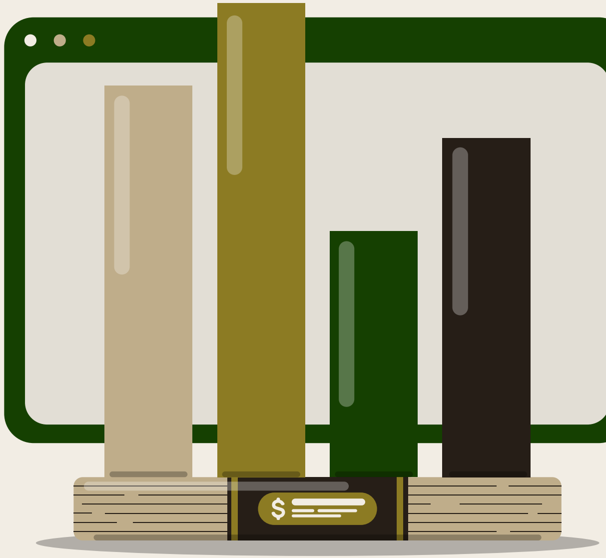
Bitcoin Fraudulent Transaction Detection

Austin Yeh, Haoxiang Yi, Muhammad Ibrahim,
Luke Leon, Biagio Alessandrello

TABLE OF CONTENTS

01

Problem



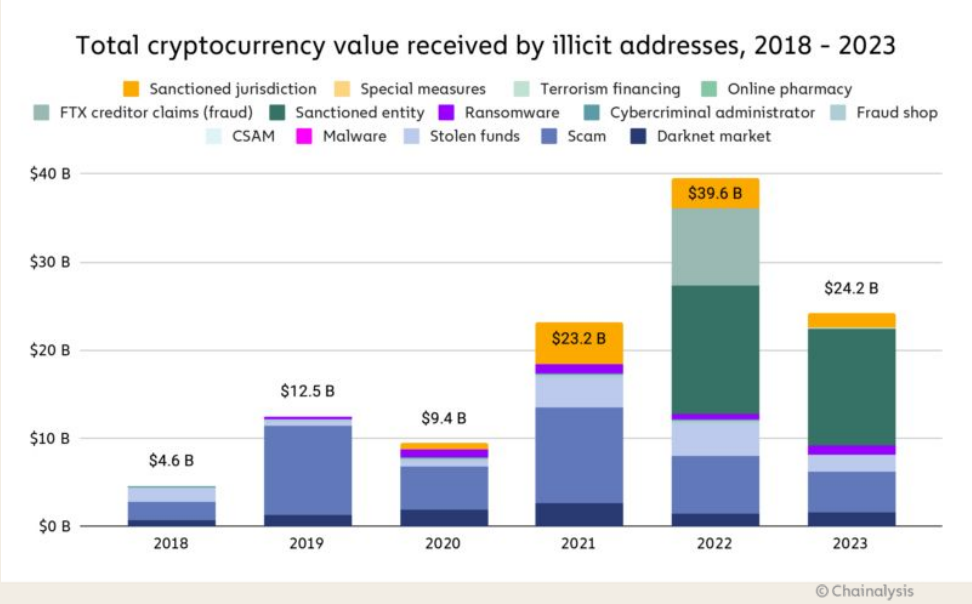
Problem Overview

In a Bitcoin network, transactions can be categorized as licit versus illicit ones. Our task is to analyze a network to detect fraudulent transactions.

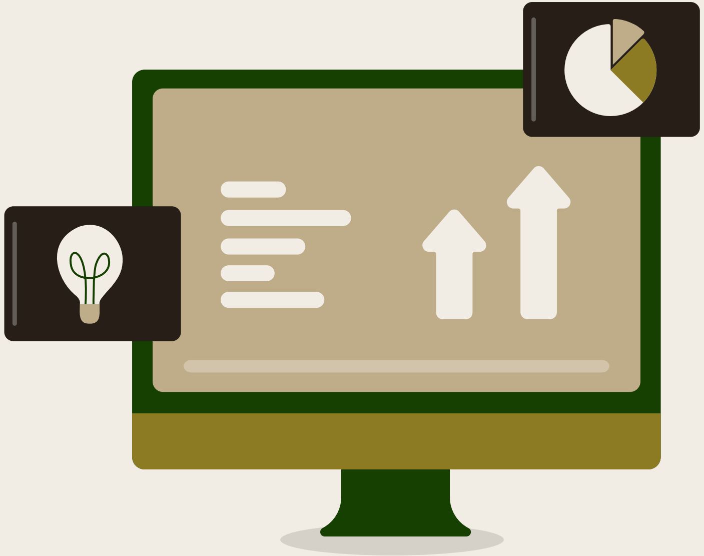
In 2023:

- $24.2 billion worth of cryptocurrencies was stolen globally

- Over 69,000 complaints to FBI from the public in the US regarding fraud using cryptocurrency



Sources: The Chainanalysis crypto crime report 2024 and FTC report on cryptocurrency scams



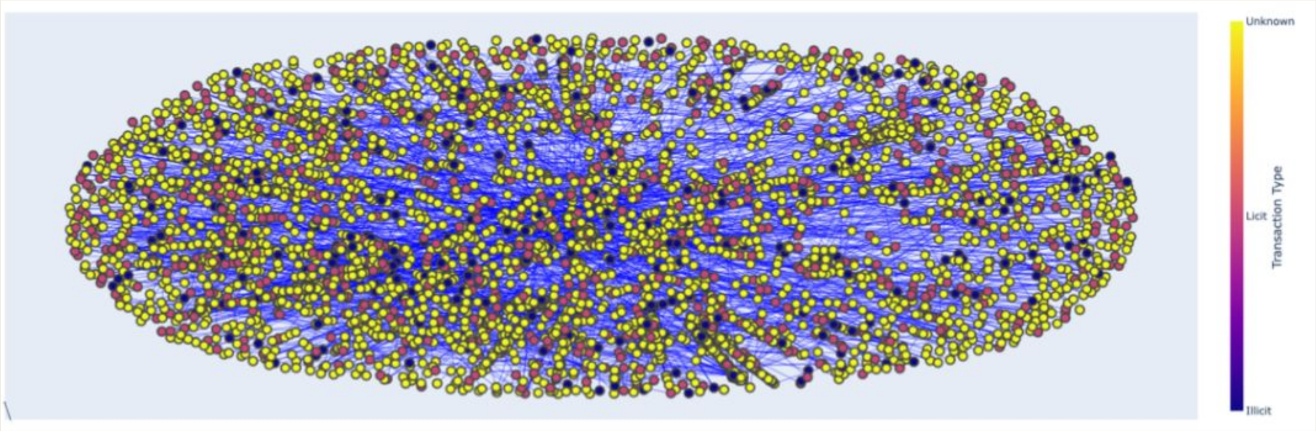
02

DATASET & EDA

What is a Bitcoin Transaction Graph?

A transaction graph from the Bitcoin blockchain, containing 203,769 nodes and 234,355 edges.

- A node represents a transaction
- An edge represents the flow of Bitcoins between two transactions



Source: Youssef Elmougy and Ling Liu. 2023. Demystifying Fraudulent Transactions and Illicit Nodes in the Bitcoin Network for Financial Forensics.

Original Dataset - Transactions

Network Structure:
- Size: 203,769 nodes and 234,355 edges

Class Distribution:
- Illicit: 2% of nodes
- Licit: 21% of nodes
- Unknown: 77%

Features:
- 166 anonymized attributes

Time
- Time steps from 1 to 49, representing 2-week increments of transactions

	txId	Time step	Local_feature_1	Local_feature_2	Local_feature_3	Local_feature_4	Local_feature_5	Local_feature_6
0	3321	1	-0.169615	-0.184668	-1.201369	-0.121970	-0.043875	-0.113002
1	11108	1	-0.137586	-0.184668	-1.201369	-0.121970	-0.043875	-0.113002
2	51816	1	-0.170103	-0.184668	-1.201369	-0.121970	-0.043875	-0.113002
3	68869	1	-0.114267	-0.184668	-1.201369	0.028105	-0.043875	-0.113002
4	89273	1	5.202107	-0.210553	-1.756361	-0.121970	260.090707	-0.113002
...	...	...	...	...	...	...	...	...
203764	158304003	49	-0.165622	-0.139563	1.018602	-0.121970	-0.043875	-0.113002
203765	158303998	49	-0.167040	-0.139563	1.018602	-0.121970	-0.043875	-0.113002
203766	158303966	49	-0.167040	-0.139563	1.018602	-0.121970	-0.043875	-0.113002
203767	161526077	49	-0.172212	-0.139573	1.018602	-0.121970	-0.043875	-0.113002
203768	194103537	49	-0.172212	-0.139573	1.018602	-0.121970	-0.043875	-0.113002

Extended Dataset - Actors

Network Structure:

- Identifies actors linked to transactions in the Bitcoin network

Features:

- 56 non-anonymized attributes (btc sent, btc received, etc.)

