



Azure ML Guide

Part 1 – Data ingestion and Designer

October 2020

Hubert Duan, Cloud Solution Architect

Architecture

Modeling and management



Azure ML
services

Designer

Automated ML

Big data development



Azure
Databricks

Feature engineering

Spark

Data



Azure Data Lake Gen 2

Getting started

Azure provisioning

This workshop will assume services are organized within a resource group, so go ahead and create a new RG along with these services:

- Azure Data Lake Gen 2
 - Hierarchical namespace enabled
- Azure Databricks
 - Premium tier
- Azure Machine Learning
 - Enterprise edition

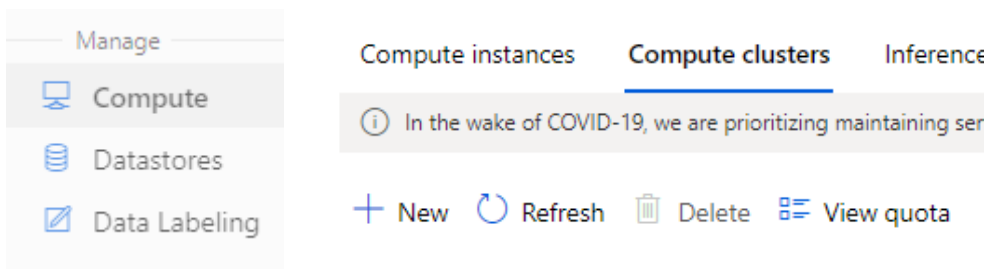
1. Compute

Azure Machine Learning compute

Let's create an Azure ML compute

- Go to Azure ML studio UI (ml.azure.com)

Compute → Computer clusters → Create / New



- Give it a name and virtual machine size. Enter 1 as the maximum number of nodes.

Compute name * ⓘ 👁

hudua *

Region * ⓘ

westus2 *

Virtual machine type *

CPU (Central Processing Unit) *

Virtual machine priority * ⓘ

Dedicated Low priority

Virtual machine size * ⓘ

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

Minimum number of nodes * ⓘ

0 *

Maximum number of nodes * ⓘ

1 *

Idle seconds before scale down * ⓘ

120 *

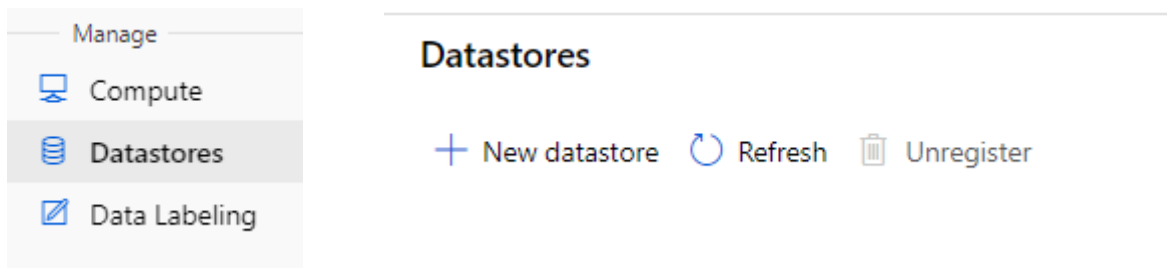
2. Data Ingestion

Azure Machine Learning datastore



Let's create an Azure ML datastore


- Go to Azure ML studio UI (ml.azure.com)

Datastores → New datastore



- Fill in the details and select the storage account name, and container. Name should correspond to container name



