How to use Azure Machine Learning with Python

Kinfey Lo

關於我

Port-Cloud Information Technology Limited

Microsoft Tech Community Regional Director

Microsoft Most Valuable Professional

Xamarin Most Valuable Professional

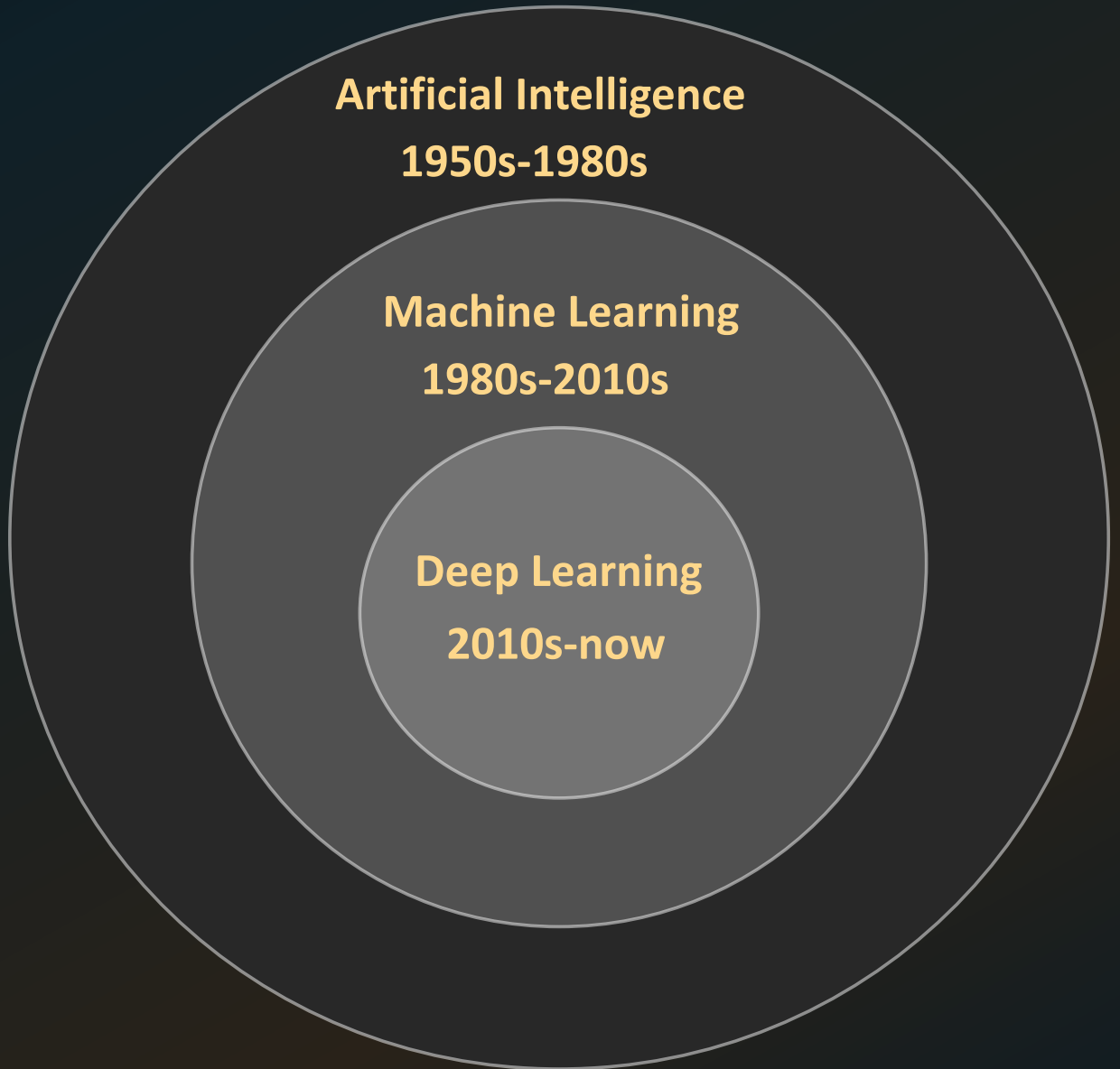
Microsoft Teched/Ignite/Teched Summit Speaker

