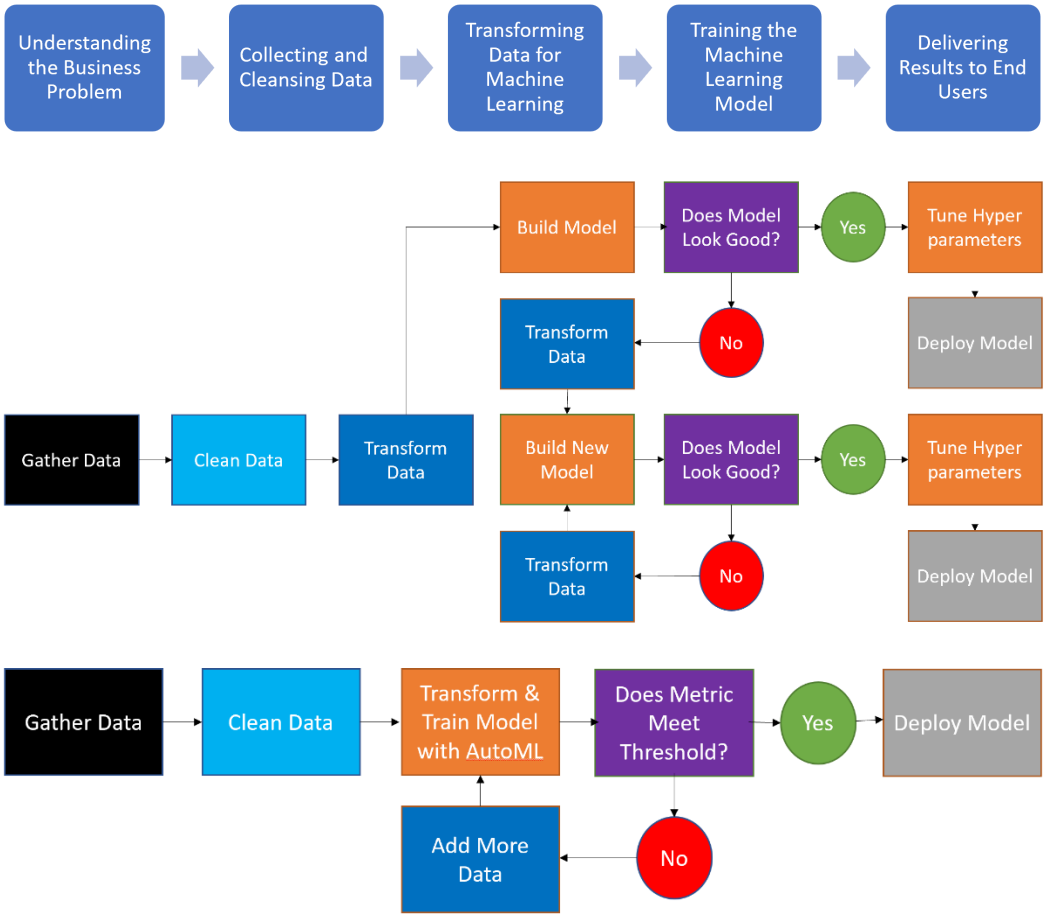
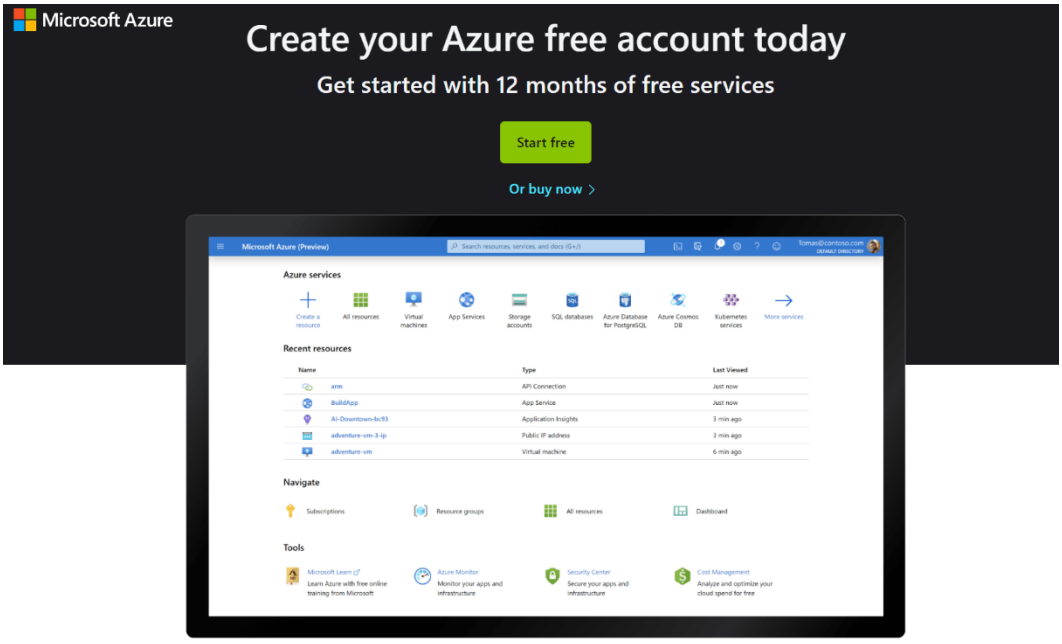
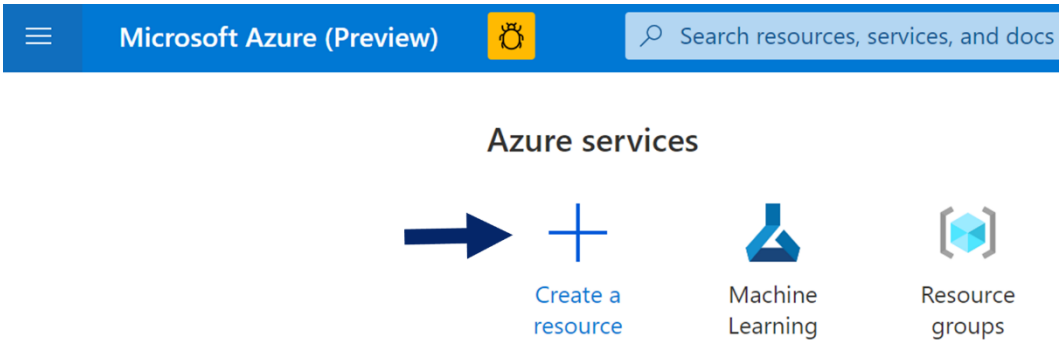
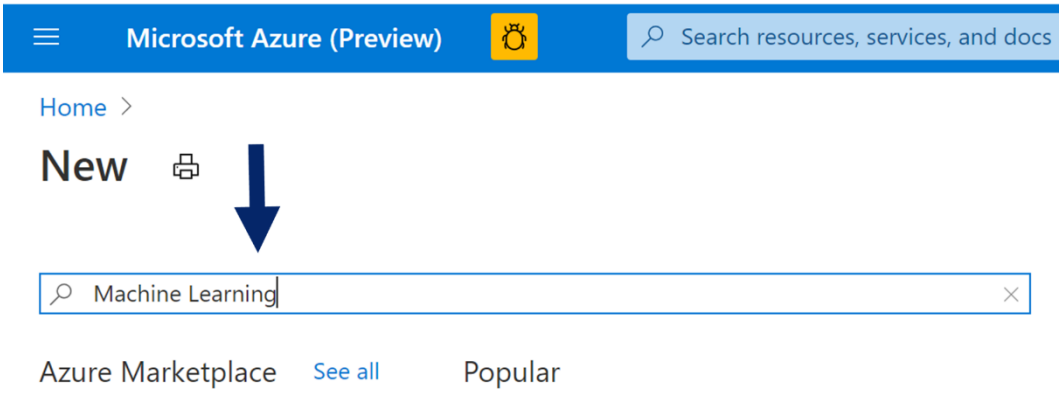
Chapter 1: Introducing AutoML

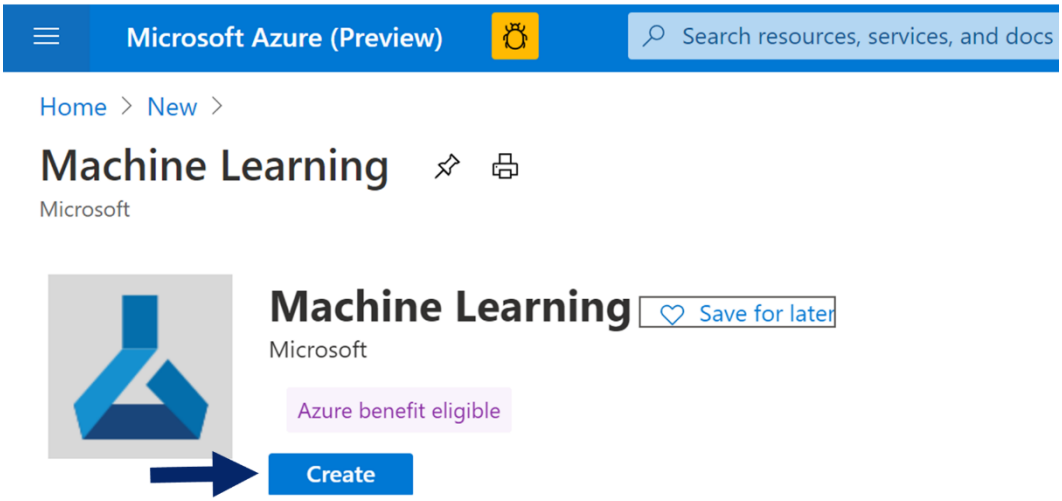


Chapter 2: Getting Started with Azure Machine Learning Service




Microsoft Azure (Preview)
Search resources, services, and docs
Azure services
Create a resource
Machine Learning
Resource groups


Microsoft Azure (Preview)
Search resources, services, and docs
Home
New
Machine Learning
Azure Marketplace
See all
Popular


Microsoft Azure (Preview)
Search resources, services, and docs
Home
New
Machine Learning
Microsoft
Machine Learning
Save for later
Microsoft
Azure benefit eligible
Create

Basics Networking Advanced Tags Review + create

Project details

Select the subscription to manage deployed resources and costs. Use resource groups like folders to organize and manage all your resources.

Subscription * ⓘ — Dennis Sawyers Internal MS Learning Account

Resource group * ⓘ — auto-ml-example-resource-group
Create new

Workspace details

Specify the name and region for the workspace.

Workspace name * ⓘ — automl-example-workspace ✓

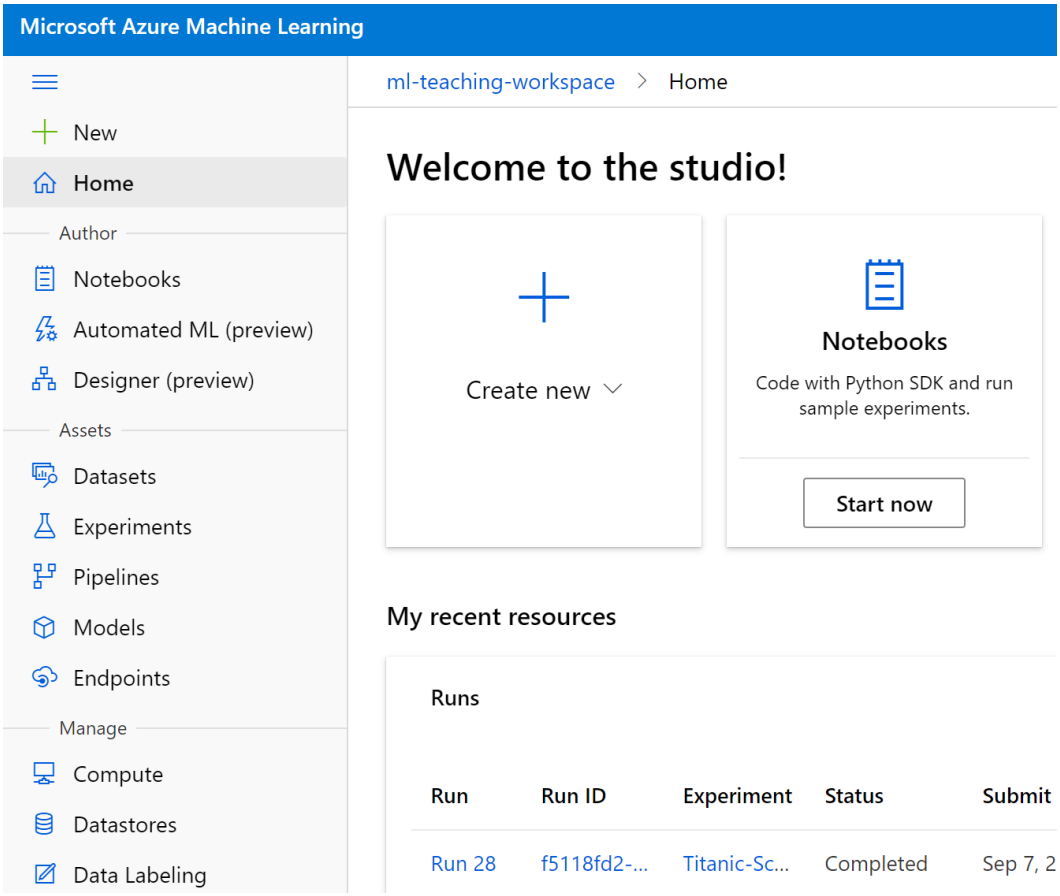
Region * ⓘ — North Central US

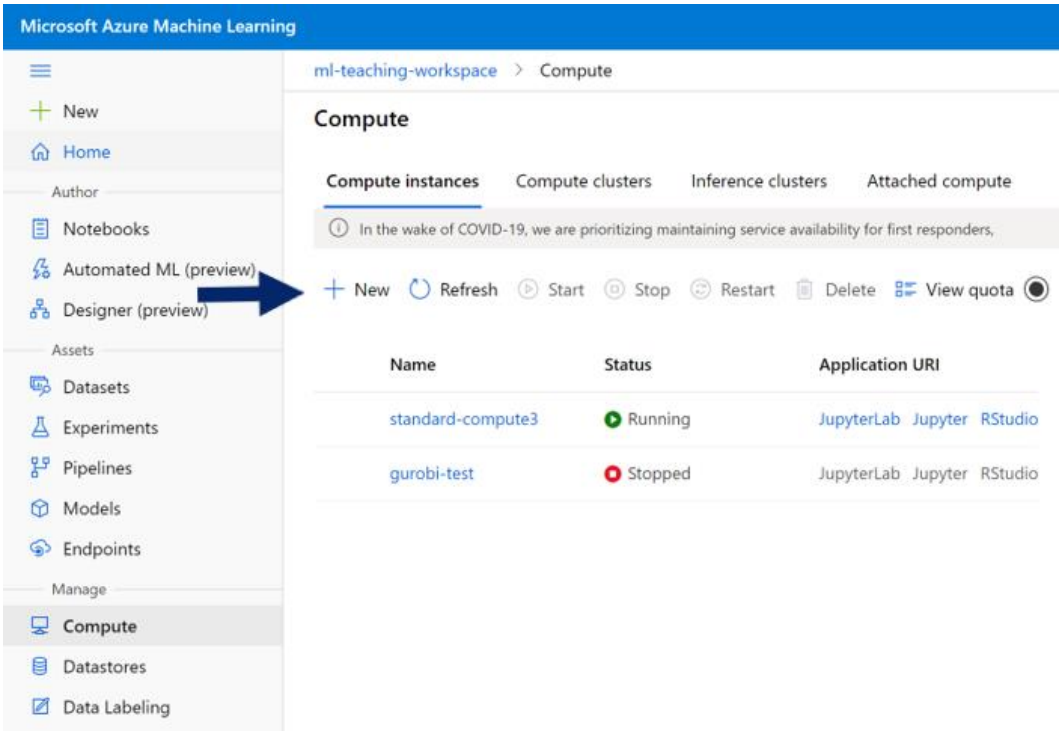
Storage account * ⓘ — (new) automlexamplew1005528502
Create new

Key vault * ⓘ — (new) automlexamplew3892816466
Create new

Application insights * ⓘ — (new) automlexamplew6936691182
Create new

Container registry * ⓘ — None
Create new


Microsoft Azure Machine Learning
New
Home
Author
Notebooks
Automated ML (preview)
Designer (preview)
Assets
Datasets
Experiments
Pipelines
Models
Endpoints
Manage
Compute
Datastores
Data Labeling
ml-teaching-workspace
Home
Welcome to the studio!
Create new
Notebooks
Code with Python SDK and run sample experiments.
Start now
My recent resources
Runs
Run
Run ID
Experiment
Status
Submit
Run 28
f5118fd2-...
Titanic-Sc...
Completed
Sep 7, 2



Select virtual machine

Select the virtual machine size you would like to use for your compute instance. Please note that only the creator can access the compute instance by default. Alternatively, you can assign the access to another user in the advanced settings section.

Region

northcentralus

Virtual machine type

(•) CPU ( ) GPU

Virtual machine size

(•) Select from recommended options ( ) Select from all options

Total available quota: 600 cores

	Name ↑	Category	Workload types	Av...	Cost
( )	Standard_DS2_v2 2 cores, 7GB RAM, 14GB storage	General purpose	Development on Notebooks (or other IDE) and light weight testing	100 c...	$0.15/hr
(•)	Standard_DS3_v2 4 cores, 14GB RAM, 28GB storage	General purpose	Classical ML model training, AutoML runs, pipeline runs (default compute)	100 c...	$0.29/hr
( )	Standard_DS12_v2 4 cores, 28GB RAM, 56GB storage	Memory optimized	Training on large datasets (>1GB) parallel run steps, batch inferencing	100 c...	$0.37/hr
( )	Standard_F4s_v2 4 cores, 8GB RAM, 32GB storage	Compute optimized	Real-time inferencing and other latency-sensitive tasks	100 c...	$0.17/hr

Configure Settings

Configure compute instance settings for your selected virtual machine size.

Name	Category	Cores	Available quota	RAM	Storage	Cost/Hour
Standard_DS3_v2	General purpose	4	100 cores	14 GB	28 GB	$0.29/hr

Compute name * ⓘ

automl-example-compute

Enable SSH access ⓘ

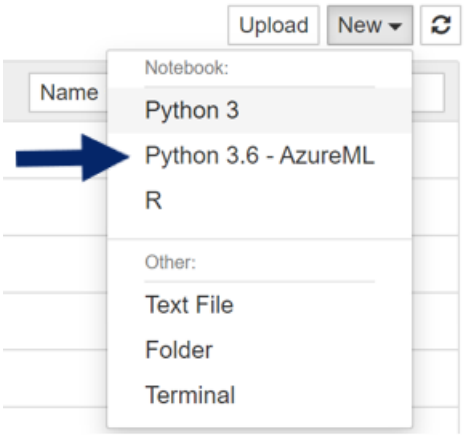
Show advanced settings

Compute

Compute instances | Compute clusters | Inference clusters | Attached compute

+ New | Refresh | Start | Stop | Restart | Delete | View quota | Show all

Name	Status	Application URI
automl-example-comp...	Running	JupyterLab Jupyter VS Code RStudio Terminal



Compute

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

Scale your compute cluster from a single node to a multi node workload

Create a single or multi node compute cluster for your training, batch inferencing or reinforcement learning workloads. Learn more

Create

View Azure Machine Learning tutorials

Select virtual machine

Select the virtual machine size you would like to use for your compute cluster.

Region

northcentralus

Virtual machine priority

(•) Dedicated ( ) Low priority

Virtual machine type

(•) CPU ( ) GPU

Virtual machine size

(•) Select from recommended options ( ) Select from all options

Total available quota: 596 cores

	Name ↑	Category	Workload types	Av...	Cost
( )	Standard_DS2_v2 2 cores, 7GB RAM, 14GB storage	General purpose	Development on Notebooks (or other IDE) and light weight testing	96 co...	$0.15/hr
(•)	Standard_DS3_v2 4 cores, 14GB RAM, 28GB storage	General purpose	Classical ML model training, AutoML runs, pipeline runs (default compute)	96 co...	$0.29/hr
( )	Standard_DS12_v2 4 cores, 28GB RAM, 56GB storage	Memory optimized	Training on large datasets (>1GB) parallel run steps, batch inferencing	96 co...	$0.37/hr
( )	Standard_F4s_v2 4 cores, 8GB RAM, 32GB storage	Compute optimized	Real-time inferencing and other latency-sensitive tasks	100 c...	$0.17/hr

Back | Next | Download a template for automation | Cancel

Configure Settings

Configure compute cluster settings for your selected virtual machine size.

