

AZURE MACHINE LEARNING WORKSHOP

AGENDA

1. [Introduction to Azure Machine Learning](#)
2. [Create your Machine Learning Workspace](#)
3. [Create Compute](#)
4. [Create Dataset](#)
5. [Automatic Machine Learning](#)
6. [Azure Machine Learning Designer](#)

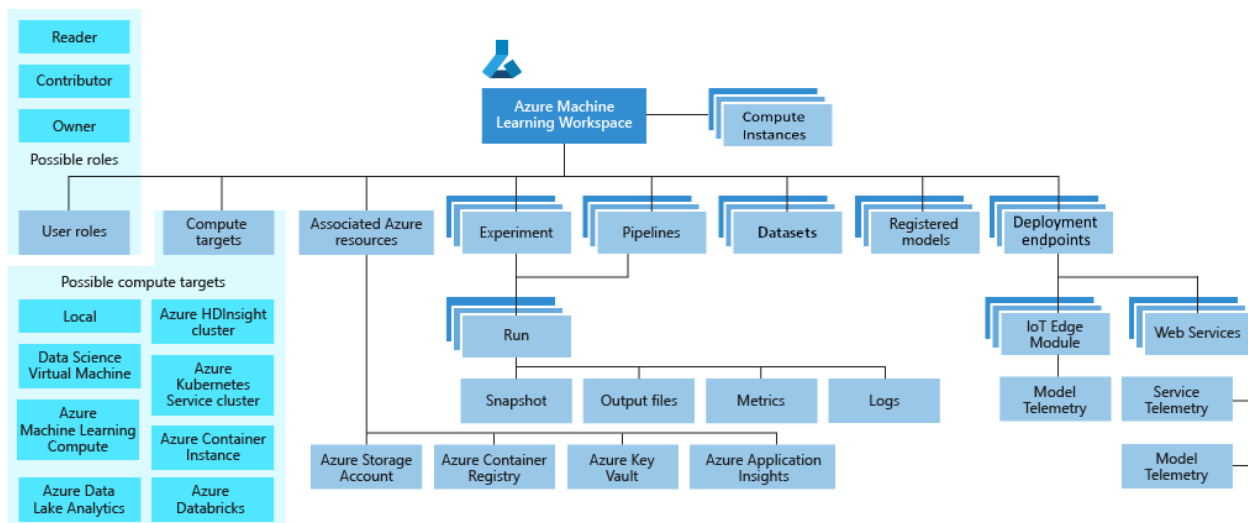
Introduction to Azure Machine Learning

Azure Machine Learning can be used for any kind of machine learning, from classical ml to deep learning, supervised, and unsupervised learning. Whether you prefer to write Python or R code or zero-code/low-code options such as the designer, you can build, train, and track highly accurate machine learning and deep-learning models in an Azure Machine Learning Workspace.

Azure Machine Learning provides all the tools developers and data scientists need for their machine learning workflows, including:

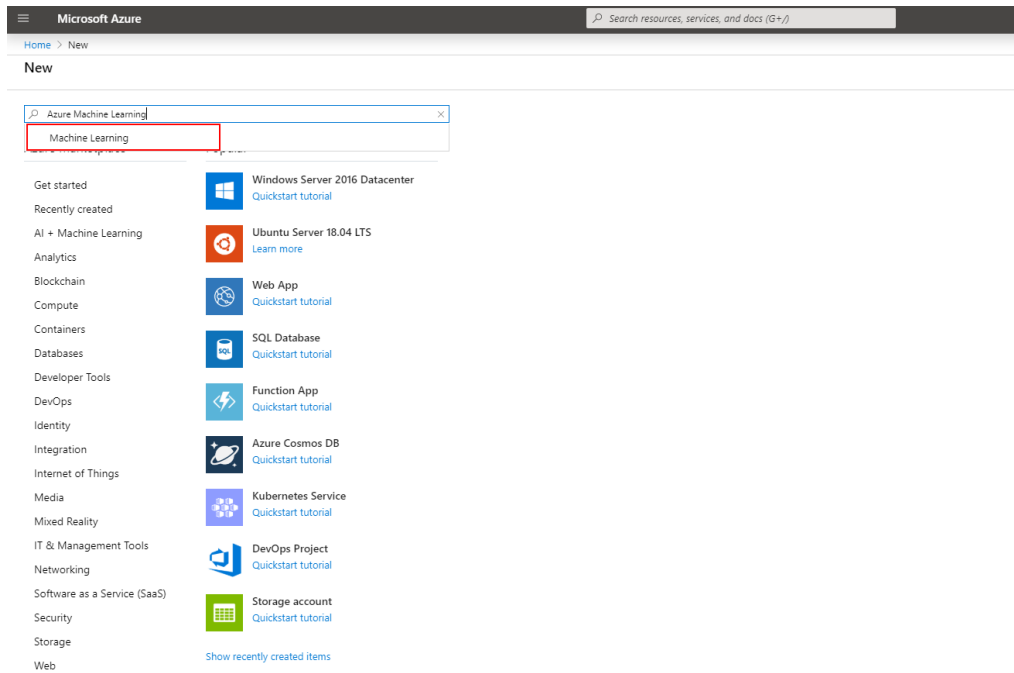
- The [Azure Machine Learning designer \(preview\)](#): drag-n-drop modules to build your experiments and then deploy pipelines.
- Jupyter notebooks: use our [example notebooks](#) or create your own notebooks to leverage our [SDK for Python](#) samples for your machine learning.
- R scripts or notebooks in which you use the [SDK for R](#) to write your own code, or use the R modules in the designer.
- [Visual Studio Code extension](#)
- [Machine learning CLI](#)
- Open-source frameworks such as PyTorch, TensorFlow, and scikit-learn and many more
- You can even use [MLflow to track metrics and deploy models](#) or Kubeflow to [build end-to-end workflow pipelines](#).

A [taxonomy](#) of the workspace is illustrated in the following diagram:



Azure Machine Learning Workspace

- Sign in to the [Azure portal](#) by using the credentials for your Azure subscription.
- In the upper-left corner of Azure portal, select **+ Create a resource**.
- Use the search bar to find **Machine Learning**
- Select **Machine Learning**



- In the **Machine Learning** pane, select **Create** to begin.

Microsoft Azure

Search resources, services, and docs (G+)

Home > Resource groups > rg-cbui-course532 > New > Machine Learning > Machine Learning

Machine Learning

Create

Main * Tags Review *

Workspace Name *

amlwksphol

This workspace name already exists

Subscription

sb-azu-dna-sb-01

Resource group

rg-cbui-course532

Create new

Location

(US) West US

Workspace edition View full pricing details

Enterprise

For your convenience, these resources are added automatically to the workspace, if regionally available: Azure storage, Azure Application Insights and Azure Key Vault.

- When you're finished configuring the workspace, select **Review + Create**.
- Review the settings and make any additional changes or corrections. When you're satisfied with the settings, select **Create**.
- To view the new workspace, select **Go to resource**.

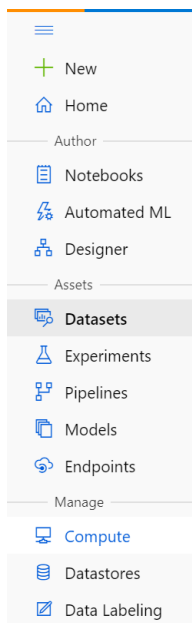
Create Compute

Create Compute Cluster

An Azure Machine Learning compute instance (preview) is a fully managed cloud-based workstation for data scientists. Compute instances make it easy to get started with Azure Machine Learning development as well as provide management and enterprise readiness capabilities for IT administrators. Use a compute instance as your fully configured and managed development environment in the cloud. Compute instances are typically used as development environments. They can also be used as a compute target for training and inferencing for development and testing. For large tasks, an [Azure Machine Learning compute cluster](#) with multi-node scaling capabilities is a better compute target choice.

