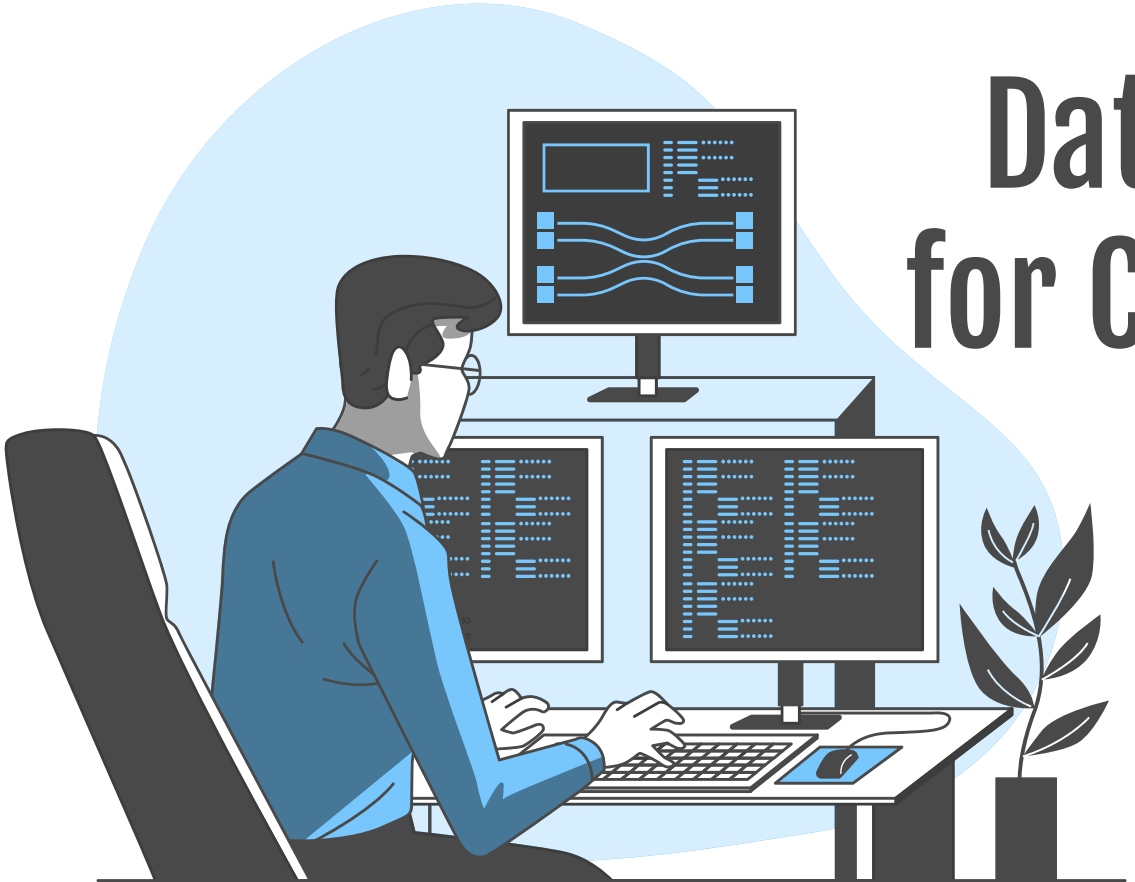


Data Science for Consulting

Broom Solutions Credit Card Fraud

Neal Xu
Maria Clarissa Fionalita,
Yutao (Eric) Li,
Andrew Leonard,
Yilin Gong

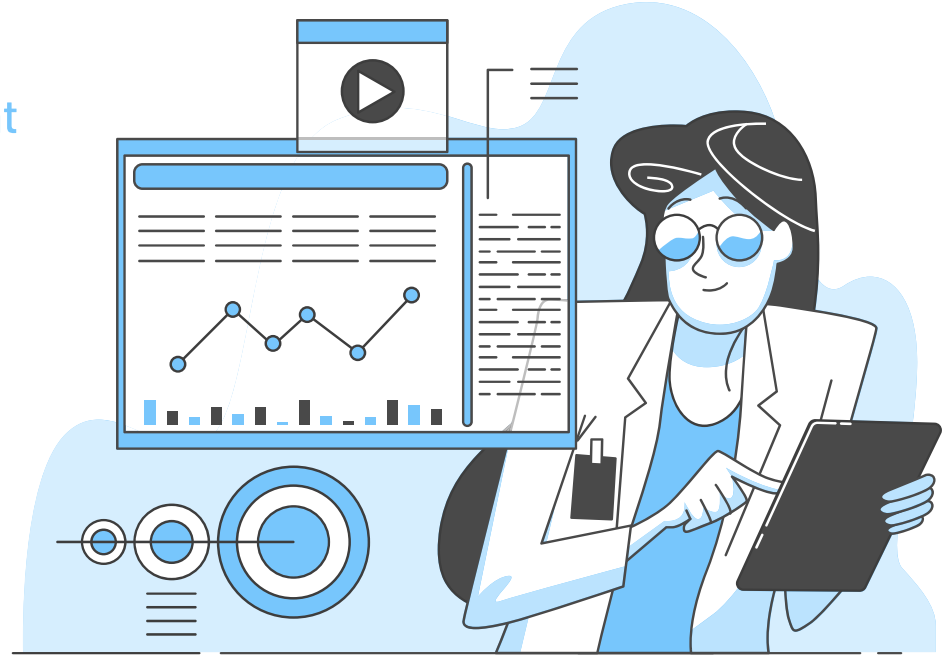
Presentation Overview

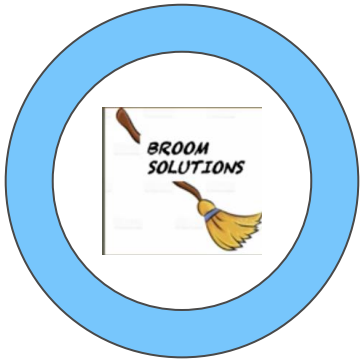
01 Problem & Solution

02 Analytic Maturity of Client

03 Modeling Methodology & Results

04 Economic Impact





Broom Solutions

Broom solutions is a family run payment processing and commerce solutions company offering a suite of services including POS systems, ecommerce payment processing, financial reporting, inventory management, and fraud detection.

!!PROBLEM!!
Too Good at Fraud Detection

Catching Criminals

Broom Solutions current fraud system catches 98% of fraudulent transactions

Flagging the Innocent

