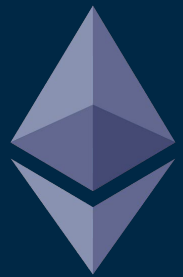
STAYING AHEAD OF FRAUDULENT ETHEREUM WALLETS

Nick Kim



FRAUD IN NUMBERS

PlusToken

$3 billion Ponzi Scheme [1]

180k *BTC*, 6.4m *ETH* and more swindled from +3 million users

Google Chrome Extensions

49 Apps Removed from Webstore [2] around Apr 2020

Browser extensions disguised leading to a loss of 1.4m *XRP* before getting removed

Tokensniffer.com

Over **46,000 tokens** [3] identified as scam

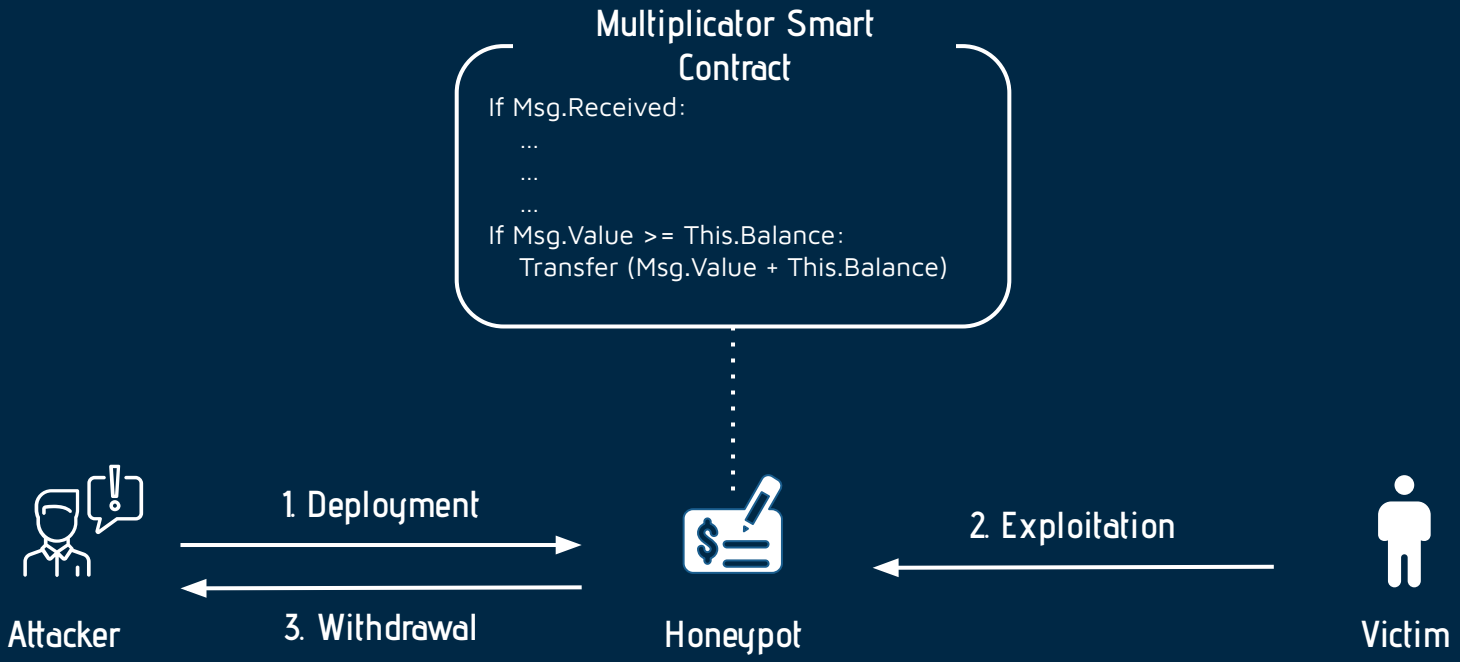
as of Oct 2021

Pranksy

NFT artwork from the artist Banksy impersonator sold at $300k [4]

