Bank Account Application Fraud Detection

Hadley Campbell, Pengwei Wang, Brian Rodriguez, Payton Stauble

December 15, 2023

Meet the Team

For inquiry, please contact the team directly

Hadley Campbell

Hacampbell@deloitte.com

Brian Rodriguez

briarodriguez@deloitte.com

Pengwei Wang

pengwwang@deloitte.com

Payton Stauble

Pstauble@deloitte.com

Agenda

Business Mission

Representation and Range of Data

Overview of Process and Techniques

Supporting Visualizations

Next Steps and Continued Development

Business Mission

Imbalanced Fraud Detection

Increase **visibility** & **efficiency**

Understand **key predictors** of bank account application fraud

How will our clients benefit?

Positive impacts of the Bank Account Application Fraud Detection ML model

Baseline Model

Employers can utilize our solution as a **baseline model for specific use cases.**

Intelligence

Employing our Bank Account Application Fraud model will **increase visibility** and understanding of **driving factors of fraud** specific to the business needs.

Cost Savings

Identifying high-risk bank account application features decreases the chances of fraud and ultimately **remediation costs** and **financial loss.**

Efficiency

Supervised learning models **reduce administrative burden** by **automating detection processes** and allowing time for growth in additional areas.

Data Understanding

A glimpse into the data used for training, testing, and model building

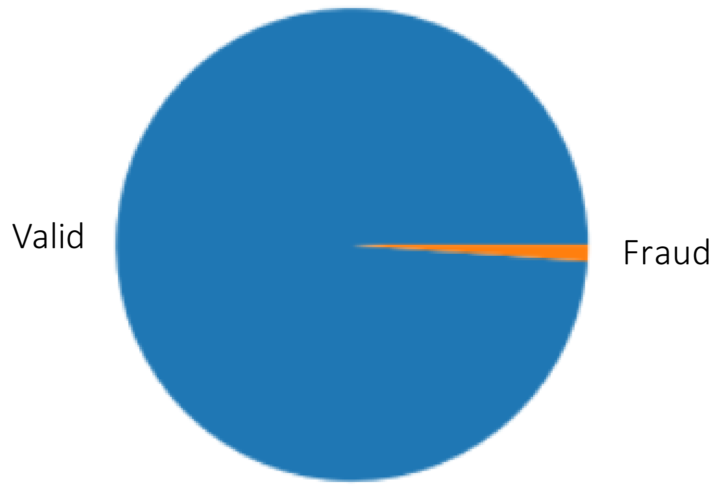
30 Variables | 25 Numerical | 5 Categorical

All user, device, and application data used

1 Million Bank Account Applications Analyzed

Highly Imbalanced Target Label

Proportion of Fraudulent vs Valid Observations



Data Overview

0 fraud_bool
1 income
2 name_email_similarity
3 prev_address_months_count
4 current_address_months_count
5 customer_age
6 days_since_request
7 intended_balcon_amount
8 payment_type
9 zip_count_4w
10 velocity_6h

11 velocity_24h
12 velocity_4w
13 bank_branch_count_8w
14 date_of_birth_distinct_emails_4w
15 employment_status
16 credit_risk_score
17 email_is_free
18 housing_status
19 phone_home_valid
20 phone_mobile_valid
21 bank_months_count

22 has_other_cards
23 proposed_credit_limit
24 foreign_request
25 source
26 session_length_in_minutes
27 device_os
28 keep_alive_session
29 device_distinct_emails_8w
30 device_fraud_count
31 month

	fraud_bool	income	name_email_similarity	prev_address_months_count	current_address_months_count	customer_age
count	1000000.000000	1000000.000000	1000000.000000	1000000.000000	1000000.000000	1000000.000000
mean	0.011029	0.562696	0.493694	16.718568	86.587867	33.689080
std	0.104438	0.290343	0.289125	44.046230	88.406599	12.025799
min	0.000000	0.100000	0.000001	-1.000000	-1.000000	10.000000
25%	0.000000	0.300000	0.225216	-1.000000	19.000000	20.000000
50%	0.000000	0.600000	0.492153	-1.000000	52.000000	30.000000
75%	0.000000	0.800000	0.755567	12.000000	130.000000	40.000000
max	1.000000	0.900000	0.999999	383.000000	428.000000	90.000000

Building a Successful Model

Steps taken to create the Bank Account Application Fraud ML model from data analysis to measuring success.

EDA

- Check Nulls
- Outlier Detection
- Correlations
- Univariate Analysis

Data Processing

- Train/Test/Split
- Resampling Methods
- Standardization
- Feature Selection

Modeling Techniques

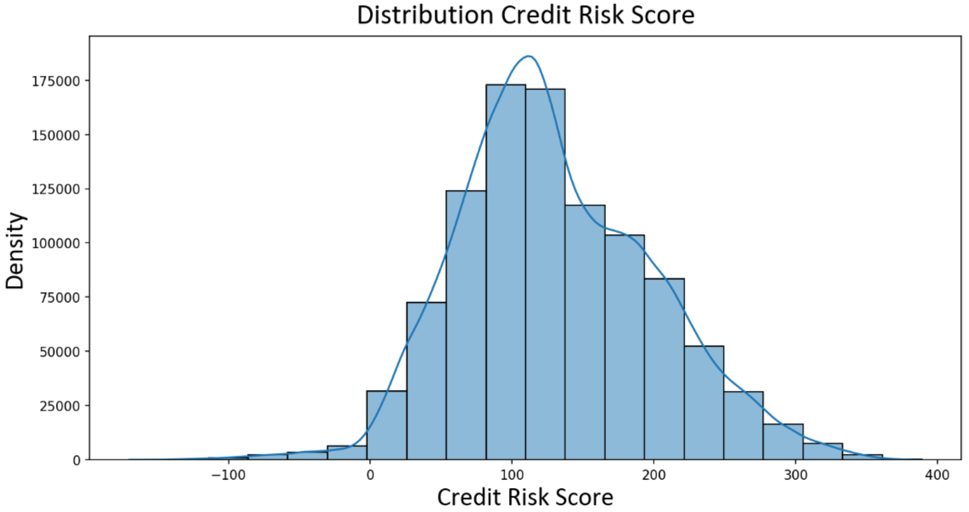
- Logistic Regression
- Ensemble Methods
- Hyperparameter Tuning
- Cost-Sensitive Learning