Non fraudulent transactions flagged as fraud 6% of the time

Client Complaints

Retail and ecommerce clients complain of losing long term business from customer frustration

Our Solutions

<$600

Allow more fraud transactions

- fewer false positives

>=$600

Increase / keep the current fraud detection precision level

Economic Value

Potential Economic Value = Correctly Predicted Transaction * Average Transaction Amt

Logistic Regression

- False positive predicted number * Average non-fraud amt
- = 51,393 (Q) * 68 (P) = **$3,494,724**

Decision Tree

- False positive predicted number * Average non-fraud amt
- = 39,205 (Q) * 68 (P) = **$2,665,940**

$2,411,756

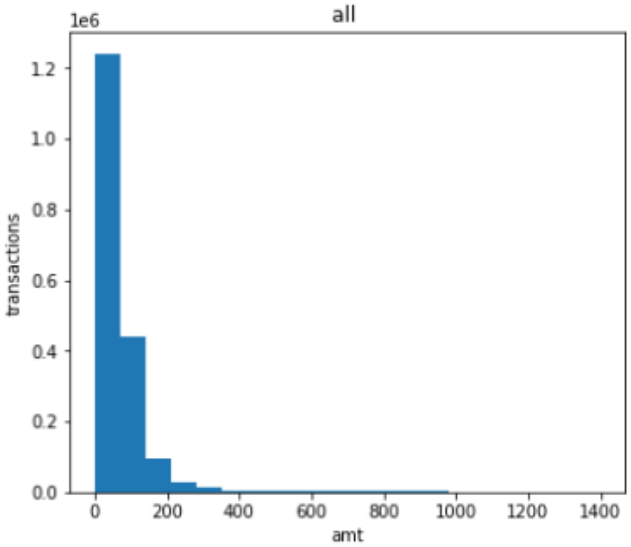
Random Forest

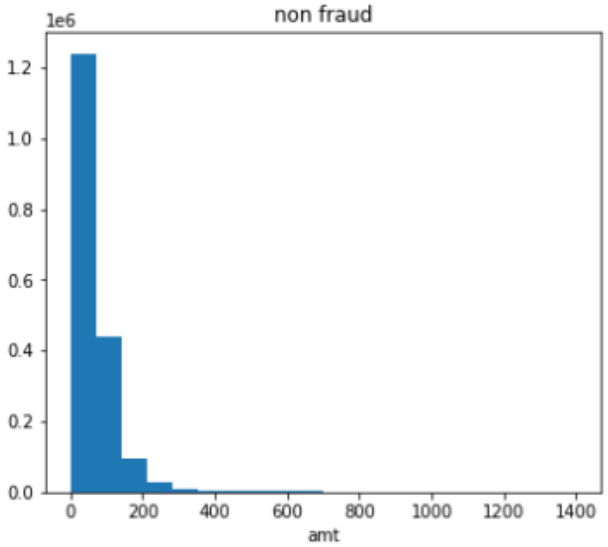
- False positive predicted number * Average non-fraud amt
- = 31,462 (Q) * 68 (P) = **$2,139,416**

