Intelligent Financial Fraud Detection

Dmitry Petukhov, $\langle \Omega, \mathfrak{U}, \mathbb{P} \rangle$

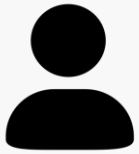
Machine Learning Consultant, Cloud Solutions Architect,
Microsoft Most Valuable Professional in AI && Coffee Addicted

#AI #Fraud #FinTech

Evolution

Hacking Everything

Fat Cats (Customer)

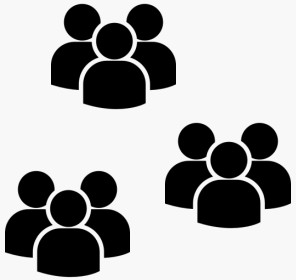


Loss: ~$1 per user
1 attack – 1 user
Hacker profit: $1,000

Thousands of customers

Loss: ~$1 per user
1 attack – >10000 users
Hacker profit: $10,000

Companies



Loss: ~$100,000 per company
1 attack – 1 company
Hacker profit: $100,000

Banks

Loss: ~$1M per bank
1 attack – 1 bank
Hacker profit: ~$1M

Central Banks

Max loss: **$1b per CB**
1 attack – 1 CB
Hacker profit: **$81M**

Hacking Everything

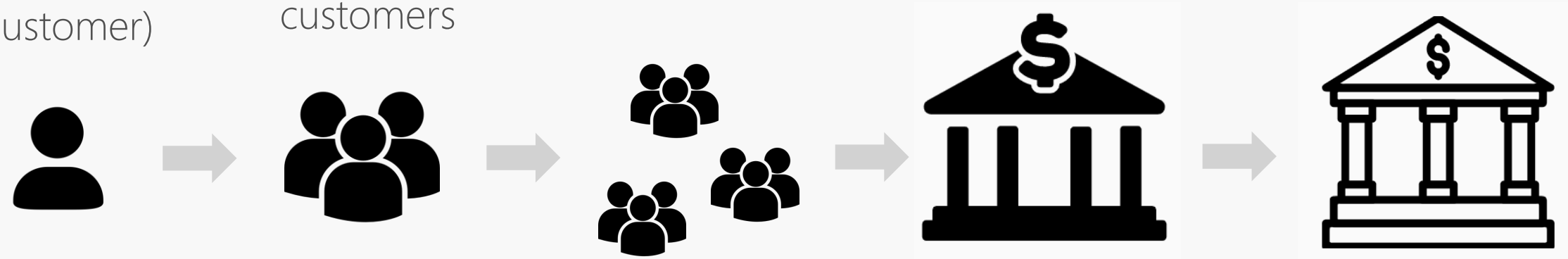
Fat Cats (Customer)

