AutoML (V8.0)

March 14, 2022

The DataRobot v8.0.0 release includes many new AutoML features and enhancements described in this section. See also the new features described in the time series (AutoTS) and MLOps release notes.

Release v8.0 provides updated UI string translations for the following languages:

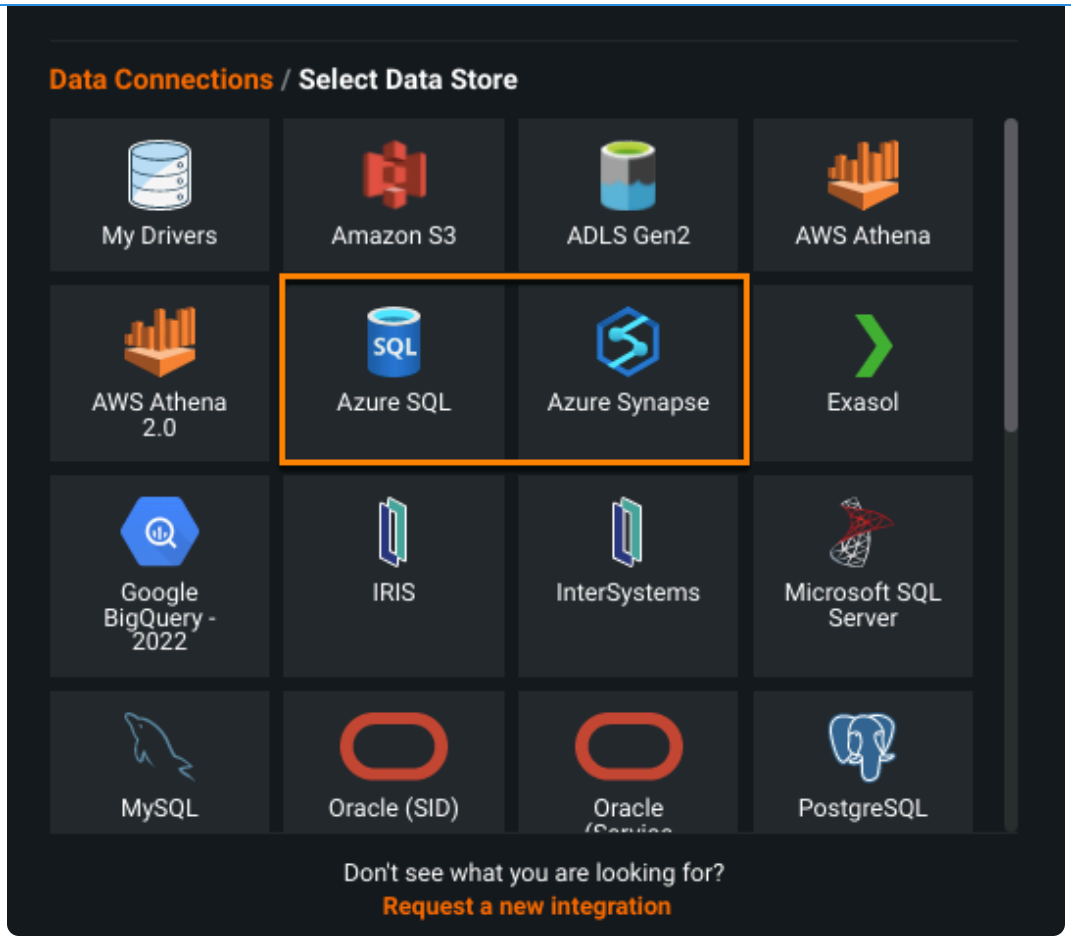
- Japanese
- French
- Spanish
- Korean

See these important deprecation announcements for information about changes to DataRobot's support for older, expiring functionality. This document also describes DataRobot's fixed issues.