Name	Category	Cores	Available quota	RAM	Storage	Cost/Hour
Standard_DS3_v2	General purpose	4	96 cores	14 GB	28 GB	$0.29/hr

Compute name * ⓘ

automl-compute

Minimum number of nodes * ⓘ

0

Maximum number of nodes * ⓘ

4

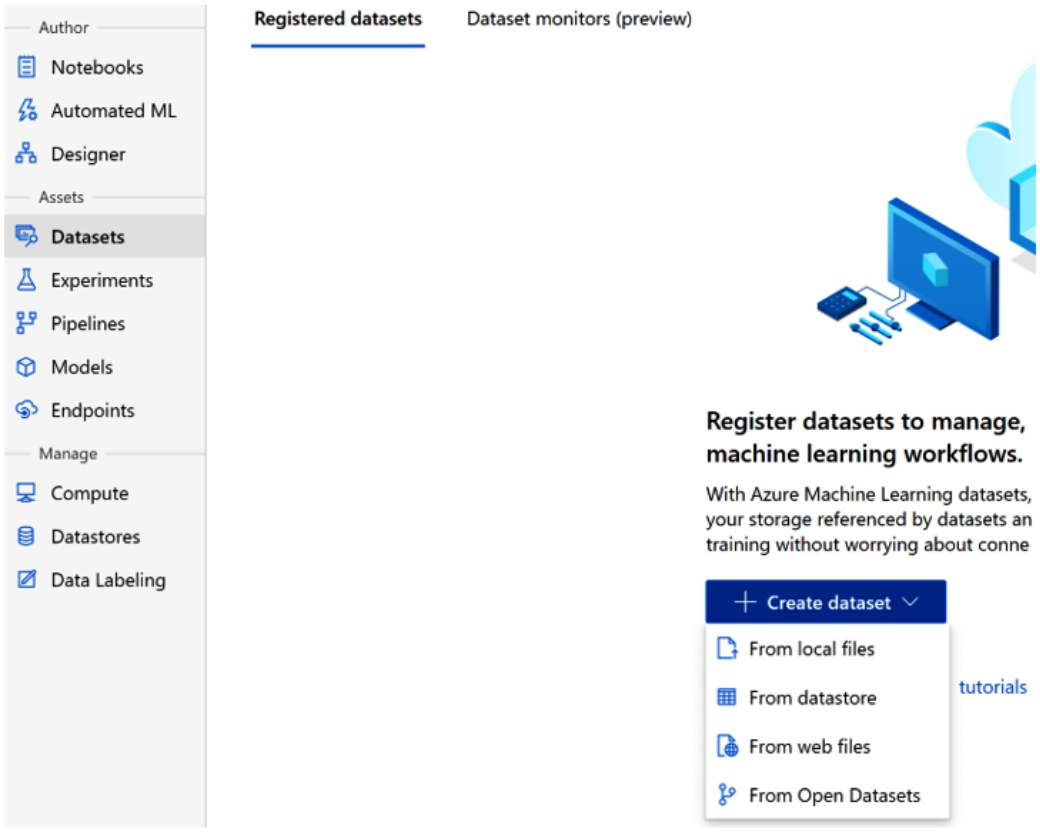
Idle seconds before scale down * ⓘ

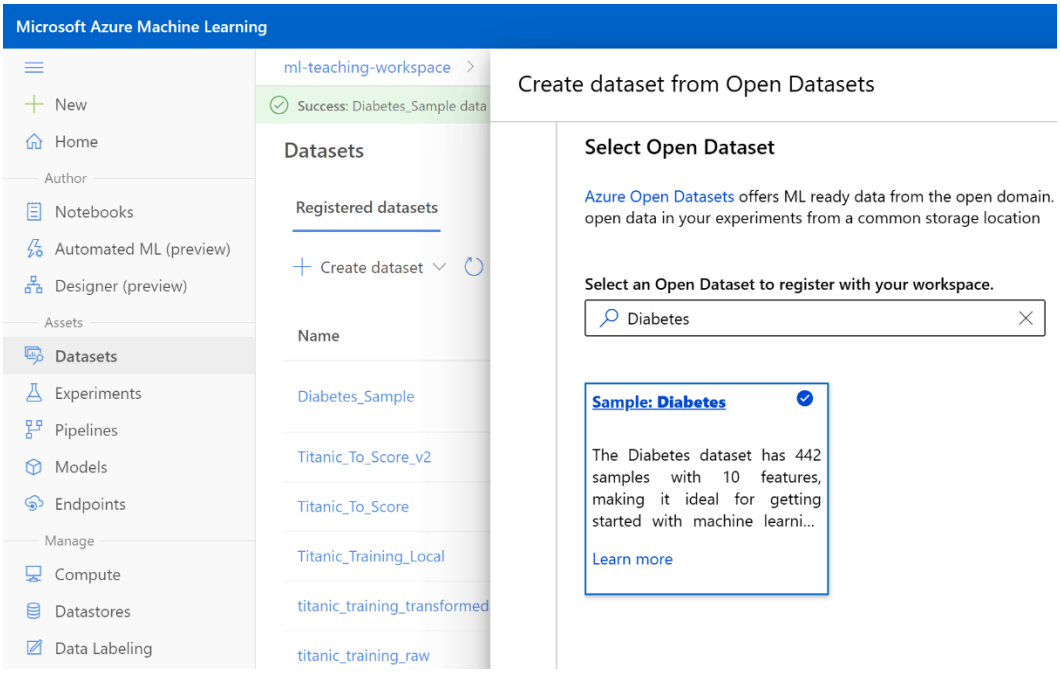
1200

Enable SSH access ⓘ

> Advanced settings

Back | Create | Download a template for automation | Cancel


Author
Notebooks
Automated ML
Designer
Assets
Datasets
Experiments
Pipelines
Models
Endpoints
Manage
Compute
Datastores
Data Labeling
Registered datasets
Dataset monitors (preview)
Register datasets to manage, machine learning workflows.
With Azure Machine Learning datasets, your storage referenced by datasets an training without worrying about conne
Create dataset
From local files
From datastore
From web files
From Open Datasets
tutorials



Dataset details

Azure Open Datasets offers ML ready data from the open domain. Registering open datasets in the workspace lets you data in your experiments from a common storage location without creating a copy of the data in your storage account.

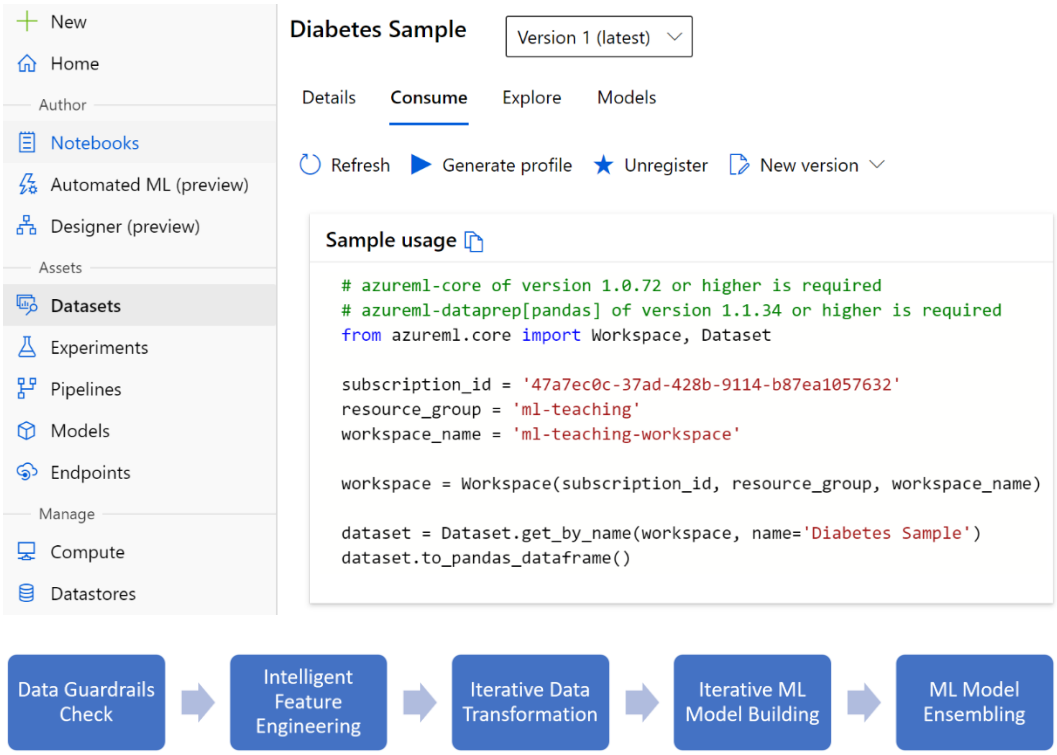
Register Sample: Diabetes

Name *

Diabetes Sample

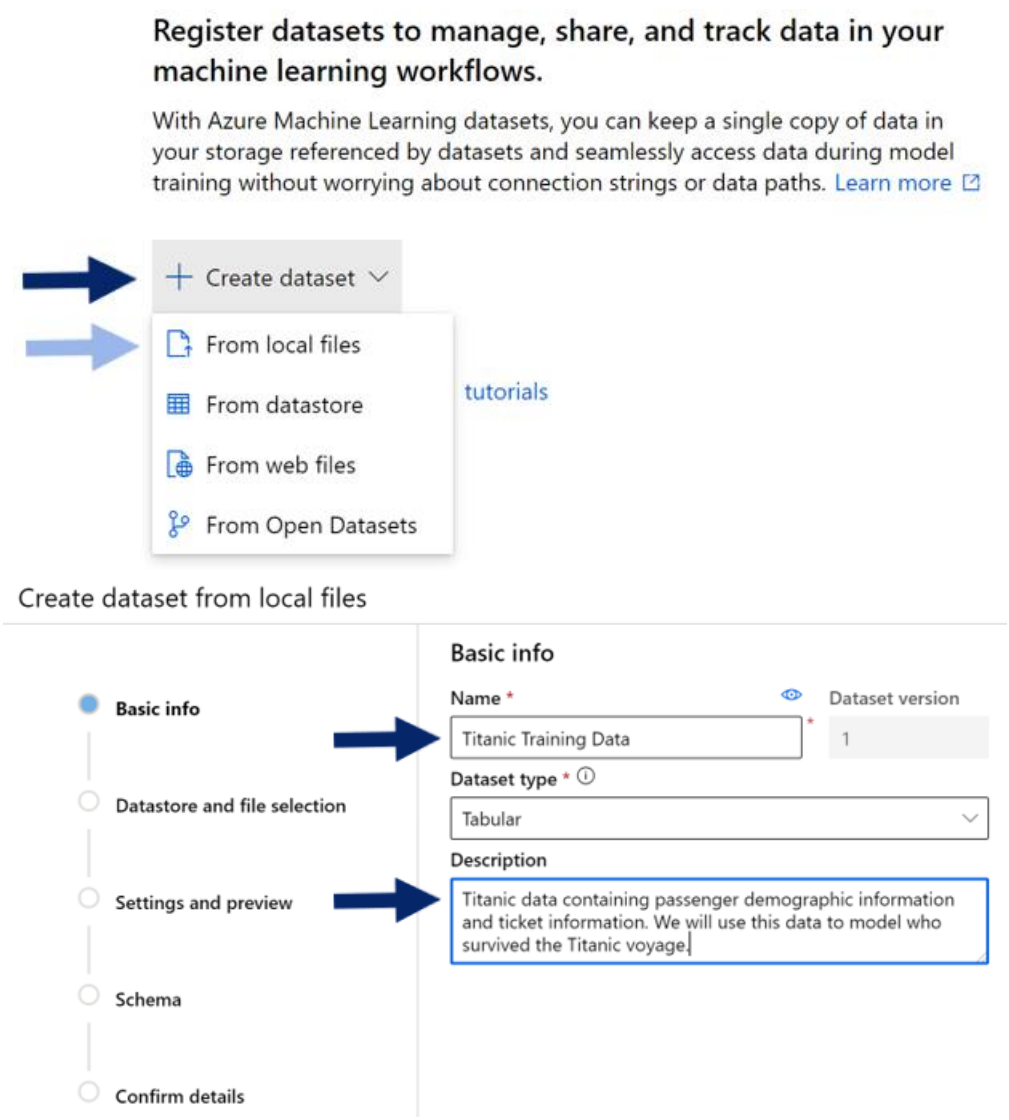
Dataset version

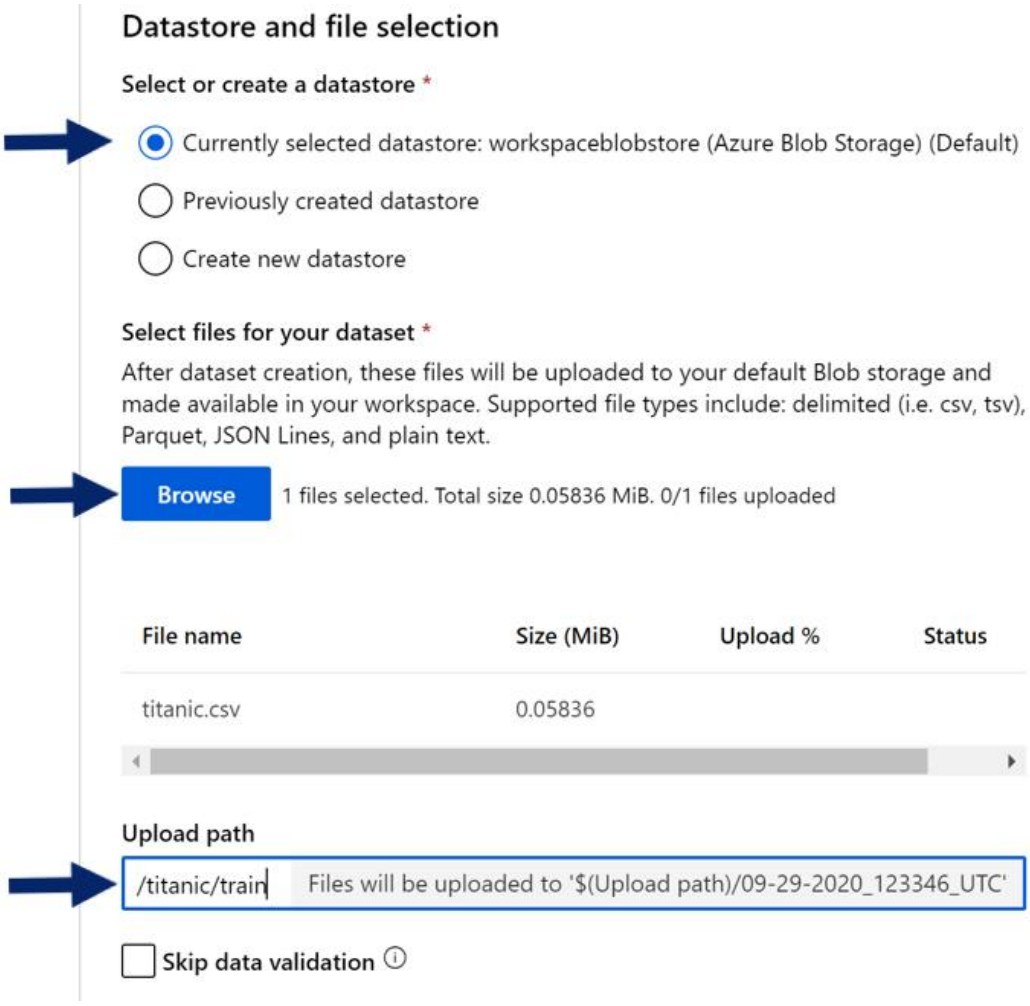
1

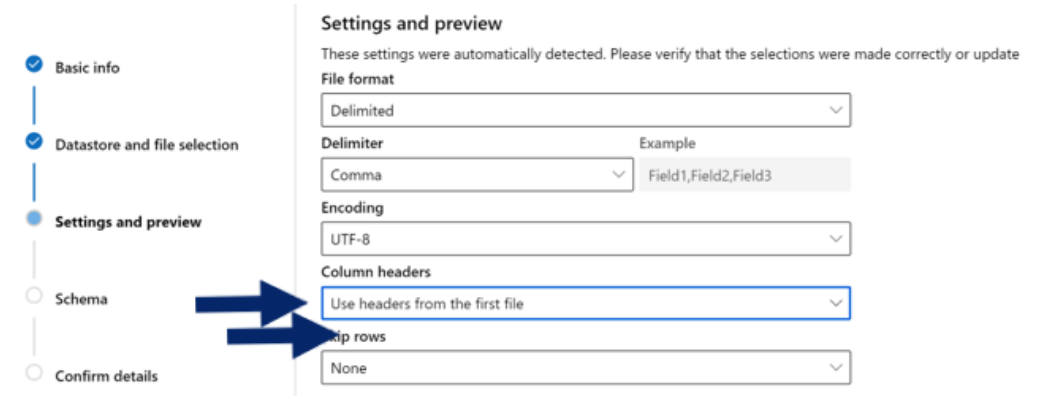

New
Home
Author
Notebooks
Automated ML (preview)
Designer (preview)
Assets
Datasets
Experiments
Pipelines
Models
Endpoints
Manage
Compute
Datastores
Diabetes Sample
Version 1 (latest)
Details
Consume
Explore
Models
Refresh
Generate profile
Unregister
New version
Sample usage
# azureml-core of version 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 = '47a7ec0c-37ad-428b-9114-b87ea1057632'
resource_group = 'ml-teaching'
workspace_name = 'ml-teaching-workspace'
workspace = Workspace(subscription_id, resource_group, workspace_name)
dataset = Dataset.get_by_name(workspace, name='Diabetes Sample')
dataset.to_pandas_dataframe()

Data Guardrails Check
Intelligent Feature Engineering
Iterative Data Transformation
Iterative ML Model Building
ML Model Ensembling

Chapter 3: Training Your First AutoML Model




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
1 files selected. Total size 0.05836 MiB. 0/1 files uploaded
File name
Size (MiB)
Upload %
Status
titanic.csv
0.05836
Upload path
/titanic/train
Files will be uploaded to '$(Upload path)/09-29-2020_123346_UTC'
Skip data validation



Path	Not applicable to selected type	String
PassengerId	Not applicable to selected type	Integer
Survived	Not applicable to selected type	Boolean
Pclass	Not applicable to selected type	Integer
Name	Not applicable to selected type	String
Sex	Not applicable to selected type	String

Confirm details

Basic info

Name
Titanic Training Data

Dataset version
1

Dataset type
Tabular

Description
Titanic data containing passenger demographic information and ticket information. We will use this data to model who survived the Titanic voyage.