Xgboost

- False positive predicted number * Average non-fraud amt
- = 15,926 (Q) * 68 (P) = **$1,082,968**

Data

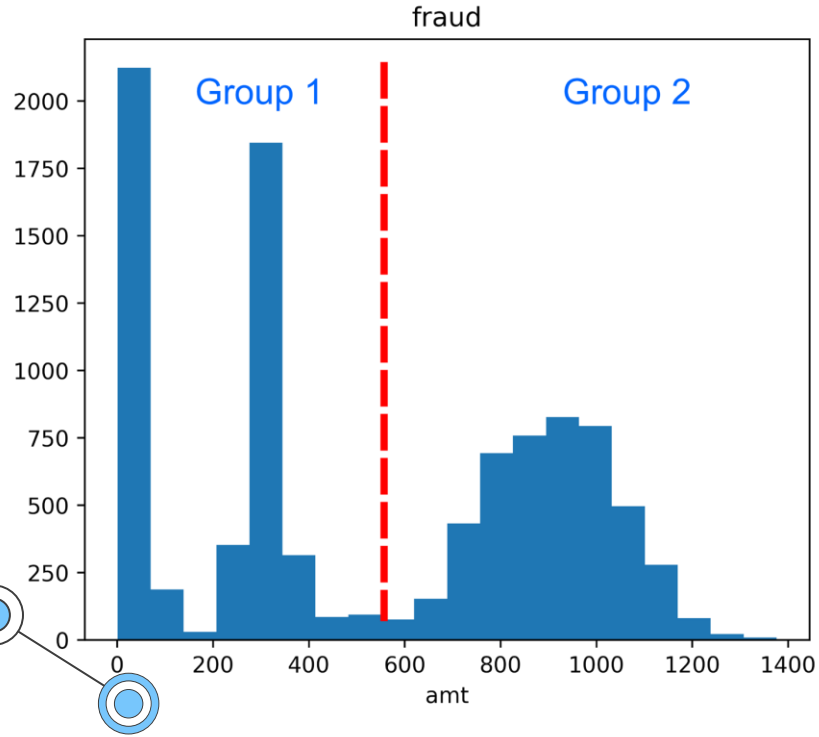




'trans_date_trans_time', 'cc_num', 'merchant', 'category', 'amt',
'first', 'last', 'gender', 'street', 'city', 'state', 'zip', 'lat',
'long', 'city_pop', 'job', 'dob', 'trans_num', 'unix_time', 'merch_lat',
'merch_long', 'is_fraud'

Exploratory Data Analysis

Dividing the Customers



We should focus on the "AMOUNT"

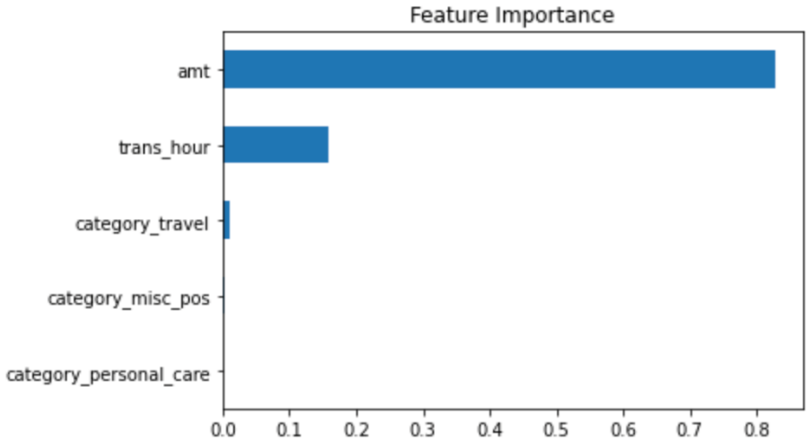


Figure above developed from EDA, showing the most relevant variables for fraud detection.