PyCon 2019 Breakout Session Speaker

Mail: lokinfey@outlook.com

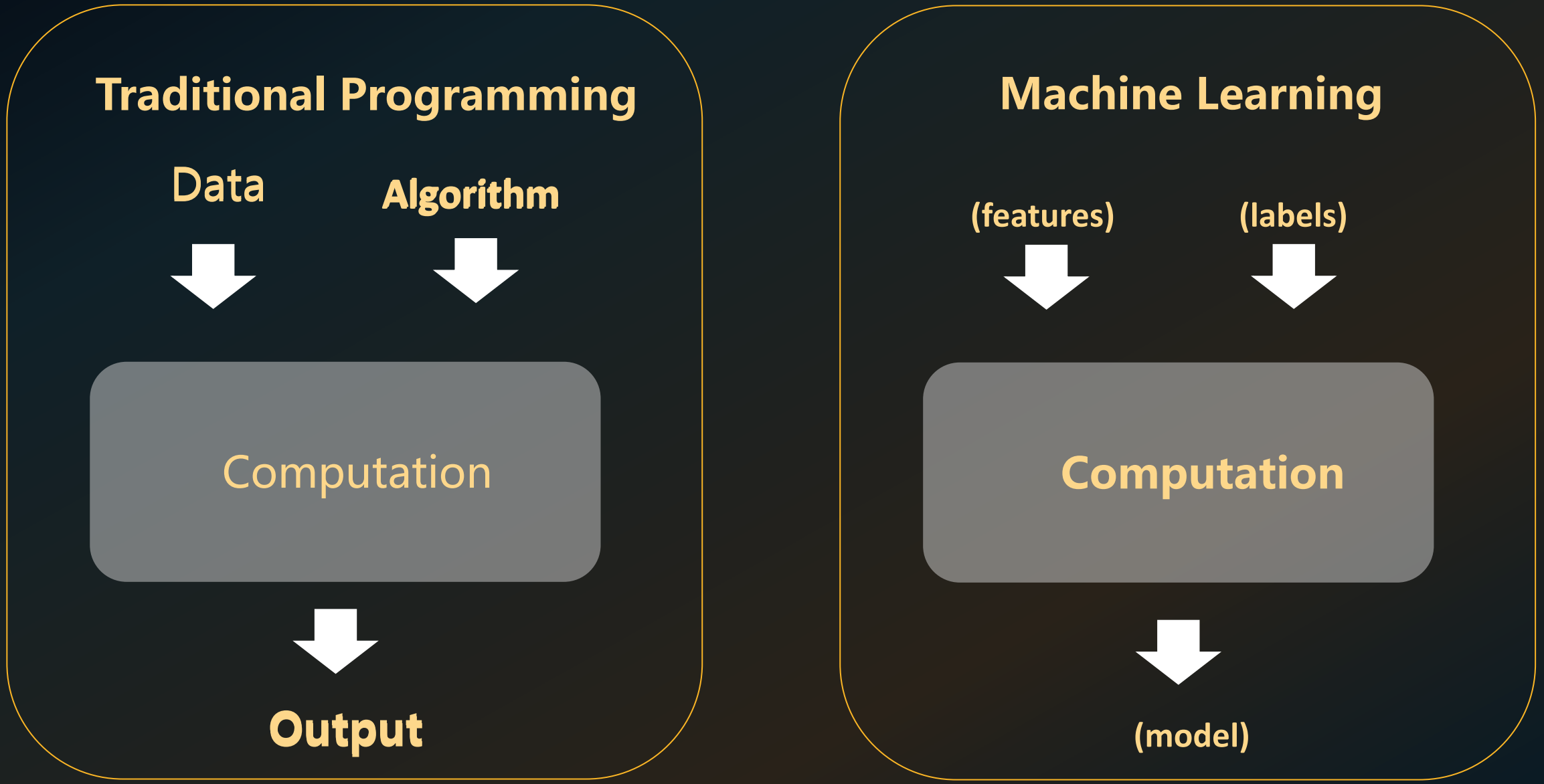
Github : https://github.com/lokinfey

Artificial Intelligence 發展歷程

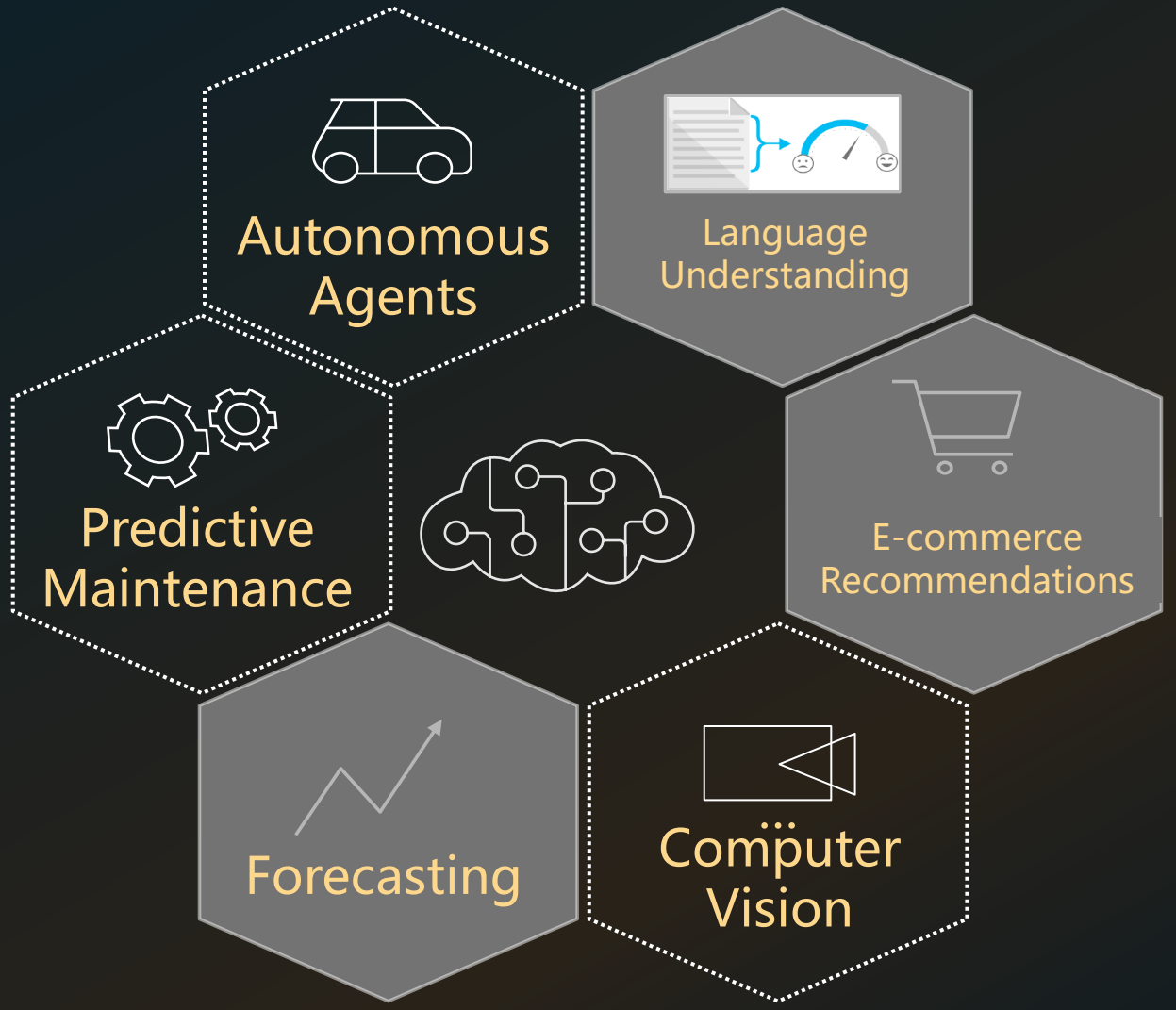


AI、Machine Learning 與 Deep Learning 關係

- Artificial Intelligence 是一種用電腦模仿人類的技術 (目標)
  - Machine Learning 讓電腦能夠依賴經驗更好去處理任務的多項技術(過程)
    - Deep Learning 通過Neutral Network 作為基礎，讓電腦自我訓練(方法)


Traditional Programming
Data
Algorithm
Computation
Output
Machine Learning
(features)
(labels)
Computation
(model)

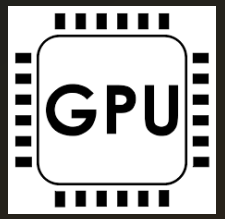
Artificial Intelligence的應用場景



Artificial Intelligence Stack

Language

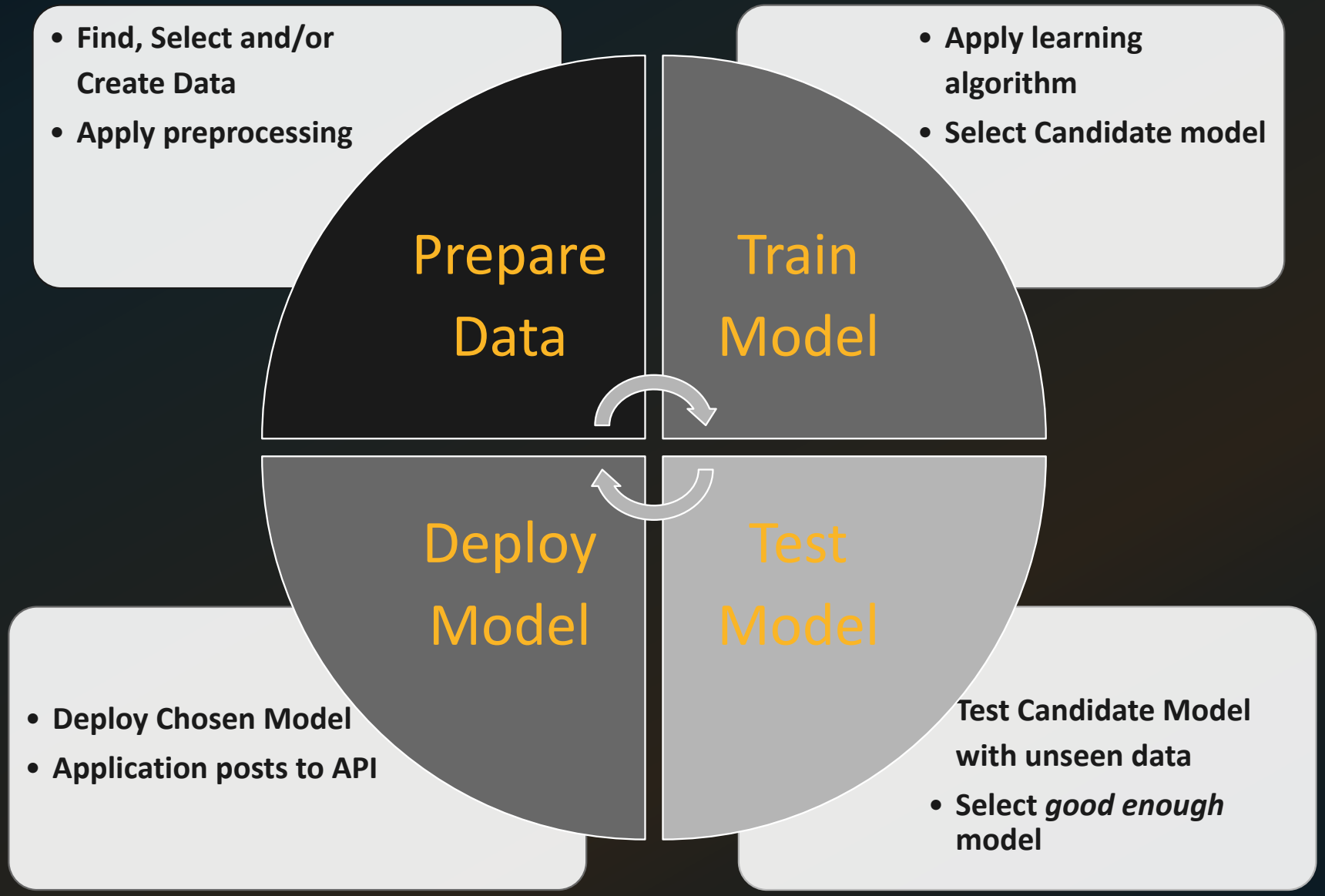
Framework



Hardware

Bye Bye AI ! ! !