We will do several actions that require a compute target to be executed on. We will start by creating a cluster of CPUs VMs.

- Sign in at <https://ml.azure.com>.
- Select Compute in the Assets section of the left pane.



- Navigate to Compute > Compute Clusters in the Manage section and click New.

Preview

Microsoft Azure Machine Learning

New

Home

Author

Notebooks

Automated ML

Designer

Assets

Datasets

Experiments

Pipelines

Models

Endpoints

Manage

Compute

Datastores

Data Labeling

amlwksphol > Compute

Compute

Compute instances

Compute clusters

Inference clusters

Attached compute

In the wake of COVID-19, we are prioritizing maintaining service availability for first responders, health and emergency manage

Create a single or multi-node compute cluster for your training or batch workloads across single node, distributed training or b

+ New

Refresh

Delete

Name	Provisioning state	Virtual machine size	Created on ↓
------	--------------------	----------------------	--------------

- **Name** the cluster cpuccluster.
- **Virtual machine type** CPU
- For **machine size** choose Standard_DS2_v2.
- Set the **Minimum number of nodes** to 0 and the **Maximum number of nodes** to 4. That way the cluster will scale automatically to up to 4 nodes as jobs require them.
- Set the **Idle seconds before scaling down** to 600. That means that nodes will be kept around for 10 minutes before they are spun down. That way, during our workshop, jobs will not have to wait for spin-up. Make sure that number is lower if you are using a more expensive VM size.

New compute cluster ⓘ



ⓘ Customers should not include personal data or other sensitive information in fields marked with ⓘ because the content in these fields may be logged and shared across Microsoft systems to facilitate operations and troubleshooting. [Learn more](#)

Compute name * ⓘ



Region * ⓘ

Virtual machine type *

Virtual machine priority * ⓘ

Dedicated Low priority

Virtual machine size * ⓘ

Standard_DS2_v2 2 Cores, 7 GB (RAM), 14 GB (Disk)

Minimum number of nodes * ⓘ

Maximum number of nodes * ⓘ

Idle seconds before scale down * ⓘ

> Advanced settings

Create

Cancel

- Click Create.
- Wait for the provisioning state to be “Succeeded”

Compute

Compute instances **Compute clusters** Inference clusters Attached compute

- ① In the wake of COVID-19, we are prioritizing maintaining service availability for first responders, health and emergency management services, critical government operations, and other essential services.
- ① Create a single or multi-node compute cluster for your training or batch workloads across single node, distributed training or batch inferencing.

+ New Refresh Delete

Name	Provisioning state	Virtual machine size	Created on ↓
cpucluster	✔ Succeeded (0 nodes)	STANDARD_DS2_V2	May 18, 2020 2:32 PM

Create Compute Instance

Next, we will create a Compute Instance. The Compute Instance will serve as an interactive workstation in the cloud that serves as a Jupyter server, but also hosts and instance of RStudio server and can run TensorBoard, Bokeh, Shiny or other apps used during the development work of a Data Scientist.

- Navigate to Compute Instances tab in the Compute section and click on New.

The screenshot shows the Azure ML interface. On the left is a sidebar with navigation options: New, Home, Author, Notebooks, Automated ML, Designer, Assets, Datasets, Experiments, Pipelines, Models, Endpoints, Manage, Compute (selected), Datastores, and Data Labeling. The main panel shows the 'Compute' section with tabs for 'Compute instances', 'Compute clusters', 'Inference clusters', and 'Attached compute'. The 'Compute instances' tab is active, displaying three informational messages and a toolbar with buttons: + New, Refresh, Start, Stop, Restart, Delete, and a toggle for 'Show created by me only'. Below the toolbar is a table with columns 'Name', 'Status', and 'Application URI'.

- Choose some sufficiently unique name, keep the default VM type (STANDARD_DS3_V2 - a fairly inexpensive machine type costing).

New compute instance ✕

ⓘ Customers should not include personal data or other sensitive information in fields marked with ⓘ because the content in these fields may be logged and shared across Microsoft systems to facilitate operations and troubleshooting. [Learn more](#) ✕

Compute name * ⓘ 🔍

amlhol

Region * ⓘ

westus

Virtual machine type *

CPU (Central Processing Unit)

Virtual machine size * ⓘ

Standard_DS3_v2 4 Cores, 14 GB (RAM), 28 GB (Disk)

☐ Enable SSH access ⓘ

[Advanced settings](#)

Create

Cancel

- Click Create.
- Wait for it to change the status to “Running”.

amlwksphol > Compute

Compute

Compute instances Compute clusters Inference clusters Attached compute

ⓘ In the wake of COVID-19, we are prioritizing maintaining service availability for first responders, health and emergency management services, critical government infrastructure, and existing paid customer use. To balance the best possible experience, we are temporarily limiting the number of new compute instances that can be created in some regions. We will continue to monitor the situation and update this message as needed.

ⓘ Get started with AzureML notebooks by creating a new compute instance. Choose from a selection of CPU or GPU instances pre-installed with the most common tools or just use our interactive notebooks

ⓘ Compute instance is replacing the Notebook VM. You cannot create new Notebook VMs, but you can still use existing Notebook VMs. [Learn More](#).

+ New Refresh Start Stop Restart Delete ☒ Show created by me only

Name	Status	Application URI	Virtual machine size	Created on ↓
amlhol	Running	JupyterLab Jupyter RStudio SSH	STANDARD_DS3_V2	May 18, 2020 2:34 PM

- Note that this machine will keep running until you stop it!

Create Dataset

With Azure Machine Learning datasets, you can:

- Keep a single copy of data in your storage, referenced by datasets.
- Seamlessly access data during model training without worrying about connection strings or data paths.
- Share data and collaborate with other users.

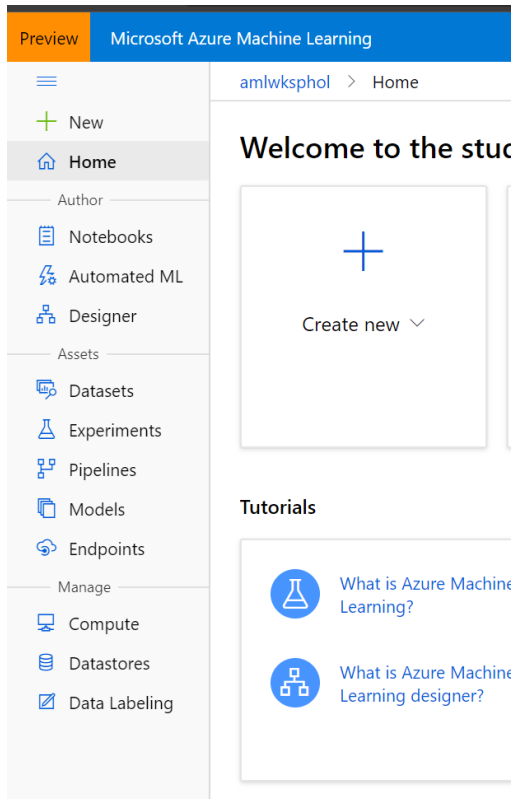
There are two dataset types, based on how users consume them in training:

- [TabularDataset](#) represents data in a tabular format by parsing the provided file or list of files. This provides you with the ability to materialize the data into a Pandas or Spark DataFrame. You can create a TabularDataset object from .csv, .tsv, .parquet, .jsonl files, and from SQL query results. For a complete list, see TabularDatasetFactory class.
- The [FileDataset](#) class references single or multiple files in your datastores or public URLs. By this method, you can download or mount the files to your compute as a FileDataset object. The files can be in any format, which enables a wider range of machine learning scenarios, including deep learning.

