McGill

INSY 695 | Group 2 | Winter 2025

AI-Powered Financial Fraud Detection:

Leveraging Machine Learning for Enhanced Security

Presented by:

Ali - Arial - Hazel - Julian - Margot

GitHub Repository: Financial-fraud

Aly-Farhat, arial0322, galaxyhikes, Julian-Oppedisano, mgerard1903

Introduction

Context and Significance

In today's fast-evolving financial landscape, digital transactions are at an all-time high, making fraud detection a critical challenge for financial institutions.
As financial crimes become more sophisticated, traditional rule-based fraud detection systems struggle to keep up, leading to billions in annual losses due to fraudulent activities.

According to the **Nasdaq Global Financial Crime Report:**[1]

- In 2023, an estimated **$3.1 trillion** in **illicit funds** flowed through the global financial system.
- **Fraud scams** and bank fraud schemes resulted in **$485.6 billion in losses** globally in 2023.
- U.S. businesses **lose** an average of **5%** of their gross **revenues** to **fraud**, with smaller businesses being more vulnerable due to fewer fraud prevention measures.

[1] *Nasdaq, 2024*

Problem Statment

Traditional fraud detection models rely heavily on **rule-based systems** that flag anomalies based on pre-defined conditions. However, fraudsters continually evolve their tactics, exploiting weaknesses in static systems.

As a result:

- Many fraudulent transactions go undetected.
- Legitimate transactions are falsely flagged, causing customer dissatisfaction.
- Financial institutions struggle to scale fraud prevention efforts efficiently.

Objective & Solution Approach

Our project leverages machine learning and advanced analytics to develop a predictive fraud detection system that can:

- Identify fraudulent transactions based on historical patterns.
- **Minimize false positives**, reducing disruptions to legitimate customers.
- Continuously adapt to new fraud schemes through self-learning capabilities.
- Enhance financial security while optimizing fraud prevention costs.

Hypothesis

- ***Null Hypothesis (H_0):***

"There is no significant relationship between transaction characteristics and fraud occurrence."

- ***Alternative Hypothesis (H_1):***

"Certain transaction characteristics increase the likelihood of fraud."

Data Description & Sources

1. Data Sources & Collection

Our project integrates three datasets to build a fraud detection model:

1. **Transactions Dataset**
   - Contains details of financial transactions, including transaction amount, payment method, merchant details, and timestamps.
2. **Fraud Labels Dataset**
   - A JSON file containing fraud labels (**Yes/No** → later mapped to **1/0**) linked to transactions by **transaction ID**.
3. **Users Dataset**
   - Includes **demographic** and **financial attributes** of customers, such as age, income, credit score, and number of credit cards.

We merged these datasets based on:

- **Transaction ID (id)** to link transactions with fraud labels.
- **Client ID (client_id)** to merge user profiles with transaction records.

2. Key Features

- **Transaction Features:** Transaction ID - amount - Payment method (Chip, Swipe, Online) - merchant_id - merchant_state - transaction_hour (Extracted from date; Feature Engineered)
- **User Features:** client_id - current_age - credit_score - yearly_income - number_credit_cards
- **Target Variable:** Binary label (0 = Legitimate Transaction, 1 = Fraudulent Transaction)

Data Preprocessing & Cleaning

1. Standardizing & Converting Data Types

- Datetime Conversion:
  - Transformed date column into datetime format for feature extraction.
  - Later dropped date after extracting transaction_hour.
- Currency Standardization:
  - Removed $ and , from financial columns (amount, per_capita_income, yearly_income, total_debt) and converted to float for numerical analysis.
- Target Encoding:
  - Converted Target (fraud indicator) from 'Yes'/'No' to binary (1 = Fraud, 0 = Legitimate).
- Categorical Standardization:
  - use_chip converted to categorical (Swipe Transaction, Chip Transaction, Online Transaction).

2. Handling Missing Values

- Imputation Strategy:
  - Merchant State: If merchant_city was "ONLINE", assigned "ONLINE" as merchant_state to maintain transaction type clarity.
  - ZIP Codes (Merchant-Based Mode Imputation):
    - Filled missing ZIP codes with the most frequent ZIP per merchant_id.
    - If a merchant had no known ZIPs, used the most common ZIP in the dataset as a fallback.

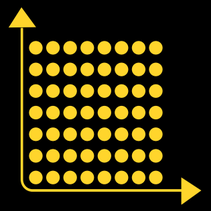
3. Dropped Columns

- errors: Had too many missing values, making it uninformative.

Data Challenges & Considerations

Class Imbalance

- Fraud cases are rare (**0.15%**) compared to legitimate transactions.



