X

VIEW

Resources > Alliance Market Insight > View

The feature store: Examining the rise of a new repository for machine learning

February 12 2021 by Krishna Roy

Introduction

Some of the largest AI-driven tech companies – including Uber, Google, Netflix, Facebook and Twitter – are already familiar with feature stores, having built their own. For many enterprises, the term is just another piece of industry jargon that is easy to dismiss in all the buzz surrounding machine learning, but specialist startups are seeking to build a business around them. Furthermore, with the debut of Amazon SageMaker Feature Store, the first big-name vendor has entered the sector to validate it, signaling that feature stores are now a growing product category and worthy of closer inspection.

The 451 Take

In 451 Research's AI & Machine Learning, Use Cases 2021 survey, 86% of respondents agree that the pandemic has caused or will cause their organization to invest in new AI initiatives,

which paints a more optimistic picture than when we asked that question in 2020. Our data from that report in August 2020 indicated that 58% of organizations surveyed expected COVID-19 to have a negative impact on their existing AI initiatives, and 19% said the pandemic had led them to stop work on these projects. Off-the-shelf feature stores have therefore hit the market at the right time to capitalize on this positive market sentiment. Furthermore, it will be interesting to see how the feature store sector evolves. On the one hand, there are clear stand-alone opportunities for organizations that want to purchase one from a specialist. However, we also expect the feature store to become an integral part of many enterprise data science platforms, so evaluating both approaches to ascertain the best fit is a good idea before making an investment.

A feature what?

To fully comprehend the business value of implementing a feature store, it is necessary to first understand what a feature is and why it is critical to machine learning. In machine learning projects, a feature is an individual observable phenomenon that can be quantified, as well as recorded, so it can be used to build a machine learning model. For example, a machine learning model trying to determine the probability of heart disease in patients could have gender, age, height, weight, blood pressure and resting heart rate as features.

Choosing informative, discriminating and independent features is a crucial step for effective machine learning – the accuracy of the machine learning model is based on a precise set and composition of features. Having irrelevant features in data can decrease the accuracy of the models and make them learn based on irrelevant features. At its most basic level, a feature store acts as a central vault for the best features.

Benefits

Storing documented, curated and access-controlled features for machine learning models, the feature store was designed to provide organizations with a faster machine learning development cycle, lower time to production and easier adoption of machine learning across teams.

Additionally, the features in the store are reusable and shareable, so they don't have to be recreated for every machine learning initiative, bringing fresh productivity and speed to the model-building process. It is also easier to standardize feature definitions and conventions, thereby making the model-building process more systemized and rigorous. Furthermore, it is easier to achieve consistency between models and data as the feature store helps build the data pipeline for machine learning.

Users

In theory anyone involved in machine learning, including data engineers, developers and IT staff, could make use of the feature store. However, features stores are typically marketed to data scientists for when they start a new project, so that they can find the features they are looking for, and in so doing save valuable time and effort in model building by not having to start from scratch.

Instead of writing features to a custom storage location after completing a data engineering pipeline, the data scientist essentially writes the features to the feature store, which means they are documented and versioned, too, which is valuable for adherence to model lineage, governance and compliance processes.

Offline, online or both

Feature stores come in two flavors: for offline and for online features. Offline features are essentially calculated as part of a batch job and are mainly used for offline processes. Average monthly spending is an example of an offline feature.

Online features are more complex – they need to be calculated swiftly as they are typically served up in milliseconds. An example of an online feature is a z-score (or standard score) for real-time fraud detection, which is designed to show how far from the mean the fraudulent activity is, at the time it occurs.

Some vendors provide separate feature stores for offline and online features, in recognition of their distinctive requirements. Others handle both feature types within the same feature store.

Table stake capabilities

Feature store functionality varies from vendor to vendor, as alluded to earlier. However, a feature store generally houses five main functions: feature storage, feature serving, feature monitoring, a feature registry and data transformation.

Indeed, it is the data transformation aspect that sets them apart from a simple data storage service because this functionality essentially turns feature engineering into a first-class construct. Well-conceived feature engineering, as noted earlier, is a critical part of successful machine learning.

Vendor landscape

Amazon SageMaker Feature Store is arguably the best-known product in this sector. Feature stores are also available from specialist startups. Tecton.ai, Scribble Data, StreamSQL, Molecula and Logical Clocks are notable specialist providers.

Open source feature stores are another alternative to consider for organizations that are open source devotees. Feast is possibly the most well known in this category. It was jointly developed by Google Cloud and Gojeck. Open source feature stores also exist for specific use cases. For example, Wix.com's open source feature store is for discovering all the features a company can use to design and manage its website online.

Finally, we expect other purveyors of enterprise data science platforms, aside from AWS and Iguazio, to embrace feature stores in order to fully address the machine learning lifecycle from within an enterprise data science stack, using the feature store as one aspect of that.