Linkage:

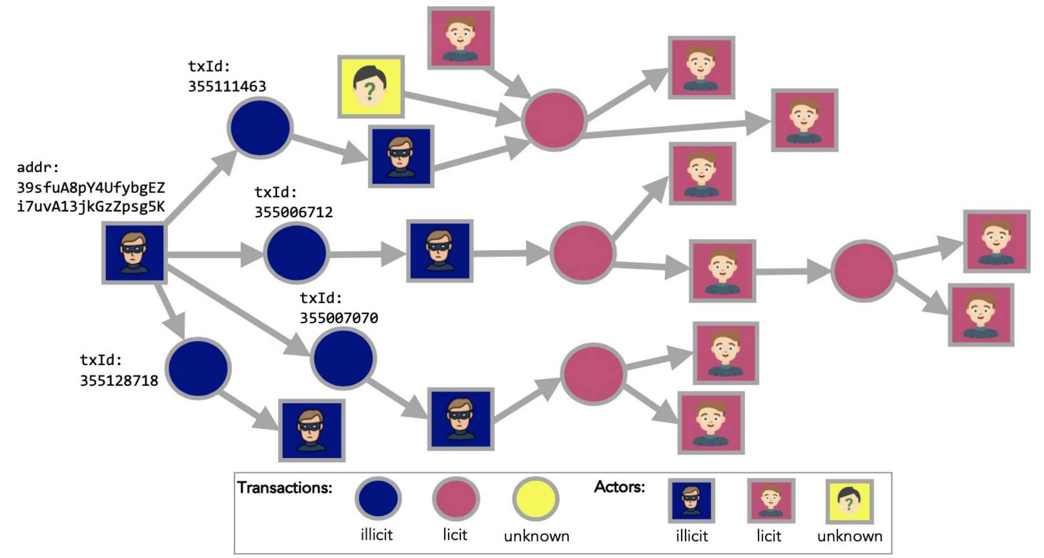
- Can map transactions (txId) to their associated actors

Class Distribution:

- Illicit:
- Licit:
- Unknown:

	input_address	txId
0	14YRXHHof4BY1TVxN5FqYPcEdpmXiYT78a	230325127
1	13Lhad3SAmu2vqYg2dxbNcxH7LE77kJu2w	230325139
2	1MAQQZn7EHP6J3erXByCciFiVcgS8ZhWqz	86875675

	txId	output_address
0	230325127	1GASxu5nMntiRKdVtTVRvEbP965G51bhHH
1	230325127	14YRXHHof4BY1TVxN5FqYPcEdpmXiYT78a
2	230325139	1GFdrdgtG34GChM8SMpMwcXFc4nYbH1A5G



Source: Youssef Elmougy and Ling Liu. 2023. Demystifying Fraudulent Transactions and Illicit Nodes in the Bitcoin Network for Financial Forensics.

Overview of Datasets

Original Dataset: 203,769 transactions, 49 time steps, licit/illicit/unknown labels, 166 anonymized features.

Extended Dataset: 822,942 wallet addresses, address-to-transaction mappings, temporal interactions.

Transactions

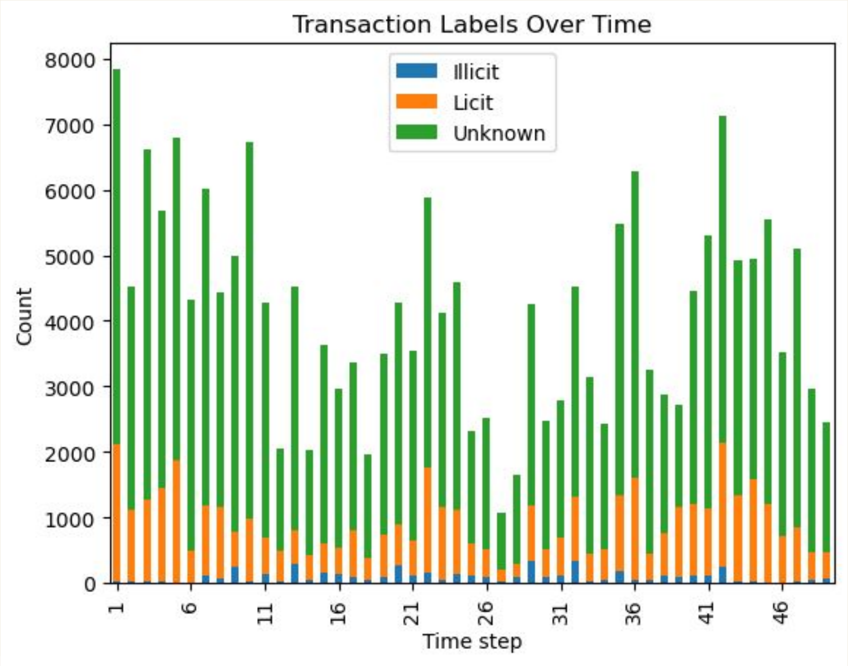
# Nodes (transactions)	203,769
# Edges (money flow)	234,355
# Time steps	49
# Illicit (class-1)	4,545
# Licit (class-2)	42,019
# Unknown (class-3)	157,205
# Features	183