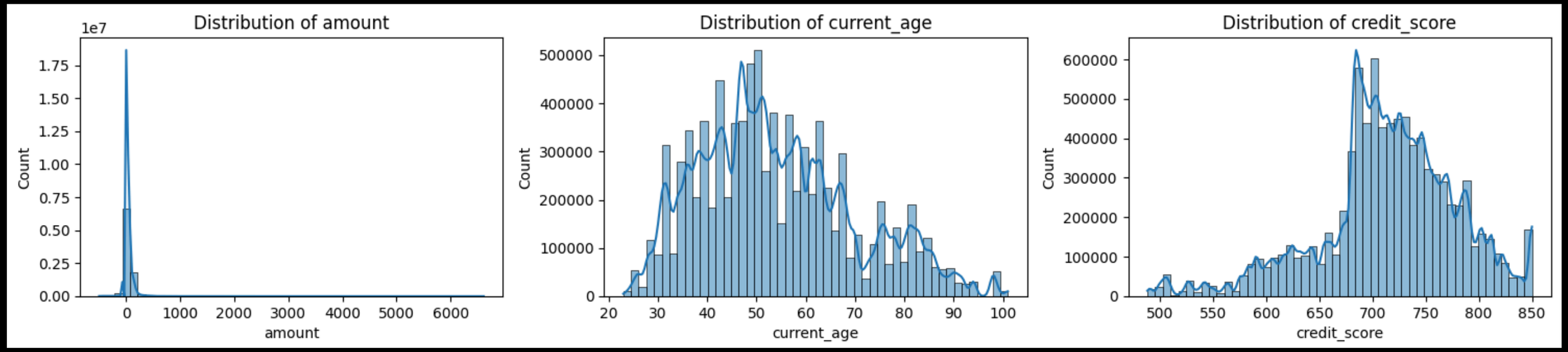
High Cardinality

- merchant_city was dropped due to excessive unique values.
- merchant_state was grouped into the top 9 states, with others categorized as "Other."

Exploratory Data Analysis

Univariate Analysis

We examined the distribution of all key numerical features such as:



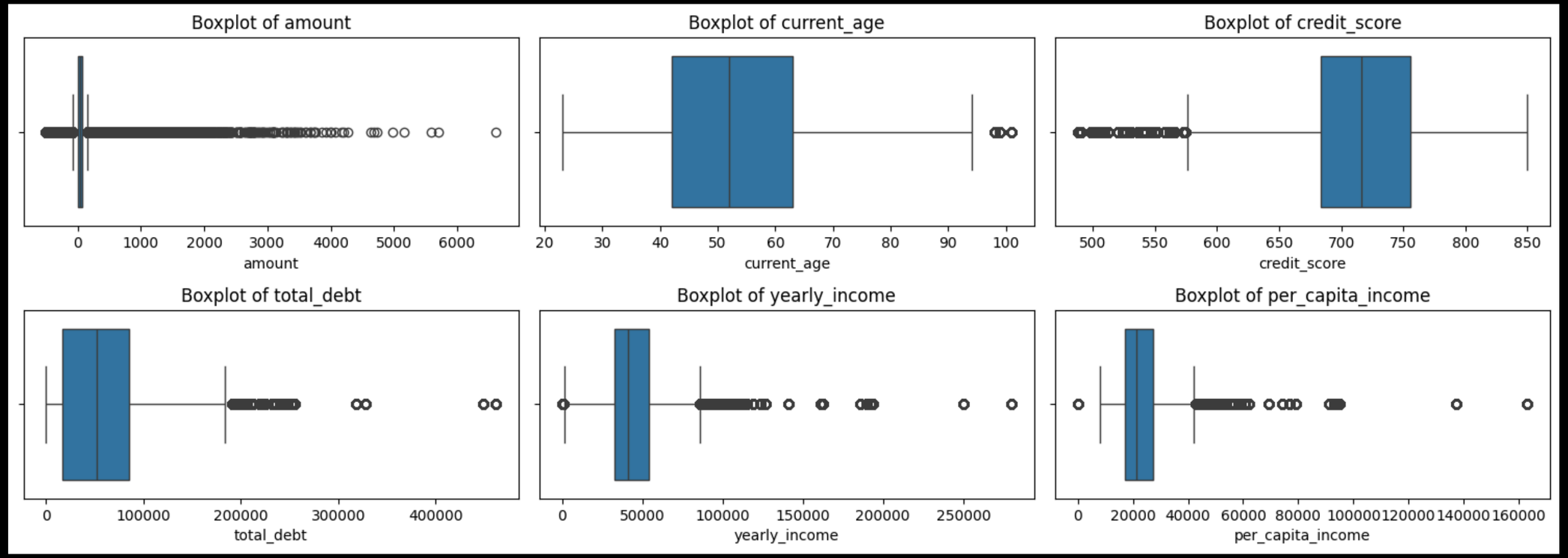
Key Observations:

- Some features exhibited high skewness such as transaction amount.
- Credit Score: Normally distributed but shows a drop-off after a threshold.