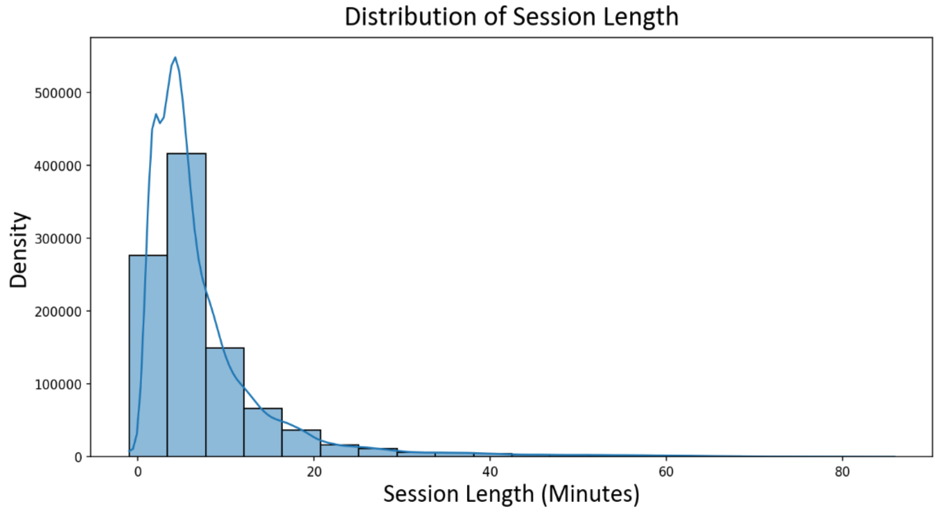
Success Metrics

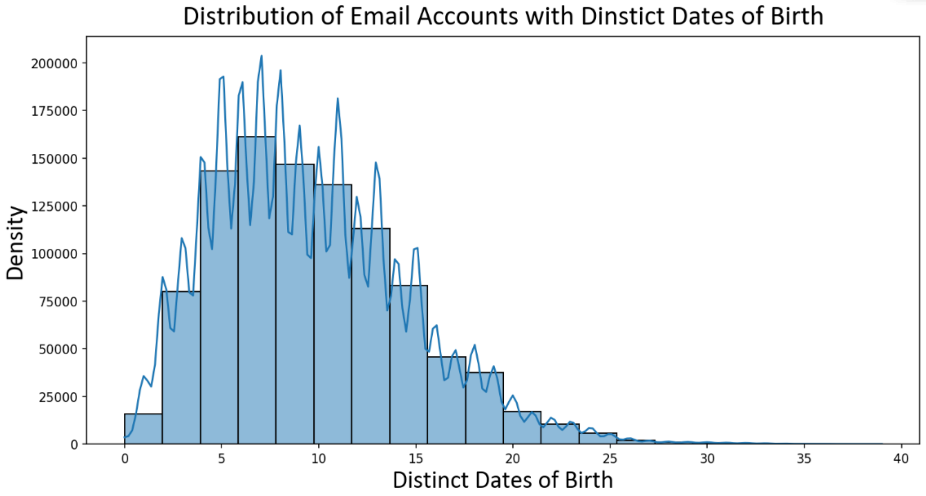
- ROC + AUC
- Precision/Recall
- Confusion Matrix

