

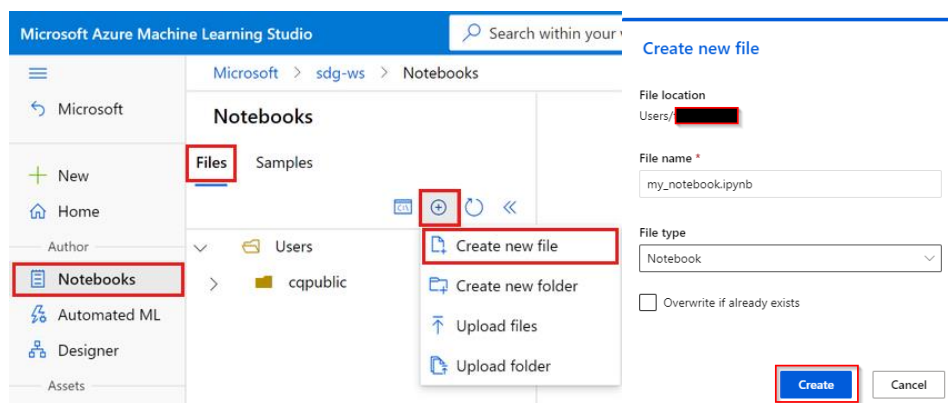
Azure Machine Learning (ML) is a platform for operating machine learning workloads in the cloud. With the Azure ML Studio, you work in a machine learning workspace. A workspace defines the boundary for a set of related machine learning assets. You can use workspaces to group machine learning assets based on projects, deployment environments (for example, test and production), teams, or some other organizing principle.

1. Notebook features

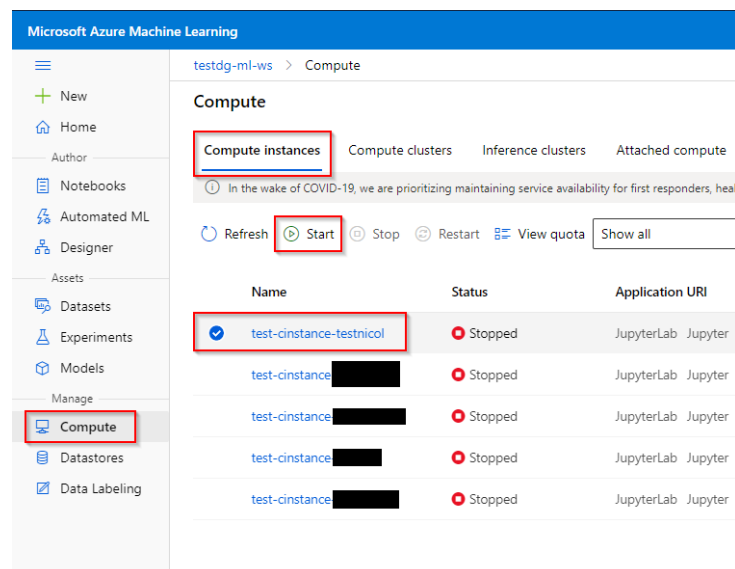
Azure ML Notebooks are Jupyter Notebooks with Python and R kernels integrated with Azure ML Studio. You can run your explorative analysis, pre-processing and post-processing actions on these Jupyter Notebooks using an user-specific compute instance (a personal Virtual Machine).

Although you can train small machine learning models on the user-specific compute instance, when higher computational power is needed, we recommend to use the compute cluster (a scalable compute resource) provided to you and connected to the workspace.

- **Creating a Notebook**



- **Starting/Stopping compute instance**



2. Data provision

Datasets are versioned packaged data objects that can be easily consumed in experiments. Datasets are the recommended way to work with data, and are the primary mechanism for advanced Azure Machine Learning capabilities like data labeling and data drift monitoring.

- **Registering a dataset from (a local file – Data Lake GEN 2 – a SQL Database)**

A registered dataset is a dataset that is registered in the workspace of the Azure ML Studio. A registered dataset can be referenced, either graphically or programmatically, when creating an experimental run using a Machine Learning model.

- types of datasets: Tabular (e.g. csv-file) and File (e.g. images)

Microsoft Azure Machine Learning

ml_workspace_test > Datasets

Create dataset from datastore

Datasets

Registered datasets Dataset monitors (pn)

+ Create dataset Refresh ☆ Unre

Version

1

Basic info

Name * diabetes_dataset Dataset version 1

Dataset type * Tabular

Description Dataset description

Create dataset from local files

Datastore and file selection

Select or create a datastore *

☒ Currently selected datastore: workspaceblobstore (Azure Blob Storage) (Default)

☐ Previously created datastore

☐ Create new datastore

Select files for your dataset *

After dataset creation, these files will be uploaded to your default Blob storage and made available in your workspace. Supported file types include: delimited (i.e. csv, tsv), Parquet, JSON Lines, and plain text.

Browse

File name	Size (MiB)	Upload %	Status
diabetes (1).csv	0.5033	100	✓

Upload path

UI Files will be uploaded to '\$(Upload path)/11-06-2020_042058_UTC'

☐ Skip data validation ⓘ

Create dataset from datastore

Basic info

Datastore selection

Settings and preview

Schema

Confirm details

Datastore selection

Select or create a datastore *

Currently selected datastore: azuredatalakestorage (Azure Data Lake Storage Gen2) (Default)

Previously created datastore

Create new datastore

Path *

diabetes.csv

To include files in subfolders, append "/" after the folder name like so: "{Folder}/".