Exploratory Data Analysis

Univariate Analysis

Outlier Detection via Boxplots



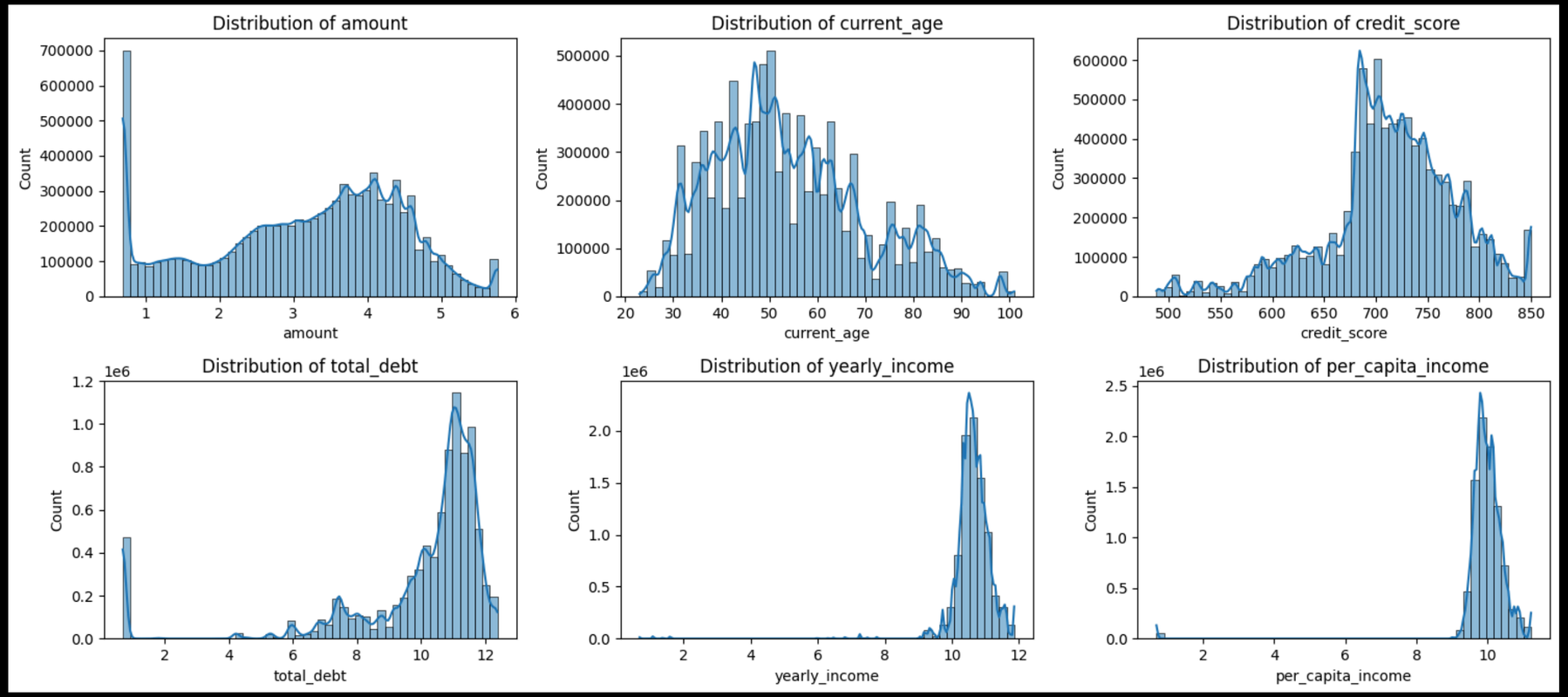
- Amount, total debt, and income variables showed extreme values.
- Outliers in credit score indicate potential fraudulent activity in specific score ranges.

Exploratory Data Analysis

Univariate Analysis

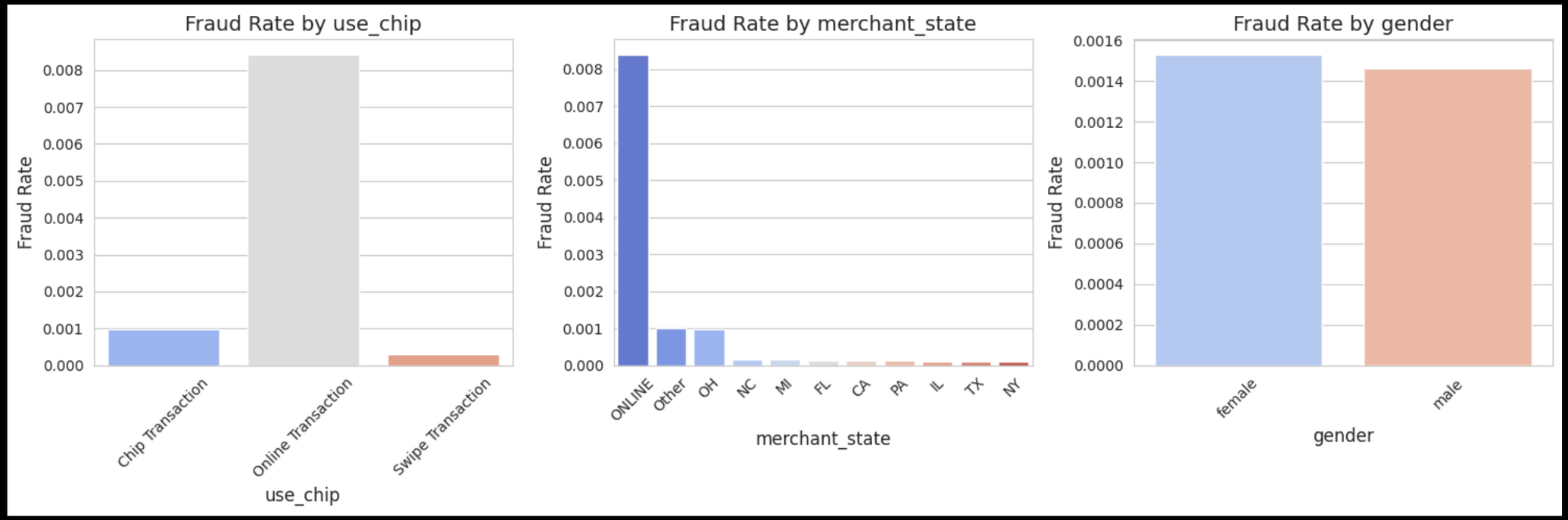
Skewness Handling (Winsorization & Log Transformation)

- Winsorization was applied to cap extreme values at the top 1%.
- Log transformation was applied to skewed features: Amount, Total Debt, Yearly Income, Per Capita Income.



