



Credit Card Fraud Detection

Christopher Chung

Springboard Capstone Project 2

November 2021

Background

- ▷ In 2018:
 - Over 200,00 breached accounts
 - Merchants lost \$2.94 for every \$1 in fraud

- ▷ Credit card fraud affects:
 - Customer
 - Merchant
 - Issuing bank



Understanding the problem

		Predicted label	
		Negative (0)	Positive (1)
Actual label	Negative (0)	True negative: The transaction is predicted as valid and is actually valid	False positive: The transaction is predicted as fraud but is actually valid
	Positive (1)	False negative: The transaction is predicted as valid but is actually fraud	True positive: The transaction is predicted as fraud and is actually fraud

Metrics

Precision of fraud class

		Predicted	
		0	1
Actual	0	True negative	False positive
	1	False negative	True positive

Recall of fraud class

		Predicted	
		0	1
Actual	0	True negative	False positive
	1	False negative	True positive

f1-score of fraud class

Key results

Model: XGBoost

Precision: 85%

Recall: 81%

f1-score: 83%



The Data

Simulated credit card transaction dataset from [Kaggle](#)

Transaction Date/Time	Credit Card Number	Merchant	Category	Amount	...	is_fraud
2019-01-01 00:00:18	27031...	Rippin, Kub and Mann	misc_net	\$4.97	...	0
2019-01-01 00:00:44	63042...	Heller, Gutmann and Zieme	grocery_pos	\$107.23	...	0
2019-01-01 00:00:51	38859...	Lind-Buckridge	entertainment	\$220.11	...	0
2019-01-03 22:58:44	49227...	Mosciski Group	travel	\$4.50	...	1
...

DATA EXPLORATION

Imbalanced classes

Transaction Date/Time

Amount

Category

Data Exploration

Imbalanced classes

Transaction Date/Time

Amount

Category

	Number of Transactions	Percentage
VALID	1,842,592	99.479%
FRAUD	9,651	0.521%

Data Exploration

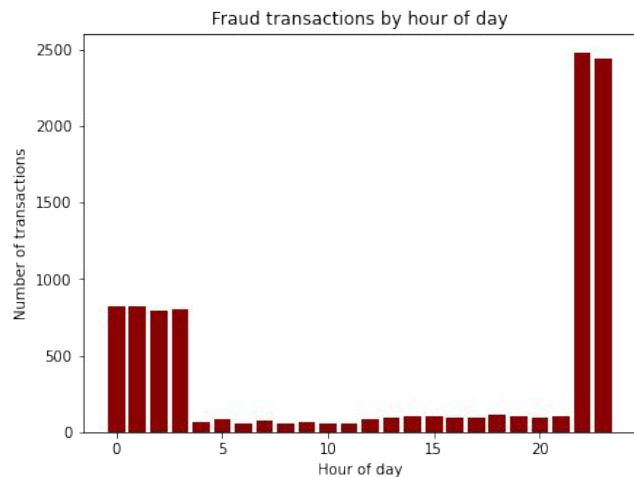
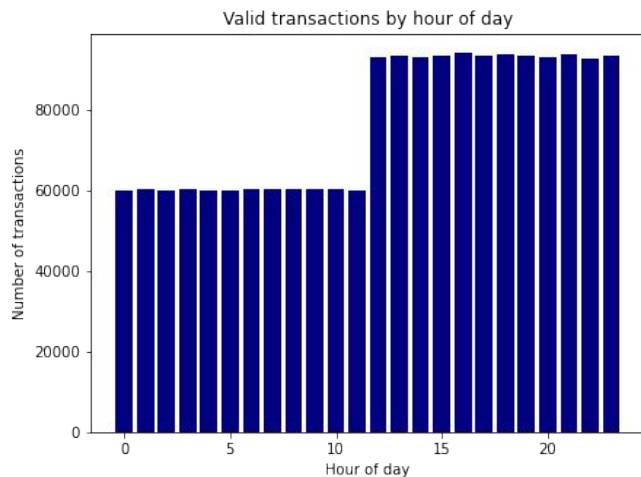
Imbalanced classes