Thousands of customers

Companies

Banks

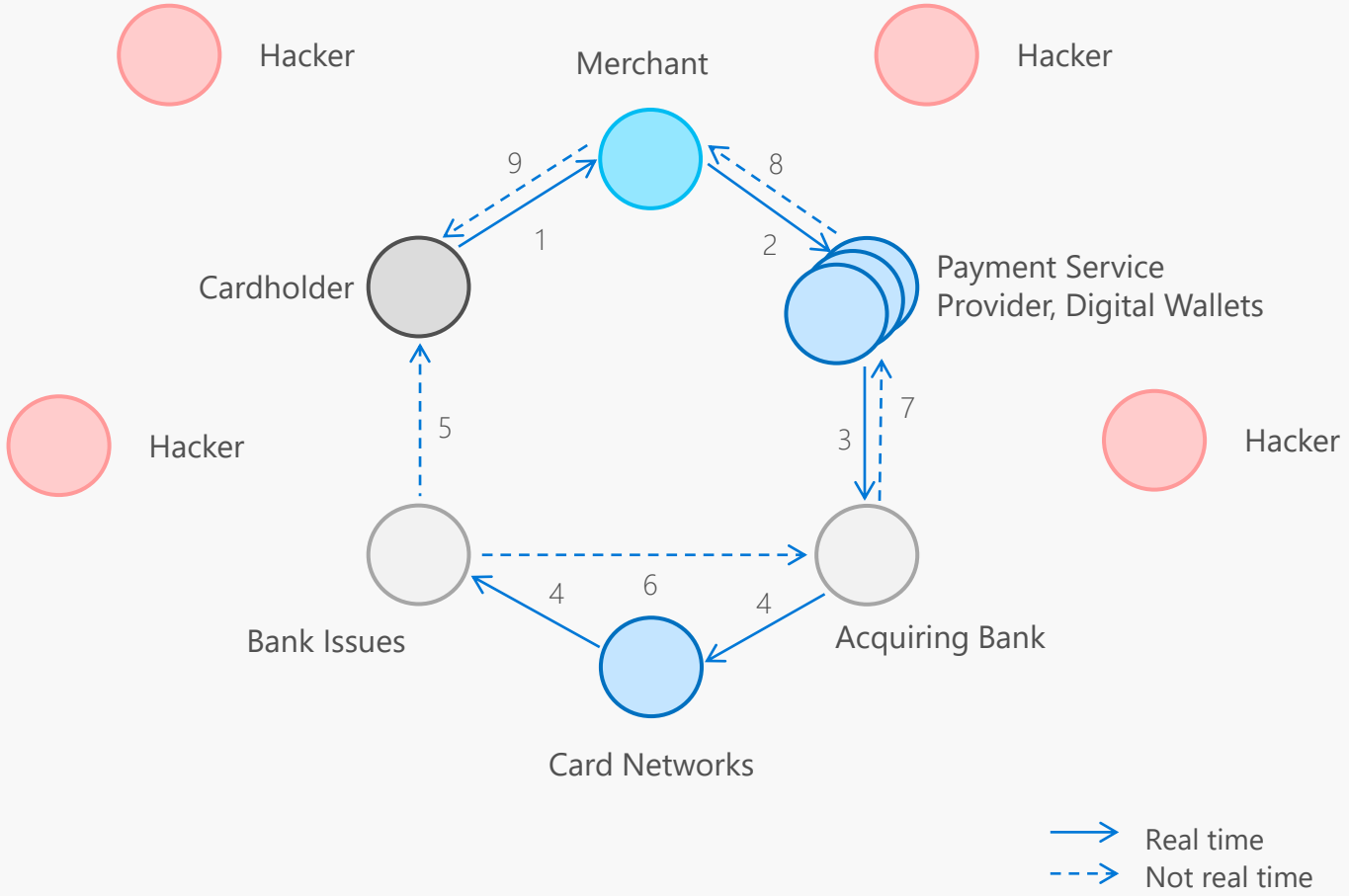
Central Banks



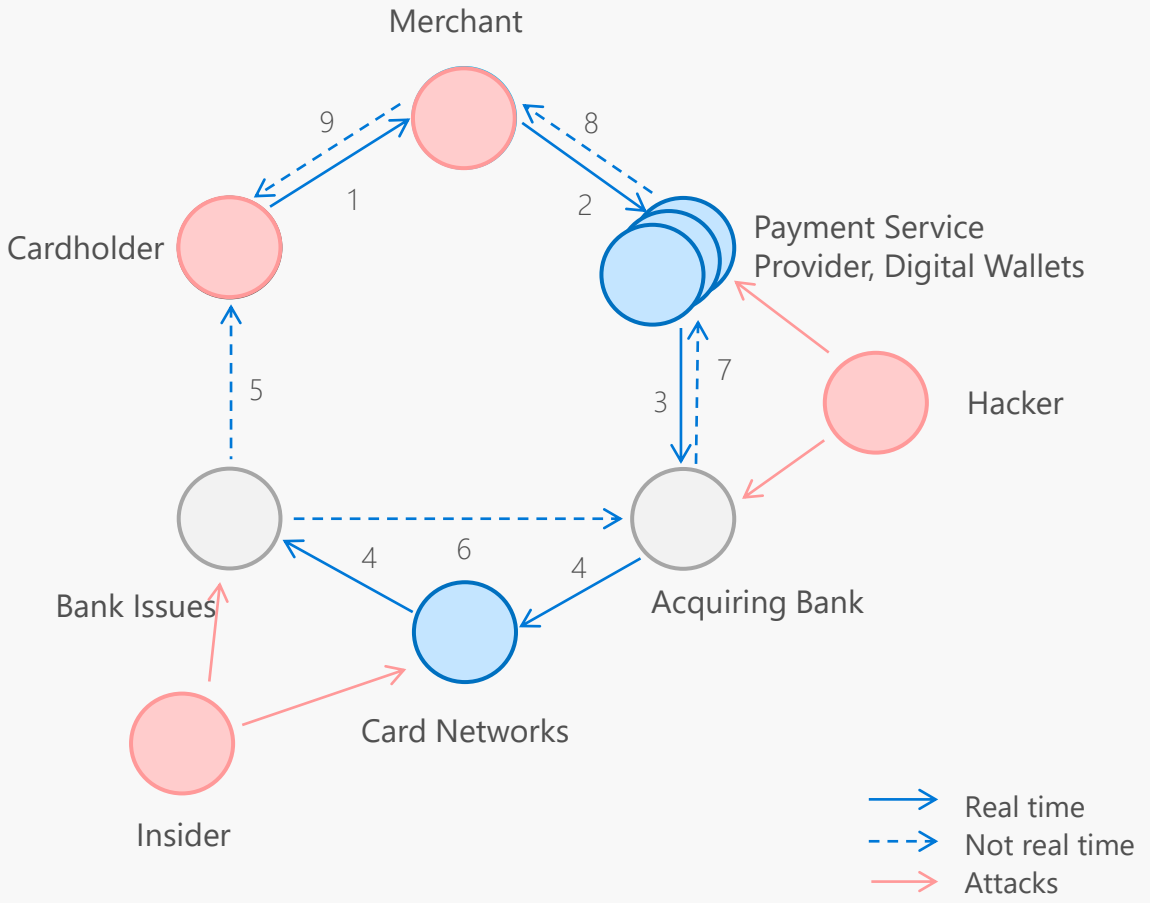
Trojan Backdoor Keylogger Spam
DDoS Spoofing Backdoor Device cloning
Spyware Cryptojacking Scam call Exploits
Malicious apps Identity theft
Proxy/VPN/TOR Social engineering Pharming
Ryuk ransomware Trickbot

Challenges

Challenge #1: Too Many Participants



Challenge #1: Too Many Participants



Challenge #2: Too Many Interaction Points

Payments Security

- Credit card fraud detection:
  - Card-present antifraud
  - Card-not-present antifraud
- P2P (Card-to-Card) antifraud
- Frictionless payments:
  - Payment risk scoring for 3DS 2.0
  - E-comm payments conversion increase
  - Avoiding 'Black Friday' syndrome
- Anti-laundering:
  - Detecting money laundering patterns
- Anomaly detection:
  - Anomaly ATM/POS activity detection
  - Anomaly mobile-/web-bank activity detection
- Fraud patterns detection:
  - Chargebacks/malicious payment chains prevention
- SWIFT messaging fraud prevention.

Oversight and Insight

- Insider detection
- Information attacks detection:
  - Including via social media

Identification 2.0

- Biometric:
  - Voice recognition
  - Face recognition
  - Keyboard handwriting + accelerometer data.

