



Credit Card Fraud Detection

Agenda

- BACKGROUND
- DATA DESCRIPTION
- METHODOLOGY
- FEATURE ENGINEERING
- EXPLORATORY DATA ANALYSIS
- MODEL RESULTS
- SUMMARY AND CONCLUSION



/02

/03

Credit card fraud is bigger than ever.

\$28.65 Billion

lost worldwide
in 2020

29% increase

in the Philippines
from 2019

Increasing fraud

during COVID-19
pandemic



Data Description

Transactions Dataset - 25,535,193 total records

Customer Information

- Credit card number
- Name
- Gender
- Location
- Job

Transaction Records

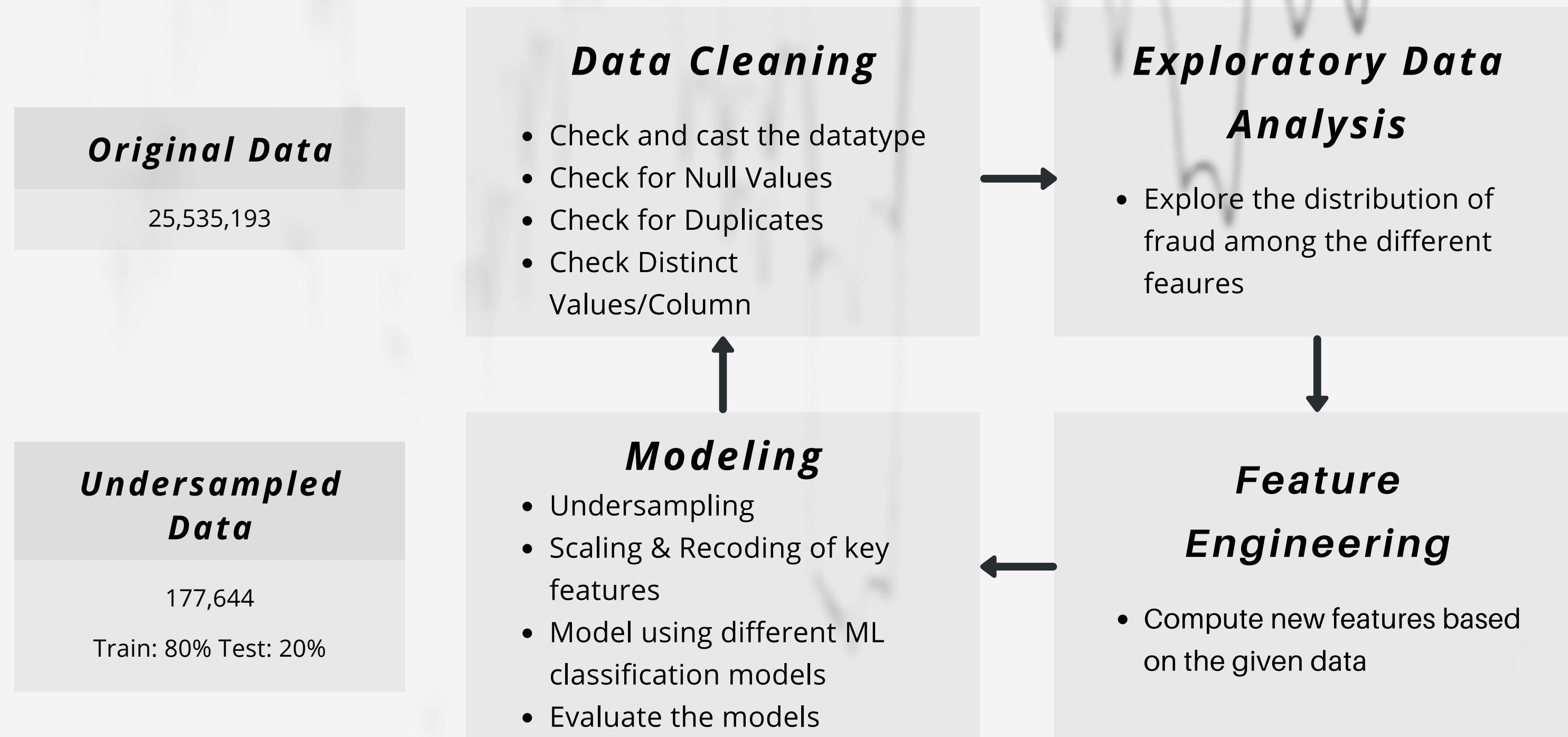
- Trans date
- Trans time
- Amount

Merchant Information

- Merchant name
- Category
- Location

Methodology

/05



/06

Feature Engineering

Age at Transaction

*Distance between
customer home and
shop*

*Elapsed time from
the last transaction*

***Z-score of the
transaction amount
in comparison to :***

- the customers 30-day total spending average
- and 30-day average spending by Category

***Online transaction
or not***

9,425

Total credit card holders

Age distribution of all credit card holders

42% are **adults** (36 to 60 years old)

22% are **young adult** (25 to 35 years old)

22% are **senior** (60 years old and above)

12% are **youth** (15 to 24 years old)

2% are **below 15 years old**

Most of their credit card transactions happen every month of **December**.



Credit Card Fraud Detection

Only 0.35%

of all transactions are tagged
as fraudulent

48 yrs old

average age of people with
fraudulent transactions

Cities with the lowest rate of fraud

- San Juan, 28%
- Naga City, 29%
- Pasay City, 29%

Cities with the highest rate of fraud

- General Santos, 42%
- Palayan City, 41%
- Vigan City, 41%

Most number of Fraud Cases based on Category

Shopping (online): **23.40 %** of the total cases

Least number of Fraud Cases based on Category

Grocery (online): **1.54 %** of the total cases

Classification of Fraud Cases

Online Transactions: **26.38%** of the total cases

Point of Sale (pos): **73.36 %** of the total cases

/10

ML MODEL

machine learning model used



RESULTS
comparison of all models

Fraud Detection using Logistic Regression Model

| | | |
|--|-----------------------------------|-----|
| <i>Label Column</i> | <i>Accuracy</i> | — |