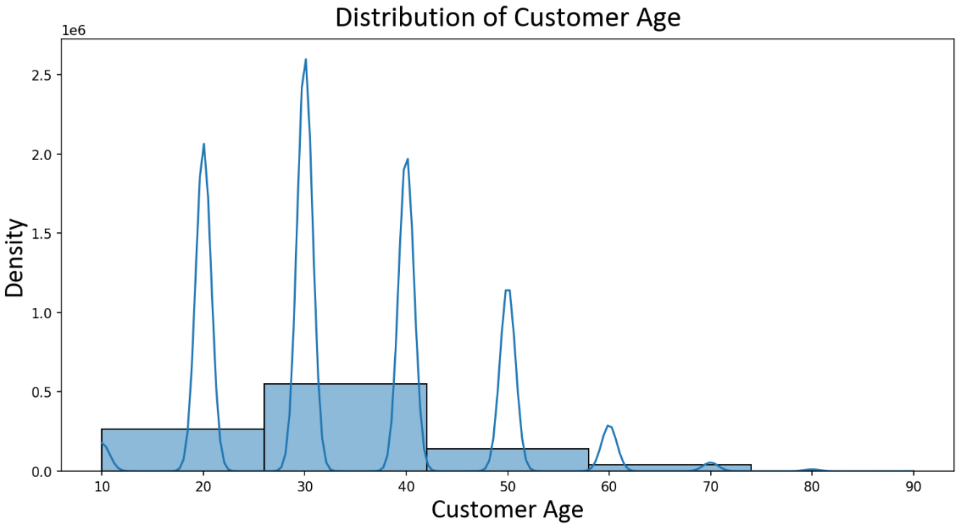
Supporting Visualizations

EDA: Key Distributions

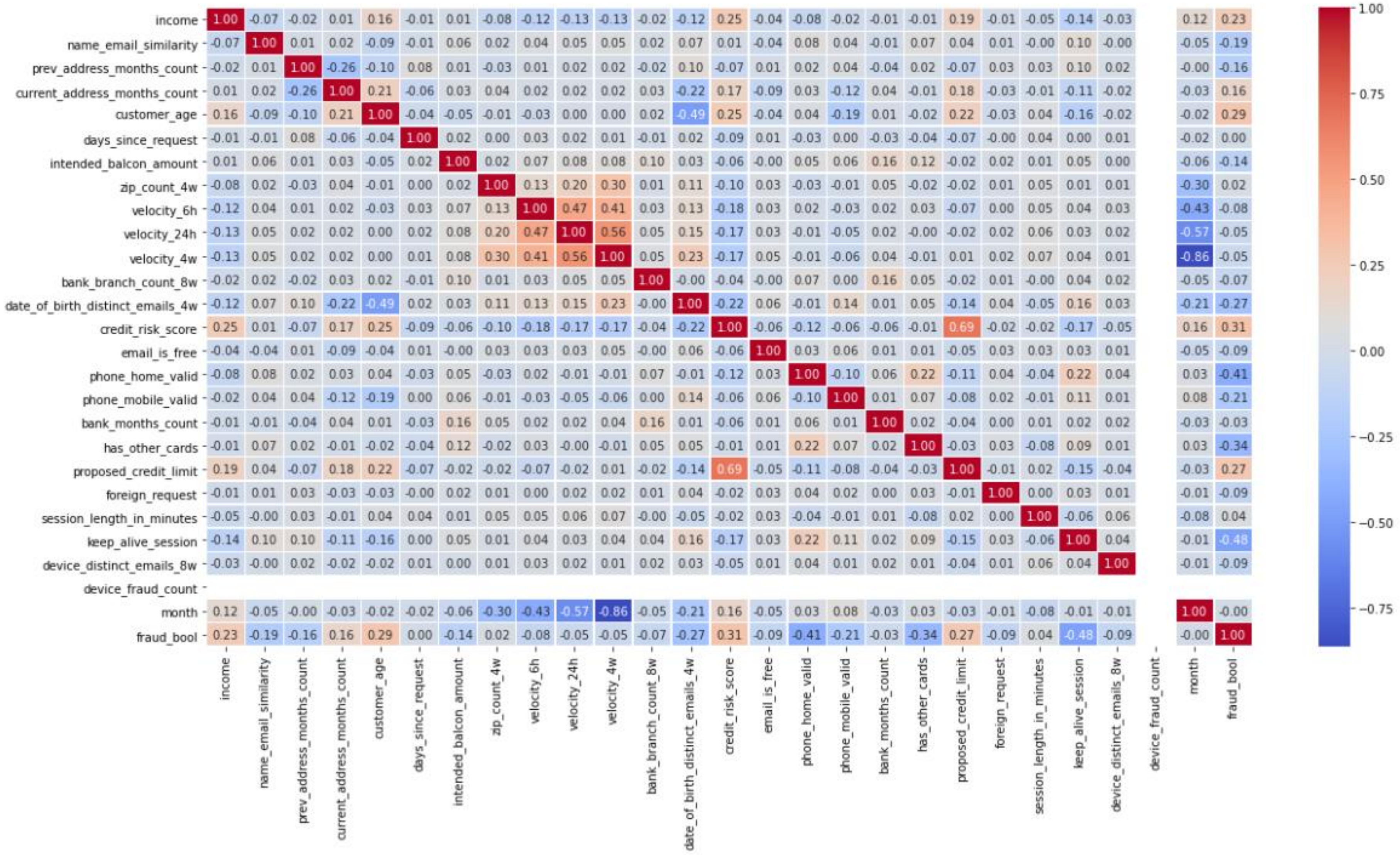








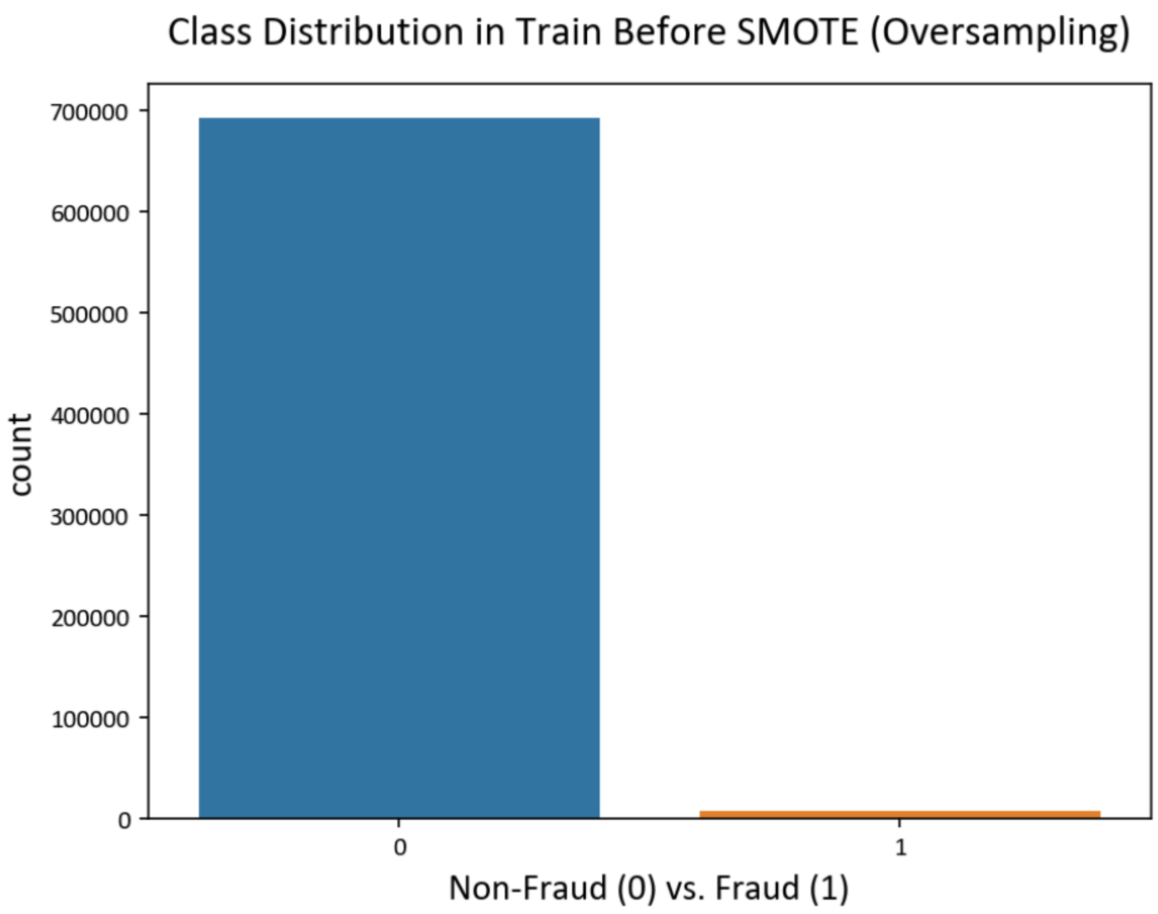
Scaled Heat Map