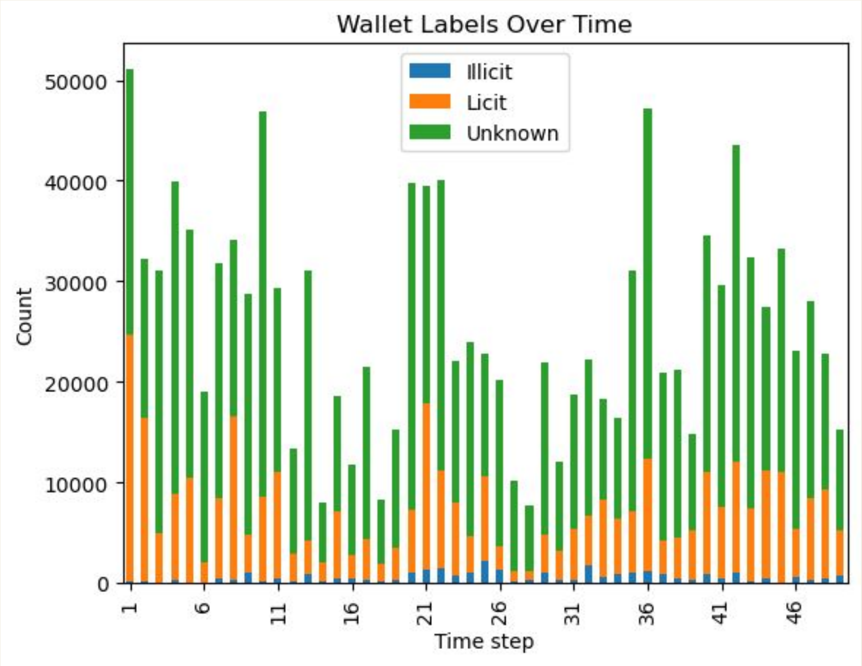
Actors

# Wallet addresses	822,942
# Nodes (temporal interactions)	1,268,260
# Edges (addr-addr)	2,868,964
# Edges (addr-tx-addr)	1,314,241
# Time steps	49
# Illicit (class-1)	14,266
# Licit (class-2)	251,088
# Unknown (class-3)	557,588
# Features	56

Source: Youssef Elmougy and Ling Liu. 2023. Demystifying Fraudulent Transactions and Illicit Nodes in the Bitcoin Network for Financial Forensics.

Class Labels Over Time



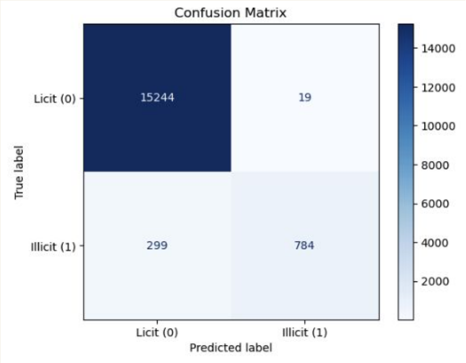


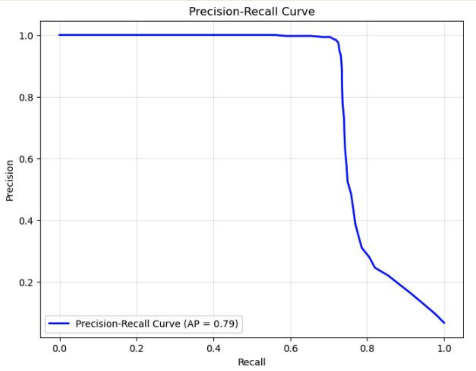
03

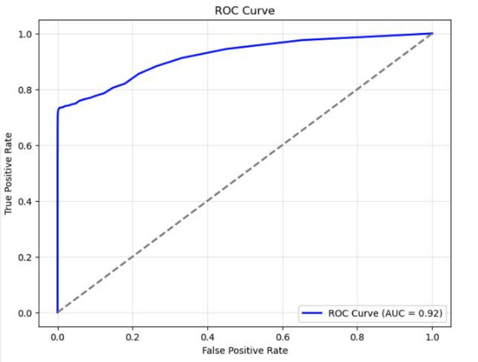
Methodology

A Simplistic Approach

1. Drop all unknown transactions
   a. ~157k out of ~204k
2. Create train/test split based on time step
   a. 1-34 train, 35-49 test
3. Run various models
   a. LR, RF, MLP, XGB
4. Select one with best metrics on test set
   a. RF (shown on slide)