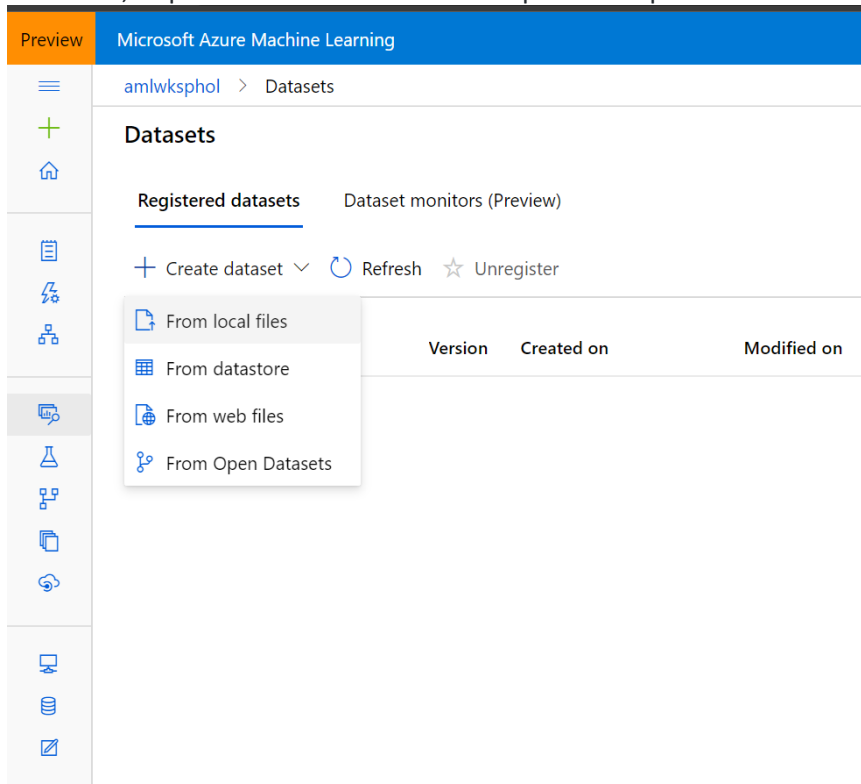
Once you have created an instance of Azure Machine Learning, select 'Launch Azure Machine Learning studio' (or browse to <https://ml.azure.com/>)

By creating a dataset, you create a reference to the data source location, along with a copy of its metadata. To create a dataset in the studio:

- Sign in at <https://ml.azure.com>.
- Select Datasets in the Assets section of the left pane.



- Select Create Dataset to choose the source of your dataset. This source can be local files, a datastore, or public URLs. For our workshop we will upload the data from local files.



- Select Tabular for Dataset type and give your dataset a name.

Create dataset from local files


Basic info

Datastore and file selection

Settings and preview

Schema

Confirm details

Customers should not include personal data or other sensitive information in fields marked with  because the content in these fields may be logged and shared across Microsoft systems to facilitate operations and troubleshooting. [Learn more](#)

×

Basic info

Name *

employeeattrition

Dataset version

1

Dataset type *

Tabular

Description

Dataset description

- Select Next to open the Datastore and file selection form. On this form you select where to keep your dataset after creation, as well as select what data files to use for your dataset. Select Browse and upload the file from your local desktop. Keep rest of the default values.

Create dataset from local files

Basic info

Datastore and file selection

Settings and preview

Schema

Confirm details

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, and plain text.

Browse

1 files selected. Total size 0.2140 MiB. 0/1 files uploaded

File name	Size (MiB)	Upload %	Status
Employee-Attrition.txt	0.2140		

Upload path

UI

Files will be uploaded to '\$(Upload path)/05-18-2020_071936_UTC'

☐ Skip data validation ⓘ

- Select Next to populate the Settings and preview and Schema forms; they are intelligently populated based on file type and you can further configure your dataset prior to creation on these forms. Change the Column header to “All files have same header”. Ensure the rest of the setting and preview of the data.

Create dataset from local files

Basic info

Datastore and file 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 All files have same headers

Skip rows None

	Age	Attrition	BusinessTravel	DailyRate	Department	DistanceFrom...	Education
1	41	1	Travel_Rarely	1102	Sales	1	2
2	49	0	Travel_Frequently	279	Research & Develop...	8	1
3	37	1	Travel_Rarely	1373	Research & Develop...	2	2
4	33	0	Travel_Frequently	1392	Research & Develop...	3	4
5	27	0	Travel_Rarely	591	Research & Develop...	2	1
6	32	0	Travel_Frequently	1005	Research & Develop...	2	2
7	59	0	Travel_Rarely	1324	Research & Develop...	3	3
8	30	0	Travel_Rarely	1358	Research & Develop...	24	1
9	38	0	Travel_Frequently	216	Research & Develop...	23	3
10	36	0	Travel_Rarely	1299	Research & Develop...	27	3
11	35	0	Travel_Rarely	809	Research & Develop...	16	3
12	29	0	Travel_Rarely	153	Research & Develop...	15	2
13	31	0	Travel_Rarely	670	Research & Develop...	26	1
14	34	0	Travel_Rarely	1346	Research & Develop...	19	2
15	28	1	Travel_Rarely	103	Research & Develop...	24	3
16	29	0	Travel_Rarely	1389	Research & Develop...	21	4
17	32	0	Travel_Rarely	334	Research & Develop...	5	2
18	22	0	Non-Travel	1123	Research & Develop...	16	2
19	53	0	Travel_Rarely	1219	Sales	2	4
20	38	0	Travel_Rarely	371	Research & Develop...	2	3

Back

Next

Cancel

- Select Next to review the Schema and Data type.

Create dataset from local files

Basic info

Datastore and file 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"/>	Age	Not applicable to selected type	Integer	41, 49, 37
<input checked="" type="checkbox"/>	Attrition	Not applicable to selected type	Integer	1, 0, 1
<input checked="" type="checkbox"/>	BusinessTravel	Not applicable to selected type	String	Travel_Rarely, Travel_Frequently, Travel_Rarely
<input checked="" type="checkbox"/>	DailyRate	Not applicable to selected type	Integer	1102, 279, 1373
<input checked="" type="checkbox"/>	Department	Not applicable to selected type	String	Sales, Research & Development, Research & Development
<input checked="" type="checkbox"/>	DistanceFromHome	Not applicable to selected type	Integer	1, 8, 2
<input checked="" type="checkbox"/>	Education	Not applicable to selected type	Integer	2, 1, 2
<input checked="" type="checkbox"/>	EducationField	Not applicable to selected type	String	Life Sciences, Life Sciences, Other
<input checked="" type="checkbox"/>	EmployeeCount	Not applicable to selected type	Integer	1, 1, 1
<input checked="" type="checkbox"/>	EmployeeNumber	Not applicable to selected type	Integer	1, 2, 4
<input checked="" type="checkbox"/>	EnvironmentSatisfaction	Not applicable to selected type	Integer	2, 3, 4
<input checked="" type="checkbox"/>	Gender	Not applicable to selected type	String	Female, Male, Male
<input checked="" type="checkbox"/>	HourlyRate	Not applicable to selected type	Integer	94, 61, 92
<input checked="" type="checkbox"/>	JobInvolvement	Not applicable to selected type	Integer	3, 2, 2
<input checked="" type="checkbox"/>	JobLevel	Not applicable to selected type	Integer	2, 2, 1
<input checked="" type="checkbox"/>	JobRole	Not applicable to selected type	String	Sales Executive, Research Scientist, Laboratory Technician

Back

Next

Cancel

- Select Next to review the Confirm details form. Check your selections and create an data profile for your dataset. Learn more about [data profiling](#).

