A Supervised Machine Learning Approach for Bank Fraud Detection

University of Chicago MS in Applied Data Science
ADSP 31008 Linear & Non-Linear Models – Utku Pamuksuz, Professor
Presented 3/9/2024

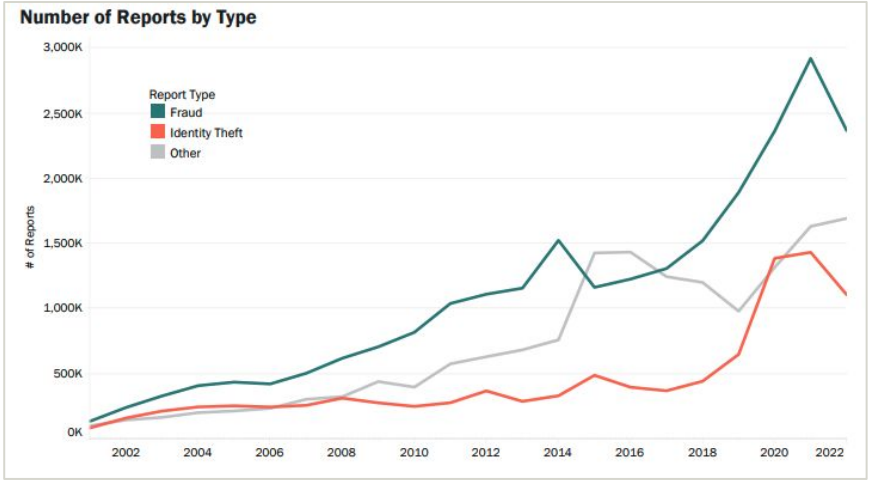
Prashant Kulkarni
Steve Veldman
Dharti Seagraves

Agenda

- Bank Fraud Detection Introduction
- Data Overview
- Project Design
- Model Implementation
- Explainable AI (XAI) Methods and Causal Inference Techniques
- Results and Discussion
- Q&A

Introduction

- Since the early 2000s there has been a drastic increase in the reported types of fraud (including bank fraud, part of ‘Identity Theft’) as reported by the Federal Trade Commission
- Creating a tool to better detect these types of fraud would benefit the customer and the bank
- Goal of this project was to evaluate multiple Linear/Non-Linear Model(s) to increase the fraud prediction rate while decreasing the number of authentic accounts flagged as fraud
- Dataset: [Kaggle Bank Fraud Dataset Suite](),
  - “Base.csv” is 1 of 6 datasets, and is most representative of the original data



*https://www.ftc.gov/system/files/ftc_gov/pdf/CSN-Data-Book-2022.pdf

Business Goals

Use case: Opening a new checking account

- Make it easy for customer to use E-Banking, to open a new account
- Reduce fraud in account opening
  - Reduce cost/liability associated with fraud
  - Reduce cost of fraud detection through automated/AI tools.

- If a fraud is detected, customer is not rejected directly:
  - Refer to human representative
  - Have them visit the nearest branch in person
- Few advantages here:
  - Fraudster will not do so
  - Genuine customer will, bank will have opportunity up selling
- Minimize the occurrences of fraudulent account classified as authentic

Business Implications

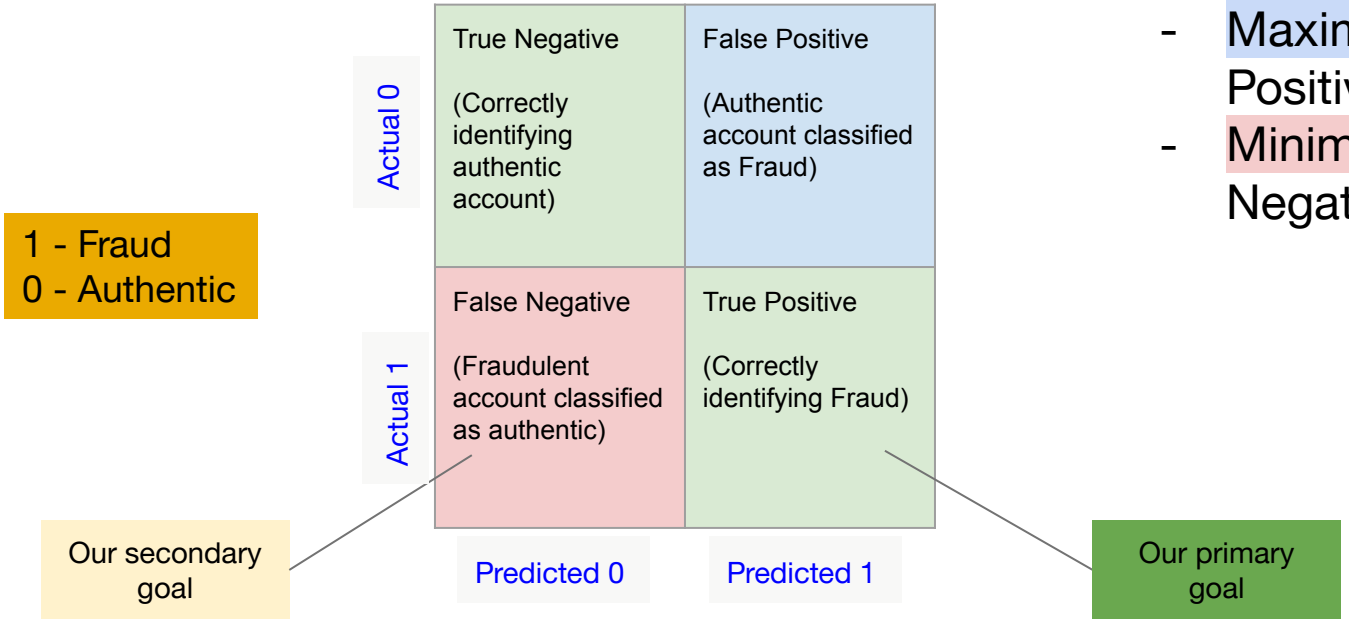
- Most of the related literature on this topic addresses fraud at the transactional level (such as individual credit card charges). In this case the cost of a single missed fraud case is likely lower than the cost of losing a customer by irritating them by declining a genuine transaction.

- In this use case FPR is still important, but the cost of failing to detect an individual instance of fraud represents a higher cost than failing to detect a single fraudulent credit card transaction.

- In both use cases, the cost of a False Positive can be mitigated by adding a layer of verification (text alerts/approvals in the case of credit card transaction, or human verification/interview in the case of bank account applications).

- Total fraud predictions (TP+FP) should represent a small portion of actual bank account applications. Focusing human intervention on these applications will allow our model to reduce a bank's costs while also increasing fraud detection.

Defining Success



- Maximize 'True Positives'
- Minimize 'False Negatives'

Defining Success

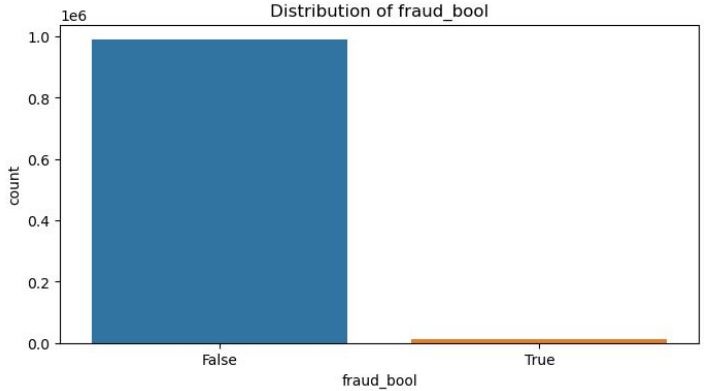
- In practice, fraud only represents ~1% of bank account applications
- Accuracy is not a meaningful metric - a model that classifies everything as authentic will have 99% but will let 100% of fraud cases through
- Recall (True Positive Rate) represents the percentage of fraud cases “caught” by our model.
- False Positive Rate represents percentage of authentic applications incorrectly classified as fraud
- AUC is a meaningful metric for how well the model balances these two goals

