Summary Analytics

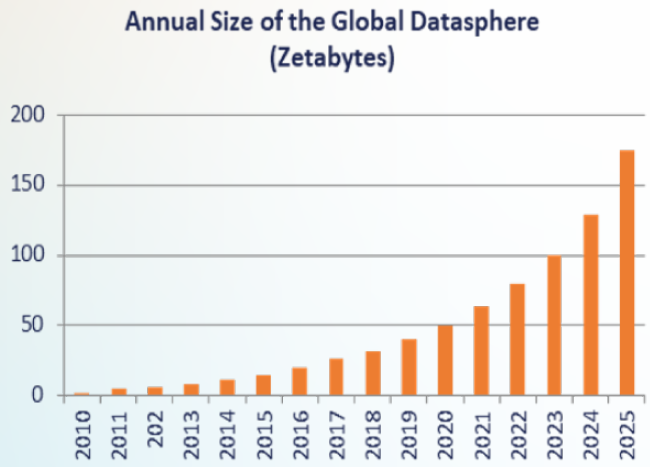
www.smr.ai
415-234-3553
info@smr.ai

Summary Analytics for Cybersecurity

Research firm ESG recently found that among IT executives, "Cybersecurity topped the list of problematic skills shortage areas, just as it has for the past 9 years."[1] The study also found that 51% of large mid-market and enterprise organizations are using machine learning (ML) analytics today in their cybersecurity operations. The report concludes, "CISOs want machines to crunch and analyze more data and help them improve security staff productivity."

Of course with pressure like this from customers, artificial intelligence (AI) and machine learning have become crucial components of many vendor cybersecurity products as well. Yet, 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 AI/ML projects stalling at some point before deployment.[2] Pactera Technologies's survey showed that 85% of AI projects ultimately fail.[3] How do you get the benefits of AI when the odds are against you?

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.



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

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[4] 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 create a data hierarchy to focus on the important data, whether old or new.

Summary Analytics for Cybersecurity

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.

As alert fatigue is a huge problem among overworked cybersecurity analysts, Summary Analytics can also be used to reduce and prioritize alerts, whether from IDS, endpoint threat detection, SIEM, or other cybersecurity monitoring systems. ML and AI are impacting cybersecurity from end to end – from analyzing threat intelligence to prioritizing remediation. But developing the ever-increasingly complex AI models, training these models, and running them with real time production data flows, are all getting more expensive while customers are expecting cost reductions. Summary Analytics works great on all kinds of cybersecurity data including structured data such as network logs, sensor data, and IDS alerts; unstructured data such as research papers, blogs, and internet chatter; and we even work on images, audio, and video streams – the bigger or more redundant the data, the more Summary Analytics can reduce time and costs and dramatically improve your odds of AI success. If you're integrating AI/ML into your internal cyber defenses or you develop commercial cybersecurity products using AI/ML, then Summary Analytics can help make your staff or customers more productive.

Bigger data? Bring it on!

[1]ESG Blog, 1/27/20; https://www.esg-global.com/blog/cisos-are-finding-ways-to-address-the-cybersecurity-skills-shortage

[2]https://content.alegion.com/dimensional-researchs-survey

[3]https://www.techrepublic.com/article/why-85-of-ai-projects-fail

[4]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