



ETHEREUM FRAUD DETECTION

By Ethan Kunin

CRYPTOCURRENCY BACKGROUND

- Bitcoin invented in 2009
- Laszlo Hanyecz pays 10,000 BTC for 2 boxes of pizza
- Mt. Gox is hacked, 850,000 BTC are stolen
- Ethereum proposed in 2013, completes crowdfunding in 2014
- Ethereum is minted in 2015 by Vitalik Buterin
- 2017 ICO bubble crashes
- August 2020, Michael Saylor invests \$250 million of MicroStrategy's corporate reserves in Bitcoin
- Decentralized Finance gains momentum and protocols are being executed on the Ethereum blockchain

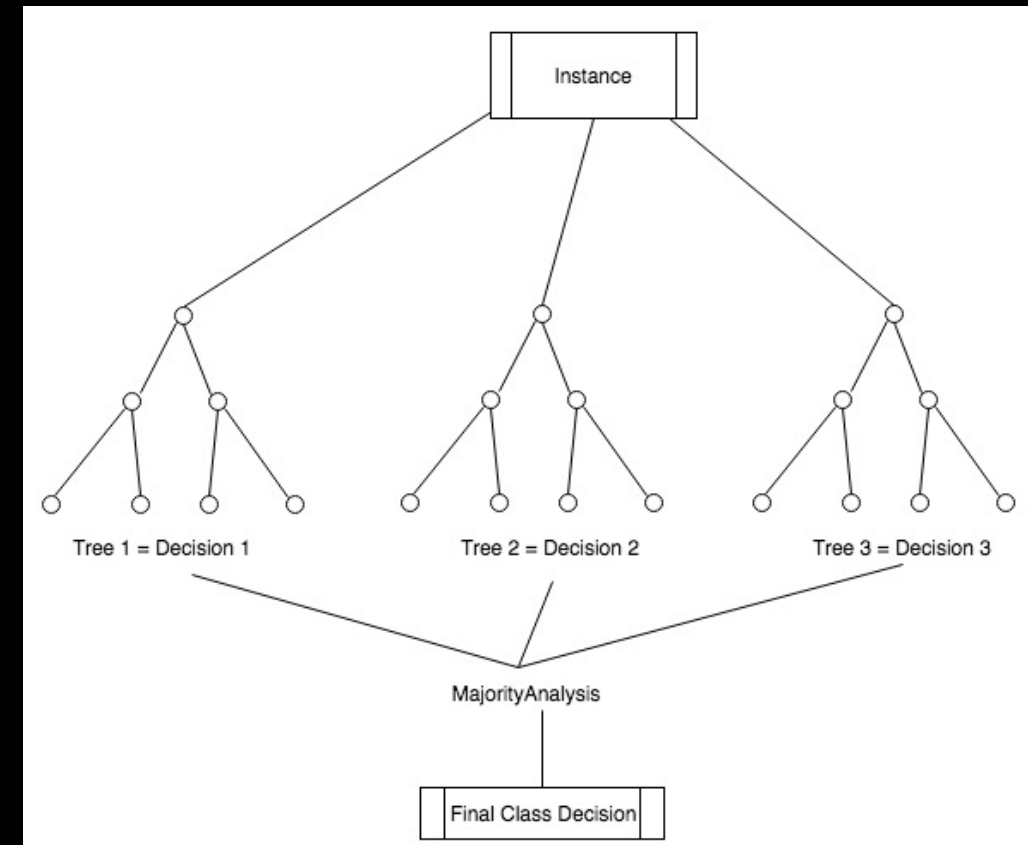
CRYPTOCURRENCY FRAUD

- Fraud is common in cryptocurrency transactions because they are **decentralized** and **anonymous**
 - Investing in fraudulent projects
 - Buying a good or service and not receiving it in return
 - Exchanges can be attacked
 - Flash loans