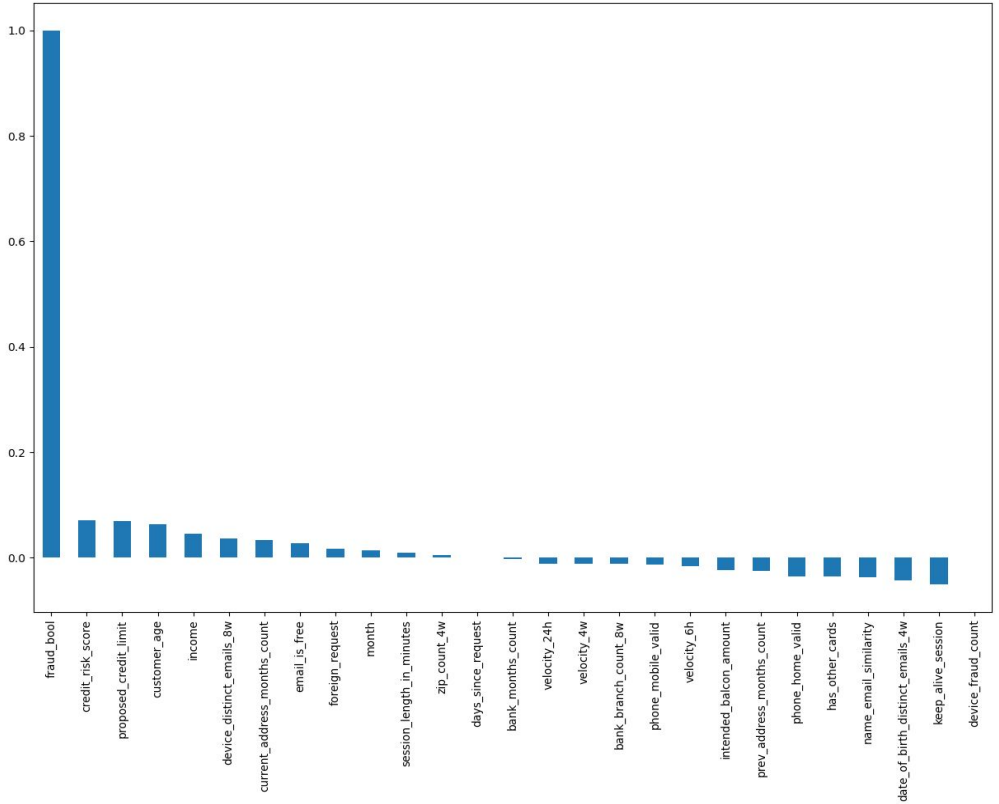
Data Overview

- Base.csv dataset: 1,000,000 records
- Synthetic dataset
  - Realistic - based on present-day real-world data
  - Biased - had a distinct type of controlled bias
  - Imbalance - low balance of cases identified as fraud
  - Dynamic - temporal data and observed distribution shifts
  - Privacy preserving - authors applied differential privacy techniques, feature encoding, and trained a generative model (CTGAN)
  - Turning the Tables: Biased, Imbalanced, Dynamic Tabular Datasets for ML Evaluation (NeurIPS 2022)
- Target/response variable: ‘fraud_bool’
- Predictor variables: 31

Exploratory Data Analysis

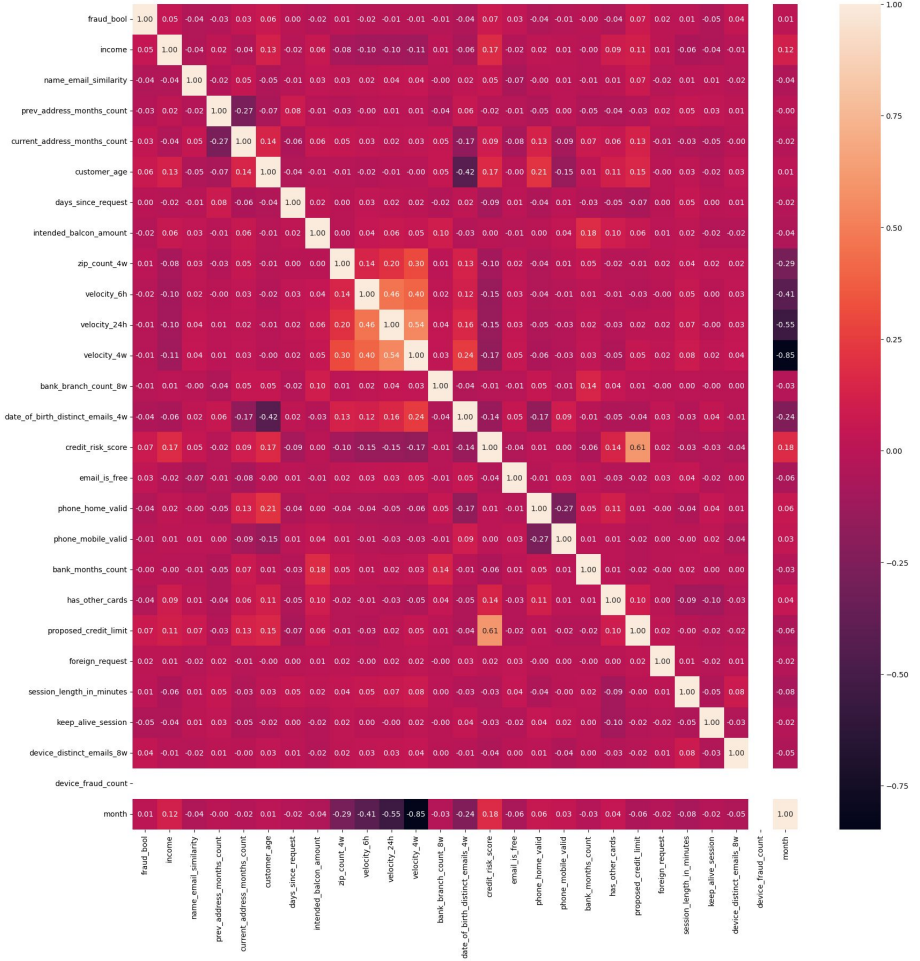
- Variable Correlation Plot (*right*)
- Fraud Distribution (*below*)





EDA (part 2)

- Highly correlated
  - Velocity_6h & Velocity_24h & Velocity_4w
  - Proposed_credit_limit & credit_risk_score
- Negatively correlated
  - Date_of_birth_distinct_emails_4w & customer_age
  - Current_address_months_count & pre_address_months_count
- No multicollinearity



Balancing Techniques:

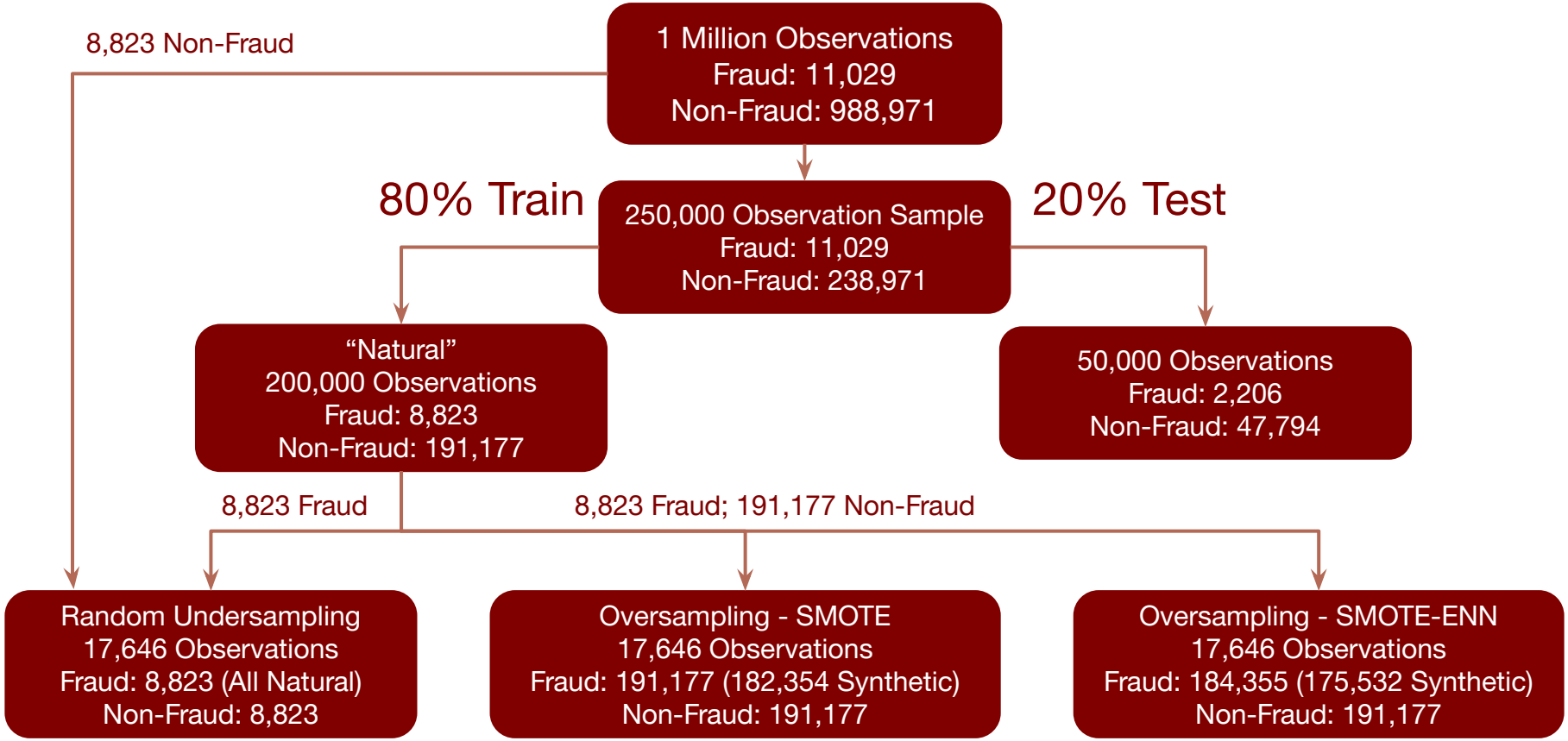
For computational efficiency, we created a sample from original dataset:

- 250,000 observation sample drawn from original dataset's 1 million observations:
  - Kept all 11,029 fraud observations
  - Remaining observations randomly sampled from authentic observations in parent dataset
  - This technically represents a "gentle" undersampling of the original dataset, but is very similar in characteristics to full 1 million observation dataset (minority class went from 1.1% to 4.4%).