完成一個Artificial Intelligence 過程



Azure AI

AI Application & Service

Azure Bot Service
Azure Cognitive Services

Machine Learning

Azure Databricks
Azure Machine Learning

Knowledge mining

Azure Cognitive Search

What is Azure Machine Learning Service?

Set of Azure Cloud Services

Python SDK

You can :

- ✓ Prepare data
- ✓ Build model
- ✓ Train model
- ✓ Manage Model
- ✓ Track Experiments
- ✓ Model Registration

Azure Machine Learning

Provide some predefined models
Can lower the development threshold


Vision


Speech


Language


Search

Compatible with different development tools
Quickly complete model development and simplify the development process


PyCharm


Jupyter


Visual Studio Code


Command line

Support for artificial intelligence development framework
Create deep learning solutions based on your needs


Pytorch


TensorFlow


Scikit-Learn


Onnx

Provide production services
Provide data and development environment for development team


Azure Databricks


Azure Machine Learning


Machine Learning VMs

Powerful hardware architecture support
Accelerate deep learning environment


CPU


GPU


FPGA

Azure Machine Learning service

- **Is an AI development platform that enhances productivity**

Improve data science productivity

Improve your training efficiency

Deploy and manage models anywhere

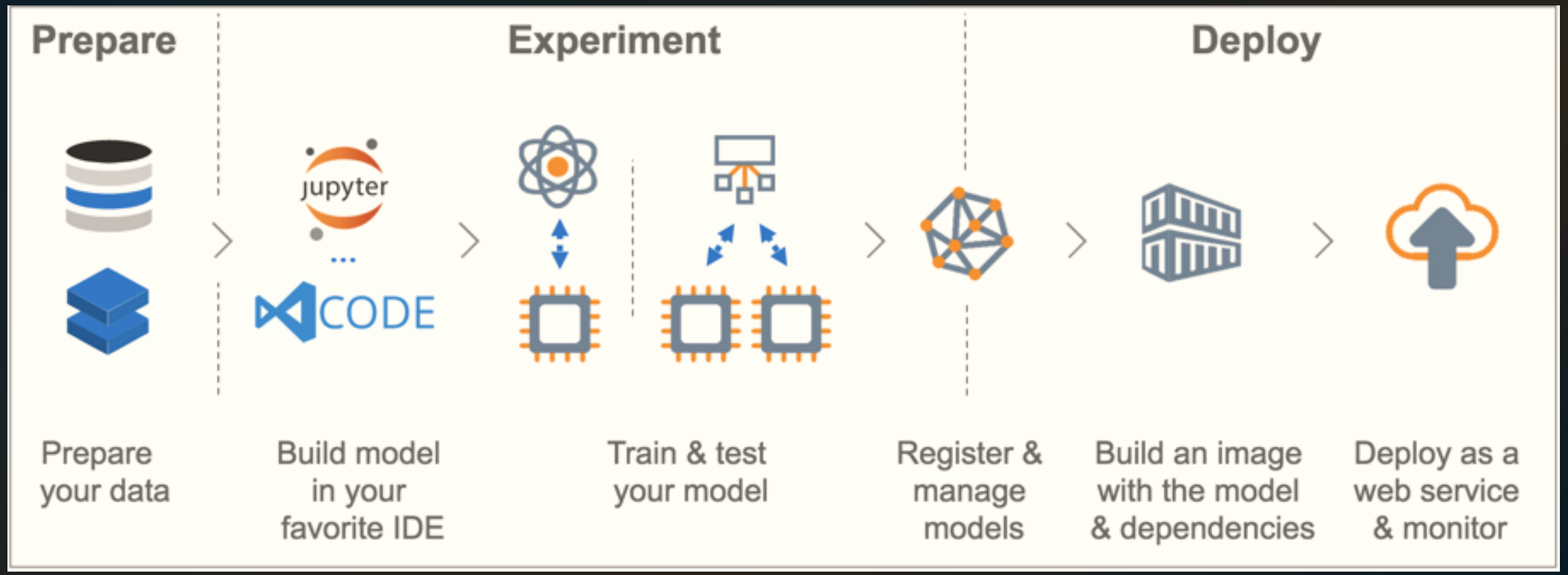
Create as needed

- Automated ML
- Managed Compute
- Deploy
- DevOps
- Support open source framrwork
- Python SDK

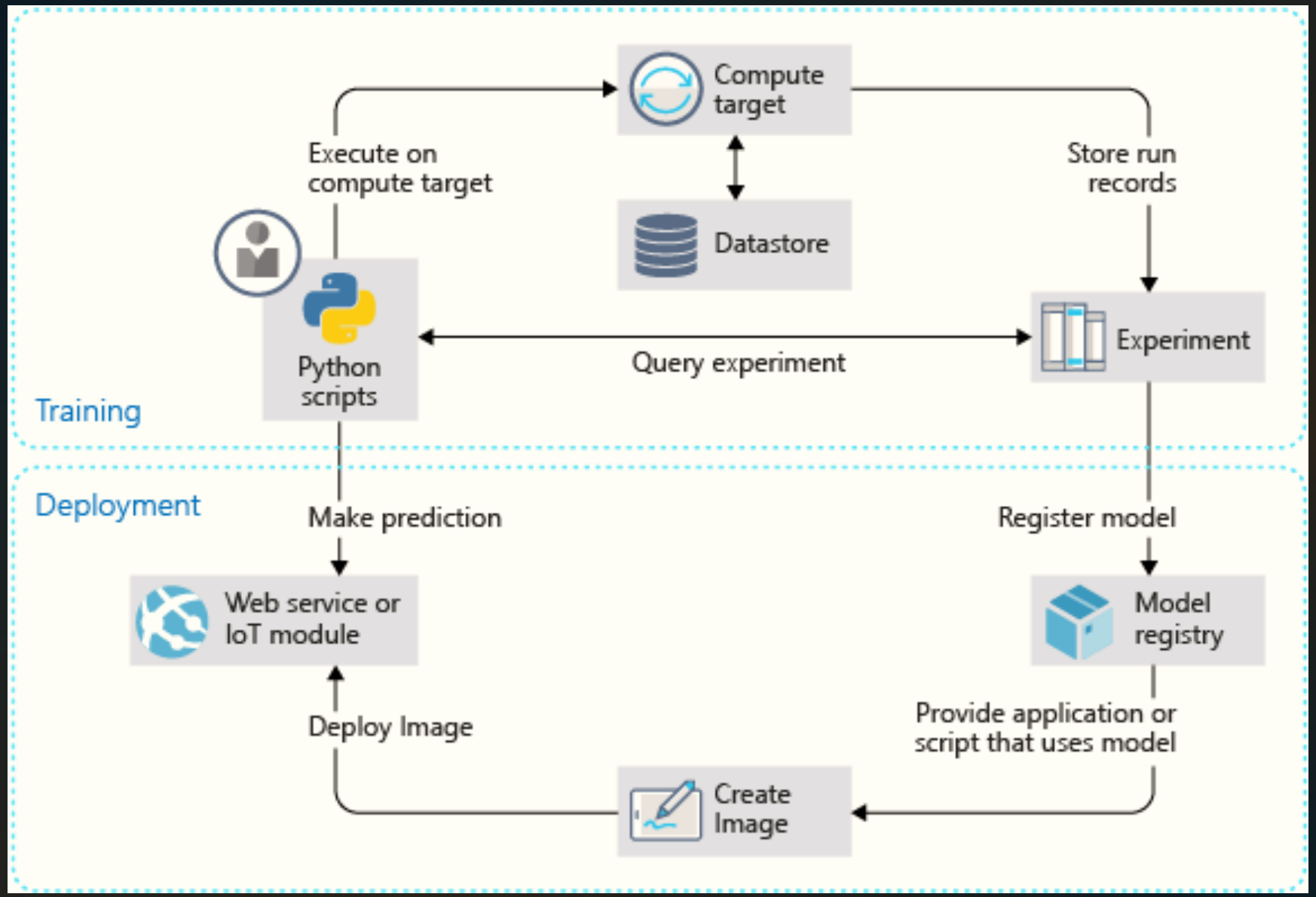
Seamlessly integrated with the Azure Portfolio

