

Online Payment Fraud Prediction

Team 9:

Chuheng (Kevin) Yu
Yuhan (Sherry) Wang
Linting (Linsey) Wang
Tyler Christoforo

Motivation

Online payments are a substantial part of modern business components, it is paramount to ensure transaction safety for entities on both ends

Objective

- Evaluate online payment attributes
- Determine fraudulent payment patterns within existing dataset
- Use machine learning to learn patterns and predict fraud

Why does fraud matter?

Characteristic

- Fraudulent or unauthorized transactions
- Lost or stolen merchandise
- False requests for a refund, return or bounced checks



Result

Money lost!

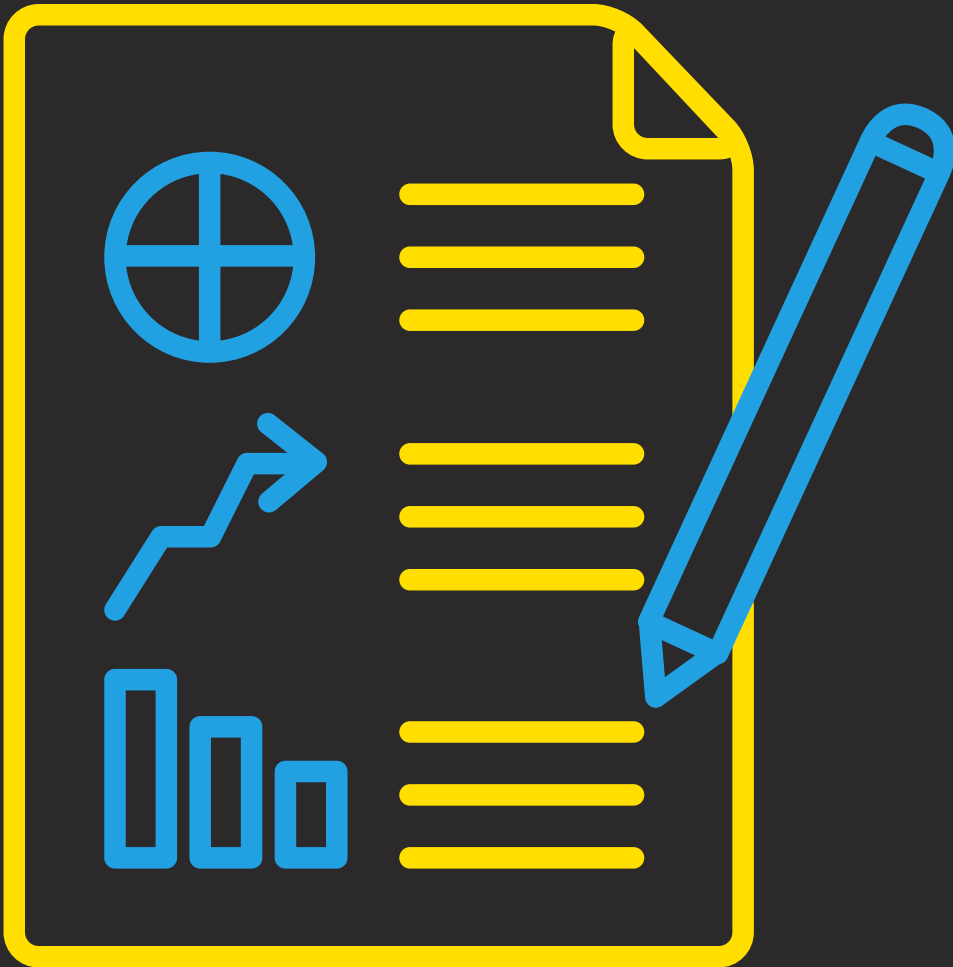
Data Source

A dark gray background with a yellow triangle in the bottom right corner. Numerous yellow dollar bills of various denominations (1, 2, 5, 10, 20, 50, 100) are scattered across the frame, appearing to fall from the top. The bills are oriented in various directions, some horizontally and some vertically.

Data-set was collected from Medium.com
made public by Edgar Alonso Lopez-Rojas

Data Description

variable	type	description
step	int64	1 step means 1 hour
type	object	type of online transaction
amount	float64	the amount of the transaction
nameOrig	object	customer starting the transaction
oldbalanceOrg	float64	balance before the transaction
newbalanceOrig	float64	balance after the transaction
nameDest	object	recipient of the transaction
oldbalanceDest	float64	initial balance recipient before the transaction
newbalanceDest	float64	the new balance recipient after the transaction
isFraud	int64	fraud transaction



	step	type	amount	nameOrig	oldbalanceOrg	newbalanceOrig	nameDest	oldbalanceDest	newbalanceDest	isFraud	isFlaggedFraud
0	1	PAYMENT	9839.64	C1231006815	170136.0	160296.36	M1979787155	0.0	0.0	0	0
1	1	PAYMENT	1864.28	C1666544295	21249.0	19384.72	M2044282225	0.0	0.0	0	0
2	1	TRANSFER	181.00	C1305486145	181.0	0.00	C553264065	0.0	0.0	1	0
3	1	CASH_OUT	181.00	C840083671	181.0	0.00	C38997010	21182.0	0.0	1	0
4	1	PAYMENT	11668.14	C2048537720	41554.0	29885.86	M1230701703	0.0	0.0	0	0