Datastore name *  

Datastore type * 

Account selection method
☒ From Azure subscription ☐ Enter manually

Subscription ID *

Storage account *

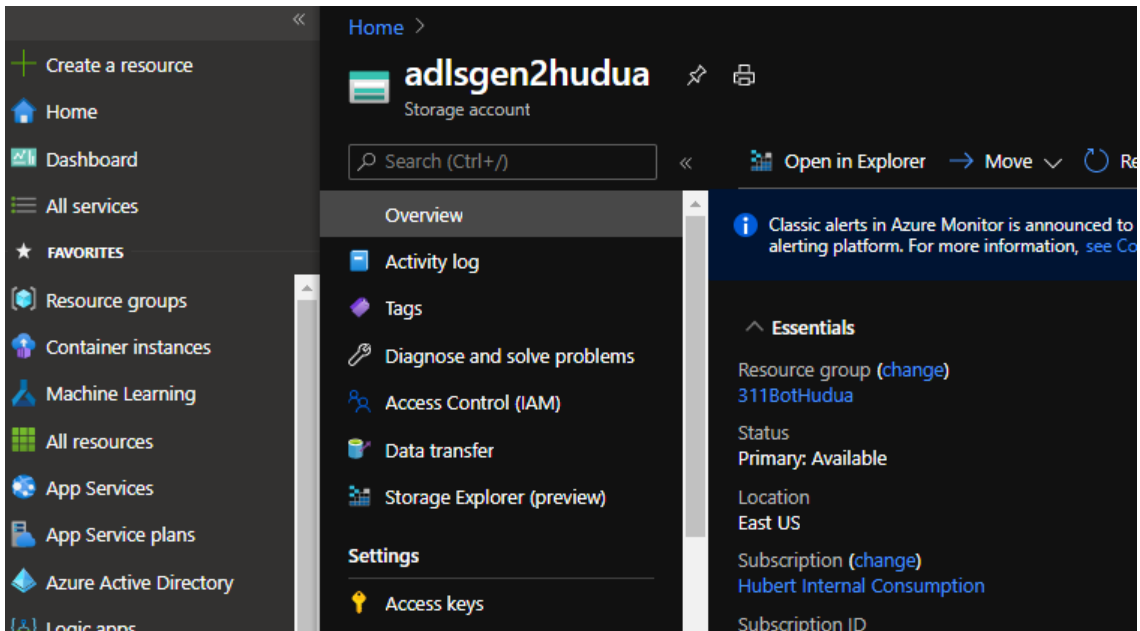
 Note: Attaching storage "adlsgen2hudua" can cause higher latencies and bandwidth co... 

Blob container *

2. Data Ingestion

Azure Machine Learning datastore

- To get the storage account key, in a new tab, go to Azure Portal, find the storage account



- Go to Access keys and copy and paste key1 to the datastore creation menu

A screenshot of the Azure Machine Learning studio 'Datastore' creation form. The form has a title 'key1' and a 'Key' label. The 'Authentication type' dropdown is set to 'Account key'. The 'Account key' input field is empty, and a red arrow points to it from the 'key1' text in the table above. Below the input field is a checkbox labeled 'Use workspace managed identity for data preview and profiling in Azure Machine Learning studio', which is currently unchecked. At the bottom, there are 'No' and 'Yes' buttons. A note at the very bottom states: 'Note: Azure Machine Learning service does not validate whether the underlying data source ...'.

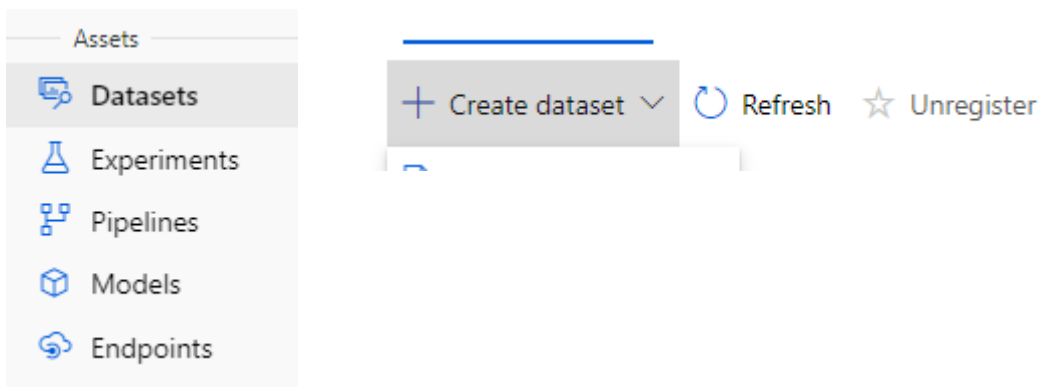
2. Data Ingestion

Azure Machine Learning dataset

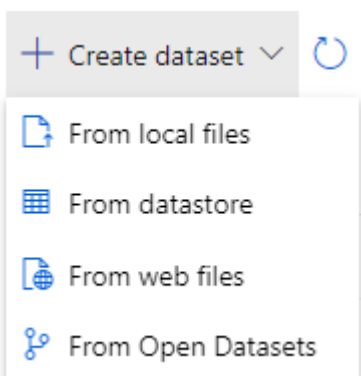
Let's create an Azure ML dataset

- On Azure ML studio UI (ml.azure.com)

