

Azure AI | Machine Learning Studio

npunext-1680261963256
Juhio07

Create data asset

1 Data type

2 Data source

3 Destination storage type

4 File or folder selection

5 Settings

6 Schema

7 Review

Set the name and type for your data asset

Name *

Diabetes_Dataset

Description

Data asset description

Type *

Tabular

Use cases for data types

When should I use File type?

The File type is recommended in most scenarios when you are working with a single data file of any type (including tabular data). This type allows you to specify a file location by URI in a storage location on your local computer, an attached Datastore, blob/ADLS storage, or a publicly available http(s) location. There are many types of supported URIs. In the Azure Machine Learning CLI v2 or Python SDK v2, this data type is called `uri_file`. [Learn more about the uri_file type](#)

When should I use Folder type?

The Folder type has all the same

Back

Next

Cancel

Azure AI | Machine Learning Studio

npunext-1680261963256
Juhio07

Create data asset

✓ Data type

✓ Data source

✓ Destination storage type

4 File or folder selection

5 Settings

6 Schema

7 Review

Choose a file or folder

Choose files or folders to upload from your local drive. If you upload multiple folders or files, they will be stored in a containing folder.

Upload path

azureml://subscriptions/8611f143-8cf4-47cb-b7d4-8b18969153e1/res...

Upload

☐ Overwrite if already exists

Upload list

diabetes.csv 505.62 KB/505.6

Information

What file types can I use?

Supported file types include: delimited (such as csv or tsv), Parquet, JSON Lines, and plain text.

Where are files uploaded?

Files will be uploaded to the selected datastore and made available in your workspace.

Back

Next

Cancel

Azure AI | Machine Learning Studio

🕒 🔔 ⚙️ ↩️ 13 ? 😊

npunext-1680261963256

Julis07

Create data asset

✓ Data type

✓ Data source

✓ Destination storage type

✓ File or folder selection

5 Settings

6 Schema

7 Review

Settings

These settings determine how the data is parsed. The initial settings are automatically detected; you can change them as needed to reparse the data.

File format

Delimiter

Example

Encoding

Delimited

Comma

Field1,Field2,Field3

UTF-8

Column headers

Skip rows

All files have same headers

None

☐ Dataset contains multi-line data ⓘ

ⓘ Note: Processing tabular files with multi-line data is slower because multiple CPU cores cannot be used to ingest the data in parallel. Checking this option may result in slower processing times.

PatientID	Pregnan...	Plasma...	Diastoli...	TricepsT...	Seruml...	BMI	Diabete...	Age	Diabetic
1354778	0	171	80	34	23	43.51	1.213	21	0
1147438	8	92	93	47	36	21.241	0.158	23	0
1640031	7	115	47	52	35	41.512	0.079	23	0
1883350	9	103	78	25	304	29.582	1.283	43	1

Back

Next

Review

Cancel

Azure AI | Machine Learning Studio

🕒 🔔 ⚙️ ↩️ 13 ? 😊

npunext-1680261963256

Julis07

Create data asset

✓ Data type

✓ Data source

✓ Destination storage type

✓ File or folder selection

✓ Settings

6 Schema

7 Review

Schema

Column types are auto-detected based on the initial subset of the data and can be updated here. Values not aligning with the specified column type will fail conversion and would be either null-filled or replaced with error value. Any conversions preview errors are non-blocking and you can proceed.

🔍 Search column name

Incl...	Column name	Type	Example values	Date format ⓘ	Properties ⓘ
<input type="checkbox"/>	Path	String		Not applicable to s...	Not applicable t...
<input checked="" type="checkbox"/>	PatientID	Integer	1354778, 1147438, 16...	Not applicable to s...	Not applicable t...
<input checked="" type="checkbox"/>	Pregnancies	Integer	0, 8, 7	Not applicable to s...	Not applicable t...
<input checked="" type="checkbox"/>	PlasmaGlucose	Integer	171, 92, 115	Not applicable to s...	Not applicable t...
<input checked="" type="checkbox"/>	DiastolicBloodPressure	Integer	80, 93, 47	Not applicable to s...	Not applicable t...

Back

Next

Cancel

Azure AI | Machine Learning Studio

1

13

?

😊

npunext-1680261963256

Juhis07

Create data asset

✓ Data type

✓ Data source

✓ Destination storage type

✓ File or folder selection

✓ Settings

✓ Schema

7 Review

Review

Review the settings for your data asset and make any changes as needed.