Source: github.com/codez0mb1e/**FinArt.AI**

Challenge #3: **Data Quality**

Data

Big Data volume

…but *not enough marked* data

Data is incompleteness and noisy

Outliers

Number of levels for categorical features

Highly imbalanced classes

Payments are either white (legitimate), either black (fraud) – FALSE

Process evolution and Data degeneration

Disproportional misclassification cost

Data is sensitive

Ethics and law (GDPR)

Anti-Fraud Models

Anti-Fraud: Main Approaches

Rule-based anti-fraud:

- Black/white lists
- Statistics-based approach.

ML-based anti-fraud:

- Fraud prediction using Supervised Learning
- Anomaly detection using Unsupervised Learning
- Semi-supervised Learning.

Fraud Prediction using Supervised Learning

Classic methods

Logistic Regression

SVM

Trees methods

Random Forest

Extremely Randomized Trees

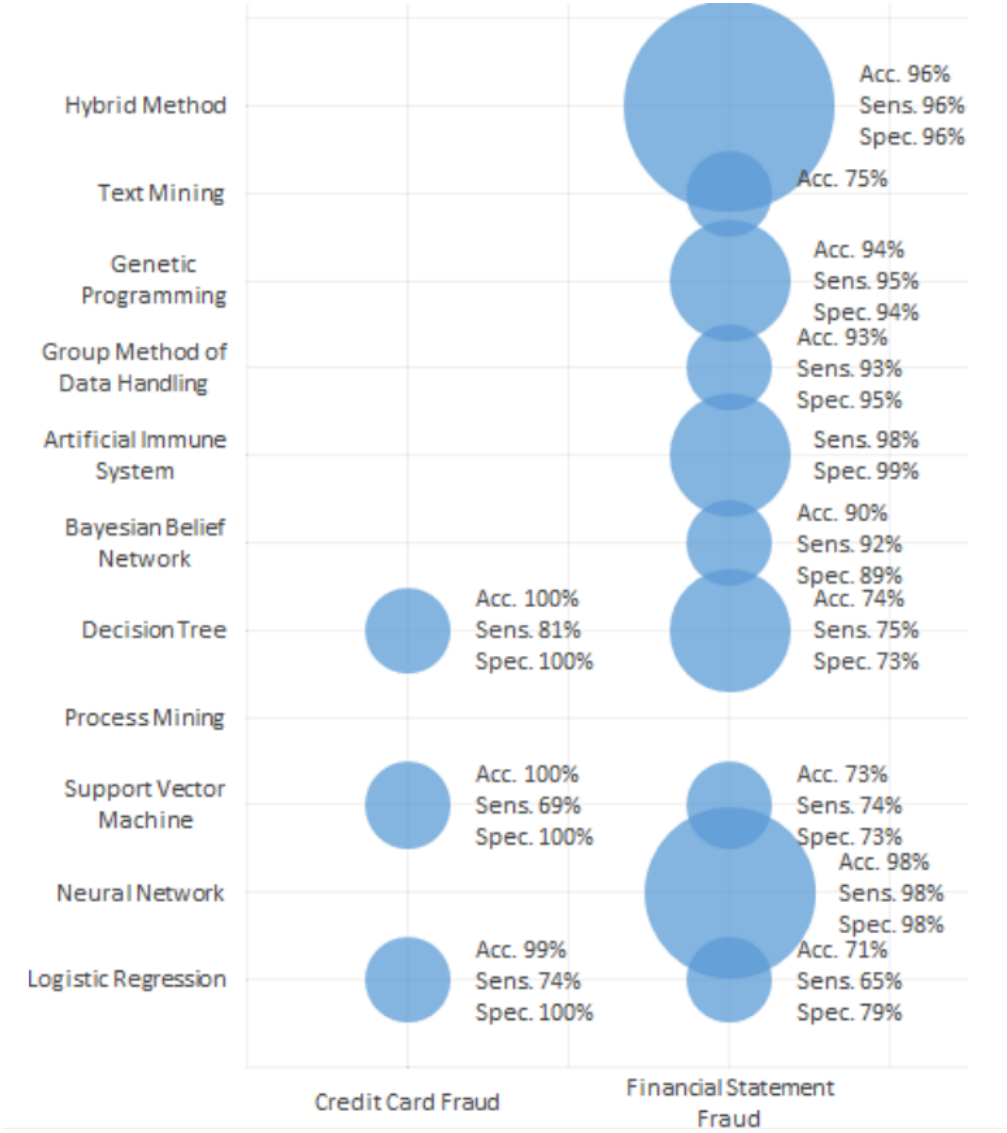
Gradient Boosted Trees

Neural Networks methods

Feedforward Neural Network

Hybrid approaches

~~Ensembles~~ (do not use it in production ☺)