Exploratory Data Analysis

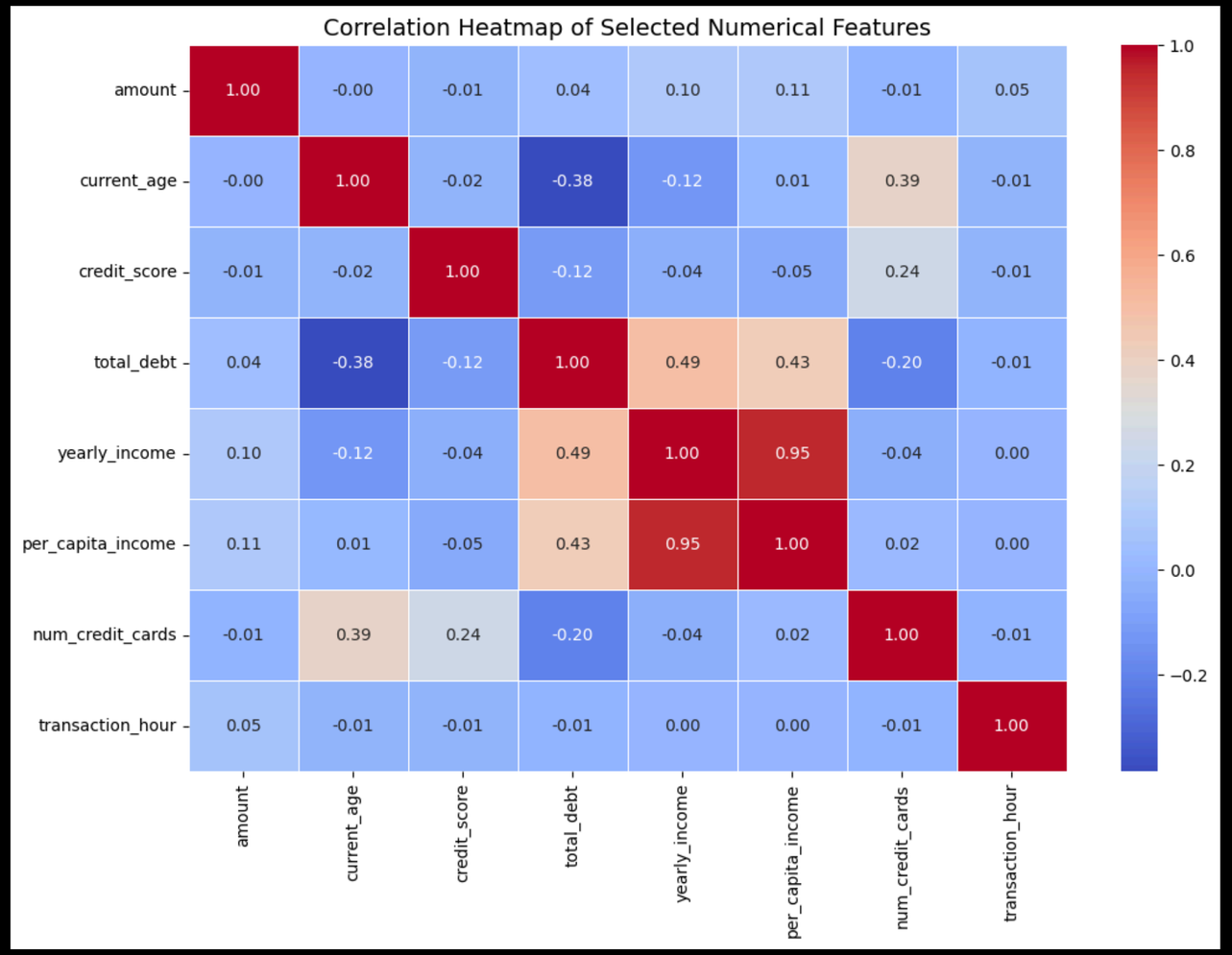
Bivariate Analysis



- Online transactions have the highest fraud rate, significantly exceeding chip and swipe transactions.
- Fraud rates vary by location, with ONLINE transactions leading by a large margin.
- Minimal difference in fraud rate between male and female customers. This means that gender might not be a strong fraud predictor.

Exploratory Data Analysis

Correlation Matrix



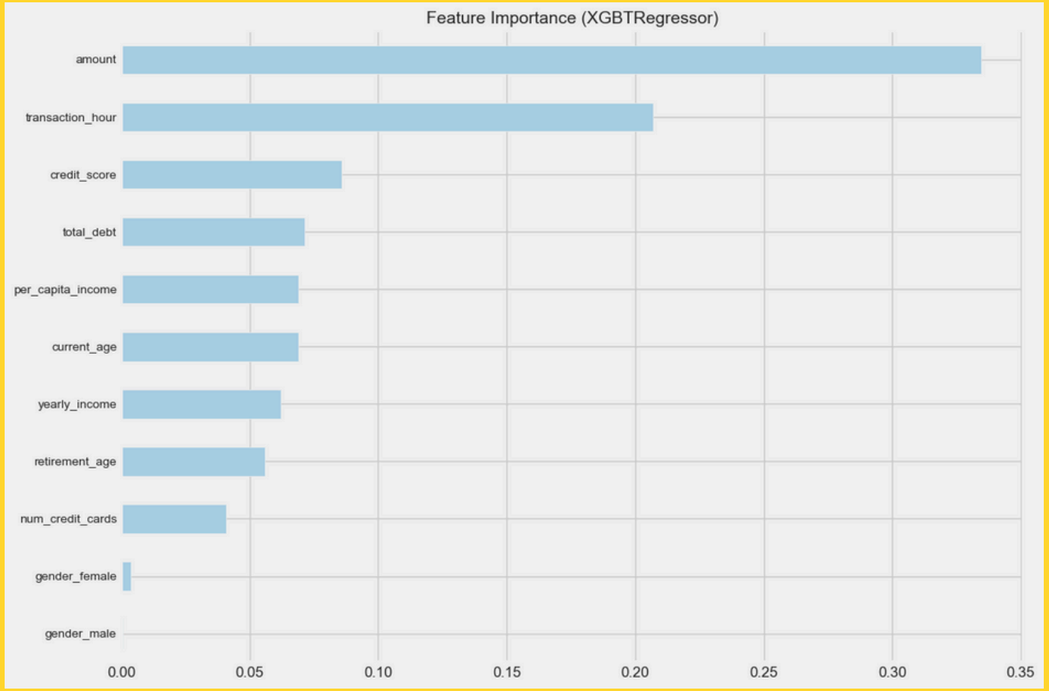
- High correlation between yearly_income and per_capita_income (0.95).
- Total debt and yearly income show moderate correlation (0.49), indicating a potential financial dependency.
- Credit score has a weak correlation with most features, meaning it might contribute independently to fraud detection.
- Transaction hour has near-zero correlation with all features, suggesting fraud occurs at varying hours.

