


EDM Webinar

Breaking All the Rules: Why Finance Is Turning to Machine Learning to Manage Data



A conversation with:

Dan Waldner
Principal Strategist
Traction on Demand

Alex Batchelor
Strategic Sales Engineer
Tamr


Tamr

EDM Council

Moderated by **Mike Meriton**
Co-Founder & COO, EDM Council

- Joined EDM Council full-time 2015 to lead Industry Engagement
- EDM Council Co-Founder & First Chairman (2005-2007)
- EDM Council Finance Board Chair (2007-2015)
- Former CEO GoldenSource (2002-2015)
- Former Executive for D&B Software and Oracle
- FinTech Innovation Lab – Executive Mentor (2011 – Present)

EDMCouncil

Today's panel:
Why finance is turning to ML to manage data

EDMCouncil

Moderator

Mike Meriton
Co-Founder and COO
EDM Council

Dan Waldner
Principal Strategist Traction on Demand

Alex Batchelor
Strategic Sales Engineer
Tamr

/ ABOUT YOUR SPEAKERS - DAN WALDNER

Principal Strategist, Financial Services
Traction on Demand

- 26 years experience in Technology sector and 15 years experience in Financial Services
- Global Salesforce owner for Scotiabank across Wealth, Commercial, and GBM (5000+ users)
- Broad experience across many diverse technologies (AI, ML, Blockchain)

“Any sufficiently advanced technology is indistinguishable from magic.”

- Arthur C. Clarke

/ DATA CONTINUES TO GROW EXPONENTIALLY

THE VASTNESS OF CREATION - AND WASTE - CANNOT BE UNDERSTATED

Every year, Google fields around 2 *trillion* searches.

Every day, Facebook receives 100 *terabytes* of data.

Every minute, Youtube receives 300 *hours* of new video.

In 2021, we will create 40 zettabytes of new data - 10x more than the entirety of human history

99.5% of all data collected never gets used or analyzed

Source: Forbes

/ AND IN ALL DIMENSIONS

SIZE OF DATA IS NOT LIMITED TO ROWCOUNT

In most Financial Services organizations:

- Growth of number of applications is slowing (but still growing)
- Growth of application/platform consolidation activities increasing
- Number of total database tables increasing
- Number of total relationships increasing
- Number of total columns increasing

While we seek simplicity from an architecture perspective, data concepts are often unaffected

