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.

0.8 – $2T

Source: https://www.unodc.org/unodc/en/money-laundering/globalization.html


$16M
2004
556 X
$8.9B
2014

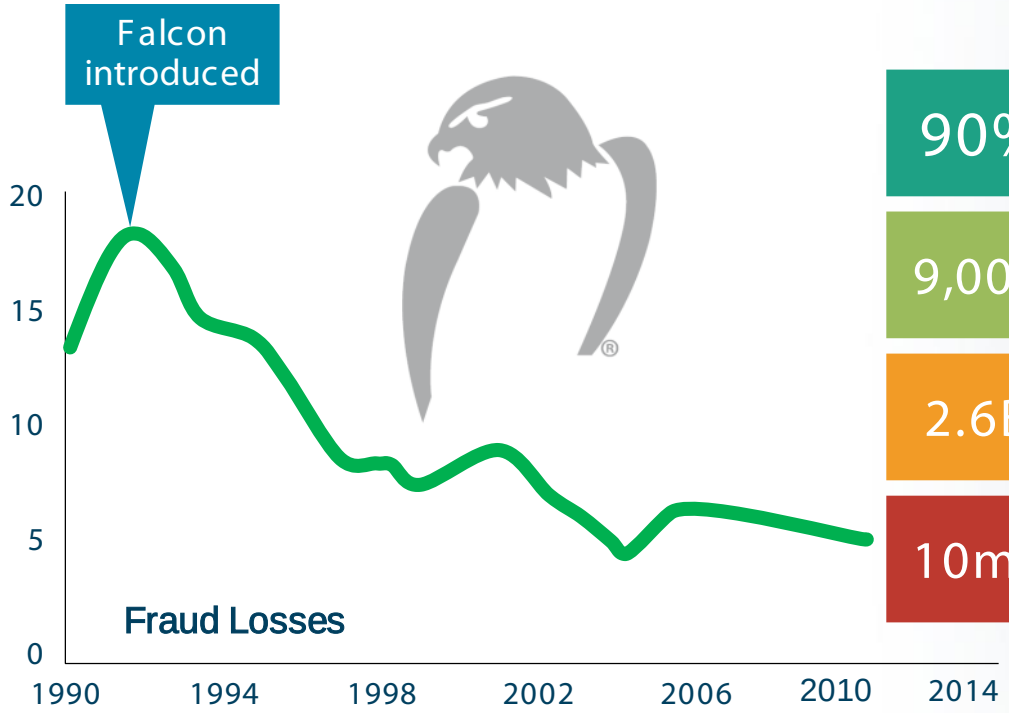
ML IMPACT

narcotics trafficking

terrorist financing

human trafficking, the second most profitable crime after narcotics

FICO® Falcon® Fraud Manager



90%	Percentage of the US payment cards protected by FICO fraud solutions
9,000	Banks participating in FICO's fraud data consortium – driving insight and analytic innovation
2.6B	Active financial accounts protected by FICO worldwide – providing a global perspective
10ms	Average response time for fraud decisions rendered by FICO's low-latency real-time engine

Combating Money Laundering Today

- Ascertain compliance risk
- KYC
- Observe-and-Report
- SARs
- Subjective
- Rule-based

Vocabularies to describe transaction behavior

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

Patents 14/796,547 (USA) ,14/613,300 (USA)

