



MCCA  
GLOBAL TEC  
FORUM

**Rise of the Machines:  
The Next Generation of Data Modeling**

April 12<sup>th</sup> 2:45-4:00



# Agenda

1

- Overview of Data Modeling and Predictive Technology Key Concepts and Definitions

2

- Current State of Predictive Coding (TAR 2.0)

3

- How is Predictive Modeling Being Used Now by Corporations and Law Firms?

4

- What Does the Future Hold for Predictive Modeling Left of the EDRM?

# Panelists



A.B. Cruz III

- Senior Vice President, USAA; Board Chair at MCAA



Paula Schauwecker

- Shareholder, Beveridge & Diamond, P.C.



David Kim

- Partner, Stinson, Leonard Street, LLP



Jeanne Somma

- Director of Analytics and Managed Review, RVM

# Data Modeling Overview 1



## Data Modeling Defined

- Data modeling is the analysis of data objects and their relationships to other data objects.

# Data Modeling Overview 2



## Conceptual and Logical Data Models

**Data modeling:** *The analysis of data objects and their relationships to other data objects.*

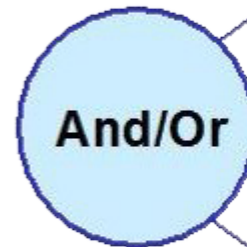
- The requirements are first recorded as a conceptual data model.
- The conceptual model is then translated into a logical data model.
  - Note: Implementation of one conceptual data model may require multiple logical data models.
- The last step is transforming the logical data model into a physical data model that organizes the information into tables and accounts for access, performance, and storage details.

# Data Modeling Overview 3

## Physical Data Models

- A complete physical data model will include all the database artifacts required to create relationships between tables or to achieve performance goals.

### Physical Data Model Options



#### Message Format

- Standards Reference
- Message Type(s)
- Message Fields with Representations
- Map from LDM to Message Fields

#### File Structure

- Standards Reference
- Record and File Descriptions
- Map from LDM to Record Fields

#### Physical Schema

- DDL or ERD Notation (with sufficient detail to generate the schema)
- Map from LDM to PDM with Rationale

#### Other Options

- 
- 
-

# Data Modeling and Predictive Technology

## Evolution of Access to Data Modeling

- With the growing needs for efficiency in business, predictive analytics has been on the rise.
- Gartner observed, “Predictive analytics vendors are trying to reach a broader audience than traditional statisticians and data scientists by adding more exploration and visualization capabilities for novices and business users,” (Linden et al, 2014).
- The solution to this problem came through automating the data modeling process through Machine Learning and incorporating elements of Natural Language Processing.

## Machine Learning

- Machine Learning algorithms help to determine patterns between existing and future data.
- When this technology is applied to data models, users can effectively create new models from previously existing ones that address specific business problems or use cases.



# Data Modeling and Predictive Technology

- **Types of Machine Learning Algorithms**

- **Deep Learning**

- Modeled on the structure of the human brain, deep learning allows the machine to use a multi-faceted approach in order to gain insight and context from unstructured data.

- **Ensemble Learning**

- The combination of individual models that are continuously using the output of the whole to refine the structures of the individual models, making each iteration better while increasing the accuracy of final output.

- **Bootstrap Aggregating**

- Leveraging output from models that use randomly generated samples of the initial dataset to improve the stability of the regression and classification analysis in the original underlying algorithm. With each cycle the output would have higher predictive confidence level.

