Vanguard

# ML-Driven New Account Fraud Early Detection System for Vanguard Australia

Will Li, Senior Advisor, Vanguard

Jose Martins, Senior Fraud Specialist, Vanguard

January, 2022

# Agenda

- Business Requirements
- Technical Solution
- Business Outcome
- Fraud Transformation Vision

# Vanguard Overview

- 1975: Vanguard commences operations May 1, 1975
- $8.0T: Total assets under management
- 30M+: More than 30 million investors
- 417: Number of funds offered worldwide
- 17,300: Approximate number of crew (employees) worldwide

Vanguard is one of the world's most respected investment management companies, offering a broad selection of investments, advice, retirement services, and insights to individual investors, institutions, and financial professionals. We operate under a unique, investor-owned structure* and adhere to a simple purpose: To take a stand for all investors, to treat them fairly, and to give them the best chance for investment success. Shareholders in Vanguard funds benefit from our client focus, experience, stability, and long-term, disciplined investment approach.

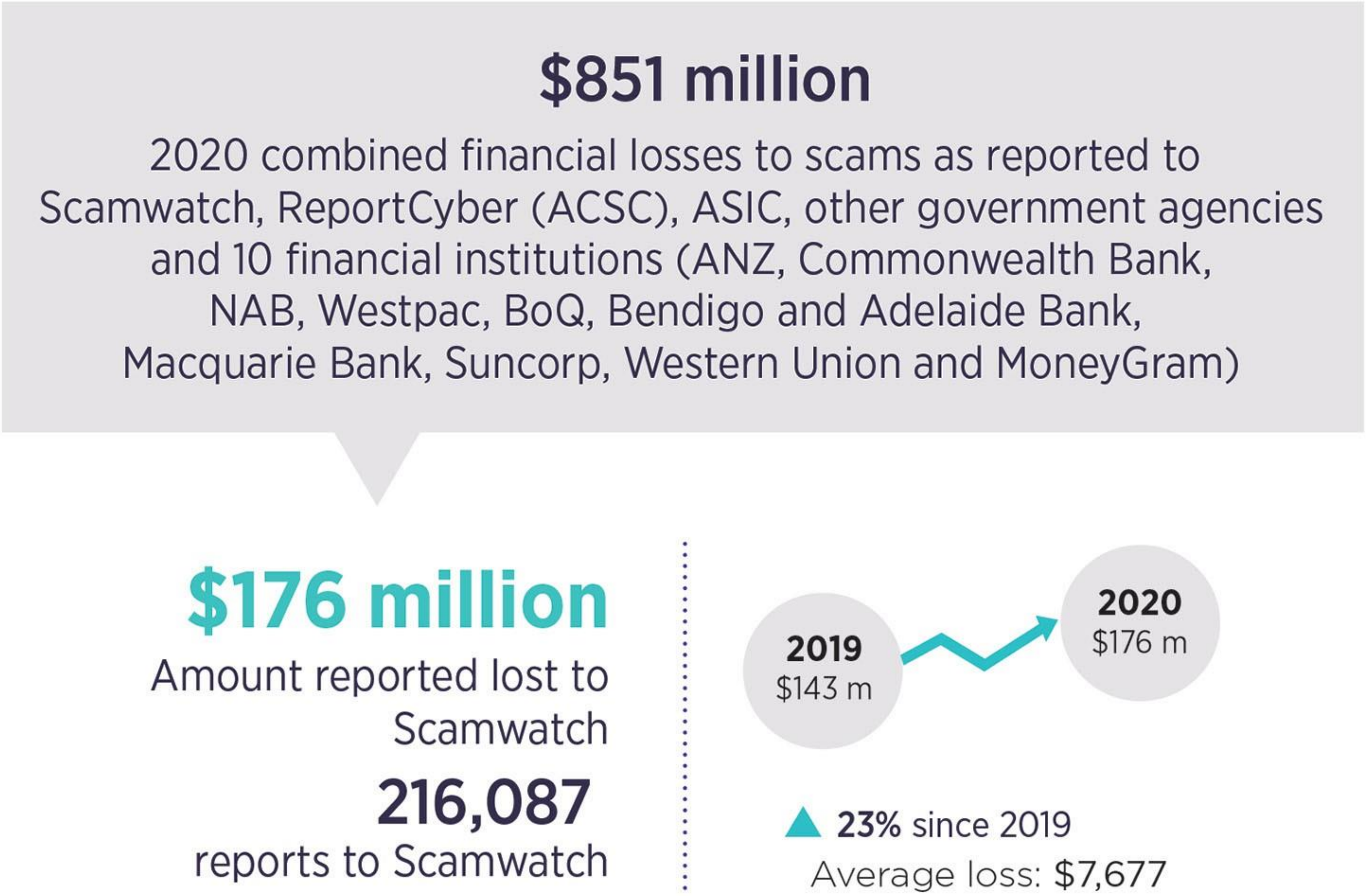
* As of September 30th, 2021

https://corporate.vanguard.com/content/corporatesite/us/en/corp/who-we-are/sets-us-apart/facts-and-figures.html