☐ Skip data validation ⓘ

Advanced settings

Partition format

/[column_name]/[time_column_name:yyyy/MM/dd]/*. *

☐ Ignore unmatched file paths

Create dataset from datastore

Basic info

Datastore selection

Settings and preview

Schema

Confirm details

Datastore selection

Select or create a datastore *

Currently selected datastore: sqldatabase (Azure SQL database)

Previously created datastore

Create new datastore

SQL query *

SELECT * FROM diabetes

☒ Skip data validation ⓘ

Advanced settings

- File format: delimited, plain text, parquet file or JSON Lines

Create dataset from datastore

Basic info

Datastore selection

Settings and preview

Schema

Confirm details

Settings and preview

These settings were automatically detected. Please verify that the selections were made correctly or update

File format

Delimited

Delimiter

Comma

Example

Field1.Field2.Field3

Encoding

UTF-8

Column headers

Use headers from the first file

Skip rows

None

	Id	PatientID	Pregnancies	PlasmaGlucose	DiastolicBloodPr...	T
1		1354778	0	171	80	34
2		1147438	8	92	93	47
3		1640031	7	115	47	52
4		1883350	9	103	78	25
5		1424119	1	85	59	27
6		1619297	0	82	92	9
7		1660149	0	133	47	19
8		1458769	0	67	87	43
9		1201647	8	80	95	33
10		1403912	1	72	31	40
11		1012922	1	99	97	36

Back

Next

Create dataset from datastore

Basic info

Datastore selection

Settings and preview

Schema

Confirm details

Schema

Include	Column name	Properties	Type	Format settings and example
<input type="checkbox"/>	Path	Not applicable to selected type	String	
<input checked="" type="checkbox"/>	PatientID	Not applicable to selected type	Integer	1354778, 1147438, 1640031
<input checked="" type="checkbox"/>	Pregnancies	Not applicable to selected type	Integer	0, 8, 7
<input checked="" type="checkbox"/>	PlasmaGlucose	Not applicable to selected type	Integer	171, 92, 115
<input checked="" type="checkbox"/>	DiastolicBloodPressure	Not applicable to selected type	Integer	80, 93, 47
<input checked="" type="checkbox"/>	TricepsThickness	Not applicable to selected type	Integer	34, 47, 52
<input checked="" type="checkbox"/>	SerumInsulin	Not applicable to selected type	Integer	23, 36, 35
<input checked="" type="checkbox"/>	BMI	Not applicable to selected type	Decimal	43.50972593, 21.24057571, 41.5...
<input checked="" type="checkbox"/>	DiabetesPedigree	Not applicable to selected type	Decimal	1.213191354, 0.158364981, 0.07...
<input checked="" type="checkbox"/>	Age	Not applicable to selected type	Integer	21, 23, 23
<input checked="" type="checkbox"/>	Diabetic	Not applicable to selected type	Integer	0, 0, 0

Back

Next

Cancel

Create dataset from local files

Basic info

Datastore and file selection

Settings and preview

Schema

Confirm details

Confirm details

Basic info

Name
test

Dataset version
1

Dataset type
Tabular

File settings

File format
Delimited

Delimiter
Comma

Encoding
UTF-8

Column headers
No headers

Skip rows
None

Datastore and file selection

Datastore
workspaceblobstore

Selected files (1)
diabetes (1).csv

Path
UI/11-06-2020_042058_UTC/diabetes (1).csv

☐ Profile this dataset after creation

Back

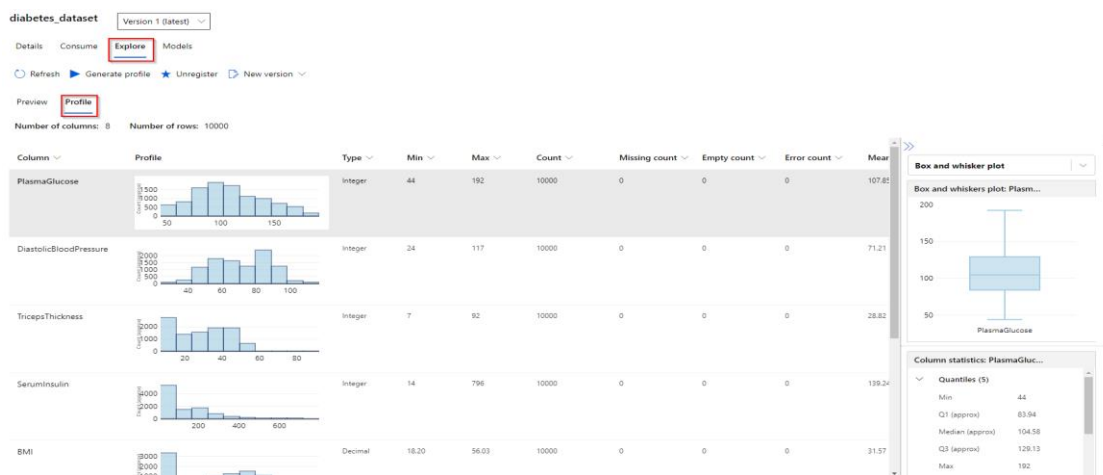
Create

- **Viewing the dataset**

After a dataset is registered, you can obtain information from this dataset in the workspace.

The screenshot shows the Microsoft Azure Machine Learning interface. On the left is a navigation pane with options like New, Home, Notebooks, Automated ML, Designer, Datasets, Experiments, Pipelines, Models, Endpoints, Compute, Datastores, and Data Labeling. The main area displays the 'diabetes_dataset' details for Version 1 (latest). The 'Details' tab is selected, showing a table of registered datasets with columns Name, Version, and Data source. The 'diabetes_dataset' is listed with Version 1 and Data source 'azuredatalakestorage'. To the right, the 'Attributes' section provides metadata: Properties (Tabular), Description, Created by (TestUser Ze Wen), Datastore (azuredatalakestorage), Relative path (diabetes.csv), Profile run dataset_01950805-fb6c-41b8-af56-3483d34d7b84, Files in dataset (1), Total size of files in dataset (515.4 KiB), Current version (1), Latest version (1), Created time (Oct 20, 2020 5:02 PM), and Modified time (Oct 20, 2020 5:02 PM).