Datasets → New datastore



- If you have existing data in the storage account container, you can pick create dataset from datastore, or upload local file



2. Data Ingestion

Azure Machine Learning dataset

- We will select upload local file. Enter a name of dataset and click on next

The screenshot shows the 'Basic info' tab selected in the left sidebar. The main area contains the following fields:

- Name ***: A text input field with the placeholder 'Dataset name'.
- Dataset version**: A text input field with the value '1'.
- Dataset type ***: A dropdown menu with 'Tabular' selected.
- Description**: A text area with the placeholder 'Dataset description'.

- Select Previously created datastore

The screenshot shows the 'Previously created datastore' section. It includes a 'Refresh' button and a search bar labeled 'Search to filter items...'. Below is a table with the following data:

Name	Type	Storage ac...
example	Azure Blob Storage	adlsgen2h...

At the bottom, there are two buttons: 'Select datastore' (blue) and 'Cancel' (white).

- Upload by browse the provided sample.csv

The screenshot shows the 'Browse' button and the file upload progress table. The table has the following data:

File name	Size (MiB)	Upload %	Status
sample.csv	0.004394	100	✓

Below the table is a progress bar. At the bottom, there is an 'Upload path' section with a text input field containing 'UI' and a tooltip that says 'Files will be uploaded to '\$(Upload path)/10-05-2020_030155_UTC''.

2. Data Ingestion

Azure Machine Learning datastore

- If you reach an encoding error, try to check off “Skip data validation

☒ Skip data validation ⓘ

- Click next, and choose “Use headers from the first file”

Settings and preview

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

File format

Delimited

Delimiter

Comma

Example

Field1,Field2,Field3

Encoding

UTF-8

Column headers

Use headers from the first file

Skip rows

None

Id	sepal_length	sepal_width	petal_length	petal_width	class
1	5.1	3.5	1.4	0.2	Iris-setosa
2	4.9	3	1.4	0.2	Iris-setosa
3	4.7	3.2	1.3	0.2	Iris-setosa
4	4.6	3.1	1.5	0.2	Iris-setosa

Back

Next

Cancel

- For schema, you can edit as needed. Click on next to confirm

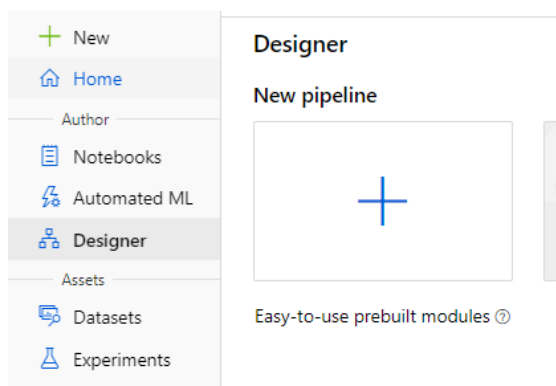
3. Modeling

Azure Machine Learning Designer

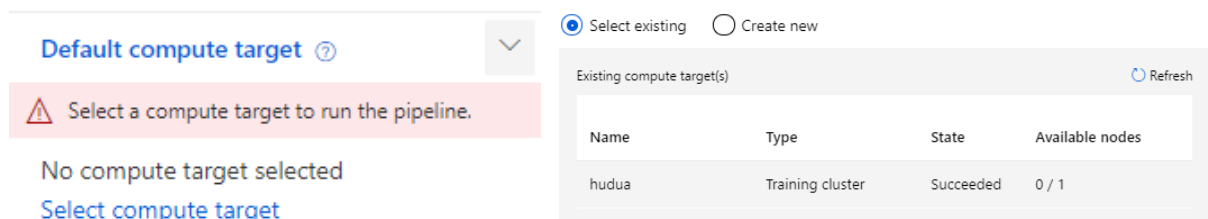
Let's create an Azure ML Designer experiment (drag-and-drop)

