

PROFISEE MDM AND MACHINE LEARNING

Master Data Management and Machine Learning are both data-intensive technologies and can enhance and enable each other. MDM can improve the quality of data used to perform machine learning, reducing data preparation efforts while improving the accuracy of the model through better data. Conversely, Machine Learning can also be used to automate MDM, reducing the burden on administrators and data stewards.

MDM-ASSISTED MACHINE LEARNING

Machine Learning and Artificial Intelligence applications are very powerful additions to the analysts arsenal allowing inferences to be drawn from large datasets with more speed and accuracy than is possible with only human analysts.

Machine learning approaches can improve outcomes in a wide variety of situations such as these:



Personalized eCommerce – boost loyalty and engagement with customized interactions based on previous purchases



Fraud detection/Anti-Money Laundering – detect anomalies in spending patterns and alert possible fraudulent or illegal activity



Healthcare Outcomes – determine which patients are more likely to develop complications based on medical records

Machine Learning relies on high-quality, trusted data. If 'normal' IT is a garbage-in, garbage-out proposition, then ML/AI is garbage-in, garbage-out on steroids. If the incoming data is inconsistent or incomplete, then training will produce inaccurate models.

The common solution to this problem is increasingly sophisticated 'data prep', which is where data scientists spend most of their time. While some data prep is necessary, much of it is invested in redundant efforts to mask data quality issues in the most important data, the master data that represents the customers or products associated with a big data set.

The better and more scalable solution to create a managed set of high-quality trusted Master Data as a cornerstone to Machine Learning. By mastering critical data domain-by-domain, inconsistent, missing and duplicated data can be systematically eliminated, leaving data that is the ideal foundation for ML and AI applications.



MACHINE LEARNING-ASSISTED MDM

As much as MDM can assist Machine Learning, Machine Learning can also assist MDM to simplify administration and reduce the burden on data stewards.

Matching – This is one of the most critical capabilities of an MDM solution. Profisee MDM uses sophisticated machine learning techniques to enable intelligent matching of data between and across applications.

Matching and grouping duplicate data is a clustering problem, with some unique challenges in the context of MDM:

- Variability of domains and rules requires flexibility in model construction
- Most organizations lack high-quality training data, requiring a more sophisticated unsupervised training approach
- Support for stewardship of ML generated results
- Retraining of the model based on the actions of data stewards
- High performance for initial match, as well as ongoing incremental matching
- Results must be 'explainable' or risk not being trusted

Taken together, these requirements represent quite a challenge. To solve for this Profisee developed its own proprietary machine learning matching engine, using a unique combination of a co-sine similarity distance algorithm, couple with an all-pairs similarity algorithm to efficiently process large sets of data. Taken together, these approaches meet the distinct needs of the Master Data Management use case.

For the technically-minded: Profisee's ML matching algorithm begins by Featurizing the input identifying attributes into a sorted vector of n-grams per attribute. Profisee then builds an ML model based on these Features which is automatically maintained in memory to support high performance in both initial and ongoing matching scenarios. This model is continuously re-trained as stewardship occurs, allowing the engine to continue learning from data stewardship actions.

For the less-technically-minded: Profisee MDM has ML-assisted matching which enable high performance and accuracy!

Data Stewardship – In addition, machine learning can be used to actively assist data stewards in resolving data issues by 'learning' from previous manual corrections and suggesting future corrections – thus saving time and effort from human experts.

And of course, the faster and more effective the data stewardship, the more data and domains can be mastered, and the better the overall data available to drive business intelligence, operations and of course ML-based predictive analytics.