Source: arxiv.org/ftp/arxiv/papers/1510/1510.07165.pdf

Anomaly Detection using Unsupervised Learning

Clustering methods

K-means

Hierarchical clustering

Anomaly detection methods

One-class SVM

Isolation Forest

Local Outlier Factor

Neural Networks methods

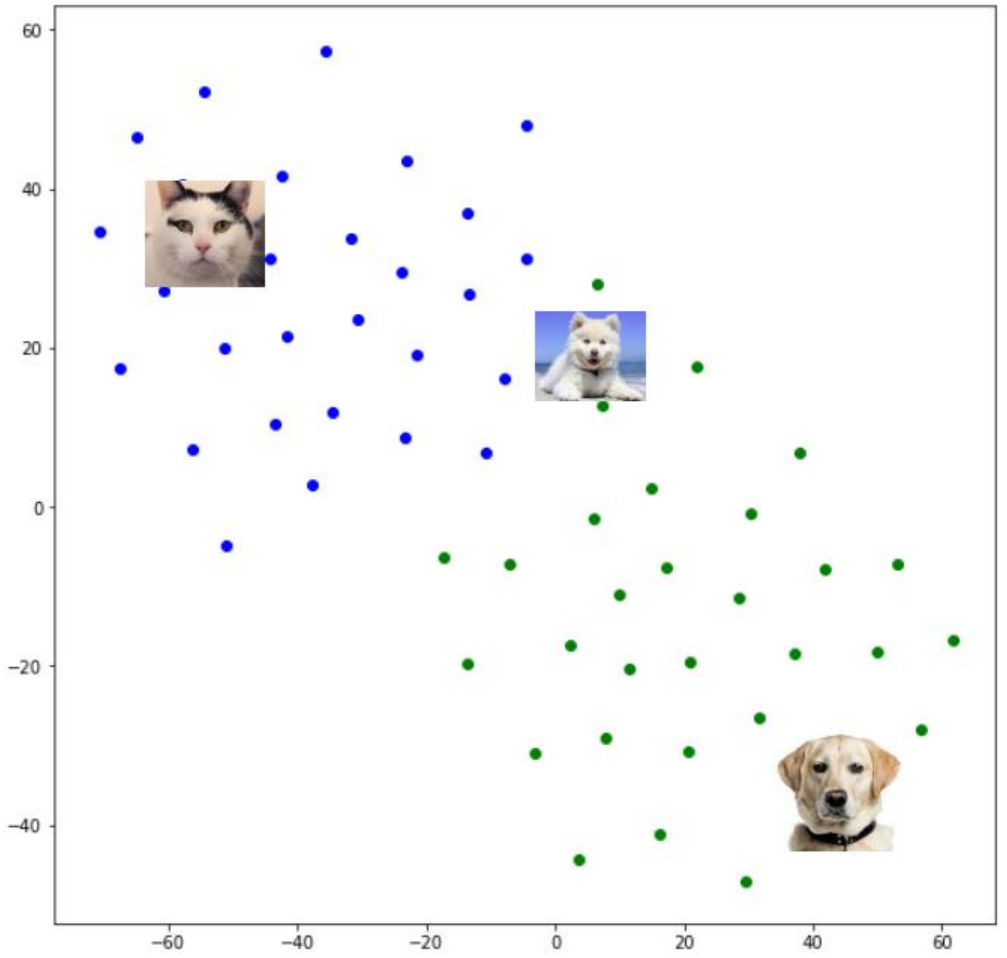
Autoencoder

Variational Autoencoder

Deep Belief Network

Generative Adversarial Networks