[1] THE 2020 STATE OF CRYPTO CRIME; [2] Google Removes 49 Phishing Extensions That Steal Cryptocurrency Data; [3] A whole new world of rug pulls, honeypots and crypto Ponzi schemes; [4] Someone paid more than $300K for a fake Banksy NFT — and the scammer gave it all back

FRAUD IN ILLUSTRATION: HONEYPOT



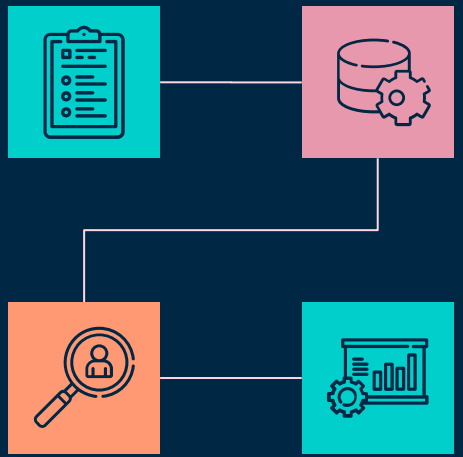
Source: https://www.usenix.org/conference/usenixsecurity19/presentation/ferreira

SOLUTION FOR THE ETHEREUM ORGANIZATION

- Build a classification model from known fraud accounts
- Provide functionality to **scan counterparty account address** to check for fraud likelihood