Azure Machine Learning Service

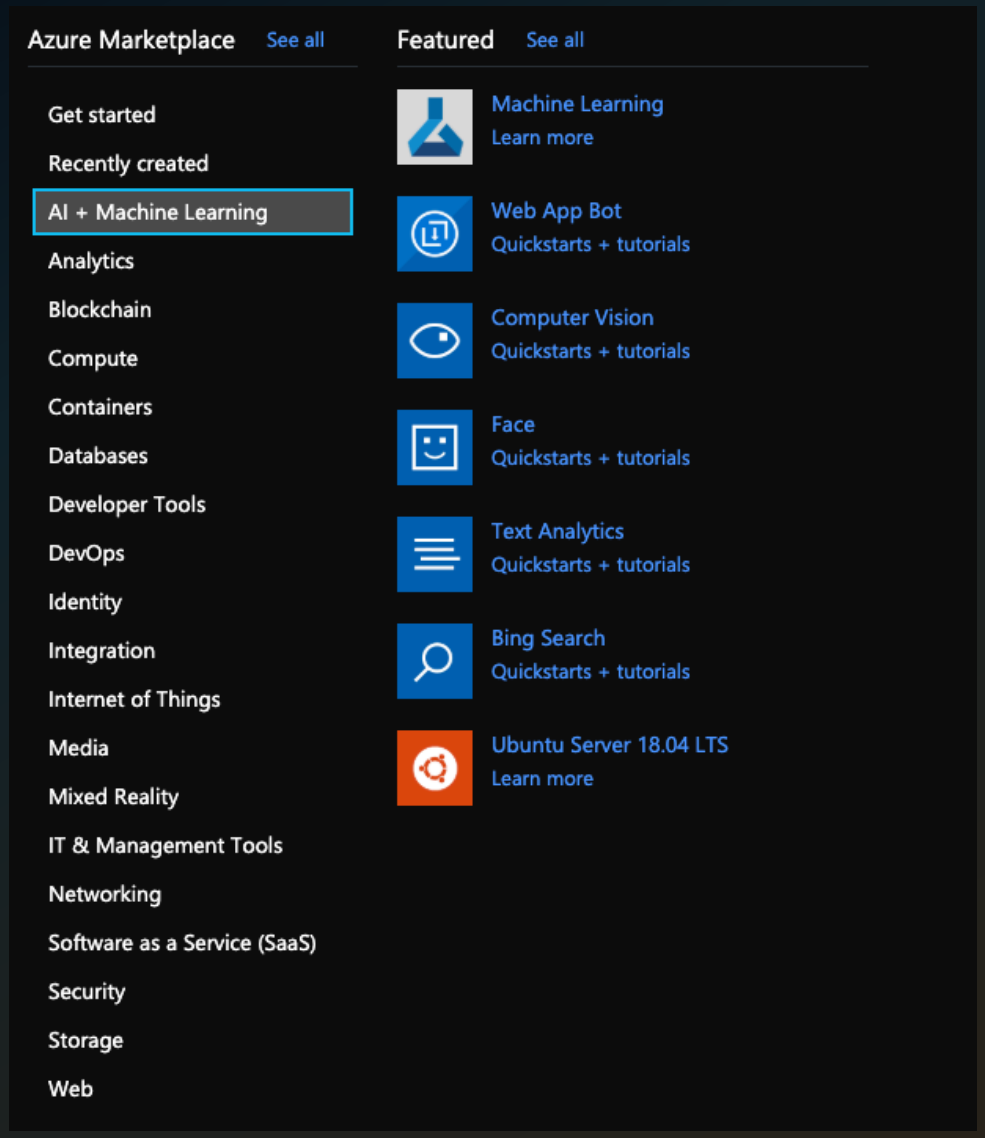
Azure Machine Learning Service provides a cloud-based environment that you can use to prepare data, train, test, deploy, manage, and track machine learning models.