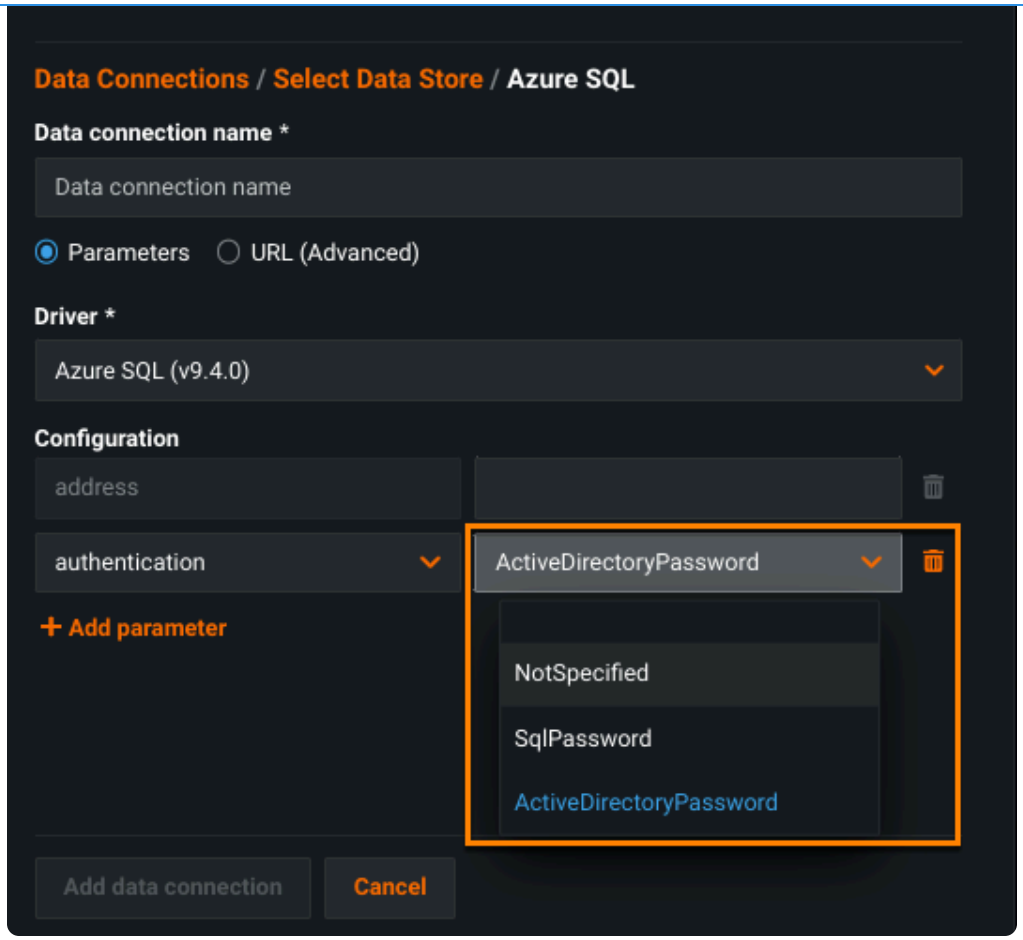
Data enhancements

Active Directory support added for Azure Synapse and SQL

DataRobot now supports Microsoft Azure Synapse and Azure SQL as a data source. When adding a new data connection, both tiles will be listed among the available stores.