METHODOLOGY

Dataset of Ethereum accounts

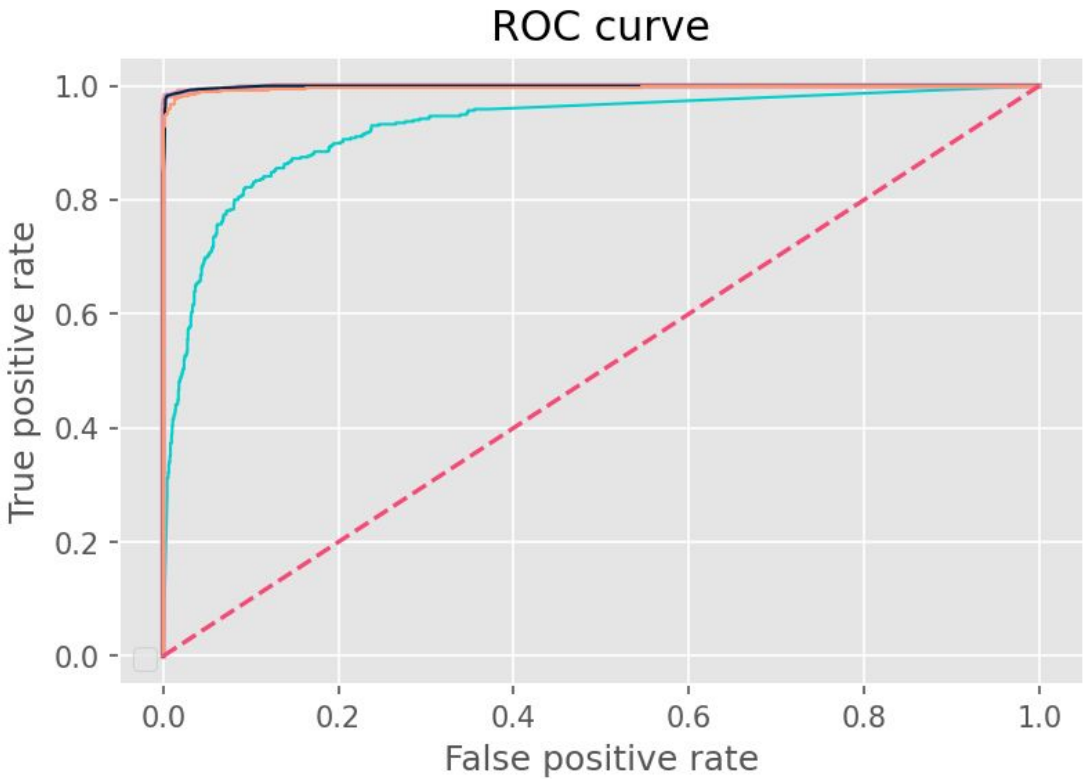


Preprocessing

Classify fraud accounts

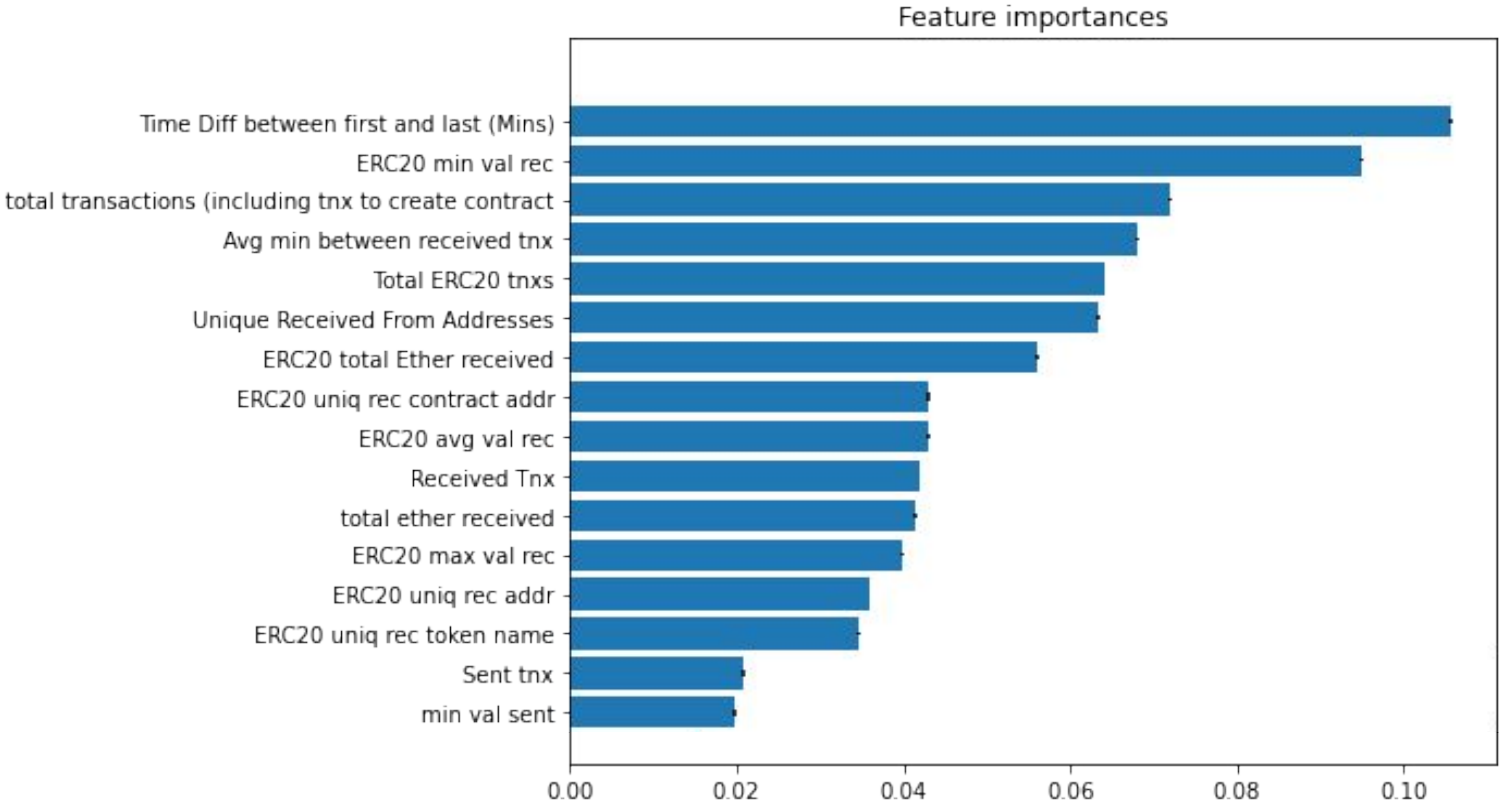
Evaluate performance

TREE-BASED MODELS



