

Money Laundering: A Predictive Model Approach

Sarah Hollingsworth

Department of Data Analytics,
Franklin University

DATA 695: Capstone

Dr. Deanna Larson

November 1, 2022



Business Understanding



Exploratory Data Analysis



Modeling



Recommendation

Money Laundering

Illegal Activities

- Drug Trafficking
- Gambling
- Organized Crime

Placement

- Structuring
- Blending
- Smurfing

Insertion

- Real Estate
- Legitimate Businesses
- Investments & Securities

Placement



Structuring

- Transactions lower than reporting threshold amounts

Blending

- Combining legitimate and illegitimate proceeds through a legitimate cash-based business

Smurfing

- Deposit funds across multiple accounts below reporting thresholds

Insertion

Once the funds have been placed into the financial system, they are then used to purchase other assets in order to appear legitimate:

- Real Estate
- Legitimate Businesses
- Investments & Securities



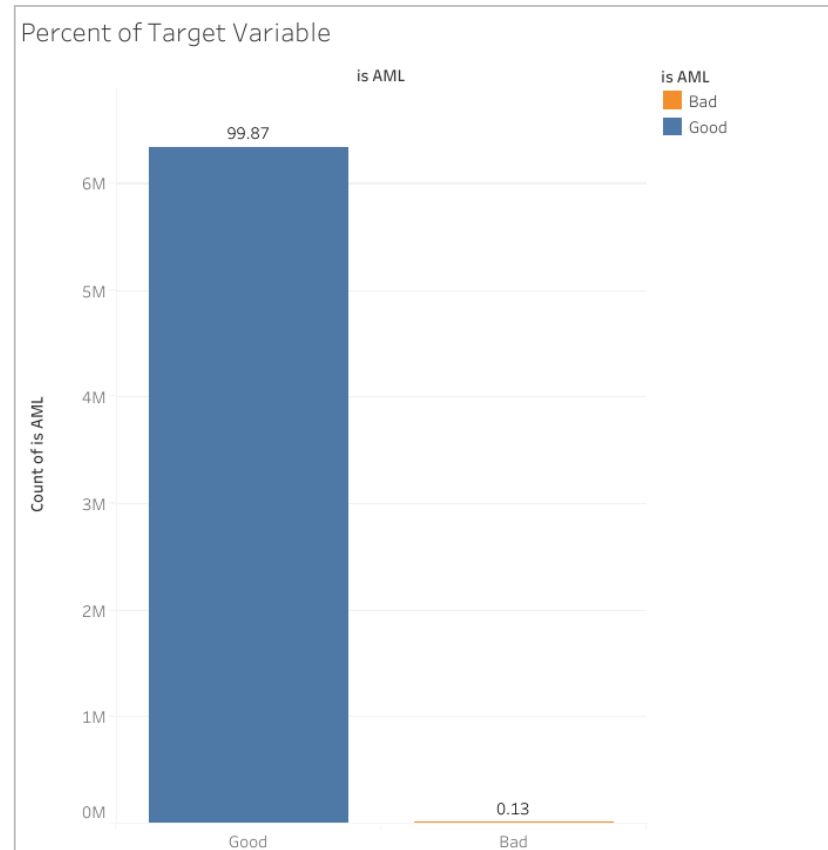
Predictive Model Approach

- Once the funds have passed into the financial system, how can they be tracked?
- Proposal: A predictive machine learning model to identify money movement in the Financial Institution
 - Identify Potential Money Laundering Transactions
 - Perform a Reasonable Investigation
 - Referral to Law Enforcement