Transaction Date/Time

Amount

Category

Hour of day



Data Exploration

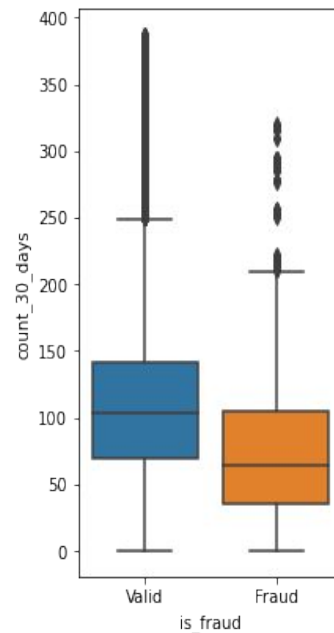
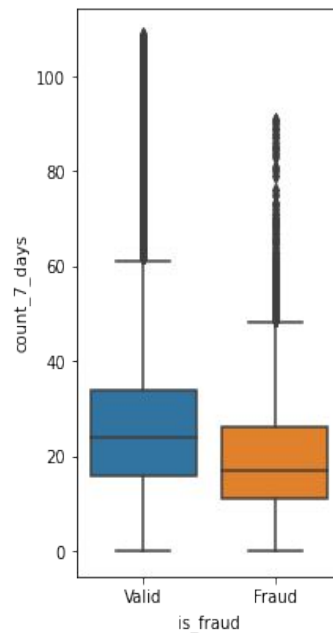
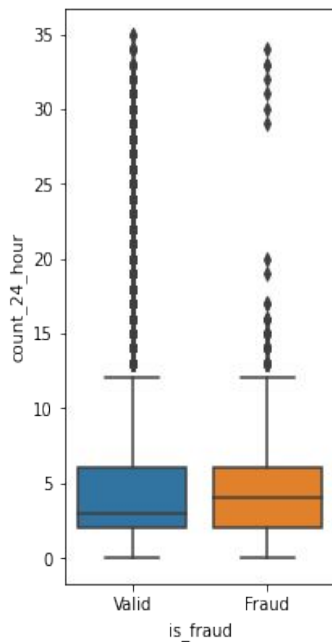
Imbalanced classes