Datastore and file selection

Datastore
workspaceblobstore

Selected files (1)
titanic.csv

Path
titanic/train/09-29-2020_123346_UTC/titanic.csv

File settings

File format
Delimited

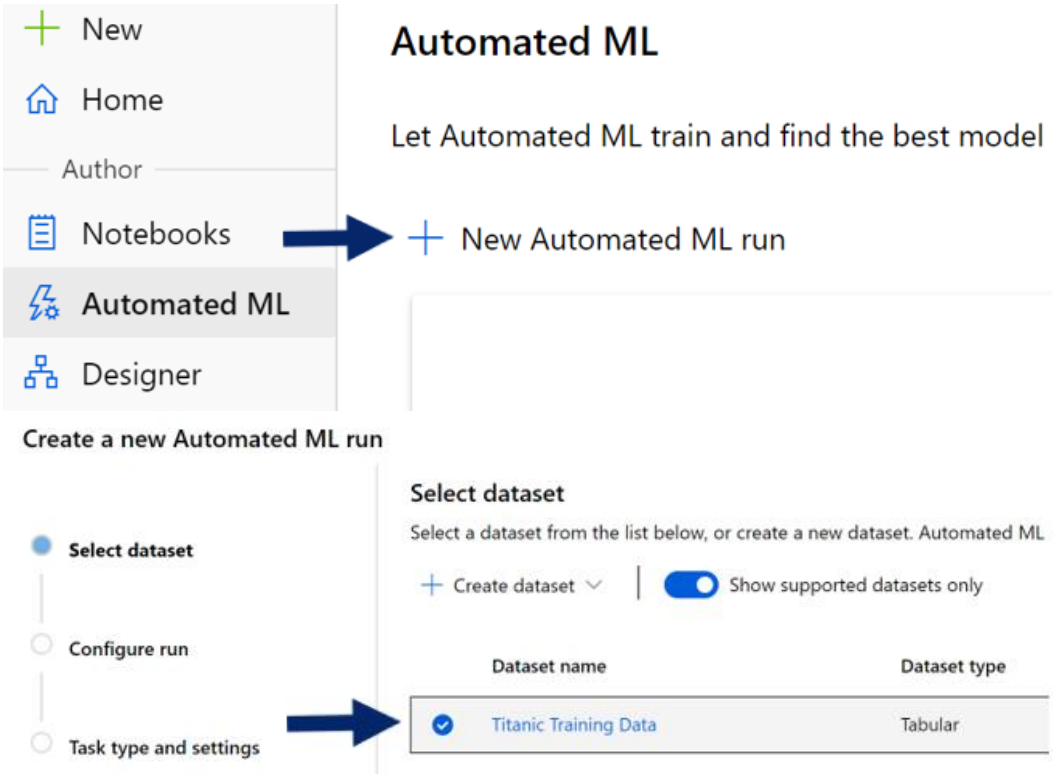
Delimiter
Comma

Encoding
UTF-8

Column headers
Use headers from the first file

Skip rows
None

☐ Profile this dataset after creation


New
Home
Author
Notebooks
Automated ML
Designer
Automated ML
Let Automated ML train and find the best model
New Automated ML 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
Create dataset
Show supported datasets only
Dataset name
Dataset type
Titanic Training Data
Tabular

Configure run

Configure the experiment. Select from existing experiments or define a new name, select the target column and the training compute experiment

Dataset

Titanic Training Data (View dataset)

Experiment name *

◯ Select existing ◉ Create new

New experiment name

Titanic-Training

Target column *

Survived

Select compute cluster *

compute-cluster

Create a new compute Refresh compute

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**

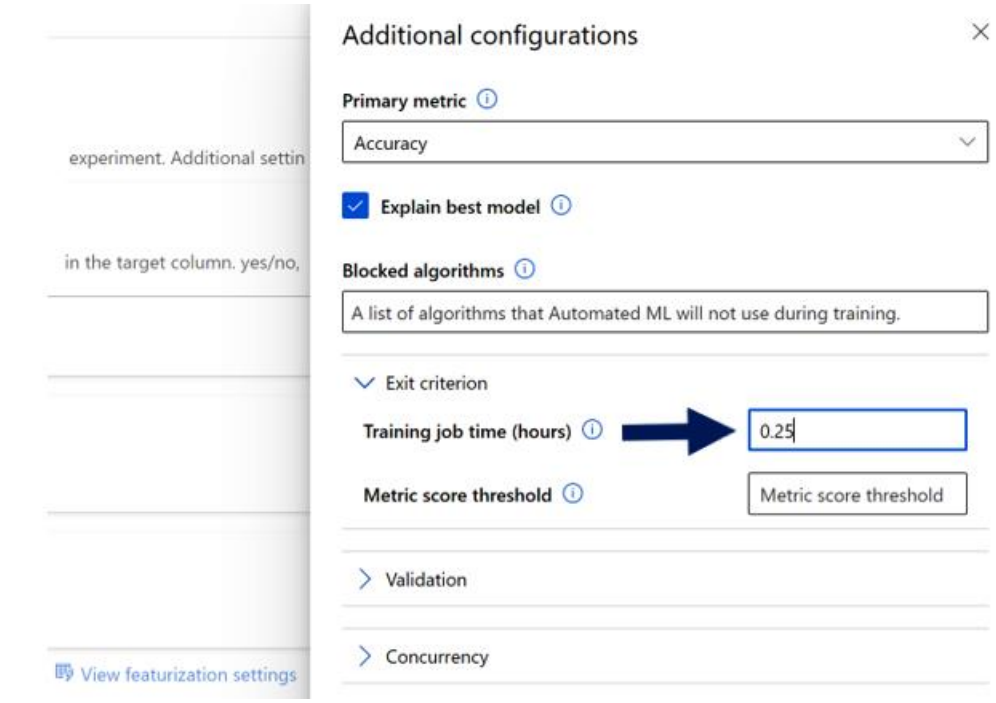
Regression

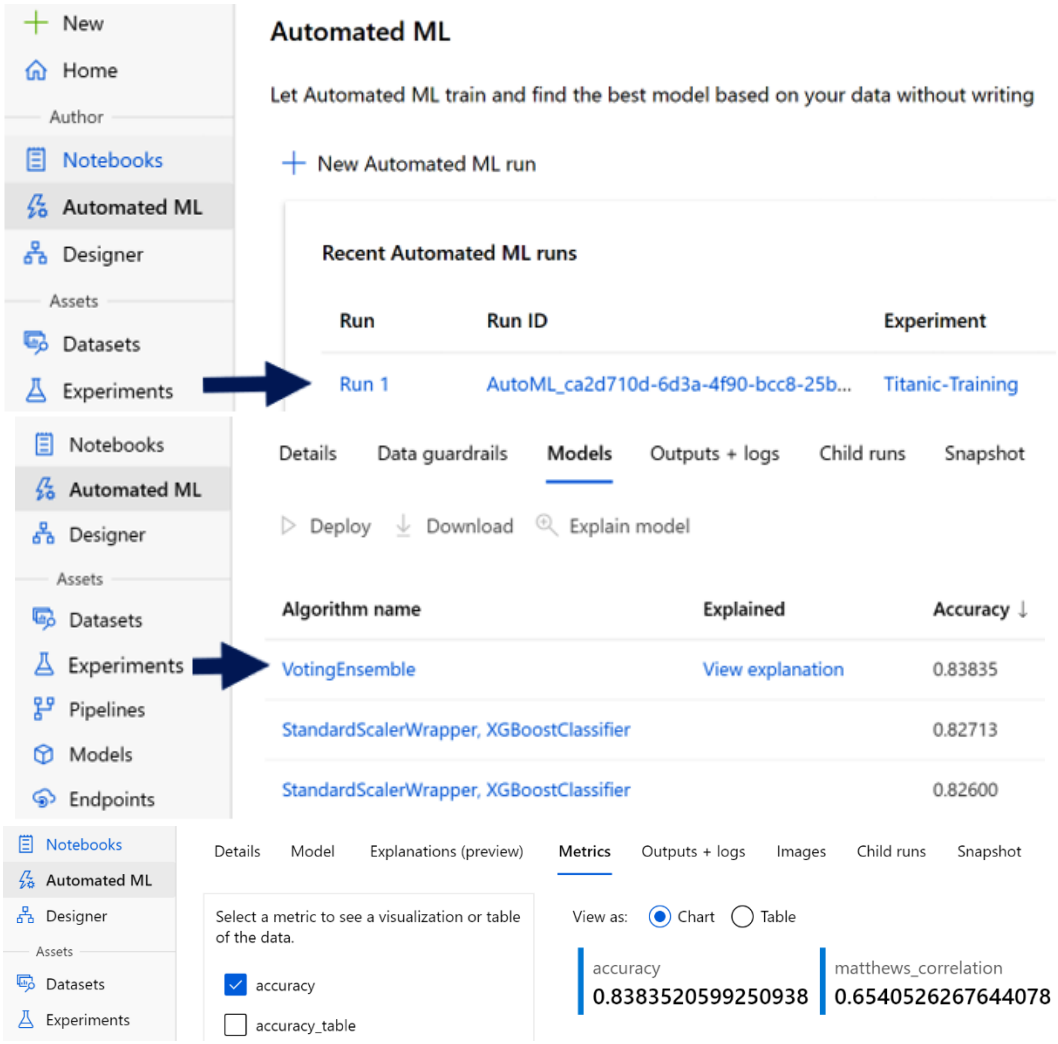
To predict continuous numeric values

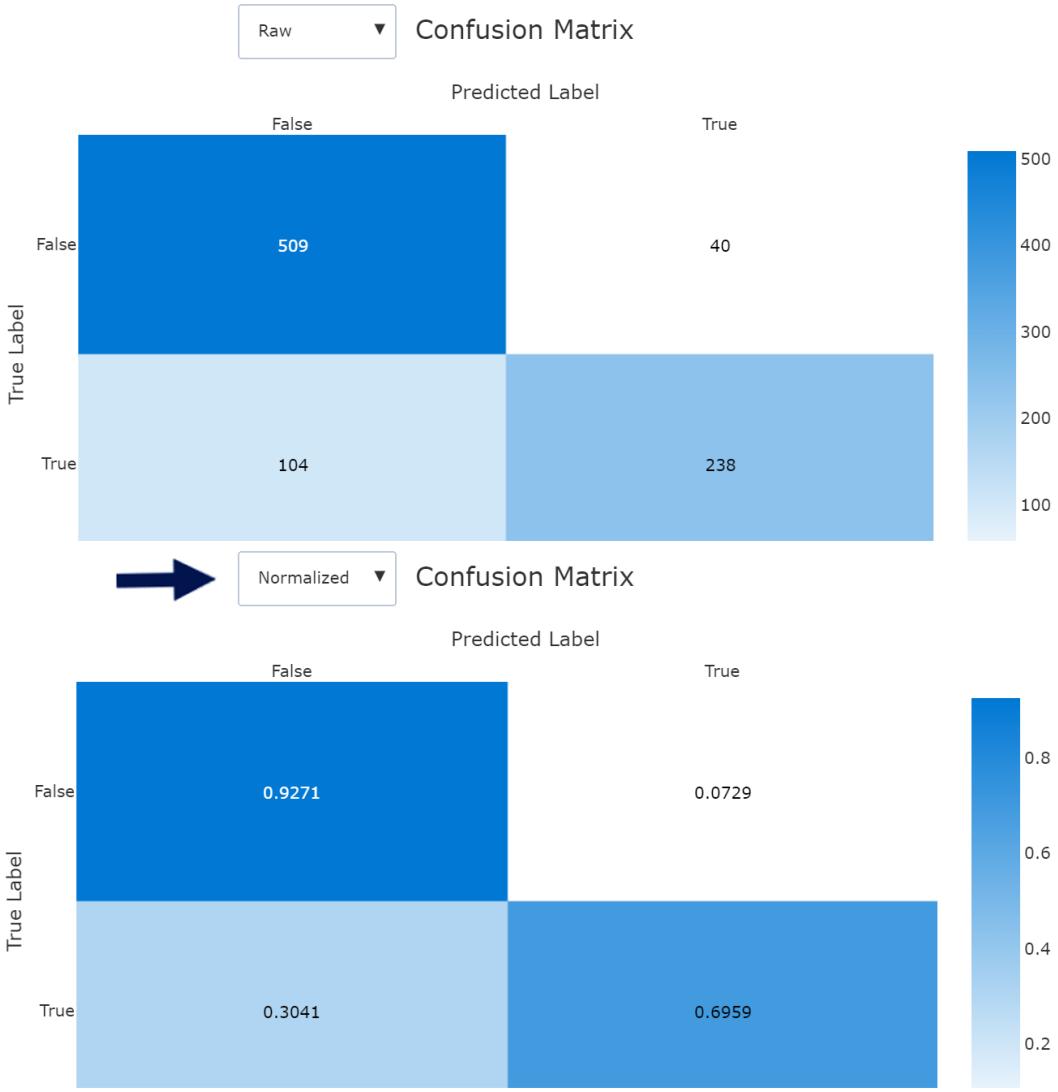
Time series forecasting

To predict values based on time

View additional configuration settings View featurization settings




New
Home
Author
Notebooks
Automated ML
Designer
Assets
Datasets
Experiments
Automated ML
Let Automated ML train and find the best model based on your data without writing
New Automated ML run
Recent Automated ML runs
Run
Run ID
Experiment
Run 1
AutoML_ca2d710d-6d3a-4f90-bcc8-25b...
Titanic-Training
Details
Data guardrails
Models
Outputs + logs
Child runs
Snapshot
Deploy
Download
Explain model
Algorithm name
Explained
Accuracy
VotingEnsemble
View explanation
0.83835
StandardScalerWrapper, XGBoostClassifier
0.82713
StandardScalerWrapper, XGBoostClassifier
0.82600
Notebooks
Automated ML
Designer
Assets
Datasets
Experiments
Pipelines
Models
Endpoints
Details
Model
Explanations (preview)
Metrics
Outputs + logs
Images
Child runs
Snapshot
Select a metric to see a visualization or table of the data.
accuracy
accuracy_table
View as:
Chart
Table
accuracy
0.8383520599250938
matthews_correlation
0.6540526267644078


Raw
Confusion Matrix
Predicted Label
False
True
False
True
True Label
509
40
104
238
500
400
300
200
100
Normalized
Confusion Matrix
Predicted Label
False
True
False
True
True Label
0.9271
0.0729
0.3041
0.6959
0.8
0.6
0.4
0.2

Select Explanation

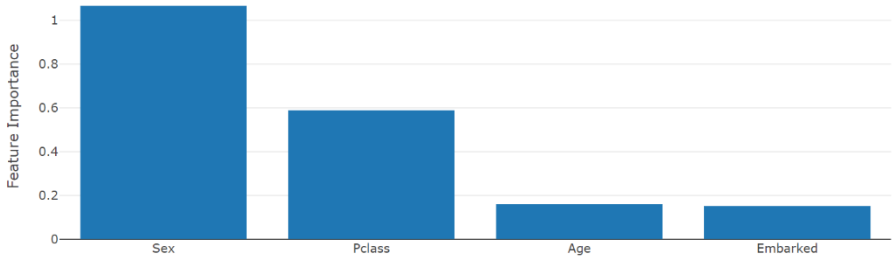
tabular | mimic.lightgbm | raw | classification | 4381fdb3-8165-4f04-8bcc-8e77d2e41987 | 9/29/2020, 10:39:44 PM

Explainer: mimic.lightgbm

Global Importance

Summary Importance

Top K Features:



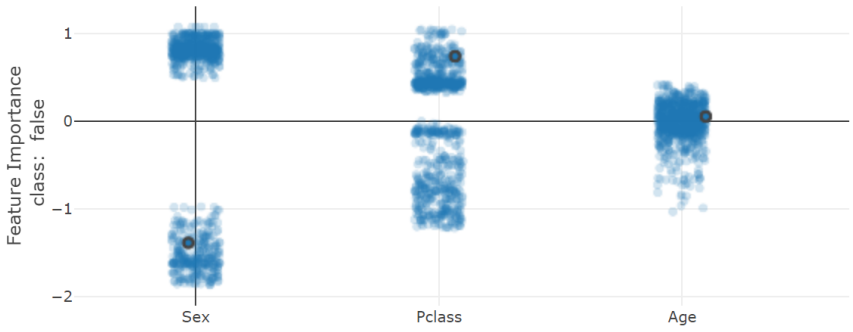
Global Importance

Summary Importance

Chart type:

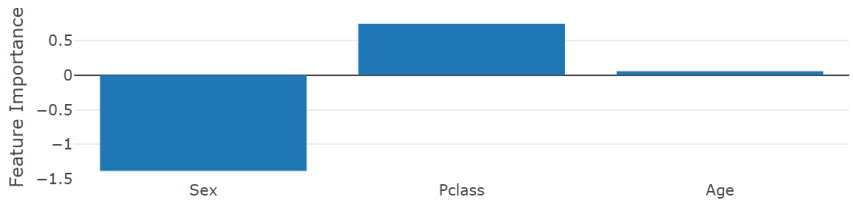
Swarm

Top K Features:



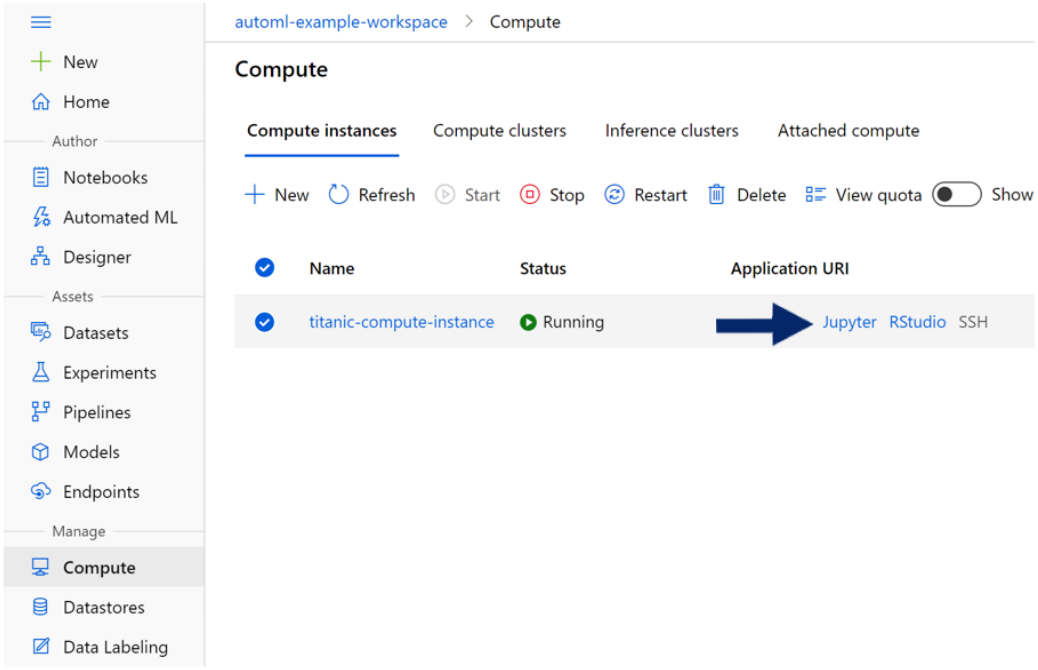
Local Feature Importance

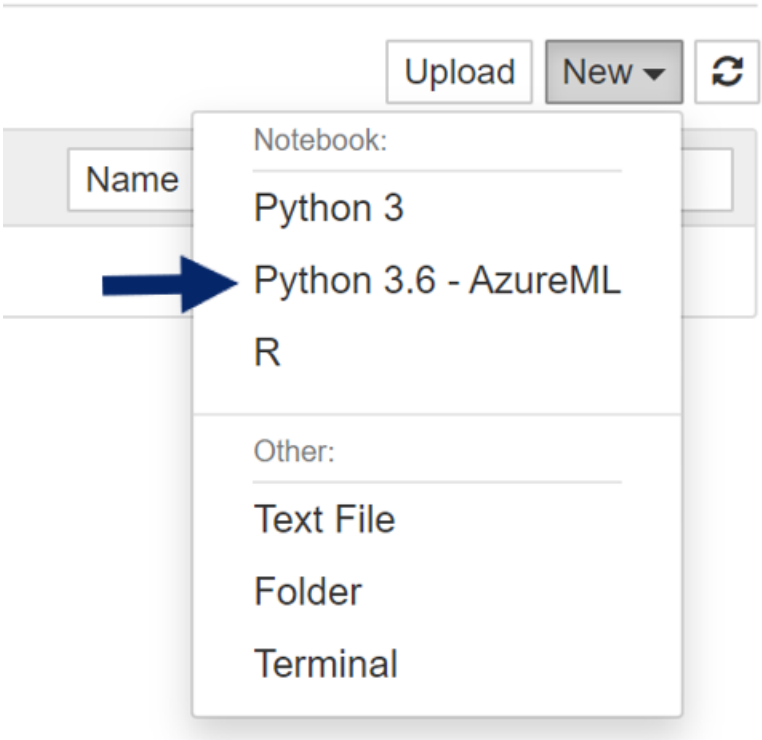
Top K Features:

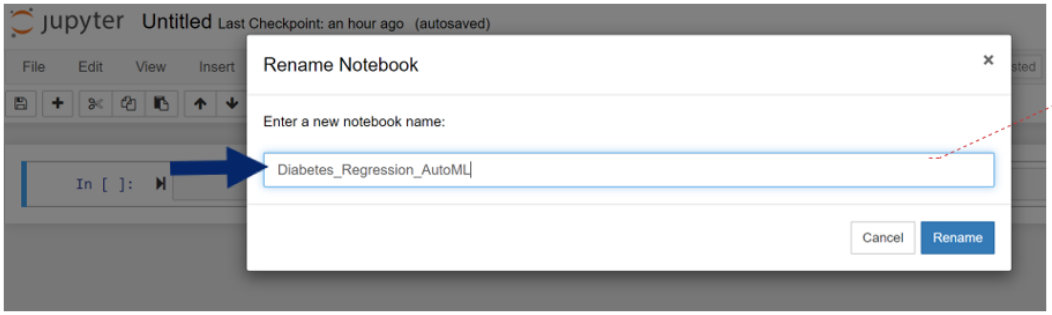


Chapter 4: Building an AutoML Regression Solution

Standard regression algorithms	Tree algorithms	Gradient boosting algorithms	Nearest neighbor algorithms	Optimization algorithms
Elastic net	Decision tree	XGBoost	KNN	SGD
LARS Lasso	Random forest	LightGBM		Online gradient descent regressor
	Extremely randomized trees	Gradient boosting		Fast linear regressor




Upload
New
Name
Notebook:
Python 3
Python 3.6 - AzureML
R
Other:
Text File
Folder
Terminal


jupyter
Untitled
Last Checkpoint: an hour ago (autosaved)
File
Edit
View
Insert
In [ ]:
Rename Notebook
Enter a new notebook name:
Diabetes_Regression_AutoML
Cancel
Rename

In [28]:
```
# View your dataset by converting to pandas
dataset.take(10).to_pandas_dataframe()
```

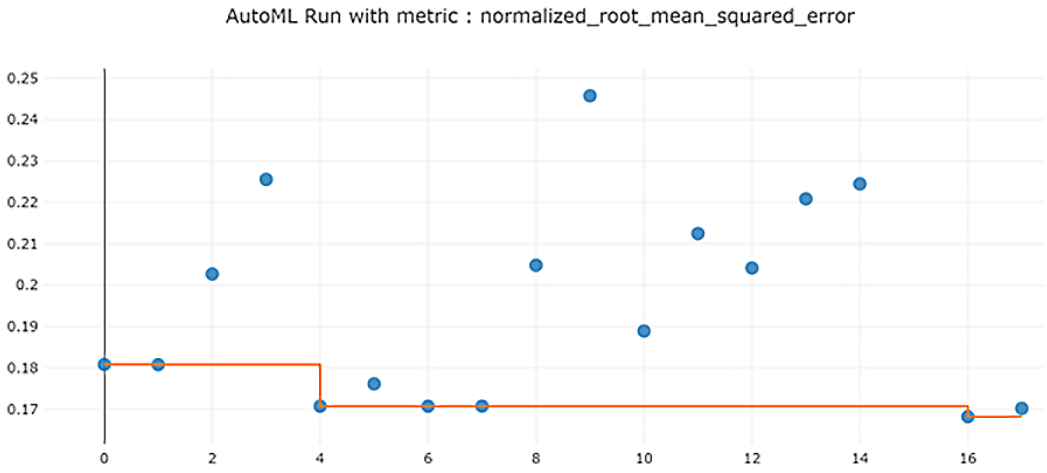
Out[28]:

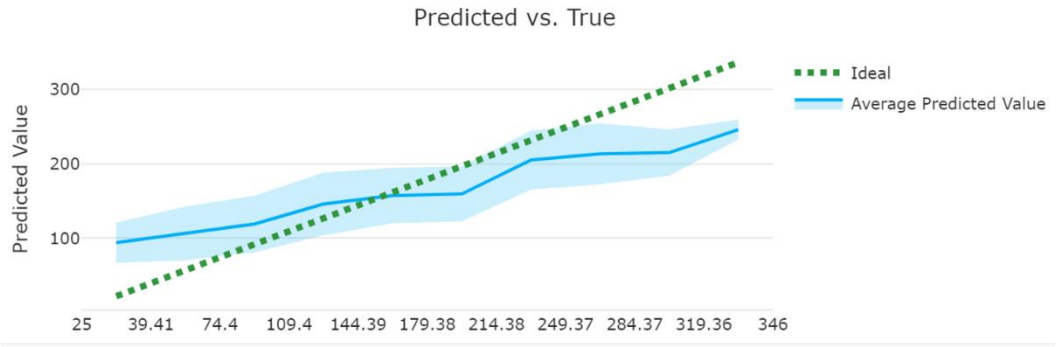
	AGE	SEX	BMI	BP	S1	S2	S3	S4	S5	S6	Y
0	59	2	32.1	101.0	157	93.2	38.0	4.00	4.8598	87	151
1	48	1	21.6	87.0	183	103.2	70.0	3.00	3.8918	69	75
2	72	2	30.5	93.0	156	93.6	41.0	4.00	4.6728	85	141
3	24	1	25.3	84.0	198	131.4	40.0	5.00	4.8903	89	206
4	50	1	23.0	101.0	192	125.4	52.0	4.00	4.2905	80	135
5	23	1	22.6	89.0	139	64.8	61.0	2.00	4.1897	68	97
6	36	2	22.0	90.0	160	99.6	50.0	3.00	3.9512	82	138
7	66	2	26.2	114.0	255	185.0	56.0	4.55	4.2485	92	63
8	60	2	32.1	83.0	179	119.4	42.0	4.00	4.4773	94	110
9	29	1	30.0	85.0	180	93.4	43.0	4.00	5.3845	88	310

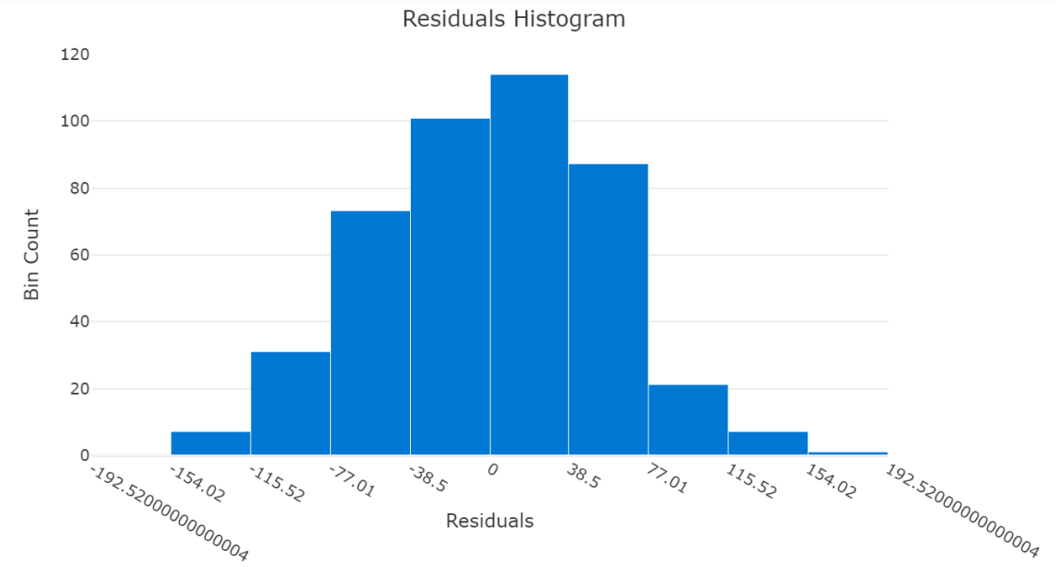
```
****************************************************************************************************
DATA GUARDRAILS:

TYPE:         Missing feature values imputation
STATUS:       PASSED
DESCRIPTION:  No feature missing values were detected in the training data.
              Learn more about missing value imputation: https://aka.ms/AutomatedMLFeaturization

****************************************************************************************************

TYPE:         High cardinality feature detection
STATUS:       PASSED
DESCRIPTION:  Your inputs were analyzed, and no high cardinality features were detected.
              Learn more about high cardinality feature handling: https://aka.ms/AutomatedMLFeaturization

****************************************************************************************************
```

```
*********************************************************************************************
ITERATION: The iteration being evaluated.
PIPELINE: A summary description of the pipeline being evaluated.
DURATION: Time taken for the current iteration.
METRIC: The result of computing score on the fitted pipeline.
BEST: The best observed score thus far.
*********************************************************************************************
```

ITERATION	PIPELINE	DURATION	METRIC	BEST
0	MaxAbsScaler LightGBM	0:00:33	0.1808	0.1808
1	MaxAbsScaler XGBoostRegressor	0:00:25	0.1808	0.1808
2	MaxAbsScaler DecisionTree	0:00:30	0.2027	0.1808
3	MinMaxScaler DecisionTree	0:00:31	0.2255	0.1808
4	StandardScalerWrapper LassoLars	0:00:29	0.1708	0.1708
5	RobustScaler ElasticNet	0:00:35	0.1761	0.1708
6	StandardScalerWrapper LassoLars	0:00:32	0.1708	0.1708
7	RobustScaler ElasticNet	0:00:36	0.1708	0.1708
8	RobustScaler DecisionTree	0:00:31	0.2048	0.1708
9	MinMaxScaler DecisionTree	0:00:30	0.2458	0.1708
10	RobustScaler DecisionTree	0:00:31	0.1889	0.1708
11	MinMaxScaler DecisionTree	0:00:32	0.2125	0.1708
12	StandardScalerWrapper DecisionTree	0:00:29	0.2041	0.1708
13	StandardScalerWrapper DecisionTree	0:02:19	0.2209	0.1708
14	RobustScaler DecisionTree	0:00:31	0.2244	0.1708
15		0:00:08	nan	0.1708
16	VotingEnsemble	0:00:52	0.1682	0.1682
17	StackEnsemble	0:00:52	0.1702	0.1682

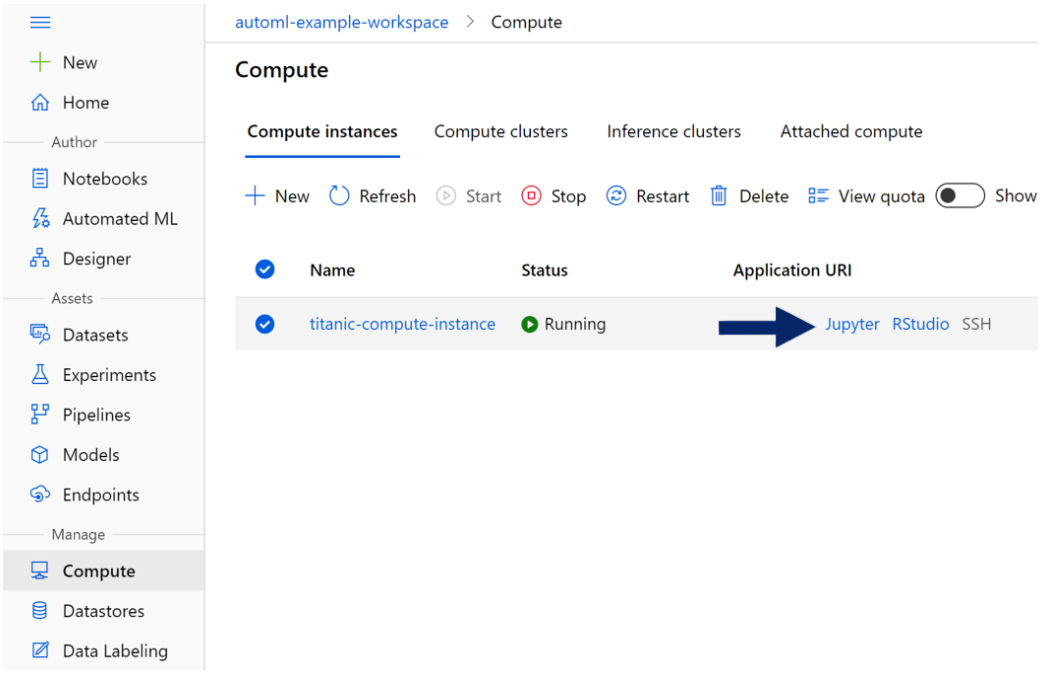


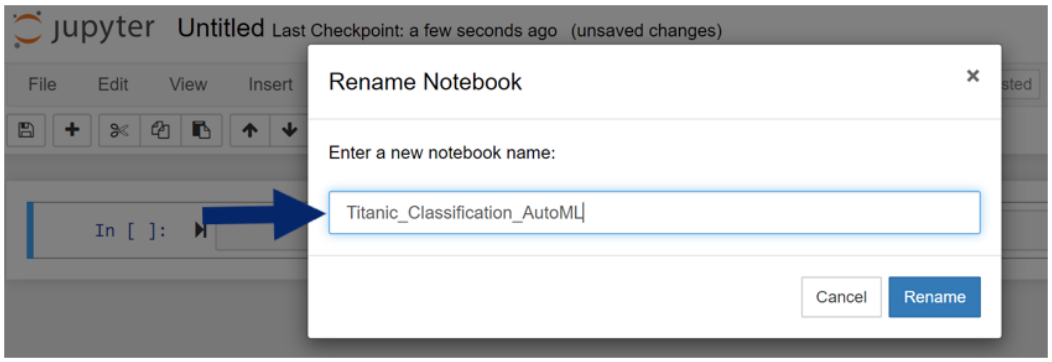





```
(25.0, 75.128]
(75.128, 152.133]
(152.133, 229.139]
(229.139, 346.0]
```

Chapter 5: Building an AutoML Classification Solution





In [7]:
```
# View your dataset by converting to pandas
dataset.take(10).to_pandas_dataframe()
```

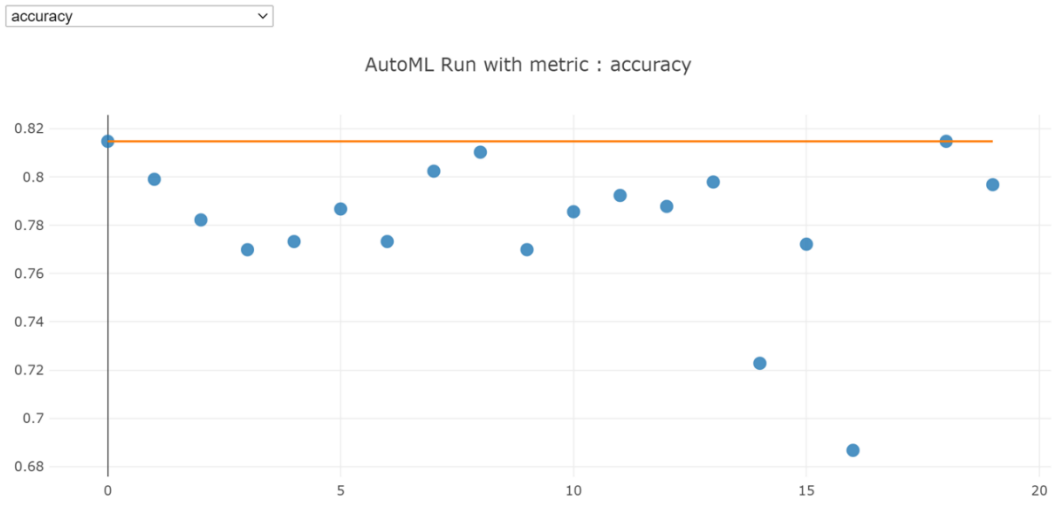
Out[7]:

	Survived	Pclass	Sex	Age	SibSp	Parch	Fare	Embarked
0	False	3	male	22.0	1	0	7.2500	S
1	True	1	female	38.0	1	0	71.2833	C
2	True	3	female	26.0	0	0	7.9250	S
3	True	1	female	35.0	1	0	53.1000	S
4	False	3	male	35.0	0	0	8.0500	S
5	False	3	male	NaN	0	0	8.4583	Q
6	False	1	male	54.0	0	0	51.8625	S
7	False	3	male	2.0	3	1	21.0750	S
8	True	3	female	27.0	0	2	11.1333	S
9	True	2	female	14.0	1	0	30.0708	C

```
****************************************************************************************************
DATA GUARDRAILS:

TYPE:         Class balancing detection
STATUS:       PASSED
DESCRIPTION:  Your inputs were analyzed, and all classes are balanced in your training data.
              Learn more about imbalanced data: https://aka.ms/AutomatedMLImbalancedData

****************************************************************************************************

TYPE:         Missing feature values imputation
STATUS:       DONE
DESCRIPTION:  If the missing values are expected, let the run complete. Otherwise cancel the current run and use a script to
customize the handling of missing feature values that may be more appropriate based on the data type and business requiremen
t.
              Learn more about missing value imputation: https://aka.ms/AutomatedMLFeaturization
DETAILS:
+---------------------------------+---------------------------------+
|Column name                      |Missing value count              |
+=================================+=================================+
|Embarked                         |2                                |
+---------------------------------+---------------------------------+

****************************************************************************************************

TYPE:         High cardinality feature detection
STATUS:       PASSED
DESCRIPTION:  Your inputs were analyzed, and no high cardinality features were detected.
              Learn more about high cardinality feature handling: https://aka.ms/AutomatedMLFeaturization

****************************************************************************************************
```

```
*********************************************************************************************
ITERATION: The iteration being evaluated.
PIPELINE: A summary description of the pipeline being evaluated.
DURATION: Time taken for the current iteration.
METRIC: The result of computing score on the fitted pipeline.
BEST: The best observed score thus far.
*********************************************************************************************
```

ITERATION	PIPELINE	DURATION	METRIC	BEST
0	MaxAbsScaler LightGBM	0:00:33	0.8148	0.8148
1	MaxAbsScaler XGBoostClassifier	0:00:29	0.7991	0.8148
2	MaxAbsScaler RandomForest	0:00:28	0.7822	0.8148
3	MaxAbsScaler RandomForest	0:00:30	0.7699	0.8148
4	MaxAbsScaler SGD	0:00:29	0.7733	0.8148
5	MaxAbsScaler SGD	0:00:33	0.7868	0.8148
6	MaxAbsScaler ExtremeRandomTrees	0:00:34	0.7733	0.8148
7	MaxAbsScaler ExtremeRandomTrees	0:00:29	0.8024	0.8148
8	MaxAbsScaler ExtremeRandomTrees	0:00:29	0.8103	0.8148
9	MaxAbsScaler ExtremeRandomTrees	0:00:31	0.7699	0.8148
10	MaxAbsScaler SGD	0:00:32	0.7856	0.8148
11	MaxAbsScaler SGD	0:00:33	0.7924	0.8148
12	MaxAbsScaler RandomForest	0:00:32	0.7878	0.8148
13	StandardScalerWrapper ExtremeRandomTrees	0:00:31	0.7979	0.8148
14	MaxAbsScaler RandomForest	0:00:29	0.7228	0.8148
15	MaxAbsScaler SGD	0:00:32	0.7722	0.8148
16	MaxAbsScaler RandomForest	0:00:33	0.6868	0.8148
17		0:00:08	nan	0.8148
18	VotingEnsemble	0:00:53	0.8148	0.8148
19	StackEnsemble	0:00:56	0.7968	0.8148



