Domino Data gives its enterprise data science platform a makeover with model velocity in mind

January 27 2022

by Krishna Roy

The provider of a platform used to code machine learning models, place them in production, and keep them up to date has accelerated the rate at which new and enhanced models are delivered in a bid to better support the prevailing transition to model-driven decision-making.
Introduction

Domino Data has identified machine learning model velocity – the speed at which models move from conceptualization to operationalization by placing them in production and monitoring them to maintain their relevancy – as a major issue stymieing corporate efforts to successfully run a model-driven business. To address model velocity issues, the vendor has released the latest version of its enterprise data science platform.

THE TAKE

Domino's focus on model velocity should boost its offering's appeal because the enhancements we explore below tackle several key issues. Autoscaling of clusters frees up data scientists' valuable time so they don't have to handle the plumbing required for each machine learning (ML) model. The debut of prebuilt connectors – as well as a self-service framework for creating and reusing connectors – addresses data access headaches, which are a major problem. Almost one-quarter (23%) of respondents to 451 Research's Voice of the Enterprise: AI & Machine Learning, Infrastructure 2021 survey cited accessing or integrating data as a challenge in AI initiatives. Furthermore, model monitoring improvements demonstrate that Domino has recognized the challenges that occur once models have been operationalized, and thus the requirement to provide familiar tools – such as metrics, alerts and dashboards – along with automated insight as a core platform capability, rather than a separate add-on. Nonetheless, we still think the company could do more to meet the needs of nonexpert business users, who are key personnel involved in the model-driven decision-making process.

Details

From the start, Domino Data's strategy has been to offer flexible 'open' code-first data science to teams of expert data scientists in large organizations. The vendor's openness is all about enabling data scientists to utilize their favorite programming languages and tools to code ML models – as well as existing compute environment and infrastructure to run them on – to preserve these existing investments and lower the barrier to enterprise-ready data science. Domino 5.0 is all about cementing this strategy by addressing model velocity so enterprises can essentially get more models into production running successfully in a repeatable and scalable way, and in so doing run their business on them. The latest version also builds on capabilities introduced in the previous two releases, examined here, to facilitate this endgame.

Domino has concentrated in the past year on the management of computing resources needed to provide the horsepower for models. With Domino 5.0, the ability to spin up distributed clusters in the form of Apache Spark, Ray or Dask clusters has moved from a point-and-click to an automated process courtesy of the introduction of autoscaling of these clusters. The move means a data scientist no longer must specify the number of executors – i.e., the compute and storage set aside for each computer instance for each model – as the platform automatically monitors what's required and adjusts executors accordingly. Therefore, Domino 5.0 will power up for a resource-hungry deep learning model, for example, and power down again for an ML model that requires less resources.
The vendor’s new connector strategy present in Domino 5.0 paves the way for more user-friendly connectivity as well as hooks into data that can be securely shared and reused across common data access patterns. Data access is a challenge when initially creating an ML model as well as refreshing the model with new data once it is in a production environment, such as embedded within an app. Domino has crafted prebuilt connectors to Snowflake, Amazon S3 and Redshift – with more in the works according to customer demand – as well as delivered a DIY connector environment for data scientists to create their own, in order to address data access issues. The DIY connectors are shareable, based on user permissions for security purposes, so a data scientist can build a connector to a specific data source and share it with another team member.

Finally, Domino has integrated Model Monitor, which was previously a separately priced module, into its platform to morph machine learning model monitoring into a standard basic feature to make it easier to deploy models. Additional model monitoring capabilities that are charged for by capacity are also available, if required. Automated insight in Domino 5.0 make it easier to dig into model quality issues to find out why a model has drifted, for example. Additionally, the company’s automated insight uses cohort analysis to ascertain the features that have the most impact on a model. The generation of customizable reports so that nonexperts such as business decision-makers can understand model quality issues – including why a model is being replaced, for example – is another aspect of its automated insight functionality.