Comparing Software and Use Cases

Thursday, February 15, 2018   4:20 PM

Projecting 2-3 years into the future

Capability	Planned Volume	Current Use Case	Potential Software	Hardware (min)
Read from SQL Table	~ 10 million rows	Read data from the Data Science Data Mart	SQL	16 GB RAM, 2+ CPUs
Basic Statistics and Visualizations	~10 million rows	Analysis of 3 million rows of episode data	R, Tableau	32 GB RAM, 4+ CPUs
Text Mining	~2 million rows	Automating validation of Verbal Notifications (reducing overhead and improving compliance)	R, Python	32 GB RAM, 4+ CPUs
Text Analysis	~ 2 million rows	Finding unsupervised patterns using matrices to define counts, sentiment, synonyms, anonyms, sentence structure	R	4+ GPUs, 256 GB RAM, 10 Gigabit Ethernet
Automated, standardized reporting	20+ automated tasks per week	Quarterly updates for SNF Performance Reporting	Windows Task Manager, Python	8 GB RAM, 2+ CPUs
Supervised Learning - Matrix Based	?	Update model weights for nH Identify using linear relationships	R Server, Python, H20ai	64 GB RAM, 4+ CPUs, 10 Gigabit Ethernet
Supervised Learning - Tree Based	?	Update model weights for nH Identify using non-linear relationships	H20ai, Python	64 GB RAM, 4+ CPUs, 10 Gigabit Ethernet
Unsupervised Learning	?	Allowing the machine to find the best predictors to improve our risk prediction models and/or enhance other products with supplemental offerings (e.g., nH Perform)	R, Python, H20ai	64 GB RAM, 4+ CPUs, 10 Gigabit Ethernet
Deep Learning	?	Use Neural Nets and Tensor Flow to find non-linear relationships in social determinants	Python	256+ GB RAM, 8 GPUs, 10 Gigabit Ethernet
Image to	?	Translate the UM Authorization PDF	Open OCR	256+ GB

Text translation		image requests to text in order to pre-separate which authorization requests to go to whom.		RAM, 8 GPUs, 10 Gigabit Ethernet

Total All-Encompassing Solution: Windows OS server with 256+ GB RAM, 8 GPUs, 10 Gigabit Ethernet connection for 10+ users containing Anaconda (Python 3.6), R Server, H20 server, OCR Open Tessaract, and with connections to SQL and Tableau.

This assumes people use it as a Virtual Machine rather than an engine connected to their desktop IDEs.