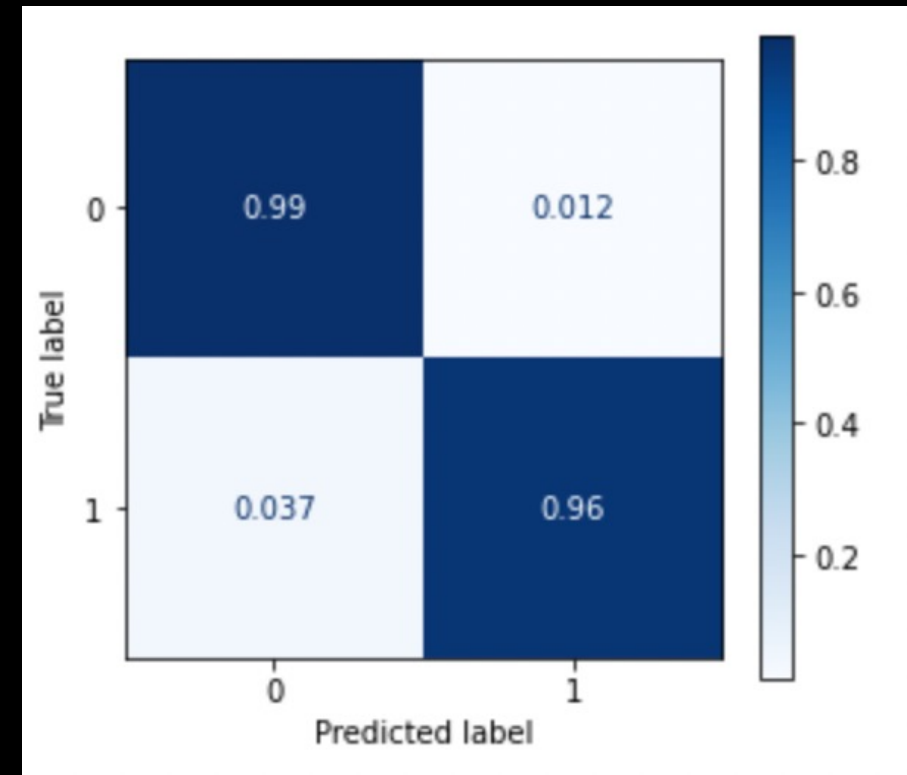
BUILDING A RANDOM FOREST MODEL TO DETECT FRAUD

- Trained on a dataset with **10,000** observations
- Features are based on **Ethereum** and **ERC20** transactions
- Due to the nature of the model type, it is **outlier resistant**
- **98% accurate** for detecting fraud and valid transactions
 - Accurately classifies valid transactions 99% of the time
 - Accurately classifies fraudulent transactions 96% of the time



MODEL CONTINUED...

- Purpose of the model is to optimize for recall
 - True Fraud transactions divided by True Fraud transactions + Falsely classified Fraud transactions
- Cost of notifying customer that a valid transaction may not be valid is very low
- Cost of not notifying a customer that a fraudulent transactions is fraudulent is very high