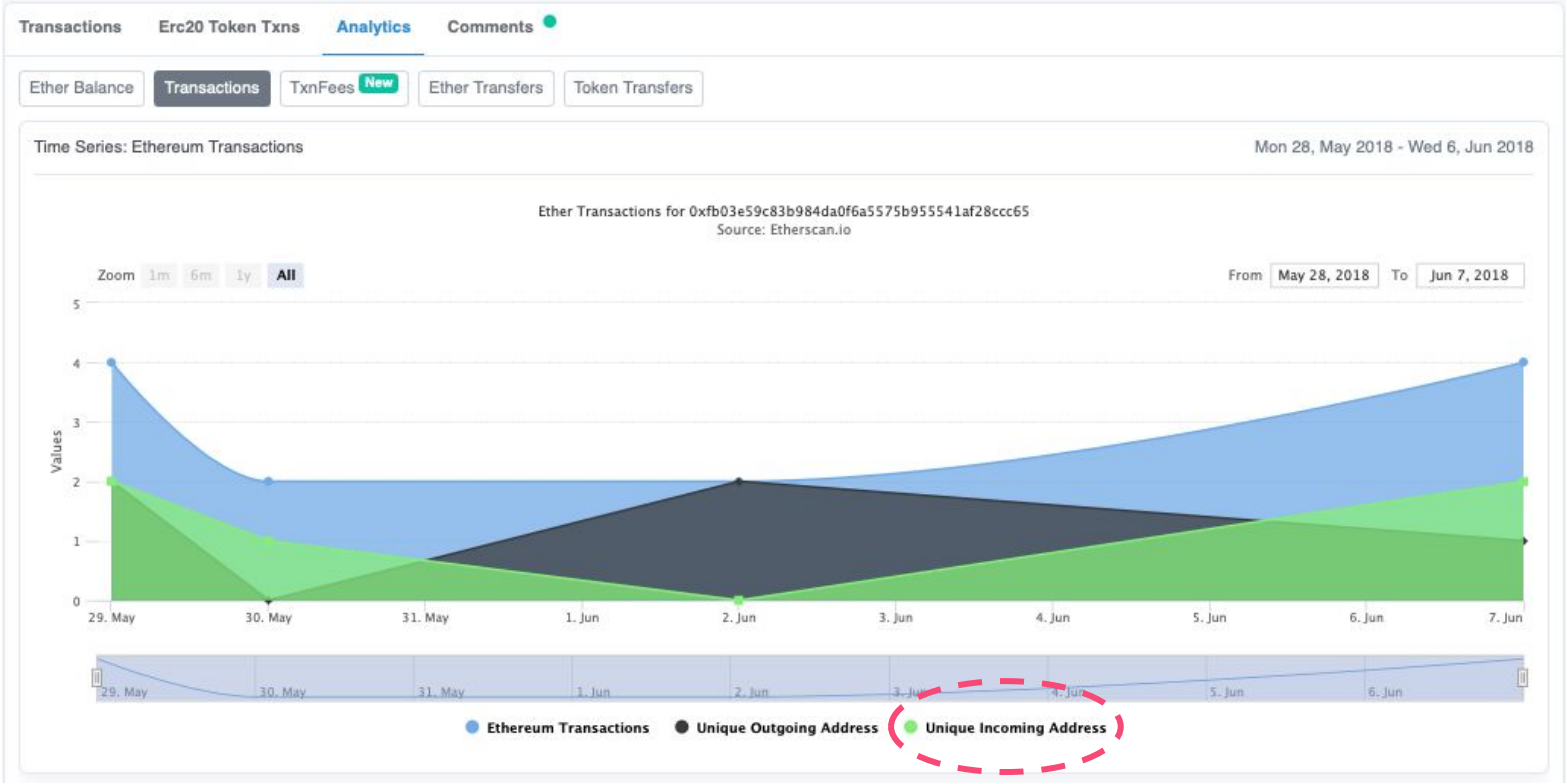
Model	AUC	(Speed)
RandomForest	**0.999**	**(10.5x)**
Extra-Trees	**0.999**	**(1.8x)**
XGBoost	0.997	(fastest)
kNN	0.926	(78.0x)

WHAT DOES THE MODEL LOOK AT?