# “Last year, scam victims reported the biggest losses we have seen” - ACCC

## Losses Reported by ACCC

### (Australian Competition and Consumer Commission)



**$851 million**

2020 combined financial losses to scams as reported to Scamwatch, ReportCyber (ACSC), ASIC, other government agencies and 10 financial institutions (ANZ, Commonwealth Bank, NAB, Westpac, BoQ, Bendigo and Adelaide Bank, Macquarie Bank, Suncorp, Western Union and MoneyGram)

**$176 million**

Amount reported lost to Scamwatch

**216,087**

reports to Scamwatch

**2019** $143 m

**2020** $176 m

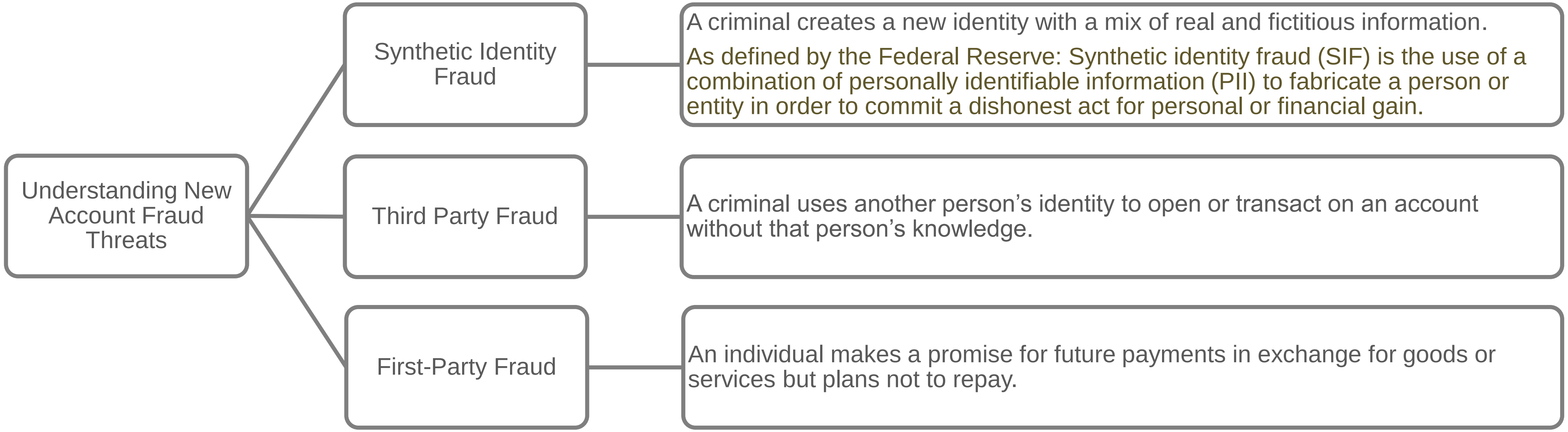
▲ **23%** since 2019

Average loss: $7,677

https://www.scamwatch.gov.au/news-alerts/scammers-capitalise-on-pandemic-as-australians-lose-record-851-million-to-scams

# Business Problem

- The greatest challenge faced by online financial products is the increase in identity theft
  - One in four Australians have been a victim of personal identifiable information misuse of some sorts as reported by the Australian Institute of Criminology (AIC)
- Rule-based fraud detection systems are becoming less effective against rising fraud tide
- We need more effective tools to detect new account fraud activities that usually follow identify theft incidences or synthetic identity creation and use



Understanding New Account Fraud Threats

Synthetic Identity Fraud

A criminal creates a new identity with a mix of real and fictitious information.