- Simple characteristics of your attributes: histogram distribution, type, min-max value, mean, standard deviation



- **Unregister a dataset**

The screenshot shows the Microsoft Azure Machine Learning interface with the 'Unregister dataset' dialog box open. The dialog has a title bar 'Unregister dataset' and a close button. The main text reads: 'You have requested to unregister the dataset: **diabetes_dataset**'. Below this, it states: 'Unregistering a dataset will permanently delete the metadata (description and tags) associated with all versions of this dataset. The underlying storage and its data will not be deleted. Are you sure you wish to unregister this dataset?'. At the bottom right, there are two buttons: 'Unregister' (highlighted with a red box) and 'Cancel'.

3. Model training

This section provides user instructions on when training a machine learning model by using a registered dataset. There are three ways to construct and train machine learning models in the workspace:

- Training a fast ML model set using the **AutoML feature**.
- Pre-process, train and post-process your data using the drag-and-drop **Designer feature**.
- you can train a custom machine learning model using Python/R script, using the **Notebook feature**.

3.1 AutoML

Automated Machine Learning (AutoML) enables you to try multiple algorithms and preprocessing transformations with your data. This, combined with scalable cloud-based compute makes it possible to find the best performing model for your data without the time-consuming manual trial and error that would otherwise be required.

The AutoML feature can handle the following machine learning tasks:

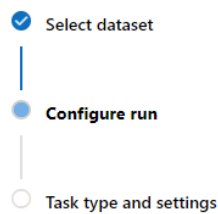
- Classification
- Regression
- Time-series forecasting

AutoML automatically pre-processes the provided data (scaling and normalization). The AutoML feature currently only supports TabularDataset.

The image shows the Microsoft Azure Machine Learning interface. The top navigation bar is blue with the text "Microsoft Azure Machine Learning". Below it, a sidebar contains a menu with options: "New", "Home", "Author", "Notebooks", "Automated ML" (highlighted with a red box), "Designer", "Assets", and "Datasets". The main content area is titled "Automated ML" and shows a "New Automated ML run" button (highlighted with a red box) and a "Refresh" button. Below this, there is a section for "Recent Automated ML runs" with columns for "Run" and "Run ID".

Below the main interface, a dialog titled "Create a new Automated ML run" is shown. It has a progress bar with three steps: "Select dataset" (selected), "Configure run", and "Task type and settings". The "Select dataset" step is expanded, showing a "Select dataset" section. This section includes a "Create dataset" button and a "Show supported datasets only" toggle. Below this, there is a "Dataset name" field with a dropdown menu. The dropdown menu is open, showing two options: "test" and "diabetes_dataset" (highlighted with a red box).

- target column: the column in the dataset that you want the model to predict



Configure run

Configure the experiment. Select from existing experiments or define

Dataset

diabetes dataset ([View dataset](#))

Experiment name *

☒ Select existing ☐ Create new

Existing experiment

diabetes-training

Target column * ⓘ

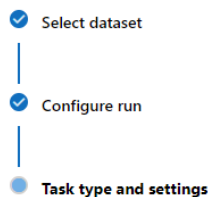
Diabetic

Select compute cluster * ⓘ

aml-cluster

- In the classification category, you can specify deep learning capabilities of AutoML. In the time-series forecasting category, additional information such as time column and forecast horizon need to be provided.

- Additional configuration settings can be provided at the bottom, such as primary metric to train, validation type, exit criteria, etc. Finally, you can also specify to include feature engineering capabilities.



Select task type

Select the machine learning task type for the experiment. Additional settings are available to



Classification

To predict one of several categories in the target column, yes/no, blue, red, green.



Regression

To predict continuous numeric values



Time series forecasting

To predict values based on time

The time series forecasting method requires some additional information.

Time column * ⓘ

Select a time column...

Time series identifier(s) ⓘ

Select column(s)...

