DATA SHEET

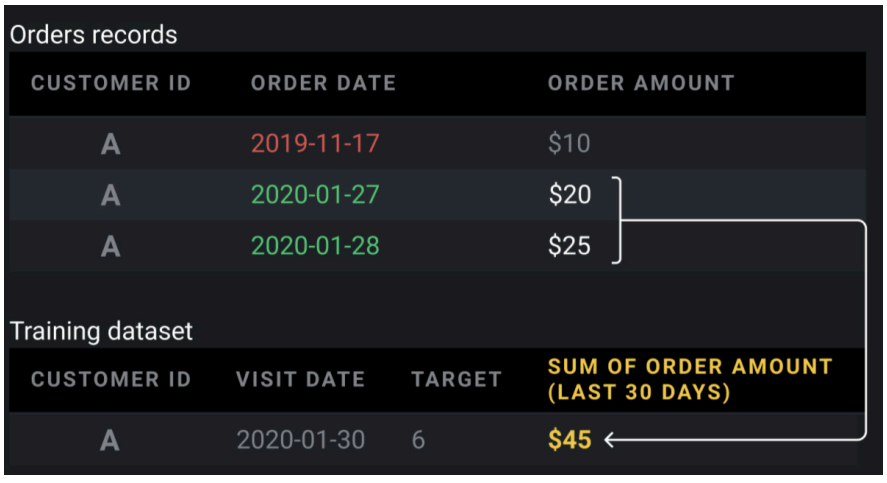
Next-Generation Automated Feature Engineering with DataRobot and Snowflake

Adopting AI across an organization has significant hurdles. AI is complex, takes time to create, and releasing poor quality models into production can pose an enormous risk to the business. DataRobot's Feature Discovery streamlines this process by offering automated feature engineering that enables the creation of valuable new features for your machine learning models. This capability is integrated with the Snowflake Data Cloud making the process both faster and more cost-effective. Snowflake and DataRobot have invested time and resources into advancing the capabilities and benefits of one of the most critical tasks in AI - feature engineering.

The Best Features from Your Snowflake Data

Feature engineering is one of the most critical tasks in AI. The features you create often determine the success or failure of your machine learning projects. The issue is that raw data rarely has all the right features your models need, and if you're using multiple data sources you need to consolidate your data into a single table to train models and make predictions.

This usually involves joining multiple tables and exploring many aggregations (i.e., sum, max, avg, count, entropy, etc.) on different derivation windows (e.g., last 30 days, last week, etc.).



(e.g., Sum of order amounts, among last 30 days orders, by customer_id)

WHEN PERFORMED MANUALLY, FEATURE ENGINEERING IS:

Time-consuming:

the more data sources you use, the more time you need to explore transformations and evaluate the results.

Error prone:

An error such as a missing filter can have a huge impact on the results and introduce target leakage.

"We used the DataRobot Feature Discovery tool on our high-frequency physiological data, which helped us find some new features that ended up being high impact throughout all of our subsequent analysis.

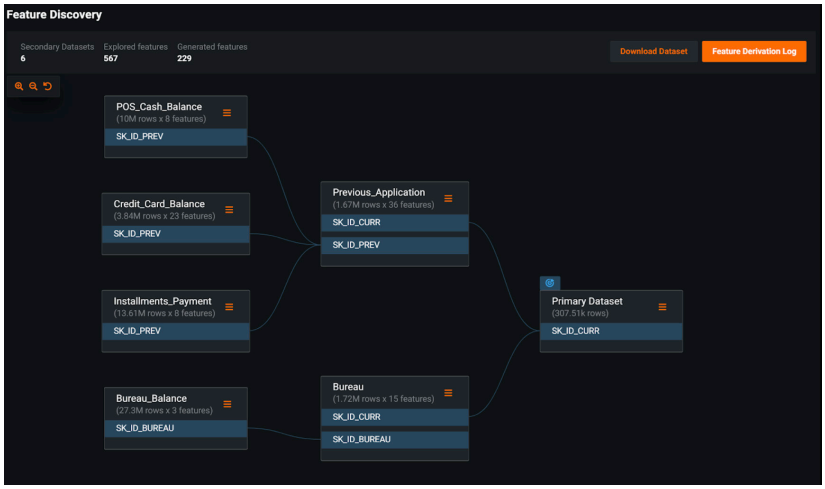
Dr. Austin Chou, PhD

Lead Data Scientist, UCSF Brain and Spinal Injury Center Zuckerberg San Francisco General Hospital and Trauma Center

DataRobot's Automated Feature Discovery simplifies and accelerates this feature engineering process through the automation of expert data science best practices.