After 80/20 train/test split, we moved forward with 4 sets of training data to explore with our models:

- **"Natural" Training Data:** Simple train/test split of sample data with no additional balancing techniques
- **SMOTE:** Oversampling of fraud observations using Synthetic Minority Oversampling TEchnique
- **SMOTE-ENN:** Hybrid over/undersampling technique utilizing SMOTE and Edited Nearest Neighbors
- **Random Undersampling:** 8,823 fraud observations from training data matched with random sample of 8,823 authentic observations (out of 988,971) from parent dataset.

Balancing Process



Balancing Techniques

We tried various data balancing techniques as below

	Train (Non-Fraud)	Train (Fraud)		Test (Non-Fraud)	Test (Fraud)
Natural*	191,177	8,823		47,794	2,206
Random Undersampling (RUS)	8,823	8,823		47,794	2,206
SMOTE	191,177	191,177		47,794	2,206
SMOTE_ENN	191,177	184,355		47,794	2,206

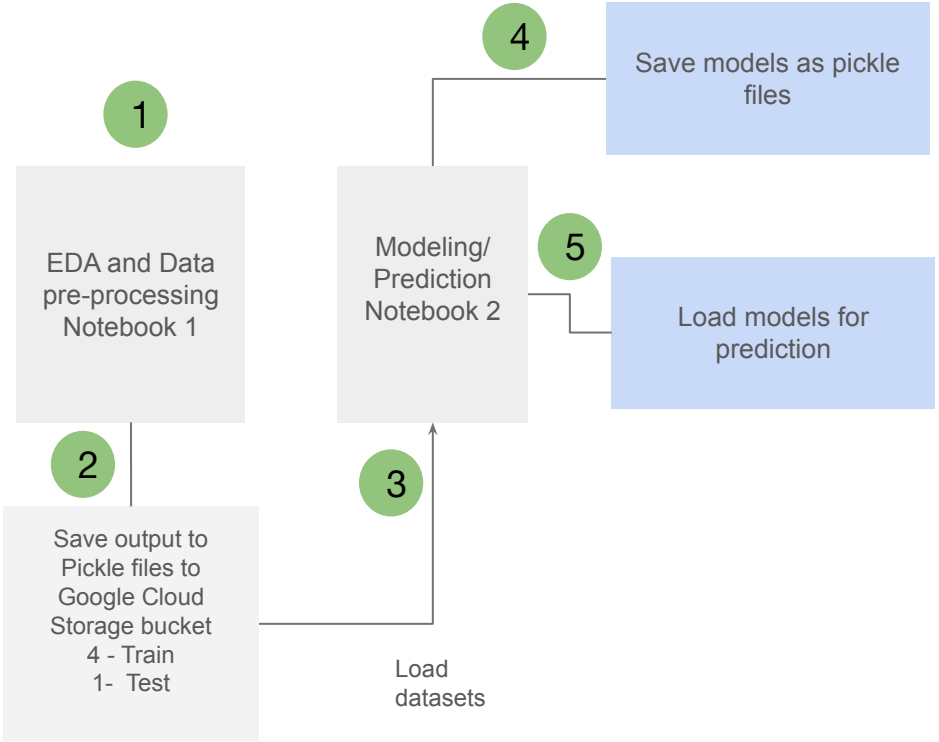
*Represents reduced sample from 1M observations for the non-fraud train data

Model selection

Nature of the Problem: The goal of this analysis is to identify fraud - a binary classification (two classes)
Size and Type of Data: Considering the volume, variety, and velocity of the data. We have a high volume data (1M rows/32 columns)
Data Quality: Highly imbalanced data (98.1/1.9%)

Algorithms considered	Algorithms selected
• Logistic regression • Decision trees/Random Forests • Gradient boosting • K-Nearest Neighbors (KNN) • Support vector machines • Other Ensemble models	• Logistic regression • Random Forests • XGBoost • Bagging • Voting Ensemble Model

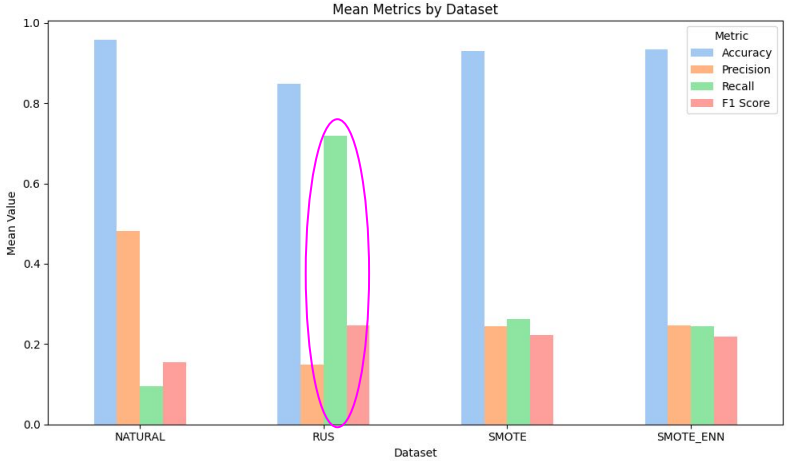
Project Design

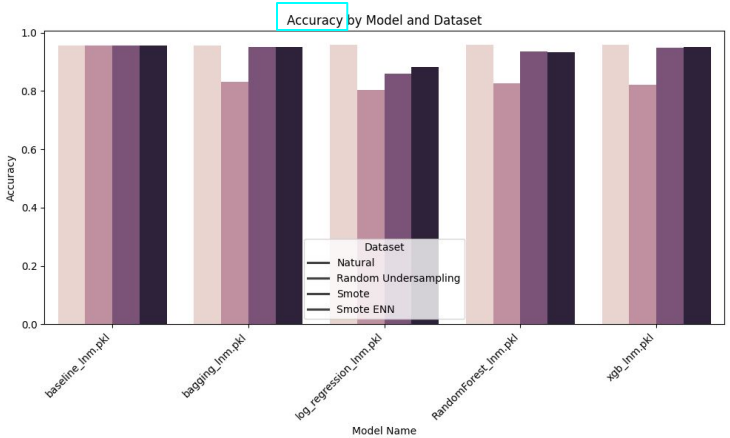


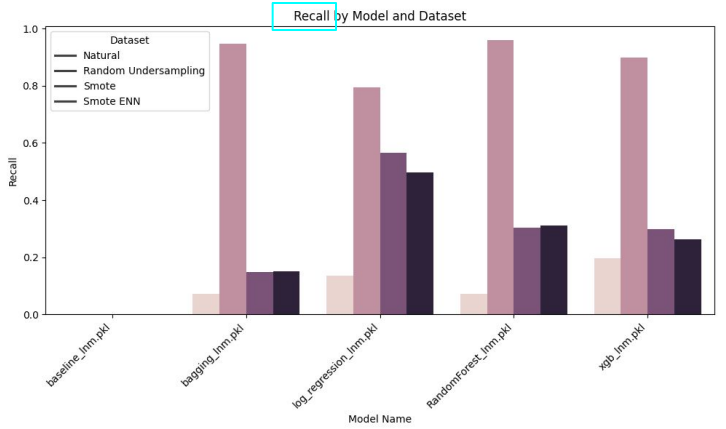
1	Exploratory data analysis, data balancing and dataset generation
2	Saved dataset to Google Cloud Storage (GCS) bucket
3	In notebook 2, load datasets from GCS bucket
4	Save models to pickle file
5	Run predictions, performance evaluation and visuals

Dataset Performance

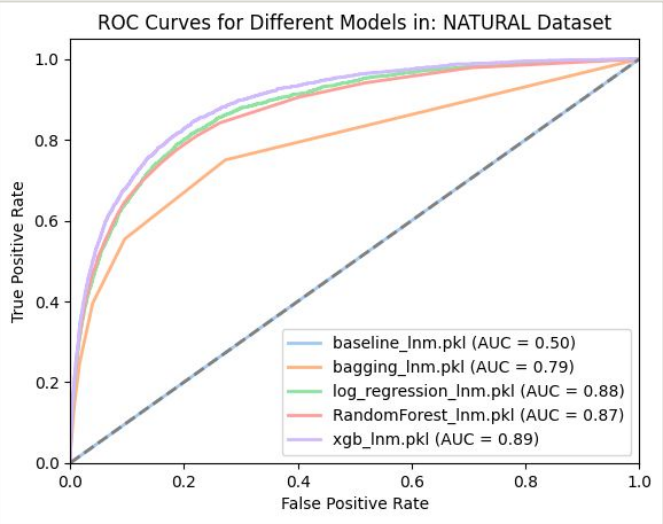
- The average accuracy for all the models was greater than 80%
- Random undersampling performed the best in our use-case, prioritizing recall to catch as many fraud cases as possible
- While the 'Natural' dataset had the best accuracy the low recall indicates a flaw for our business case
- SMOTE and SMOTE_ENN performed very similarly for all the metrics





