FICO Machine Learning AML Solution

Scott Zoldi
Chief Analytics Officer, FICO
@ScottZoldi
Scott Zoldi, Chief Analytics Officer

- Responsible for analytic development of FICO's product and technology solutions, including Falcon Fraud Manager
- 18 years at FICO
- Author of 79 patents
  - 39 granted and 40 in process
- Recent focus on self learning analytics AI for real-time detection of Cyber Security attacks, AML detection, and mobile device analytics
- Ph.D. in theoretical physics from Duke University
Money Laundering: The process of creating the appearance that illicit funds obtained through illegal activity originated from legitimate sources.
$16M 2004 556X $8.9B 2014
human trafficking, the second most profitable crime after narcotics
Combating Money Laundering Today

- As certain compliance risk
- KYC
- Observe-and-Report
- SARs
- Subjective
- Rule-based
“Increasingly, regulators recognize that rules alone are not an effective manner of detection and are pressuring banks to include more sophisticated analytics.”

Aite Group LLC, “Global AML Vendor Evaluation” 2015
Think of transaction behavior and events as words from a vocabulary

Current Account
Amounts
Wire Transfer
Country
Access Channel

Example word:
“$500-$750_Wire_FRA”

Word

| “$500-$750_Wire_FRA” |
| “$100-250_EFT” |
| “$250-$500_EFT” |
| “$500-$750_Wire_ITA” |

The stream of behavior is seen as the sequence of words
Learning archetypes from transactions: Collaborative Profiling

Customer’s data stream:

From many other customers

- Bayesian Learning
  - Unsupervised
  - Learn archetypes from millions of customers.

Learned Archetypes
(≈ 10’s)
Clustering archetypes: Misalignment with clusters is suspicious

Scenario: Existing customer moves out of cluster:
- Sleeper account being activated
Real-World AML Example: SAR distribution in archetype space

- Many SARs are outliers from normal customers along certain archetypes
Autoencoders for unsupervised anomaly scoring

- Autoencoders are deep neural nets trained to represent/compress input by minimizing reconstruction error.

![Diagram of an autoencoder network with three layers: input layer, hidden layer, and output layer. The target is the input, and the network is trained to minimize the difference between input and output through "bottleneck layers." ]
Autoencoders: Reconstruction error measures similarity to training data

- For anomaly scoring, this reconstruction error indicates how much a sample differs from the training population.
Real-world AML application: Autoencoder finds outlier in archetype space

- Autoencoder trained on Collaborative Profiling archetypes
- High scores when autoencoder finds archetype mixtures very different from training set.
Thank you!

Scott M Zoldi
Chief Analytics Officer, FICO
@ScottZoldi