

Making ‘Competing On Analytics’ A Reality In Healthcare: Using Machine Learning To Bring Order To Data Chaos

Developing Analytic Prowess Requires Skillful Data Management

A multinational medical devices, pharmaceutical, and consumer packaged goods organization set out to better compete and grow through the use of analytics. They wanted to leverage all of their data, both internally generated through decades of business operations and externally purchased, to optimize their business operations from sales to marketing to supply chain functions. Their goal was to answer key business questions like “which product is selling best in which geographic region?” and “what is our best sales channel for a particular sub-brand?”

The company had created a best-in-class analytics platform to analyze their existing data, but soon realized that the data feeding these analytic tools varied too much in structure to be easily tied together and analyzed. This is primarily because company sales are driven through thousands of global retail partners, with each partner sending the organization point-of-sale data in a unique format with regional product naming conventions and other variations. Moreover, the company purchases enrichment data from third parties with their own unique formats and naming conventions – creating even more extreme variety and serving to exacerbate data management issues.

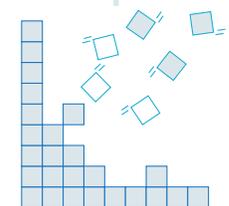
At the core of the company’s challenge was the need to unify the disparate datasets to provide a single view of all customers and products that would feed downstream analytic tools. Each customer and product in the wide range of datasets needed to be uniquely identified and classified for the data to be effectively used.

Tamr Builds A Centralized Data Management Capability

To achieve this vision, the organization needed to initially build a centralized capability for managing the company’s product data. Their internal data included everything from SAP ERP data to finance data, while their external data encompassed a range of sources from retailer point of sale datasets to market data. The ultimate goal was to unify and categorize both the internal and external data such that each product (and customer) could be uniquely identified as well as classified into a five-tier product taxonomy consisting of 38 categories. In this way, the company can compare, for instance, product pricing, sales trends and market share across a variety of geographies.

The internal data team decided to implement a Data Lake strategy that would create a centralized data staging and integration layer for large scale rapid data processing, data mining, and analytics. Unfortunately, their efforts to unify the variety of datasets within the repository were ad-hoc and completed through manual coding. This resulted in integration processes that were not only too slow, but also not granular enough to meet business intelligence requirements. For instance, the company did not have unified, detailed views of their sub-brands, limiting the strength of their downstream analyses.

“What is our best sales channel for a particular sub-brand?”



The company needed to unify disparate datasets to provide a single view of all customers and products

The business needed granular insight in a very tight timeframe to execute on strategic decisions. They gave the internal data team only a few months to determine how to solve the issue of unifying and classifying all of the internal and external data within the Data Lake. The team knew that a new approach was needed if they were to achieve results in such a short period of time.

Understanding that manual coding was not going to enable them to reach their goal, the team evaluated other traditional data management techniques, such as Master Data Management, against Tamr’s machine learning-based approach. After testing each of the approaches, the organization selected Tamr due to the advantages demonstrated after a head-to-head evaluation process, which included:

- + Time to Value - While legacy data management technologies exist in most large organizations, they are often burdened by a rules-based approach. Mastering entities such as customers or products and categorizing them to a multi-tier taxonomy takes too long using the construction of rules. Tamr’s human guided, machine learning-based approach delivered significantly improved time to value, as algorithms recommend customer and product matches as well as classifications while an internal expert validates the matches when needed.
- + Scalability - The desire to integrate an increasing amount of disparate internal and external data across many formats requires a scalable approach. In addition to speed, Tamr’s technology enabled the data team to integrate as many sources as desired for analysis. Tamr incorporated feedback from the organization’s experts into the technology so that the next series of data from known sources could be integrated with minimal human intervention. In this way, the company could gain as much visibility and granularity as desired with minimum effort needed.
- + Flexibility - Legacy data management solutions often require the purchase of multiple products to cover multiple data domains. Companies are forced to purchase products to master and classify each organizational entity - whether products, customers, suppliers, or others. Tamr was a compelling option for the data team because of its multi-domain support and its superior classification capabilities. This flexibility will allow the organization to continue to bring in new data, adapt to new goals - such as mastering across different data domains - and efficiently respond to changes in internal taxonomies.

Driving Business Outcomes By Innovating In Data Management

With Tamr added to the company’s data management capabilities, the value realized from their Data Lake investment has increased dramatically. The internal data team met their aggressive project deadlines to deliver trustworthy, unified data sets to their business stakeholders, thereby enabling them to focus their efforts on analysis and action rather than data preparation. The company’s Tamr-powered data pipelines continue refresh these analytics as the underlying source data is updated, ensuring that the business is always working with the most current information. Finally, the data team is able to incorporate new sources into the mix quickly and easily using Tamr’s flexible, dynamic, machine learning-powered data models. Tamr has proven that its approach to data unification is fundamentally different and more compelling than traditional rules-based approaches, and is now considered an essential part of the company’s next-generation data management capabilities.

