

## Summary Analytics for ML/AI Development

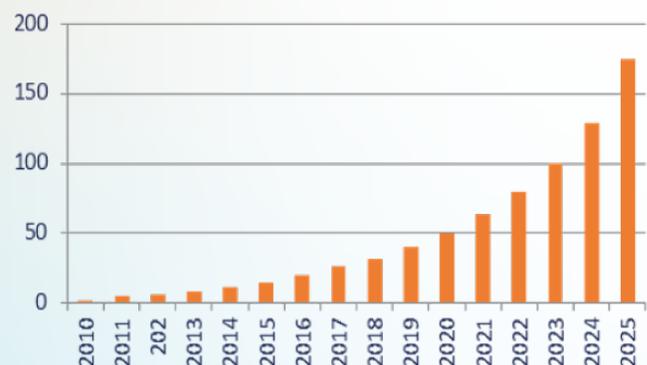
Are you feeling pressure to scale your artificial intelligence (AI) projects quickly and roll them into production? You are not alone. Based on a recent Accenture study, an amazing 75% of C-level executives believe they risk going out of business in the next five years if they can't scale their artificial intelligence (AI) efforts aggressively.<sup>1</sup> And according to Gartner, AI augmenting human decision-making will create US\$2.9-trillion of business value and 6.2-billion hours of worker productivity globally in 2021.<sup>2</sup> Machine learning (ML) and AI teams are feeling intense pressure to get going while hiring skilled employees is nearly impossible. The results are not pretty. Dimensional Research found that 96% of companies have run into training related problems – including data quality, labeling required to train an AI system, and building model confidence – with 78% of their ML/AI projects stalling at some point before deployment.<sup>3</sup> And Pactera Technologies's survey showed that 85% of AI projects ultimately fail.<sup>4</sup> How do you get the benefits of AI, and get them quickly, when the odds are against you?

To get more done with your limited resources, you must get more productive. Developing an effective AI model requires extensive repetition with trial and error analysis of historical data. But often the historical datasets are overwhelming in size and need manual labeling before the models can be tested. With Summary Analytics's mathematically proven artificial intelligence techniques, you can shrink the datasets through summarizing and prioritizing without loss of fidelity – delivering better insight while reducing time and cost, and significantly reducing the amount of manual data labeling required. This minimizes the common problem of operator fatigue errors in data labeling and the resultant errors in the models.

Likewise, Summary Analytics can help with training your AI models. The computational power required to train state-of-the-art AI models is doubling every 3.4 months<sup>5</sup> as Moore's Law continues losing steam, no longer doubling processor performance every 18-months. So far, this problem has been addressed with machine learning algorithmic advances and increased parallel compute power. These help, but more is needed to stop runaway AI analytics costs and delays. A new complementary tool is needed, adding "informational efficiency" to the process. That tool is Summary Analytics. Our software-as-a-service (SaaS) offering summarizes and prioritizes data sets before running expensive analytics. Summary Analytics enables early model testing on significantly reduced and prioritized datasets, while saving larger (but still reduced) datasets to be used for final optimization of the model.

Summary Analytics eliminates redundancies in your data. More than just deduplication, we do this even among massive numbers of unique records. We eliminate the unnecessary and shrink the haystack so finding the needle of insight is faster and less expensive. Of course your data is more complex and dynamic than a single needle in a haystack, with new streams and data churn constantly adding new and removing old hay and needles. Worried about what to do with old data? We help

Annual Size of the Global Datasphere  
(Zetabytes)



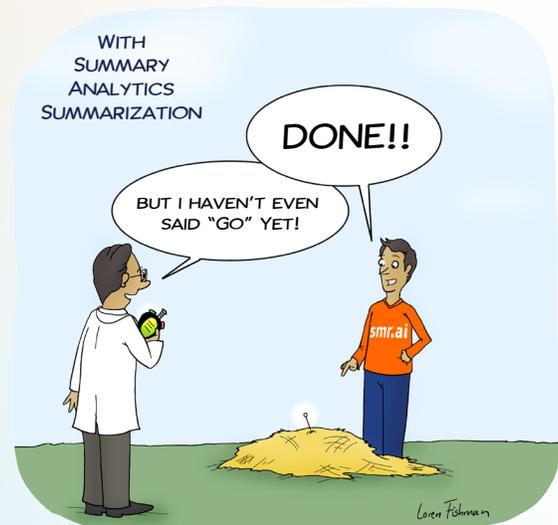
Source: Data Age 2025, from IDC Global DataSphere, Nov. 2018

## Summary Analytics for ML/AI Development

create a data hierarchy to focus on the important data, whether old or new. Now you can quickly focus your resources on what's most important to your business, accelerating decision-making while reducing alert and decision fatigue. expensive analytics. Summary Analytics enables early model testing on significantly reduced and prioritized datasets, while saving larger (but still reduced) datasets to be used for final optimization of the model.

How does it work? Professor Jeff Bilmes from University of Washington in Seattle developed proprietary calibrated submodular (CaSM) functions which mathematically analyze and order data along the lines of diminishing marginal returns. We automatically prioritize the data in terms of its biggest contribution to the information content of the entire data set, and then relegate redundant data to the end. CaSM functions are extremely processor efficient – orders of magnitude faster than typical AI algorithms. They don't replace AI algorithms, our CaSM functions just make machine learning run much faster since the data sets are vastly smaller but still contain all the important information. And we work on any kind of data, whether health records, customer profiles, network logs, biological signals, sensor data, or even images, audio, and video streams.

Imagine developing your stock trading AI model in one tenth the time, or training your cybersecurity application to analyze email or network traffic with 90% less hardware resources. We can help with data bias analysis, too. The bigger or more redundant the data, the more Summary Analytics can reduce costs and make your ML/AI team more productive.



**Bigger data? Bring it on!**

<sup>1</sup><https://newsroom.accenture.com/news/failure-to-scale-artificial-intelligence-could-put-75-percent-of-organizations-out-of-business-accenture-study-shows.htm>

<sup>2</sup><https://www.gartner.com/en/newsroom/press-releases/2019-08-05-gartner-says-ai-augmentation-will-create-2point9-trillion-of-business-value-in-2021>

<sup>3</sup><https://content.alegion.com/dimensional-researchs-survey>

<sup>4</sup><https://www.techrepublic.com/article/why-85-of-ai-projects-fail/>

<sup>5</sup>[https://www.technologyreview.com/s/614700/the-computing-power-needed-to-train-ai-is-now-rising-seven-times-faster-than-ever-before/?utm\\_source=newsletters&utm\\_medium=email&utm\\_campaign=+the\\_download.unpaid.engagement](https://www.technologyreview.com/s/614700/the-computing-power-needed-to-train-ai-is-now-rising-seven-times-faster-than-ever-before/?utm_source=newsletters&utm_medium=email&utm_campaign=+the_download.unpaid.engagement)