Learning archetypes from transactions: Collaborative Profiling

From many other customers

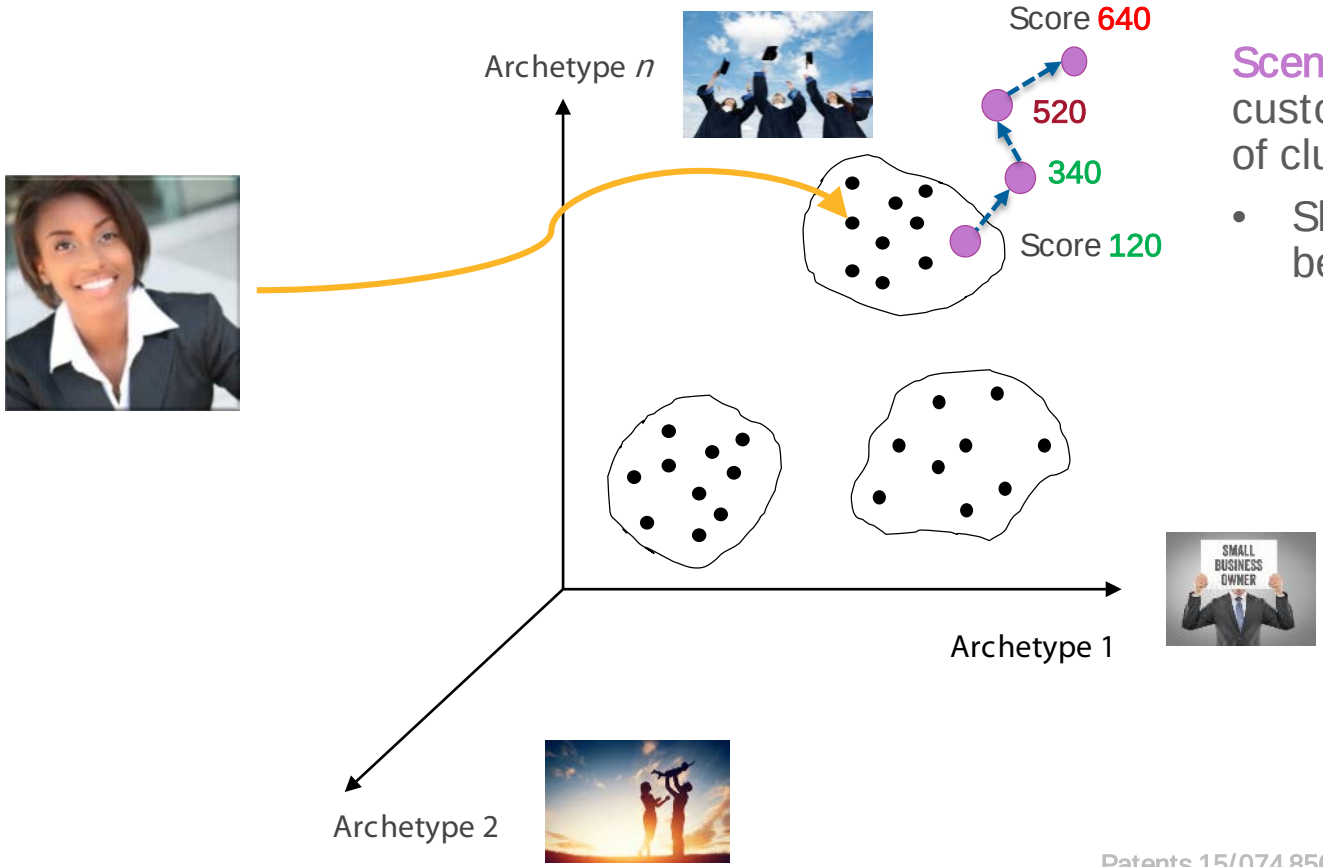
Customer's data stream:

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

Learned Archetypes
(~10's)

Patents 14/796,547 (USA) ,14/613,300 (USA)

Clustering archetypes: Misalignment with clusters is suspicious



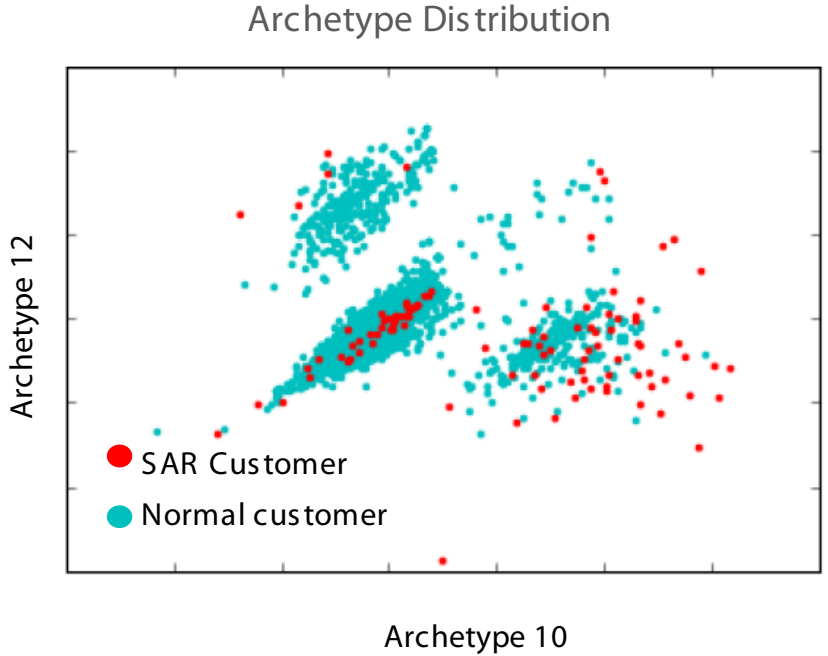
Scenario: Existing customer moves out of cluster:

- Sleeper account being activated

Patents 15/074,856 (USA) ,14/074,977 (USA)

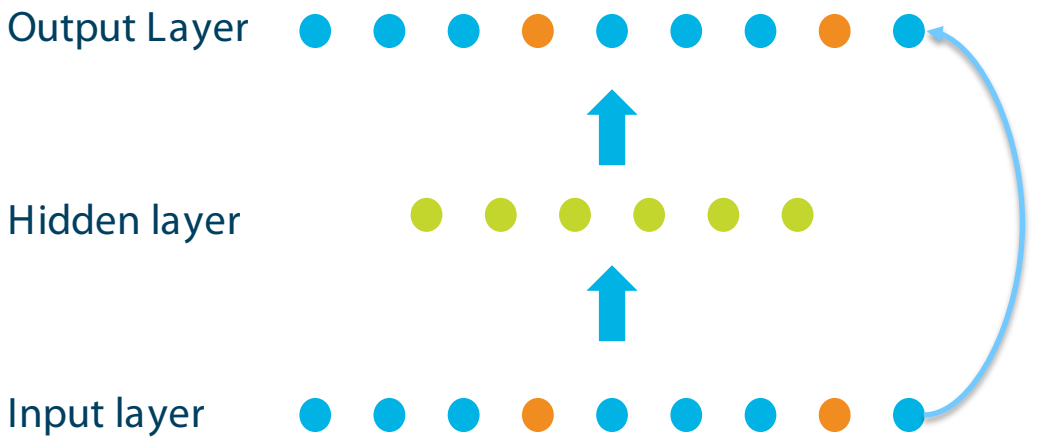
Real-World AML Example: SAR distribution in archetype space

- 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.

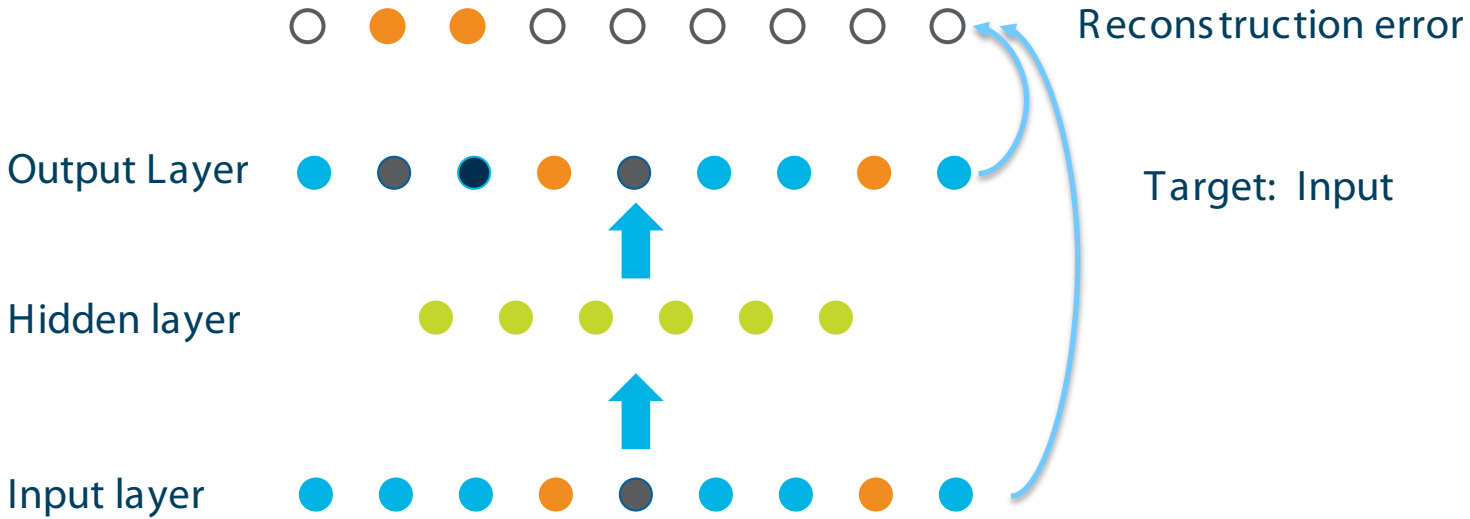


Target: Input

Train to minimize 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.