Self-Organized Maps



Source: Image clustering using Transfer learning

Anomaly Detection using Unsupervised Learning

Clustering methods

K-means

Hierarchical clustering

Anomaly detection methods

One-class SVM

Isolation Forest

Local Outlier Factor

Neural Networks methods

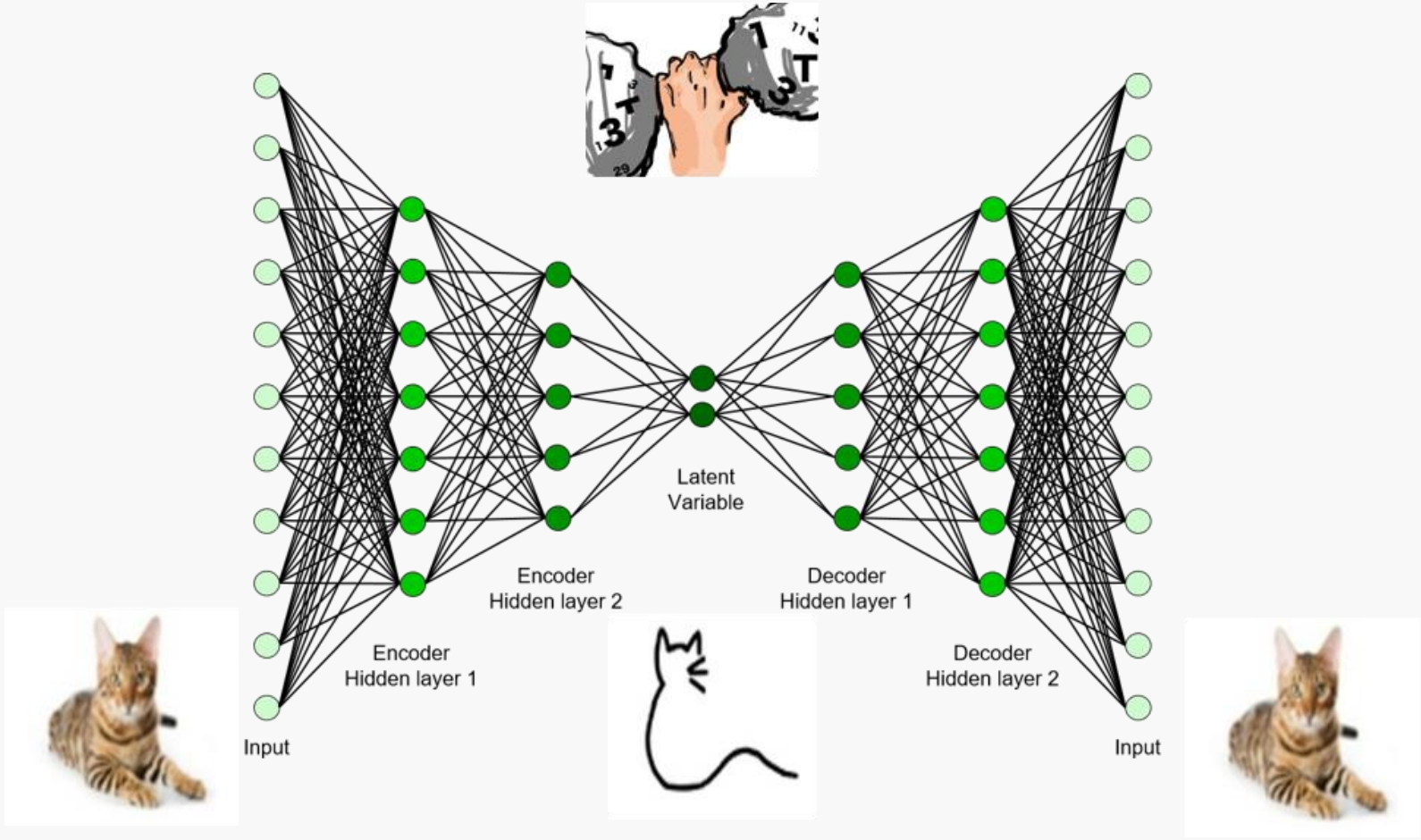
Autoencoder

Variational Autoencoder

Deep Belief Network