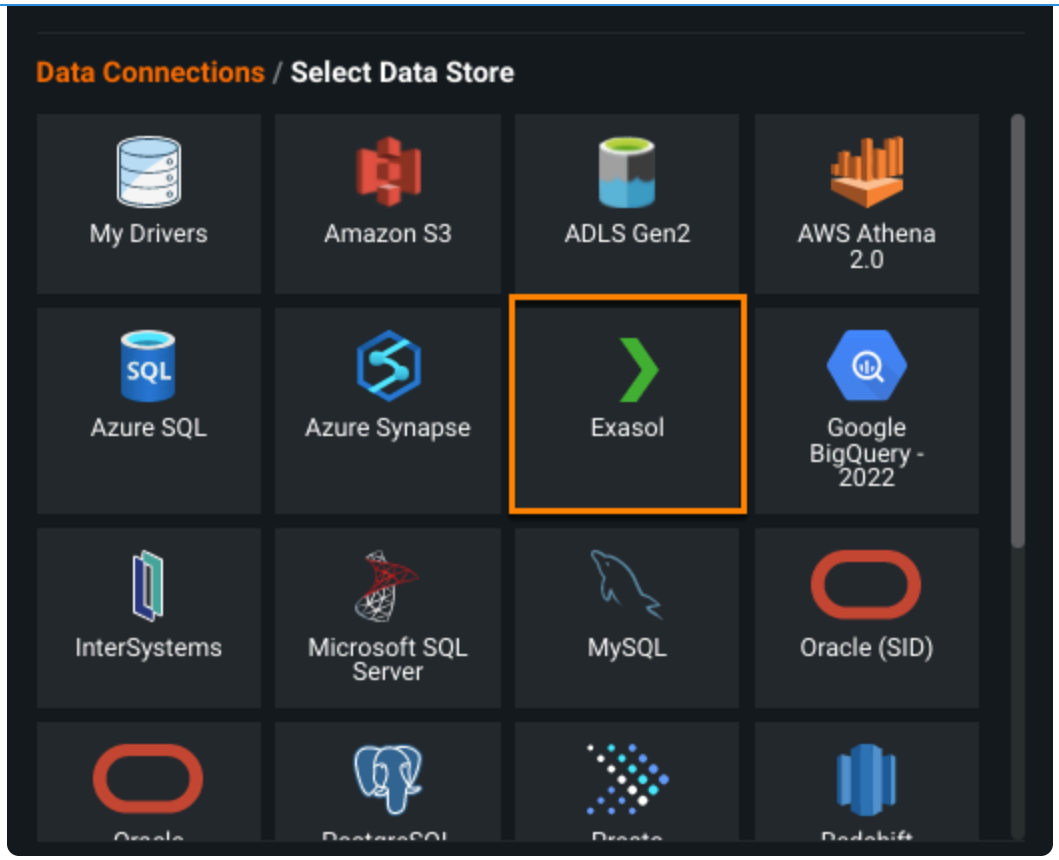


When defining the parameters of the connection, you can specify the authentication method as **SqlPassword** or **ActiveDirectoryPassword**. Selecting ActiveDirectoryPassword allows you to use your Azure identity instead of credentials defined in the database. For information on Active Directory, see the client setup requirements.



Exasol JDBC driver supported in DataRobot and batch predictions

DataRobot now supports the latest version of the Exasol JDBC Driver as a data source. When adding a new data connection, an Exasol tile will be listed among the available stores. After the data connection is set up, you can create batch prediction jobs that score to and from your Exasol database.



Google Cloud and Azure Storage SDK upgraded for improved reliability

Storage is a fundamental part of the DataRobot infrastructure because it is used to store datasets, models, insights, etc. To keep the storage subsystem reliable and performant, DataRobot upgraded Google Cloud and Azure Storage SDK versions for Self-Managed AI Platform installations.

Verified and updated list of data sources for batch predictions

DataRobot verified existing data sources that support batch predictions and added support for new data sources. See Data sources supported for

Feature Discovery features

Improvements to Feature Discovery feature derivation process

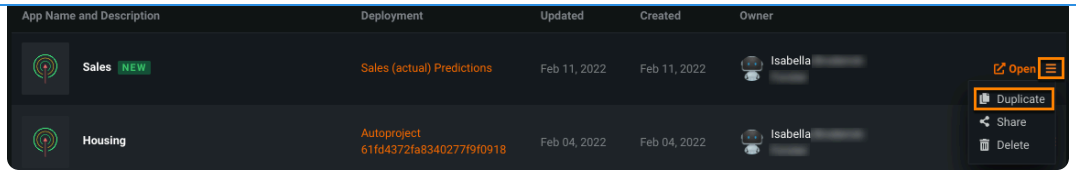
DataRobot reduced the likelihood of not generating features after defining relationships when the feature derivation window (FDW) is too large, causing the computation to be too complex, or when the column is set as both the time-index and join column.

Modeling features

Duplicate applications in the App Builder

With this release, you can create a copy of an existing application, so new users can leverage the existing work without having to spend the time and effort recreating every aspect of the app's charts, predictions, scenarios, simulations, and more. When an application is shared, any changes made by the new user affect the original owner's application; however, you can now duplicate an application and share the copy—allowing new users to access pre-existing content without disrupting or changing the work of the original owner's app.

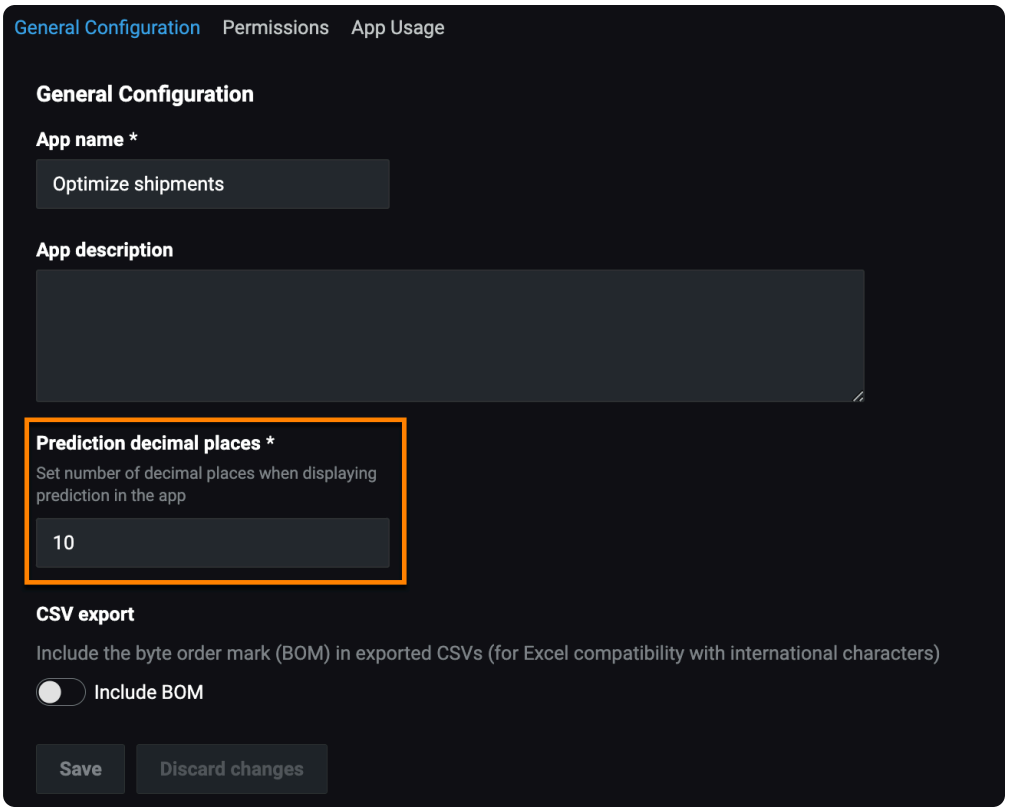
To duplicate an application, go to **Applications > Current Applications**. Click the menu icon of the app you want to copy and select **Duplicate**. For more information, see Duplicate applications.