Data type

Name
Diabetes_Dataset

Description
--

Type
tabular

Data source

Type
Local

File selection

Upload path
azureml://subscriptions/8611f143-8cf4-47cb-b7d4-8b18969153e1/resourcegroups/trainingml/workspa
ces/Juhis07/datasets/workspaceblobstore/cathel

Schema

PatientID Integer

Pregnancies Integer

PlasmaGlucose Integer

DiastolicBloodPre... Integer

TricepsThickness Integer

(showing 5 of 11 columns)

Back

Create

Cancel

Azure AI | Machine Learning Studio

1

13

?

😊

npunext-1680261963256

Juhis07

Authoring

Notebooks

Automated ML

Designer

Prompt flow PREVIEW

Assets

Data

Jobs

Components

Pipelines

Environments

Models

Endpoints

Manage

Compute

Monitoring PREVIEW

Data Labeling

Unext > Juhis07 > Data > Diabetes_Dataset

Diabetes_Dataset

Version: 1 (latest) ☆

Details

Consume

Explore

Models

Jobs

New version

Refresh

Generate profile

Archive

Attributes

Type ⓘ
Table (mltable)

Dataset type (from Azure ML v1 APIs)
Tabular

Created by
Shellunext unextIDA57

Profile
[View profile](#)
Job: --

Files in dataset
1

Total size of files in dataset ⓘ
505.6 KiB

Current version
1

Tags

ⓘ No data

Description

ⓘ Click edit icon to add a description

Data sources

Datastore
[workspaceblobstore](#)

Relative path
UI/2023-10-04_065150_UTC/diabetes.csv

Actions
[View in datastores browse](#)
[View in Azure Portal](#)

Azure AI | Machine Learning Studio

Unext > Juhis07 > Data

Data

Data assets Datastores Dataset monitors PREVIEW

Data assets are immutable references to your data that can be created from datastores, local files, public URLs, or Open Datasets. Data assets created with AzureML v2 APIs cannot be deleted, but you can up-version or archive them for easy referencing and reuse in machine learning tasks. Deleting data assets created with v1 APIs will permanently delete the data asset and all metadata. [Learn more about data assets](#)

[+ Create](#) Refresh Archive View options Show latest version only Include archived

Search

Name	☆	Source	Version	Created on ↓	Modified on	Type
Diabetes_Dataset		This workspace	1	Oct 4, 2023 12:25 PM	Oct 4, 2023 12:25 PM	Tab

< Prev Next > 25/Page

Azure AI | Machine Learning Studio

Unext > Juh-workspace > Designer > Authoring

Configure & Submit

Save Pipeline interface

Columns to be cleaned

Select columns ☒ With rules ☐ By name

Allow duplicates and preserve column order in selection ☐

Include Column names Age X AnnualIncome X SpendingScore X

Save Cancel

Generate missing column

Generate missing value indicator column ☐ *

False

Output settings

Azure AI | Machine Learning Studio

Unext > juhi-workspace > Designer > Authoring

Undo Redo Validate Show lineage Clone AutoSave Configure & Submit

Pipeline-Created-on-10-04-2023 Save Pipeline interface

Customer_data
Customer_data
Data output
Dataset
Clean Missing Data
clean_missing_data
Cleaned datas... Cleaning tran...

Parameters

Clean Missing Data

Columns to be cleaned *

Column names: Age,AnnualIncome,SpendingScore

Minimum missing value ratio *

0.0

Maximum missing value ratio *

1.0

Cleaning mode *

Custom substitution value

Replacement value

0

Generate missing value indicator column *

False

Output settings

Azure AI | Machine Learning Studio

Unext > juhi-workspace > Designer > Authoring

Undo Redo Validate Show lineage Clone ... Configure & Submit

Pipeline-Created-on-10-04-2023 Save Pipeline interface

Search: tune

Tags: All Add filter

Data Component

2 Most relevant

Tune Model Hyperparameters
Microsoft
Perform a parameter sweep on the model to determine the optimum parameter settings. [Learn ...]
azureMLDesigner:true 1/10/2023

Train PyTorch Model
Microsoft
Train pytorch model from scratch or fine-tune it. [Learn More](https://aka.ms/aml/train-pytorch-mod...)
azureMLDesigner:true 1/10/2023

Customer_data
Customer_data
Data output
Clean Missing Data
clean_missing_data
Cleaned datas... Cleaning tran...
Filter Based Feature Selection
filter_based_feature_selection
Filtered data... Features
Dataset
Split Data
split_data
Results data... Results data...
Neural Network Regression
neural_network_regression
Untrained model
Untrained mod... Dataset
Train Model
train_model
Trained model
Trained model
Score Model
score_model
Scored dataset
Scored dataset... Scored dataset...
Evaluate Model
evaluate_model
Tune Model Hyperparameters
tune_model_hyperparameters
Sweep results Trained best ...

Parameters