| is_fraud | 77% | |
| <i>Features</i> | <i>AUC</i> | |
| gender, city, age_at_trans, distance, rolling_z-score, cat_rolling_z-score, elapsed_trans_time, online | 0.76 | |
| | <i>False Positive Rate</i> | /11 |
| | 14% | |

/12

| <i>Model</i> | <i>Logistic Regression</i> | <i>Decision Tree</i> | <i>Naïve Bayes</i> | <i>Support Vector Machine</i> |
|---------------------|--|--|--|--|
| Features | gender, city, is_fraud, age_at_trans, distance, rolling_z-score, cat_rolling_z-score, elapsed_trans_time, online | gender, city, is_fraud, age_at_trans, distance, rolling_z-score, cat_rolling_z-score, elapsed_trans_time, online | gender, city, is_fraud, age_at_trans, distance, rolling_z-score, cat_rolling_z-score, elapsed_trans_time, online | gender, city, is_fraud, age_at_trans, distance, rolling_z-score, cat_rolling_z-score, elapsed_trans_time, online |
| Accuracy | 0.82 | 0.86 | 0.75 | 0.68 |
| False Positive Rate | 0.17 | 0.15 | 0.14 | 0.13 |
| AUC | 0.81 | 0.86 | 0.75 | 0.77 |

/13

Summary



Credit Card Fraud

is increasing in an alarming rate worldwide especially during the pandemic where people are mostly using cashless transactions

ML Algorithms

are efficient in detecting credit card fraud but there's still more to develop to reduce false positive detection

ML Models

Logistic Regression and Decision trees are one of the top performing models in classifying fraud and non-fraud transactions

Recommendations

/14

Model Recommendations

- Check first with the credit card holder whether a transaction is fraud before acting on it
- Run the model using Decision Tree classifier to confirm that it can output better accuracy
- Exploring and focusing on the Hyperparameter tuning and using different models is highly recommended given that the processing power is not an issue
- Use other features that can monitor monthly frequency of transactions per category

Feature Recommendations

- Providing additional features like family status, purchasing history and basic income can improve the models accuracy and prediction. (purchasing behavior)
- A more sophisticated feature engineering can be done in choosing which features are important in the modelling process

/15

Thank you!

Credit Card Fraud Detection