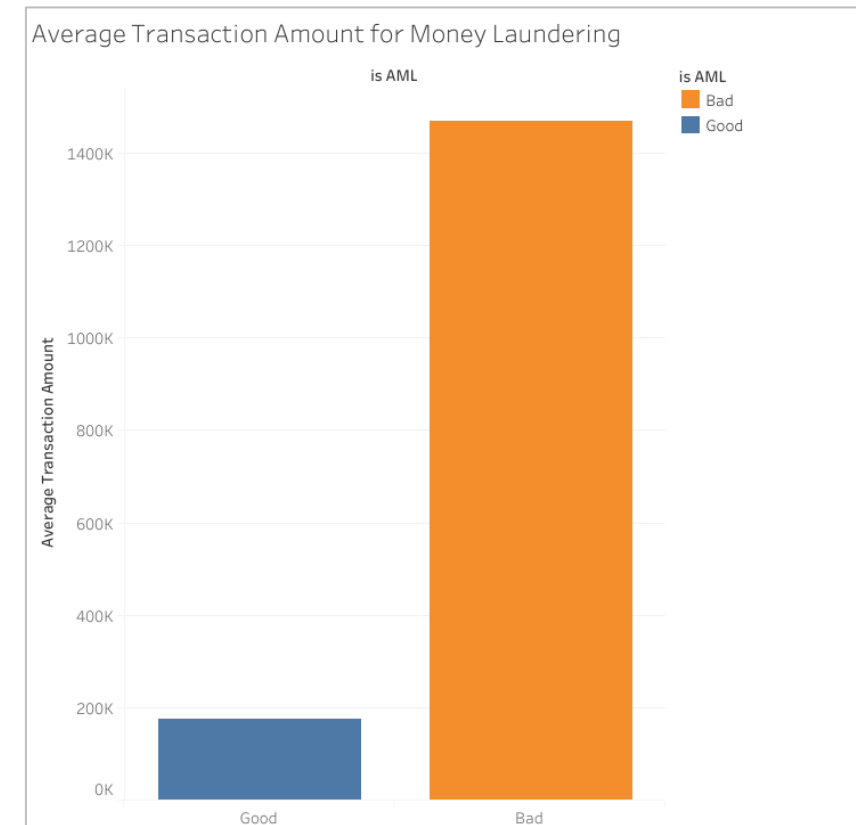
Exploratory Data Analysis

Target Variable



0.13% of all Transactions are in the
Target Class

Average Transaction Amount for
Money Laundering Transactions is
7X greater



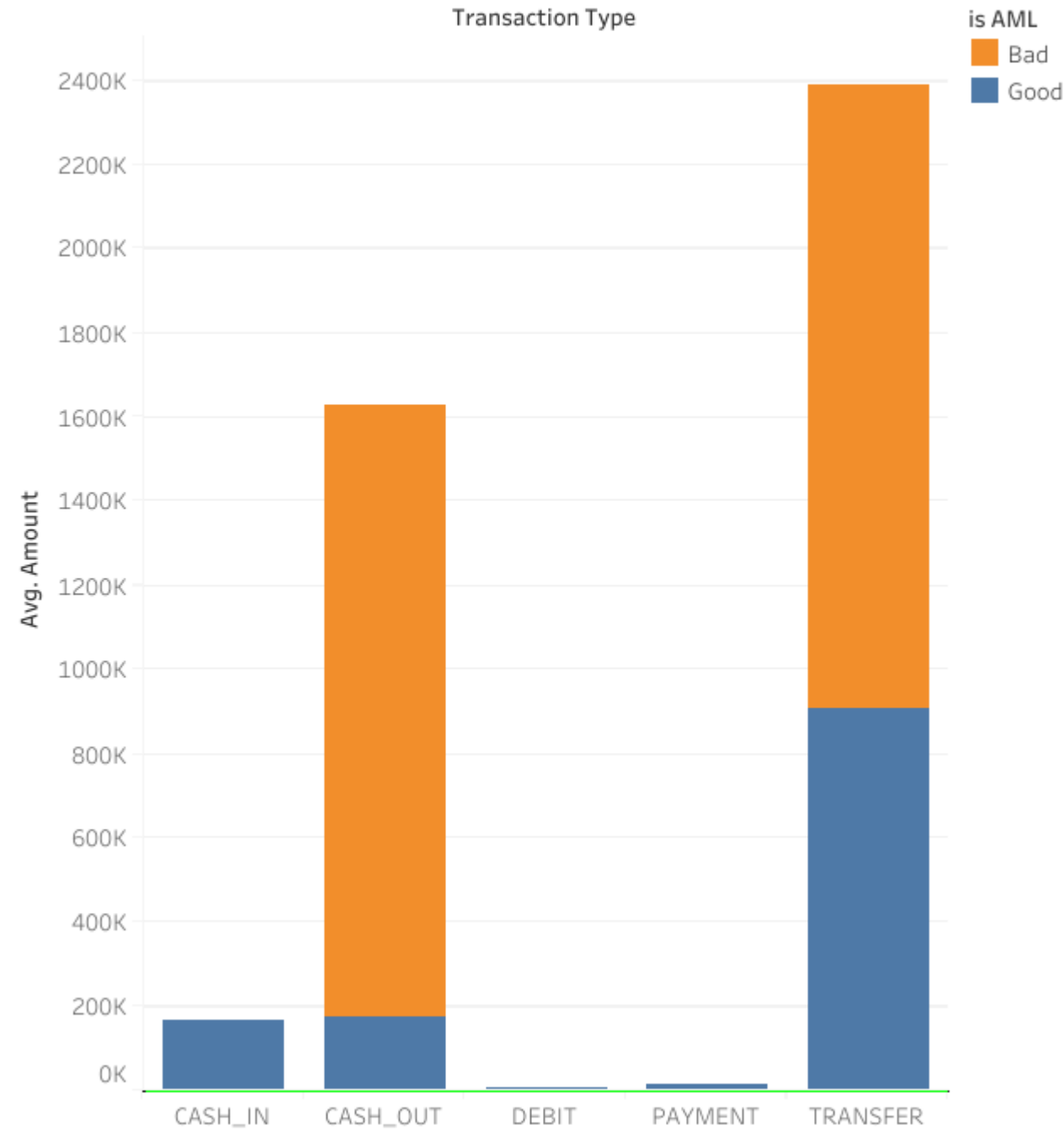
Transaction Type Matters

All Money Laundering Transactions are in Transfer and Cash Out Categories

Cash Out Money Laundering Transactions are 7X greater

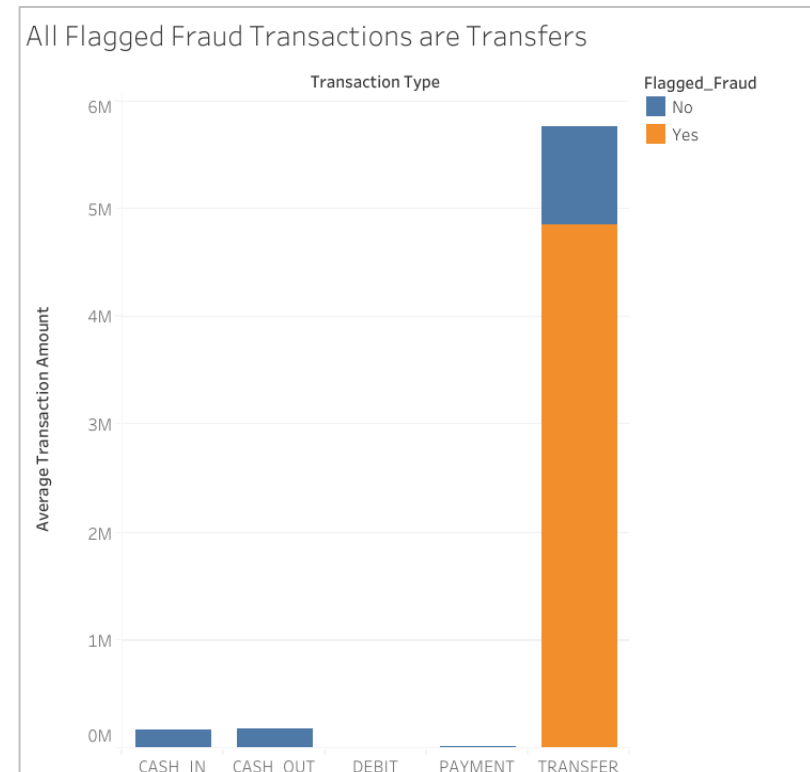
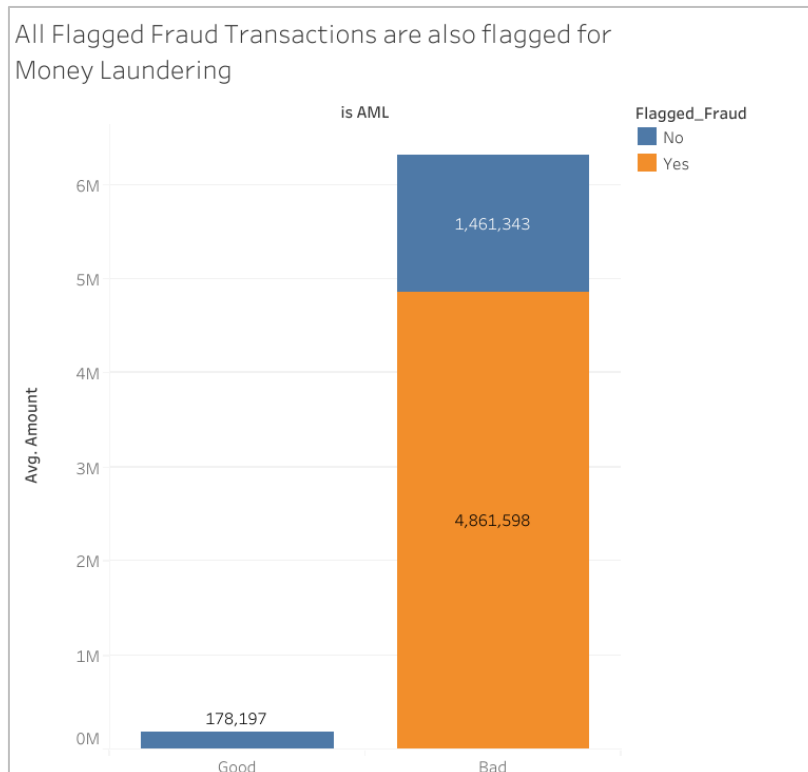
Transfer Money Laundering Transactions are 1.5X greater

Average Transaction Amount by Transaction Type



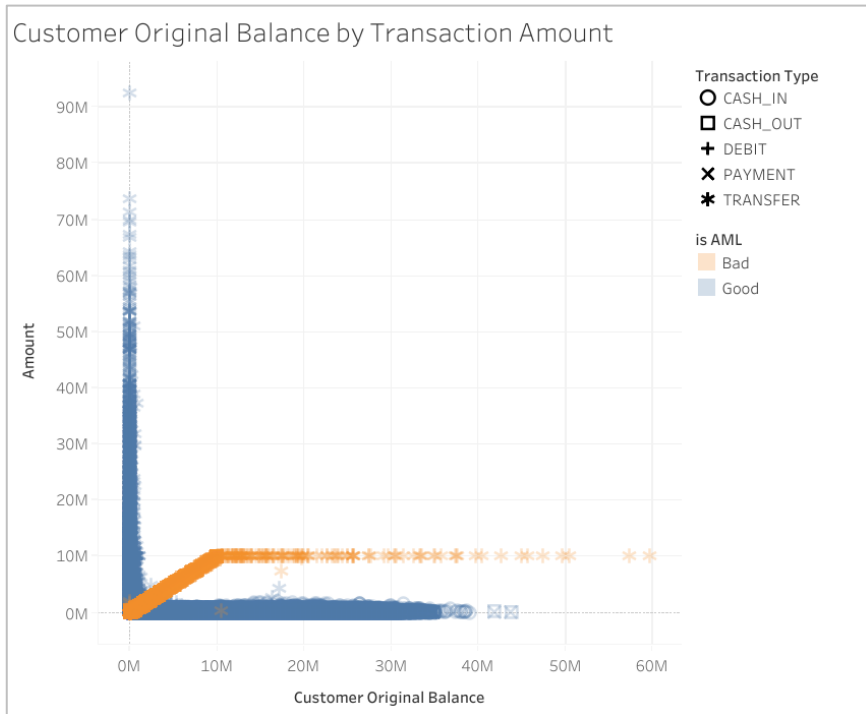
Does Flagged Fraud equal Money Laundering?