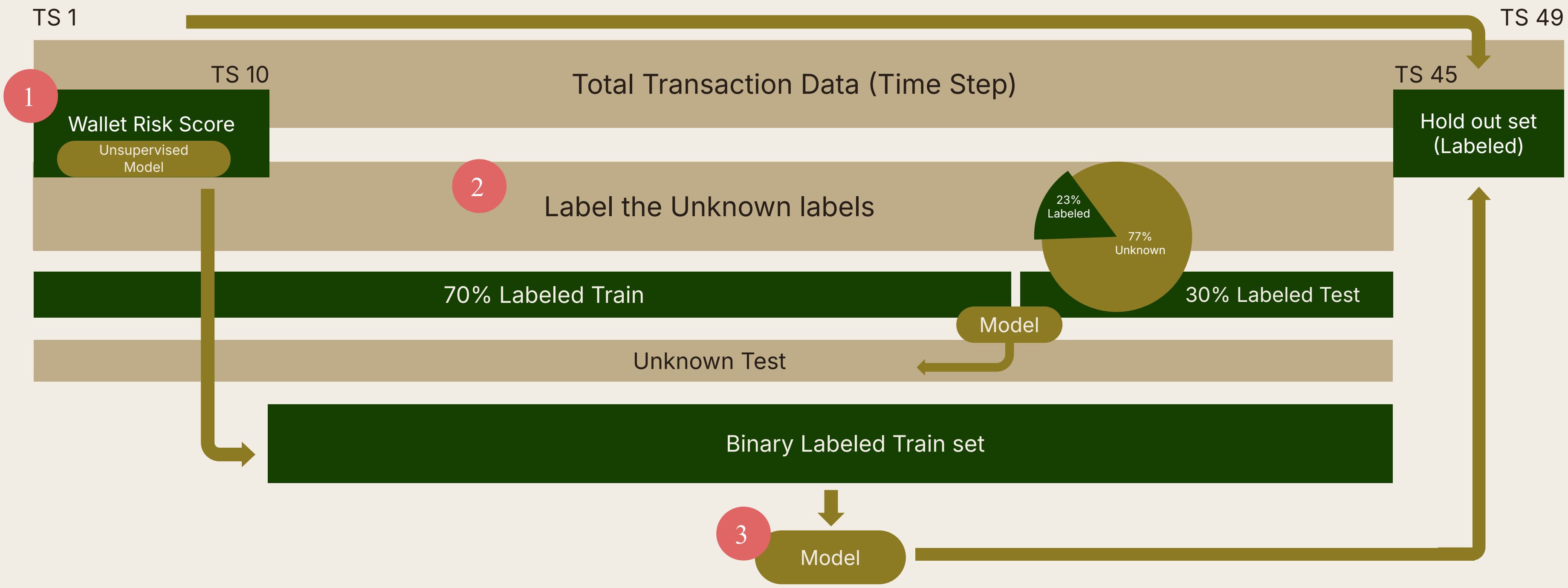
Code/approach adapted from: Youssef Elmougy and Ling Liu. 2023. Demystifying Fraudulent Transactions and Illicit Nodes in the Bitcoin Network for Financial Forensics.

We Wanted to Do Something More Involved

Make Use of All Unlabeled Data

Develop a Unique Methodology

Modeling Approach



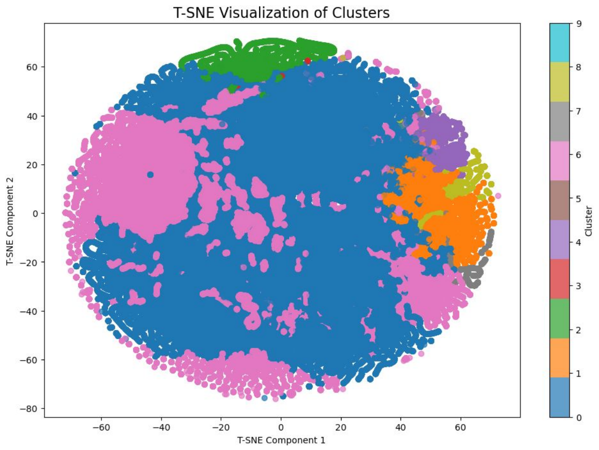
1 Unsupervised (Wallets)