As defined by the Federal Reserve: Synthetic identity fraud (SIF) is the use of a combination of personally identifiable information (PII) to fabricate a person or entity in order to commit a dishonest act for personal or financial gain.

Third Party Fraud

A criminal uses another person’s identity to open or transact on an account without that person’s knowledge.

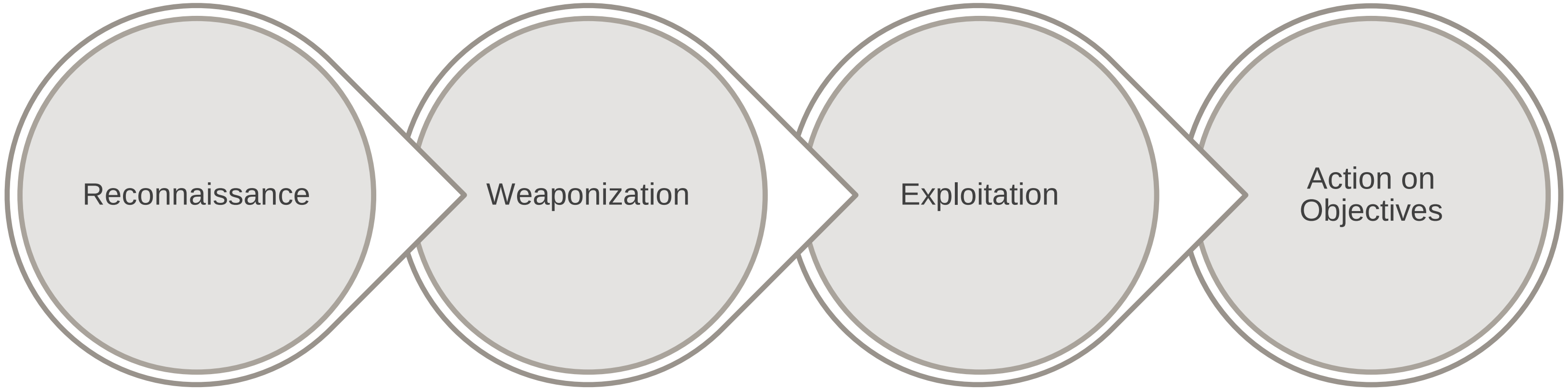
First-Party Fraud

An individual makes a promise for future payments in exchange for goods or services but plans not to repay.

https://fedpaymentsimprovement.org/strategic-initiatives/payments-security/synthetic-identity-payments-fraud/synthetic-identity-fraud-defined/

# Project Technical Requirements

- Need to reduce False Positives to remove client relationship barriers
- Need to reduce False Negatives to improve fraud detection
- Prioritized loss prevention ahead of recovery by striking early at the Fraud Kill Chain
  - Model development should not rely on account withdrawal data



- Bank accounts
- Credentials
- PII

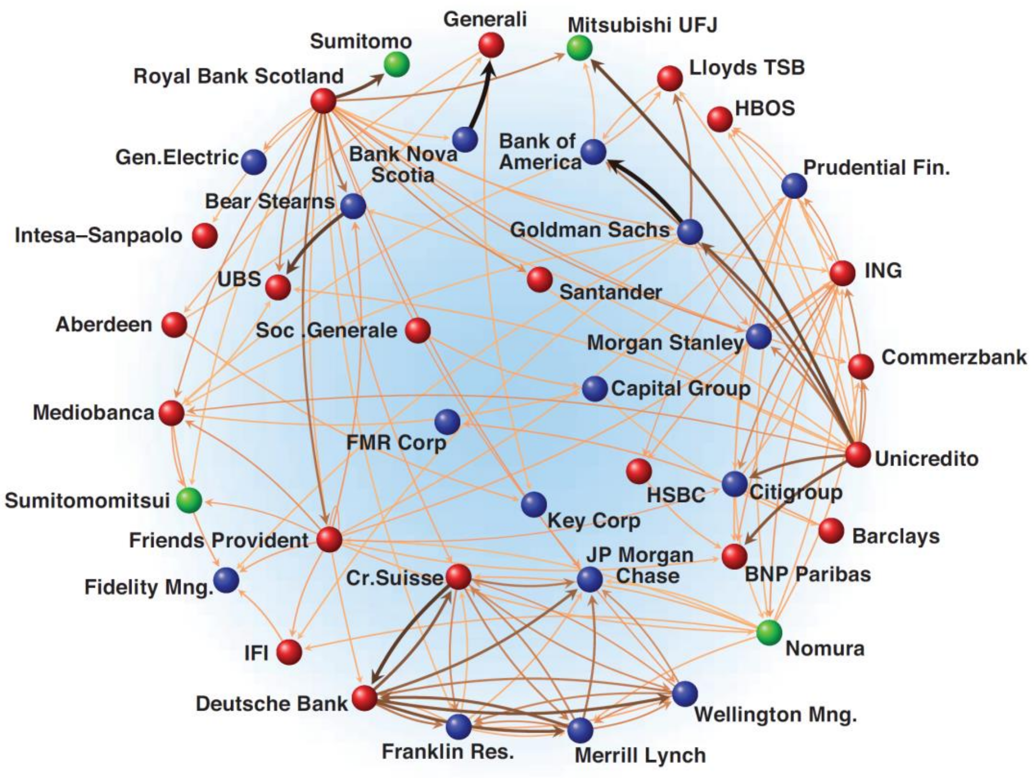
- Aggregation of data
- Augmentation and enrichment
- Distribution