Model Evaluation

Causal ML using EconML

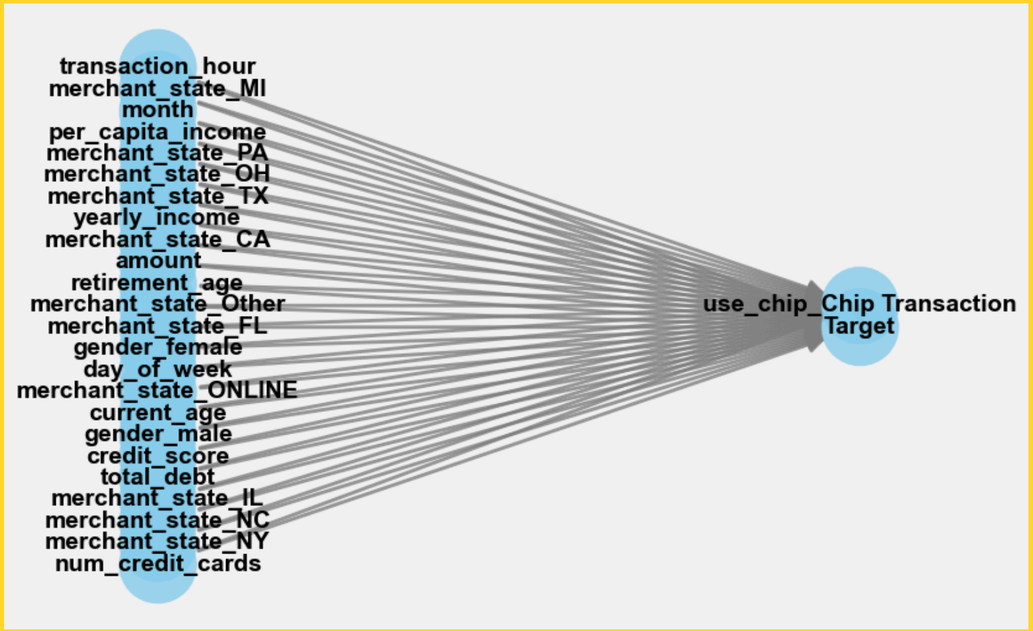
	Target	Treatement	ATE	95 CI %
LRSRegressor	Target	use_chip_Chip Transaction	-0.02	-0.02
XGBTRegressor	Target	use_chip_Chip Transaction	-0.06	-0.6

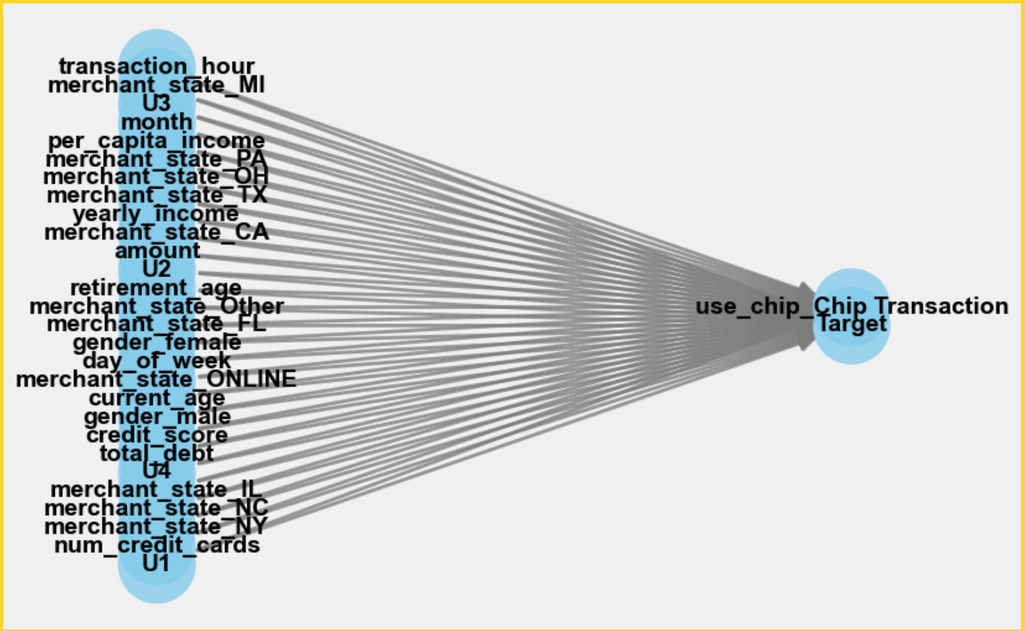
*Confounders: amount, transaction_hour, current_age, retirement_age, per_capita_income, yearly_income, total_debt, credit_score, num_credit_cards, gender_female, gender_male