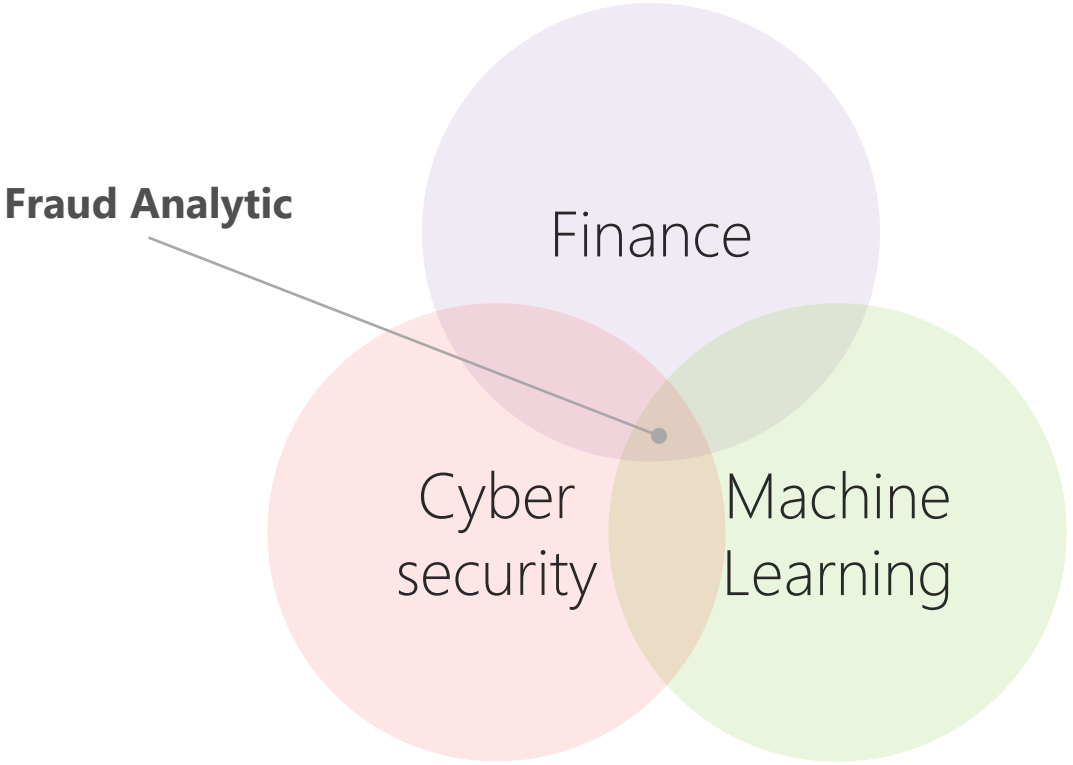
Generative Adversarial Networks

Self-Organized Maps



Source: Autoencoder by Prof. Seungchul Lee

Conclusion



Thank you!

Q&A

Now or later (see contacts below)

Stay connected

Facebook: @codez0mb1e

Telegram: @codez0mb1e

All contacts: http://0xCode.in/@codez0mb1e

GitHub repo: github.com/codez0mb1e/FinArt.AI

Download slides from