- On Azure ML studio UI (ml.azure.com)

Designer → New pipeline



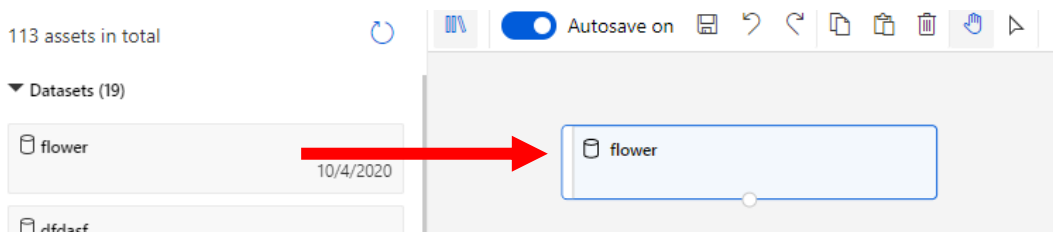
- Select the created compute



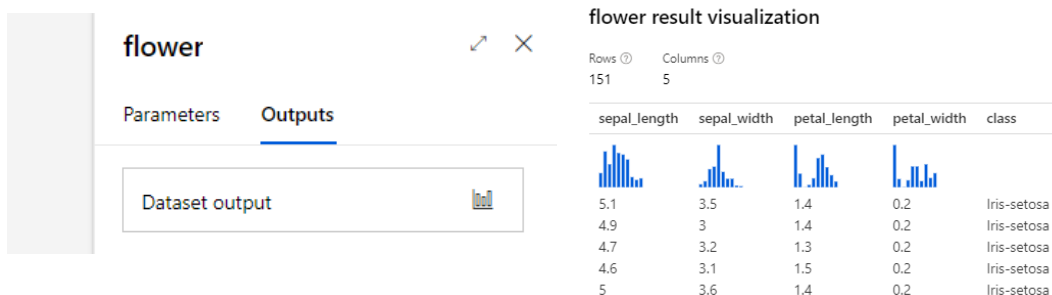
3. Modeling

Azure Machine Learning Designer

- Under dataset, select the newly created dataset, here called “flower” and drag it over