Model Evaluation

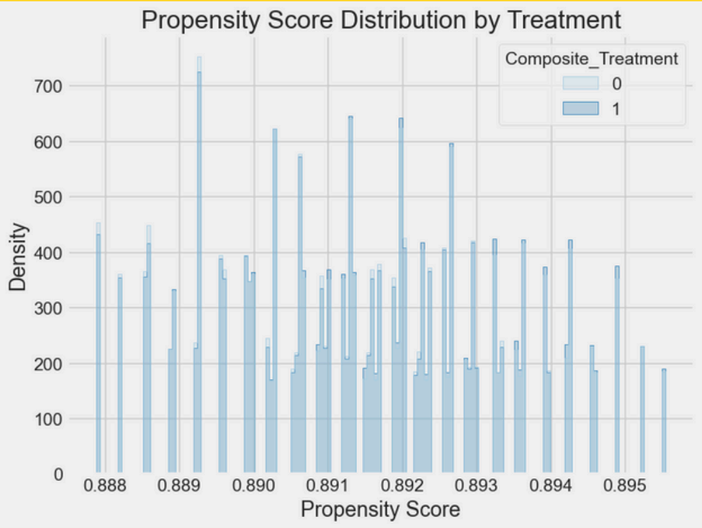
Causal ML using DoWhy

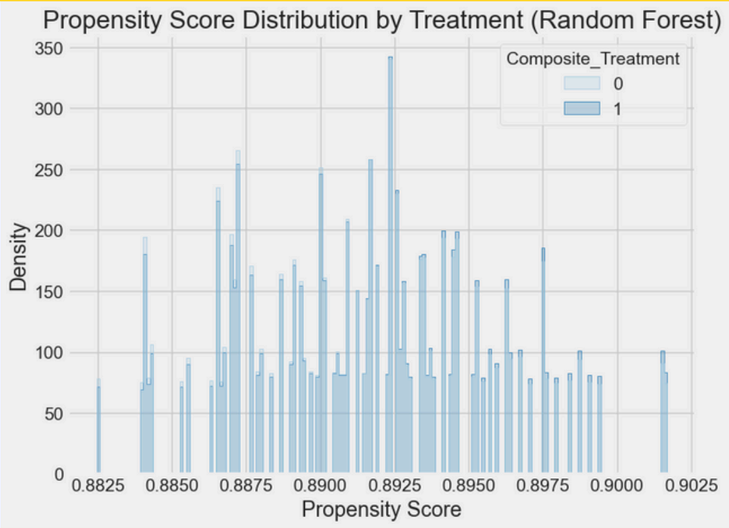




Model Evaluation

Understanding the results





1) Logistic Regression-based Propensity Score
2) Inverse Probability Weighting (IPW) using backdoor method
3) Feature Selection for Confounders removing highly correlated confounders
4) Random Forest-based Propensity Score Estimation

Feature Engineering

Client specific features

Global metrics

- Fraud rate
- Transaction frequency
- Weekday hour category
- Time between transactions
- Amount : Yearly income
- Debt : Yearly income

Feature Engineering

Client specific features

Global metrics

Fraud rate

Transaction frequency

Weekday hour category

Time between transactions

Amount : Yearly income

Debt : Yearly income

Nature of transaction

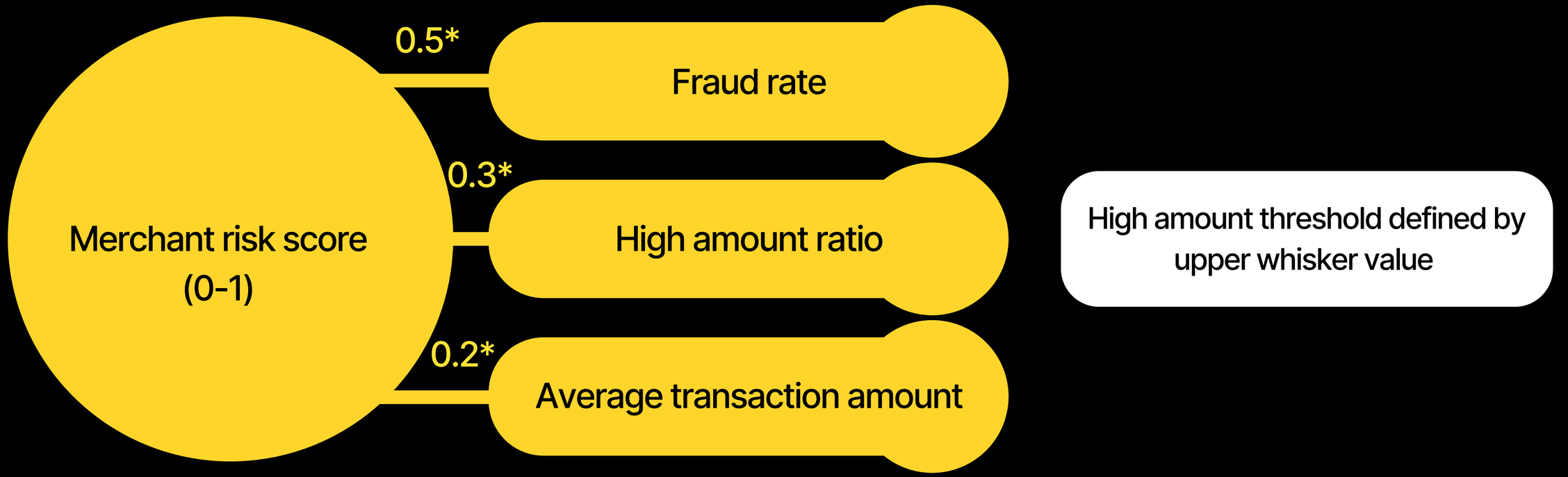
Distance : Yearly income

Amount : Average value

Typicality score

Suspicious transaction activity

Feature Engineering



Calculated separately for each transaction type

Modelling

Train-test split
- Training 80%
- Testing 20%

Resample training set
- Oversample fraud to 20%
- Undersample non-fraud, balance to 1:3 ratio