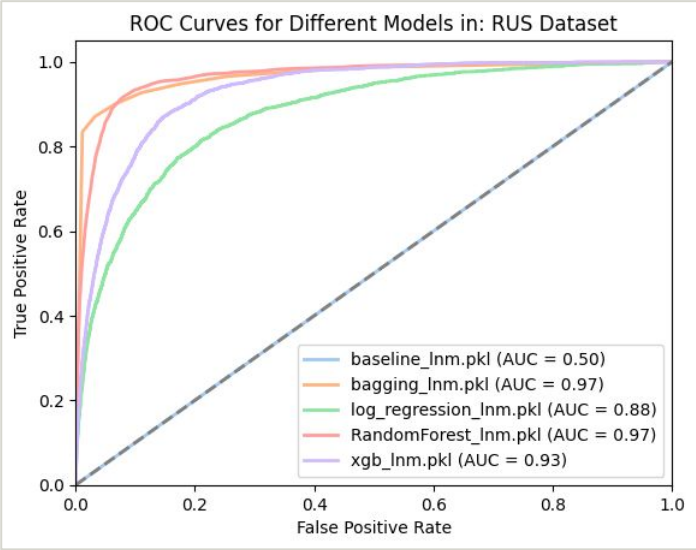


Natural dataset



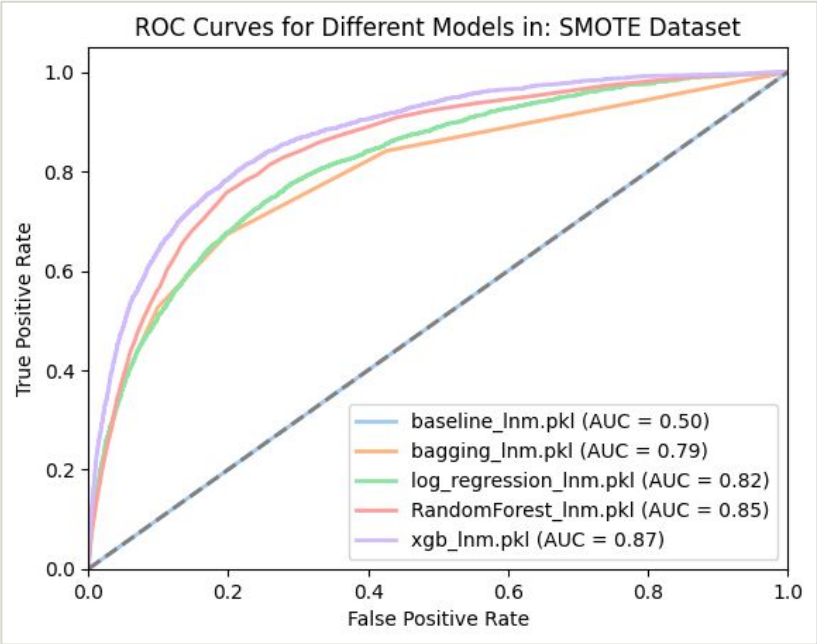
	Accuracy	Precision	Recall	F1 Score
baseline	0.96	0.00	0.00	0.00
bagging	0.95	0.27	0.03	0.05
log_regression	0.96	0.77	0.02	0.04
RandomForest	0.96	**1.00**	**0.00**	0.00
xgb	0.96	0.79	0.02	0.03

RUS dataset



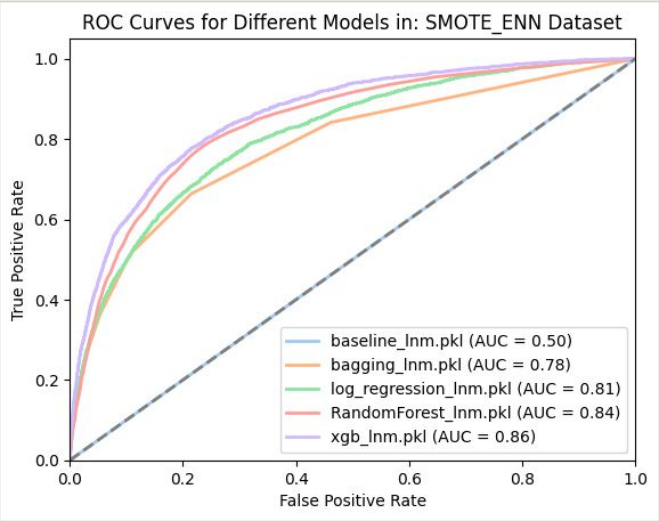
	Accuracy	Precision	Recall	F1 Score
baseline	0.96	0.00	0.00	0.00
bagging	0.83	0.20	**0.95**	0.33
log_regression	0.80	0.16	**0.79**	0.26
RandomForest	0.83	0.20	**0.96**	0.33
xgb	0.82	0.19	**0.90**	0.31

SMOTE Dataset



	Accuracy	Precision	Recall	F1 Score
baseline	0.96	0.00	0.00	0.00
bagging	0.95	0.37	0.15	0.21
log_regression	0.86	0.17	0.56	0.26
RandomForest	0.94	0.28	0.30	0.29
xgb	**0.95**	**0.40**	**0.30**	**0.34**

Smote ENN Dataset



	Accuracy	Precision	Recall	F1 Score
baseline	0.96	0.00	0.00	0.00
bagging	0.95	0.37	0.15	0.21
log_regression	0.88	0.19	0.50	0.27
RandomForest	0.93	0.28	0.31	0.29
xgb	**0.95**	**0.40**	**0.26**	**0.32**

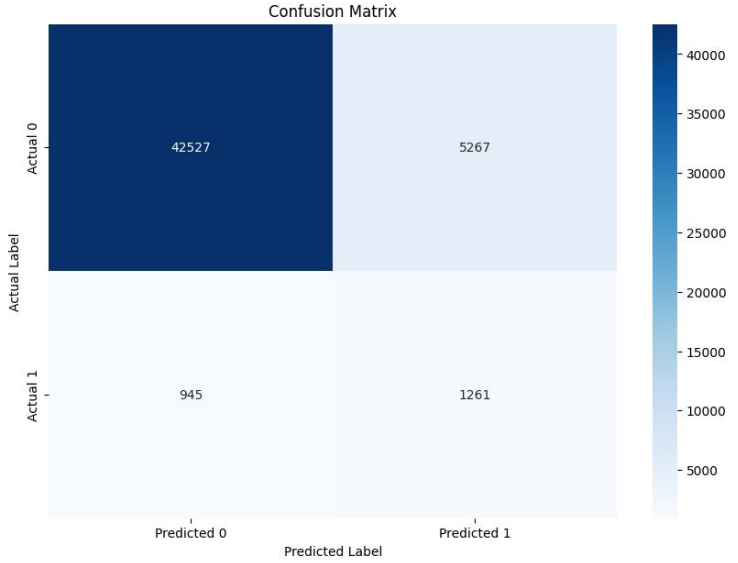
Summary

Model Name	Accuracy	Precision	Recall	F1 Score	Dataset
bagging	0.95	0.27	0.03	0.05	Natural
log_regression	0.96	0.77	0.02	0.04	Natural
RandomForest	0.96	1	0	0	Natural
xgb	0.96	0.79	0.02	0.03	Natural
bagging	**0.83**	**0.20**	**0.95**	**0.33**	**RUS**
log_regression	**0.80**	**0.16**	**0.79**	**0.26**	**RUS**
RandomForest	**0.83**	**0.20**	**0.96**	**0.33**	**RUS**
xgb	**0.82**	**0.19**	**0.90**	**0.31**	**RUS**
bagging	0.95	0.37	0.15	0.21	SMOTE
log_regression	0.86	0.17	0.56	0.26	SMOTE
RandomForest	0.94	0.28	0.30	0.29	SMOTE
xgb	0.95	0.4	0.30	0.34	SMOTE
bagging	**0.95**	**0.37**	**0.15**	**0.21**	**SMOTE ENN**
log_regression	**0.88**	**0.19**	**0.50**	**0.27**	**SMOTE ENN**
RandomForest	**0.93**	**0.28**	**0.31**	**0.29**	**SMOTE ENN**
xgb	**0.95**	**0.40**	**0.26**	**0.32**	**SMOTE ENN**

Voting Ensemble Model (RUS)

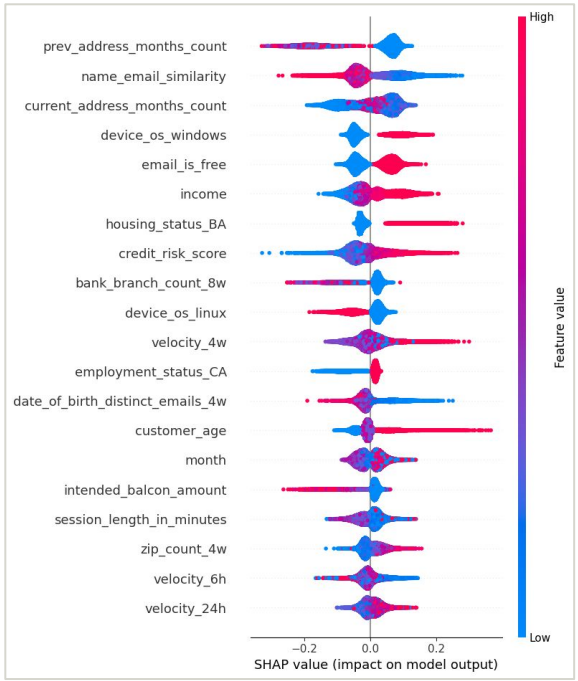
This model was riddled with a very low precision and recall, because of that Ensembling model didn't turn out to be a good choice. Also, we found ensembling is less interpretable, the output of the ensembled model was hard to predict and explain.