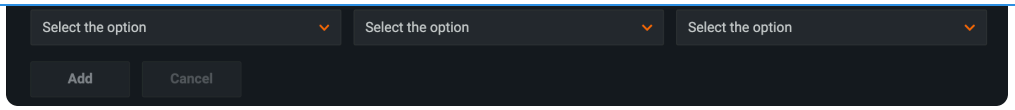
Improvements to the no code App Builder

This release introduces several improvements to the no code App Builder:

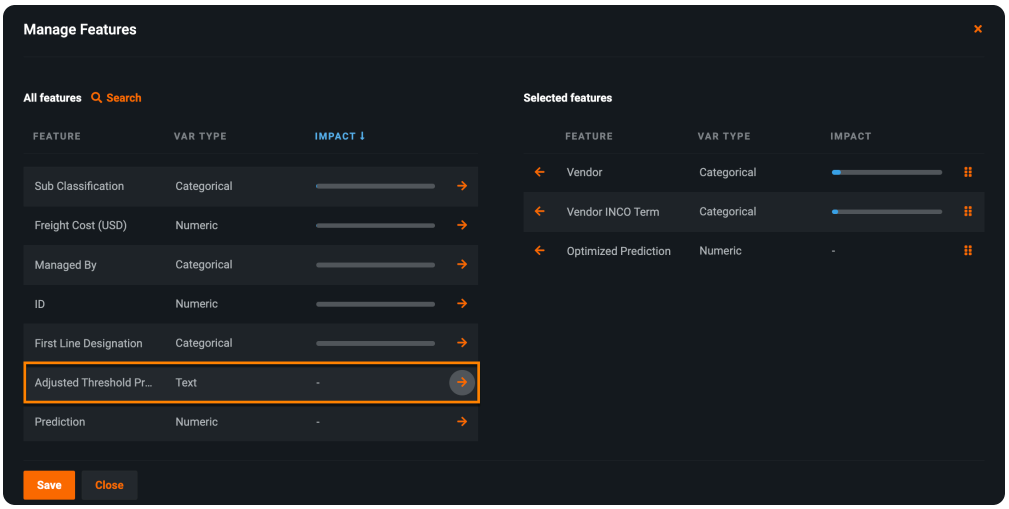
- In an application's **Settings**, you can now specify the number of decimal places to display for predictions throughout an application.



- When making single record predictions, click **Populate averages** to enter the average feature value for each visible field.



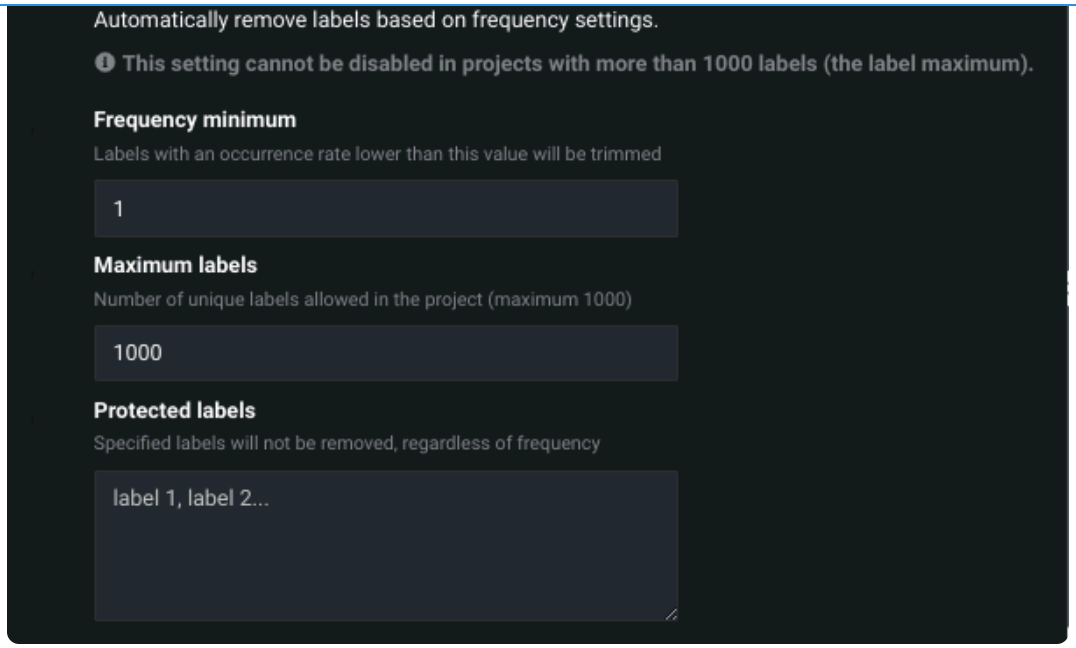
- You can now add an Adjusted Prediction Threshold column to the All Rows widget for binary classification projects. To add this column, go to **Build** mode, select the **All Rows** widget, and click **Manage** on the left. Click the orange arrow next to **Adjusted Prediction Threshold** and click **Save**.