- Click on the flower module, under outputs, you can visualize it



- Now, under Data Transformation, drag over Partition and Sample



3. Modeling

Azure Machine Learning Designer

- Under its options, change fraction of rows in first output to 0.8

Split Data ×

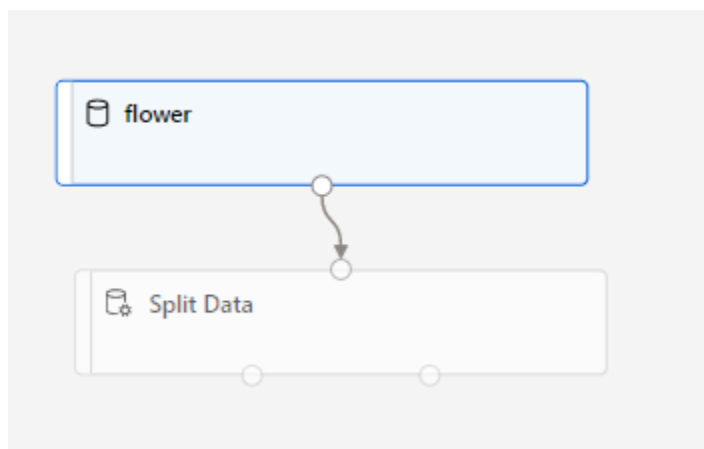
Splitting mode ? *

Split Rows ▼

Fraction of rows in the first output dataset ? *

0.8

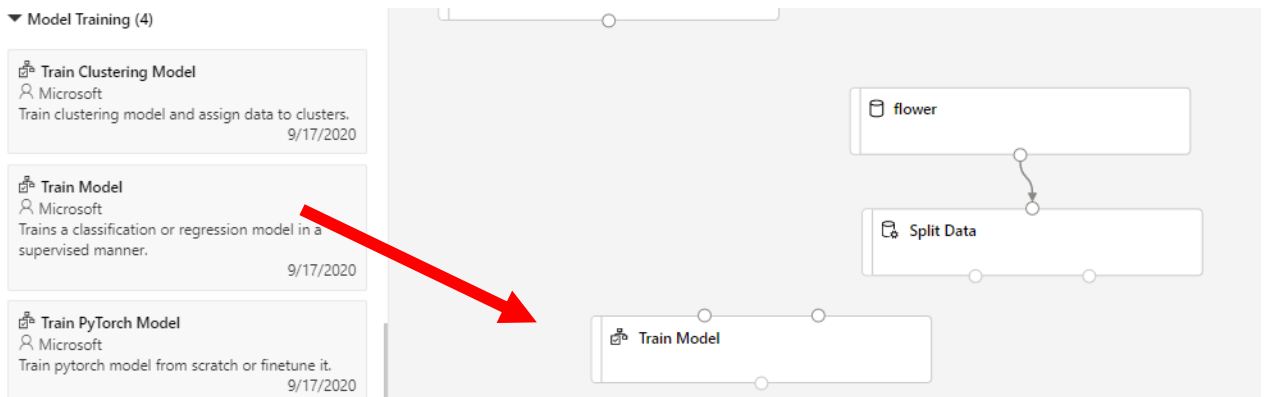
- Now, drag the end of the flower dataset module to the start of the Split Data module



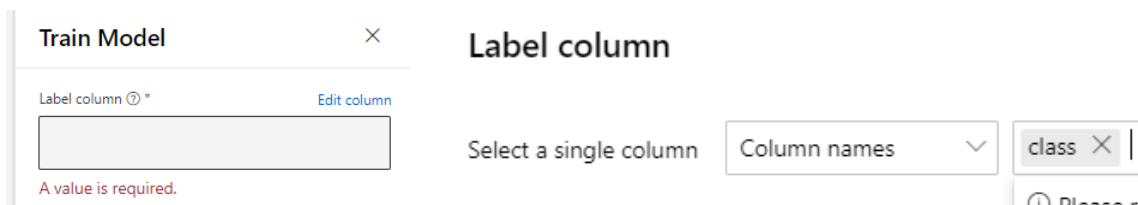
3. Modeling

Azure Machine Learning Designer

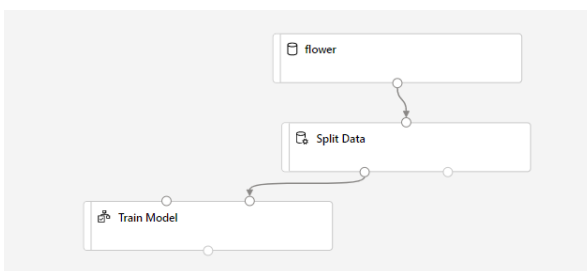
- Under model training, drag in Train Model