	Precision	Recall	F1	Support
0	0.98	0.89	0.93	47794
1	0.19	**0.57**	0.29	2206
Accuracy			0.88	50000
Macro avg	0.59	0.73	0.61	50000
Weighted avg	0.94	0.88	0.90	50000



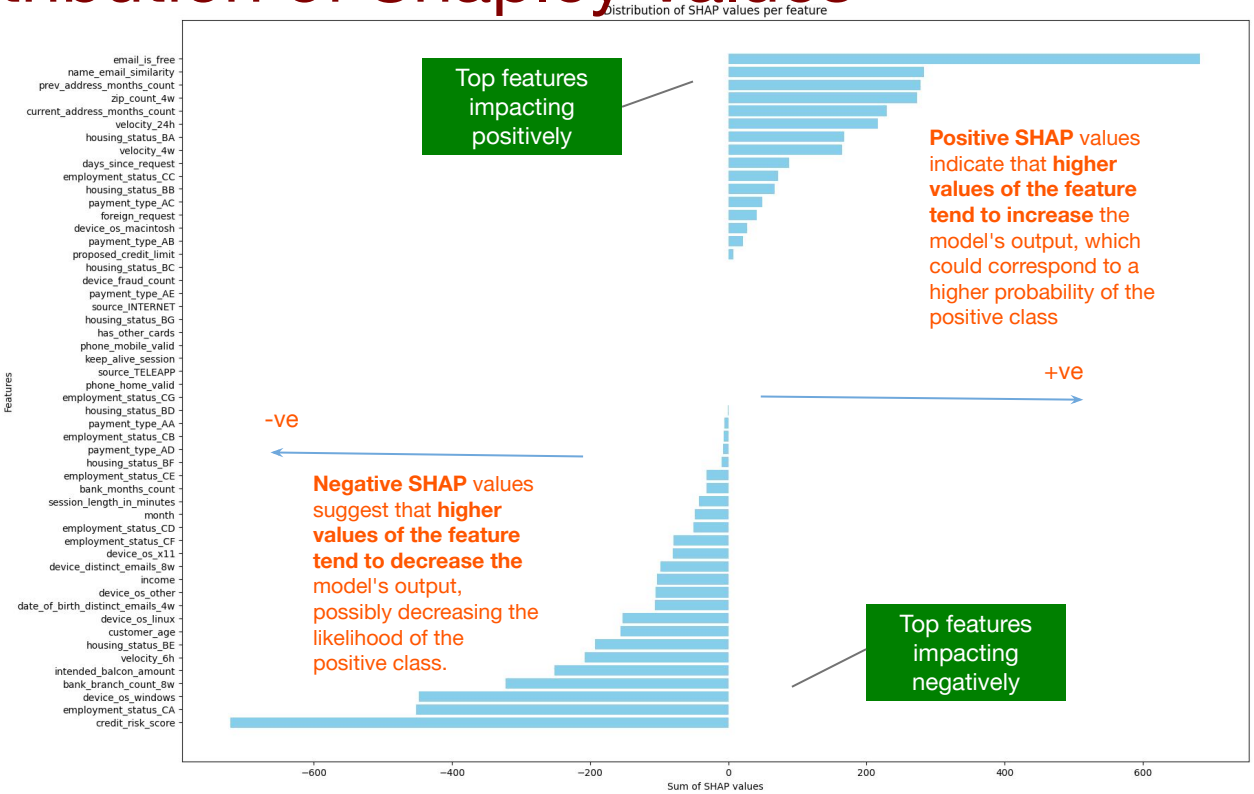
Explainable (XAI) AI Methods

This summary plot provides a visualization of the feature importances and their effects on the model's predictions. Each point on the summary plot represents a SHAP value for a feature and an instance. The position on the X-axis indicates the impact of the value on the model's output, and the color represents the value of the feature (red = high, blue = low).



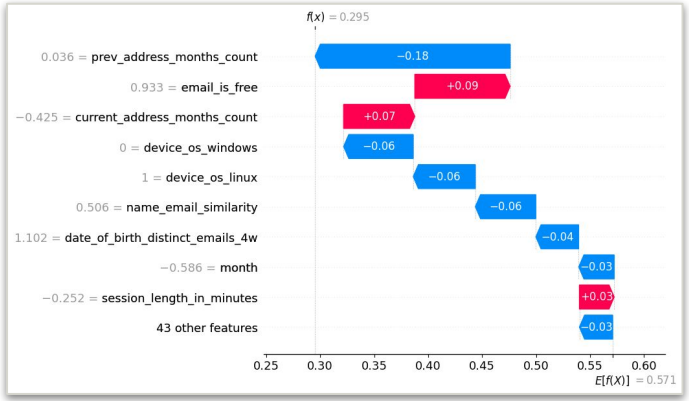
- High impact features:
  a. prev_address_months_count
  b. name_email_similarity
  c. current_address_months_count
  d. device _os_windows
  e. income
  f. housing_status_BA
- Medium impact features
  a. credit_risk_score
  b. bank_branch_count_8w
  c. device_os_linux
  d. velocity_4w
  e. employment_status_CA
- Low impact features:
  a. session_length_in_minutes
  b. zip_count_4w
  c. velocity_6h
  d. velocity_24h