Create dataset from local files

- Basic info
- Datastore and file selection
- Settings and preview
- Schema
- Confirm details**

Confirm details

Basic info

Name
employeeattrition

Dataset version
1

Dataset type
Tabular

File settings

File format
Delimited

Delimiter
Comma

Encoding
UTF-8

Column headers
All files have same headers

Skip rows
None

☒ Profile this dataset after creation

Select compute for profiling
cpcucluster Refresh

Back Create Cancel

- Select Create to complete your dataset creation.
- Dataset should be registered and available for use.

Microsoft Azure Machine Learning

amlwksphol > Datasets

Success: employeeattrition dataset created successfully

Datasets

Registered datasets Dataset monitors (Preview)

+ Create dataset Refresh Unregister

Name	Version	Created on	Modified on	Properties	Created by	Tags
employeeattrition	1	May 18, 2020 2:44 PM	May 18, 2020 2:44 PM	Tabular	Ashish Talati	

< Prev Next >

- You can register a new dataset under the same name by creating a new version. A dataset version is a way to bookmark the state of your data so that you can apply a specific version of the dataset for experimentation or future reproduction. Learn more about [dataset versions](#).

- Once the data profile is completed (profile generation), you can get a vast variety of summary statistics across your data set to verify whether your data set is ML-ready. For non-numeric columns, they include only basic statistics like min, max, and error count. For numeric columns, you can also review their statistical moments and estimated quantiles.

Statistic	Description
Feature	Name of the column that is being summarized.
Profile	In-line visualization based on the type inferred. For example, strings, booleans, and dates will have value counts, while decimals (numerics) have approximated histograms. This allows you to gain a quick understanding of the distribution of the data.
Type distribution	In-line value count of types within a column. Nulls are their own type, so this visualization is useful for detecting odd or missing values.
Type	Inferred type of the column. Possible values include: strings, booleans, dates, and decimals.
Min	Minimum value of the column. Blank entries appear for features whose type does not have an inherent ordering (e.g. booleans).
Max	Maximum value of the column.
Count	Total number of missing and non-missing entries in the column.
Not missing count	Number of entries in the column that are not missing. Empty strings and errors are treated as values, so they will not contribute to the "not missing count."
Quantiles	Approximated values at each quantile to provide a sense of the distribution of the data.
Mean	Arithmetic mean or average of the column.
Standard deviation	Measure of the amount of dispersion or variation of this column's data.
Variance	Measure of how far spread out this column's data is from its average value.
Skewness	Measure of how different this column's data is from a normal distribution.
Kurtosis	Measure of how heavily tailed this column's data is compared to a normal distribution.

amlwksphol > Datasets > employeeattrition

employeeattrition Version 1 (latest) ▾

Details Consume Explore Models

Refresh Generate profile Unregister New version ▾

Attributes

Properties
Tabular

Description
--

Created by
Ashish Talati

Datastore
workspaceblobstore

Relative path
UI/05-18-2020_074302_UTC/Employee-Attrition.txt

Profile
[Profile generation is running](#)

Files in dataset
1

Total size of files in dataset
219.1 KiB

Current version
1

Latest version
1

Created time
May 18, 2020 2:44 PM

Modified time
May 18, 2020 2:44 PM

- Now, click on the newly created dataset and click Explore. Here you can see the fields of the Tabular dataset.

amlwksphol > Datasets > employeeattrition

employeeattrition Version 1 (latest) ▾

Details Consume **Explore** Models

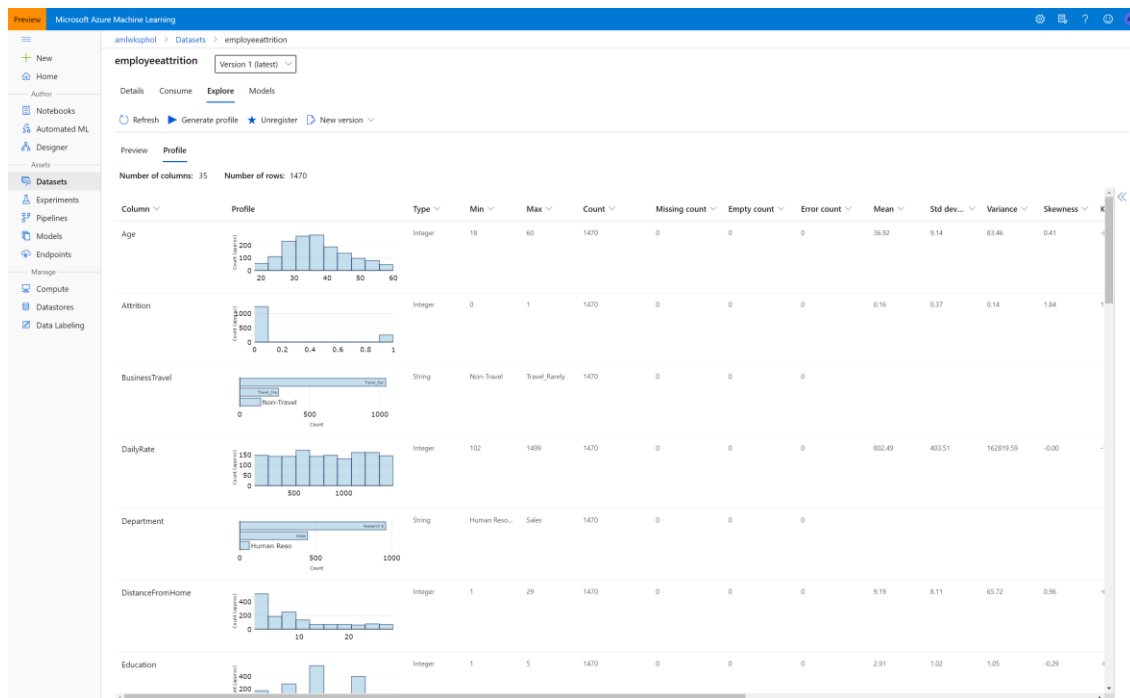
Refresh Generate profile Unregister New version ▾

Preview Profile

Number of columns: 35 Number of rows: 1000 (of 1470)

	Age	Attrition	BusinessTravel	DailyRate	Department	DistanceFrom...	Education	EducationField	EmployeeCount	EmployeeNum...	EnvironmentSa...	Gender
1	41	1	Travel_Rarely	1102	Sales	1	2	Life Sciences	1	1	2	Female
2	49	0	Travel_Frequently	279	Research & Develop...	8	1	Life Sciences	1	2	3	Male
3	37	1	Travel_Rarely	1373	Research & Develop...	2	2	Other	1	4	4	Male
4	33	0	Travel_Frequently	1392	Research & Develop...	3	4	Life Sciences	1	5	4	Female
5	27	0	Travel_Rarely	591	Research & Develop...	2	1	Medical	1	7	1	Male
6	32	0	Travel_Frequently	1005	Research & Develop...	2	2	Life Sciences	1	8	4	Male
7	59	0	Travel_Rarely	1324	Research & Develop...	3	3	Medical	1	10	3	Female
8	30	0	Travel_Rarely	1358	Research & Develop...	24	1	Life Sciences	1	11	4	Male
9	38	0	Travel_Frequently	216	Research & Develop...	23	3	Life Sciences	1	12	4	Male
10	36	0	Travel_Rarely	1299	Research & Develop...	27	3	Medical	1	13	3	Male
11	35	0	Travel_Rarely	809	Research & Develop...	16	3	Medical	1	14	1	Male
12	29	0	Travel_Rarely	153	Research & Develop...	15	2	Life Sciences	1	15	4	Female
13	31	0	Travel_Rarely	670	Research & Develop...	26	1	Life Sciences	1	16	1	Male
14	34	0	Travel_Rarely	1346	Research & Develop...	19	2	Medical	1	18	2	Male
15	28	1	Travel_Rarely	103	Research & Develop...	24	3	Life Sciences	1	19	3	Male
16	29	0	Travel_Rarely	1389	Research & Develop...	21	4	Life Sciences	1	20	2	Female
17	32	0	Travel_Rarely	334	Research & Develop...	5	2	Life Sciences	1	21	1	Male
18	22	0	Non-Travel	1123	Research & Develop...	16	2	Medical	1	22	4	Male
19	53	0	Travel_Rarely	1219	Sales	2	4	Life Sciences	1	23	1	Female
20	38	0	Travel_Rarely	371	Research & Develop...	2	3	Life Sciences	1	24	4	Male
21	24	0	Non-Travel	673	Research & Develop...	11	2	Other	1	26	1	Female
22	36	1	Travel_Rarely	1218	Sales	9	4	Life Sciences	1	27	3	Male
23	34	0	Travel_Rarely	419	Research & Develop...	7	4	Life Sciences	1	28	1	Female
24	21	0	Travel_Rarely	391	Research & Develop...	15	2	Life Sciences	1	30	3	Male
25	34	1	Travel_Rarely	699	Research & Develop...	6	1	Medical	1	31	2	Male
26	53	0	Travel_Rarely	1282	Research & Develop...	5	3	Other	1	32	3	Female