The Methodology

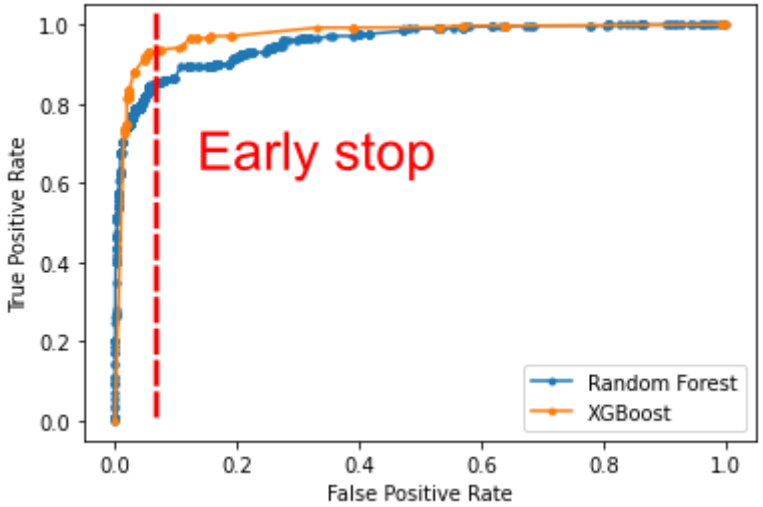
Divide the data into two groups

Oversampling for the imbalance and lack to data

Dual modeling for over-precise model

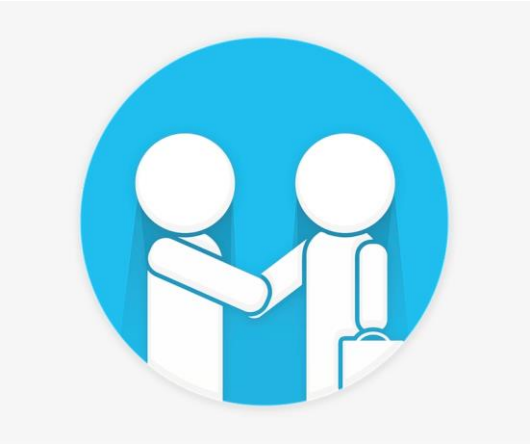
Results

```
Random Forest: ROC AUC=0.956
XGBoost: ROC AUC=0.974
```



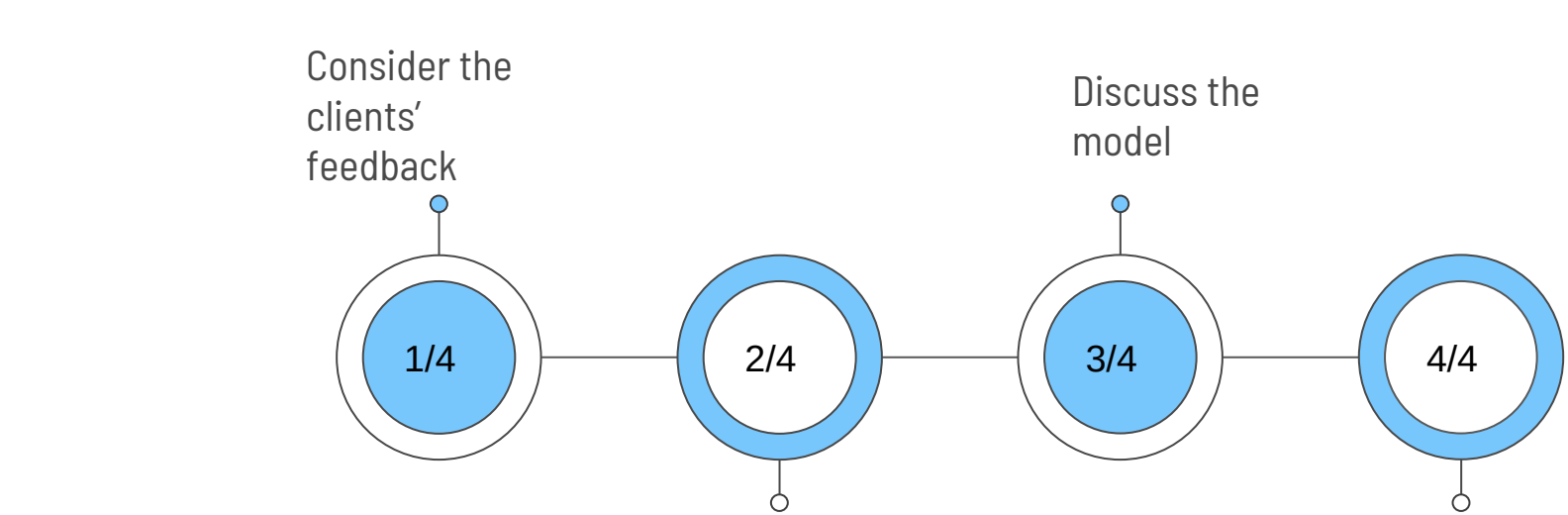
Questions to Ask the Client

- Any data regarding the the customer satisfaction?
- How much loss would you take for a higher customer satisfaction?



- Any additional objectives in addition to the increase customer satisfaction?

Next Steps


Consider the clients' feedback
Discuss the model
1/4
2/4
3/4
4/4
Impact of the model towards the customers' satisfaction
Testing the strategy