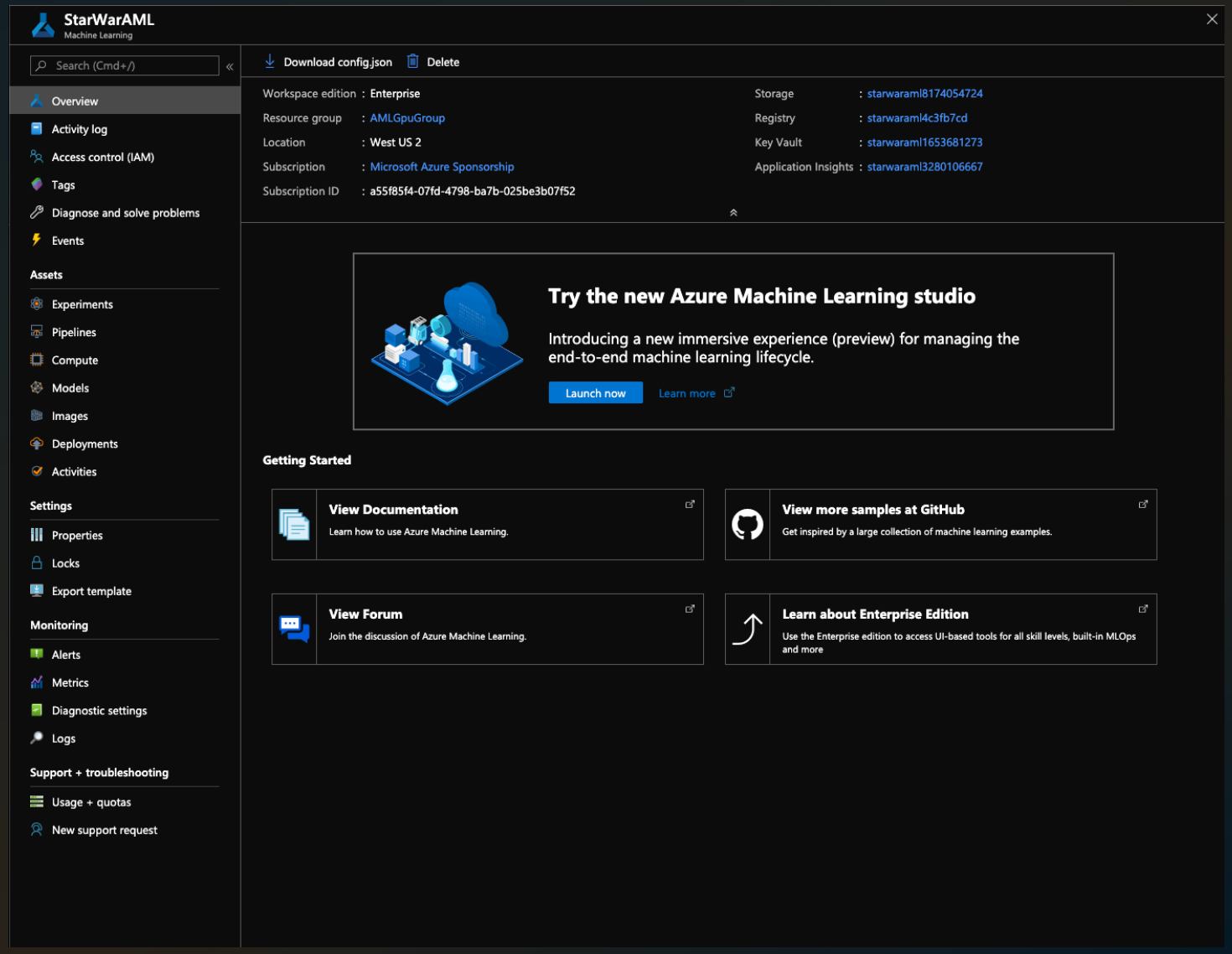


Azure Machine Learning Service

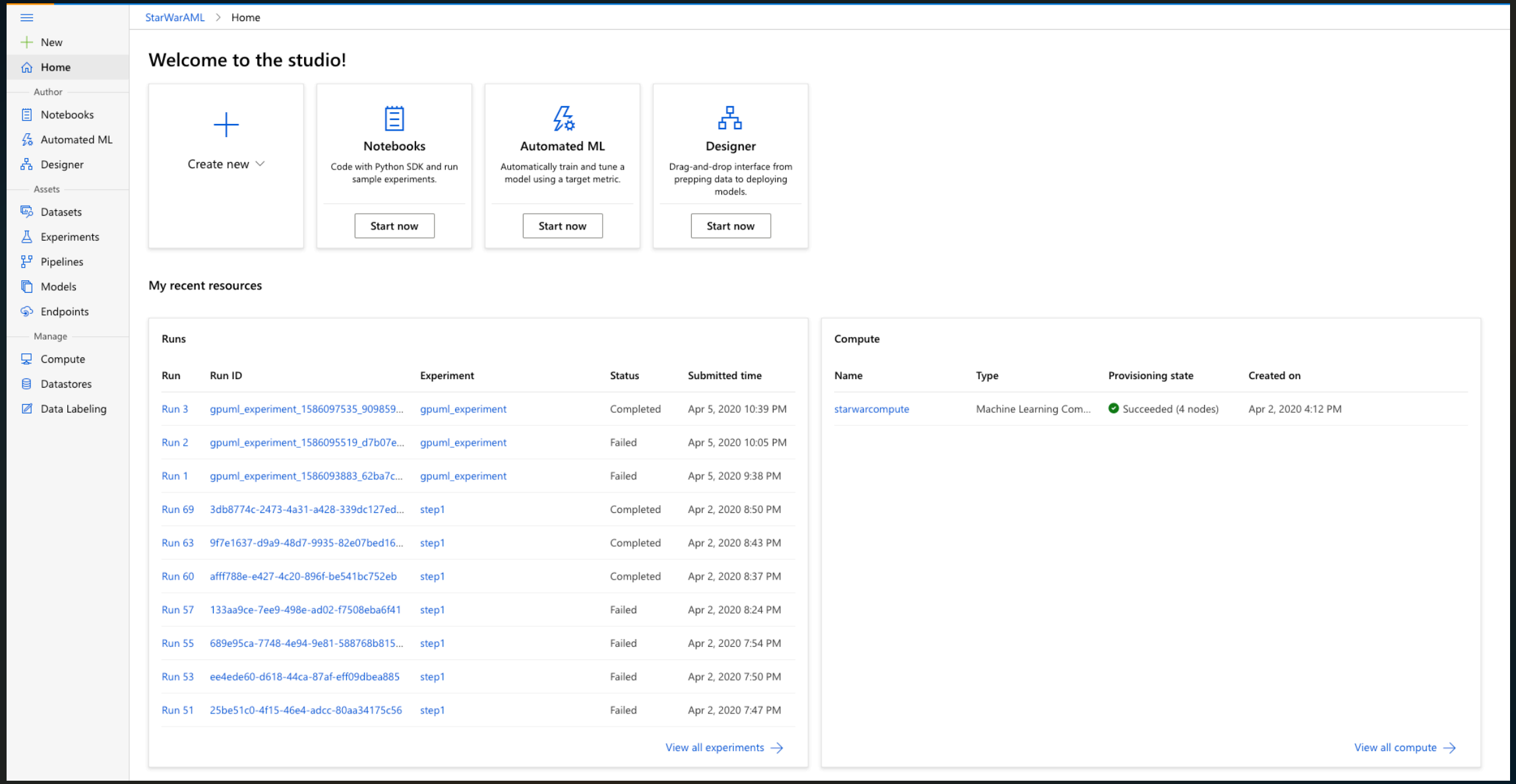


Azure Machine Learning Service





ml.azure.com


StarWarAML > Home
New
Home
Author
Notebooks
Automated ML
Designer
Assets
Datasets
Experiments
Pipelines
Models
Endpoints
Manage
Compute
Datastores
Data Labeling
Welcome to the studio!
Create new
Notebooks
Code with Python SDK and run sample experiments.
Start now
Automated ML
Automatically train and tune a model using a target metric.
Start now
Designer
Drag-and-drop interface from prepping data to deploying models.
Start now
My recent resources
Runs
Run | Run ID | Experiment | Status | Submitted time
Run 3 | gpuml_experiment_1586097535_909859... | gpuml_experiment | Completed | Apr 5, 2020 10:39 PM
Run 2 | gpuml_experiment_1586095519_d7b07e... | gpuml_experiment | Failed | Apr 5, 2020 10:05 PM
Run 1 | gpuml_experiment_1586093883_62ba7c... | gpuml_experiment | Failed | Apr 5, 2020 9:38 PM
Run 69 | 3db8774c-2473-4a31-a428-339dc127ed... | step1 | Completed | Apr 2, 2020 8:50 PM
Run 63 | 9f7e1637-d9a9-48d7-9935-82e07bed16... | step1 | Completed | Apr 2, 2020 8:43 PM
Run 60 | afff788e-e427-4c20-896f-be541bc752eb | step1 | Completed | Apr 2, 2020 8:37 PM
Run 57 | 133aa9ce-7ee9-498e-ad02-f7508eba6f41 | step1 | Failed | Apr 2, 2020 8:24 PM
Run 55 | 689e95ca-7748-4e94-9e81-588768b815... | step1 | Failed | Apr 2, 2020 7:54 PM
Run 53 | ee4ede60-d618-44ca-87af-eff09dbea885 | step1 | Failed | Apr 2, 2020 7:50 PM
Run 51 | 25be51c0-4f15-46e4-adcc-80aa34175c56 | step1 | Failed | Apr 2, 2020 7:47 PM
View all experiments
Compute
Name | Type | Provisioning state | Created on
starwarcompute | Machine Learning Com... | Succeeded (4 nodes) | Apr 2, 2020 4:12 PM
View all compute