Distribution of Shapley Values

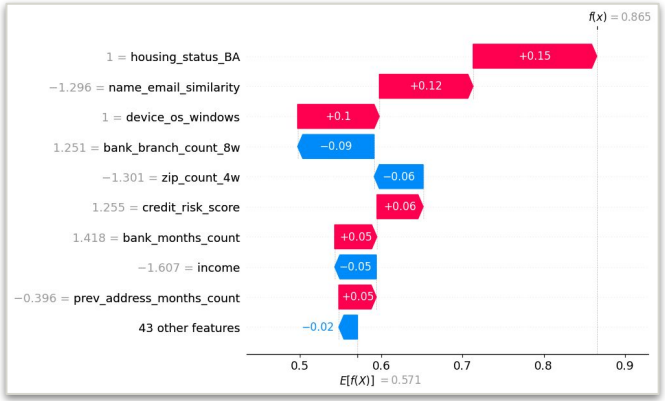


Waterfall Plots - XGBoost with RUS

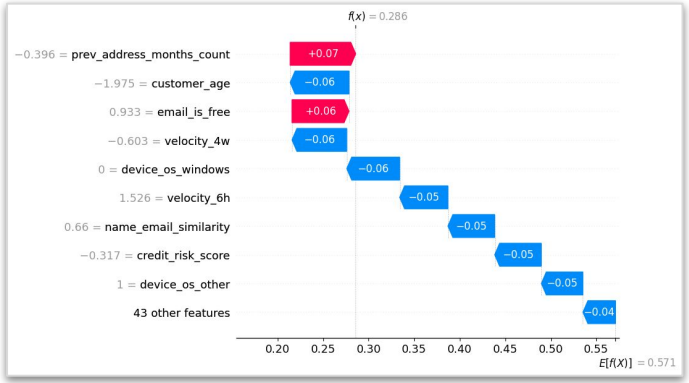
True Negative



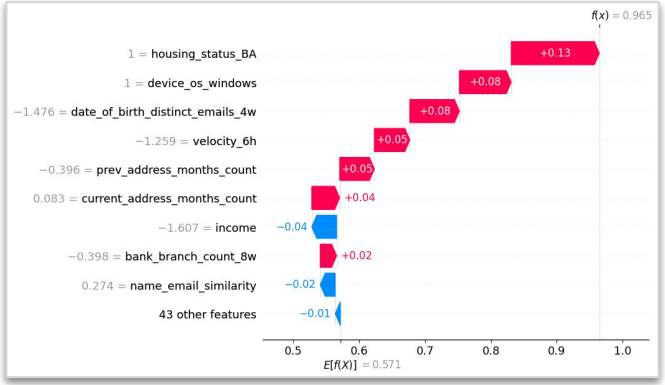
False Positive



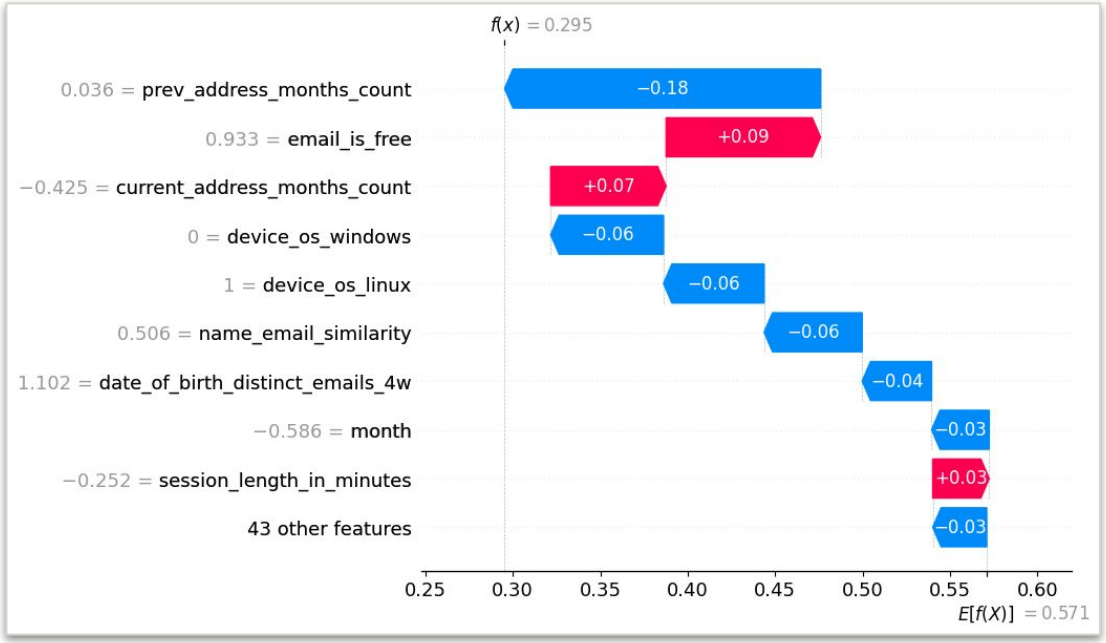
False Negative



True Positive

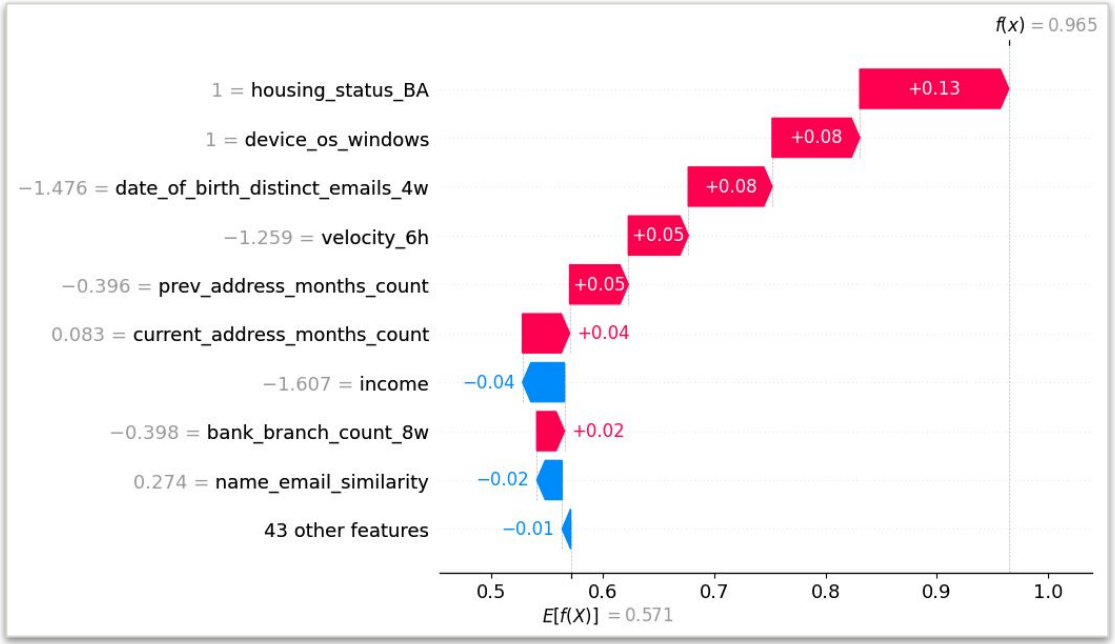


True Negative



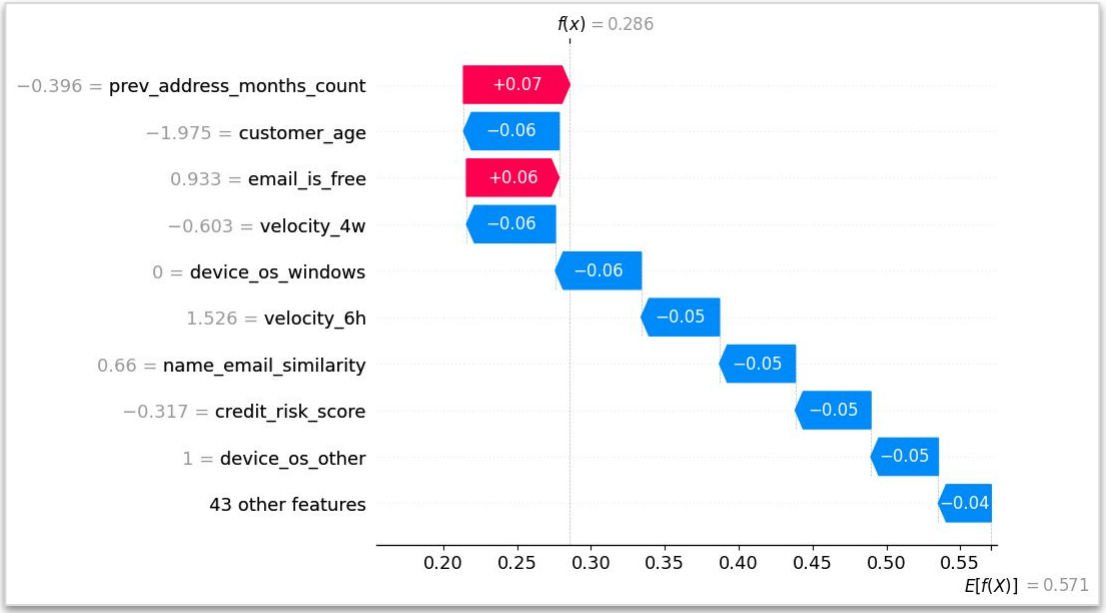

```
Actual test record:
income                                0.60
name_email_similarity                 0.64
prev_address_months_count               18
current_address_months_count            50
customer_age                            30
days_since_request                    0.03
intended_balcon_amount               10.69
payment_type                            AA
zip_count_4w                          1236
velocity_6h                       5,186.26
velocity_24h                      5,734.86
velocity_4w                       5,086.44
bank_branch_count_8w                     2
date_of_birth_distinct_emails_4w        15
employment_status                       CA
credit_risk_score                      188
email_is_free                            1
housing_status                          BC
phone_home_valid                         0
phone_mobile_valid                       1
bank_months_count                        6
has_other_cards                          0
proposed_credit_limit                500.00
foreign_request                          0
source                            INTERNET
session_length_in_minutes             5.53
device_os                            linux
keep_alive_session                       1
device_distinct_emails_8w                1
device_fraud_count                       0
month                                    2
Name: 237036, dtype: object
```

True Positive



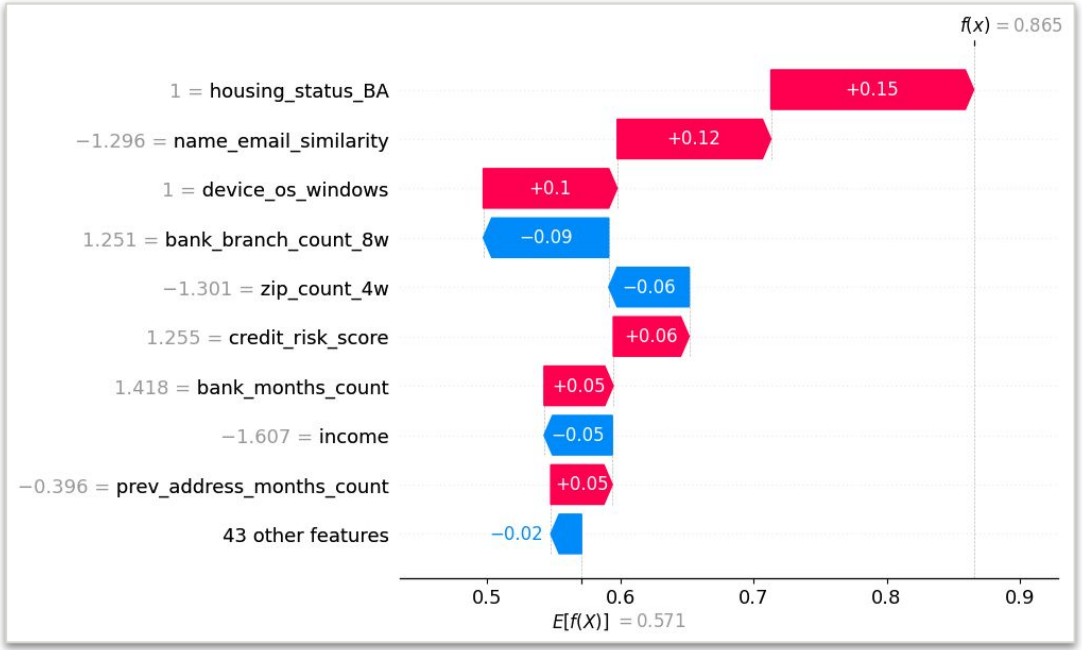

```
Actual test record:
income                                0.1
name_email_similarity                 0.569426
prev_address_months_count              -1
current_address_months_count           95
customer_age                           20
days_since_request                     0.000301
intended_balcon_amount                 -1.0
payment_type                           AC
zip_count_4w                           921
velocity_6h                            1857.472419
velocity_24h                           4596.404951
velocity_4w                            5027.706722
bank_branch_count_8w                   0
date_of_birth_distinct_emails_4w       2
employment_status                      CA
credit_risk_score                      134
email_is_free                          0
housing_status                         BA
phone_home_valid                       1
phone_mobile_valid                     1
bank_months_count                      -1
has_other_cards                        0
proposed_credit_limit                  200.0
foreign_request                        0
source                                 INTERNET
session_length_in_minutes              2.266416
device_os                              windows
keep_alive_session                     0
device_distinct_emails_8w              1
device_fraud_count                     0
month                                  4
Name: 169850, dtype: object
```