- Additional date formats are now supported in applications.

Support for unlimited labels goes GA

This release brings enhanced support for multicategorical targets, now allowing any number of labels ("unlimited multilabel modeling"). Previously, projects were limited to 100 labels. When DataRobot builds multilabel projects, it uses up to 1,000 labels in each multicategorical feature. You can either allow the application to trim extraneous labels or you can specify which labels to trim in the Feature Constraints section of advanced options. Additionally, export of labelwise Lift Charts via **Predict → Download** is now enabled.

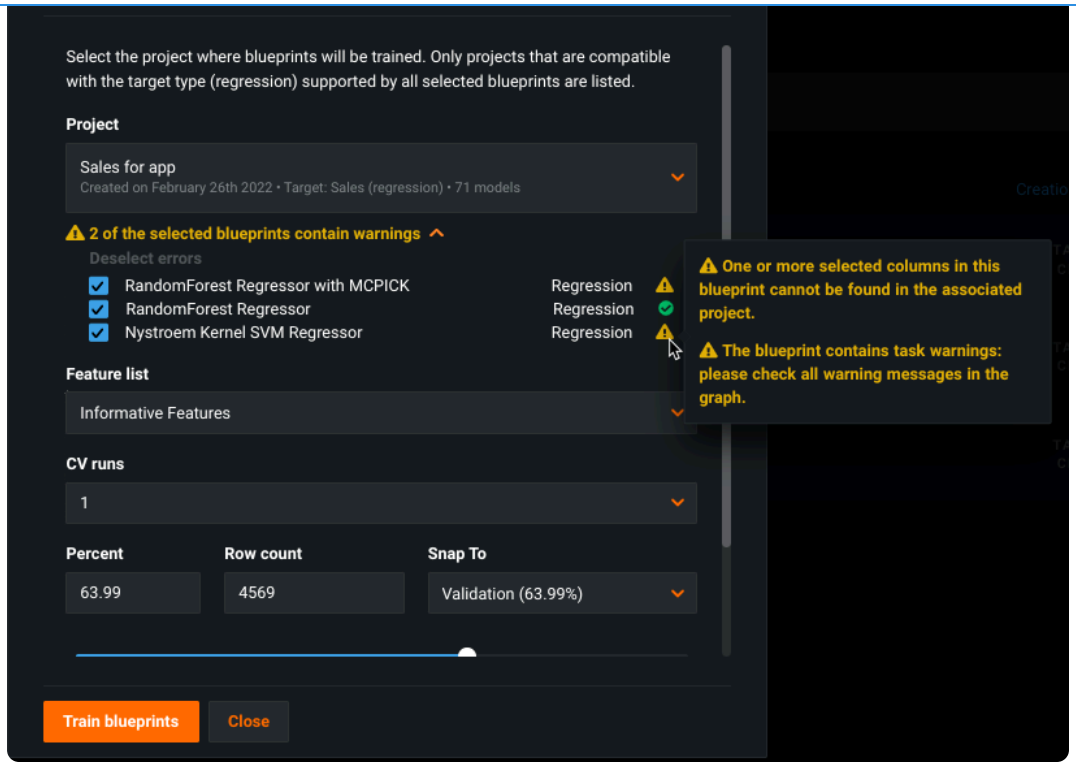


Note

Availability of multilabel modeling is dependent on your DataRobot package. If it is not enabled for your organization, contact your DataRobot representative for more information.

Improved blueprint handling in the AI Catalog

The ability to simultaneously train multiple blueprints (“bulk train”) from the AI Catalog has been improved by helping to identify errored blueprints. Now, the **Train multiple blueprints** modal displays a color-coded message that indicates status for the group of blueprints in the training request and the number of affected blueprints. You can hover over an error or warning to display a tooltip containing additional information.