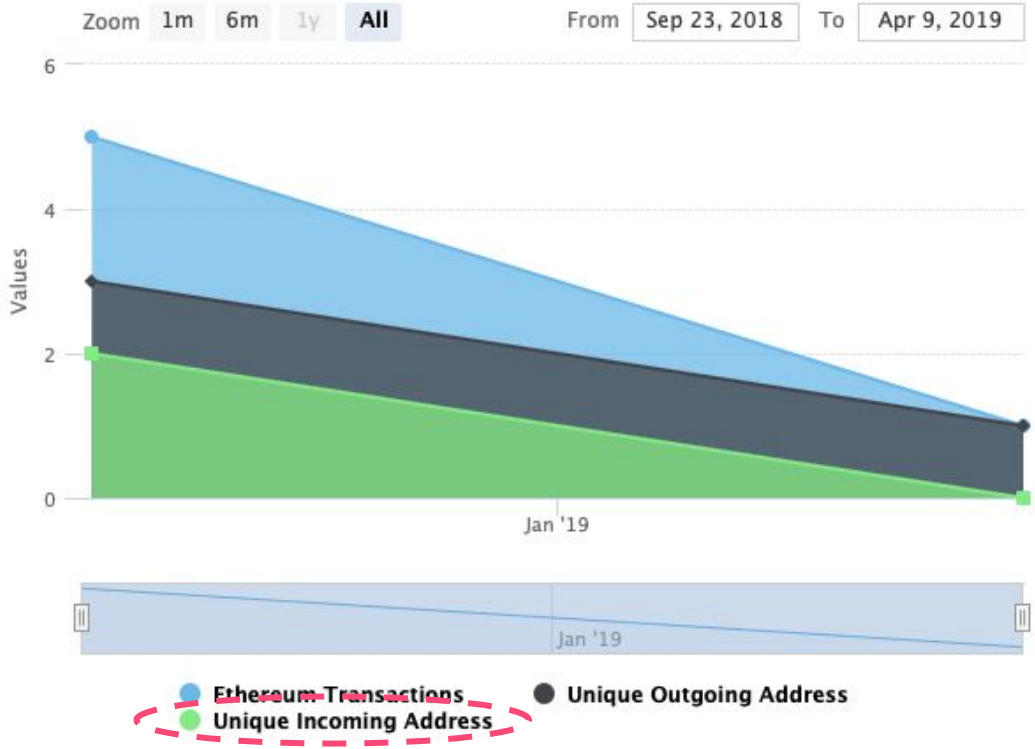
WHAT DOES THE MODEL LOOK AT?