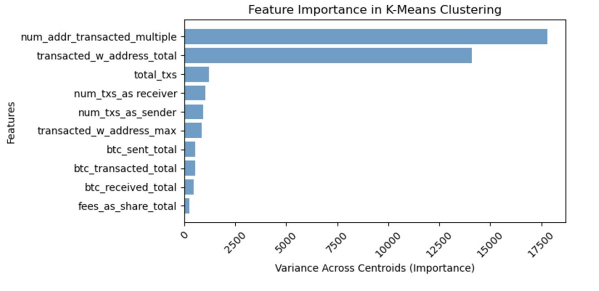
T-SNE Clustering

- Grouped into 10 buckets
- Analyzed the percentage fraud transactions in each bucket
- Most had very low percentage, but showed some distinction

K-Means Clustering

- Two of the features were substantially important in this model
  - # of address transacted multiple
  - Transacted with address total
  - Most likely correlated
- Other notable important features
  - # of receiving transactions
  - # of sender transactions
  - Amount of BTC being transferred
- Will use the model to predict category of each wallet post T10.



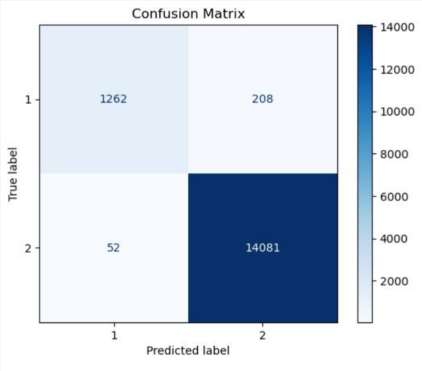


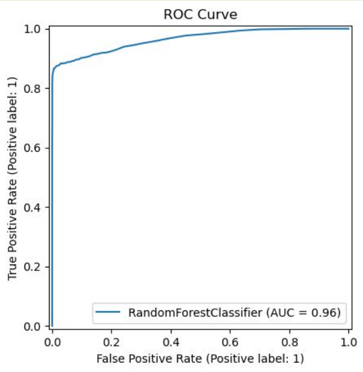
Score	Fraud Percentage
9	0.000
8	0.066
7	0.265
6	0.995
5	0.000
4	0.202
3	0.000
2	0.000
1	0.342
0	0.465

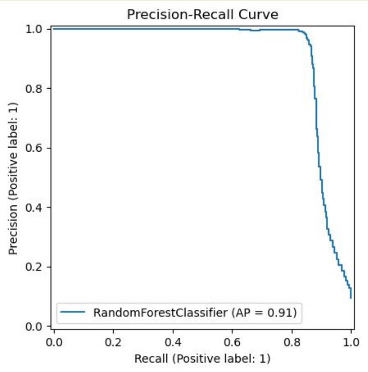
2 Labeling (TX)

Random Forest

Performance Metrics:
Accuracy: 0.98
Precision: 0.96
Recall: 0.86
F1-Score: 0.91

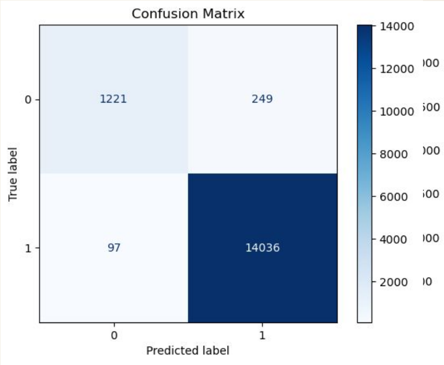


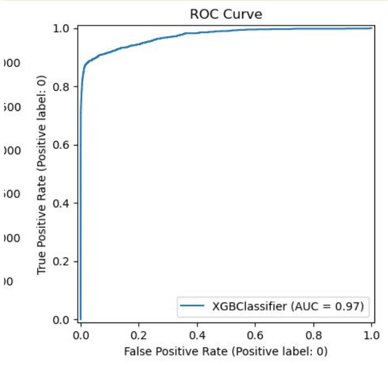


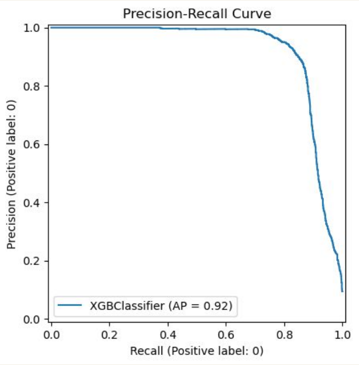


XGBoost

Performance Metrics:
Accuracy: 0.98
Precision: 0.98
Recall: 0.99
F1-Score: 0.99







3 Results

MODEL	Accuracy	Precision	Recall	AUROC	AUPRC
Simplistic*	0.98	0.97	0.72	0.92	0.79
Random Forest	0.97	0.40	0.02	0.87	0.17
Random Forest w Smote	0.97	0.33	0.02	0.88	0.13
XGBoost	0.97	0.97	1.00	0.92	0.26
XGBoost w Smote	0.58	0.95	0.93	0.91	1.00
GNN	0.97	0.94	0.97	0.76	0.99