All Flagged Fraud Transactions are in the Money Laundering Category



All Flagged Fraud Transactions are in the Transfer Category

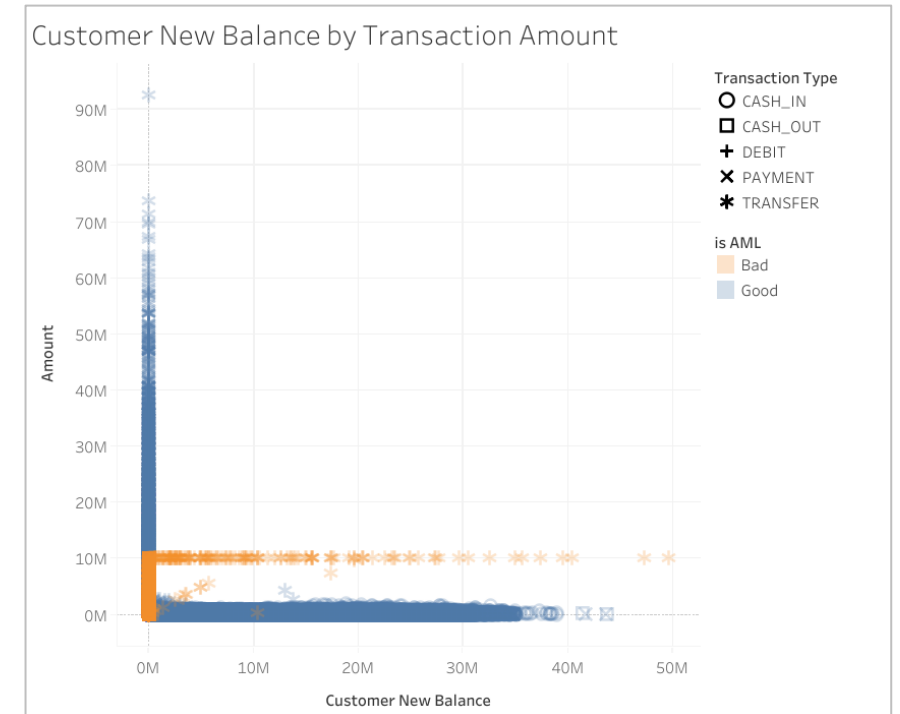
Customer Balance



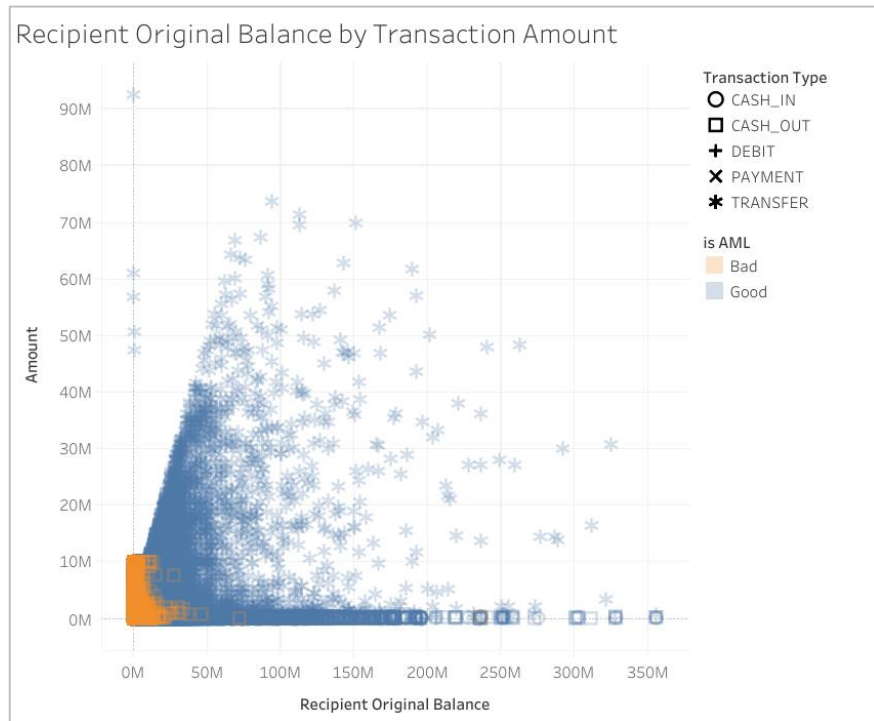
Correlation between
Customer Original Balance &
Transaction Amount

Transaction Amount is \$10
million or less for Money
Laundering Transactions

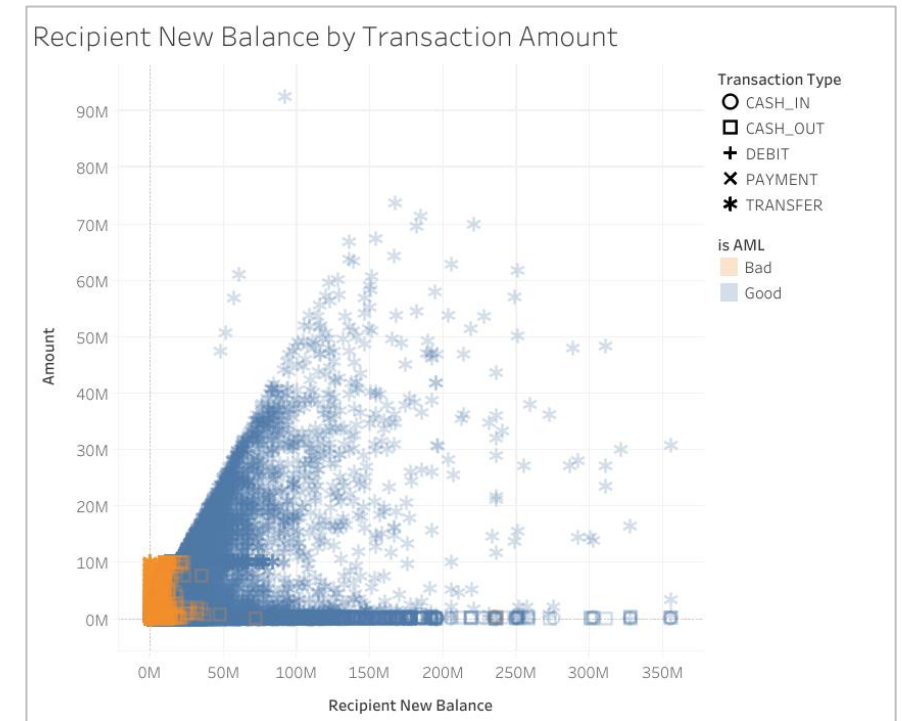
Customer New Balance is zero
for Money Laundering
Transactions



Recipient Balance



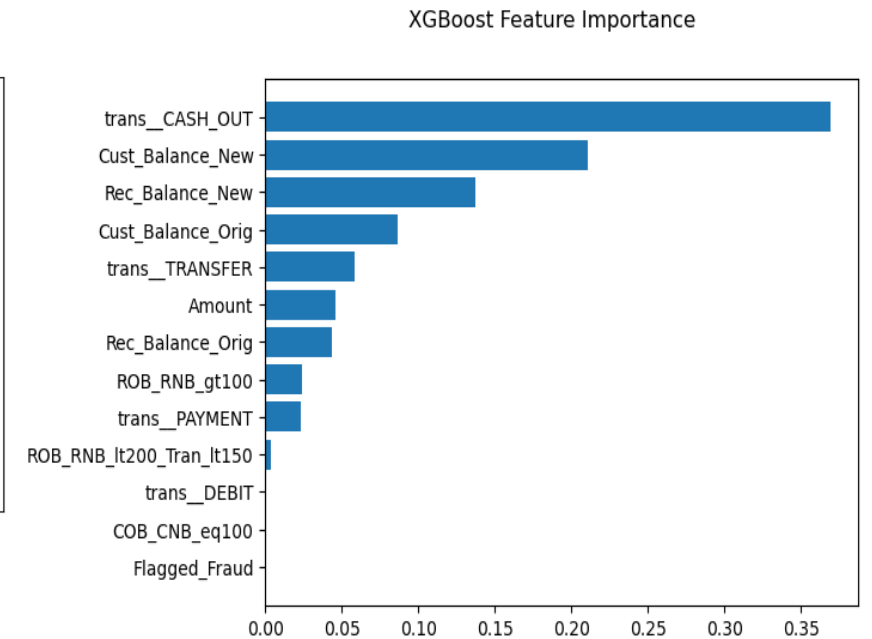
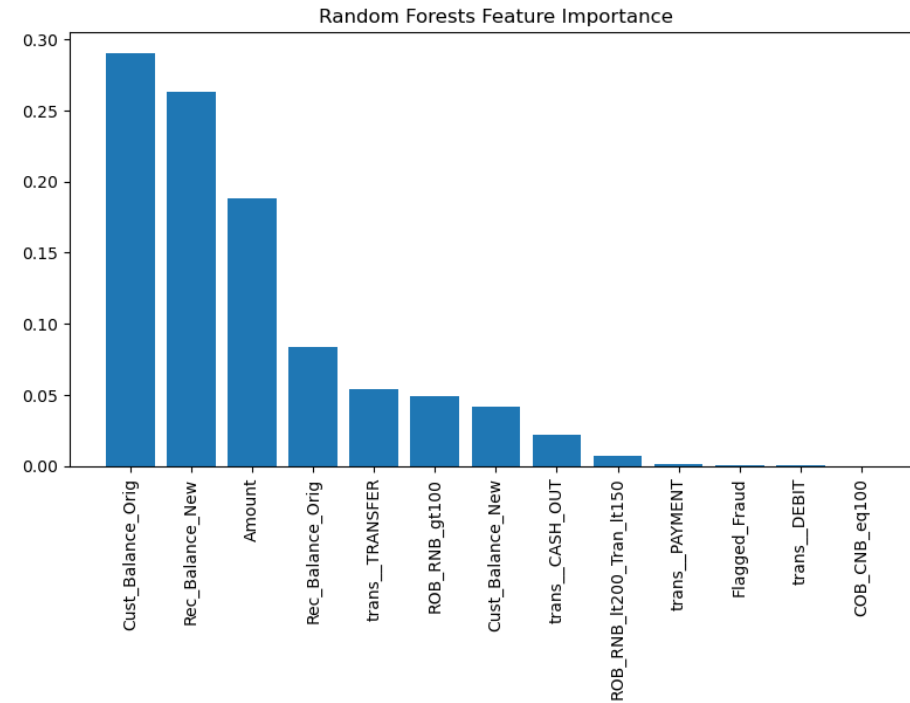
Recipient Original Balance
and New Balance
Unchanged for Money
Laundering Transactions