Azure Machine Learning Service

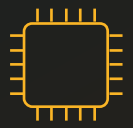
- Key Artifacts

Workspace

Models

Experiments

Pipelines



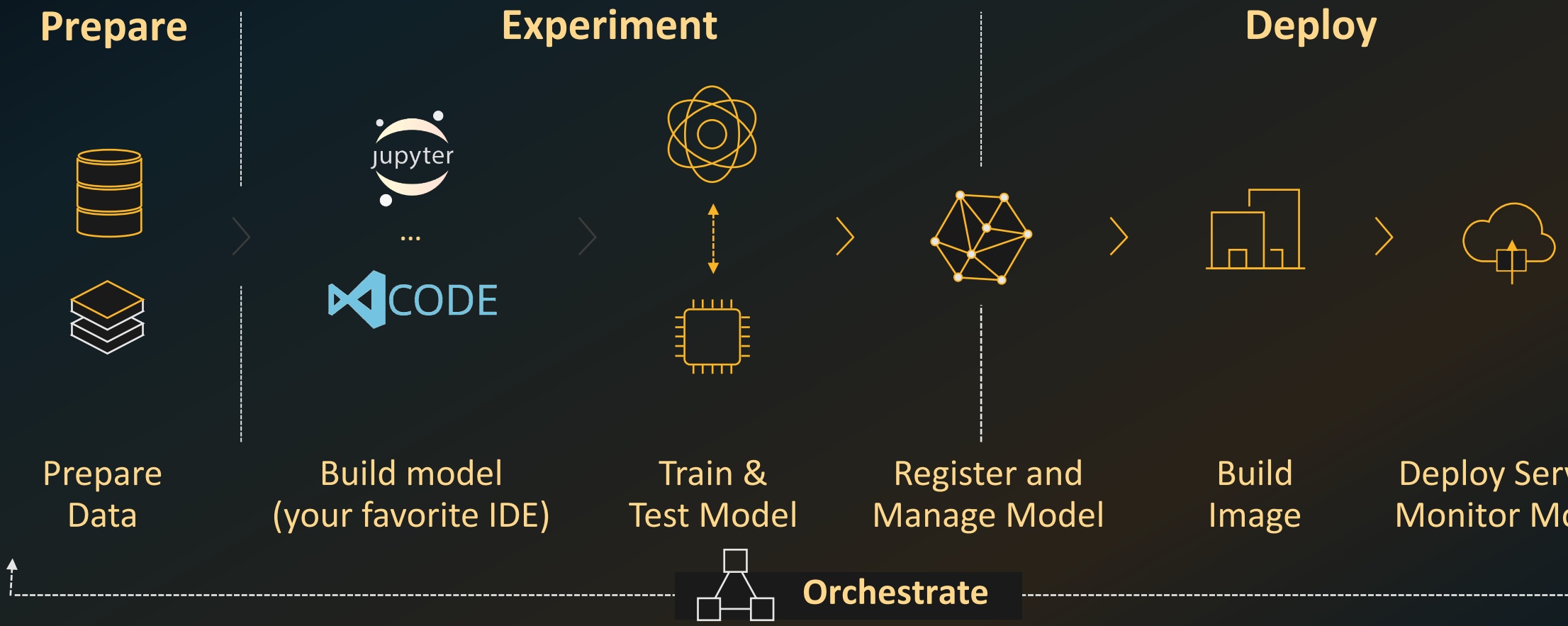
Compute Target

Images

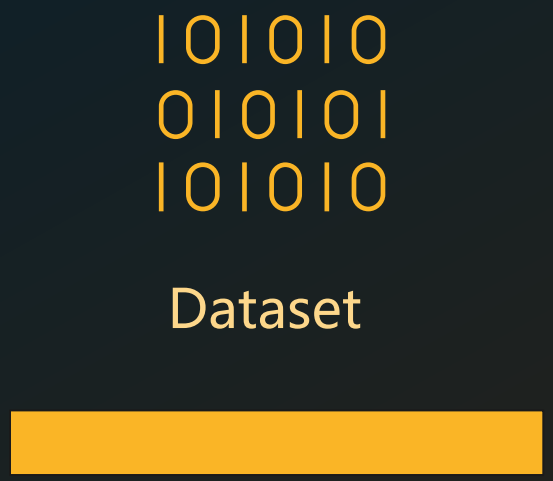
Deployment

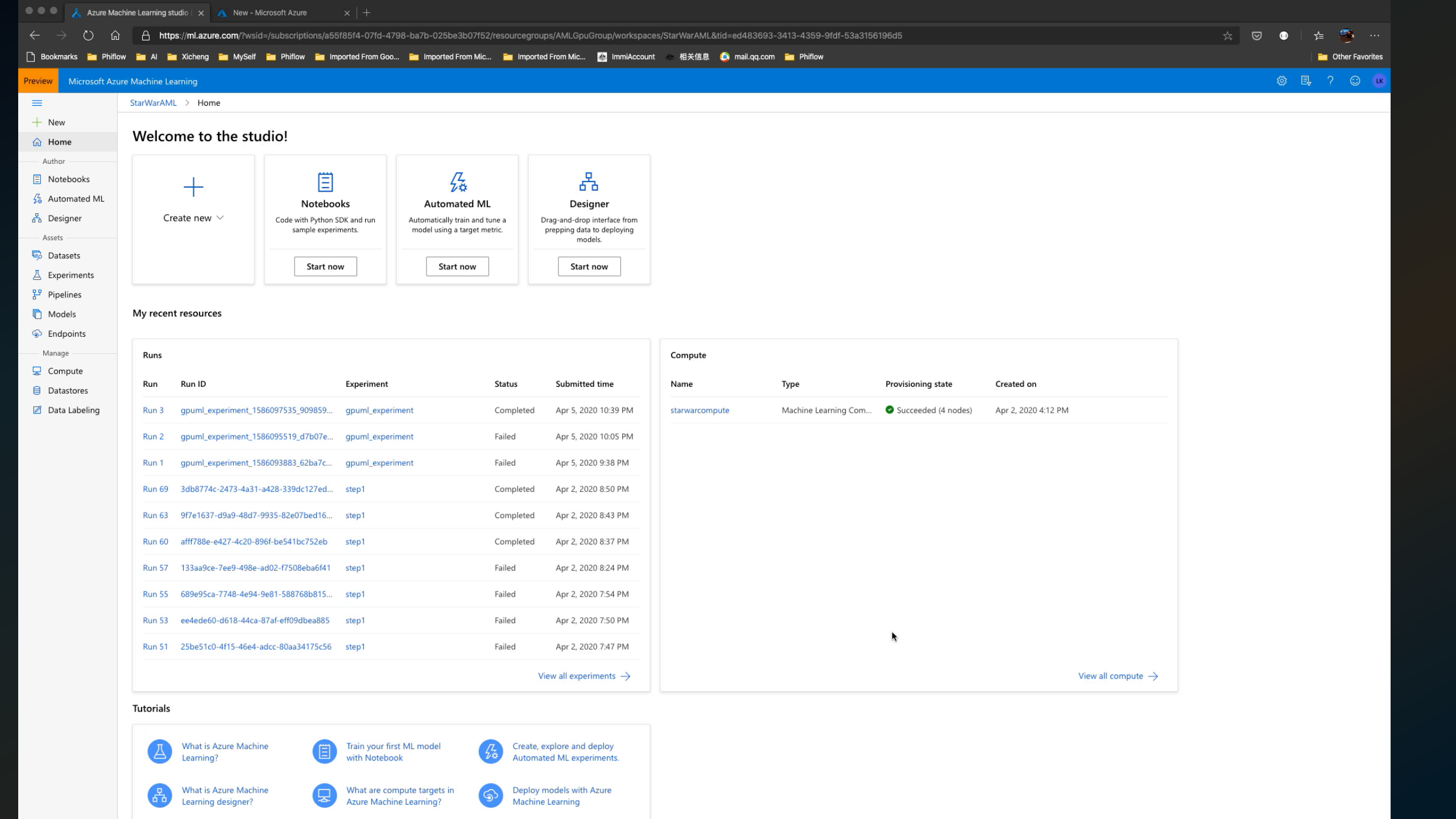
Data Stores

Azure Machine Learning



Azure Machine Learning Pipelines




Azure Machine Learning studio
New - Microsoft Azure
https://ml.azure.com/?wsid=/subscriptions/a55f85f4-07fd-4798-ba7b-025be3b07f52/resourcegroups/AMLGpuGroup/workspaces/StarWarAML&tid=ed483693-3413-4359-9fdf-53a3156196d5
Bookmarks
Phiflow
AI
Xicheng
MySelf
Phiflow
Imported From Goo...
Imported From Mic...
Imported From Mic...
ImmiAccount
相关信息
mail.qq.com
Phiflow
Other Favorites
Preview
Microsoft Azure Machine Learning
New
Home
Author
Notebooks
Automated ML
Designer
Assets
Datasets
Experiments
Pipelines
Models
Endpoints
Manage
Compute
Datastores
Data Labeling
StarWarAML
Home
Welcome to the studio!
Create new
Notebooks
Code with Python SDK and run sample experiments.
Start now
Automated ML
Automatically train and tune a model using a target metric.
Start now
Designer
Drag-and-drop interface from prepping data to deploying models.
Start now
My recent resources
Runs
Run
Run ID
Experiment
Status
Submitted time
Run 3
gpuml_experiment_1586097535_909859...
gpuml_experiment
Completed
Apr 5, 2020 10:39 PM
Run 2
gpuml_experiment_1586095519_d7b07e...
gpuml_experiment
Failed
Apr 5, 2020 10:05 PM
Run 1
gpuml_experiment_1586093883_62ba7c...
gpuml_experiment
Failed
Apr 5, 2020 9:38 PM
Run 69
3db8774c-2473-4a31-a428-339dc127ed...
step1
Completed
Apr 2, 2020 8:50 PM
Run 63
9f7e1637-d9a9-48d7-9935-82e07bed16...
step1
Completed
Apr 2, 2020 8:43 PM
Run 60
afff788e-e427-4c20-896f-be541bc752eb
step1
Completed
Apr 2, 2020 8:37 PM
Run 57
133aa9ce-7ee9-498e-ad02-f7508eba6f41
step1
Failed
Apr 2, 2020 8:24 PM
Run 55
689e95ca-7748-4e94-9e81-588768b815...
step1
Failed
Apr 2, 2020 7:54 PM
Run 53
ee4ede60-d618-44ca-87af-eff09dbea885
step1
Failed
Apr 2, 2020 7:50 PM
Run 51
25be51c0-4f15-46e4-adcc-80aa34175c56
step1
Failed
Apr 2, 2020 7:47 PM
View all experiments
Compute
Name
Type
Provisioning state
Created on
starwarcompute
Machine Learning Com...
Succeeded (4 nodes)
Apr 2, 2020 4:12 PM
View all compute
Tutorials
What is Azure Machine Learning?
Train your first ML model with Notebook
Create, explore and deploy Automated ML experiments.
What is Azure Machine Learning designer?
What are compute targets in Azure Machine Learning?
Deploy models with Azure Machine Learning

dataset.ipynb* — mldemo

EXPLORER

OPEN EDITORS
- datasets_init.ipynb
- dataset.ipynb*

MLDEMO
- datasets
- resized_dataset
- aml.ipynb
- config.json
- dataset.ipynb
- datasets_init.ipynb
- train.py