Transaction Date/Time

Amount

Category

Transaction Counts



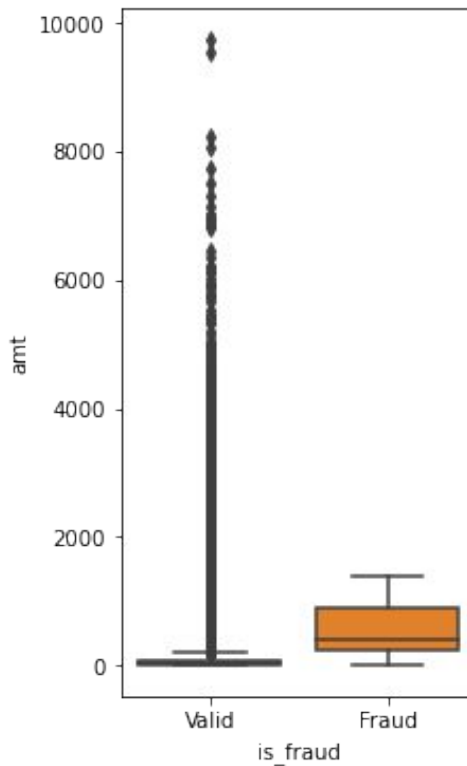
Data Exploration

Imbalanced classes

Transaction Date/Time

Amount

Category



Mean amount

VALID	\$66.88
-------	---------

FRAUD	\$530.66
-------	----------

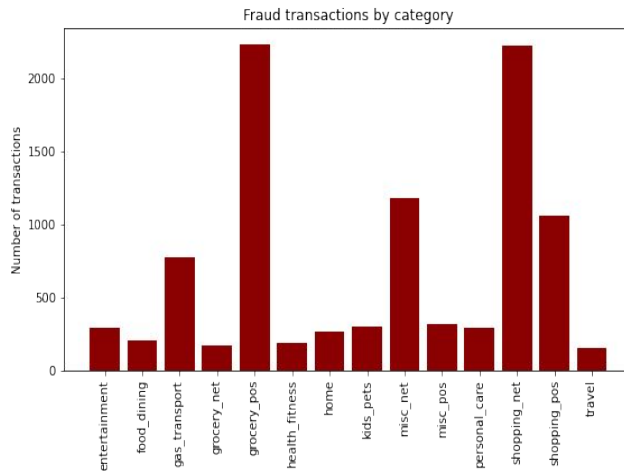
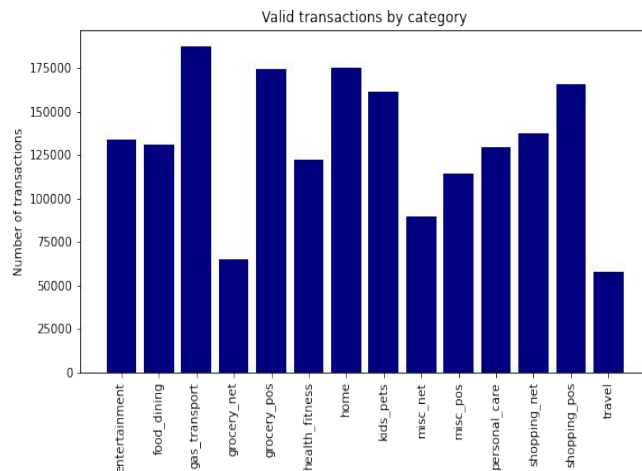
Data Exploration

Imbalanced classes

Transaction Date/Time

Amount

Category



Top categories for fraud:

- grocery_pos
- shopping_net
- misc_net
- shopping_pos
- gas_transport

MODELING

Modeling Overview

Class Balancing and Algorithms

Comparisons

Best Model

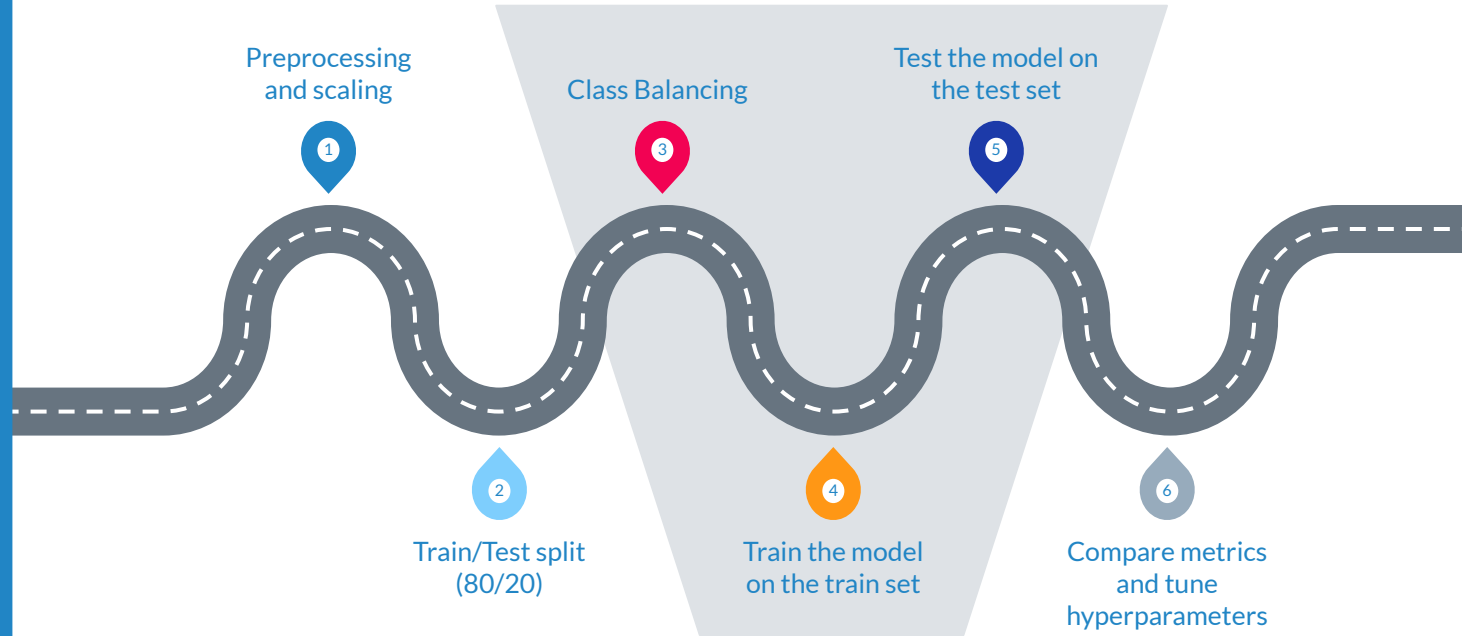
Modeling

Modeling overview

Class Balancing and Algorithms

Comparisons

Best Model



Modeling

Modeling overview

Class Balancing and Algorithms

Comparisons

Best Model

Class Balancing Techniques

- Random undersampling
- Random oversampling
- SMOTE
- Balancing class_weight parameter in algorithm