Stacking

CatBoost
Fine-tuning with precision score

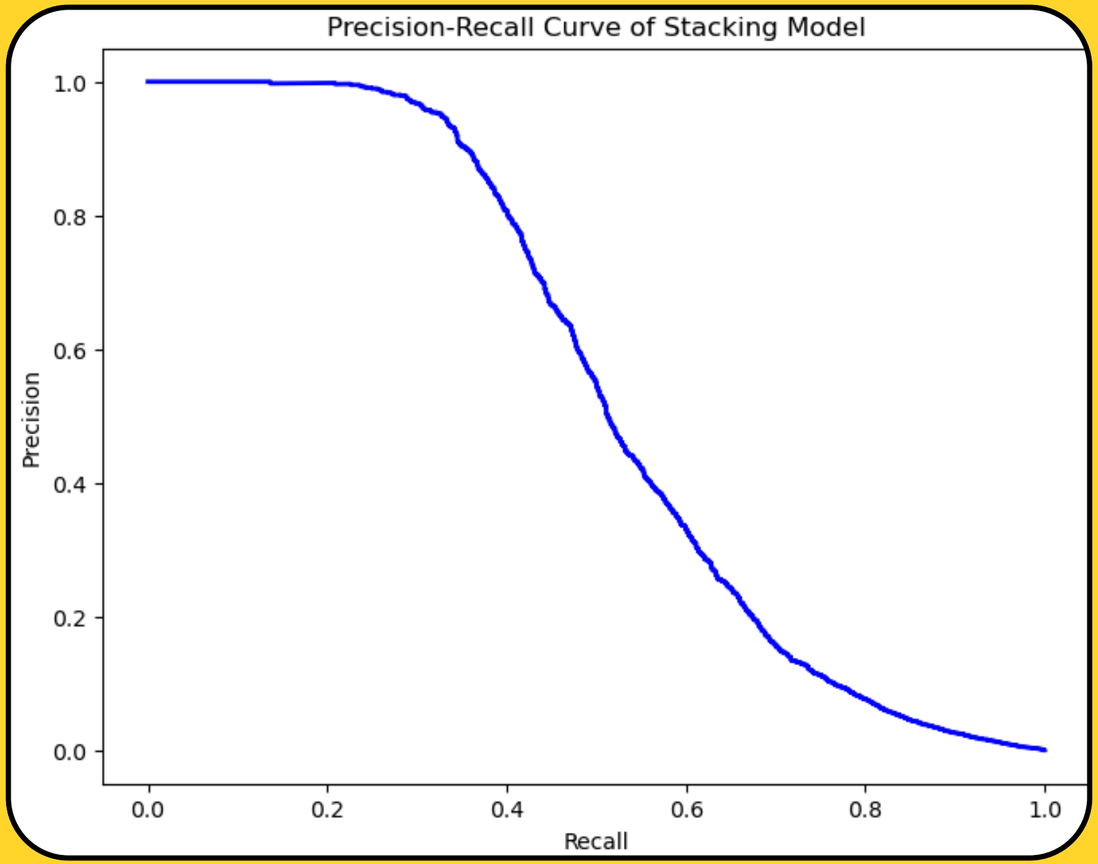
TabNet
Fine-tuning with PR-AUC

Meta Model: Logistic Regression

Final Prediction
Tune decision threshold with f1 score

Model Evaluation
- Precision, Recall, F1 (specific to Fraud class)
- Feature importance

Model Evaluation



Coefficients of Logistic Regression:

CatBoost - 5.08, TabNet - 3.90

Best Decision Threshold: 0.51

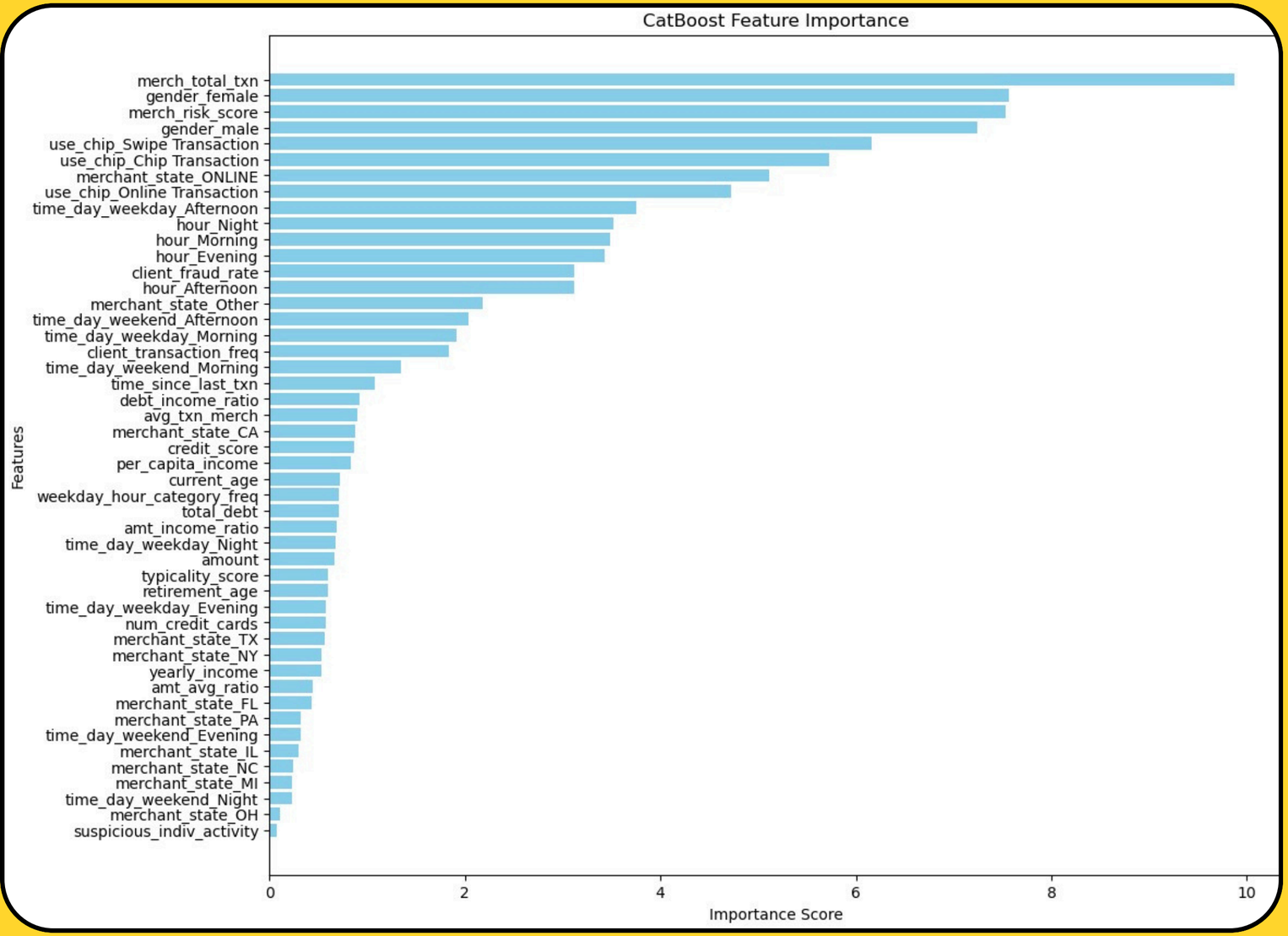
	Precision	Recall	F1-Score	PR-AUC
Stacking	0.70	0.44	0.54	0.54
CatBoost	0.33	0.64	0.44	0.58
TabNet	0.13	0.61	0.22	0.46