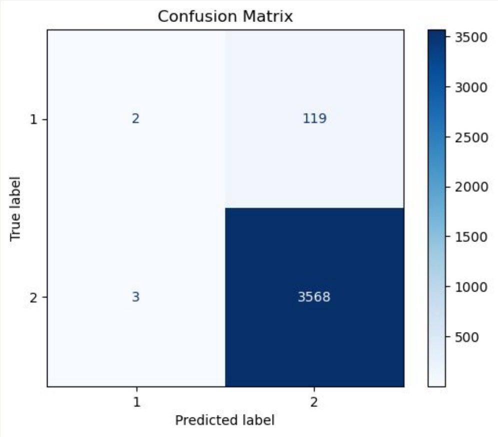
Models not mentioned

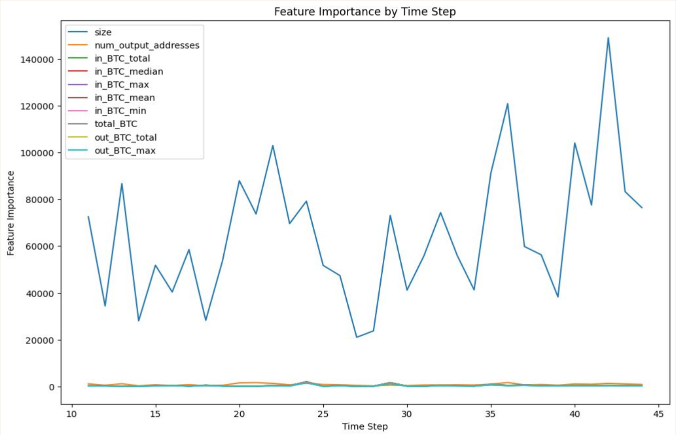
- Neural Network
- Convolutional Neural Net
- Support Vector Classification
- Simple Classification

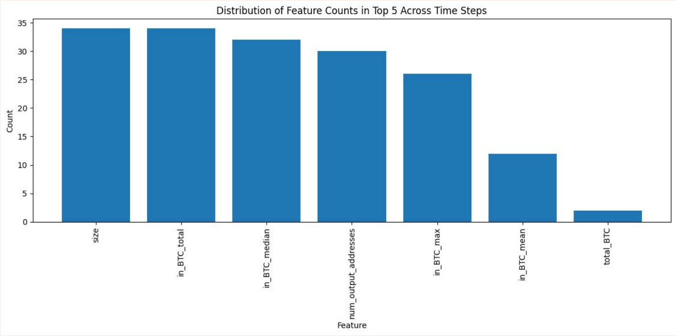
* Uses different train/test split and methodology than the rest of the models

Random Forest

- Dropped Class, TxId, and Time Step
- Fit on times steps 11 through 44
- Predicted time steps 45 to 49
- Results
  - Accuracy: 0.97
  - Precision: 0.40
  - Recall: 0.02
  - F1-Score: 0.03
- With Smote Results
  - Accuracy: 0.97
  - Precision: 0.33
  - Recall: 0.02
  - F1-Score: 0.03