Other usability improvements include:

- A new bulk delete feature allows you to select multiple blueprints for deletion and confirm, via modal, the specific blueprints to ensure an accidental deletion does not occur.
- When selecting blueprints from the AI Catalog, your selections persist as you page through the inventory. From any page, you can apply bulk actions such as training, validating, or deletion.

Blueprint editor enhancements

This release brings improvements to the blueprint editor used for composable ML.

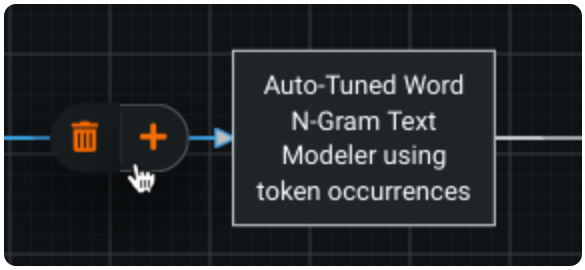
Add and edit blueprint objects

you needed to click a node to access actions for the node. Now, you only need to hover over a node to access the actions. You can now perform actions directly on connectors, as well—a more intuitive approach.

Hover over a node to access the actions described below:

ACTION ICON	DESCRIPTION
	Modify a node.
	Add a node.
	Connect nodes.
	Remove a node. Removing nodes removes downstream nodes.

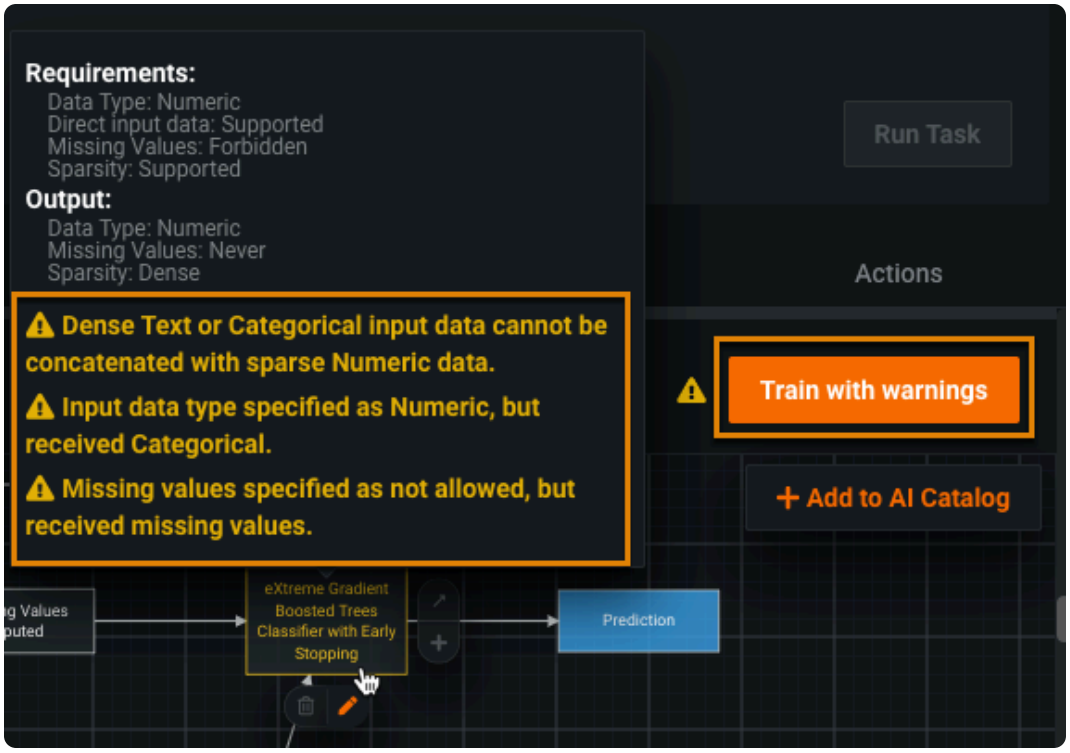
Hover over a connector to access the actions described below.



	Add a node.
	Remove a connector.

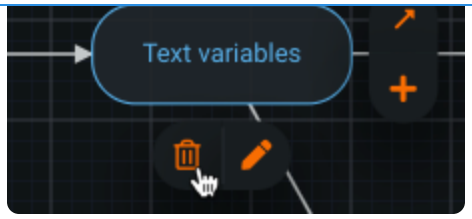
Blueprint validation

Blueprint validation has also been enhanced. When you hover over a node that contains warnings (highlighted in yellow), the warning messages display. You can now train on blueprints that contain warnings. To do so, click **Train with warnings**.