Source: McKinsey

A HYPOTHETICAL OLD PROBLEM THAT PERSISTS

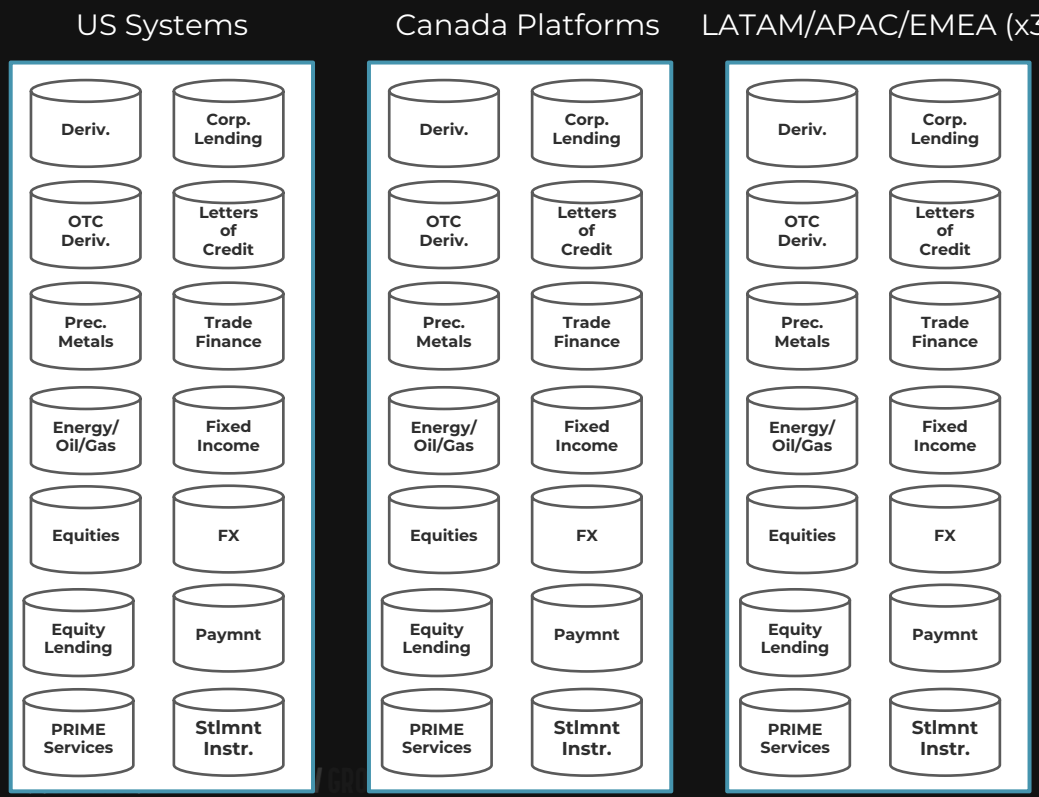
CREATING A MASTER CUSTOMER REPOSITORY

Large Multinational Bank (LMB) has a problem with their institutional customer data.

General Situation:

- A multitude of systems to manage different aspects of their ecosystem
- Swivel chair interface for many of them; few automated integrations
- No consistency between data points
- No supremacy between systems
- No guarantee of static schemas

WELCOME TO THE THUNDERDOME



- Over 100 large systems, due to:
  - Multiple data sovereignty issues in sensitive markets (APAC/LATAM) requiring separate instances
  - Inorganic growth inherits systems in different jurisdictions, making integration and consolidation a long-term project
  - A large footprint of countries that LMB does business in.
- Widely varying degrees of data quality – highly heterogeneous data.
- Large amount of data – measured in petabytes
- Numerous aggregation and enrichment systems to attempt a piecemeal solution.

- **If I added all the integration points, you wouldn't be able to see anything useful on the screen.**

T2V SERIES

/ RULES WERE MADE TO BE BROKEN

AT THIS SCALE, RULES JUST WON'T WORK

Consider that you need to:

- Clearly eliminate exceptions (where Legal Entity Name ≠ 'Test Account')
- Express the concept of unique identity (LEI, AVID, DUNS, Name & Address, etc.)
- Relate this concept to each set (System A -> System B, System A -> System C, etc.)
- Account for missing data points (System A has DUNS, System C does not)
- Account for inconsistent data (System A has old data, System C has new data)

/ DEATH BY 1,000 PAPERCUTS

REMEMBER, DATA IS GROWING AND CHANGING

Once you hit ~20 systems:

- The number of rules grows to wildly unmanageable levels
- Data schemas are constantly in flux
- Temporality of data has a way of breaking the system
- Missing data bends the system
- Inconsistent data breaks it

The system becomes unwieldy, brittle, and ineffective.

If your data is growing at an exceptional rate, a static approach to mastery will not work effectively for very long.