To view the profile of the dataset we generated in the previous step, click the Profile tab. If you want to regenerate a profile (or you created the dataset without selecting the profile option), you can click Generate profile and select a cluster to generate profile information for the dataset.



- In the Consume tab we can find a short code snippet for consuming the dataset from Notebooks.

Sample usage

```
# azureml-core of version 1.0.72 or higher is required
# azureml-dataprep[pandas] of version 1.1.34 or higher is required
from azureml.core import Workspace, Dataset

subscription_id = '5e22d967-997b-49c7-8ca1-7ccfbf37e621'
resource_group = 'rg-cbui-course532'
workspace_name = 'amlwksphol'

workspace = Workspace(subscription_id, resource_group, workspace_name)

dataset = Dataset.get_by_name(workspace, name='employeeattrition')
dataset.to_pandas_dataframe()
```

Dataset tutorials

- Create and run machine learning pipelines with Azure Machine Learning SDK
- Run batch inference on large amounts of data by using Azure Machine Learning
- Train models with automated machine learning in the cloud
- Train with datasets in Azure Machine Learning
- [GitHub] Azure Machine Learning datasets notebooks

- For more information on datasets, see the [how-to for more information on creating and using Datasets](#).

Automated Machine Learning

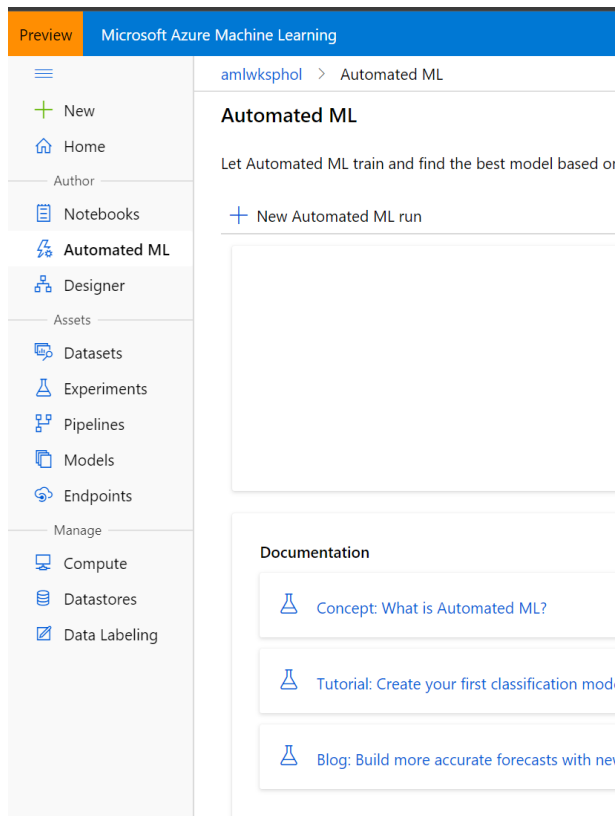
Automated machine learning (automated ML) builds high quality machine learning models for you by automating model and hyperparameter selection. Bring a labelled dataset that you want to build a model for, and automated ML will give you a high quality machine learning model that you can use for predictions.

If you are new to Data Science, automated ML will help you get jumpstarted by simplifying machine learning model building. It abstracts you from needing to perform model selection, hyperparameter selection and in one step creates a high-quality trained model for you to use.

If you are an experienced data scientist, automated ML will help increase your productivity by intelligently performing the model and hyperparameter selection for your training and generates high quality models much quicker than manually specifying several combinations of the parameters and running training jobs. Automated ML provides visibility and access to all the training jobs and the performance characteristics of the models to help you further tune the pipeline if you desire.

Follow the instructions in the [documentation](#) for a full overview of the user interface.

- Navigate to the left pane of your workspace. Select Automated ML under the Author section.



- Select New Automated ML run and select the employeeattrition dataset that you had created earlier.

amlakgphol > Automated ML > Start run

Create a new Automated ML run

Select dataset

Configure run

Task type and settings

Select dataset

Select a dataset from the list below, or create a new dataset. Automated ML currently only supports tabular data for authoring runs.

+ Create dataset ☒ Show supported datasets only

Search to filter items...

Dataset name	Dataset type	Created on	Modified
employeeattrition	Tabular	May 18, 2020 2:44 PM	May 18, 2020 2:44 PM

Back Next Cancel

- Select Next
- Enter the Experiment name, select the target column Attrition (this is what we want to predict), and select the cluster cpuccluster which you've created earlier.

amlakgphol > Automated ML > Start run

Create a new Automated ML run

Select dataset

Configure run

Task type and settings

Configure run

Configure the experiment. Select from existing experiments or define a new name, select the target column and the training compute to use. [Learn more on how to configure the experiment](#)

Dataset
employeeattrition [View dataset](#)

Experiment name *

☐ Select existing ☒ Create new

New experiment name

Target column *

Select training cluster *

[Create a new compute](#) [Refresh compute](#)

Back Next Cancel

- Select Next
- Keep the task type as “classification”
- Select view additional configuration.

amlwksphol > Automated ML > Start run

Create a new Automated ML run

- Select dataset
- Configure run
- Task type and settings**

Select task type

Select the machine learning task type for the experiment. Additional settings are available to fine tune the experiment if needed.

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

☐ Enable deep learning (preview) ⓘ

Regression
 To predict continuous numeric values

Time series forecasting
 To predict values based on time

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

- Set the Primary [metric](#) to AUC_weighted, the training job time to 0.25 and Max concurrent iterations to 4 (sufficient for the workshop). The concurrency parameter defines how many nodes of your cluster will be used for training.

Additional configurations ×

Primary metric ⓘ
 AUC weighted

☒ Explain best model ⓘ

Blocked algorithms ⓘ
 A list of algorithms that Automated ML will not use during training.

Exit criterion

Training job time (hours) ⓘ
 0.25

Metric score threshold ⓘ
 Metric score threshold

Validation

Validation type ⓘ
 Auto

Concurrency

Max concurrent iterations ⓘ
 4

- Click Save
- Hit Finish and wait for the training job to start. You'll be able to see the models which are created during the run, click on any of the models to open the detailed view of that model, where you can analyze the [graphs and metrics](#).
- Once the run is completed(Should take about 17 minutes) , click Deploy best model to create a deployed endpoint from the model.