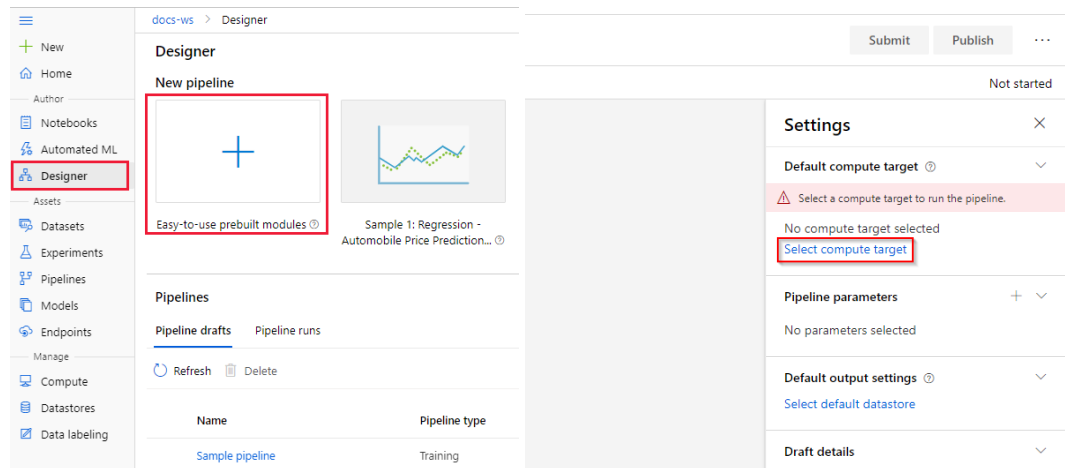
Forecast horizon * ⓘ

☒ Autodetect

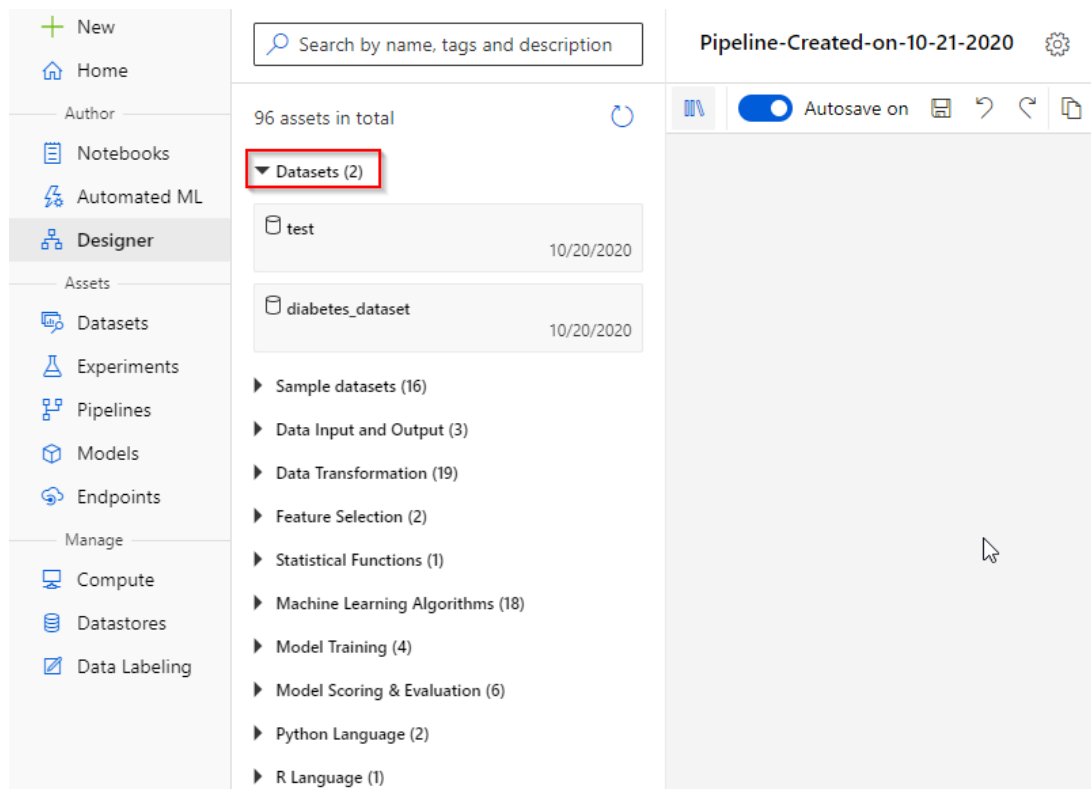
[View additional configuration settings](#) [View featurization settings](#)

3.2 Designer

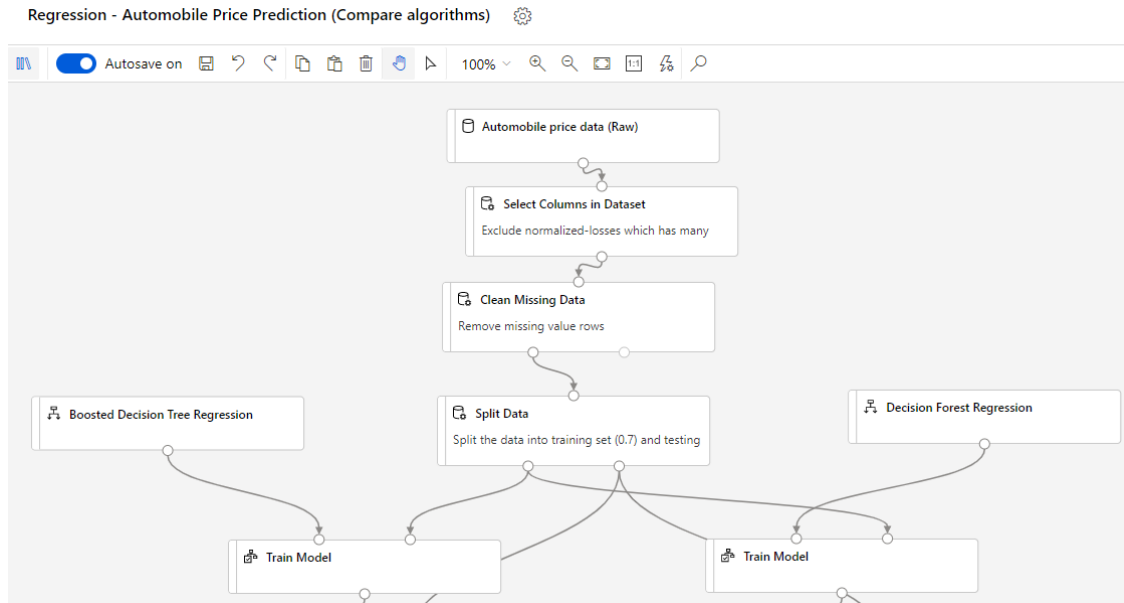
In an enterprise data science process, you'll generally want to separate the overall process into individual tasks, and orchestrate these tasks as pipelines of connected steps. Azure Machine Learning designer lets you visually connect datasets and modules on an interactive canvas to create machine learning models.



- the different assets/building blocks you can use to pre-process, train and post-process your dataset. In the “Datasets” asset, you will see all the registered datasets in the workspace.



- the Automobile Price Prediction template, which compares the results of two different, trained machine learning models on the Automobile price data.



Set up pipeline run

Experiment

☒ Select existing ☐ Create new

Existing experiment *
automobile-exp

Run description *
Regression - Automobile Price Prediction (Compare algorithms)

Compute target
Default aml-cluster

Settings

Default compute target
aml-cluster
[Select compute target](#)

Pipeline parameters
No parameters selected

Default output settings
[Select default datastore](#)

Draft details

Draft name
Regression - Automobile Price Prediction (C...)

Draft description (optional)
This sample shows how to compare two regression models' performance on automobile price prediction.