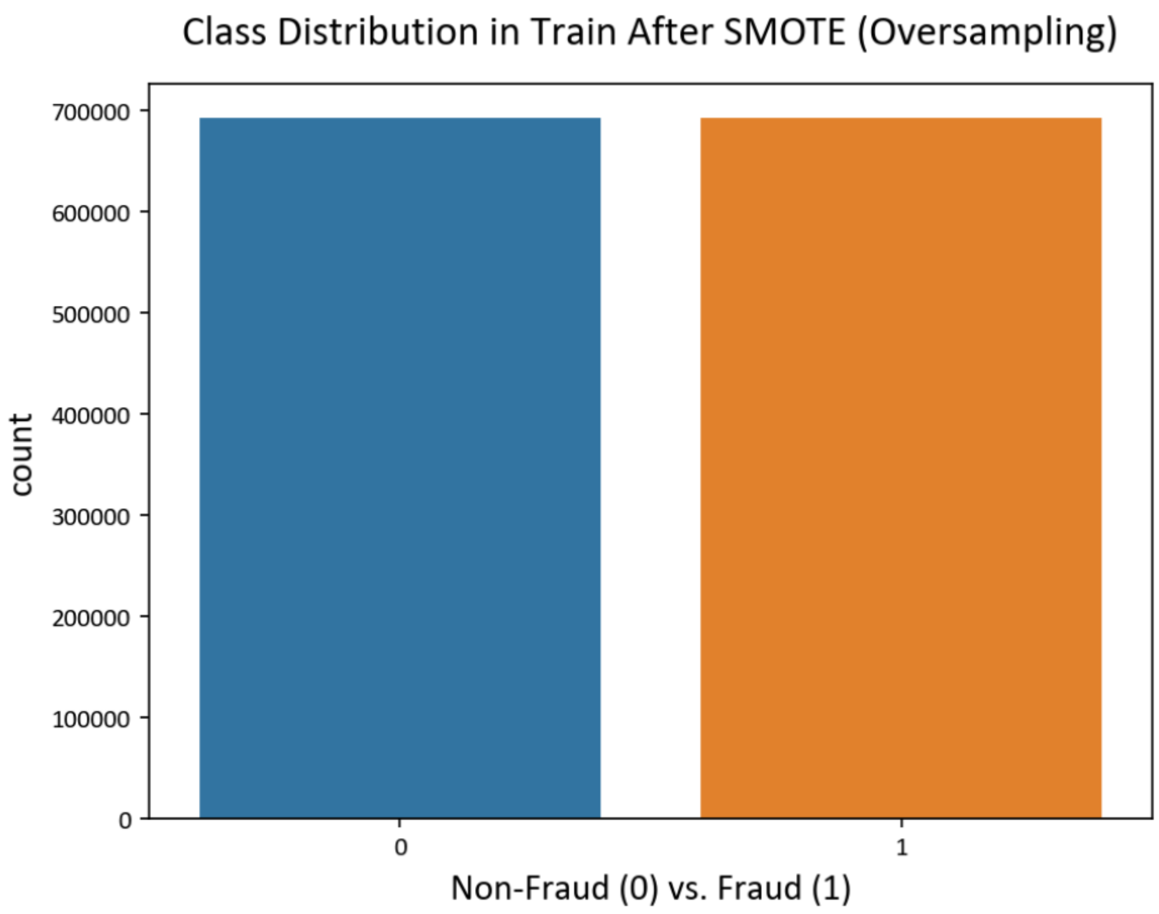


Highly Correlated Numeric Features

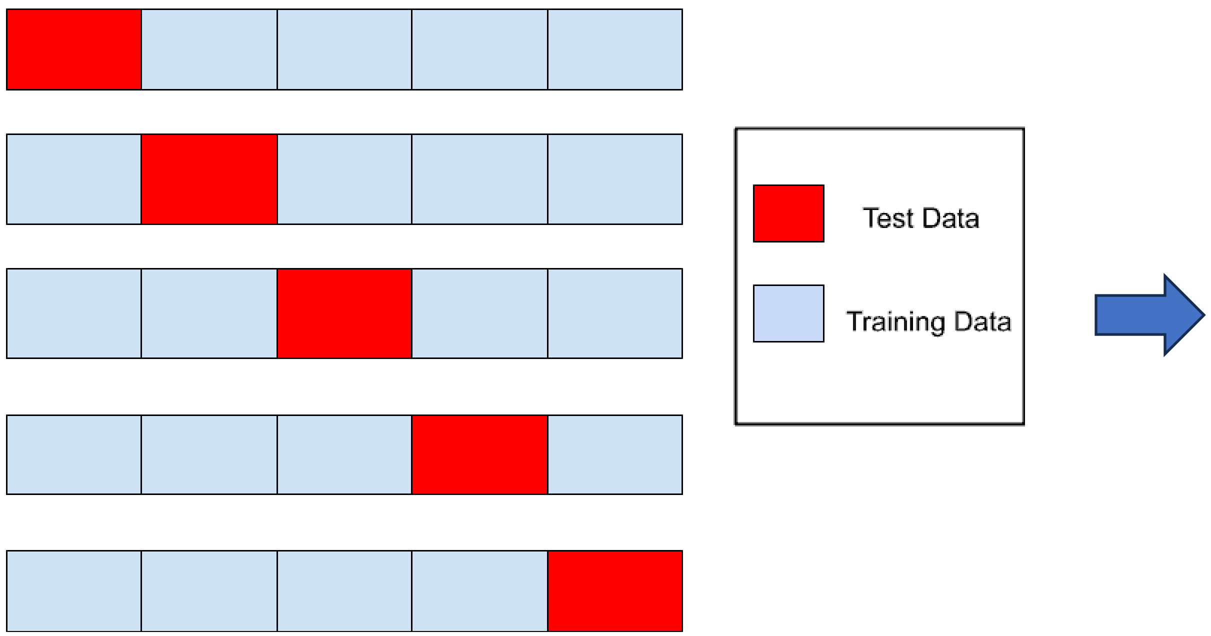
- Income
- Customer Age
- Proposed Credit Limit
- Credit Risk Score