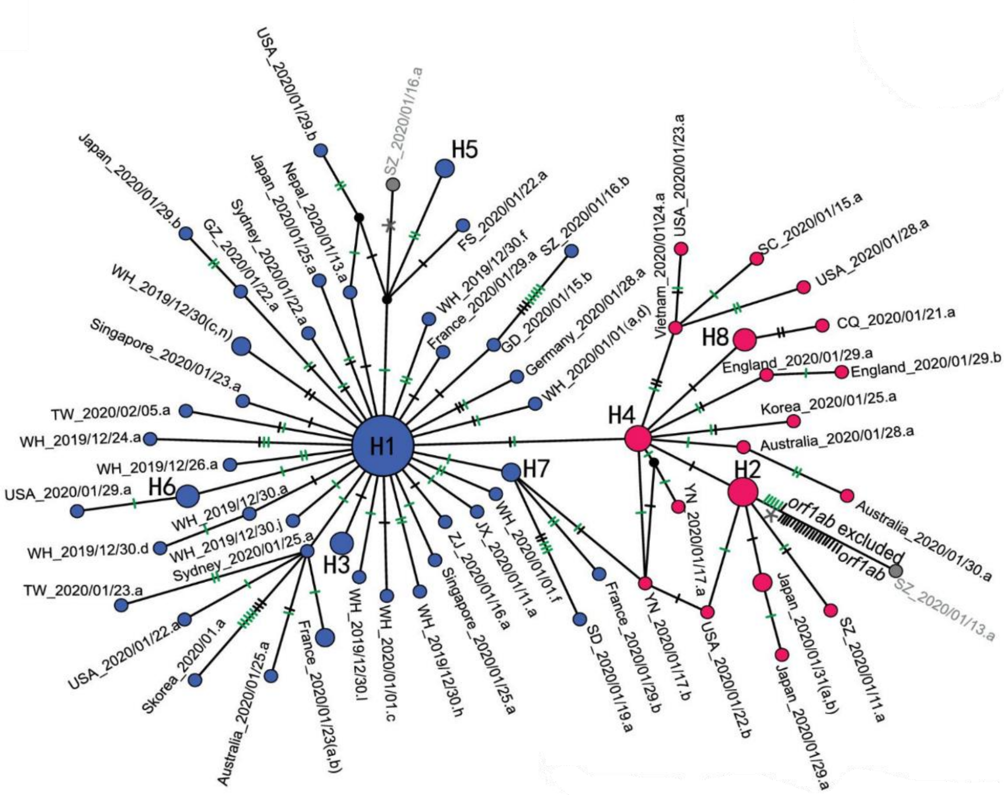
- Secure drop accounts
- Compromise victim’s account
- Make fraudulent payments

- Transfer stolen funds
- Cash out

# Solutions, Part I – Graph Theory

- Challenge: Complex relationships can not be easily detected using tabular data format