- For option, click on Edit column for Label column, and enter “class” for name



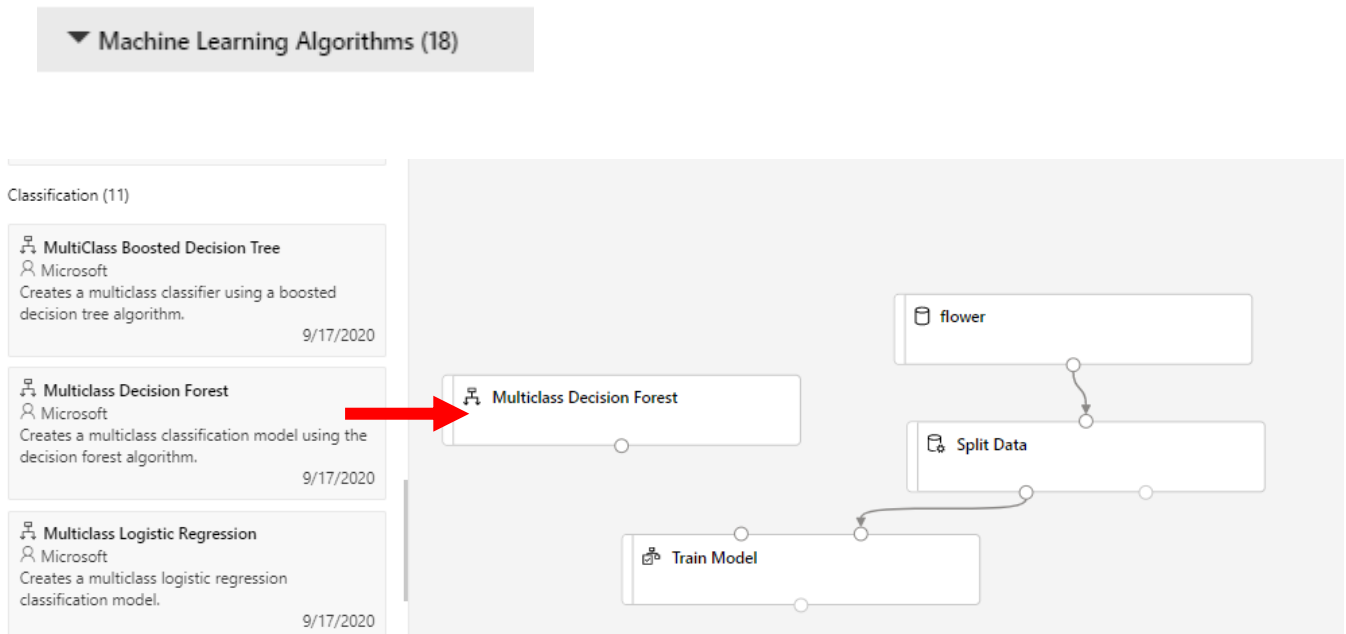
- Drag the first end of Split Data to the second top button of Train Model



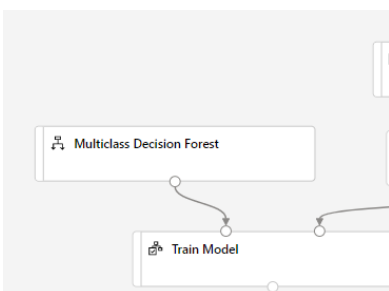
3. Modeling

Azure Machine Learning Designer

- Under Machine Learning algorithms, Classification, select Multi-Class Decision Forest and drag it in



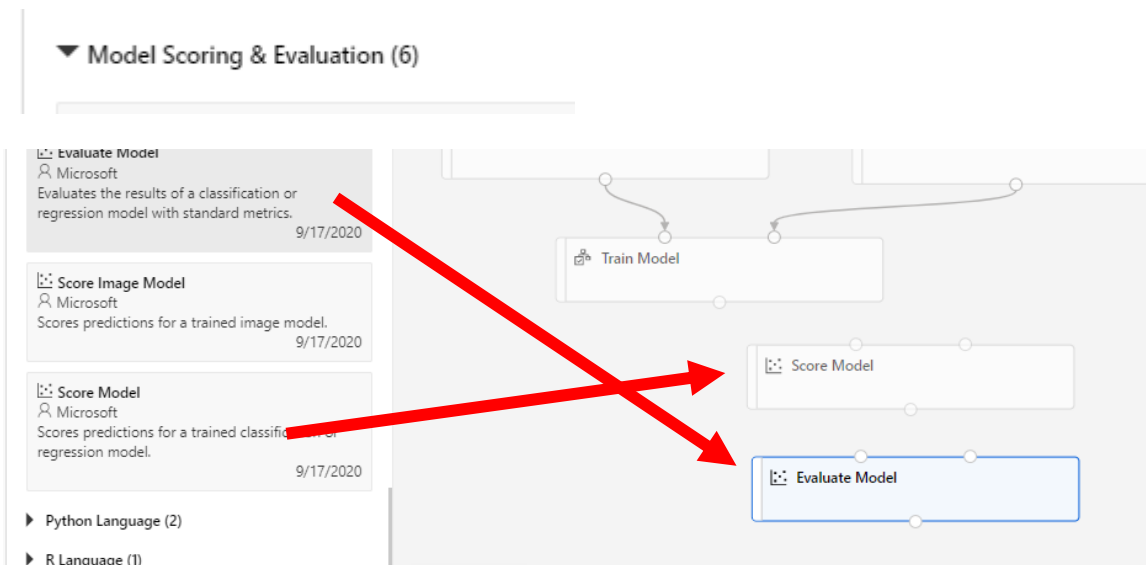
- Drag the bottom of Multiclass Decision Forest to the first button of Train Model