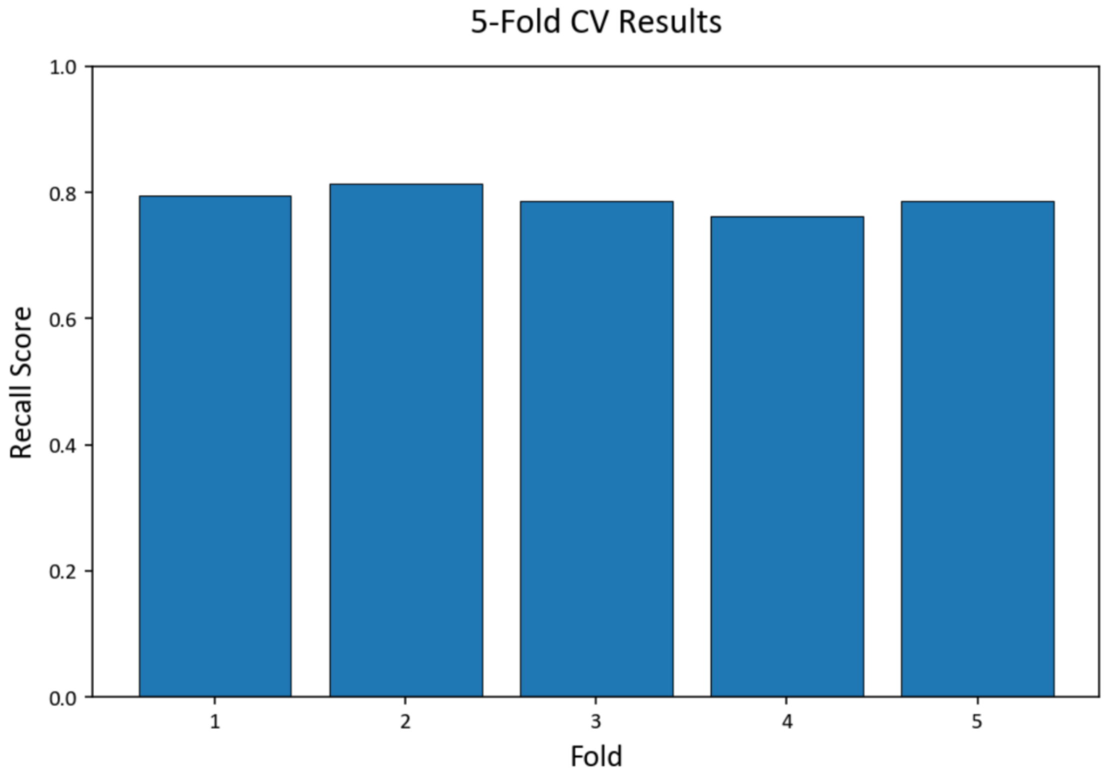
Target Class Proportions



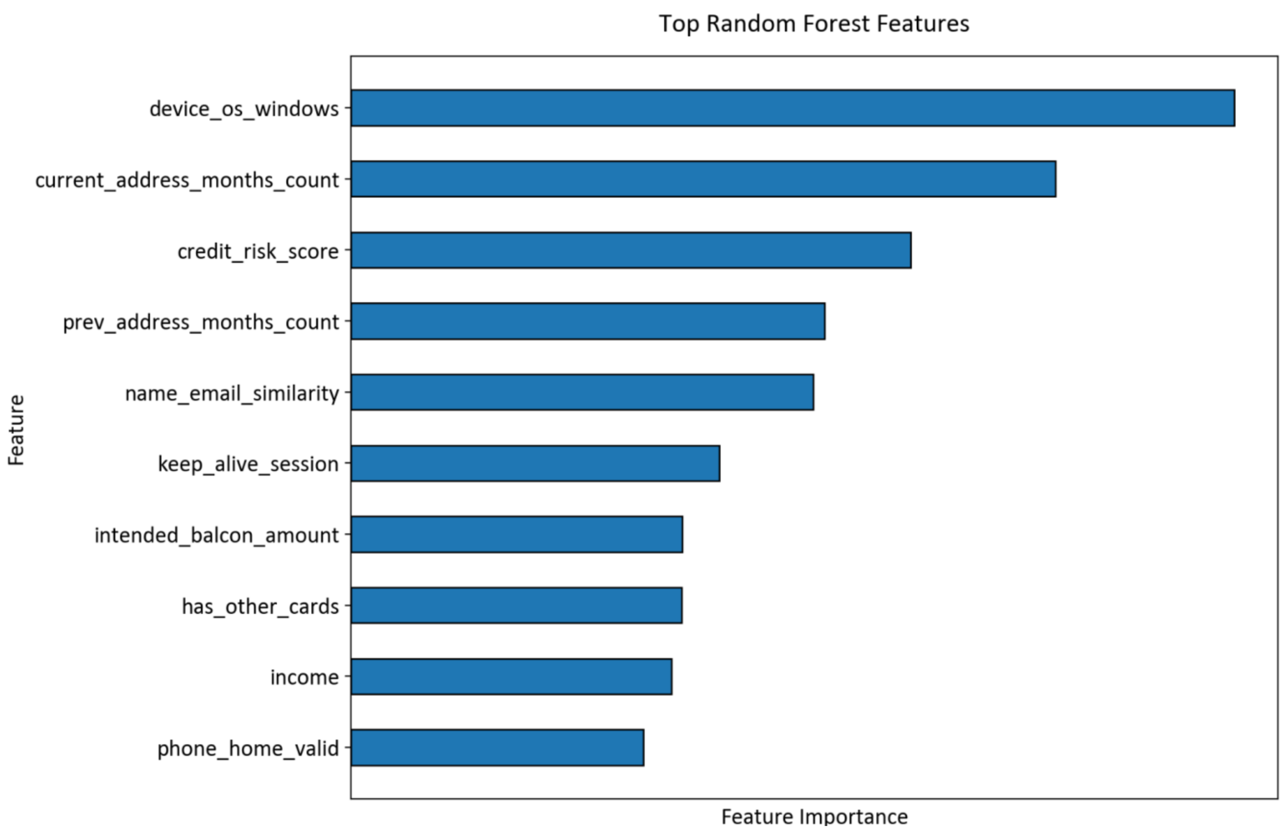


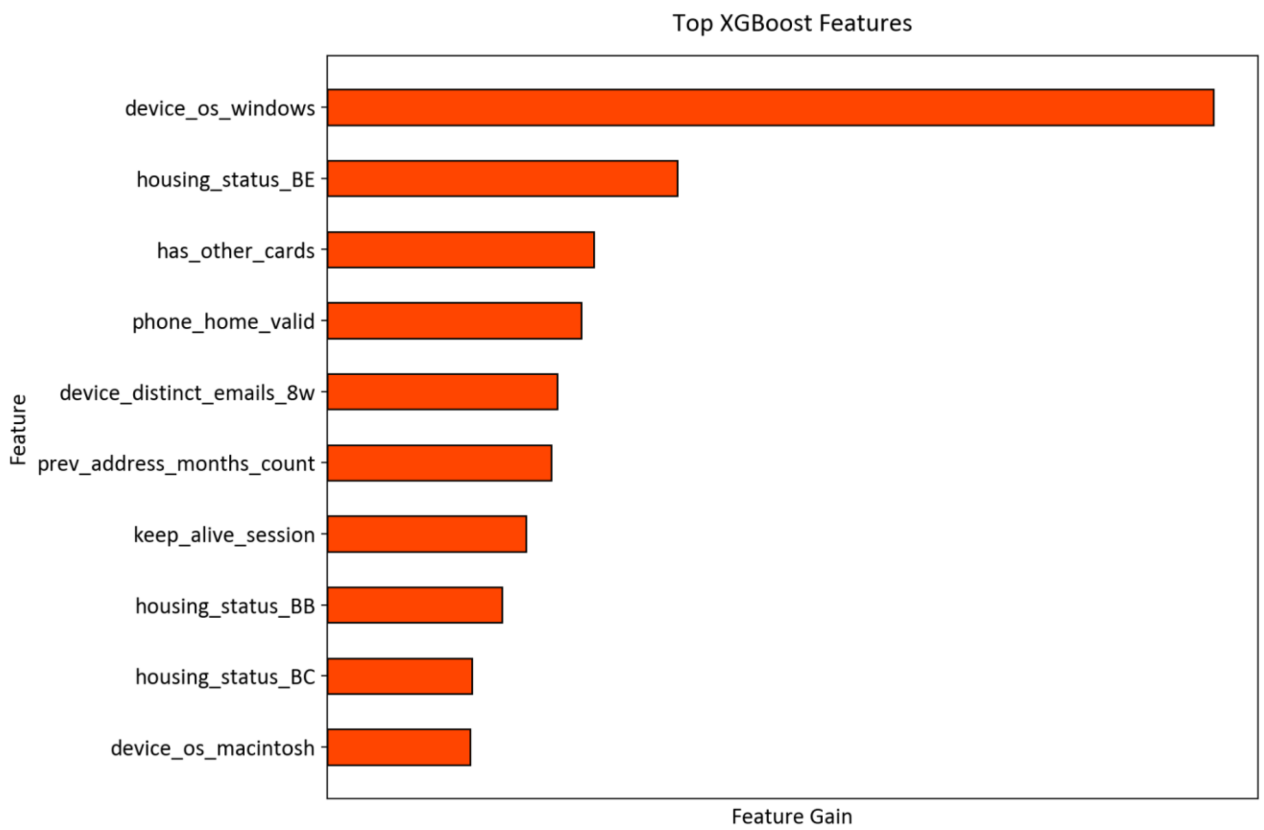
Cross Validation: Recall



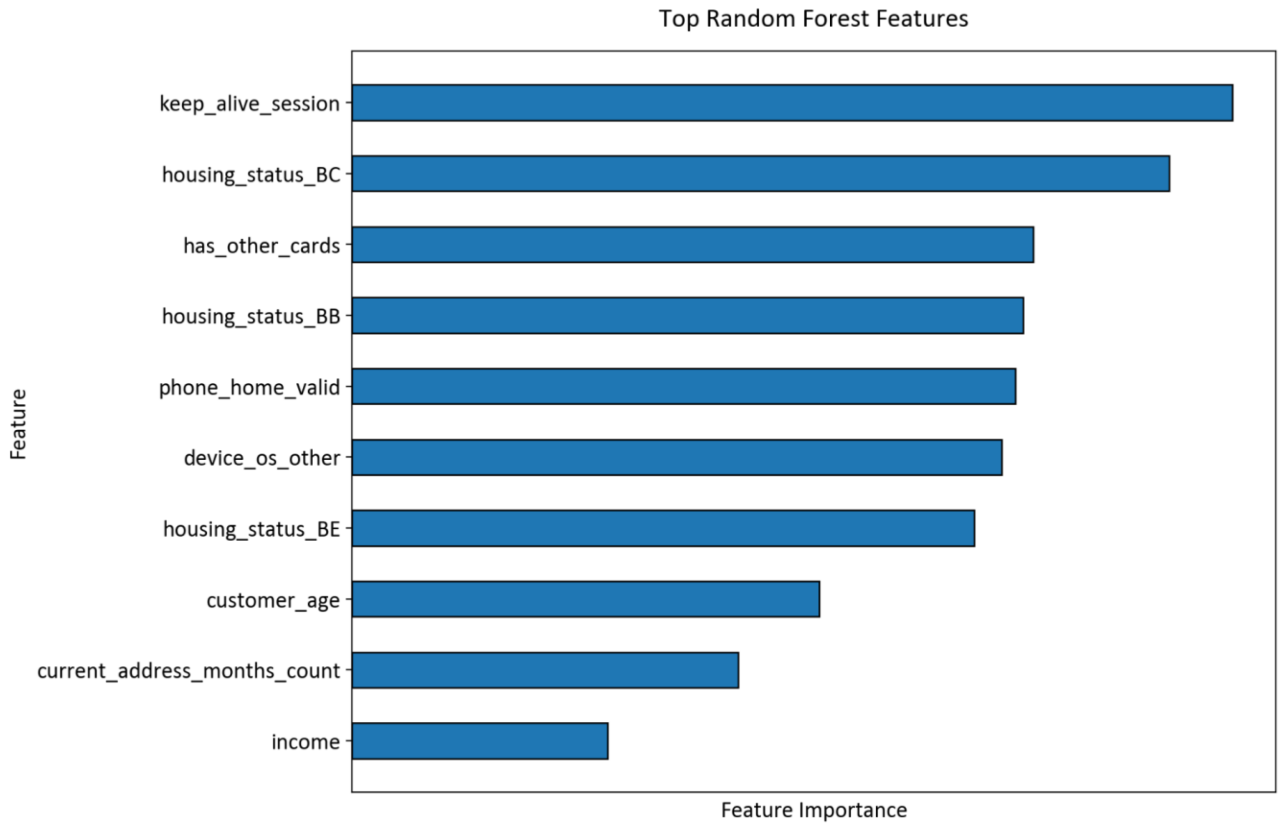