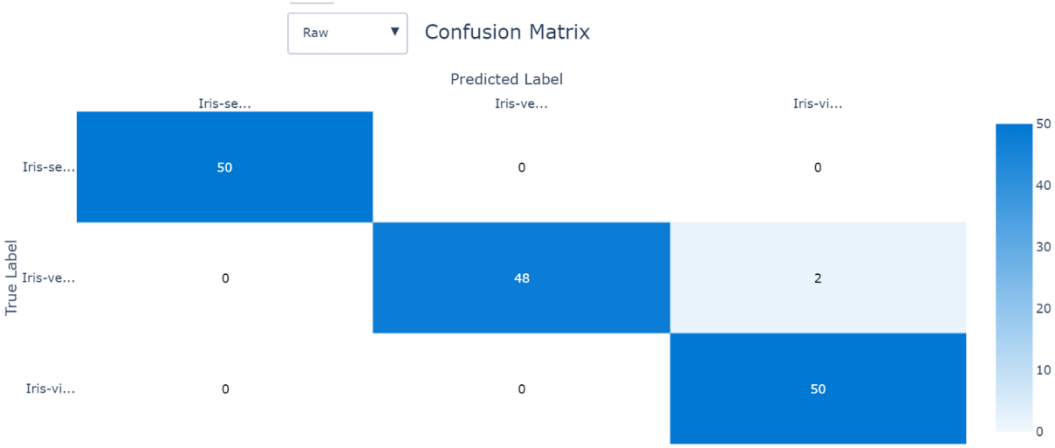
In [8]:
```
# View your dataset by converting to pandas
dataset.take(10).to_pandas_dataframe()
```

Out[8]:

	sepal_length	sepal_width	petal_length	petal_width	species
0	5.1	3.5	1.4	0.2	Iris-setosa
1	4.9	3.0	1.4	0.2	Iris-setosa
2	4.7	3.2	1.3	0.2	Iris-setosa
3	4.6	3.1	1.5	0.2	Iris-setosa
4	5.0	3.6	1.4	0.2	Iris-setosa
5	5.4	3.9	1.7	0.4	Iris-setosa
6	4.6	3.4	1.4	0.3	Iris-setosa
7	5.0	3.4	1.5	0.2	Iris-setosa
8	4.4	2.9	1.4	0.2	Iris-setosa
9	4.9	3.1	1.5	0.1	Iris-setosa

ITERATION	PIPELINE	DURATION	METRIC	BEST
0	MaxAbsScaler LightGBM	0:00:42	0.9467	0.9467
1	MaxAbsScaler XGBoostClassifier	0:00:49	0.9467	0.9467
2	MinMaxScaler RandomForest	0:00:55	0.9733	0.9733
3	MinMaxScaler RandomForest	0:00:40	0.9467	0.9733
4	MinMaxScaler RandomForest	0:00:49	0.9400	0.9733
5	MinMaxScaler SVM	0:00:52	0.9467	0.9733
6	MaxAbsScaler GradientBoosting	0:00:42	0.9533	0.9733
7	SparseNormalizer XGBoostClassifier	0:00:43	0.9200	0.9733
8	StandardScalerWrapper RandomForest	0:00:49	0.9533	0.9733
9	SparseNormalizer LightGBM	0:00:45	0.9733	0.9733
10	StandardScalerWrapper LightGBM	0:00:39	0.9400	0.9733
11	RobustScaler ExtremeRandomTrees	0:00:39	0.9533	0.9733
12	SparseNormalizer XGBoostClassifier	0:00:42	0.9667	0.9733
13	SparseNormalizer XGBoostClassifier	0:00:45	0.9067	0.9733
14	SparseNormalizer XGBoostClassifier	0:00:43	0.9133	0.9733
15	RobustScaler KNN	0:00:51	0.9533	0.9733
16	VotingEnsemble	0:00:57	0.9867	0.9867
17	StackEnsemble	0:01:05	0.9867	0.9867

Iteration	Pipeline	Iteration metric	Best metric	Status	Duration	Started
16	VotingEnsemble	0.98666667	0.98666667	Completed	0:01:07	Jan 18, 2021 9:18 PM
17	StackEnsemble	0.98666667	0.98666667	Completed	0:01:16	Jan 18, 2021 9:19 PM
2	MinMaxScaler, RandomForest	0.97333333	0.97333333	Completed	0:01:07	Jan 18, 2021 9:04 PM
9	SparseNormalizer, LightGBM	0.97333333	0.97333333	Completed	0:00:55	Jan 18, 2021 9:11 PM
12	SparseNormalizer, XGBoostClassifier	0.96666667	0.97333333	Completed	0:00:52	Jan 18, 2021 9:14 PM



Standard Classification Algorithms	Tree Algorithms	Gradient Boosting Algorithms	Support Vector Algorithms	Optimization Algorithms
Logistic regression	Decision tree	XGBoost	SVC	SGD
Naïve Bayes	Random forest	LightGBM	Linear SVC	Averaged perceptron classifier
K-nearest neighbors	Extremely randomized trees	Gradient boosting	Linear SVM classifier	

Chapter 6: Building an AutoML Forecasting Solution

```
In [47]: df.head(10)
```

Out[47]:

	WeekStarting	Store	Brand	Quantity	Advert	Price	Revenue
0	1990-06-14	1000	dominicks	12003	1	2.59	31087.77
1	1990-06-21	1000	dominicks	10239	1	2.39	24471.21
2	1990-06-28	1000	dominicks	17917	1	2.48	44434.16
3	1990-07-05	1000	dominicks	14218	1	2.33	33127.94
4	1990-07-12	1000	dominicks	15925	1	2.01	32009.25
5	1990-07-19	1000	dominicks	17850	1	2.17	38734.50
6	1990-07-26	1000	dominicks	10576	1	1.97	20834.72
7	1990-08-02	1000	dominicks	9912	1	2.26	22401.12
8	1990-08-09	1000	dominicks	9571	1	2.11	20194.81
9	1990-08-16	1000	dominicks	15748	1	2.42	38110.16

```
****************************************************************************************************
DATA GUARDRAILS:

TYPE:         Frequency detection
STATUS:       PASSED
DESCRIPTION:  The time series was analyzed, all data points are aligned with detected frequency.


****************************************************************************************************

TYPE:         Missing feature values imputation
STATUS:       PASSED
DESCRIPTION:  No feature missing values were detected in the training data.
              Learn more about missing value imputation: https://aka.ms/AutomatedMLFeaturization

****************************************************************************************************

TYPE:         Short series handling
STATUS:       PASSED
DESCRIPTION:  Automated ML detected enough data points for each series in the input data to continue with training.
```

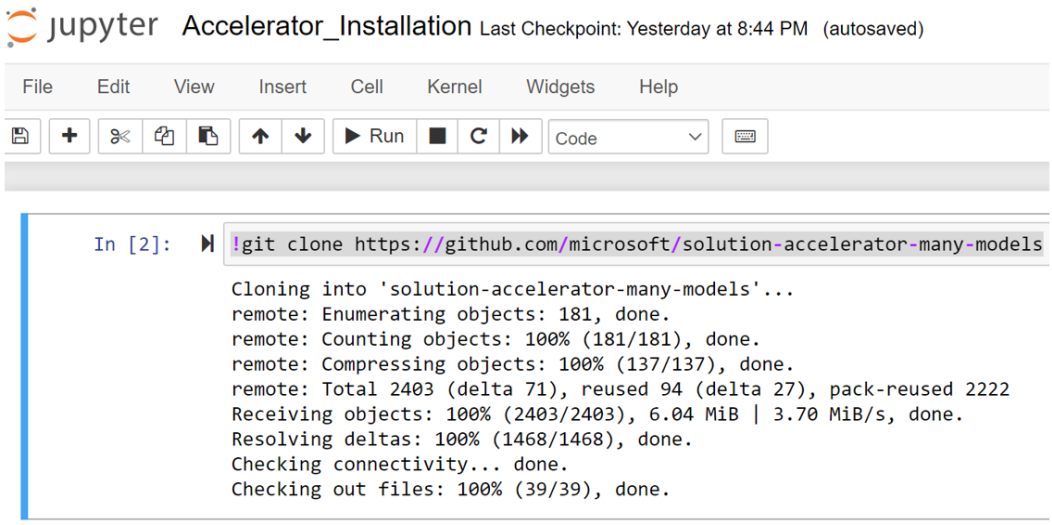
```
********************************************************************************
ITERATION: The iteration being evaluated.
PIPELINE: A summary description of the pipeline being evaluated.
DURATION: Time taken for the current iteration.
METRIC: The result of computing score on the fitted pipeline.
BEST: The best observed score thus far.
********************************************************************************
```

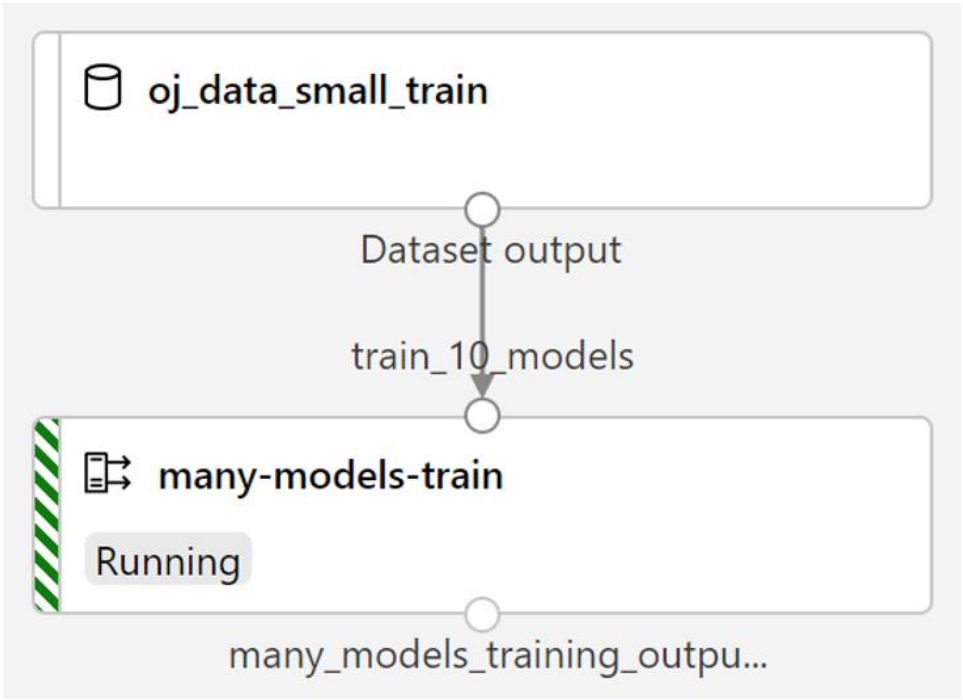
ITERATION	PIPELINE	DURATION	METRIC	BEST
0	MaxAbsScaler DecisionTree	0:00:33	0.0608	0.0608
1	MinMaxScaler DecisionTree	0:00:27	0.0761	0.0608
2	StandardScalerWrapper LassoLars	0:00:33	0.0244	0.0244
3	RobustScaler ElasticNet	0:00:31	0.1121	0.0244
4	StandardScalerWrapper LassoLars	0:00:33	0.0244	0.0244
5	RobustScaler ElasticNet	0:00:29	0.0241	0.0241
6	RobustScaler DecisionTree	0:00:33	0.0648	0.0241
7	MinMaxScaler DecisionTree	0:00:31	0.0525	0.0241
8	RobustScaler DecisionTree	0:00:36	0.0890	0.0241
9	MinMaxScaler DecisionTree	0:00:31	0.0847	0.0241
10	StandardScalerWrapper DecisionTree	0:00:34	0.0621	0.0241
11	StandardScalerWrapper DecisionTree	0:00:28	0.0395	0.0241
12	RobustScaler DecisionTree	0:00:34	0.0410	0.0241
13	RobustScaler ElasticNet	0:00:31	0.0762	0.0241
14	StandardScalerWrapper DecisionTree	0:00:37	0.0563	0.0241
15	MinMaxScaler DecisionTree	0:00:40	0.0458	0.0241
16	MinMaxScaler DecisionTree	0:00:28	0.0586	0.0241
17		0:00:29	nan	0.0241
18	VotingEnsemble	0:00:41	0.0164	0.0164

ITERATION	PIPELINE	DURATION	METRIC	BEST
0	AutoArima	0:01:31	0.2910	0.2910
1	ProphetModel	0:00:58	0.0352	0.0352
2	MaxAbsScaler DecisionTree	0:00:31	0.0466	0.0352
3	MinMaxScaler DecisionTree	0:00:35	0.0702	0.0352
4	StandardScalerWrapper LassoLars	0:00:41	0.0244	0.0244
5	RobustScaler ElasticNet	0:00:33	0.1121	0.0244
6	StandardScalerWrapper LassoLars	0:00:33	0.0244	0.0244
7	RobustScaler ElasticNet	0:00:32	0.0241	0.0241
8	RobustScaler DecisionTree	0:00:28	0.0603	0.0241
9	MinMaxScaler DecisionTree	0:00:44	0.0415	0.0241
10	RobustScaler DecisionTree	0:00:34	0.0717	0.0241
11	MinMaxScaler DecisionTree	0:00:30	0.0716	0.0241
12	StandardScalerWrapper DecisionTree	0:00:42	0.0486	0.0241
13	StandardScalerWrapper DecisionTree	0:00:38	0.0590	0.0241
14	VotingEnsemble	0:00:49	0.0174	0.0174

Standard regression algorithms	**Tree algorithms**	**Gradient boosting algorithms**	**Other algorithms shared with regression**	**Forecasting-specific algorithms**
Elastic Net LARS Lasso	Decision Tree Random Forest Extremely Randomized Trees	XGBoost LightGBM Gradient Boosting	Stochastic Gradient Descent (SGD) KNN	Prophet ARIMA ForecastTCN

Chapter 7: Using the Many Models Solution Accelerator



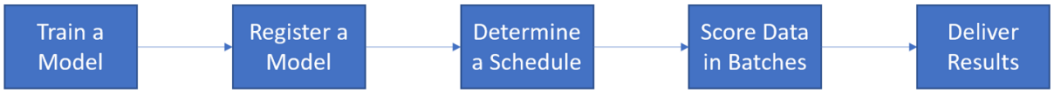


	Week Starting	Store	Brand	Quantity	Advert	Price	Revenue	Predicted
0	1992-05-28	1001	dominicks	12736	1	2.37	30184.32	12655.926953
1	1992-06-04	1001	dominicks	17512	1	2.28	39927.36	17589.279167
2	1992-06-11	1001	dominicks	17769	1	2.18	38736.42	17616.802905
3	1992-06-18	1001	dominicks	17825	1	2.64	47058.00	18321.770558
4	1992-06-25	1001	dominicks	11105	1	2.02	22432.10	11428.825570
5	1992-07-02	1001	dominicks	15763	1	1.97	31053.11	15358.724956
6	1992-07-09	1001	dominicks	10904	1	2.33	25406.32	10827.578098
7	1992-07-16	1001	dominicks	16153	1	2.27	36667.31	15773.192383
8	1992-07-23	1001	dominicks	14260	1	2.03	28947.80	14397.971780
9	1992-07-30	1001	dominicks	9284	1	2.67	24788.28	8715.212884

Prediction has 124 rows. Here the first 10.

	Date	Store	Sales	Predicted
0	2021-03-01	San Francisco	83	54.902630
1	2021-03-02	San Francisco	37	49.947213
2	2021-03-03	San Francisco	44	49.690911
3	2021-03-04	San Francisco	50	54.322844
4	2021-03-05	San Francisco	73	53.245385
5	2021-03-06	San Francisco	91	50.389134
6	2021-03-07	San Francisco	55	53.390174
7	2021-03-08	San Francisco	9	50.627610
8	2021-03-09	San Francisco	98	48.155031
9	2021-03-10	San Francisco	8	49.585355

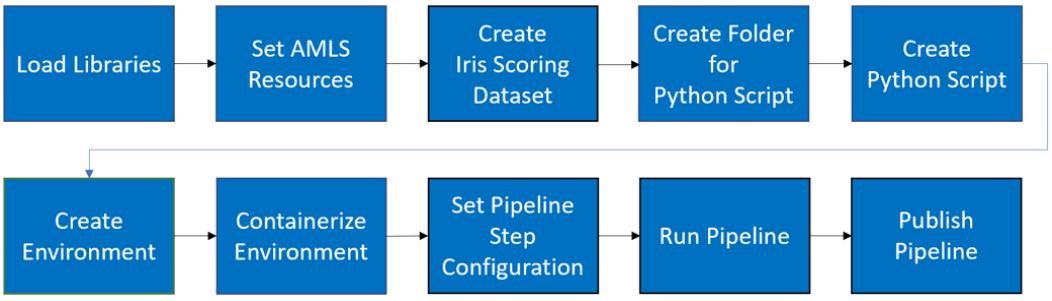
Chapter 8: Choosing Real-Time versus Batch Scoring

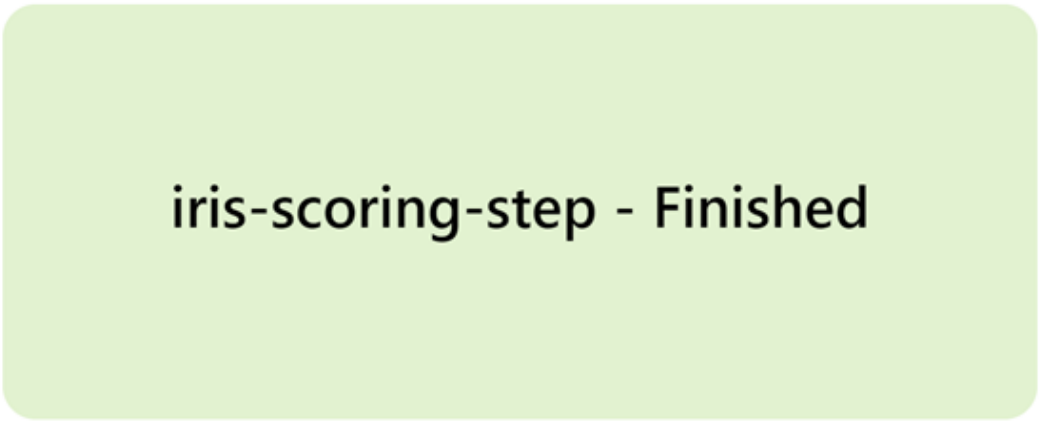


	Advantages	**Disadvantages**	**When to use**
Batch scoring	Cheap Easy to replicate Simple	Slow	When end users can wait for predictions with new data
Real-time scoring	Fast	Costly Complex Unique	When end users require predictions immediately

Scenario	**Type of solution**
Fraud detection	Real-time
Predictive maintenance	Dependent on when machines are serviced
Once a day, week, or month scoring	Batch
Recommendation engines	Real-time
Forecasting for the long term	Batch
Forecasting minute by minute	Real-time
Web-based applications	Dependent on user expectations

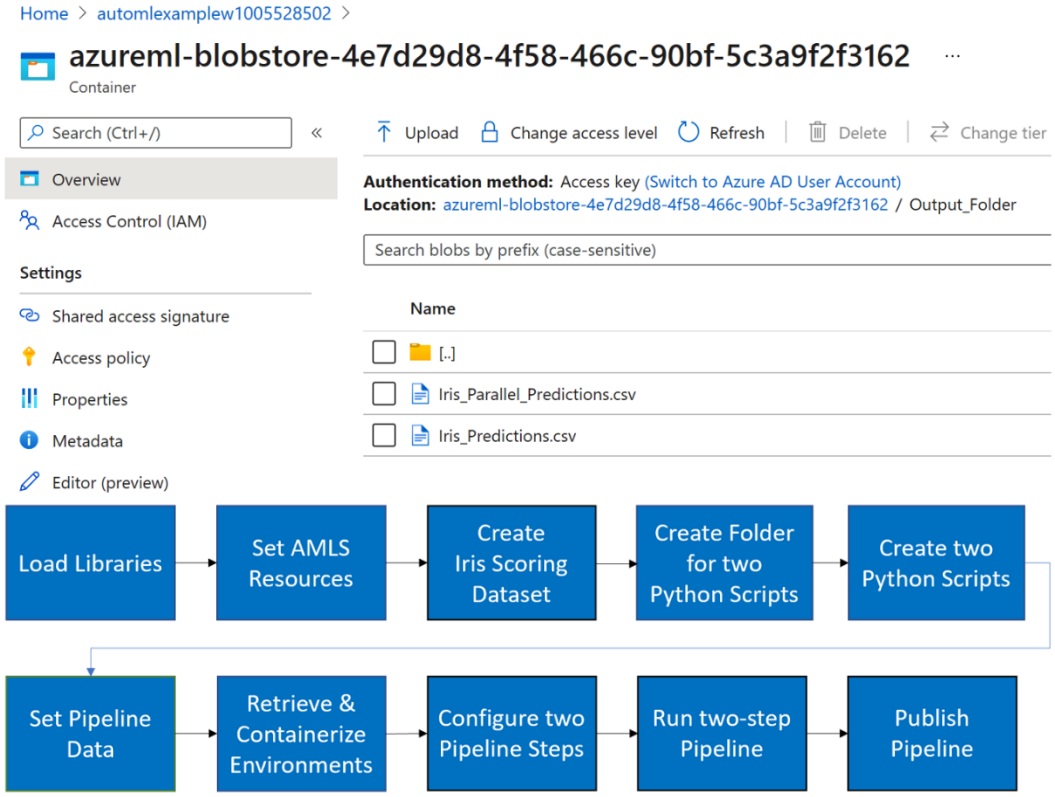
Chapter 9: Implementing a Batch Scoring Solution