Microsoft Azure Machine Learning

ml_workspace_test > Designer

Designer

New pipeline

Easy-to-use prebuilt modules

Image Classification using DenseNet

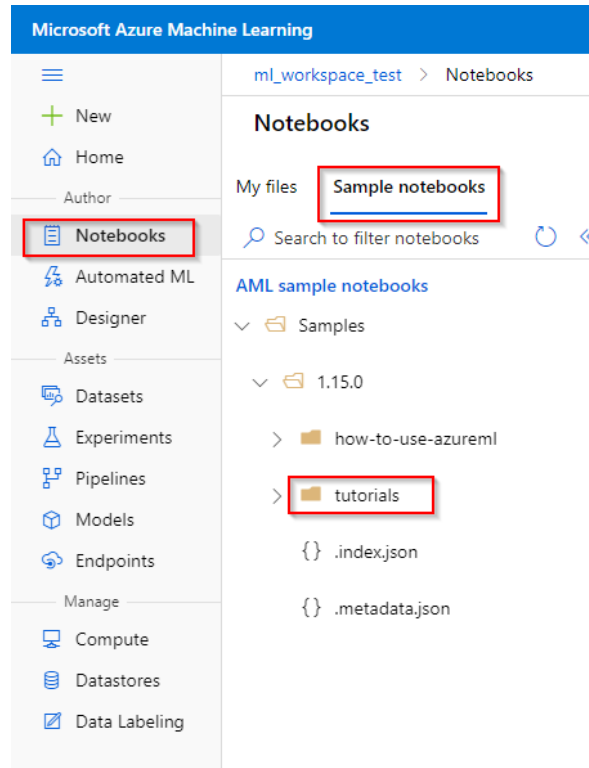
Pipelines

Pipeline runs

Name	Pipeline type
Regression - Automobile Price Prediction...	Training
Pipeline-Created-on-10-21-2020	N/A

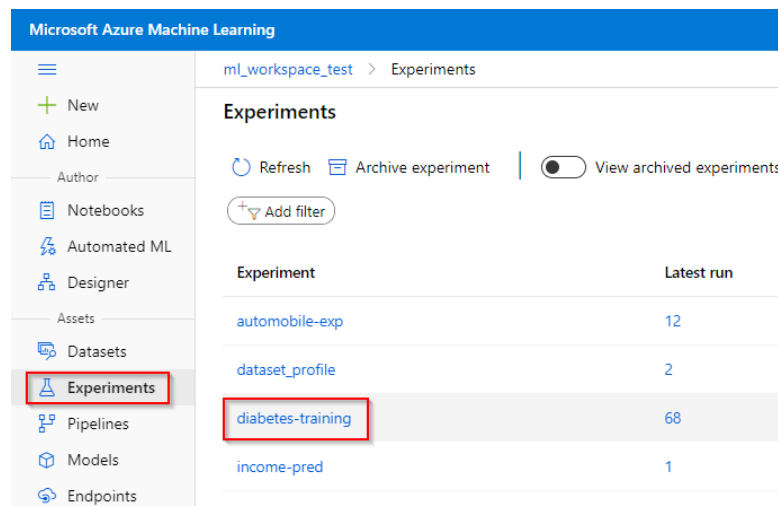
3.3 Notebooks

you can also write and run your own code in managed Jupyter Notebook servers that are directly integrated in the workspace. Microsoft has provided several Machine Learning tutorials on how to connect to the workspace resources and create and run experiments in a programmatic way.

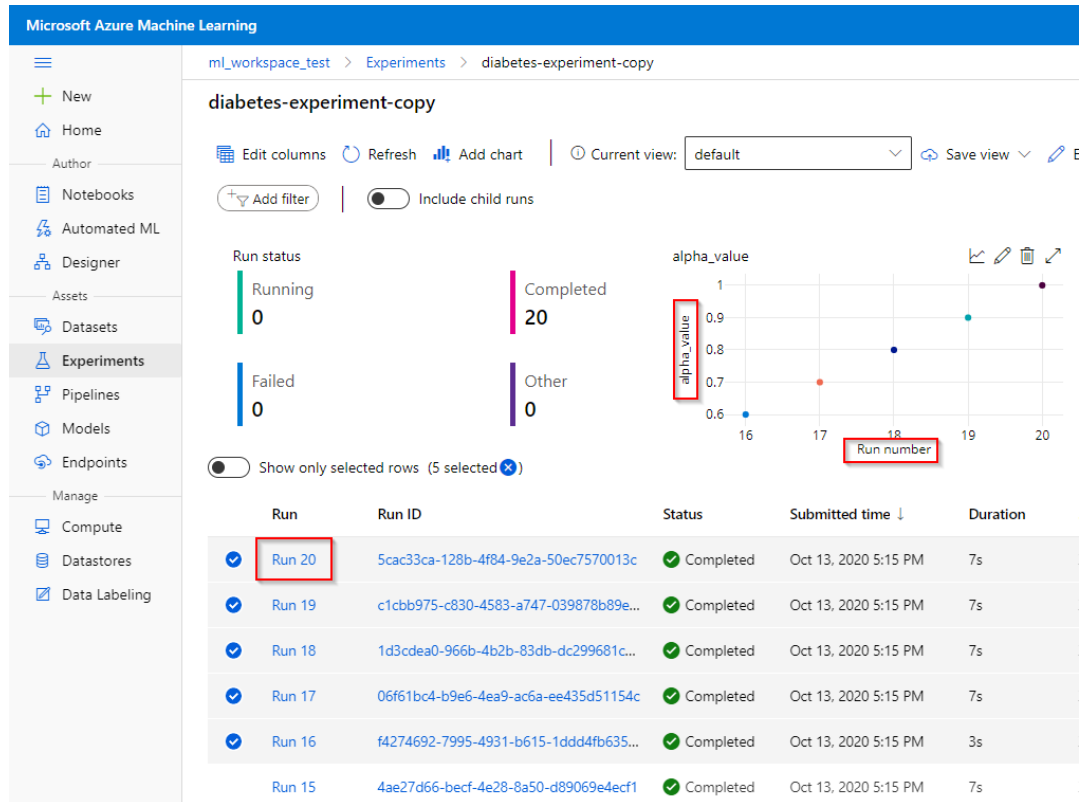


- **Exploring the Experimental runs**

The information on an Azure ML experimental run can be viewed in the “Experiment” tab under the “Assets” category.