Fraud / Positive Class



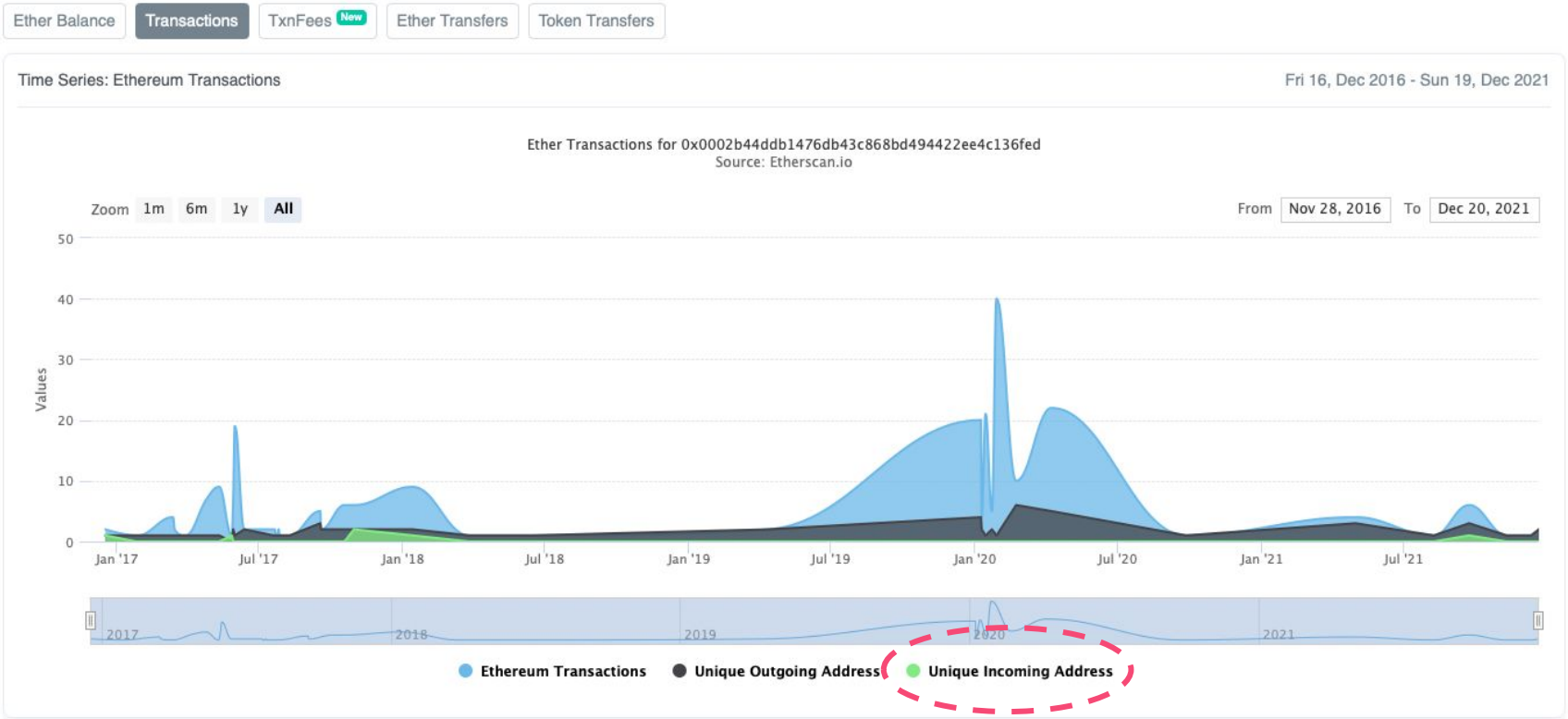
WHAT DOES THE MODEL LOOK AT?