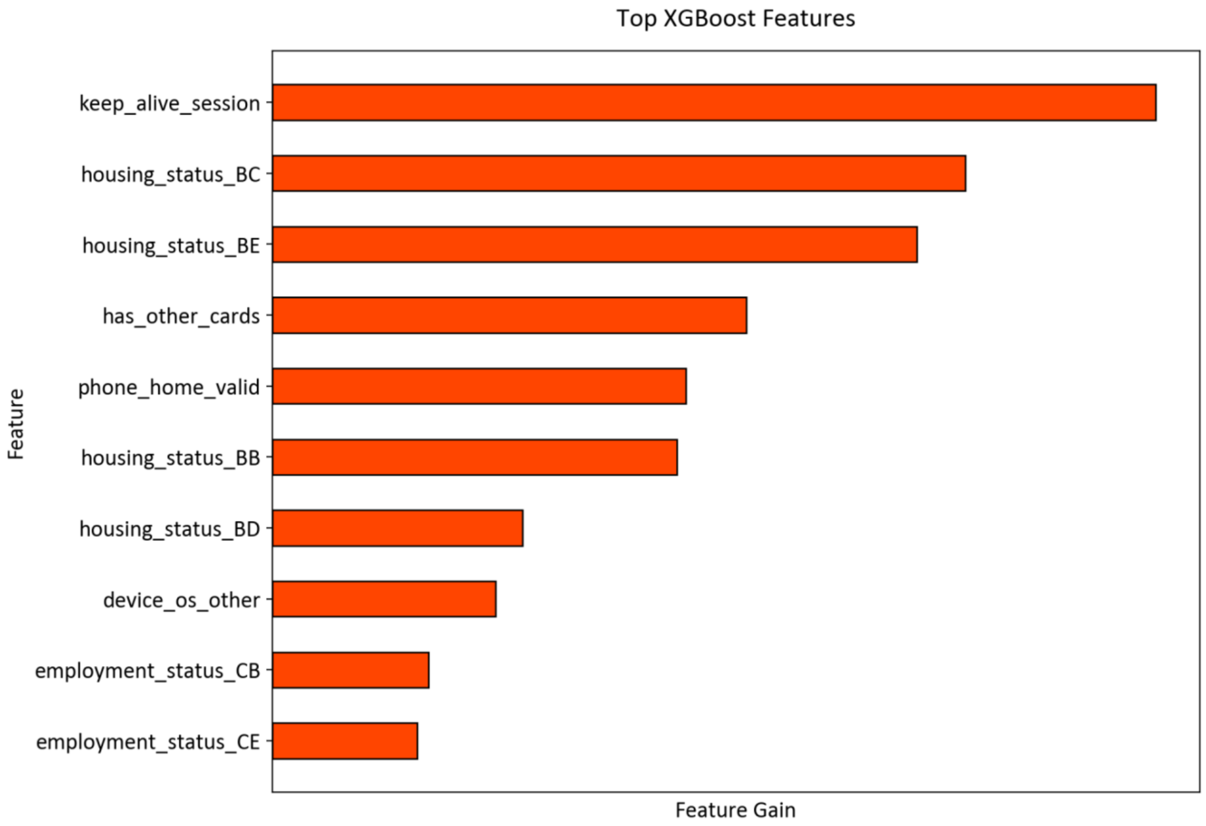
Feature Importance: Undersampling



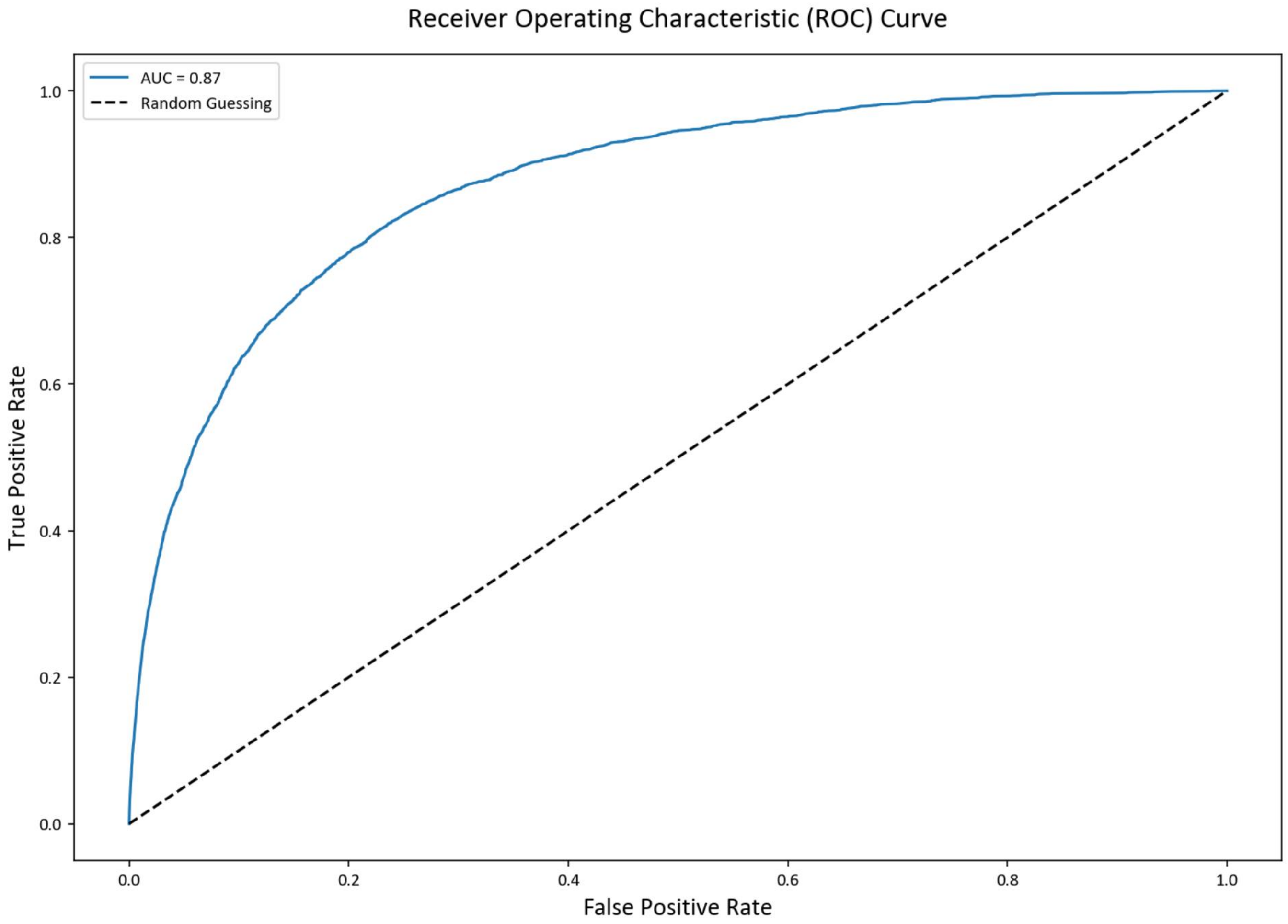


Feature Importance: Oversampling





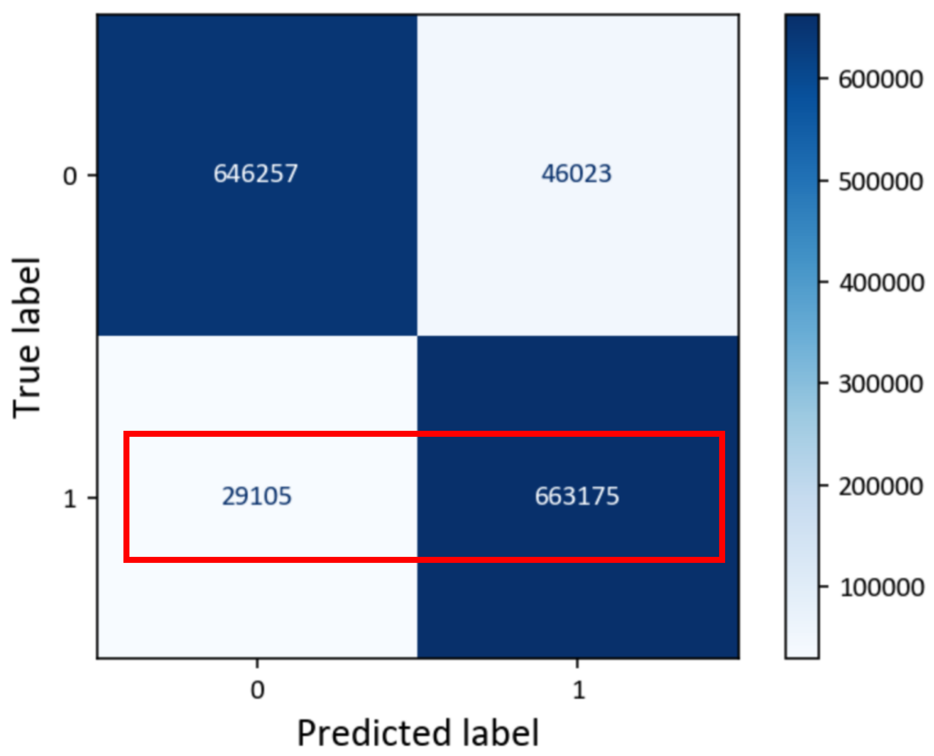