**World Financial Network**

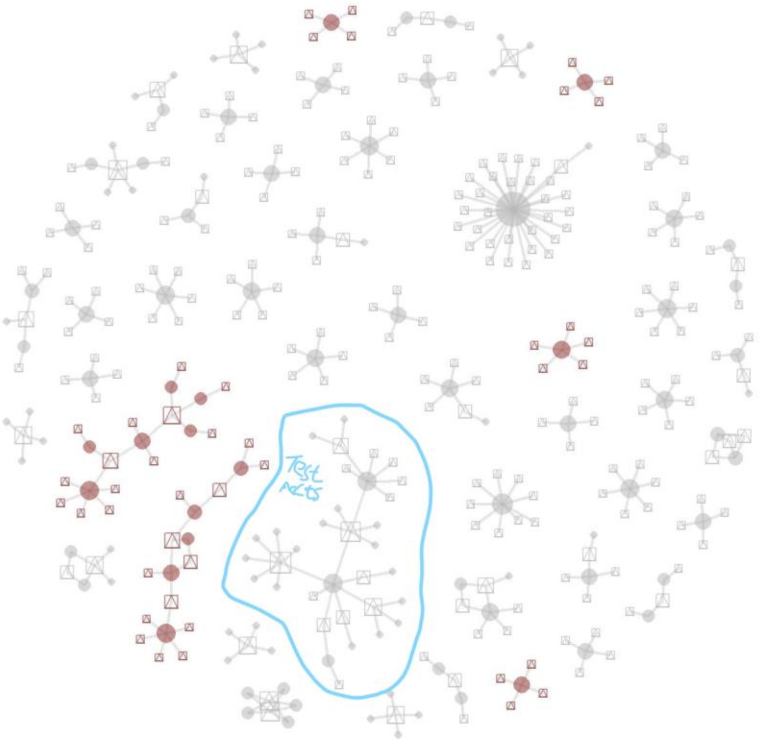
https://www.researchgate.net/publication/26692139

**COVID-19 Virus Mutation Graph**

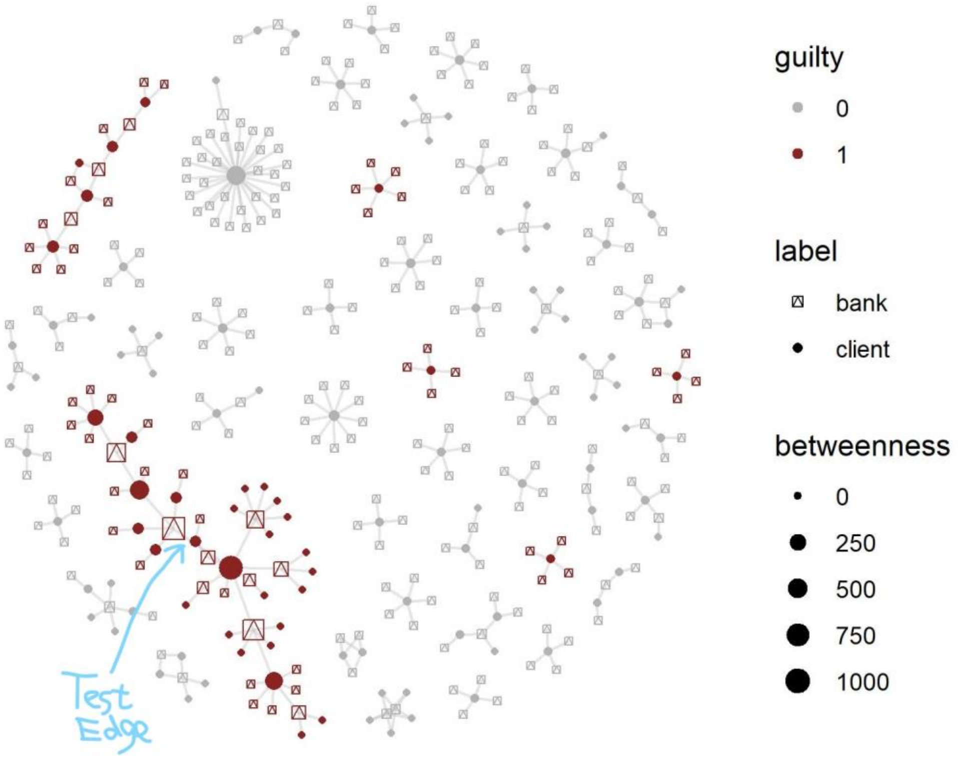
https://academic.oup.com/nsr/article/7/6/1012/5775463

# Solutions, Part I – Graph Theory, Continued

- Key Component: Graph-based Feature Engineering and “Guilt-by Association”