Feature Importance

3 of 5 Top Features

- Customer Original Balance
- Recipient New Balance
- Transfer Transactions



Modeling



LOGISTIC
REGRESSION



RANDOM
FORESTS



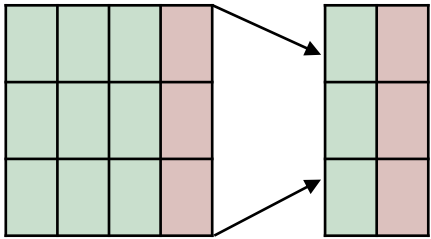
SUPPORT VECTOR
MACHINES



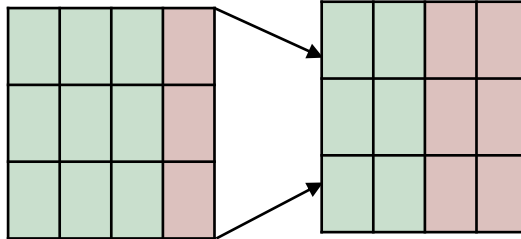
NAÏVE BAYES
CLASSIFIER

Class Imbalance

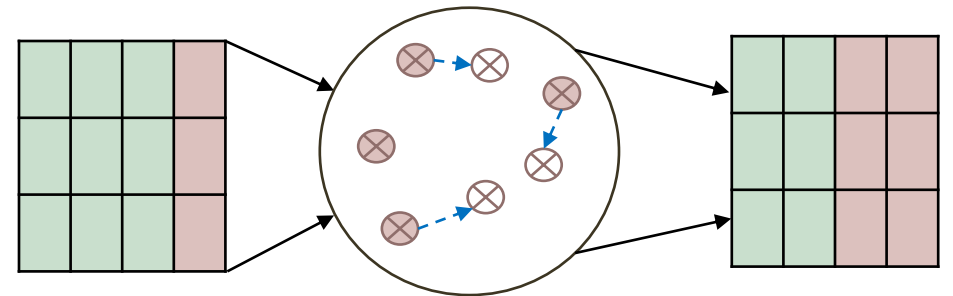
Undersampling



Oversampling



Synthetic **M**inority **O**versampling **T**echnique (SMOTE)



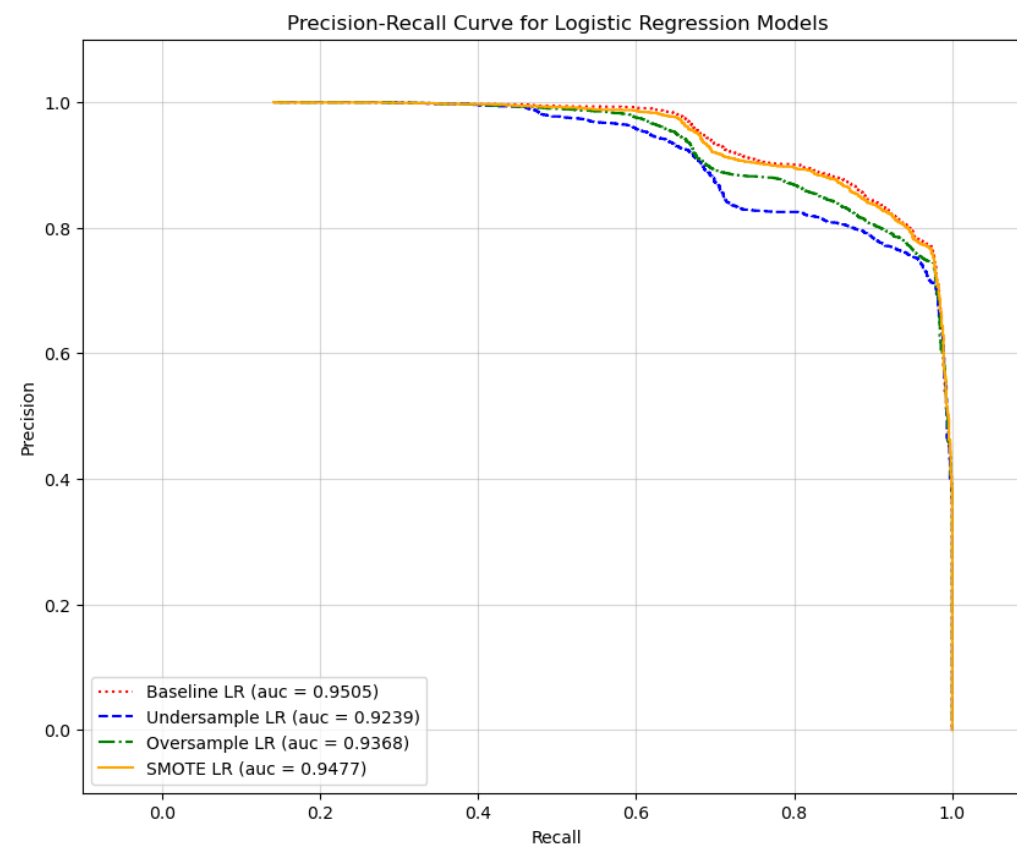
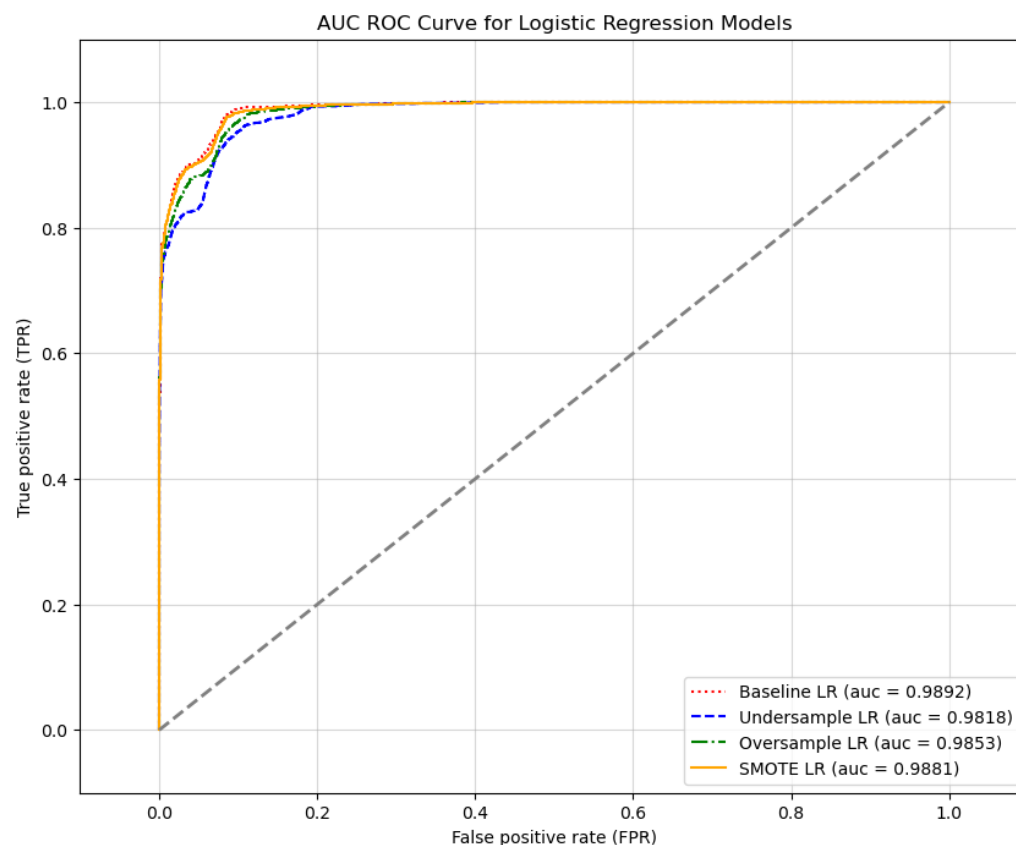
Logistic Regression

AUC ROC Curve:

- Baseline - 98.92%
- SMOTE - 98.81%

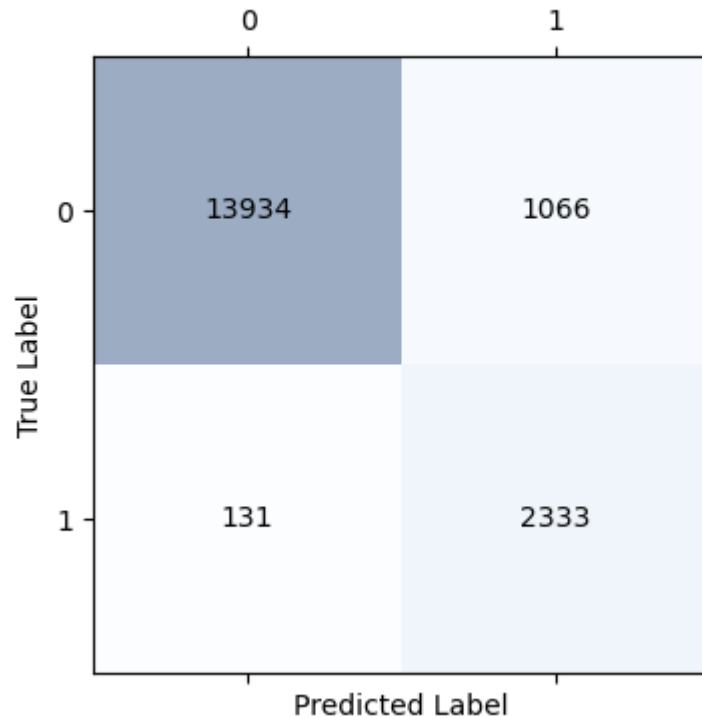
Precision-Recall Curve:

- Baseline - 95.05%
- SMOTE - 94.77%

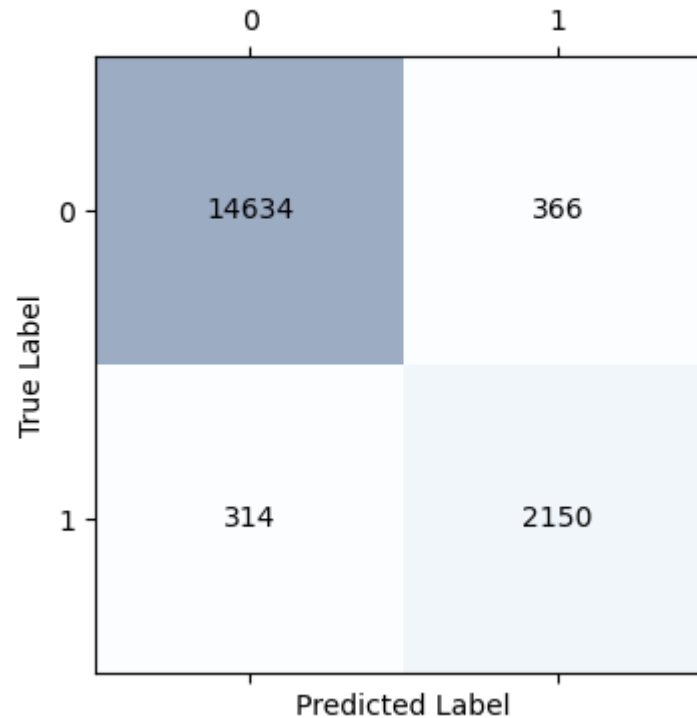


Logistic Regression

Logistic Regression Baseline Confusion Matrix



Logistic Regression SMOTE Confusion Matrix



Precision:

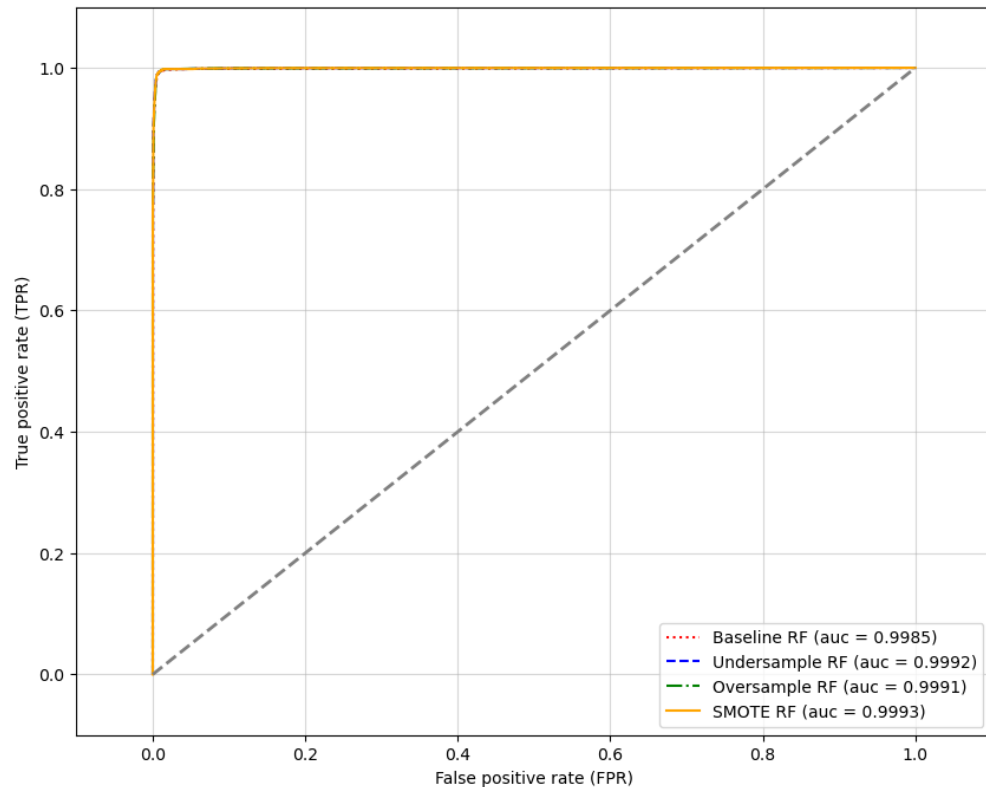
- Baseline - 68.64%
- SMOTE - 85.45%

Recall:

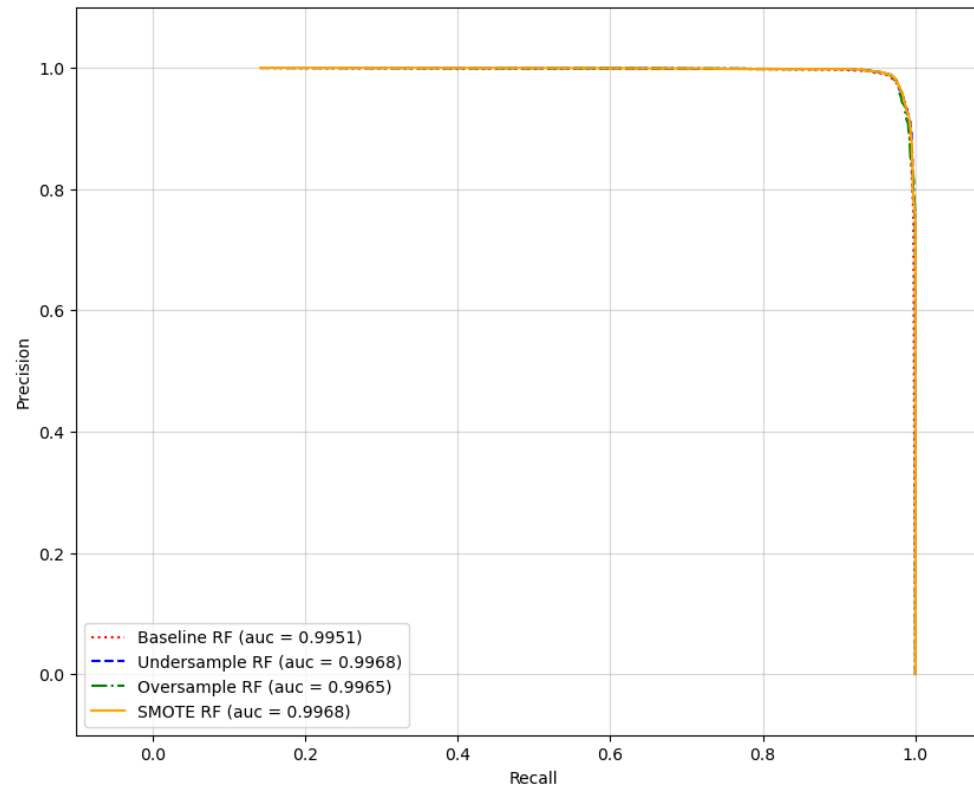
- Baseline - 94.68%
- SMOTE - 87.26%

Random Forests

AUC ROC Curve for Random Forests Models



Precision-Recall Curve for Random Forests Models



AUC ROC Curve:

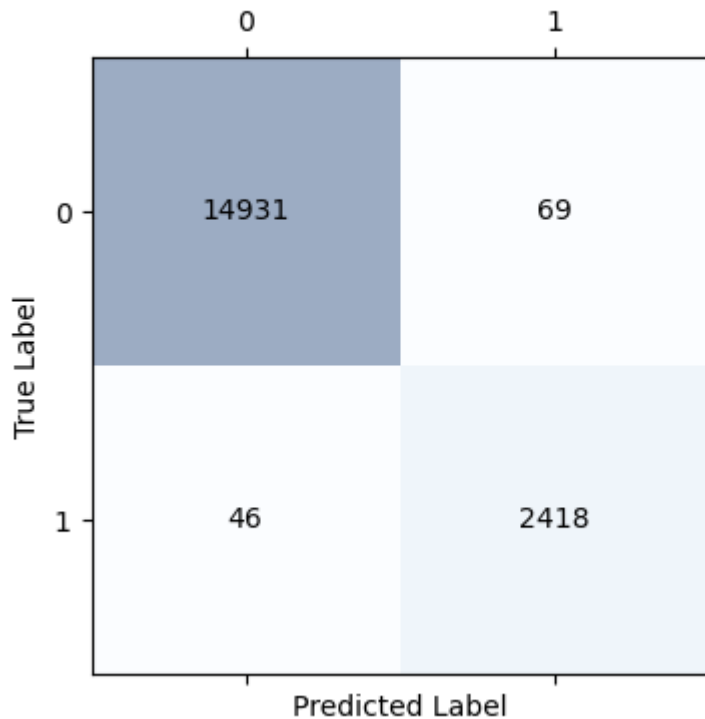
- Undersampled - 99.92%
- SMOTE - 99.93%

Precision-Recall Curve:

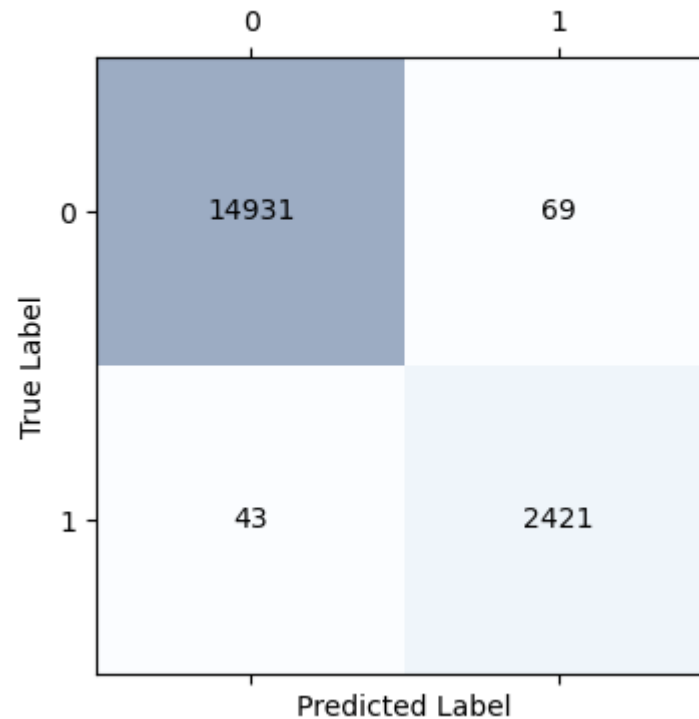
- Undersampled - 99.68%
- SMOTE - 99.68%

Random Forests

Random Forests Undersampled Confusion Matrix



Random Forests SMOTE Confusion Matrix



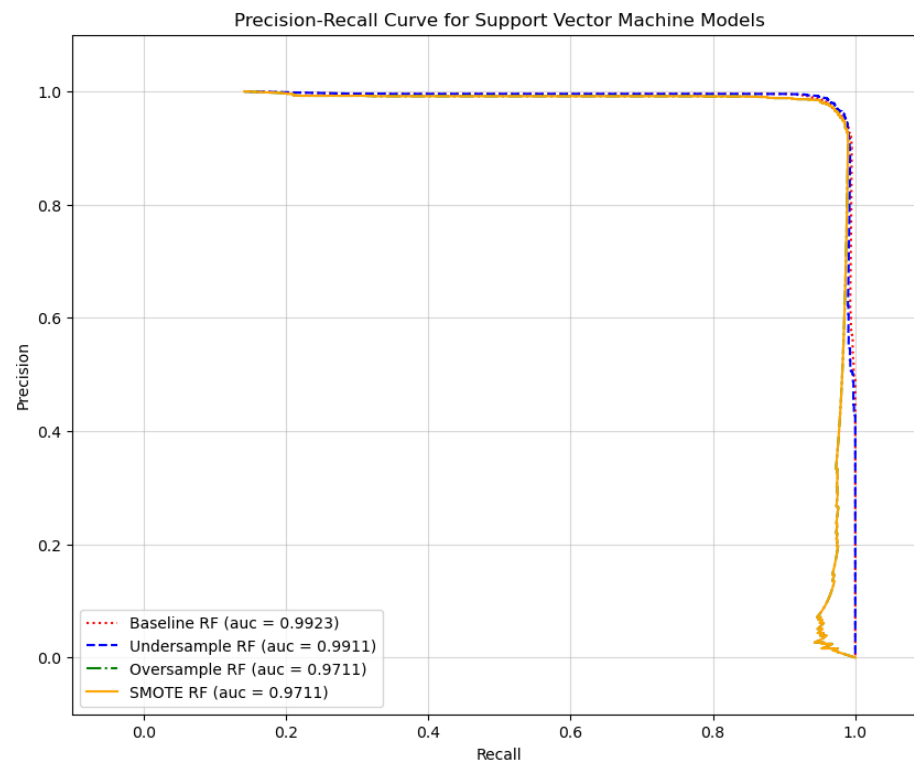
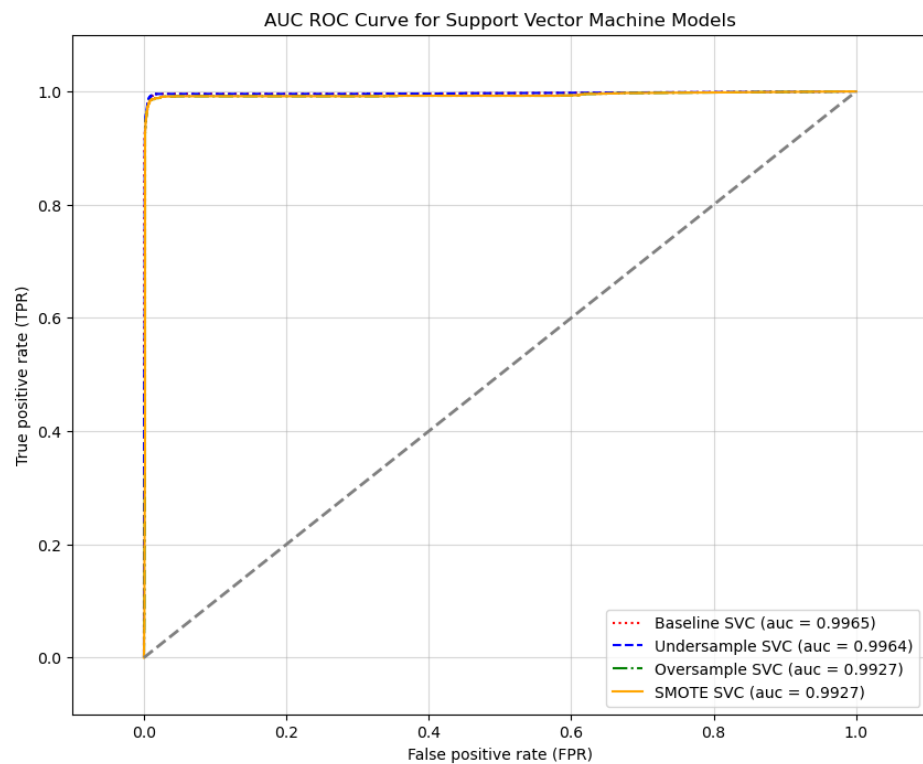
Precision:

- Undersampled - 97.23%
- SMOTE - 97.23%

Recall:

- Undersampled - 98.13%
- SMOTE - 98.25%

Support Vector Machines



AUC ROC Curve:

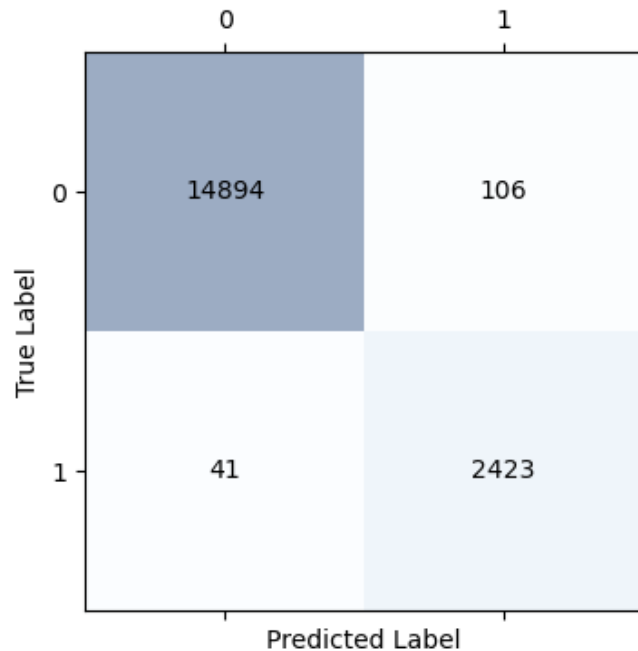
- Baseline - 99.65%
- Undersampled - 99.64%

Precision-Recall Curve:

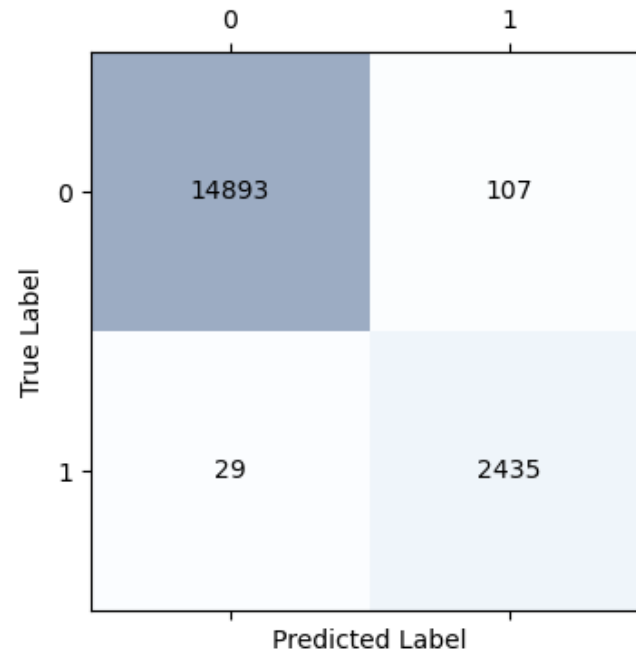
- Baseline - 99.23%
- Undersampled - 99.11%

Support Vector Machines

Support Vector Machines Baseline Confusion Matrix



Support Vector Machines Undersampled Confusion Matrix



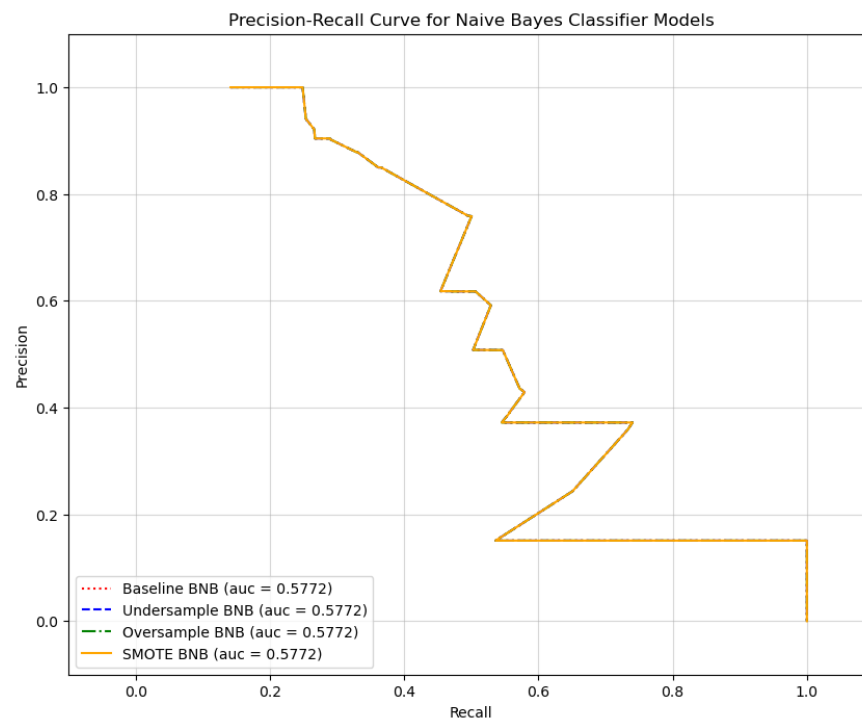
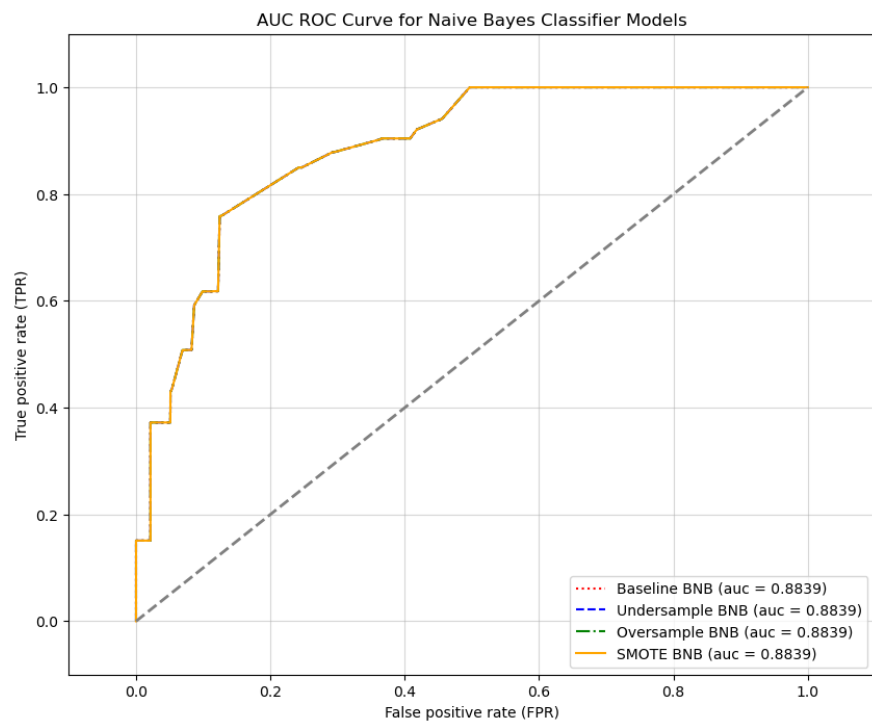
Precision:

- Baseline - 95.81%
- Undersampled - 95.79%

Recall:

- Baseline - 98.34%
- Undersampled - 98.82%

Naïve Bayes Classifier



AUC ROC Curve:

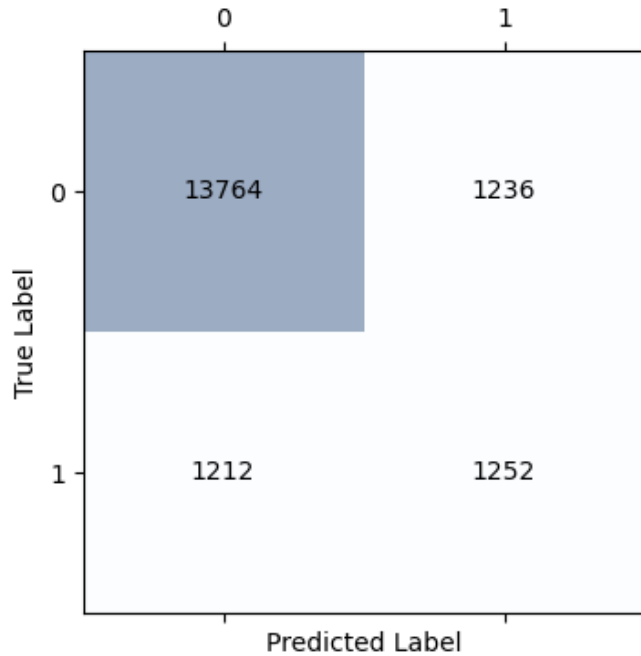
- Baseline - 88.39%
- Oversampled - 88.39%

Precision-Recall Curve:

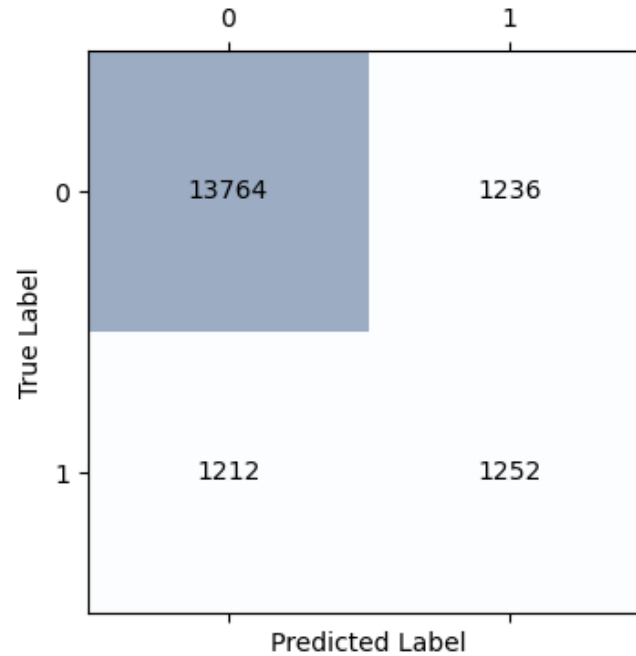
- Baseline - 57.72%
- Oversampled - 57.72%

Naïve Bayes Classifier

Naive Bayes Classifier Baseline Confusion Matrix



Naive Bayes Classifier Oversampled Confusion Matrix



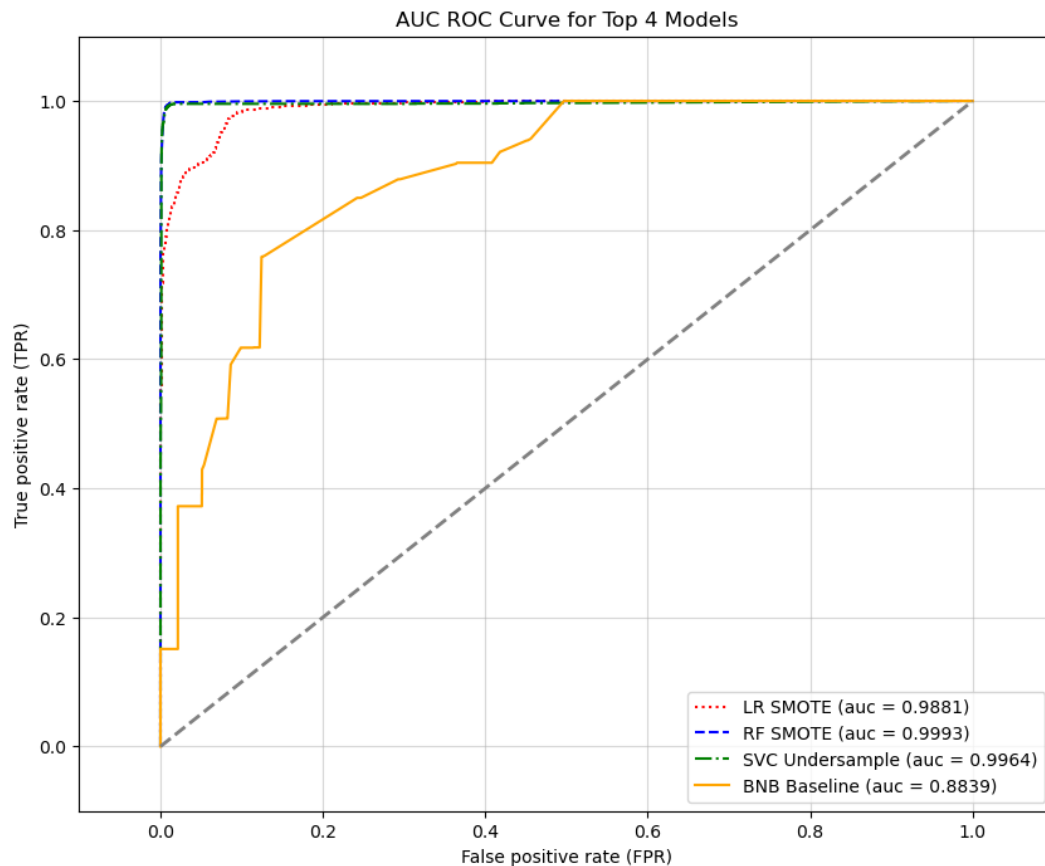
Precision:

- Baseline - 50.32%
- Oversampled - 50.32%

Recall:

- Baseline - 50.81%
- Oversampled - 50.81%

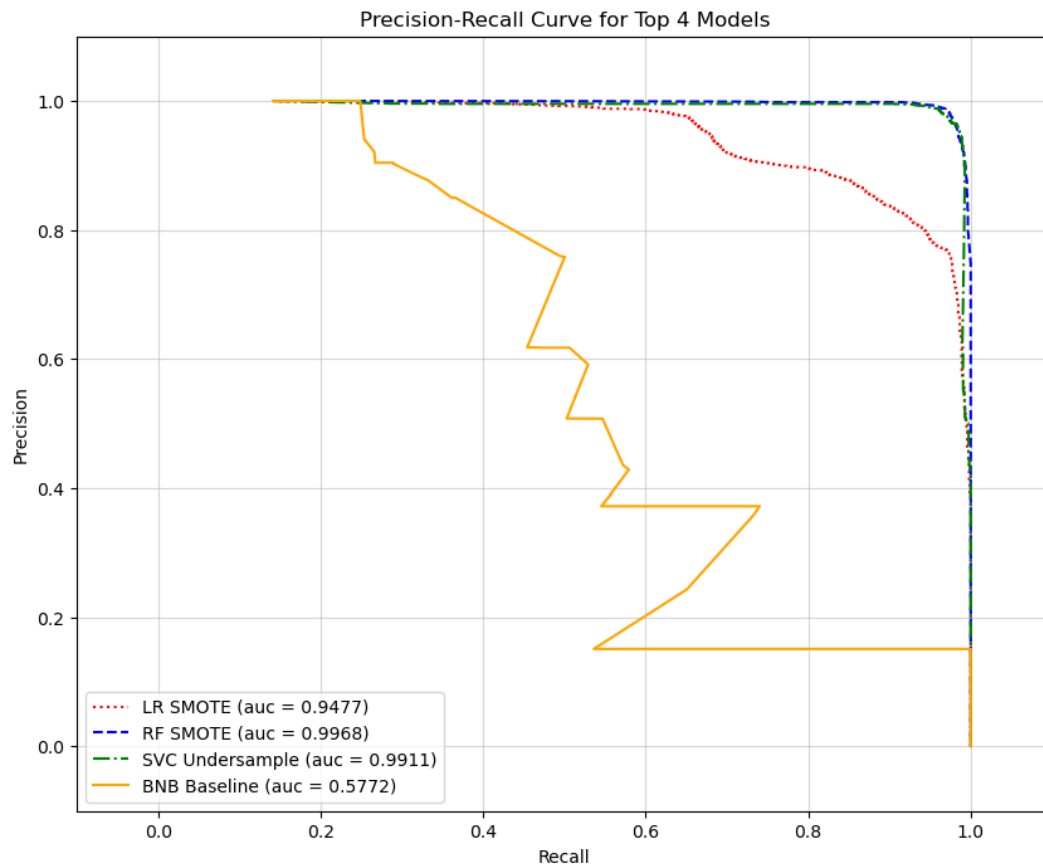
Recommendation



AUC ROC Curve:

- Logistic Regression SMOTE - 98.81%
- Random Forests SMOTE - 99.93%
- Support Vector Machines Undersampled - 99.64%
- Naïve Bayes Classifier Baseline - 88.39%

Recommendation



Precision-Recall Curve:

- Logistic Regression SMOTE - 94.47%
- Random Forests SMOTE - 99.68%
- Support Vector Machines Undersampled - 99.11%
- Naïve Bayes Classifier Baseline - 57.72%

Recommendation

	Logistic Regression SMOTE	Random Forests SMOTE	SVM Undersampled	Naïve Bayes Classifier Baseline
Test Accuracy	0.9611	0.9936	0.9922	0.8598
Test Precision	0.8545	0.9723	0.9579	0.5032
Test Recall	0.8726	0.9825	0.9882	0.5081
AUC ROC Curve	0.9881	0.9993	0.9964	0.8839
Precision-Recall Curve	0.9447	0.9968	0.9911	0.5772
Run Time	1min 41s	1min 15s	18min 5s	329 ms

Recommendation

Business Rules

- Review Flagged Fraud Transactions for Money Laundering
- Review Large Cash Out / Transfer Amounts – \$10 million
- Recipient's Original and New Balance Unchanged and Transfer / Cash Out Amounts - \$10 million

Q & A