Run 3 Completed

[Refresh](#) [Cancel](#)

Details Data guardrails **Models** Logs Outputs

[Deploy](#) [Download](#) [Explain model](#) [Search to filter items...](#)

Algorithm name	Explained	AUC weighted ↓	Sampling	Run	Created	Duration	Status
VotingEnsemble	View explanation	0.84161	100%	Run 34	May 19, 2020 10:11 PM	2m 0s	Completed
StackEnsemble		0.83891	100%	Run 33	May 19, 2020 10:11 PM	1m 44s	Completed
MaxAbsScaler, SGD		0.83285	100%	Run 9	May 19, 2020 9:57 PM	1m 34s	Completed
MaxAbsScaler, SGD		0.82740	100%	Run 14	May 19, 2020 10:01 PM	3m 6s	Completed
MaxAbsScaler, LightGBM		0.80752	100%	Run 22	May 19, 2020 10:05 PM	1m 36s	Completed
MaxAbsScaler, SGD		0.80616	100%	Run 21	May 19, 2020 10:05 PM	1m 21s	Completed
MaxAbsScaler, SGD		0.80413	100%	Run 13	May 19, 2020 10:00 PM	1m 21s	Completed
MaxAbsScaler, LightGBM		0.80389	100%	Run 6	May 19, 2020 9:55 PM	3m 4s	Completed
StandardScalerWrapper, LightGBM		0.80155	100%	Run 12	May 19, 2020 9:59 PM	1m 23s	Completed
MaxAbsScaler, SGD		0.79612	100%	Run 25	May 19, 2020 10:06 PM	1m 14s	Completed
StandardScalerWrapper, VGBowClassifier		0.79578	100%	Run 7	May 19, 2020 9:55 PM	3m 7s	Completed

-
- Once the model is deployed, you can consume that API in any client. For instance, you can call the API from Postman. Go to EndPoints in AML assets

Preview Microsoft Azure Machine Learning

amlwksphol > Endpoints

Endpoints

Real-time endpoints Pipeline endpoints

[Refresh](#) [Delete](#)

Name	Description	Created on
employeeattritionautomlaci	--	May 21, 2020 6:
predictattritionsvc	Explain prediction...	May 20, 2020 1:

-
- Go to consume tab and find the URL Endpoint and key(if authentication was enabled)


Preview Microsoft Azure Machine Learning

amlwksphol > Endpoints > employeeattritionautomlaci

employeeattritionautomlaci

Details Consume

Basic consumption info

REST endpoint
 

-
- Open Postman and create a new HTTP Post for that Rest Endpoint and send following HTTP Body

POST http://89e51907-5257-4715-af...

Untitled Request

POST

http://89e51907-5257-4715-af8f-fcb5ee84cfc2.westus.azurecontainer.io/score

Params

Authorization

Headers (9)

Body

Pre-request Script

Tests

Settings

none

form-data

x-www-form-urlencoded

raw

binary

GraphQL

JSON

15

16

17

18

19

20

21

22

23

24

25

26

27

28

29

30

31

32

33

34

35

"JobLevel": 2,

"JobRole": "Sales Executive",

"JobSatisfaction": 4,

"MaritalStatus": "Single",

"MonthlyIncome": 5993,

"MonthlyRate": 19479,

"NumCompaniesWorked": 8,

"Over18":1,

"OverTime": 1,

"PercentSalaryHike": 11,

"PerformanceRating": 3,

"RelationshipSatisfaction": 1,

"StandardHours" : 80,

"StockOptionLevel": 0,

"TotalWorkingYears": 8,

"TrainingTimesLastYear": 0,

"WorkLifeBalance": 1,

"YearsAtCompany": 6,

"YearsInCurrentRole": 4,

"YearsSinceLastPromotion": 0,

Body

Cookies

Headers (7)

Test Results

Pretty

Raw

Preview

Visualize

JSON

1

"{"result": [1]}"

```
{
  "data": [{
    "Age": 41,
    "BusinessTravel": "Travel_Rarely",
    "DailyRate": 1102,
    "Department": "Sales",
    "DistanceFromHome": 1,
    "Education": 2,
    "EducationField": "Life Sciences",
    "EmployeeCount" : 1,
    "EmployeeNumber" : 1,
    "EnvironmentSatisfaction": 2,
    "Gender": "Female",
    "HourlyRate": 94,
    "JobInvolvement": 3,
    "JobLevel": 2,
    "JobRole": "Sales Executive",
    "JobSatisfaction": 4,
    "MaritalStatus": "Single",
    "MonthlyIncome": 5993,
    "MonthlyRate": 19479,
    "NumCompaniesWorked": 8,
    "Over18":1,
    "OverTime": 1,
```

```

    "PercentSalaryHike": 11,
    "PerformanceRating": 3,
    "RelationshipSatisfaction": 1,
    "StandardHours" : 80,
    "StockOptionLevel": 0,
    "TotalWorkingYears": 8,
    "TrainingTimesLastYear": 0,
    "WorkLifeBalance": 1,
    "YearsAtCompany": 6,
    "YearsInCurrentRole": 4,
    "YearsSinceLastPromotion": 0,
    "YearsWithCurrManager": 5
  }
}

```

To learn more about automated ML, see documentation [here](#).

Optional Tasks:

- Once your model has been deployed, follow these [instructions](#) to consume the model from Power BI.
- Feature engineering: Auto ML also allows you to do feature engineering before starting the run. You can checkout this Featurization option on the start run page.

Featurization
×

Feature selection identifies the actions performed on the dataset to prepare the data for training. This will not impact the input data needed for inferencing i.e., if columns are excluded from training, the excluded columns will still be required as input for inferencing on the model. [Learn more about Automated ML's featurization](#)

☒ Enable featurization

Column name	Included	Feature type	Impute with	Data example
Age	<input checked="" type="checkbox"/>	Auto	Auto	41, 49, 37
Attrition (Target column)	<input type="checkbox"/>	Auto	Auto	1, 0, 1
BusinessTravel	<input checked="" type="checkbox"/>	Auto	Auto	Travel_Rarely, Travel_Frequently, Tra...
DailyRate	<input checked="" type="checkbox"/>	Auto	Auto	1102, 279, 1373
Department	<input checked="" type="checkbox"/>	Auto	Auto	Sales, Research & Development, Res...
DistanceFromHome	<input checked="" type="checkbox"/>	Auto	Auto	1, 8, 2
Education	<input checked="" type="checkbox"/>	Auto	Auto	2, 1, 2
EducationField	<input checked="" type="checkbox"/>	Auto	Auto	Life Sciences, Life Sciences, Other
EmployeeCount	<input checked="" type="checkbox"/>	Auto	Auto	1, 1, 1

Save

Try the [sample notebooks](#).

Azure Machine Learning Designer

Azure Machine Learning designer lets you visually connect [datasets](#) and [modules](#) on an interactive canvas to create machine learning models.

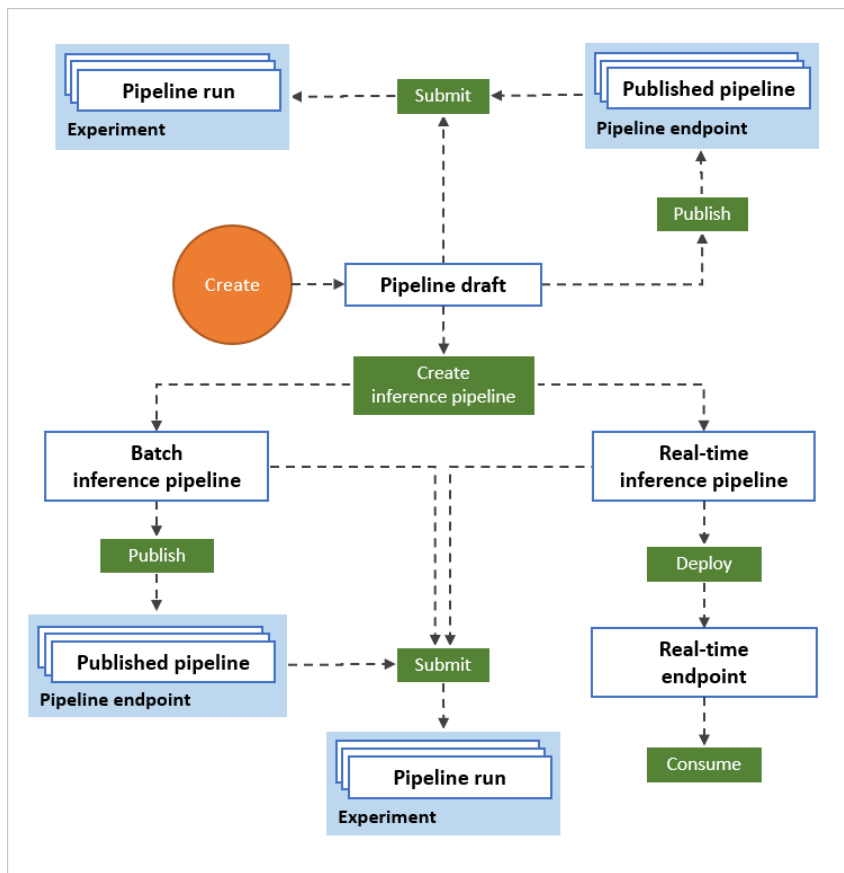
The designer uses your Azure Machine Learning workspace to organize shared resources such as:

- [Pipelines](#)
- [Datasets](#)
- [Compute resources](#)
- [Registered models](#)
- [Published pipelines](#)
- [Real-time endpoints](#)

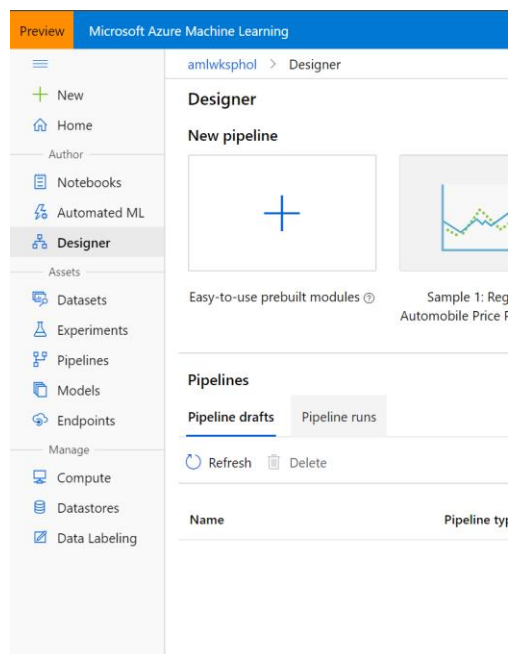
The designer gives you a visual canvas to build, test, and deploy machine learning models. With the designer you can:

Drag-and-drop datasets and modules onto the canvas.

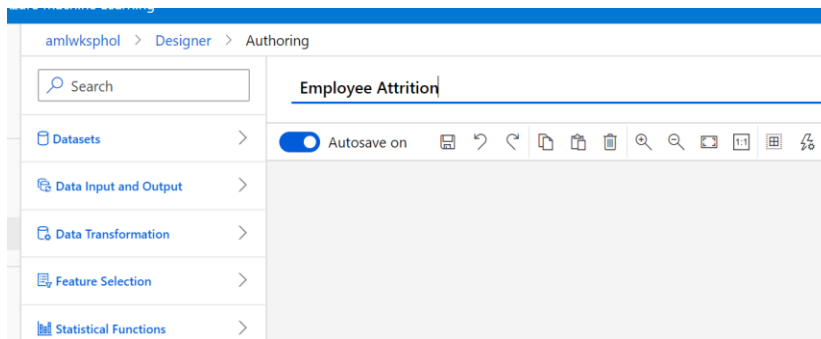
- Connect the modules together to create a pipeline draft.
- Submit a pipeline run using the compute resources in your Azure Machine Learning workspace.
- Convert your training pipelines to inference pipelines.
- Publish your pipelines to a REST pipeline endpoint to submit new pipeline runs with different parameters and datasets.
- Publish a training pipeline to reuse a single pipeline to train multiple models while changing parameters and datasets.
- Publish a batch inference pipeline to make predictions on new data by using a previously trained model.
- Deploy a real-time inference pipeline to a real-time endpoint to make predictions on new data in real time.



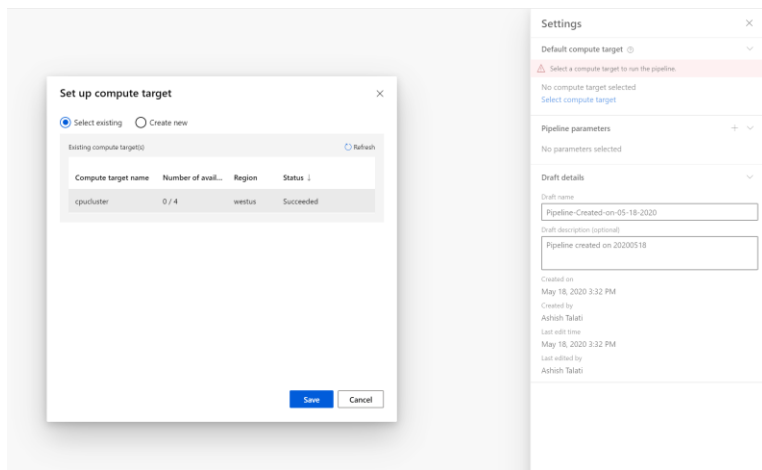
- In the AML, navigate to the Designer tab under the Author section and launch it.
- Create a new experiment by click +



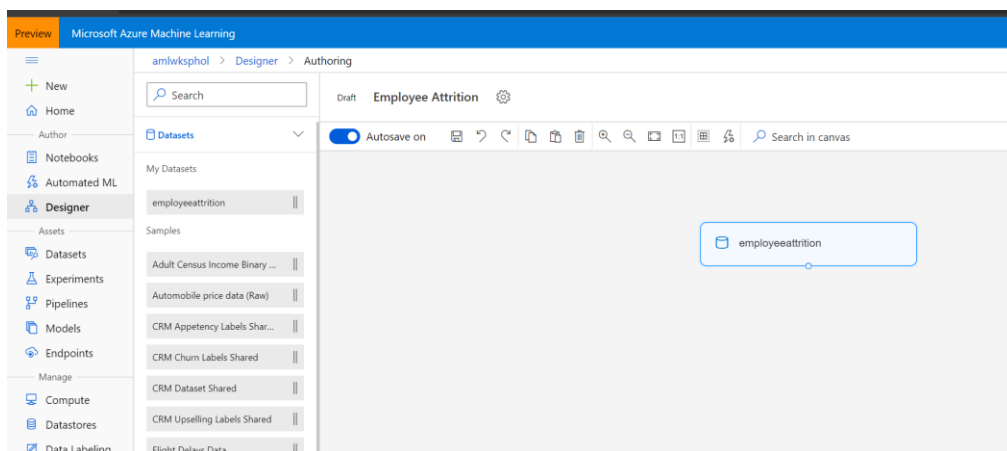
- Rename the pipeline to Employee Attrition



- Select the compute target to the one that you created earlier. Click Save.



- Click "X" to close the property window
- Expand Datasets -> My DataSets and select employeeattrition dataset that was created earlier.
- Drag the employeeattrition dataset module into the pipeline canvas.



- Add “Select Columns in Dataset” module to remove “EmployeeCount, EmployeeNumber, Over18 and StandardHours” columns.

Select columns ×

Select columns ☒ With rules ☐ By name

Allow duplicates and preserve column order in selection ☐

Include All columns +

Exclude Column names + −

EmployeeCount × EmployeeNumber ×
Over18 × StandardHours ×

Save Cancel

- Add a Split Data module to create the training and test sets. Set the fraction of rows in the first output dataset to 0.7. This setting specifies that 70% of the data will be output to the left port of the module and the rest to the right port. We use the left dataset for training and the right one for testing.