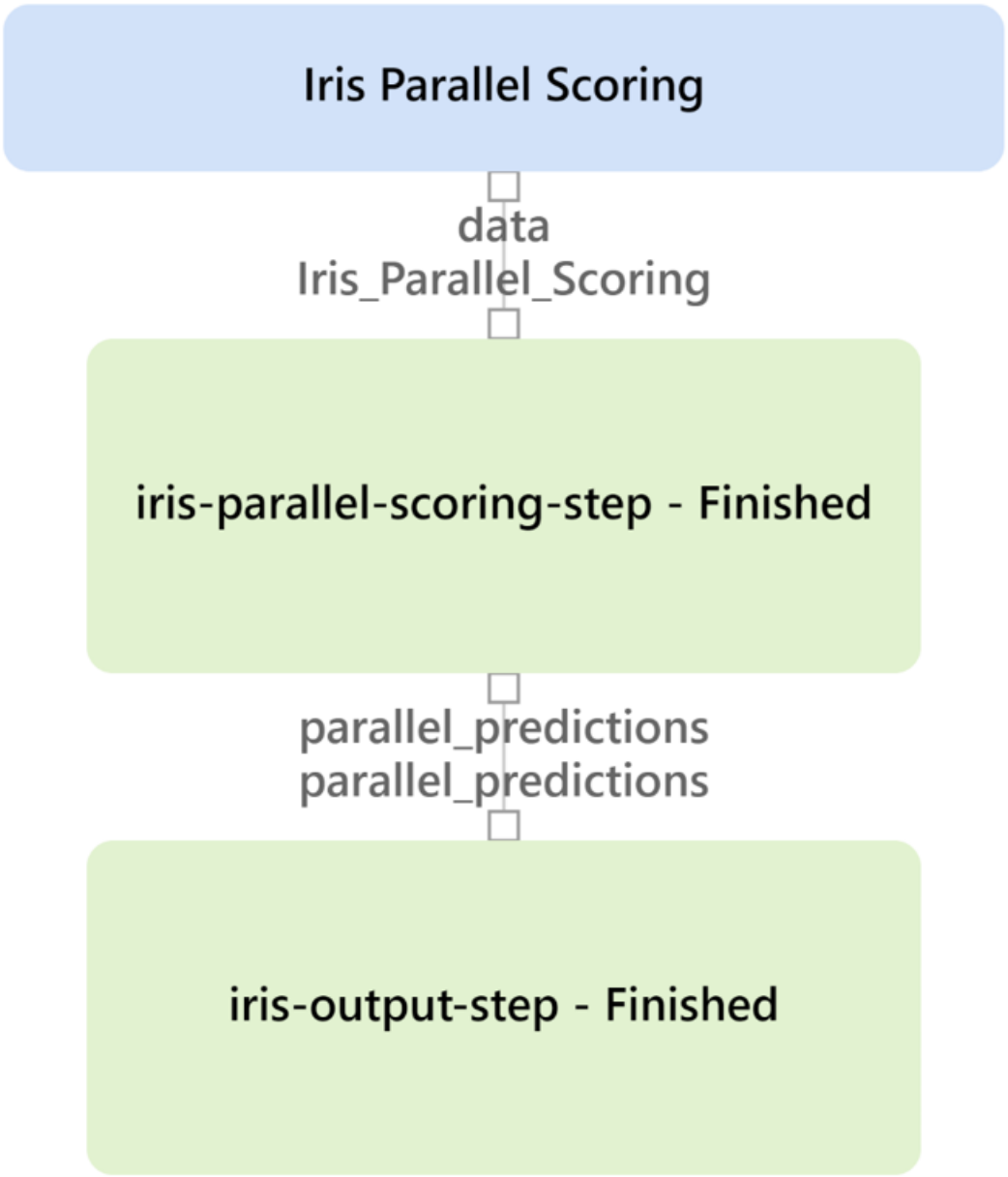



```
published_pipeline = pipeline_run.publish_pipeline(
    name='Iris-Scoring-Pipeline',\
    description='Pipeline that Scores Iris Data', version= '1.0')

published_pipeline
```

Name	Id	Status	Endpoint
Iris-Scoring-Pipeline	466fc3a2-cfcf-4a31-8fa5-21762e3c4b6e	Active	REST Endpoint

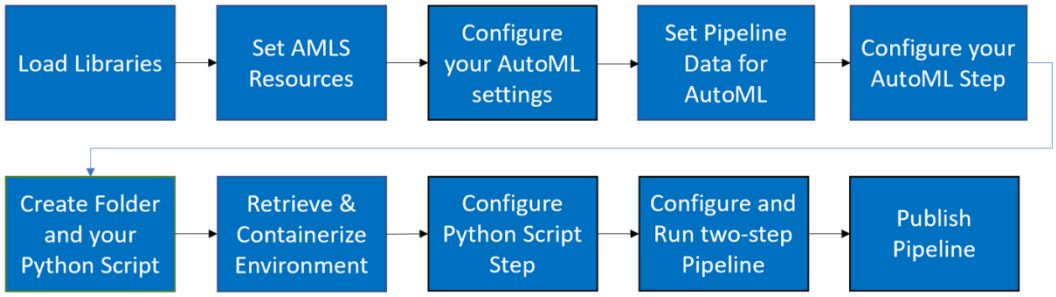

Home > automlexamplew1005528502 >
azureml-blobstore-4e7d29d8-4f58-466c-90bf-5c3a9f2f3162
Container
Search (Ctrl+/)
Overview
Access Control (IAM)
Settings
Shared access signature
Access policy
Properties
Metadata
Editor (preview)
Upload
Change access level
Refresh
Delete
Change tier
Authentication method: Access key (Switch to Azure AD User Account)
Location: azureml-blobstore-4e7d29d8-4f58-466c-90bf-5c3a9f2f3162 / Output_Folder
Search blobs by prefix (case-sensitive)
Name
[..]
Iris_Parallel_Predictions.csv
Iris_Predictions.csv

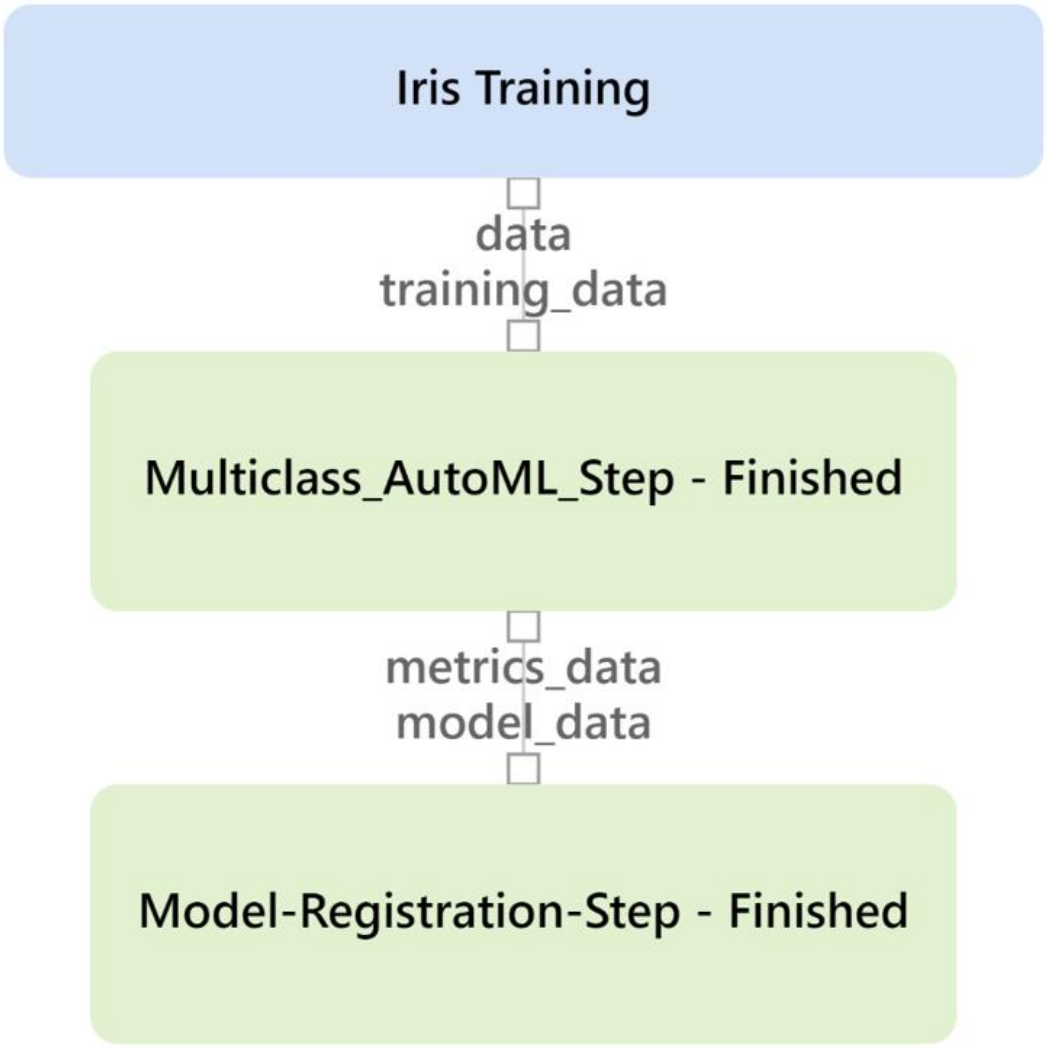
Load Libraries
Set AMLS Resources
Create Iris Scoring Dataset
Create Folder for two Python Scripts
Create two Python Scripts
Set Pipeline Data
Retrieve & Containerize Environments
Configure two Pipeline Steps
Run two-step Pipeline
Publish Pipeline


Iris Parallel Scoring
data
Iris_Parallel_Scoring
iris-parallel-scoring-step - Finished
parallel_predictions
parallel_predictions
iris-output-step - Finished

```
published_pipeline = pipeline_run.publish_pipeline(
    name='Iris-Parallel-Scoring-Pipeline',\
    description='Pipeline that Scores Iris Data in Parallel', version= '1.0')

published_pipeline
```

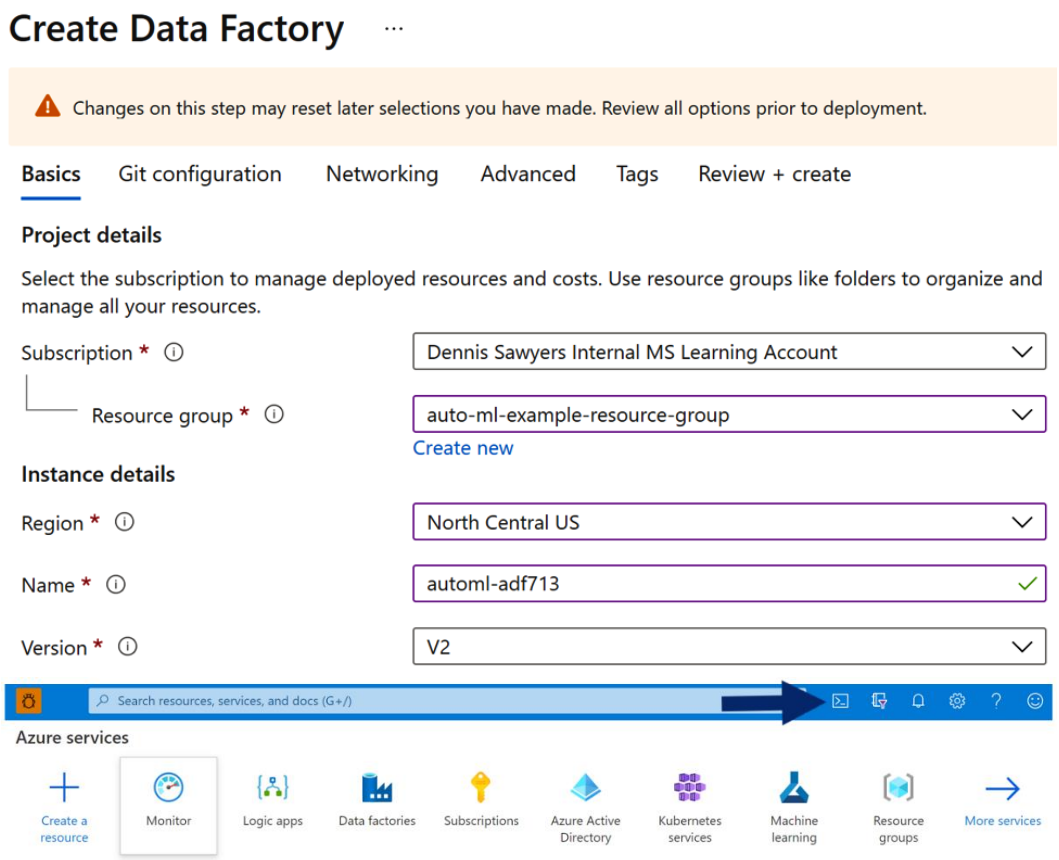
Name	Id	Status	Endpoint
Iris-Parallel-Scoring-Pipeline	c58a16ea-5abc-4d42-b25c-21fa9967895f	Active	REST Endpoint

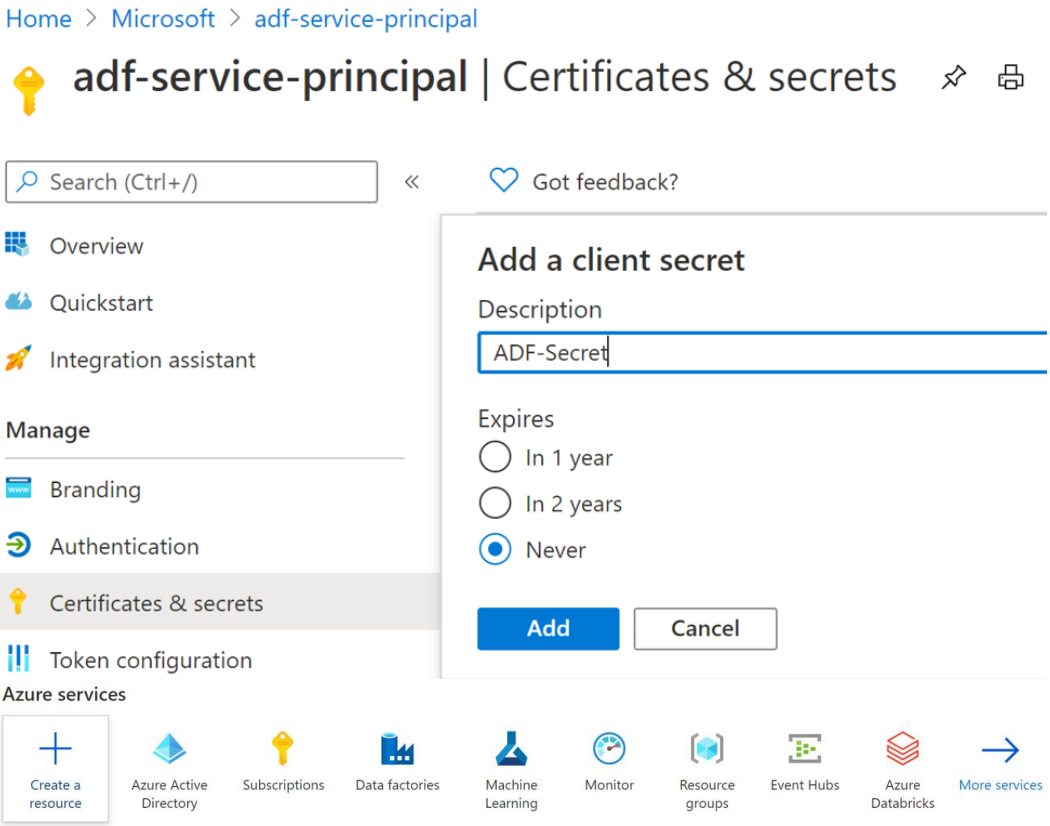




Name	Id	Status	Endpoint
Iris-AutoML-Training-Pipeline	bdafc5af-4964-40ac-bb01-50dd49b6dcb5	Active	REST Endpoint

Chapter 10: Creating End-to-End AutoML Solutions




Home > Microsoft > adf-service-principal
adf-service-principal | Certificates & secrets
Search (Ctrl+/)
Got feedback?
Overview
Quickstart
Integration assistant
Manage
Branding
Authentication
Certificates & secrets
Token configuration
Add a client secret
Description
ADF-Secret
Expires
In 1 year
In 2 years
Never
Add
Cancel
Azure services
Create a resource
Azure Active Directory
Subscriptions
Data factories
Machine Learning
Monitor
Resource groups
Event Hubs
Azure Databricks
More services