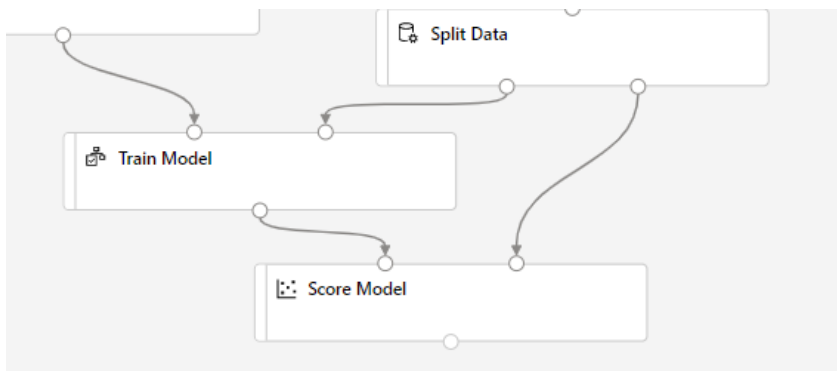
3. Modeling

Azure Machine Learning Designer

- Under Model Scoring & Evaluation, drag in Score Model and then below drag in Evaluate Model



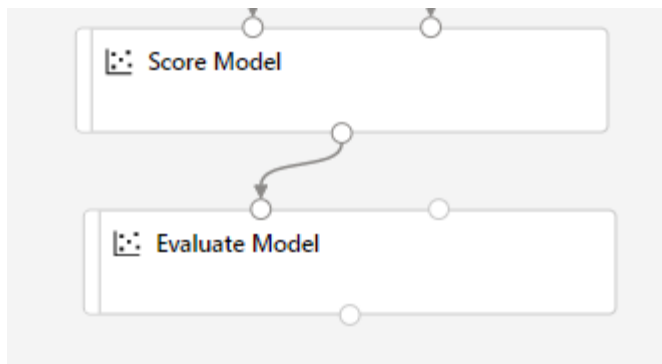
- Drag end of Train Model to first button of Score Model, and Split Data's second to second of Score Model