In the Azure ML experiments, you can see which runs are still running, failed or completed. If you assigned any custom, logged metrics to your runs (using the Notebook feature), the evolution of these metrics over different runs can be viewed here.



Microsoft Azure Machine Learning

ml_workspace_test > Experiments > diabetes-experiment-copy > Run 20

Run 20 Completed

Refresh Resubmit Cancel

Details Metrics Images Child runs Outputs + logs Snapshot Explanations (preview) Fairness (preview)

Properties

Status
Completed

Created
Oct 13, 2020 5:15 PM

Started
Oct 13, 2020 5:15 PM

Duration
6.978s

Compute target
local

Run ID
5cac33ca-128b-4f84-9e2a-50ec7570013c

Run number
20

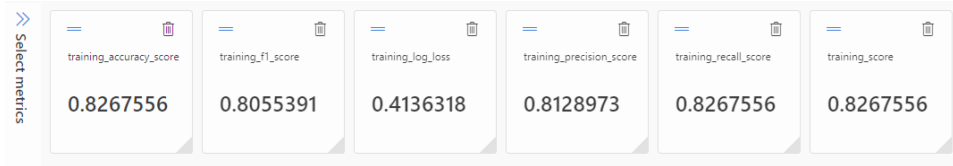
Script name
--

Microsoft > sdg-new-ws > Jobs > Default > GradientBoostingClassifier

GradientBoostingClassifier Completed

Overview **Metrics** Images Child jobs Outputs + logs Code Explanations (preview) Fairness (preview) Monitoring







 Refresh  Cancel  Create custom chart  View as...  Current view: Local  Edit view 







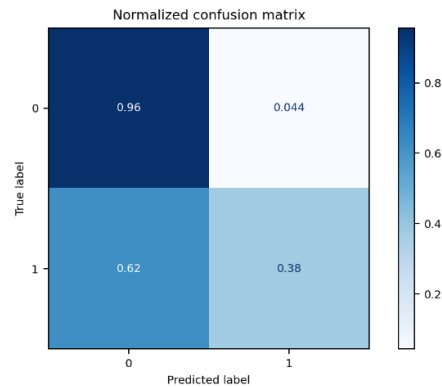
Microsoft > sdg-new-ws > Jobs > Default > GradientBoostingClassifier

GradientBoostingClassifier Completed

Overview Metrics **Images** Child jobs Outputs + logs Code Explanations (preview) Fairness (preview) Monitoring

 Refresh  Connect to compute  Resubmit  Register model  Cancel  Delete

  training_confusion_matrix.png  training_precision_recall_curve.png  training_roc_curve.png




Microsoft Azure Machine Learning

ml_workspace_test > Experiments > diabetes-experiment-copy > Run 20

Run 20 Completed


 Refresh  Resubmit  Cancel  Enable log streaming  Word wrap

Details Metrics Images Child runs **Outputs + logs** Snapshot Explanations (preview) Fairness (preview)

  model_alpha_1.0.pkl

> outputs

 model_alpha_1.0.pkl

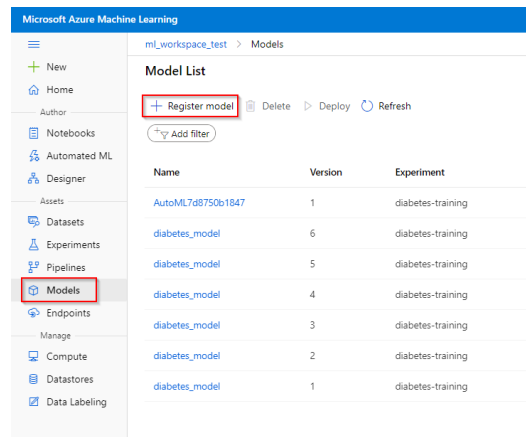
 The file is either binary or uses an unsupported text encoding

4. Model registration

After training the model, you can register it in the workspace from the trained machine learning model runs. Registering a model might be beneficial to manage the machine learning environment of your organization, e.g., be able to reference the model in the Notebook feature.

4.1 Registering a model from registered experimental runs

4.2 Registering a model from a local model file



- Giving a name to your model and specify the model framework. Also, provide the framework version (for keeping track of your models) and model file corresponding to the model framework given (e.g. the.pkl-file). You can also provide tags and properties that are associated with your model.

Register a model

Name *

diabetes_model

Description

Model framework *

Other

Framework name *

Custom

Framework version *

1

Model file or folder *

☒ Upload file ☐ Upload folder

diabetes_model (1).pkl

Browse

Add tags

Name : Value

Add tag

Origin: Intern

Add properties

Name : Value

Add property

Register Cancel

5. Data Labeling

Azure Machine Learning data labeling gives you a central place to create, manage, and monitor labeling projects. Use it to coordinate data, labels, and team members to efficiently manage labeling tasks. Machine Learning supports image classification, either multi-label or multi-class, and object identification with bounded boxes.

Microsoft Azure Machine Learning Studio

Search

This workspace

ML-docs
sdg-new-ws

Microsoft > sdg-new-ws > Data Labeling > Create project

All workspaces

Home

Authoring

Notebooks

Automated ML

Designer

Assets

Data

Jobs

Components

Pipelines

Environments

Models

Endpoints

Manage

Compute

Linked Services

Data Labeling

1 Project details

2 Add workforce optional

3 Select or create data

4 Incremental refresh optional

5 Label categories

6 Labeling instructions optional

7 Quality control (preview) optional

8 ML assisted labeling optional

Project details

1 New feature: To make labeling faster, we've added a new feature to train an ML model while you label. This feature currently supports image or text classification and L...