OUTLINE

TIMELINE

datasets_init.ipynb | dataset.ipynb*

Jupyter Server: loc… Python 3.7.5 64-bit: I…

[1]
```
from azureml.core import Workspace, Dataset
import azureml.contrib.dataset

subscription_id = ''
resource_group = 'AMLGpuGroup'
workspace_name = 'StarWarAML'

workspace = Workspace(subscription_id, resource_group, workspace_name)

dataset = Dataset.get_by_name(workspace, name='StarWarLabel-2020-04-02 11:53:10')
dataset.to_pandas_dataframe()
```

```
WARNING - Warning: Falling back to use azure cli login credentials.
If you run your code in unattended mode, i.e., where you can't give a user input, then we recommend to use ServicePrincipalAuthentication or MsiAuthentication.
Please refer to aka.ms/aml-notebook-auth for different authentication mechanisms in azureml-sdk.
```

	image_url	label	label_confidence
0	StreamInfo(AmlDatastore://UI/04-02-2020_073223...	r2d2	1.0
1	StreamInfo(AmlDatastore://UI/04-02-2020_073223...	c3po	1.0
2	StreamInfo(AmlDatastore://UI/04-02-2020_073223...	r2d2	1.0
3	StreamInfo(AmlDatastore://UI/04-02-2020_073223...	c3po	1.0
4	StreamInfo(AmlDatastore://UI/04-02-2020_073223...	c3po	1.0
5	StreamInfo(AmlDatastore://UI/04-02-2020_073223...	r2d2	1.0
6	StreamInfo(AmlDatastore://UI/04-02-2020_073223...	c3po	1.0
7	StreamInfo(AmlDatastore://UI/04-02-2020_073223...	c3po	1.0
8	StreamInfo(AmlDatastore://UI/04-02-2020_073223...	r2d2	1.0
9	StreamInfo(AmlDatastore://UI/04-02-2020_073223...	r2d2	1.0
10	StreamInfo(AmlDatastore://UI/04-02-2020_073223...	c3po	1.0
11	StreamInfo(AmlDatastore://UI/04-02-2020_073223...	c3po	1.0

[2]
```
print('Workspace name: ' + workspace.name,
      'Azure region: ' + workspace.location,
      'Subscription id: ' + workspace.subscription_id,
      'Resource group: ' + workspace.resource_group, sep='\n')
```

```
Workspace name: StarWarAML
Azure region: westus2
Subscription id: a55f85f4-07fd-4798-ba7b-025be3b07[illegible]2
Resource group: AMLGpuGroup
```

[4]
```
starwar_labels = Dataset.get_by_name(workspace, name='StarWarLabel-2020-04-02 11:53:10')
starwar_labels
```

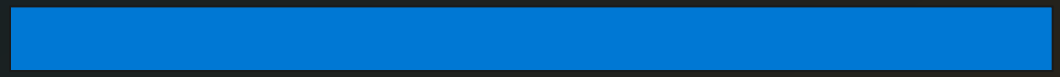

```
{
  "source": [
    "('workspaceblobstore', '/export/dataset/7cba7279-5ced-6ccf-1233-e53167d797c6/4984c20b-a101-4767-9909-082fd3766d62/c2c1a5bc-bc3f-4558-b926-5fff9b17b706/LabeledDatasetJsonLines.json')"
  ],
  "definition": [
    "GetDatastoreFiles",
    "ParseJsonLines",
    "ExpressionAddColumn",
    "DropColumns",
    "DropColumns",
    "RenameColumns"
  ],
  "registration": {
    "id": "b5212952-a08a-4b26-8ada-77138ef9fcb9",
    "name": "StarWarLabel-2020-04-02 11:53:10",
    "version": 1,
    "description": "LabeledDs_StarWarLabel Of Type ImageClassificationMultiClass, Sourced From 4984c20b-a101-4767-9909-082fd3766d62",
```