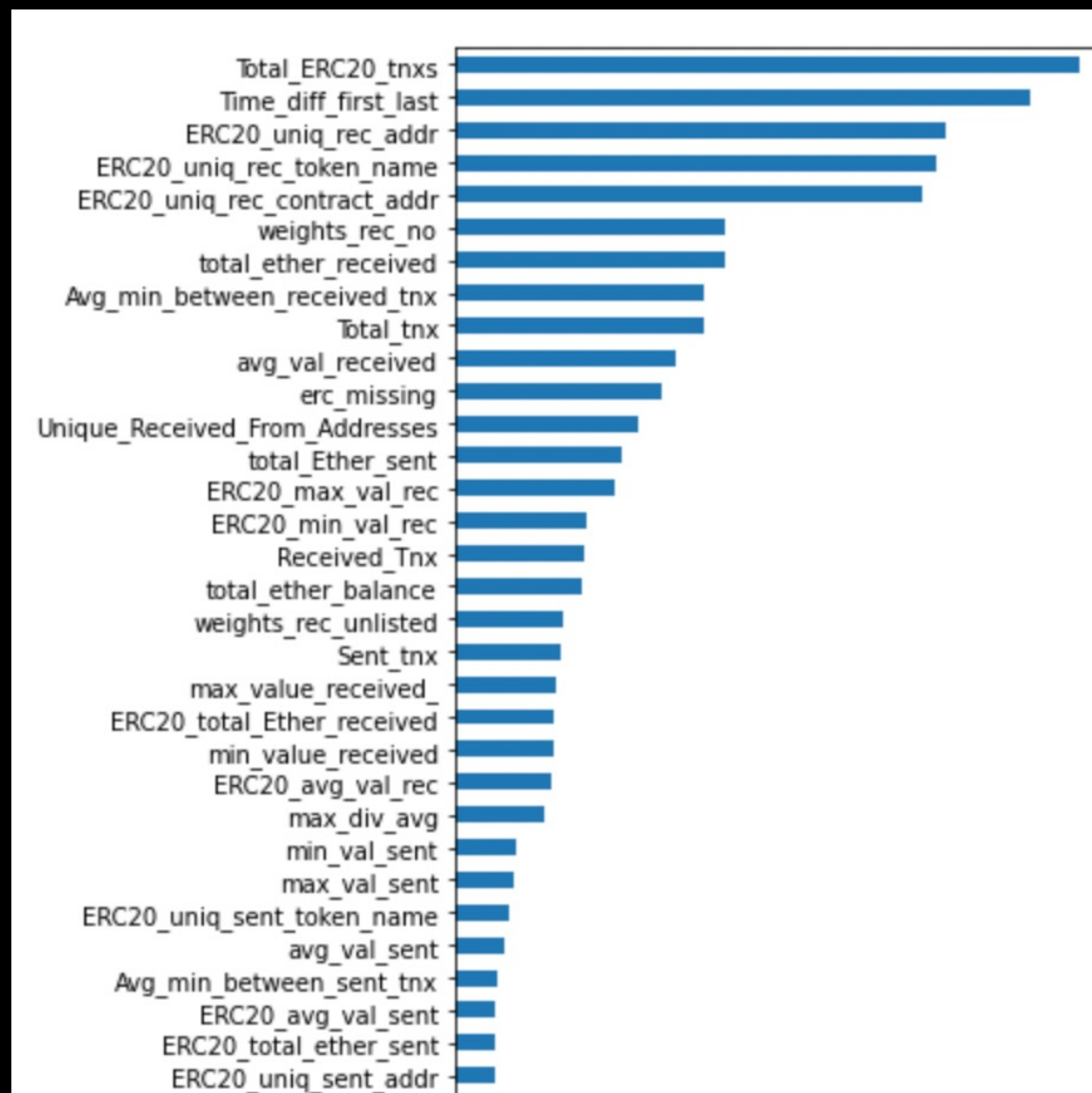
MODEL COMPARISONS

- Tried two different types of models optimizing each for recall score
- Random Forest had a top recall score of 96% compared to 93% for K-Nearest Neighbors
- Used the Random Forest because it did a better job at generalizing for new data

Model	Total Customers	Expected Valid Tnx	Expected Fraud Tnx Attempts	Defrauded Tnx	Total Cost	Avg. Expected Loss
Dummy	10,000	9,500	500	250	\$191,000	\$19.11
K-Nearest Neighbors	10,000	9,500	500	40	\$30,560	\$3.27
Random Forest	10,000	9,500	500	19	\$14,516	\$1.55

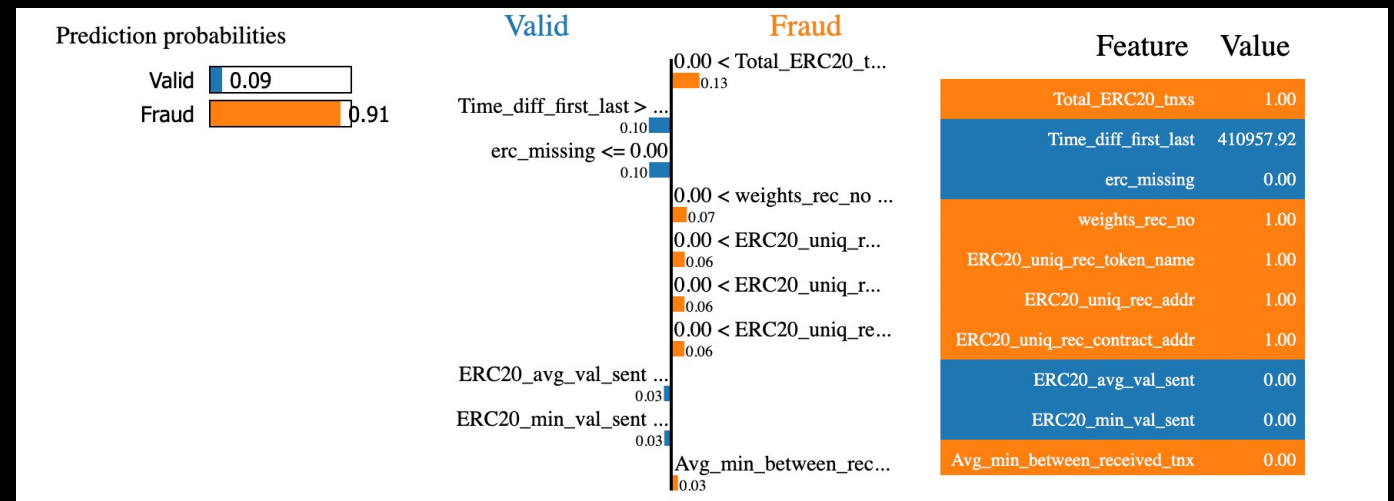
IMPORTANT FEATURES

- The most important features for classifying a transaction as fraudulent or valid are: **Total ERC20 tnx, time difference between first and last transaction, ERC20 unique received addresses, and ERC20 unique received token names**
- Greater number of transactions increases likelihood fraud
- Greater time difference between first and last transaction increases likelihood of fraud
- If a user uses ERC20 tokens they are less likely to be defrauded



MODEL DECISION MAKING PROCESS

- Random Forest Model Model
 - Each observation is processed by many different types of Decision Tree and classified as Fraud or Valid transaction
 - On the right, these are the various subcategories which the Decision Tree uses to classify the transaction type
 - The model builds 80 of these trees using different feature types so it will be able to generalize on new data
 - It then votes Fraud or Valid based on the majority of votes by the trees



WALLET IMPLEMENTATION

- Customers put faith into their wallet providers that their coins will be kept safe
- Wallets are susceptible to risk when they are connected to exchanges or information on keys get leaked
- Value add would be to integrate this model with a wallet provider so that customers can be notified when they are making possible fraudulent transactions
 - Every fraudulent transactions costs \$756 on average





THANK YOU FOR LISTENING

Are there any questions?