False Negative




```
Actual test record:
income                                0.4
name_email_similarity                 0.681183
prev_address_months_count              -1
current_address_months_count          262
customer_age                          10
days_since_request                    0.000032
intended_balcon_amount                -1.0
payment_type                          AB
zip_count_4w                          1225
velocity_6h                           10258.152865
velocity_24h                          4500.840135
velocity_4w                           4296.725517
bank_branch_count_8w                  15
date_of_birth_distinct_emails_4w      10
employment_status                     CA
credit_risk_score                     110
email_is_free                         1
housing_status                        BB
phone_home_valid                      0
phone_mobile_valid                    1
bank_months_count                     5
has_other_cards                       0
proposed_credit_limit                 200.0
foreign_request                       0
source                                INTERNET
session_length_in_minutes             4.054755
device_os                             other
keep_alive_session                    1
device_distinct_emails_8w             1
device_fraud_count                    0
month                                 5
Name: 166068, dtype: object
```

False Positive




```
Actual test record: income
0.1
name_email_similarity                  0.114247
prev_address_months_count                    -1
current_address_months_count                176
customer_age                                 30
days_since_request                     0.023155
intended_balcon_amount                     -1.0
payment_type                                 AB
zip_count_4w                                268
velocity_6h                        5442.769242
velocity_24h                       3874.614618
velocity_4w                        4094.960443
bank_branch_count_8w                        753
date_of_birth_distinct_emails_4w             18
employment_status                            CA
credit_risk_score                           221
email_is_free                                 0
housing_status                               BA
phone_home_valid                              0
phone_mobile_valid                            1
bank_months_count                            28
has_other_cards                               0
proposed_credit_limit                    1000.0
foreign_request                               0
source                                 INTERNET
session_length_in_minutes              11.85705
device_os                               windows
keep_alive_session                            1
device_distinct_emails_8w                     1
device_fraud_count                            0
month                                         5
Name: 49185, dtype: object
```

Improvements

- Data Balancing Techniques:
  - Further tune hyperparameters on SMOTE and SMOTE-ENN algorithms
    - Improve process for encoding/decoding categorical predictors
  - Refine approach for Random Undersampling
  - More sophisticated techniques:
    - Generative Adversarial Network (GAN)
    - Hybrid technique (SMOTified-GAN)
- Fine-Tune Existing Models:
  - Feature Selection
  - Feature Engineering

Recommendation for Future Work

- Pursue improvements listed on previous slide
- Attempt to recreate the impressive results achieved with random undersampling
- Stress-Test our models on additional datasets
  - Evaluate for different types of bias (important for this use case)
- Additional models (neural networks, deep learning, etc)
  - Neural Networks/Deep Learning
  - Ensemble Undersampling Methods (https://imbalanced-learn.org/stable/references/ensemble.html)

Conclusion

- The various decision tree-based ensembles (bagging, boosting, random forest) performed better than logistic regression with all of the training data options.
- Datasets balanced with synthetic data (SMOTE and SMOTE-ENN) performed notably better than the imbalance data.
- The dataset balanced with simple random undersampling performed the best by a wide margin
- Further work should include:
  - Continued exploration and evaluation of data balancing techniques
  - Incorporate additional datasets beyond “base”
  - Explore additional modeling approaches (neural networks and ensembles with built-in undersampling)

Thank You

Q&A

Appendix

Literature Review & Method Analysis

1. **Article**: A machine learning based credit card fraud detection using the GA algorithm for feature selection“[1]

   Published in: A machine learning based credit card fraud detection using the GA algorithm for feature selection | Journal of Big Data

2. **Article**: A supervised machine learning algorithm for detecting and predicting fraud in credit card transactions“[3]

   Published in: https://www.sciencedirect.com/science/article/pii/S2772662223000036#b29

3. **Article**: Credit Card Fraud Detection using Artificial Neural Network and Back Propagation

   Published In: 2020 4th International Conference on Intelligent Computing and Control Systems (ICICCS)

4. **Article**: Explainable AI: current status and future directions

   Published In: IEEE https://arxiv.org/pdf/2107.07045.pdf

5. **Article**: Semi-Supervised Learning Classification Based on Generalized Additive Logistic Regression for Corporate Credit Anomaly Detection

   Published In: IEEEXplore

Literature Review & Method Analysis

6. **Article**: General Pitfalls of Model-Agnostic Interpretation Methods for Machine Learning Models

   Published in: xxAI - Beyond Explainable AI - International Workshop, Held in Conjunction with ICML 2020

7. **Article**: Explainable Machine Learning for Fraud Detection

   Published in: IEEE Computer Special Issue on Explainable AI and Machine Learning

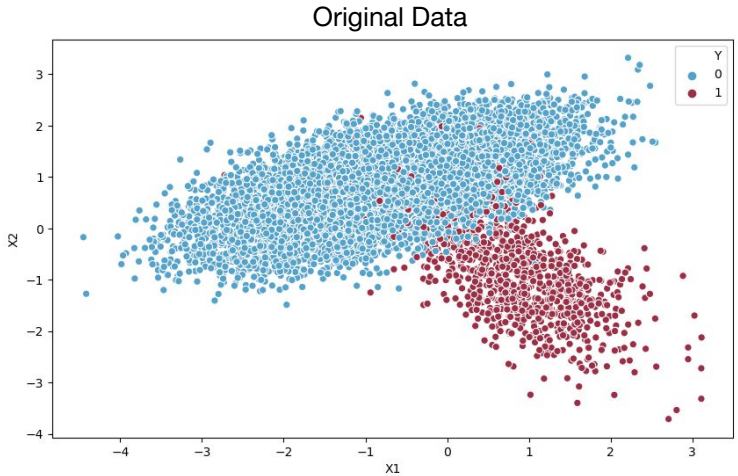
8. **Article**: Performance evaluation of class balancing techniques for credit card fraud detection

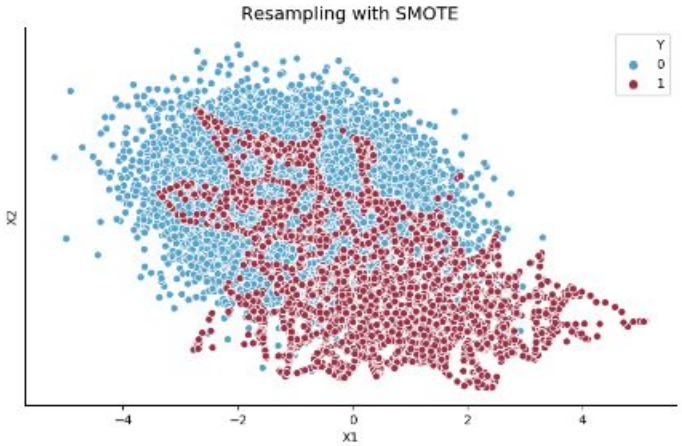
   Published In: 2017 IEEE International Conference on Power, Control, Signals and Instrumentation Engineering (ICPCSI)

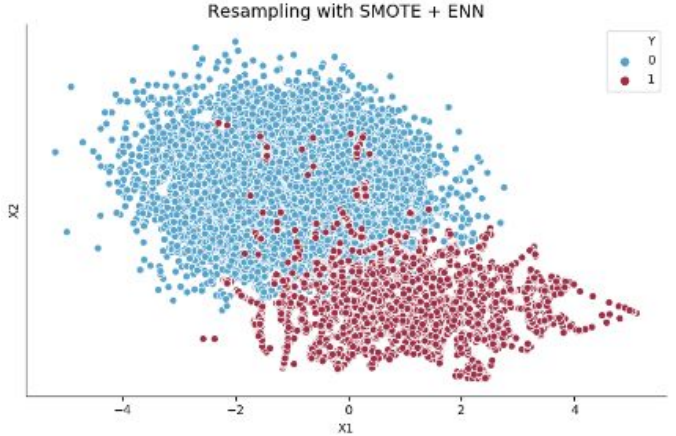
Glossary of matrices

- **True Positive**: You predicted positive, and it's true.
- **True Negative**: You predicted negative, and it's true.
- **False Positive**: (Type 1 Error): You predicted positive, and it's false.
- **False Negative**: (Type 2 Error): You predicted negative, and it's false.
- **Accuracy**: the proportion of the total number of correct predictions that were correct.
- **Positive Predictive Value or Precision**: the proportion of positive cases that were correctly identified.
- **Negative Predictive Value**: the proportion of negative cases that were correctly identified.
- **Sensitivity or Recall**: the proportion of actual positive cases which are correctly identified.
- **Specificity**: the proportion of actual negative cases which are correctly identified.
- **Rate**: It is a measuring factor in a confusion matrix. It has also 4 types TPR, FPR, TNR, and FNR.