A BETTER WAY

TAKING A MACHINE-LEARNING, PROBABILISTIC APPROACH

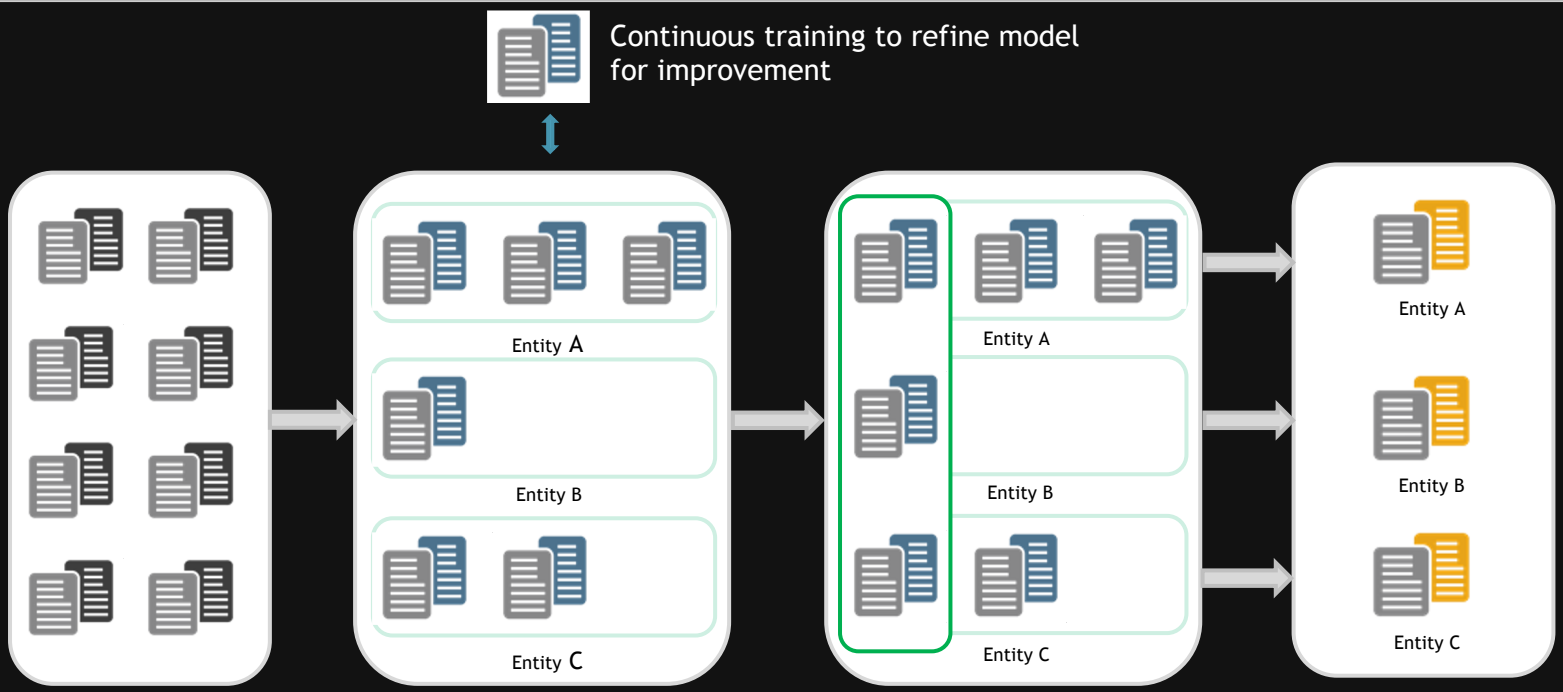
- Matching is driven by statistical models that take the entire dataset into consideration
- Model driven by ML and incorporates human training to learn over time
- Training adjust the correlation between datasets and attributes
- Produces a confidence level that determines if a match has occurred
- The output can be manually adjusted – data is complicated and inconsistent at times

/ GETTING STRONGER THROUGH DATA

THE BIGGER THE DATA, THE MORE ACCURATE THE RESULTS

- The statistical models get more refined the more data is supplied to it = better over time
- Temporality of the data gets accounted for
- Missing data is corrected in most instances
- Inconsistent data gets corrected through democratized data

A VISUALIZATION OF THE PROCESS



/ PROOF IS IN THE PUDDING

RESULTS OF THIS APPROACH OVER FIRST SIX MONTHS

35 separate systems in-scope for initial phase of project (heterogenous in every way)

3,700,000 rows ingested into the Entity Resolution tool (each row is a customer)

325,000 clusters coalesced within the system (each cluster is a matched customer)

Time to onboard a new system from landing data to mastery: 5-7 days

Process time through end-to-end for new record: 2 days maximum

Today's panel:
Why finance is turning to ML to manage data

EDMCouncil

Moderator

Mike Meriton
Co-Founder and COO
EDM Council

Dan Waldner
Principal Strategist Traction on Demand

Alex Batchelor
Strategic Sales Engineer
Tamr

Questions?

EDM Webinar









FOR MORE INFORMATION:

Alex Batchelor
Strategic Sales Engineer
Tamr
alex.batchelor@tamr.com

Tamr

EDMCouncil