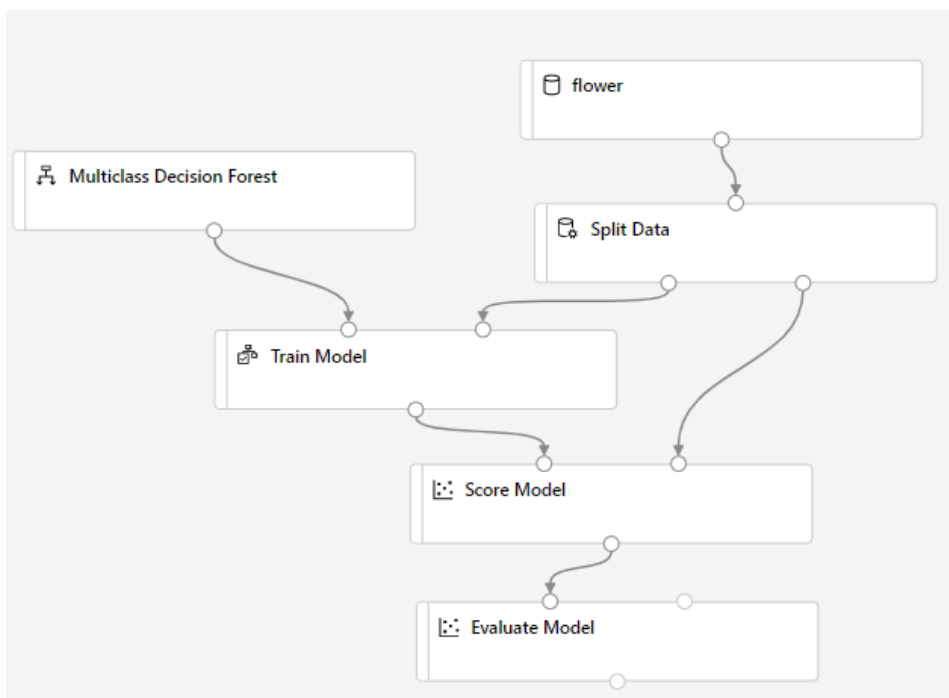
3. Modeling

Azure Machine Learning Designer

- Link Score Model's end to the first button of Evaluate Model



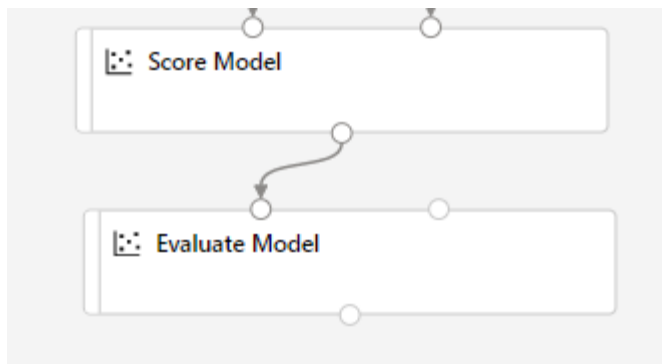
- Your entire experiment should look like this



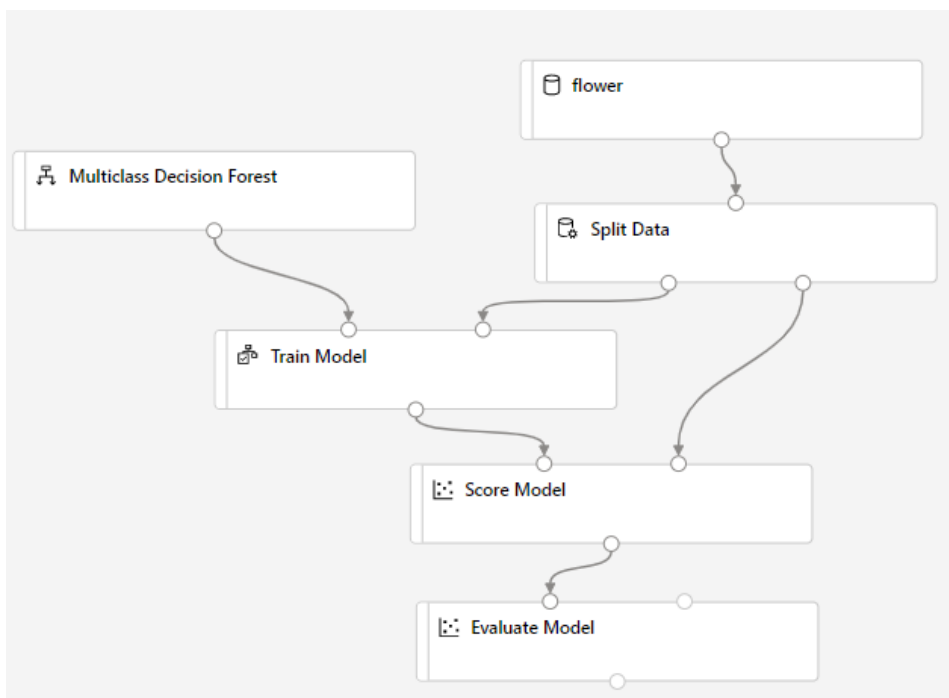
3. Modeling

Azure Machine Learning Designer

- Link Score Model's end to the first button of Evaluate Model



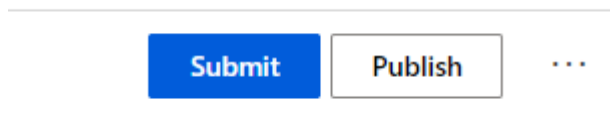
- Your entire experiment should look like this



3. Modeling

Azure Machine Learning Designer

- Now click on Submit to run, and Create new experiment to track



Experiment

☐ Select existing ☒ Create new

New experiment name *

experiment_run

Run description *




- After a few minutes, the experiment should complete, and you can click on Evaluate Model to see the results

Evaluate Model

Parameters **Outputs + logs** Details

Data outputs [Hide data outputs](#)

Evaluation results

Evaluate Model result visualization

Rows 1 Columns 5

Overall_Accuracy	Micro_Precision	Macro_Precision	Micro_F
0.933333	0.933333	0.944444	0.933333