SMOTE vs SMOTE-ENN

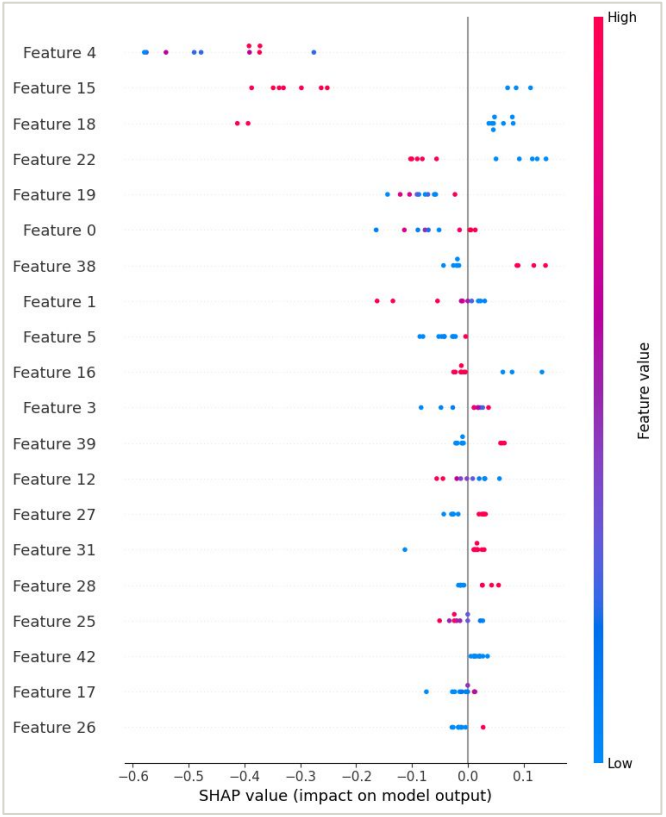






Source: [SMOTE for Imbalanced Classification with Python]

Ensemble model XAI

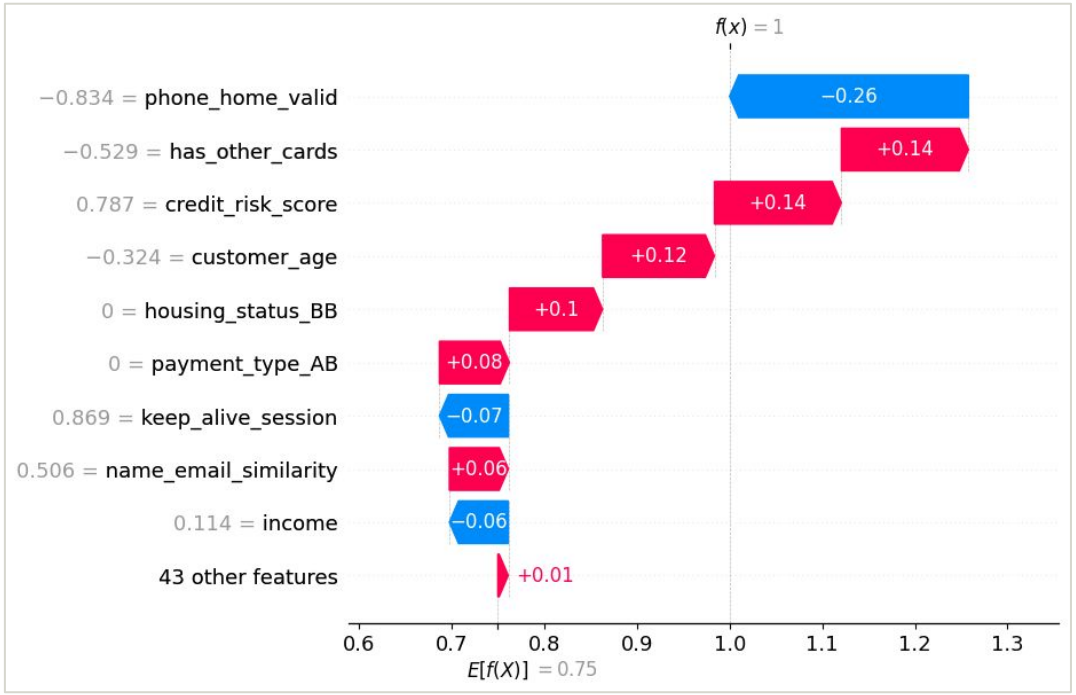



```
#   Column                              Non-Null Count  Dtype
---  ------                             --------------  -----
 0   income                             50000 non-null  float64
 1   name email similarity              50000 non-null  float64
 2   prev address months count          50000 non-null  float64
 3   current address months count       50000 non-null  float64
 4   customer age                       50000 non-null  float64
 5   days since request                 50000 non-null  float64
 6   intended balcon amount             50000 non-null  float64
 7   zip count 4w                       50000 non-null  float64
 8   velocity 6h                        50000 non-null  float64
 9   velocity 24h                       50000 non-null  float64
 10  velocity 4w                        50000 non-null  float64
 11  bank branch count 8w               50000 non-null  float64
 12  date of birth distinct emails 4w   50000 non-null  float64
 13  credit risk score                  50000 non-null  float64
 14  email is free                      50000 non-null  float64
 15  phone home valid                   50000 non-null  float64
 16  phone mobile valid                 50000 non-null  float64
 17  bank months count                  50000 non-null  float64
 18  has other cards                    50000 non-null  float64
 19  proposed credit limit              50000 non-null  float64
 20  foreign request                    50000 non-null  float64
 21  session length in minutes          50000 non-null  float64
 22  keep alive session                 50000 non-null  float64
 23  device distinct emails 8w          50000 non-null  float64
 24  device fraud count                 50000 non-null  float64
 25  month                              50000 non-null  float64
 26  payment type AA                    50000 non-null  float64
 27  payment type AB                    50000 non-null  float64
 28  payment type AC                    50000 non-null  float64
 29  payment type AD                    50000 non-null  float64
 30  payment type AE                    50000 non-null  float64
 31  employment status CA               50000 non-null  float64
 32  employment status CB               50000 non-null  float64
 33  employment status CC               50000 non-null  float64
 34  employment status CD               50000 non-null  float64
 35  employment status CE               50000 non-null  float64
 36  employment status CF               50000 non-null  float64
 37  employment status CG               50000 non-null  float64
 38  housing status BA                  50000 non-null  float64
 39  housing status BB                  50000 non-null  float64
 40  housing status BC                  50000 non-null  float64
 41  housing status BD                  50000 non-null  float64
 42  housing status BE                  50000 non-null  float64
 43  housing status BF                  50000 non-null  float64
 44  housing status BG                  50000 non-null  float64
 45  source INTERNET                    50000 non-null  float64
 46  source TELEAPP                     50000 non-null  float64
 47  device os linux                    50000 non-null  float64
 48  device os macintosh                50000 non-null  float64
 49  device os other                    50000 non-null  float64
 50  device os windows                  50000 non-null  float64
 51  device os x11                      50000 non-null  float64
dtypes: float64(52)
```

Dataset

#	Column	Count	Non-Nu	Dtype
0	income	50000	non-null	float64
1	name_email_similarity	50000	non-null	float64
2	prev_address_months_count	50000	non-null	int64
3	current_address_months_count	50000	non-null	int64
4	customer_age	50000	non-null	int64
5	days_since_request	50000	non-null	float64
6	intended_balcon_amount	50000	non-null	float64
7	payment_type	50000	non-null	category
8	zip_count_4w	50000	non-null	int64
9	velocity_6h	50000	non-null	float64
10	velocity_24h	50000	non-null	float64
11	velocity_4w	50000	non-null	float64
12	bank_branch_count_8w	50000	non-null	int64
13	date_of_birth_distinct_emails_4w	50000	non-null	int64
14	employment_status	50000	non-null	category
15	credit_risk_score	50000	non-null	int64
16	email_is_free	50000	non-null	int64
17	housing_status	50000	non-null	category
18	phone_home_valid	50000	non-null	int64
19	phone_mobile_valid	50000	non-null	int64
20	bank_months_count	50000	non-null	int64
21	has_other_cards	50000	non-null	int64
22	proposed_credit_limit	50000	non-null	float64
23	foreign_request	50000	non-null	int64
24	source	50000	non-null	category
25	session_length_in_minutes	50000	non-null	float64
26	device_os	50000	non-null	category
27	keep_alive_session	50000	non-null	int64
28	device_distinct_emails_8w	50000	non-null	int64
29	device_fraud_count	50000	non-null	int64
30	month	50000	non-null	int64

False negative (smote/Ensemble)



Actual test record:	
income	0.6
name_email_similarity	0.636561
prev_address_months_count	18
current_address_months_count	50
customer_age	30
days_since_request	0.031813
intended_balcon_amount	10.685194
payment_type	AA
zip_count_4w	1236
velocity_6h	5186.264123
velocity_24h	5734.857872
velocity_4w	5086.435013
bank_branch_count_8w	2
date_of_birth_distinct_emails_4w	15
employment_status	CA
credit_risk_score	188
email_is_free	1
housing_status	BC
phone_home_valid	0
phone_mobile_valid	1
bank_months_count	6
has_other_cards	0
proposed_credit_limit	500.0
foreign_request	0
source	INTERNET
session_length_in_minutes	5.526779
device_os	linux
keep_alive_session	1
device_distinct_emails_8w	1
device_fraud_count	0
month	2
Name: 237036,	