Project name *

Media type *

☒ Image ☐ Text

Labeling task type *

Image Classification Multi-class

Image Classification Multi-label

Object Identification (Bounding Box)

Instance Segmentation (Polygon)

Apply only a single class from a set of classes to an image

[Learn more](#)

Back Next Cancel

Microsoft Azure Machine Learning Studio

Search

This workspace

ML-docs
sdg-new-ws

Microsoft > sdg-new-ws > Data Labeling > Create project

All workspaces

Home

Authoring

Notebooks

Automated ML

Designer

Assets

Data

Jobs

Components

Pipelines

Environments

Models

Endpoints

Manage

Compute

Linked Services

Data Labeling

1 Project details

2 Add workforce optional

3 Select or create data

4 Incremental refresh optional

5 Label categories

6 Labeling instructions optional

7 Quality control (preview) optional

8 ML assisted labeling optional

Project details

1 New feature: To make labeling faster, we've added a new feature to train an ML model while you label. This feature currently supports image or text classification and L...

Project name *

Media type *

☐ Image ☒ Text

Labeling task type *

Text Classification Multi-class

Text Classification Multi-label

Text Named Entity Recognition

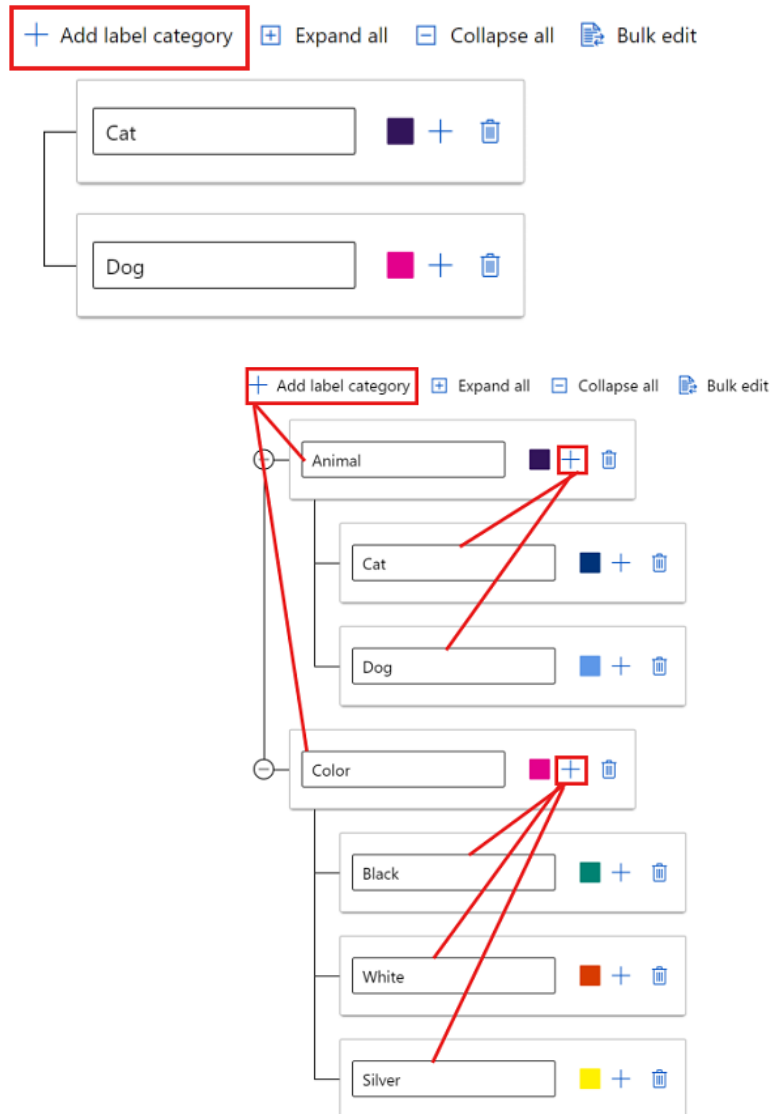
Apply only a single label from a set of classes to a piece of text

[Learn more](#)

Back Next Cancel

Label categories

Enter a hierarchy of labels. You can group labels into categories and add subclasses for each label category.

