



Machine Learning: Taking Data Management Beyond Human Levels

PRESENTED BY



SPONSORED BY



The basic premise of ML is to build algorithms that can receive input data and use statistical analysis to predict an output value within an acceptable range.

I Introduction

With more data pouring into organizations than ever before, the need to manage said information has never been more pressing. As such, business leaders are searching for help. Artificial Intelligence (AI), fortunately, can empower data management initiatives to quickly deliver business insights, support data governance and bolster return on investment. Machine learning (ML) brings even more capabilities to the data management endeavor. A look at exactly how ML differs from AI; the value that ML can bring to data management efforts; how ML compares to human data management initiatives; and industry-specific ML data management use cases illustrates how this emerging technology can enable businesses to expediently leverage vast amounts of data to make better strategic decisions.

II Artificial Intelligence vs. Machine Learning

AI is a branch of computer science dealing with the simulation of intelligent behavior in computers. With AI, human intelligence is simulated by machines, specifically computer systems. As such, AI-enabled computer systems do not just process information, they also learn (by acquiring information and rules for using the information), reason (by using the rules to reach approximate or definite conclusions), and self-correct, according to a definition from TechTarget.¹

ML takes AI a step further. ML is a type of AI that enables software applications to become more accurate in predicting outcomes without being explicitly programmed. The basic premise of ML is to build algorithms that can receive input data and use statistical analysis to predict an output value within an acceptable range.² With the development of neural networks, computers and machines can be taught to think as humans do. What's more, the addition of a feedback loop enables learning as the machine senses whether its decisions are right or wrong and then modifies its approach in the future.³

ML can be used for a variety of business purposes. So far, applications have been developed to discover how the human genome works, understand consumer behavior and construct systems for purchase recommendations, image recognition, and fraud prevention, among other uses.⁴

III Machine Learning: Bringing Value to Data Management Efforts

When applied to data management, machine learning provides significant value by detecting patterns and discovering relationships within data. Many organizations, in fact, are dealing with large volumes of data – and are often confronted with undocumented data sources. Databases, in fact, were created long ago (typically 15 – 30 years) and since then they have grown over time. The documentation and any related information was either not completed or has been lost over the years. Without this documentation, it takes longer to understand the complexity of the database in order to start any meaningful analysis. To overcome the lack of documentation, ML can efficiently understand patterns from existing data.⁵

As such, ML can provide the automated discovery of data relationships that is critical to documenting, understanding, managing and protecting data. More specifically, ML can help companies auto-discover the location, movement and flow of data throughout their organizations. This enhanced data discovery process makes it possible to efficiently uncover complex data relationships, improve data visualization and expedite the data preparation process.

Indeed, a powerful ML solution goes far beyond simply understanding data catalogs and business glossaries and intrinsically enables the data to speak for itself by identifying classifications that naturally emanate from the data.

According to a report from ISACA, a global technology association focused on helping individuals and businesses achieve the positive potential of technology, ML can “give enterprises the capability to not only discover patterns and trends from increasingly large and diverse datasets but also enables automated analysis that has traditionally been done by humans, to learn from business-related interactions and deliver evidence-based responses.”⁶

ML applications can surpass human ability to manage data because these solutions are capable of:

- Detecting certain types of characteristics that are difficult for humans to identify
- Quickly analyzing large amounts of data, even terabytes
- Identifying fuzzy patterns, where elements are given partial membership in a particular data set, instead of the absolute membership associated with crisp data sets
- Easily identifying errors in the data, making it possible to avert erroneous conclusions.

However, not all ML applications are created equally. Truly powerful ML solutions don't merely rely on an understanding of metadata. Indeed, a powerful ML solution goes far beyond simply understanding data catalogs and business glossaries and intrinsically enables the data to speak for itself by identifying classifications that naturally emanate from the data. So, instead of trying to group data into established metadata classifications, these solutions would understand the underlying data and actually define the types of metadata needed. In other words, instead of forcing square data into a round hole, it would come up with a custom-sized shape for the data.

IV Machine Learning in Action: A Case Study

A financial services company recently discovered just how valuable machine learning could be to its data management efforts. The firm was exploring options to support a new migration project, aimed at consolidating their data centers. The project would require moving 80 servers.

To proceed, the firm needed to better understand its data. More specifically, leaders wanted to understand how data flows through internal systems and how one element relates to another. Understanding how the applications were connected with one another would also help in establishing the order of database migration and would provide the insight to better position their company for making business transformation decisions backed by accurate, reliable data.

Initially, the organization employed approximately 20 subject-matter experts (SMEs) to pore through the data and work toward migrating and re-platforming 80 servers from Sybase to Oracle; understanding data relationships within and across these databases; simplifying and removing redundant data, in an effort to enable a smoother transition between platforms.

This manual migration of all 80 servers would have taken the SMEs four years to accomplish. In fact, just analyzing the sheer volume of data on those servers to reduce redundancies, eliminate identical data sets, centralize data flow management and streamline a complex maze of data relationships would have taken between 18 and 24 months. Even then, the effort would still only scratch the surface of the data relationships, as it would be impossible to manually analyze every data element. What's more, manual discovery

is prone to mistakes and limited by the knowledge of its users, even ones who are subject-matter experts in their field.

After deciding the manual approach was not realistic, the firm leveraged a machine learning data discovery solution to help uncover 'hidden' data relationships and redundant data. The ML application ran in the environment, processing 785 tables, 8,800 columns and 1.6 billion rows of data elements in just 14 days. The machine learning solution processed seven times more data than the SMEs could process manually. By relying on ML instead of human SMEs, the company cut the original 18 to 24 month planned migration time in half.

V Conclusion

Machine learning applications can help a variety of companies by making it possible to efficiently and effectively ensure data quality and manage data in a manner that will help them to move their organizations strategically forward. While some larger companies could develop the machine learning applications, doing so falls outside the capabilities of most. As such, rather than taking advantage of the raw science of machine learning, most companies should instead leverage products that have already been developed such as Io-Tahoe's platform. To learn more about how your company can leverage Io-Tahoe go to: www.io-tahoe.com

References

1. AI (Artificial Intelligence). TechTarget. <http://searchcio.techtarget.com/definition/AI>
2. Machine Learning. TechTarget. <http://whatis.techtarget.com/definition/machine-learning>
3. Marr, B. What Is the Difference Between Artificial Intelligence and Machine Learning? Bernard Marr & Company. <https://www.bernard-marr.com/default.asp?contentID=958>
4. Bell, L. Machine Learning vs. AI: What's the Difference? Wired, December 2016. <http://www.wired.co.uk/article/machine-learning-ai-explained>
5. Jain, A. Latest Trends in Data – Machine Learning & Predictive Analysis. Io-Tahoe. <https://io-tahoe.com/latest-trends-in-data-machine-learning-predictive-analy>
6. ISACA. Machine Learning Drives Big Business Benefits. <http://www.isaca.org/Knowledge-Center/Research/ResearchDeliverables/Pages/machine-learning-drives-big-business-benefits.aspx>