*All of the metrics are specific to Fraud class

Comparison to Literature

Method	Evaluation Metrics (%)		
	Precision	*Recall*	*F1-Score*
RF [11]	53.90	71.60	51.40
LR [11]	53.30	69.70	49.90
DT [12]	52.28	61.14	56.37
Voting [12]	57.19	43.08	49.14
AdaBoost [12]	18.89	55.76	28.21
TabNet	79.37	48.67	60.34
TabNet + SMOTE	79.01	49.31	60.73
TabNet + Learning Rate Scheduler + SMOTE	77.62	50.91	**61.49**

C. C. Meng, K. M. Lim, C. P. Lee and J. Y. Lim, "Credit Card Fraud Detection using TabNet," 2023 11th International Conference on Information and Communication Technology (ICoICT), Melaka, Malaysia, 2023, pp. 394-399, doi: 10.1109/ICoICT58202.2023.10262711.



Model Evaluation: Feature Importance

- Online transactions are a major risk factor
- Several engineered features emerged to be important
- Potential over-reliance on generalised features eg merchant total transactions leading to false alarms

Threats to Validity

1. **Data Privacy & Limited Access:** Strict privacy regulations often lead to restricted or anonymized data, reducing dataset richness and representativeness.
2. **Unrepresentative Samples:** Datasets may not reflect the full population (e.g., certain regions or institutions underrepresented), limiting generalizability.
3. **Labeling & Ground Truth Issues:** Fraud can be mislabeled or discovered with delays, introducing noise and bias into training data.
4. **Omitted Confounders & Missing Data:** Key drivers of fraud (e.g., changing economic conditions) might be absent or incomplete, skewing model estimates.
5. **Concept Drift & Evolving Fraud Tactics:** Fraudsters continually adapt, causing historical patterns to become outdated and reducing model performance over time.
6. **Measurement Bias:** Inconsistent data collection practices or different detection thresholds can distort the true incidence and nature of fraudulent transactions

Business Implications

- **Reduced Financial Losses –** AI-driven fraud detection minimizes chargebacks, unauthorized transactions, and financial fraud-related losses.
- **Enhanced Risk Assessment –** Immediate fraud detection ensures quicker response times, reducing institutional risk exposure.
- **Enhanced Customer Trust & Retention –** Secure transactions improve brand reputation and increase customer loyalty.
- **Operational Cost Reduction –** Automation reduces manual fraud investigations, lowering operational expenses.
- **Optimized Fraud Prevention & User Experience –** Adaptive fraud scoring minimizes false positives, ensuring seamless transactions for legitimate customers.
- **Regulatory Compliance –** Ensures adherence to AML, PCI-DSS, GDPR standards, reducing legal risks.
- **Bias & Fairness Considerations –** Prevents algorithmic bias, ensuring ethical and unbiased fraud detection.
- **Competitive Market Advantage –** AI-based fraud prevention differentiates financial institutions, enhancing industry positioning.
- **Fraud Prevention-as-a-Service (FPaaS) –** Potential monetization of fraud detection models as a business service.
- **Data Security & Privacy –** Advanced encryption and cybersecurity measures protect sensitive financial data from breaches.

Lessons Learned

- Data Quality is Critical – Handling missing values, outliers, and feature engineering significantly impacts model performance.
- Feature Selection Matters – Removing highly correlated variables prevents redundancy and improves interpretability.
- Fraud Patterns are Complex – Fraudulent transactions exhibit unique behavioral patterns, requiring both numerical and categorical insights.
- Balancing Model Accuracy & False Positives – A trade-off exists between capturing fraud and minimizing false alarms, requiring careful threshold tuning.
- Time-Based Trends in Fraud – Transaction hour and merchant type influence fraud likelihood, highlighting the need for temporal analysis.
- Model Interpretability is Key – Businesses require explainable AI to justify fraud flags, ensuring trust and compliance.
- Continuous Monitoring is Necessary – Fraud patterns evolve, demanding rapid updates and retraining of detection models.

INSY 695

Group 2

Winter 2025

Thank you