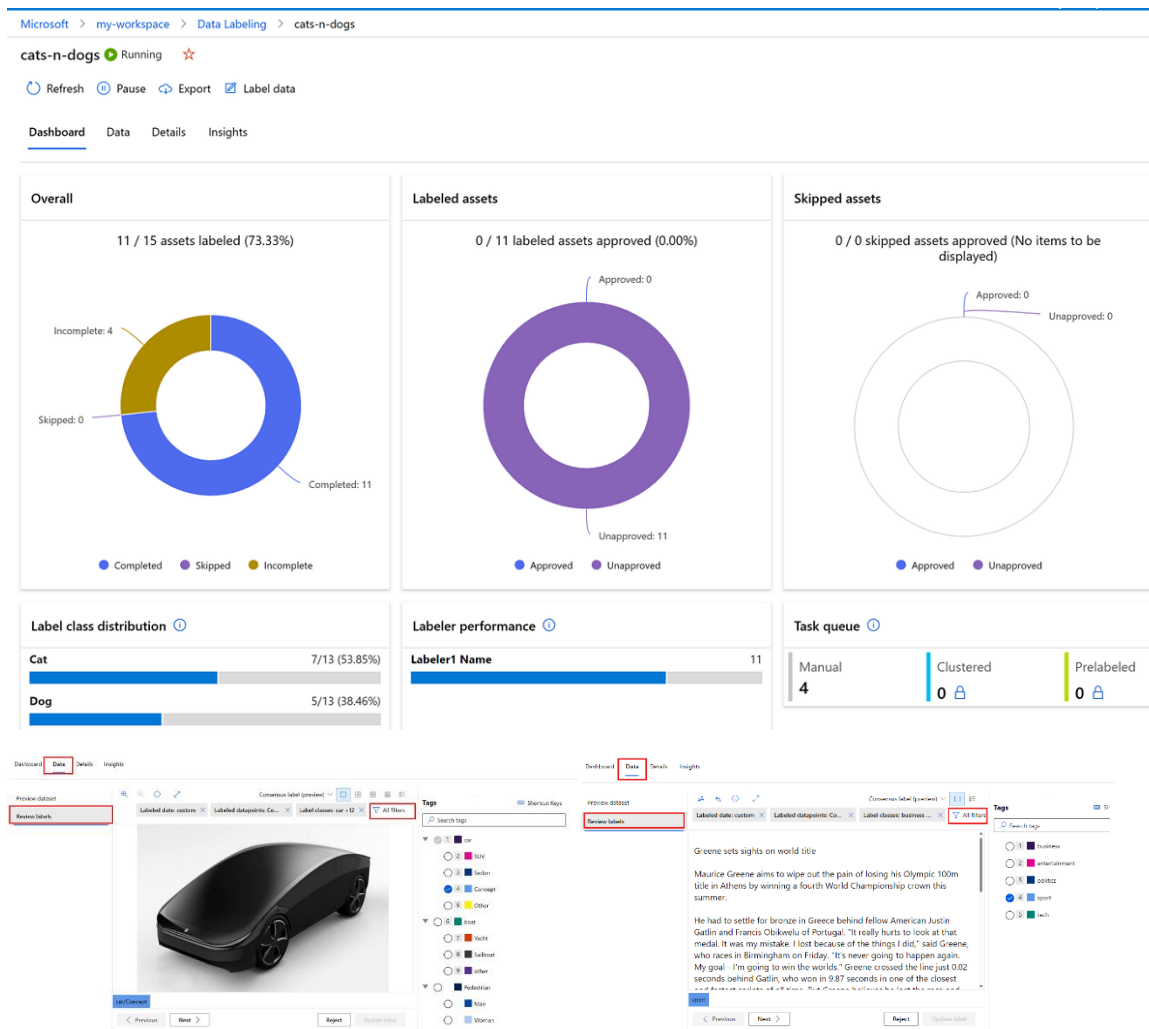


- To accelerate labeling tasks, on the ML assisted labeling page, you can trigger automatic machine learning models. For text, Machine learning (ML)-assisted labeling can handle both file (.txt) and tabular (.csv) text data inputs.

ML-assisted labeling consists of two phases:

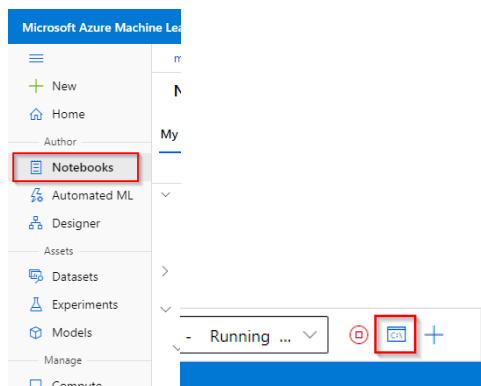
- Clustering
- Pre-labeling

- The clustering phase doesn't appear for object detection models or text classification.



6. Git integration

Azure Machine Learning fully supports Git repositories for tracking work - you can clone repositories directly onto your shared workspace file system using Git on your local workstation.

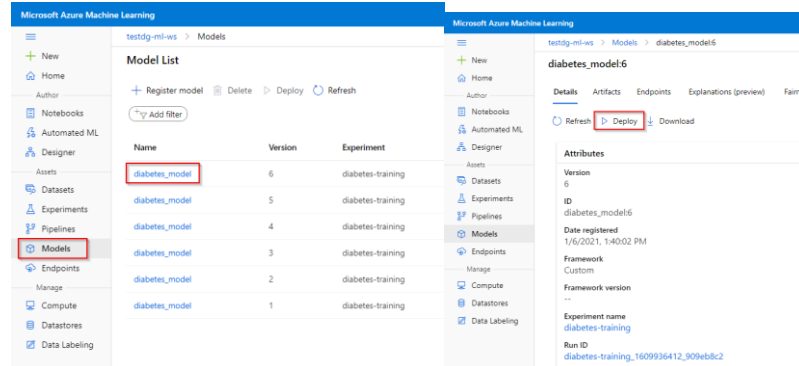


7. Model Deployment

Azure ML Studio has also the feature to deploy a trained machine learning model using a real-time Azure Container Instances (ACIs), a Web Service.

7.1 Deploying a web service

A model can only be deployed after you have registered it in the Azure ML Studio. To deploy a model.



- Provide the entry script and a conda dependencies file. The entry script is used to process the data send to the Web Service. It can be a python file containing an `init()` and a `run(data)` function. The conda dependencies file defines the environment used to run the deployed model.

Deploy a model

Name *

diabetes-service

Description

Web service to predict whether my patients are diabetic.

Compute type *

Azure Container Instance

Models: diabetes_model:6

Enable authentication

☐

Entry script file *

score_diabetes.py

Browse

Conda dependencies file *

diabetes_env.yml

Browse

Dependencies

Add File

Advanced

Deploy

Cancel

7.2 view a web service

The first screenshot shows the 'Endpoints' page in the Microsoft Azure Machine Learning portal. The left sidebar has the 'Endpoints' option highlighted. The main area shows a table with one endpoint named 'diabeteservice'.

The second screenshot shows the 'diabeteservice' details page. The 'Consume' tab is selected, and the 'Basic consumption info' section is highlighted, showing the REST endpoint URL.

The third screenshot shows the 'Endpoints' page again, but with the 'Delete' button highlighted in the top right corner of the table.

7.3 Deleting a web service

The screenshot shows the 'Endpoints' page in the Microsoft Azure Machine Learning portal. The left sidebar has the 'Endpoints' option highlighted. The main area shows a table with one endpoint named 'diabeteservice'. The 'Delete' button is highlighted in the top right corner of the table.