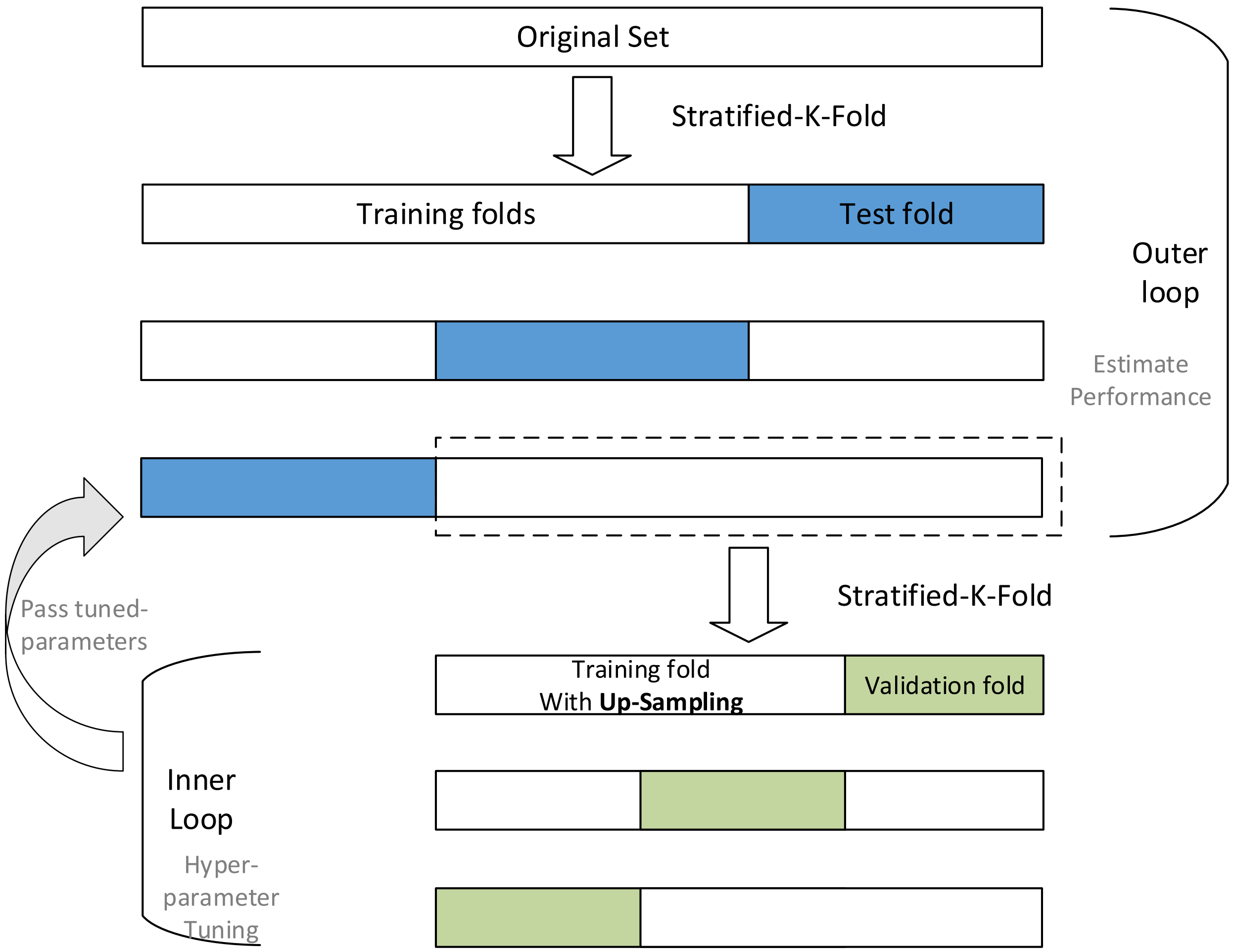
Client Relationship Views



“Guilt-by-Association” Logic

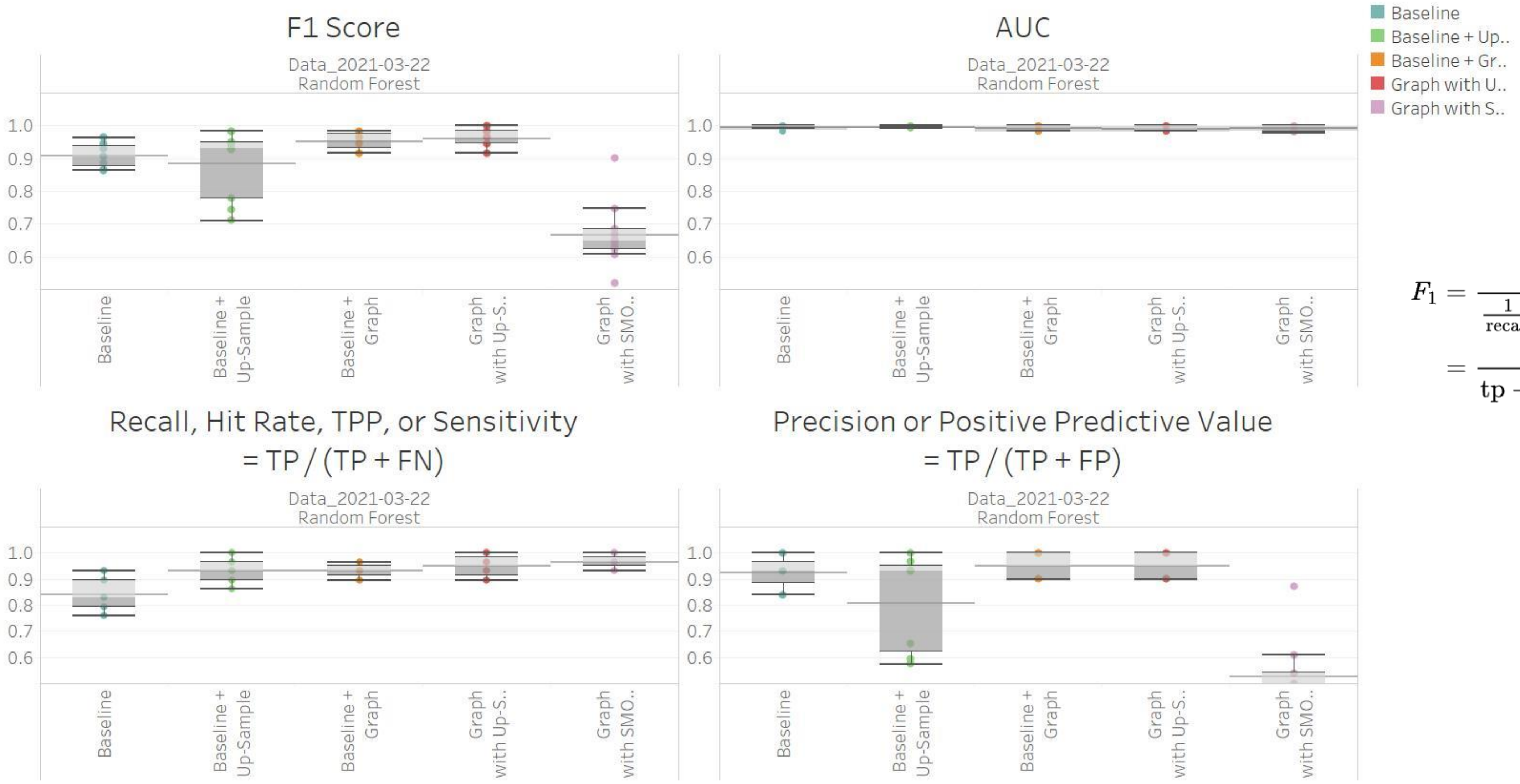
# Solutions, Part II – Machine Learning Training Techniques

- Challenge: Highly imbalanced data (1000:1) and very few positive data points
- Key Component: “Repeated Nested-Cross Validation” and Up-Sampling



# Key Result: Lift in ML Detection Efficacy from Graph and Up-Sampling

- Challenge: We have extreme data imbalance, accuracy = 99.9% if we choose to do nothing



F1 Score

AUC

Recall, Hit Rate, TPP, or Sensitivity
= TP / (TP + FN)

Precision or Positive Predictive Value
= TP / (TP + FP)