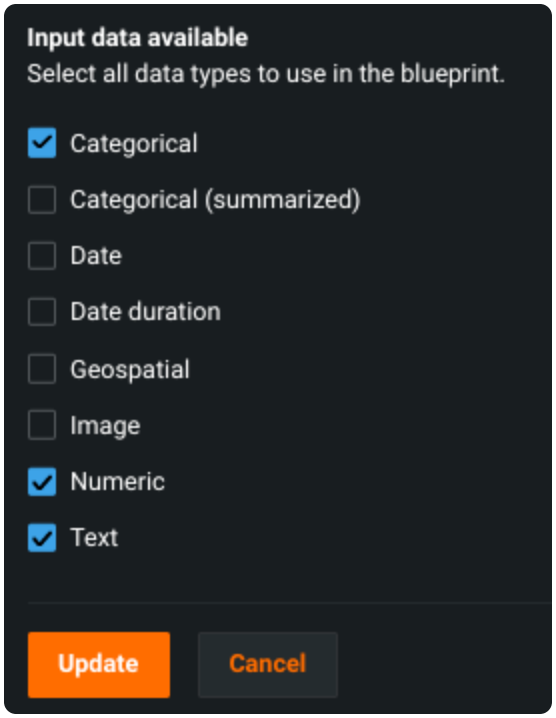


Remove data type nodes

You can now remove data type nodes directly.



In past releases, you needed to clear check boxes in the **Input data available** window to remove data type nodes.



See the documentation on modifying blueprints for details.

NLP Fine-Tuner blueprints for multi-modal datasets in any language

Natural language processing (NLP) deals with the interaction between computers and humans using the natural language and is essential for every AutoML system. Fine-tuning is a process that takes a model that has already been trained for a given task, and makes it perform a second

the first.

NLP Fine-Tuner blueprints allow you to use a model previously trained for NLP and fine-tune them, similar to existing functionality in Visual AI. Doing so increases accuracy, and lets you adjust models to a specific use case and downstream task. NLP Fine-Tuner blueprints are available in any language for multi-modal datasets, multilabel datasets, and Composable ML.

Improvements to External Predictions insights

You can now configure up to 100 external prediction column names in the **External Predictions** tab of advanced options.

Admin enhancements

Python 3 support for Hadoop clusters

Following on from our announcement regarding Python 2 deprecation, support for Python 3 on Hadoop clusters is now available. New installations will use Python 3 for projects and models. Those upgrading from previous releases will see support for both Python 2 and Python 3 side-by-side so that pre-existing projects will continue to function as expected. See the deprecation notice and Python 3 migration guide for more information. The following deprecated features are not supported with Python 3 projects, but Python 2 projects containing these features will work as expected:

- Scaleout models
- Hadoop Scoring

DataRobot now officially supports Red Hat Enterprise Linux 8.4 (RHEL 8.4) and 8.5 (RHEL 8.5) as installation targets. Additionally, CentOS Linux 8 has reached End of Life (EOL) as of December 31st, 2021 and is no longer supported.

Account and profile settings reorganized

To improve the user account management experience, the **Profile** page now includes the following tabs for individual user preferences, including the settings previously located on the **Settings** page:

TAB	SETTINGS
Account	The original **Profile** page settings.
Security	The following individual user security settings: • Change Your Password • Two-Factor Authentication
System	The following individual user system settings: • Language • Theme • CSV export
Notifications	The following individual user notification settings: • Mute all email notifications • Enable email notification when Autopilot has finished • Enable browser notification when Autopilot has finished

settings on the **Settings** page. For all other users, the **Settings** page is deprecated.

Skip the DataRobot login when SAML is enforced

If your organization has SAML authentication enforced, you can now bypass the DataRobot login screen, automatically redirecting users to the SAML login page from the application url. To skip the login screen, set the following configuration setting to `TRUE` in your config.yaml: `SKIP_LOGIN_UI_IF_SAML_SSO_IS_ENFORCED`.

API Enhancements

The following is a summary of API new features and enhancements. Go to the API Documentation home for more information on each client.

> **Tip**
>
> DataRobot highly recommends updating to the latest API client for Python and R.

New Features

API release v2.28.0 introduces new routes for computing and retrieving samples for Image Augmentation Lists:

- `POST /api/v2/imageAugmentationLists/(augmentationId)/samples/`
- `GET /api/v2/imageAugmentationLists/(augmentationId)/samples/`

Enhancements

date:

- `GET /api/v2/projects/(projectId)/models/(modelId)/clusterInsights/`

New properties have been added to a leaderboard item: `bias_mitigation` and `bias_mitigation_parent_lid`.

- `GET /api/v2/projects/(projectId)/models/(modelId)/`

API deprecation notices

The `customModelType` parameter is now deprecated in the following routes. It will be removed completely in a later release.