Success Metrics: ROC AUC



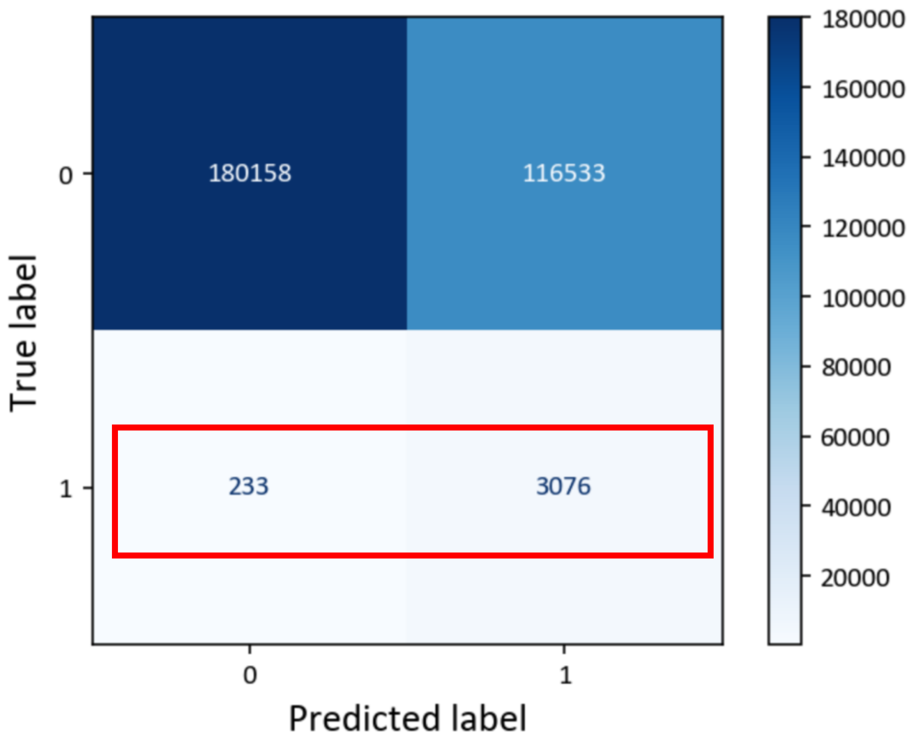
AUC in Test: 0.867

Success Metrics: Confusion Matrix

Training Accuracy: 0.95



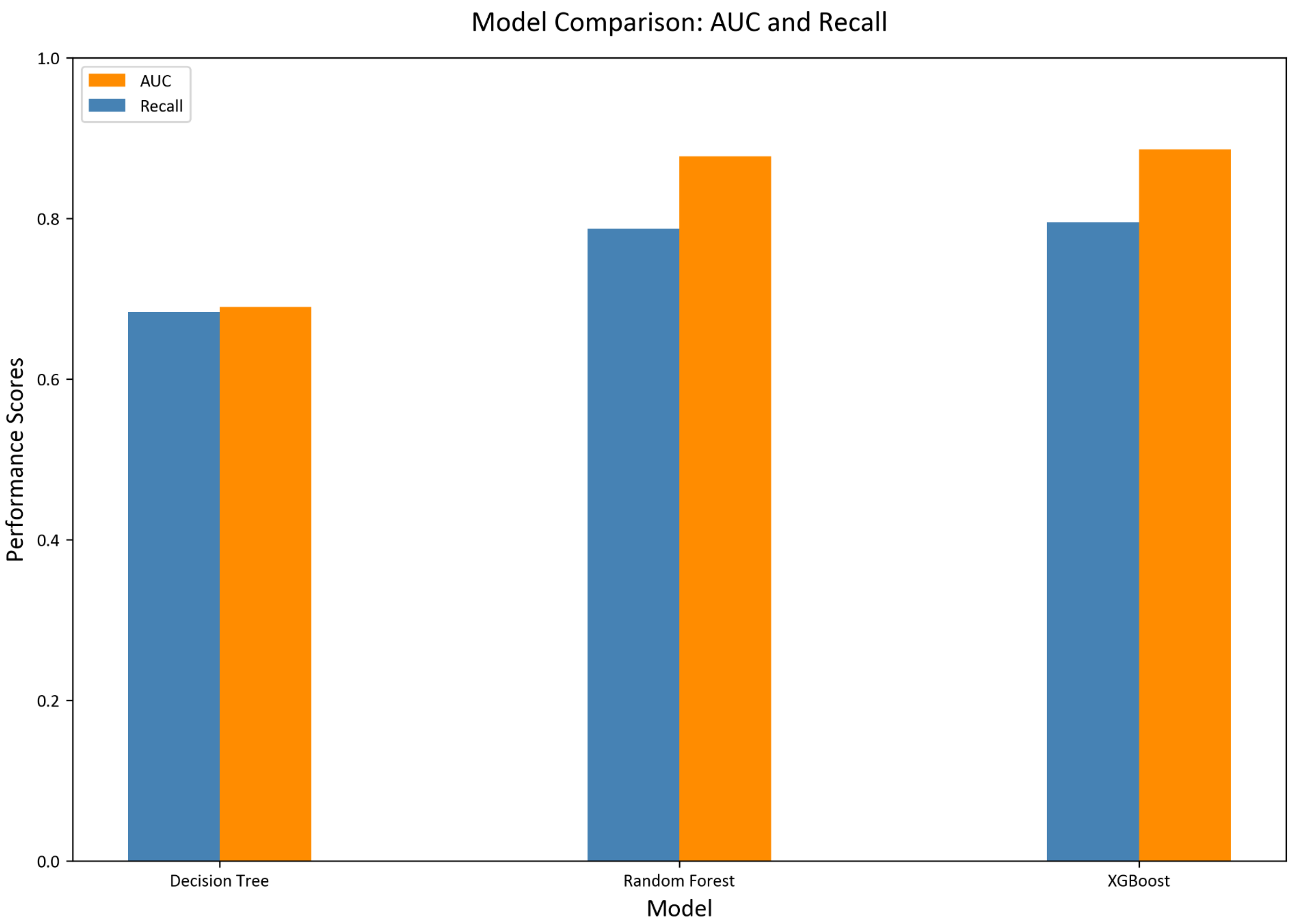
Test Accuracy: 0.90



Threshold	Recall (FPR)