Add role assignment

Role

Contributor

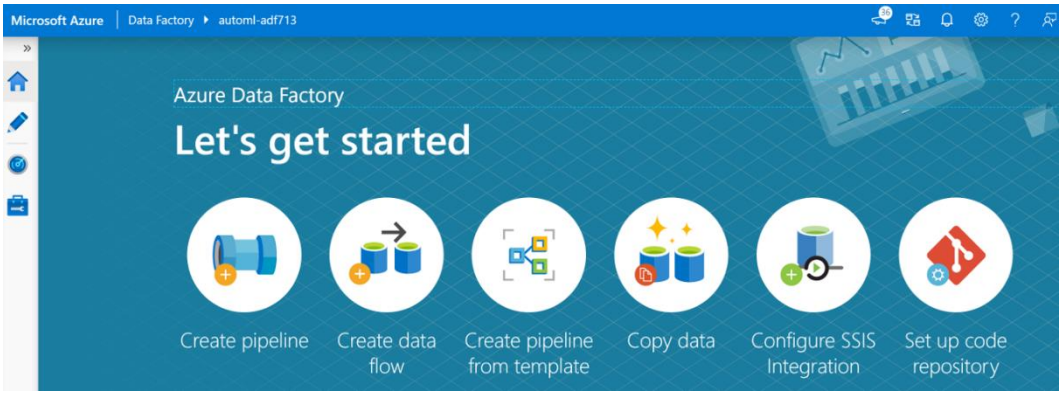
Assign access to

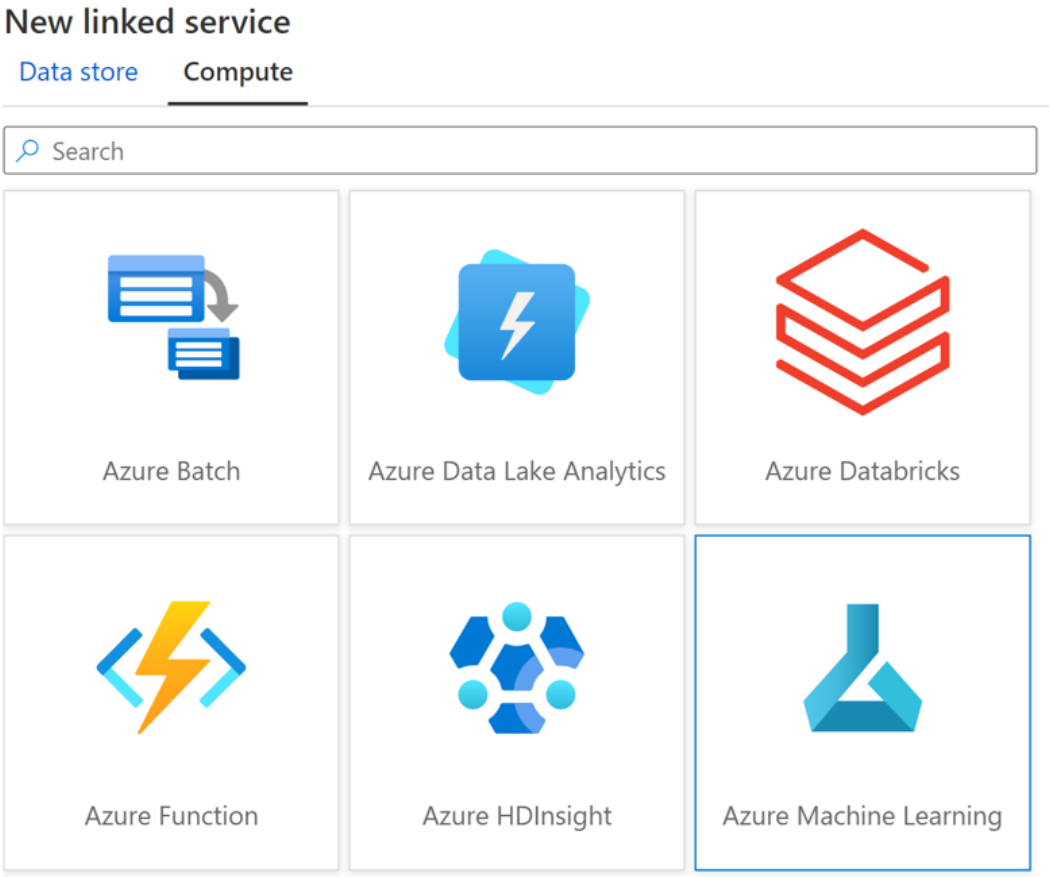
User, group, or service principal

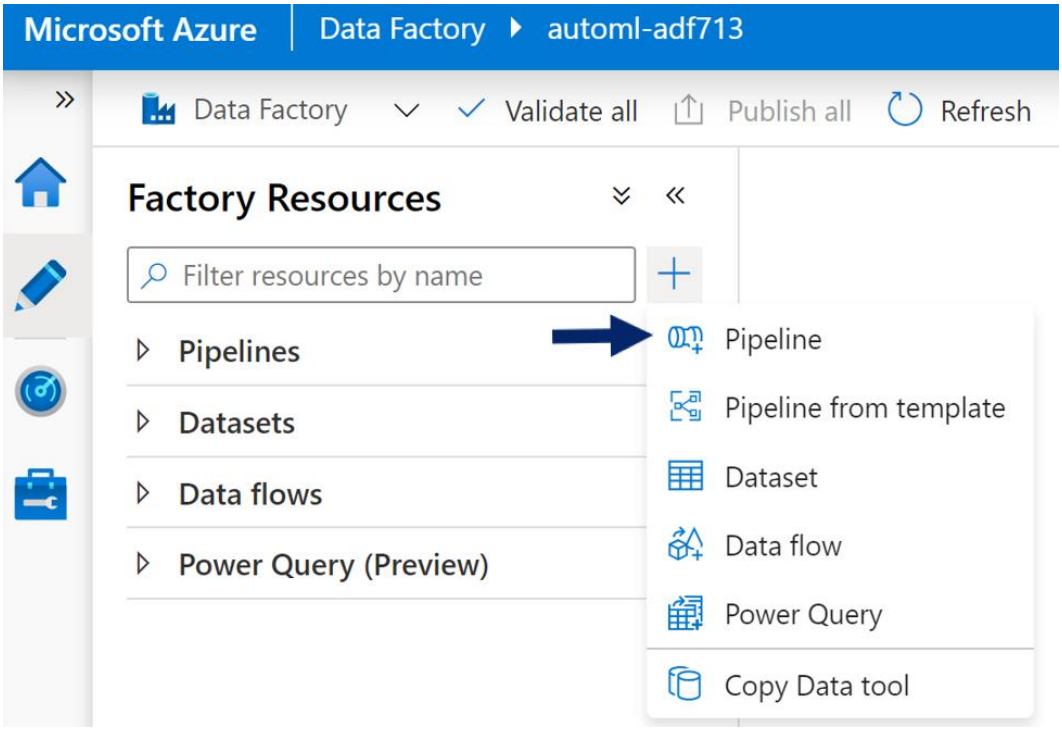
Select

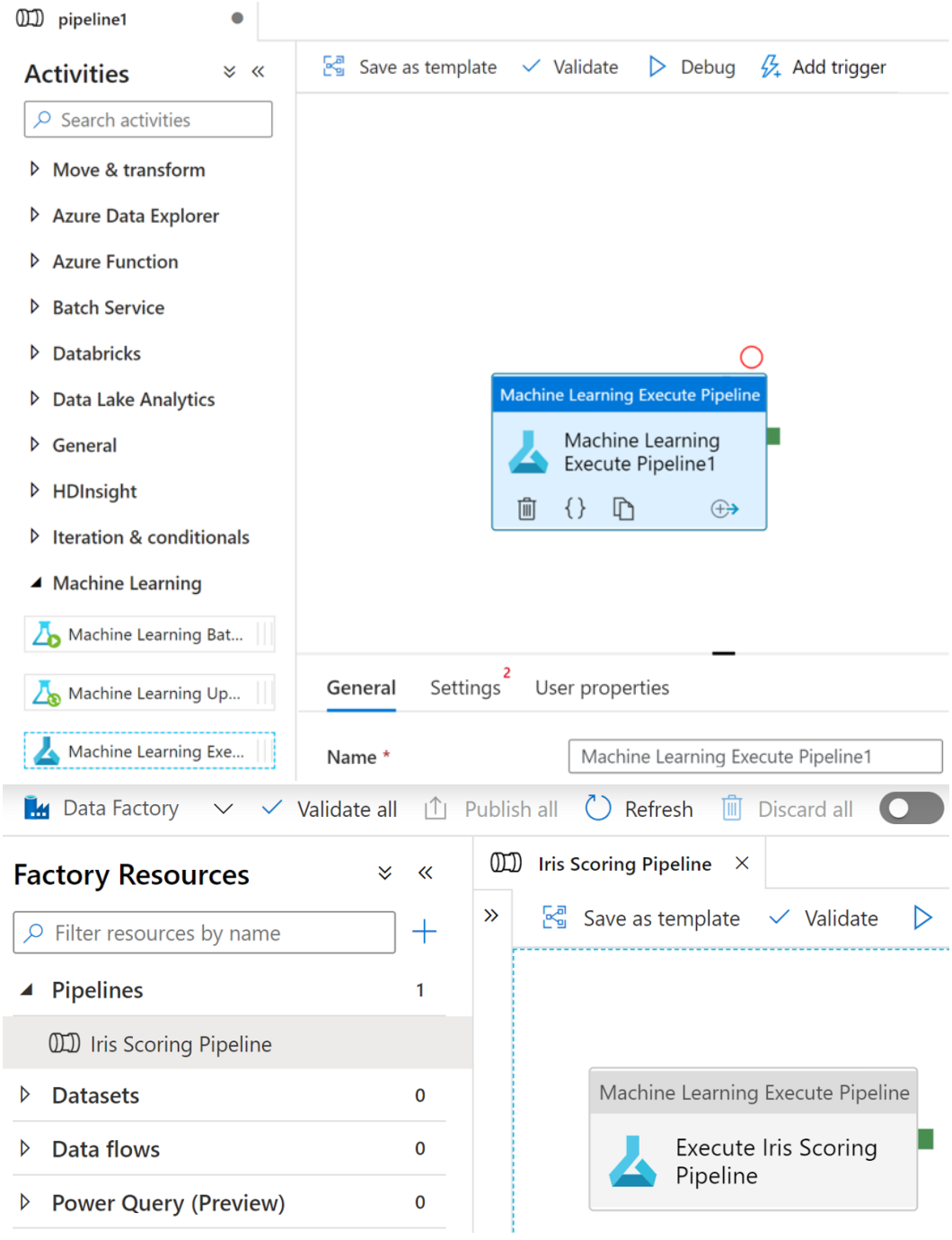
adf-service-principal

adf-service-principal




New linked service
Data store
Compute
Search
Azure Batch
Azure Data Lake Analytics
Azure Databricks
Azure Function
Azure HDInsight
Azure Machine Learning


Microsoft Azure
Data Factory
automl-adf713
Data Factory
Validate all
Publish all
Refresh
Factory Resources
Filter resources by name
Pipelines
Datasets
Data flows
Power Query (Preview)
Pipeline
Pipeline from template
Dataset
Data flow
Power Query
Copy Data tool


pipeline1
Activities
Search activities
Move & transform
Azure Data Explorer
Azure Function
Batch Service
Databricks
Data Lake Analytics
General
HDInsight
Iteration & conditionals
Machine Learning
Machine Learning Bat...
Machine Learning Up...
Machine Learning Exe...
Save as template
Validate
Debug
Add trigger
Machine Learning Execute Pipeline
Machine Learning Execute Pipeline1
General
Settings
2
User properties
Name *
Machine Learning Execute Pipeline1
Data Factory
Validate all
Publish all
Refresh
Discard all
Factory Resources
Filter resources by name
Pipelines
1
Iris Scoring Pipeline
Datasets
0
Data flows
0
Power Query (Preview)
0
Iris Scoring Pipeline
Save as template
Validate
Machine Learning Execute Pipeline
Execute Iris Scoring Pipeline

Integration runtime setup

Integration Runtime is the native compute used to execute or dispatch activities. Choose what integration runtime to create based on required capabilities. Learn more

Azure, Self-Hosted

Perform data flows, data movement and dispatch activities to external compute.

Azure-SSIS

Lift-and-shift existing SSIS packages to execute in Azure.

Integration runtimes

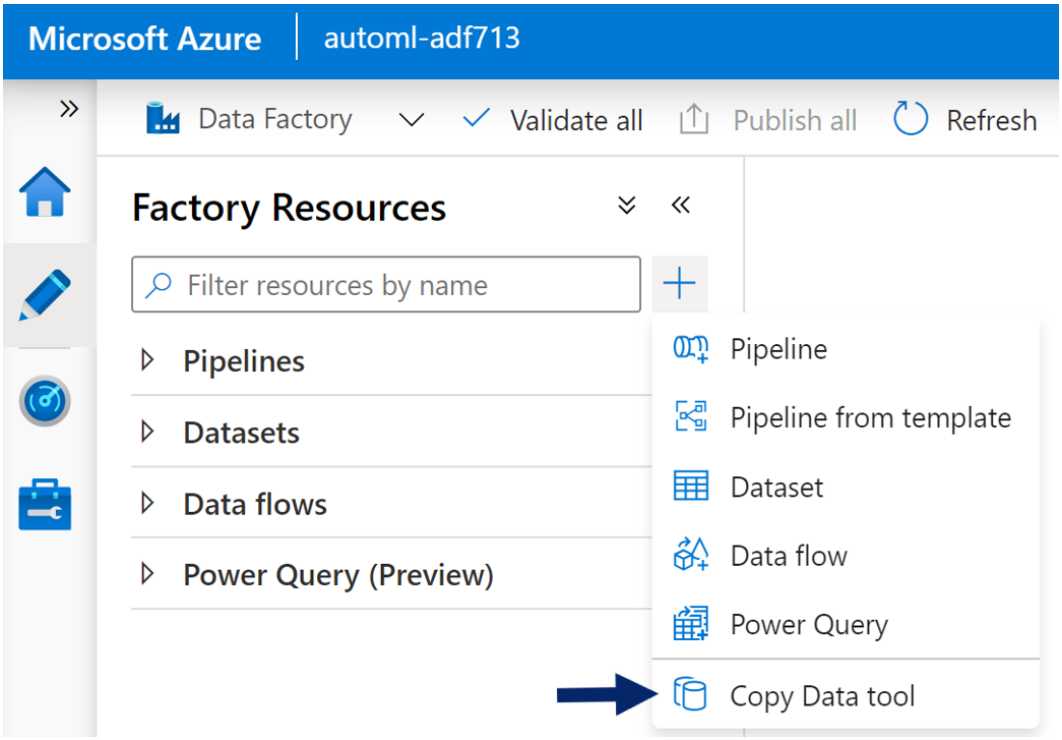
The integration runtime (IR) is the compute infrastructure to provide the following data integration capabilities

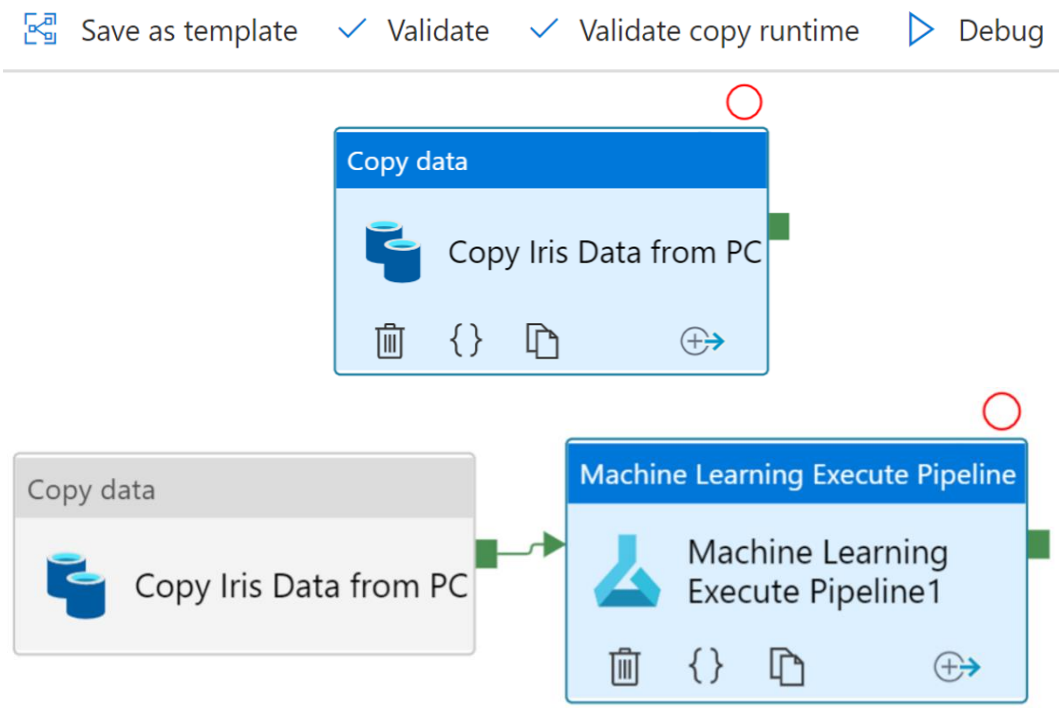
+ New Refresh

Filter by name

Showing 1 - 2 of 2 items

Name	Type	Sub-type	Status
AutoResolveIntegra...	Azure	Public	Running
IntegrationRuntime	Self-Hosted	---	Running


Microsoft Azure
automl-adf713
Data Factory
Validate all
Publish all
Refresh
Factory Resources
Filter resources by name
Pipelines
Datasets
Data flows
Power Query (Preview)
Pipeline
Pipeline from template
Dataset
Data flow
Power Query
Copy Data tool



Set properties

Name

ScoringResults

Linked service *

AMLSDatastoreLink

Connect via integration runtime *

AutoResolveIntegrationRuntime

File path

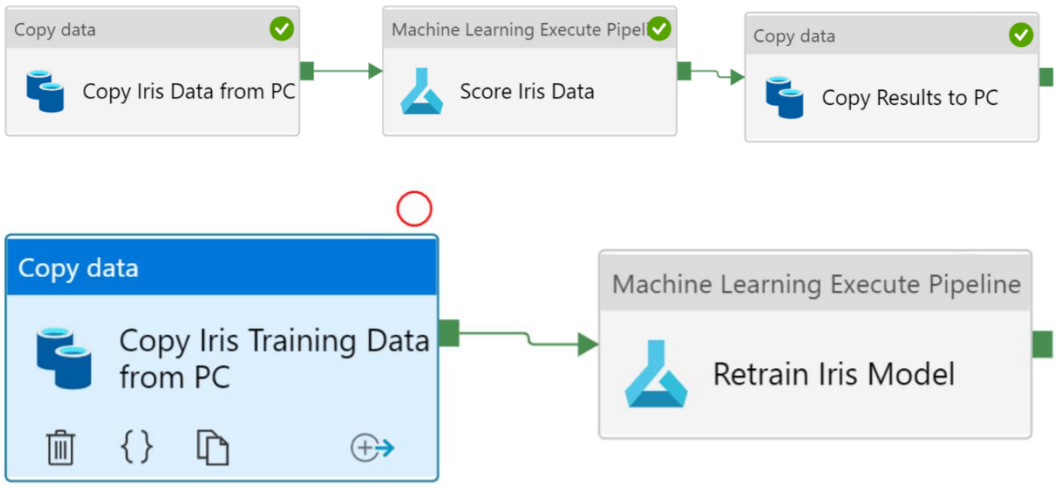
azureml-blobstore-4e7d2 / Output_Folder / Iris_Predictions.csv