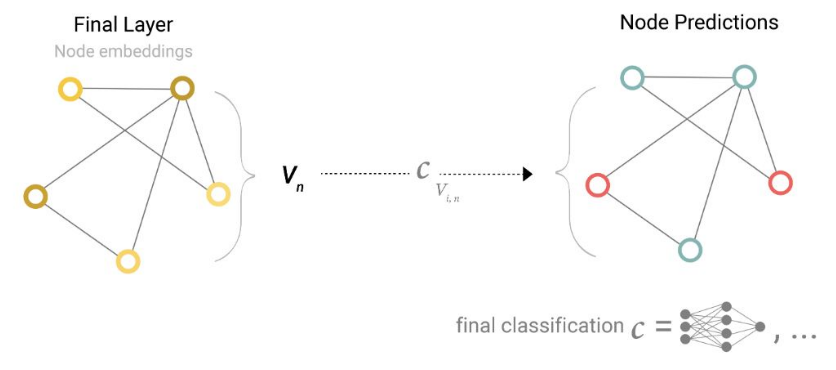


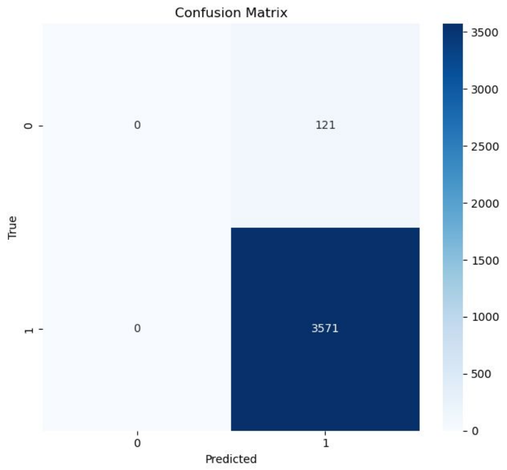


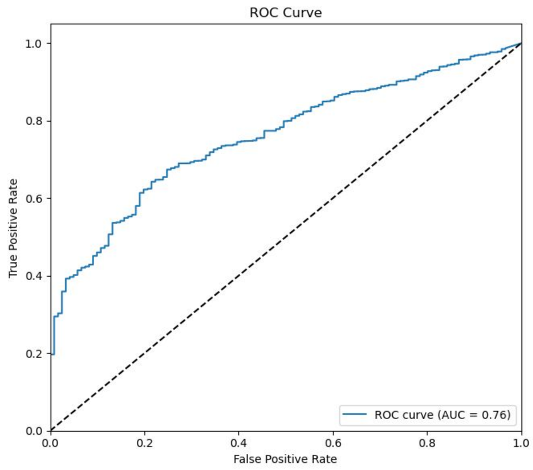


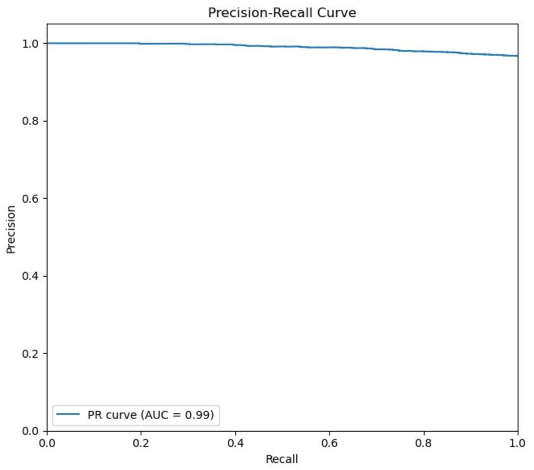
Graph Neural Net

- Most Intuitive Approach
- Results
  - Accuracy: 0.97
  - Precision: 0.94
  - Recall: 0.97
  - F1-Score: 0.98





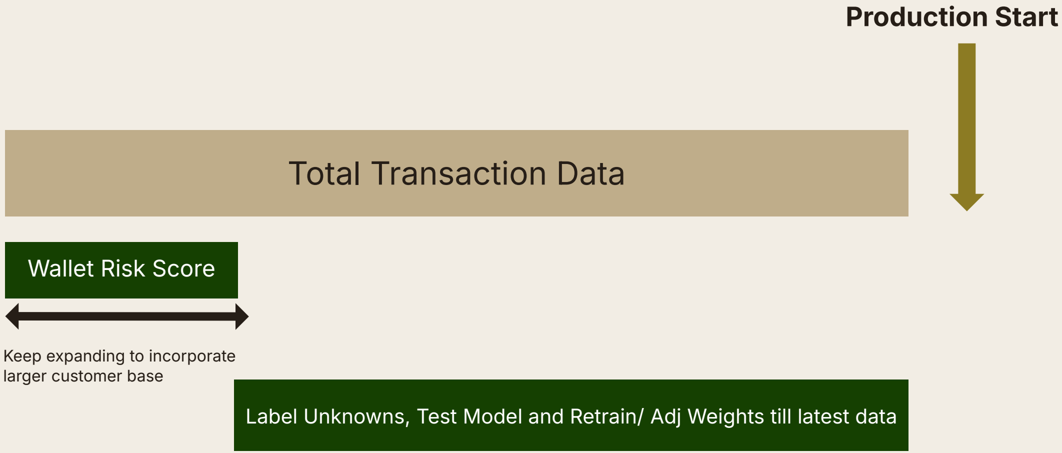




04

Real World Application

Production Pipeline



1. As new data comes in we expand our window for finding better labeling of risk scores for our wallets. Eventually we would have a larger customer lookup or better unsupervised model

2. We keep a rolling window for labeling and training on the transactions data and test on the hold out set

3. Once a good model is found we retrain till the last timestep to incorporate the latest information in the production model

05

Takeaways

Challenges / Takeaways

Complex Problem

Anonymized Features - Difficult Feature Engineering

Many Modeling Approaches

Immense Compute Power

Problems with Predicting a Proxy

Preventing Data Leakage

Experienced Leakage on Several Approaches

New Techniques

Graph Neural Networks (GNN)

Self-Supervised Learning

Thank You!

Questions?