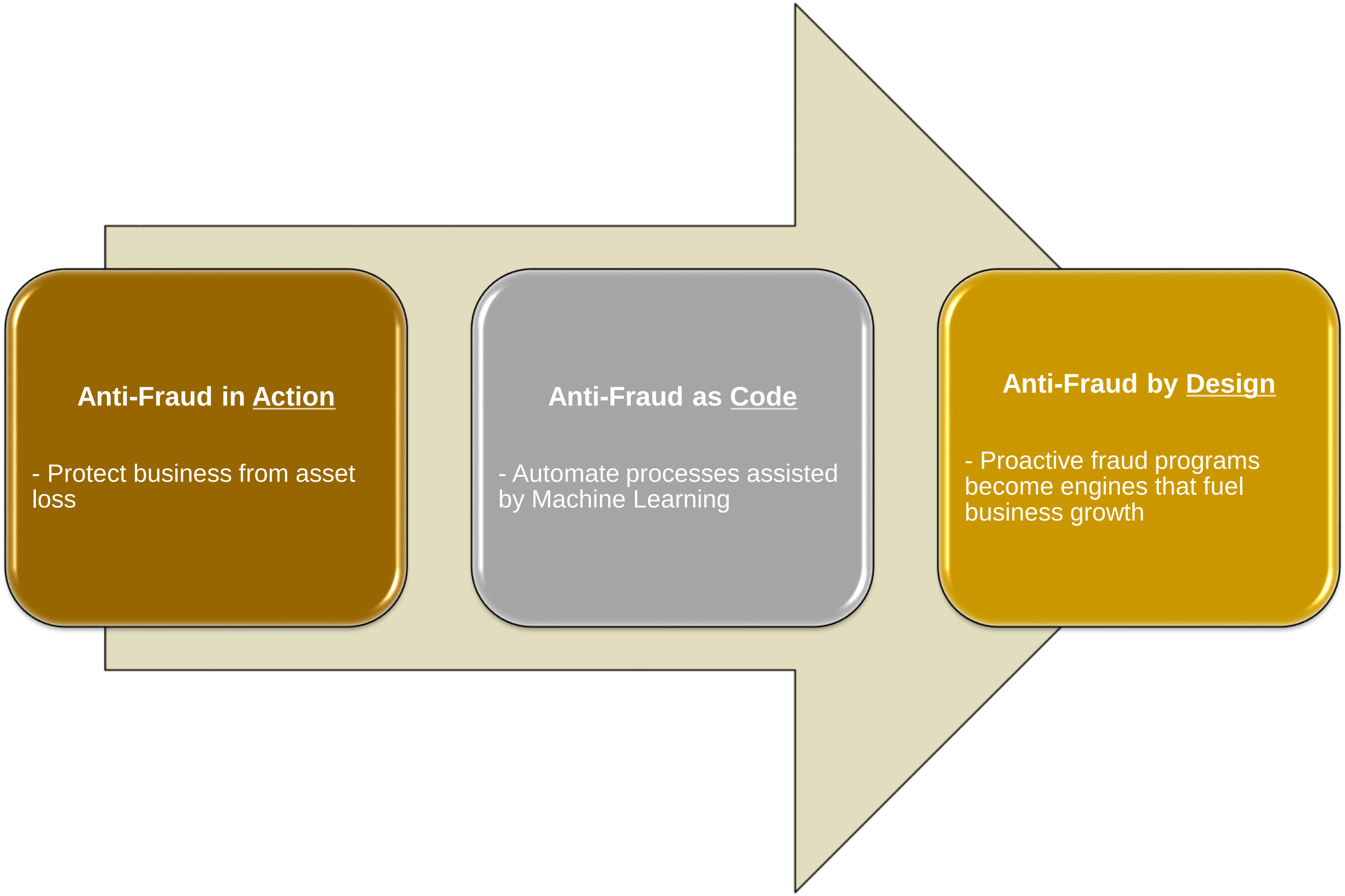
$$F_1 = \frac{2}{\frac{1}{\mathrm{recall}} \times \frac{1}{\mathrm{precision}}} = \frac{\mathrm{tp}}{\mathrm{tp} + \frac{1}{2}(\mathrm{fp} + \mathrm{fn})}$$

Vanguard

# Conclusion - Model Result and Business Outcome

- The model detected more fraud case with value exceeding $5,000,000 so far in 2021 before fraudsters were able to transfer fund out
  - One case stopped more than $3 Million potential loss before the withdrawal process
- We expect to reduce false positives and increase model precision rate overtime
- We created an ingestion and prediction pipeline, saved the model and implemented it in our custom-made alerting software

# Next Steps: Fraud Transformation Vision



**Anti-Fraud in Action**

- Protect business from asset loss

**Anti-Fraud as Code**

- Automate processes assisted by Machine Learning

**Anti-Fraud by Design**

- Proactive fraud programs become engines that fuel business growth

Vanguard®