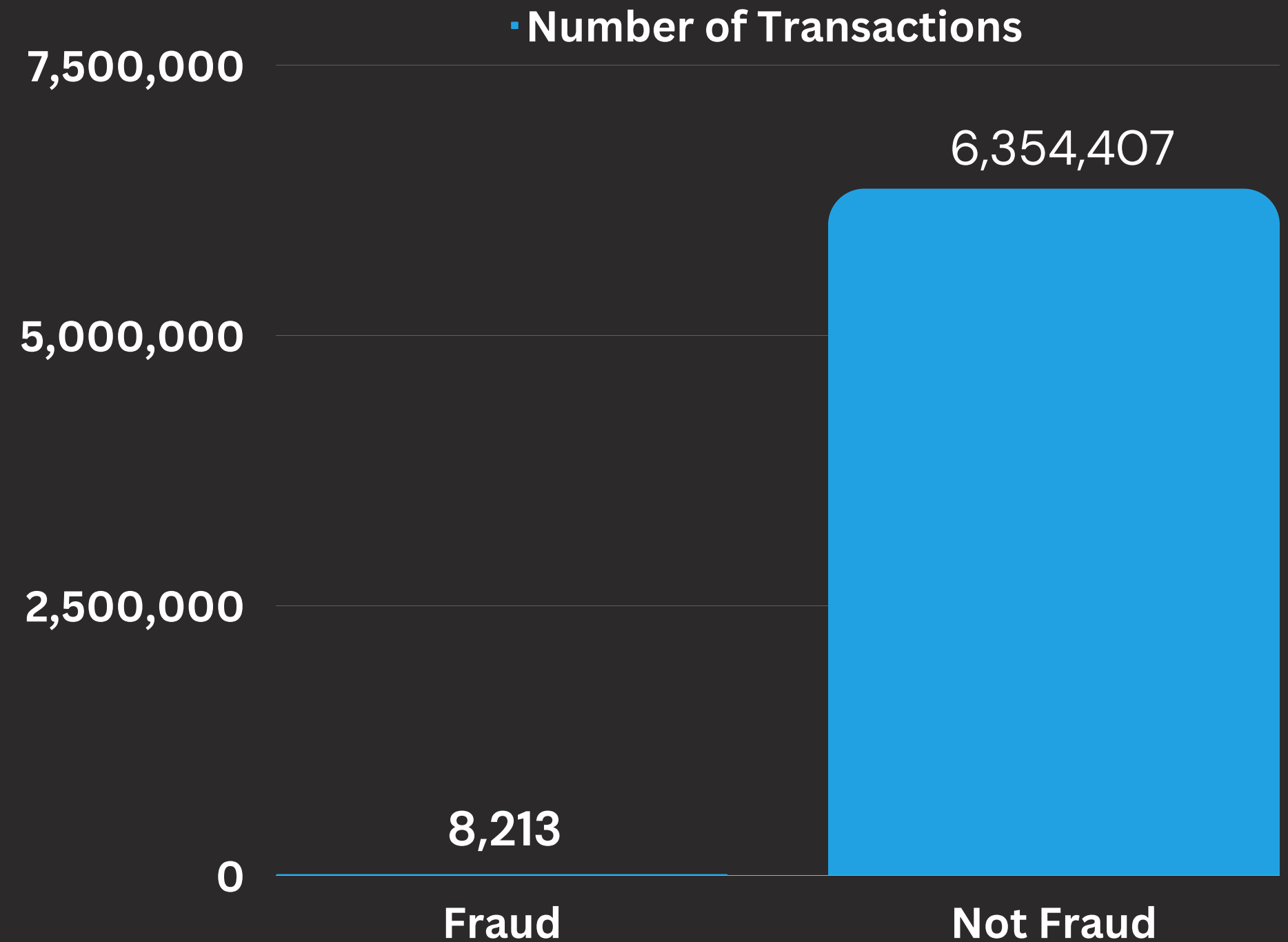
Summary Analysis

Imbalance distribution

- fraud transactions is only 0.13%

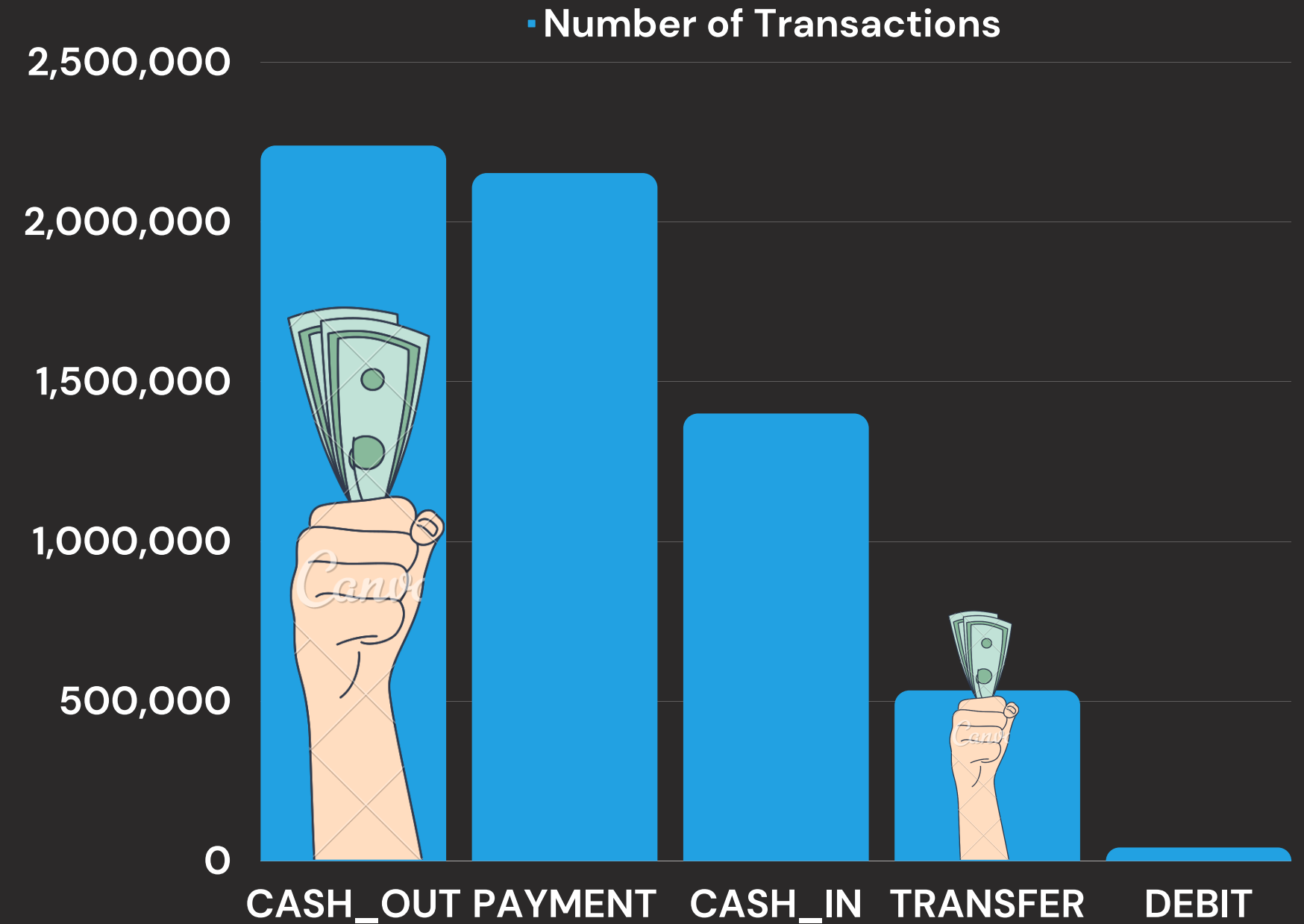
How to deal with imbalance dataset

- Oversampling
- Undersampling



Summary Analysis

- **Most transaction types**
 - Cash_out and Payment
- **Focused group:**
 - Fraud transactions only appearing in
 - Cash_out and Transfer



Correlation matrix

	step	amount	oldbalanceOrg	newbalanceOrig	oldbalanceDest	newbalanceDest	isFraud
step	1.000000	0.022373	-0.010058	-0.010299	0.027665	0.025888	0.031578
amount	0.022373	1.000000	-0.002762	-0.007861	0.294137	0.459304	0.076688
oldbalanceOrg	-0.010058	-0.002762	1.000000	0.998803	0.066243	0.042029	0.010154
newbalanceOrig	-0.010299	-0.007861	0.998803	1.000000	0.067812	0.041837	-0.008148
oldbalanceDest	0.027665	0.294137	0.066243	0.067812	1.000000	0.976569	-0.005885
newbalanceDest	0.025888	0.459304	0.042029	0.041837	0.976569	1.000000	0.000535
isFraud	0.031578	0.076688	0.010154	-0.008148	-0.005885	0.000535	1.000000

Further Analysis

- Data cleaning
 - Normalization & feature extraction
 - Conversion of categorical fields
- Model selection:
 - XGBClassifier, KNN, Decision Tree, Logistic Regression...
- Cross-validation
 - To optimize the fit of the training data
- Evaluation
 - AUC score comparison

Questions?