Split Data ×

Splitting mode * Split Rows ▼

Fraction of rows in the first output dataset * 0.7

☒ Randomized split

Random seed * 0

Stratified split * False ▼

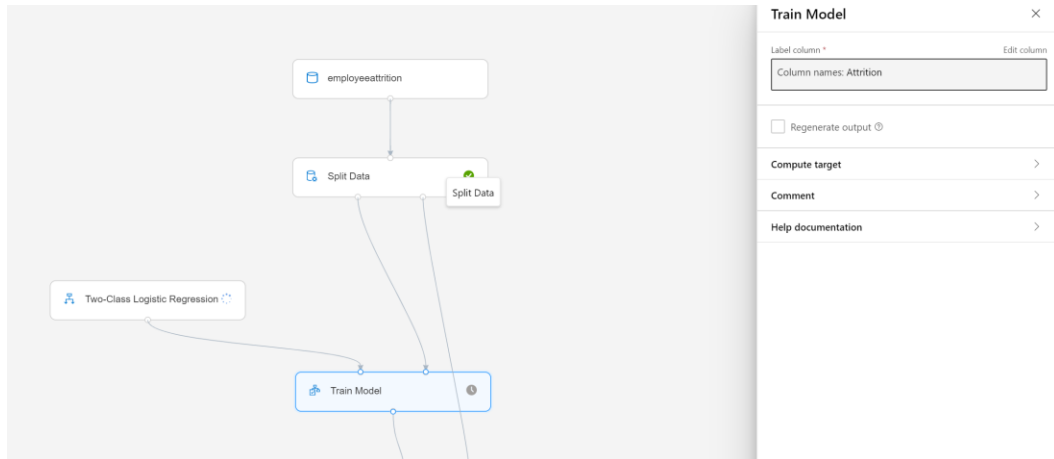
☐ Regenerate output ⓘ

[Compute target](#) >

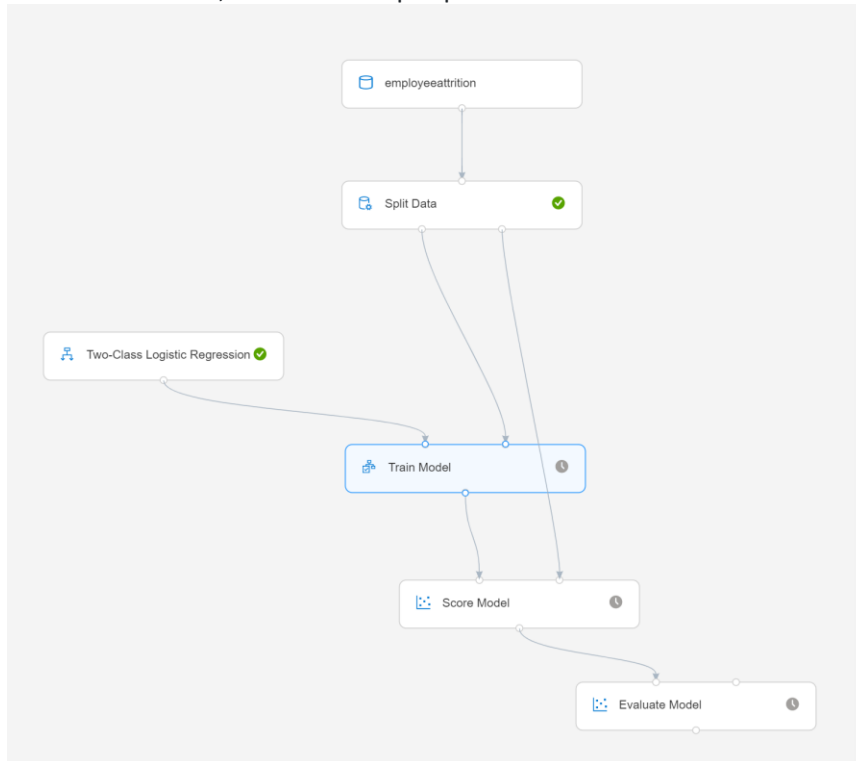
[Comment](#) >

[Help documentation](#) >

- Add a Two-Class Logistics Regression module to initialize a boosted decision tree classifier.
- Add a Train Model module. Connect the classifier from the previous step to the left input port of the Train Model. Connect the filtered dataset from Filter Based Feature Selection module as training dataset. Add the target column as Attrition in the label Column option. (**Column names: Attrition**) The Train Model will train the classifier.



- Add **Score Model** module and connect the **Train Model** module to it. Then add the test set (the output of Filter Based Feature Selection module which apply feature selection to test set too) to the **Score Model**. The **Score Model** will make the predictions. You can select its output port to see the predictions and the positive class probabilities.
- Add an Evaluate Model module and connect the scored dataset to its left input port. To see the evaluation results, select the output port of the Evaluate Model module and select Visualize.



- Click Submit at the top and give your experiment a name.

Set up pipeline run



Experiment

☐ Select existing ☒ Create new

① Customers should not include personal data or other sensitive information in fields marked with because the content in these fields may be logged and shared across Microsoft systems to facilitate operations and troubleshooting. [Learn more](#)

New experiment name *



employeeattrition

Run description *

Employee Attrition

Compute target

Default

cpucluster

Submit

Cancel

- Once you ran the experiment, you can inspect the outputs of the individual steps - check out the output of the different steps, the last one by right-clicking it and selecting Visualize Scored dataset.

Score Model result visualization



Rows ①
441

Columns ②
37

Age	Attrition	BusinessTravel	DailyRate	Department	DistanceFromHome	Education	EducationField	Employ
36	0	Travel_Rarely	1278	Human Resources	8	3	Life Sciences	1
30	0	Travel_Rarely	1358	Research & Development	24	1	Life Sciences	1
33	0	Travel_Frequently	1303	Research & Development	7	2	Life Sciences	1
42	0	Travel_Rarely	810	Research & Development	23	5	Life Sciences	1
28	0	Travel_Frequently	773	Research & Development	6	3	Life Sciences	1
31	0	Travel_Rarely	616	Research & Development	12	3	Medical	1
33	0	Non-Travel	530	Sales	16	3	Life Sciences	1
31	0	Travel_Rarely	1154	Sales	2	2	Life Sciences	1
41	0	Travel_Rarely	642	Research & Development	1	3	Life Sciences	1
30	0	Travel_Rarely	413	Sales	7	1	Marketing	1
50	0	Travel_Rarely	1126	Research & Development	1	2	Medical	1

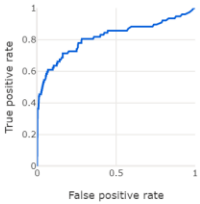
To view, select a column in the table

Close

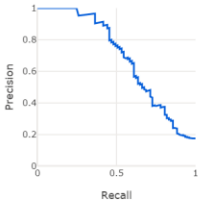
Evaluate Model result visualization

Scored dataset (left port)

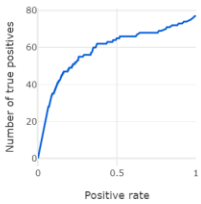
ROC curve



Precision-recall curve

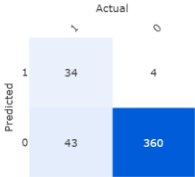


Lift curve



Threshold 0.5

Accuracy 0.893
Precision 0.895
Recall 0.442
F1 Score 0.591
AUC 0.817



Score bin ↓	Positive exam...	Negative exam...	Fraction above thresh...	Accuracy	F1 Score	Precisi...	Recall	Negative precisi...	Negative recall	Cumulative AUC
(0.900,1.000]	3	0	0.007	0.832	0.075	1.000	0.039	0.831	1.000	0.000
(0.800,0.900]	8	0	0.025	0.850	0.250	1.000	0.143	0.847	1.000	0.000

Close