Classification Algorithms

- Logistic Regression
- Random Forest
- XGBoost

Modeling

Modeling overview

Class Balancing and Algorithms

Comparisons

Best Model

Top three models

Algorithm	Class Balancing	Valid class			Fraud class		
		<i>Precision</i>	<i>Recall</i>	<i>f1</i>	<i>Precision</i>	<i>Recall</i>	<i>f1</i>
Random Forest	Random oversampling	1.00	1.00	1.00	0.93	0.71	0.80
Random Forest	SMOTE	1.00	1.00	1.00	0.77	0.80	0.79
XGBoost	SMOTE	1.00	1.00	1.00	0.85	0.81	0.83

Modeling

Modeling overview

Class Balancing and Algorithms

Comparisons

Best Model

Best model

Algorithm	Class Balancing	Valid class			Fraud class		
		<i>Precision</i>	<i>Recall</i>	<i>f1</i>	<i>Precision</i>	<i>Recall</i>	<i>f1</i>
Random Forest	Random oversampling	1.00	1.00	1.00	0.93	0.71	0.80
Random Forest	SMOTE	1.00	1.00	1.00	0.77	0.80	0.79
XGBoost	SMOTE	1.00	1.00	1.00	0.85	0.81	0.83

Modeling

Modeling overview

Class Balancing and Algorithms

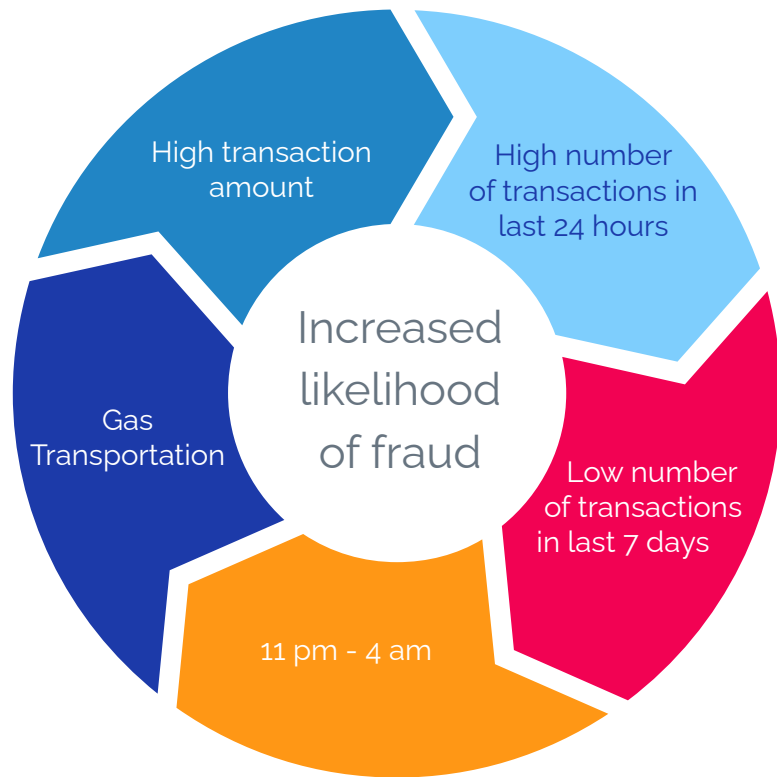
Comparisons

Best Model