First row as header

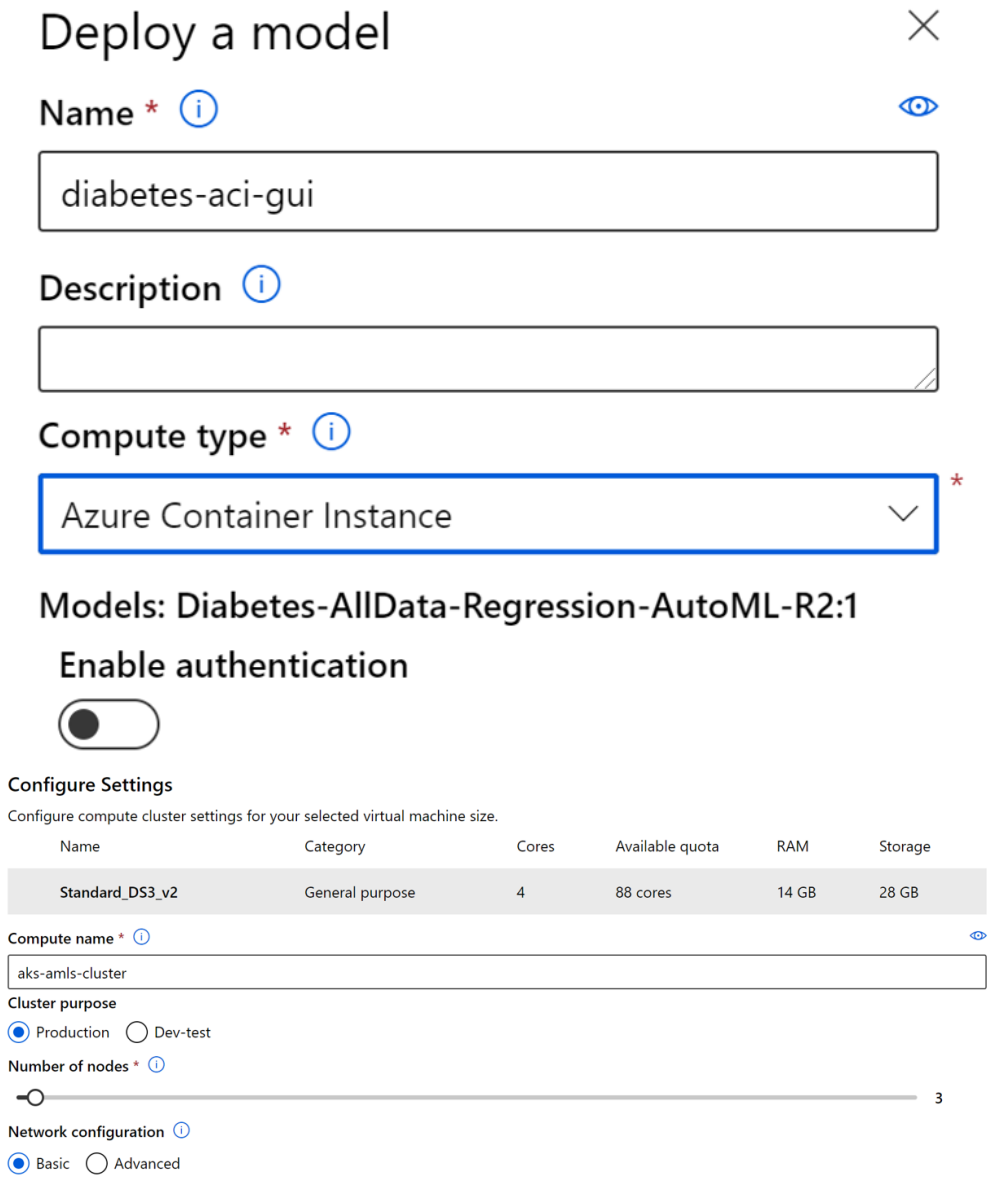
Import schema

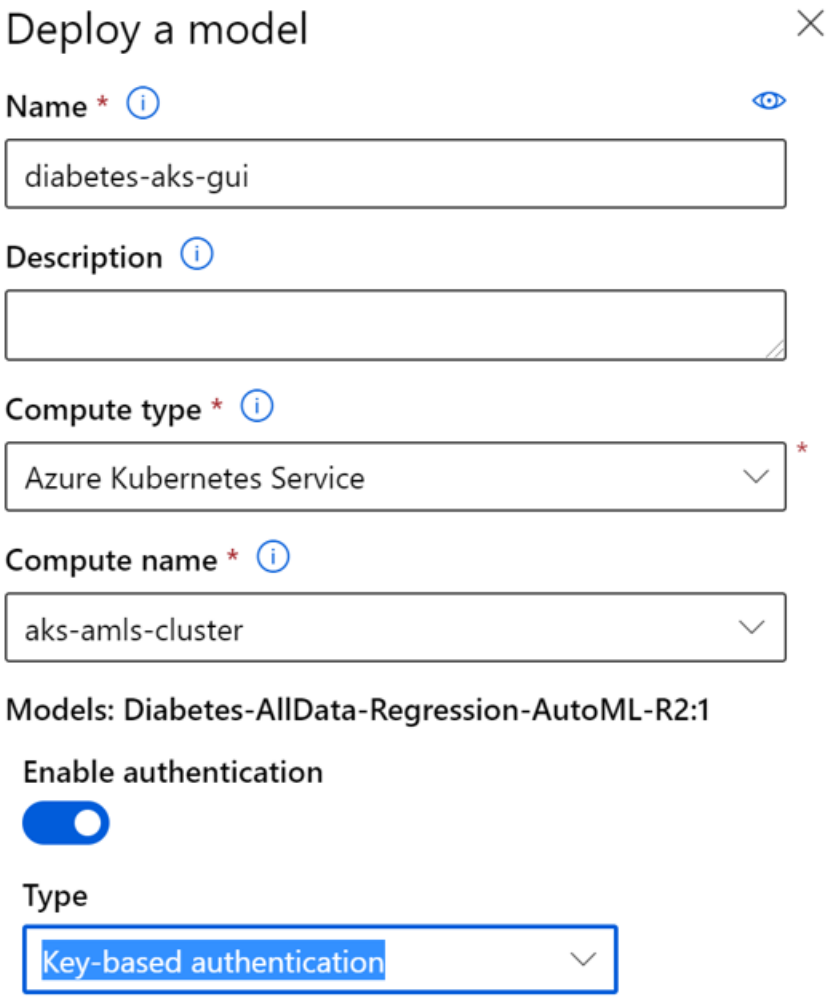
From connection/store From sample file None

Advanced


Copy data
Copy Iris Data from PC
Machine Learning Execute Pipel
Score Iris Data
Copy data
Copy Results to PC
Copy data
Copy Iris Training Data from PC
Machine Learning Execute Pipeline
Retrain Iris Model

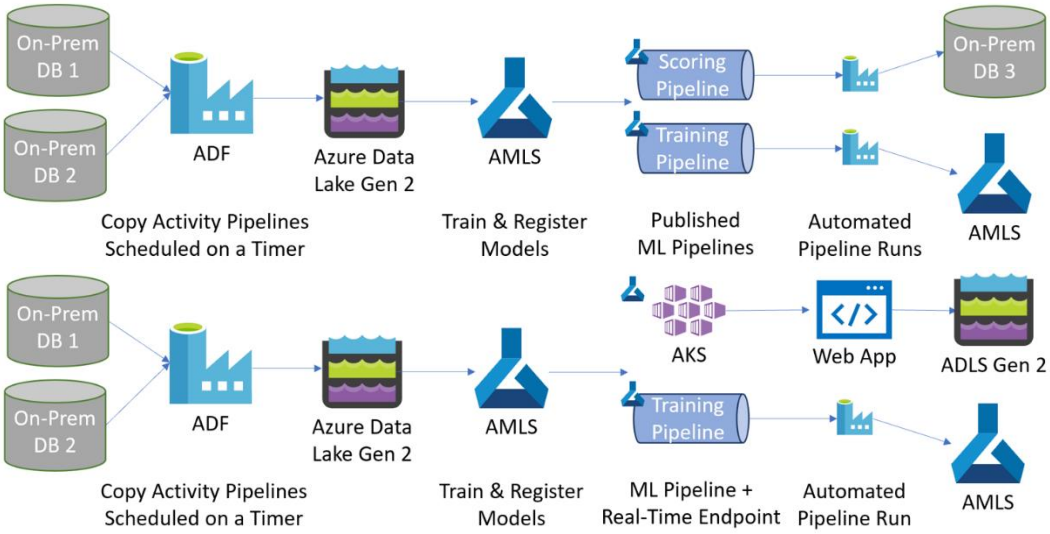
Chapter 11: Implementing a Real-Time Scoring Solution



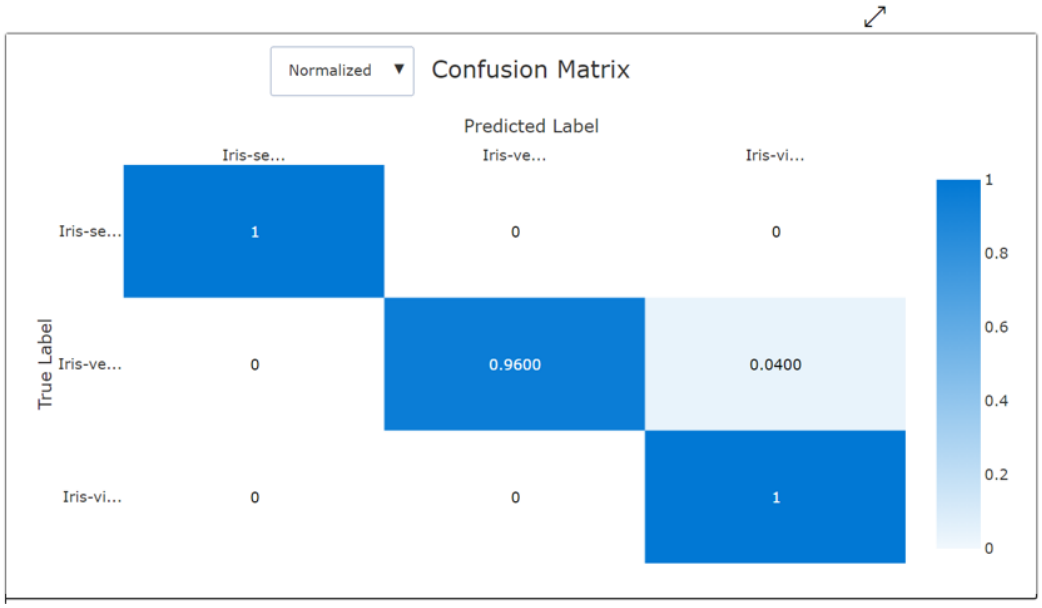


'{"data":[{"AGE":68,"SEX":1,"BMI":41.3,"BP":119,"S1":145,"S2":65.7,"S3":27,"S4":8.51,"S5":4.823,"S6":94},{"AGE":59,"SEX":1,"BMI":32.1,"BP":104,"S1":104,"S2":187.6,"S3":97,"S4":6.87,"S5":3.438,"S6":120},{"AGE":64,"SEX":1,"BMI":30.4,"BP":67,"S1":220,"S2":169.0,"S3":86,"S4":4.79,"S5":5.704,"S6":86},{"AGE":57,"SEX":1,"BMI":32.3,"BP":113,"S1":290,"S2":116.5,"S3":58,"S4":5.35,"S5":3.872,"S6":101},{"AGE":36,"SEX":1,"BMI":26.5,"BP":131,"S1":131,"S2":122.5,"S3":57,"S4":6.9,"S5":5.634,"S6":111}]}'

Chapter 12: Realizing Business Value with AutoML

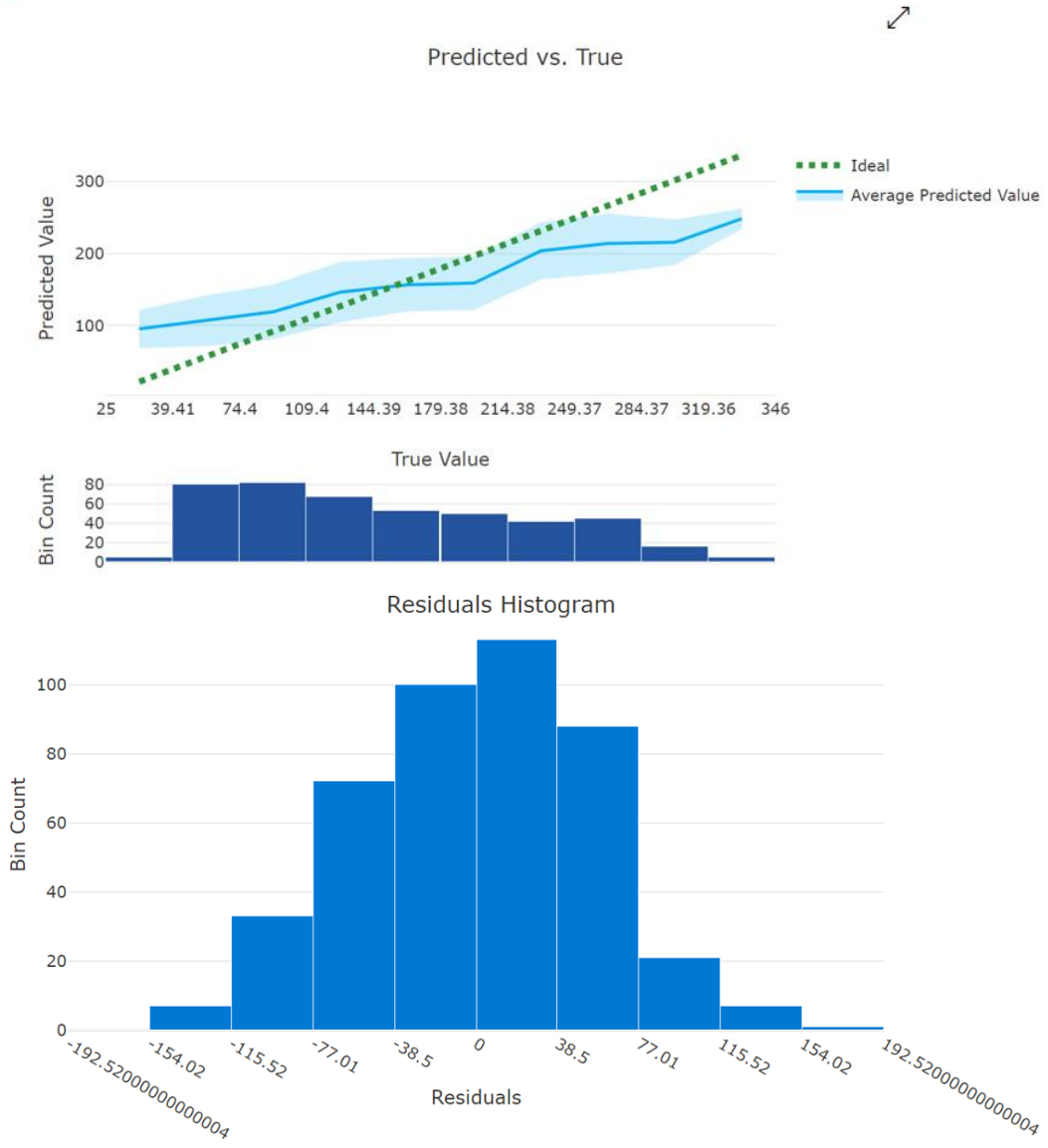


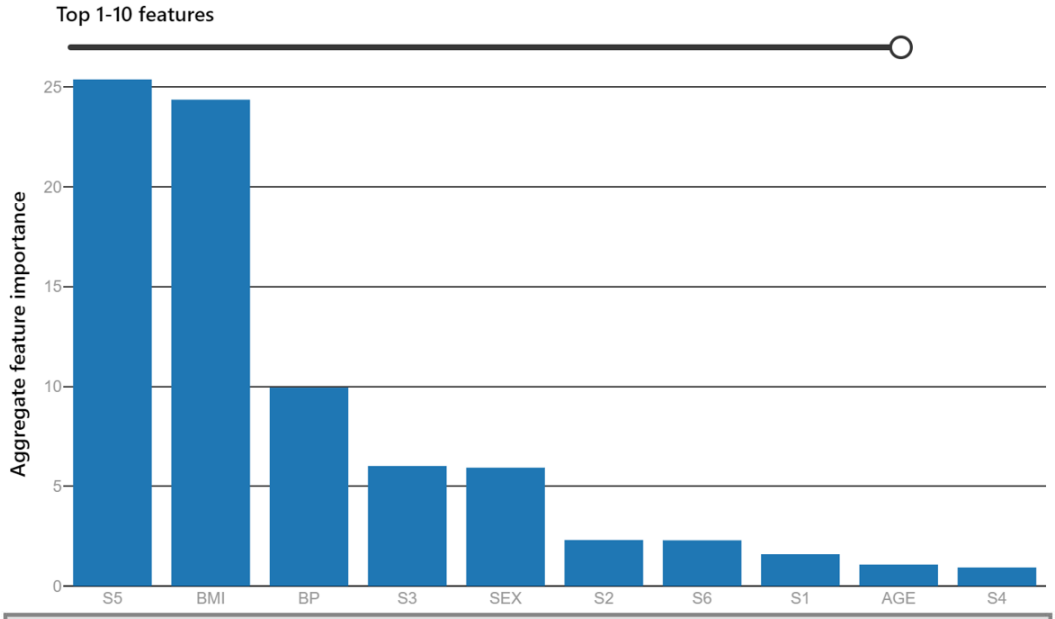
accuracy
0.987



mean_absolute_percentage_e...

38.941





New to machine learning models? Here's what you'll be doing:

1. Create and train your model
2. Improve it
3. Apply it

Select training data

Select your base data and related inputs to train your model.

Choose a model type

We'll help you pick the best model to achieve your business goals.

Train your model

The model will train on your data and report on its performance.

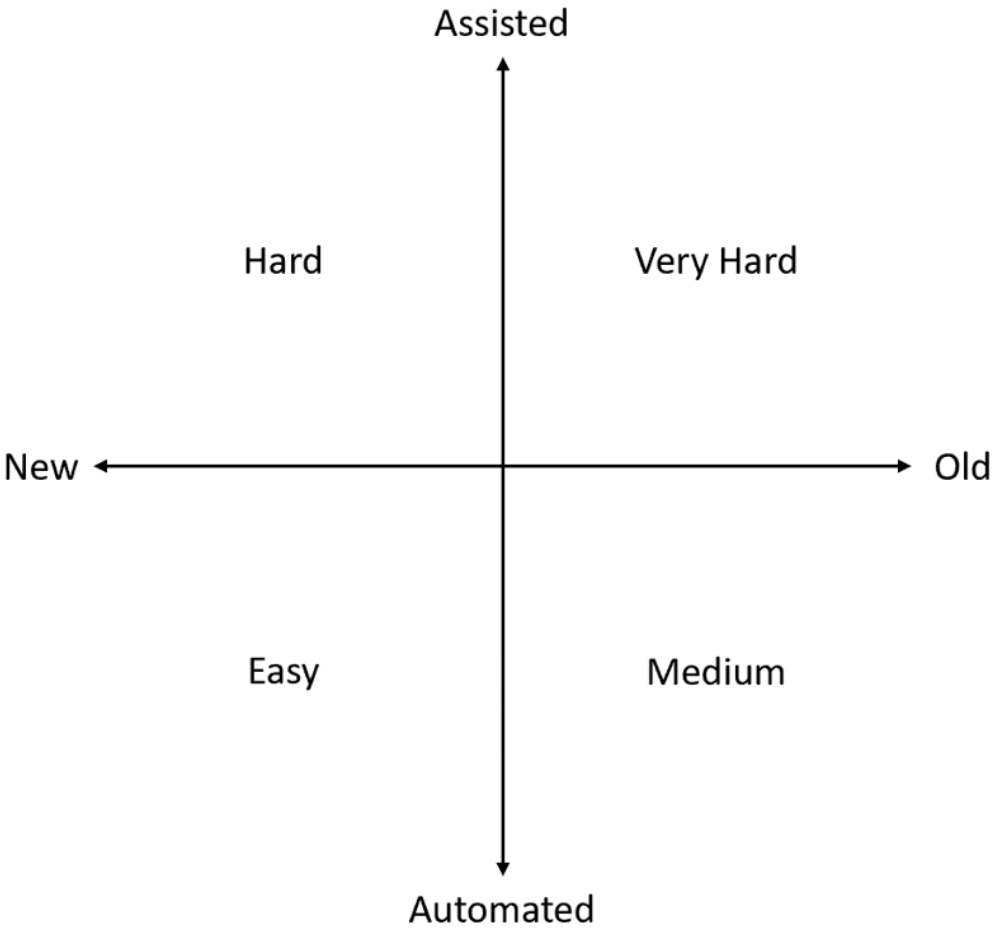
Iterate and retrain

Evaluate, customize and retrain your model until it's optimized

Apply the model

Apply your model to future data for predictive insights.

Get started


Assisted
Hard
Very Hard
New
Old
Easy
Medium
Automated

Key questions to ask	Common answers to key questions
Where is my input data coming from?	On-premise database, cloud database
How often should my solution score?	On certain days at certain times
How do end users access results?	Report generated by an on-premise database
When should I retrain my model?	Once a month as new data becomes available
How should I orchestrate my solution?	Azure Data Factory

Key questions to ask	Common answers to key questions
Where is my input data coming from?	On-premise database, cloud database
How often should I retrain my model?	Whenever model performance degrades
Where is my endpoint located?	A user-facing web app, an Azure function
How many requests are coming in at once?	A maximum of 100 requests per minute
How fast do end users expect a response?	Within 10 seconds of submitting a request
How large should my AKS cluster be?	Large VM size for faster performance

	Difficulty of gaining trust	Who is most likely to offer resistance	How to overcome resistance
Completely new automated process	Easy as the process is being created with ML in mind	Resistance is unlikely if your ML model performs as expected	Build an ML model that exceeds expectations and works well
Replacing an old, automated process	Medium as you are replacing an older process	People who built or oversee the output of the older process	Compare results of the old and new process side-by-side
Completely new assisted decision making process	Hard as many people may not trust AI generated output	People who feel their decision making is threatened by AI	Stress that AI exists as an advisor to enhance their own decisions
Replacing an old, assisted decision making tool	Extremely difficult as users will fondly recall the old tool	People who cannot understand the tech behind the new tool	Emphasize the explainability of your AutoML-built solution