Fraud / Positive Class

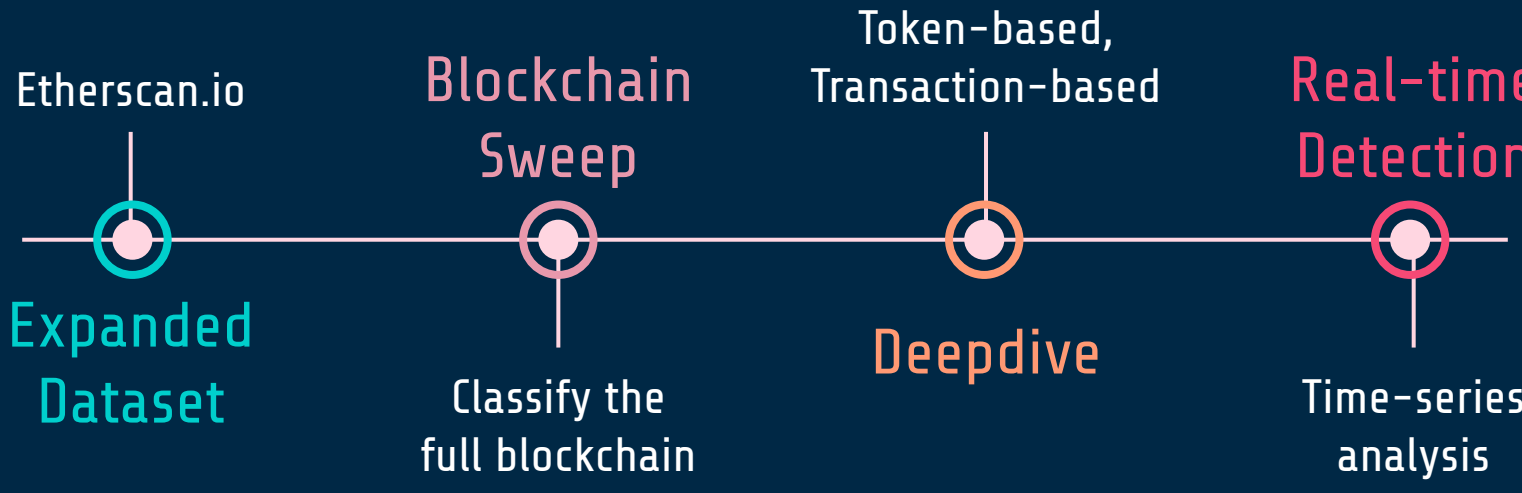


WHAT DOES THE MODEL LOOK AT?

Legitimate / Negative Class



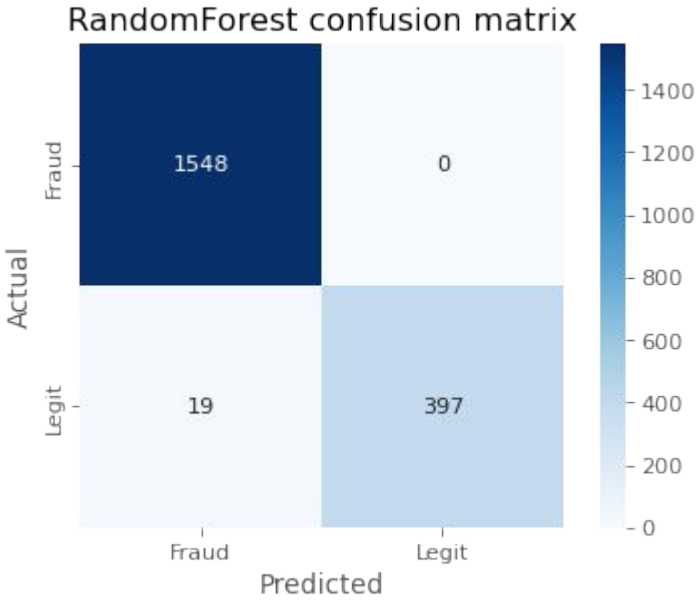
FUTURE WORK


Etherscan.io
Expanded Dataset
Blockchain Sweep
Classify the full blockchain
Token-based, Transaction-based
Deepdive
Real-time Detection
Time-series analysis

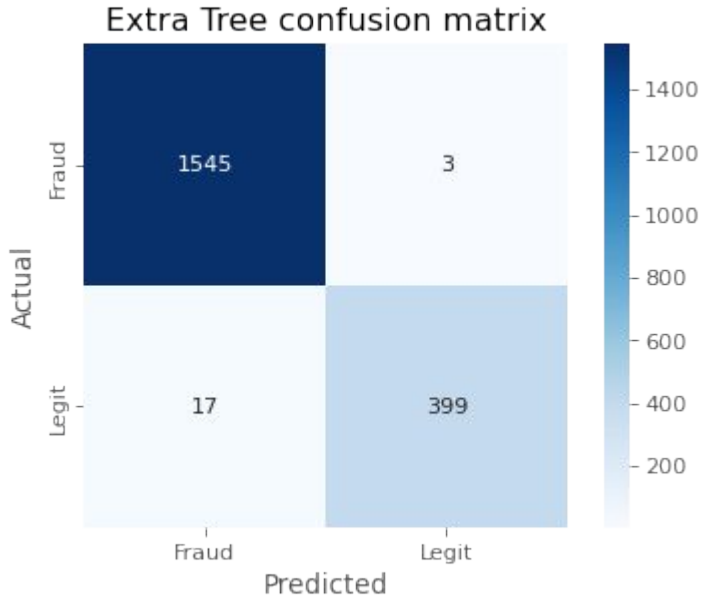
Q&A

APPENDIX

Confusion matrices of Tree-based models



Threshold	0.64
Precision	1.00
Recall	0.95
F-Beta	0.97



Threshold	0.64
Precision	0.99
Recall	0.96
F-Beta	0.97

APPENDIX

Tree-based models' evaluation metrics given changes in decision threshold