Confusion matrix

		Predicted label	
		Valid	Fraud
Actual label	Valid	368242	277
	Fraud	362	1568

Feature impact



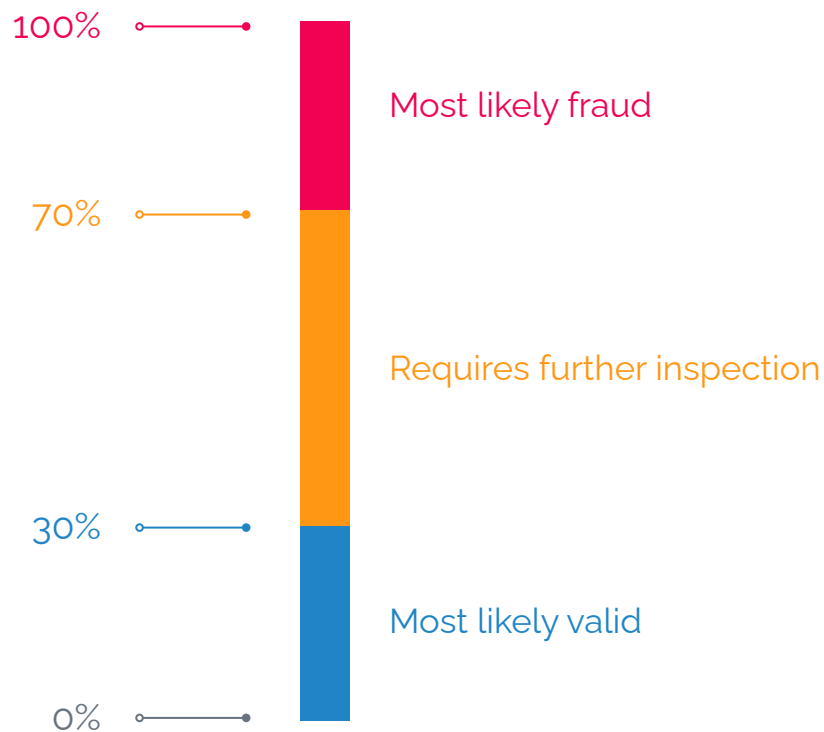
Future work

- ▷ Engineer 'jobs' and 'state' columns
- ▷ Alternative resampling methods
- ▷ Explore *cost* of false negative vs false positive
- ▷ Explore predicted probabilities of fraud



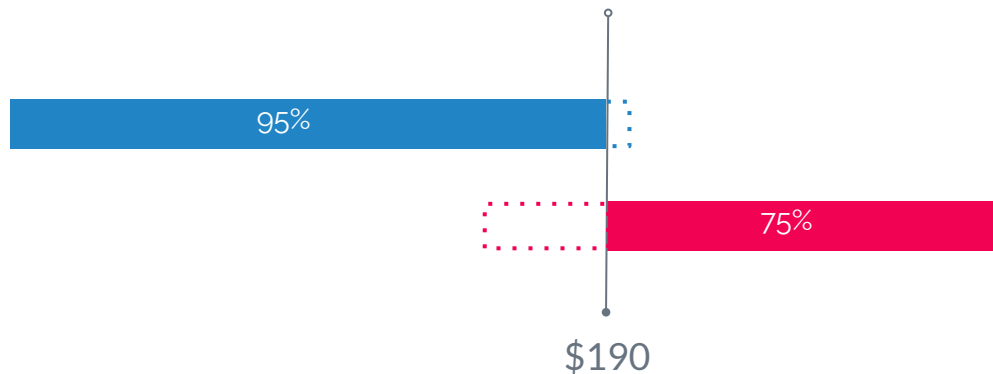
Recommendations

1. Create probability thresholds to categorize transactions



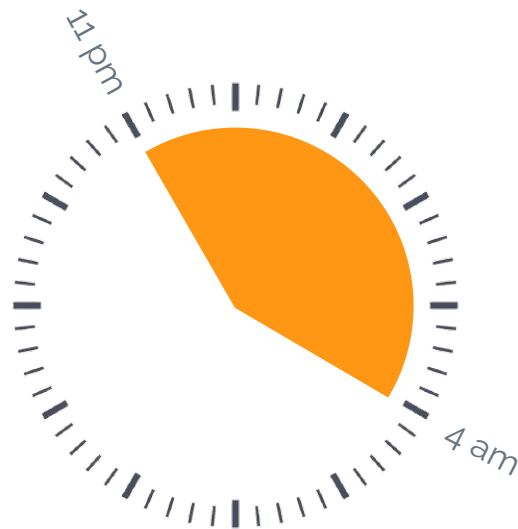
Recommendations

1. Create probability thresholds to categorize transactions
2. Inspect further if amount is greater than \$190



Recommendations

1. Create probability thresholds to categorize transactions
2. Inspect further if amount is greater than \$190
3. Inspect further if transaction takes place during abnormal hours





Thank you!

Email: cchung106@gmail.com
LinkedIn: <https://www.linkedin.com/in/cchung106/>
GitHub repo: <https://github.com/cchung106/Capstone-Two>