- **Interplay with Natural Language Processing**

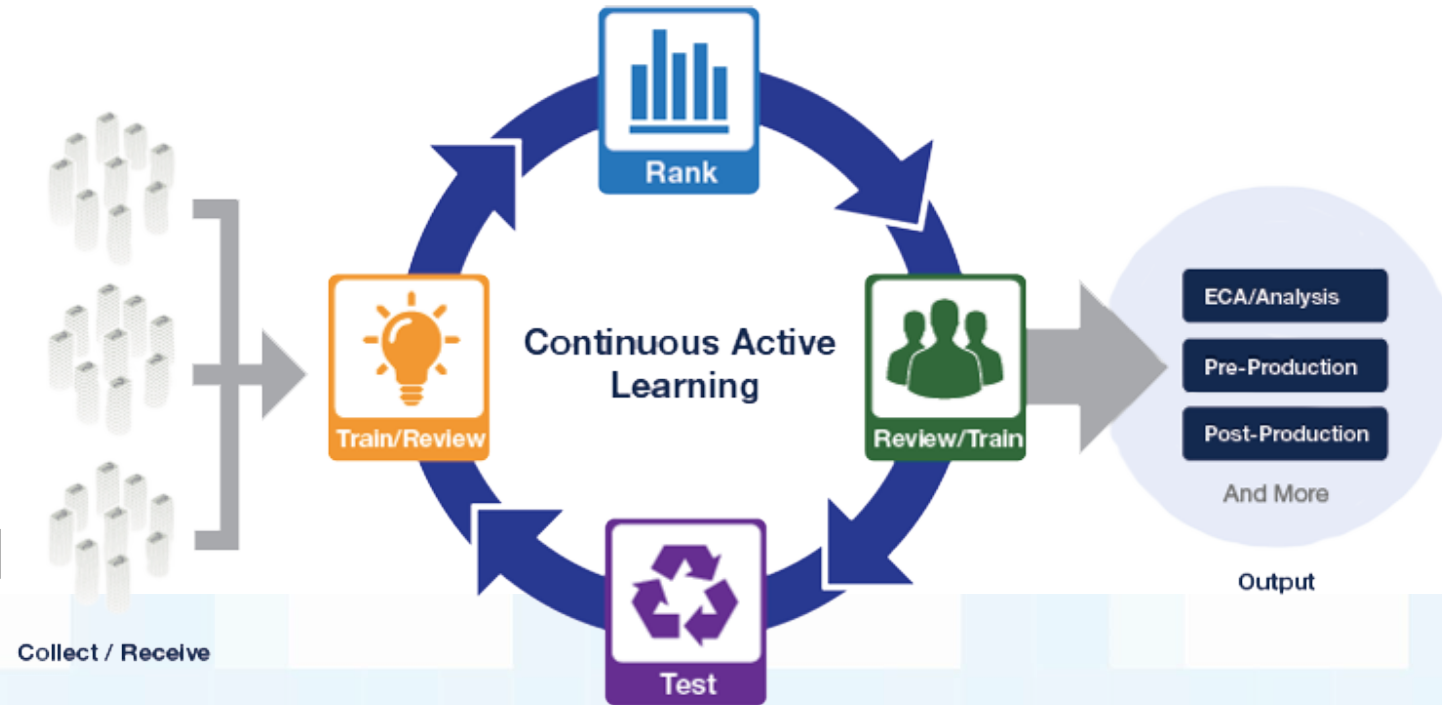
- Using Natural Language Processing, some solutions can provide query results or explanations of those results. (Modern predictive coding)

# Current State of Predictive

## Coding (TAR 2.0) 1

### TAR 2.0 Process Overview

- Sample and Code Initial Documents
  - Review Documents
  - Continue the Review Process
  - Sample the Un-Reviewed Documents
- 
- Continuous Active Learning – The Game Changer



# Current State of Predictive Coding (TAR 2.0) 2



## Acceptance of TAR By Courts

- “[N]ow black letter law” – “where the producing party wants to utilize [TAR] for document review, courts will permit it”
  - *Rio Tinto PLC v. Vale SA*, 306 F.R.D. 125 (S.D.N.Y. 2015) (Peck, J.)
- “There may come a time when TAR is so widely used that it might be unreasonable for a party to decline to use TAR. We are not there yet.”
  - *Hyles v. New York City*, No. 10CIV3119ATAJP, 2016 WL 4077114 (S.D.N.Y. Aug. 1, 2016) (Peck, J.).
- The current trend is not to disclose the seed set.

# Current Uses of Predictive Modeling by Corporations and Law Firms



- Predictive Coding
- Modeling Across Matters
- Contract Analysis
- Legal Research and Writing
- Spend Assessment

# The Future of Predictive Modeling Left of the EDRM



- Make your data work for you to:
  - Identify PII
  - Assess Privilege at the Source
  - Segregate Trade Secret and other Competitive Intelligence
  - Build a Stronger Compliance Program
  - Identify At Risk Employees and Employment Situations
- Basically if you have data on it – you can model it

# Questions?

