. The rightmost circle is a light blue-grey. They overlap in a way that creates a sense of depth and movement from left to right.

Thanks!