- `POST /api/v2/customModels/`
  - This endpoint only creates custom inference models.
  - To create a Custom Training Task (`customModelType=training`) use the dedicated customTasks endpoint `POST /api/v2/customTasks/`.
- `GET /api/v2/customModels/`
  - This endpoint only lists custom inference models.
  - To list custom training tasks (`customModelType=training`) use the dedicated customTasks endpoint `GET /api/v2/customTasks/`.

Routes for Image Augmentation Samples not related to Image Augmentation Lists are deprecated and will be removed in the following routes:

- `POST /api/v2/imageAugmentationSamples/`

augmentation list and generate samples for it using the endpoint `POST /api/v2/imageAugmentationLists/(augmentationId)/samples/`.

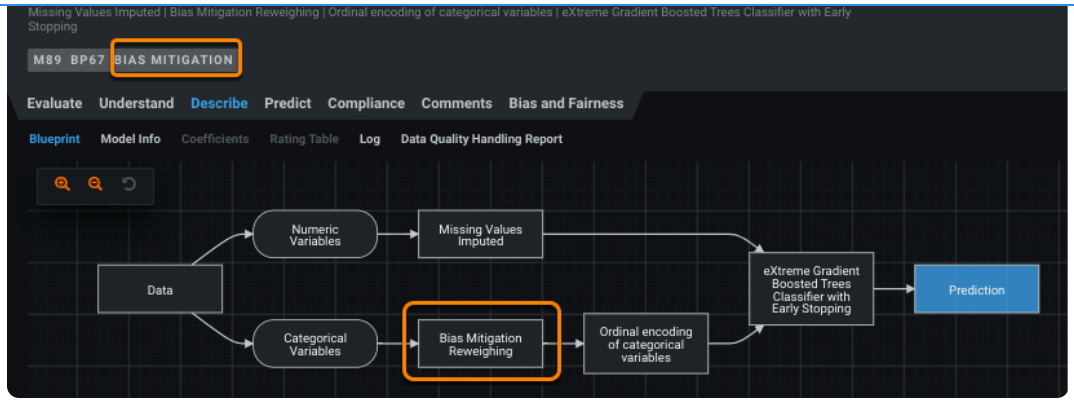
- `GET /api/v2/imageAugmentationSamples/(samplesId)/`
  - To retrieve image augmentation samples retrieve them using the endpoint `GET /api/v2/imageAugmentationLists/(augmentationId)/samples/`.

Preview features

Bias mitigation now available for binary classification projects

Bias mitigation, a technique to mitigate Leaderboard models for biased behavior, is now available as a preview feature. It works by augmenting blueprints with a pre- or post-processing task causing the blueprint to then attempt to reduce bias across classes in a protected feature. You can apply mitigation either automatically (as part of Autopilot) or manually (after Autopilot completes). When run automatically, you set mitigation criteria as a part of the Bias and Fairness advanced option settings. Autopilot then applies mitigation to the top three Leaderboard models. Or, once Autopilot completes, you can apply mitigation to any non-blender, unmitigated model available from the Leaderboard. Finally, compare mitigated versus unmitigated models from the Bias vs Accuracy insight.

For more information, see the documentation.



Deprecation notices

Note the following to better plan for later migration to new releases.

TensorFlow blueprints deprecated and soon to be removed

TensorFlow (TF) blueprints are being deprecated with this release, making them unavailable for building in new projects. They are being replaced by Keras blueprints, which in most cases outperform TF for both speed and accuracy. TF blueprints built as part of an existing project will still function normally. These blueprints are no longer searchable in user blueprints, either new or existing.

Feature Fit visualization deprecated in favor of Feature Effects

The **Feature Fit** visualization, available from the Leaderboard under the **Evaluate** tab, is deprecated and will soon be removed. For insight on a feature's impact on model predictions, use **Understand > Feature Effects** instead. Both visualizations report a feature's model-agnostic importance. Feature Fit, calculated during EDA2, charted results based on a feature's importance score. This score is still available from the **Data** page. Feature

to a feature's value would effect a model's predictions.

Feature Fit will be removed in on-premise release 9.0.0. For managed AI Platform users, Feature Fit will be removed within the next quarter.

Customer-reported fixed issues

The following issues have been fixed since release 7.3.5.

Platform

- EP-2285: Testing facts failed to set target for mongo version.
- MODEL-8321: Fixes an internal service error (ISE) when selecting a character level analyzer along with any tokenization method besides `None`.
- VIZAI-3055: Removes the ability to create multilabel projects with OTV or TS that are not supported for multilabel project types via the API.
- VIZAI-3062: Enables Feature Discovery for multilabel projects.

Predictions

- PRED-7153: Fixes an issue with frozen models causing the **Predict** tab on the Leaderboard to not render properly.
- PRED-7191: Fixes an issue with the **Make Predictions** tab on the Leaderboard when attempting to use a derived features as an optional pass-through column.