0.5	78%

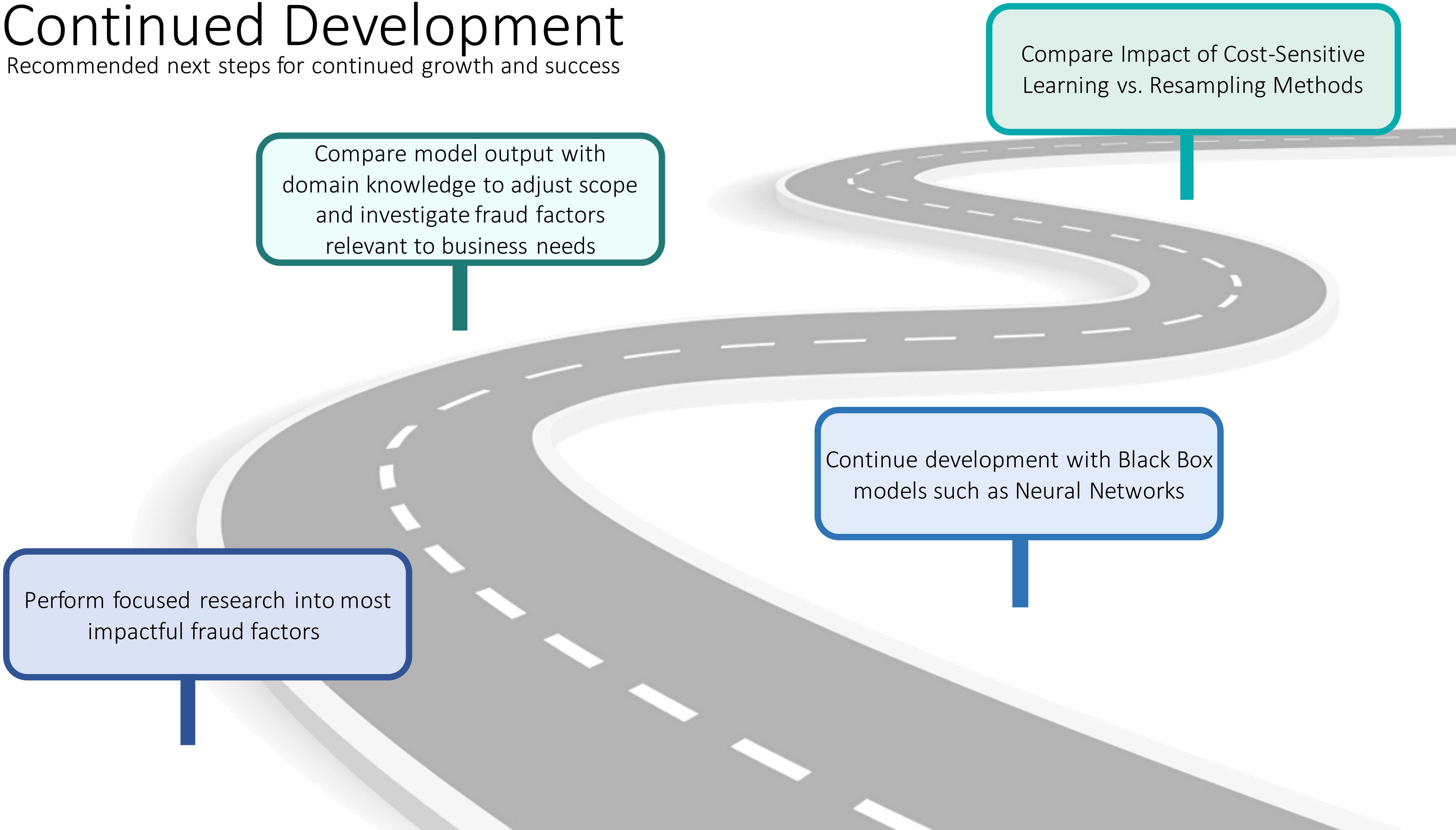
Threshold	Recall (FPR)
0.3	91%

Evaluation: Model Comparison



Continued Development

Recommended next steps for continued growth and success



Questions?