DataRobot's Feature Discovery: Next-Generation Automated Feature Engineering

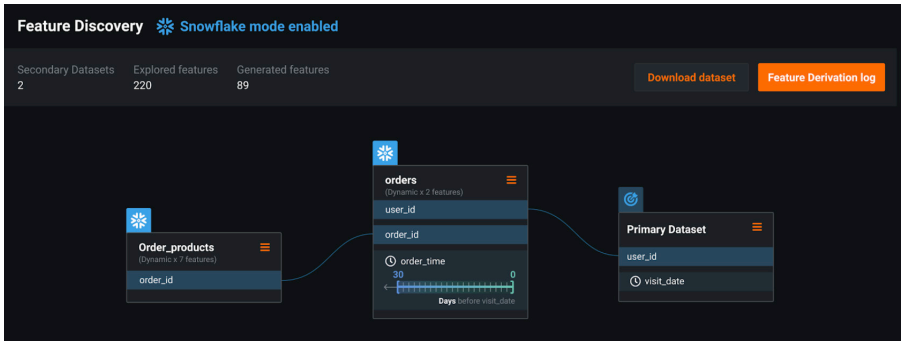
Compared to usual automated feature engineering solutions, Feature Discovery can leverage data from multiple datasets, not just one, and automatically discovers, tests, and creates hundreds of valuable new features for your machine learning models, dramatically improving their accuracy.



Now Faster with Snowflake

Exploring multiple data sources has always required transferring large amounts of data between systems which was resource-intensive and time consuming.

DataRobot's new Snowflake integration pushes Feature Discovery operations into Snowflake to minimize data movement, resulting in faster results and lower operating costs.



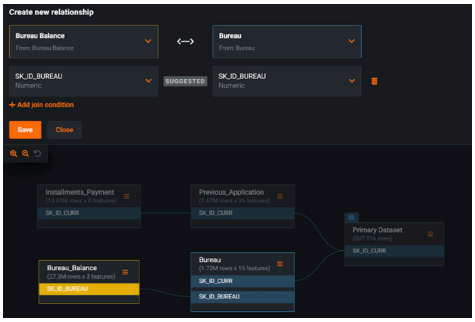
In many ways, Feature Discovery is an extension of DataRobot's AutoML (Automated Machine Learning) product in terms of the automation that it brings to the data science process. This integration now allows users to get even more accurate models from their Snowflake Data Cloud.

DataRobot's Feature Discovery

Automated feature engineering taken to a new level.

VISUAL AND INTUITIVE

Feature Discovery unlocks the art of advanced feature engineering for data scientists, data engineers, and business analysts. Using DataRobot's visual relationship editor, you can select all of the datasets you want to use in your project, then quickly declare relationships between the datasets with just a few clicks. DataRobot even suggests joins for you if you don't know the relationships in advance. Feature Discovery makes it incredibly easy for anyone to define very complex data schemas upon which to perform automated feature engineering in minutes.

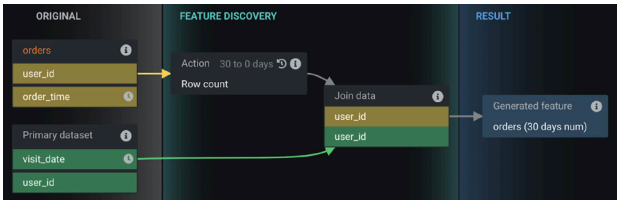


BUILT-IN AWARENESS OF TIME

DataRobot Feature Discovery is fully time-aware. If your datasets are temporal in nature, you can set derivation windows to control how much history should be used when calculating new features. For example, you can tell DataRobot to consider only 30 days of order history when predicting whether or not a customer will make a purchase when he visits a store. Feature Discovery also has built-in guardrails that avoid common leakage problems, such as ensuring that future data is excluded when generating new features.

PRACTICAL, EXPLAINABLE, AND TRACEABLE

Like every automated capability in the DataRobot AI platform, Feature Discovery is incredibly transparent. You can visualize and explore every feature generated to understand predictive potential. Full lineage is also available for every feature and is created for traceability and auditing purposes. You can access detailed logs to understand exactly which features were generated, explored, tested, and discarded. You can also download the full training dataset, with all the new derived features, for further analysis and use in other applications.



About DataRobot and Snowflake Partnership

Snowflake and DataRobot together offer an end-to-end AI experience that accelerates time to value by reducing complexity and removing the delay between data and AI insights.

Specifically, Snowflake's Data Cloud breaks down data silos, allowing users to work with any and all of their data, without limits on scale, performance or flexibility, and grants instant access to third-party data via the Snowflake Data Marketplace. DataRobot takes advantage of this seamless access to organized and documented data to massively accelerate the model development lifecycle, enabling the creation of trusted and scalable models across the business, ultimately driving a significant competitive advantage.

The combination of Snowflake and DataRobot accelerates the journey to become AI-driven.

Contact Us

DataRobot
225 Franklin Street, 13th Floor
Boston, MA 02110, USA

www.datarobot.com
info@datarobot.com