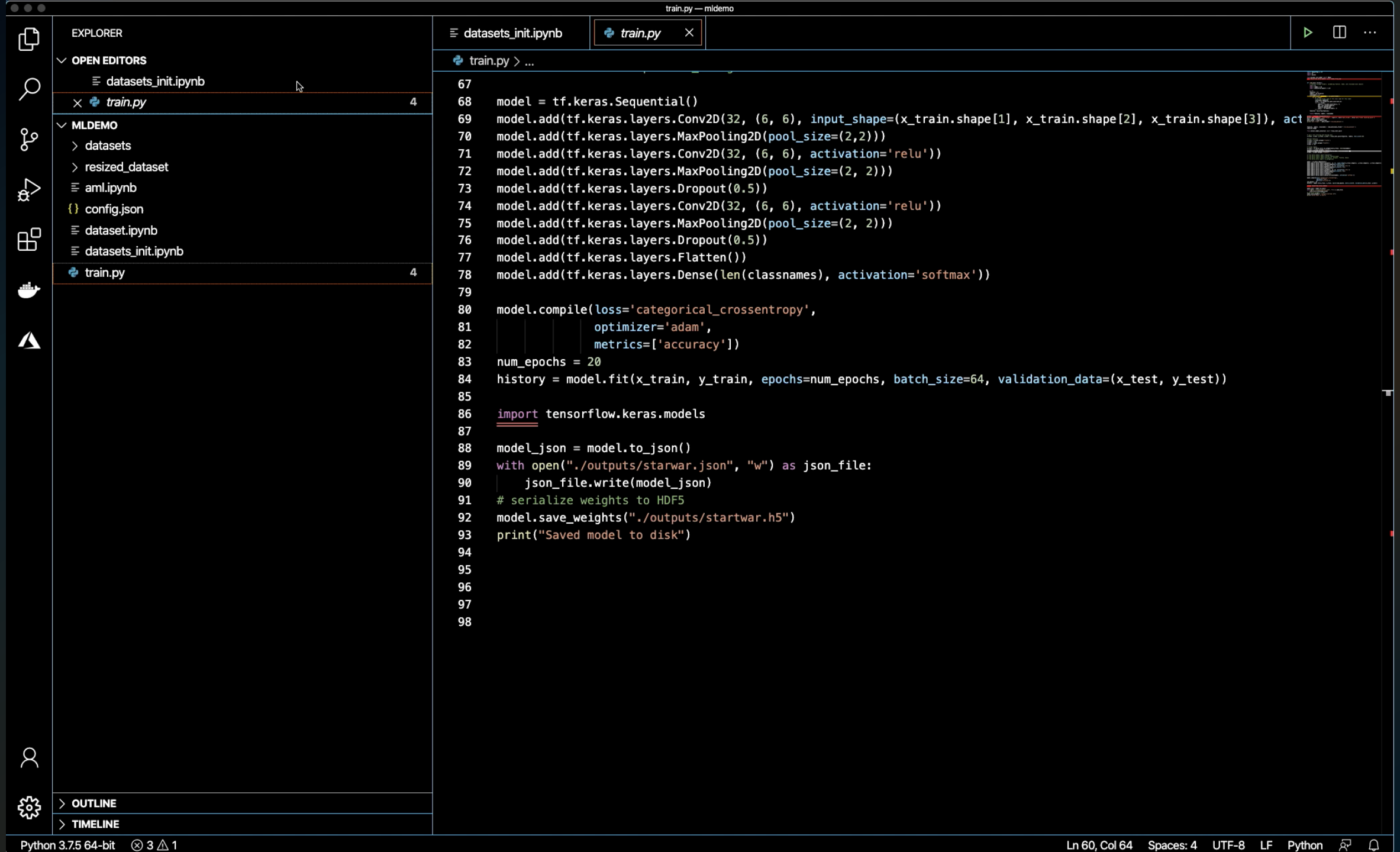

Python 3.7.5 64-bit ⊗ 0 ⚠ 0

Azure Machine Learning Pipelines

Dataset

Training




train.py — mldemo
EXPLORER
OPEN EDITORS
datasets_init.ipynb
train.py 4
MLDEMO
datasets
resized_dataset
aml.ipynb
config.json
dataset.ipynb
datasets_init.ipynb
train.py 4
OUTLINE
TIMELINE
datasets_init.ipynb
train.py
train.py > ...
67
68 model = tf.keras.Sequential()
69 model.add(tf.keras.layers.Conv2D(32, (6, 6), input_shape=(x_train.shape[1], x_train.shape[2], x_train.shape[3]), act
70 model.add(tf.keras.layers.MaxPooling2D(pool_size=(2,2)))
71 model.add(tf.keras.layers.Conv2D(32, (6, 6), activation='relu'))
72 model.add(tf.keras.layers.MaxPooling2D(pool_size=(2, 2)))
73 model.add(tf.keras.layers.Dropout(0.5))
74 model.add(tf.keras.layers.Conv2D(32, (6, 6), activation='relu'))
75 model.add(tf.keras.layers.MaxPooling2D(pool_size=(2, 2)))
76 model.add(tf.keras.layers.Dropout(0.5))
77 model.add(tf.keras.layers.Flatten())
78 model.add(tf.keras.layers.Dense(len(classnames), activation='softmax'))
79
80 model.compile(loss='categorical_crossentropy',
81 optimizer='adam',
82 metrics=['accuracy'])
83 num_epochs = 20
84 history = model.fit(x_train, y_train, epochs=num_epochs, batch_size=64, validation_data=(x_test, y_test))
85
86 import tensorflow.keras.models
87
88 model_json = model.to_json()
89 with open("./outputs/starwar.json", "w") as json_file:
90 json_file.write(model_json)
91 # serialize weights to HDF5
92 model.save_weights("./outputs/startwar.h5")
93 print("Saved model to disk")
94
95
96
97
98
Python 3.7.5 64-bit
3 1
Ln 60, Col 64
Spaces: 4
UTF-8
LF
Python

Azure Machine Learning Pipelines

Dataset

Training

Model Registration

AMLTrain.ipynb — AzureCOVID

EXPLORER

OPEN EDITORS
- AMLTrain.ipynb

AZURECOVID
- .ipynb_checkpoints
- dataset
- model
- AMLTrain.ipynb
- config.json
- test001.png
- test002.png
- test003.png
- train.py

OUTLINE

TIMELINE

AMLTrain.ipynb

Jupyter Server: lo… Python 3: Idle

[1]
```
import azureml.core
print(azureml.core.VERSION)
```
```
1.0.85
```

[3]
```
import azureml.core
from azureml.core import Workspace, Datastore

ws = Workspace.from_config()
```

[4]
```
ds = ws.get_default_datastore()
```

[5]
```
datastores = ws.datastores
for name, ds in datastores.items():
    print(name, ds.datastore_type)
```
```
workspaceblobstore AzureBlob
workspacefilestore AzureFile
```

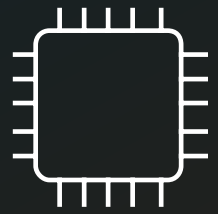
[7]
```
import azureml.data
from azureml.data.azure_storage_datastore import AzureFileDatastore, AzureBlobDatastore

ds.upload(src_dir='./dataset',
          target_path='dataset',
          overwrite=True,
          show_progress=True)
```
```
Output was trimmed for performance reasons.
To see the full output set the setting "python.dataScience.textOutputLimit" to 0.
...
person291_virus_596.jpeg, 5845 files out of an estimated total of 5978
Uploading ./dataset/train/person731_bacteria_2633.jpeg
Uploaded ./dataset/train/person58_bacteria_272.jpeg, 5846 files out of an estimated total of 5978
Uploading ./dataset/train/person860_virus_1505.jpeg
Uploading ./dataset/train/person1438_bacteria_3718.jpeg
Uploaded ./dataset/train/person731_bacteria_2633.jpeg, 5847 files out of an estimated total of 5978
Uploaded ./dataset/train/person24_bacteria_111.jpeg, 5848 files out of an estimated total of 5978
Uploaded ./dataset/train/person860_virus_1505.jpeg, 5849 files out of an estimated total of 5978
Uploaded ./dataset/train/person1438_bacteria_3718.jpeg, 5850 files out of an estimated total of 5978
Uploading ./dataset/train/person397_virus_790.jpeg
Uploading ./dataset/train/person1235_virus_2095.jpeg
Uploading ./dataset/train/NORMAL2-IM-0553-0001.jpeg
```

Python 3.7.5 64-bit ⊗ 0 ⚠ 0

Azure Machine Learning Pipelines

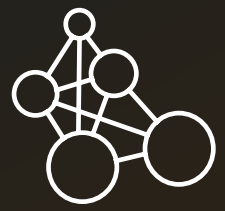
Dataset



Compute

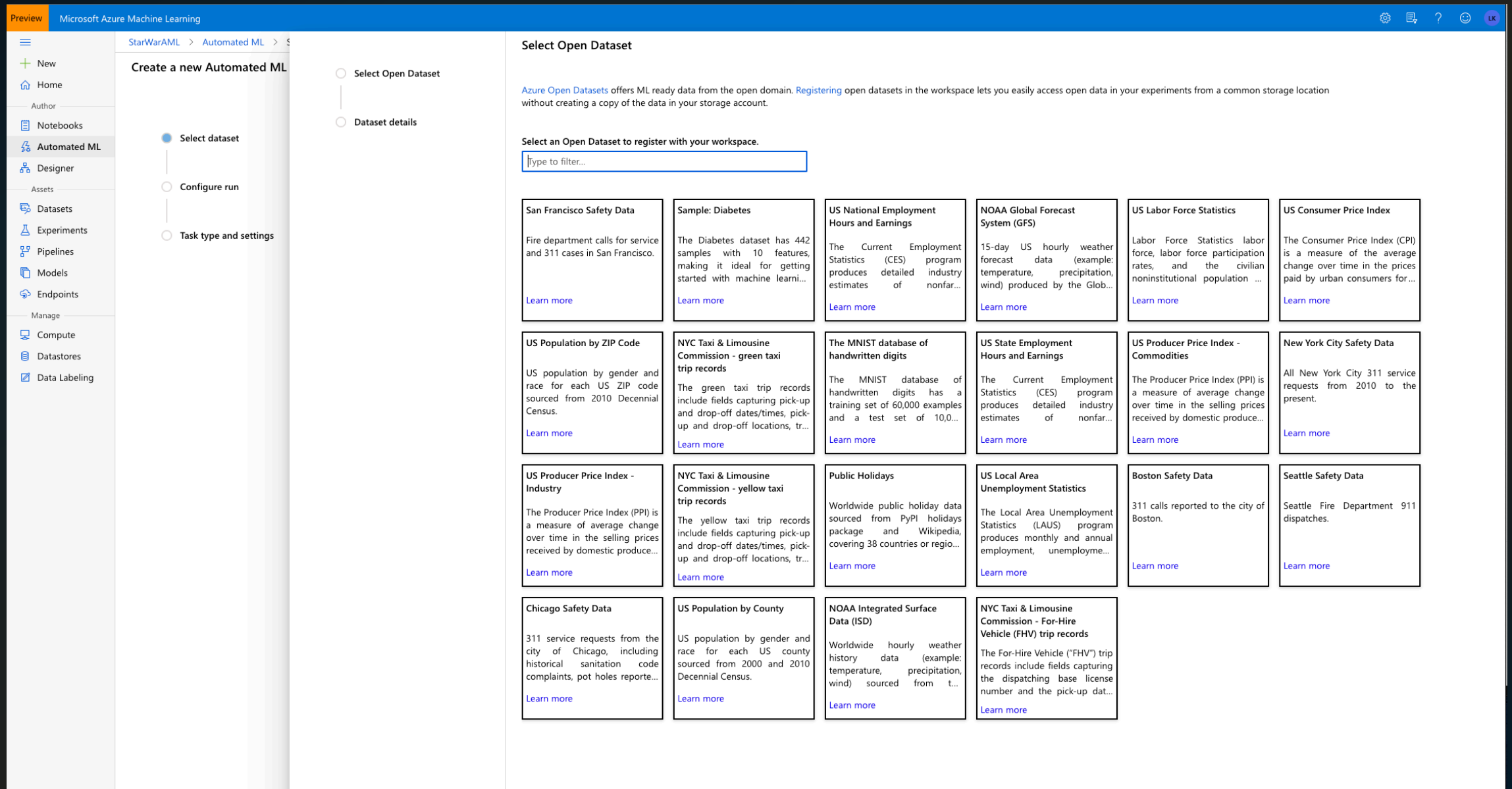
Data Preparation

Training

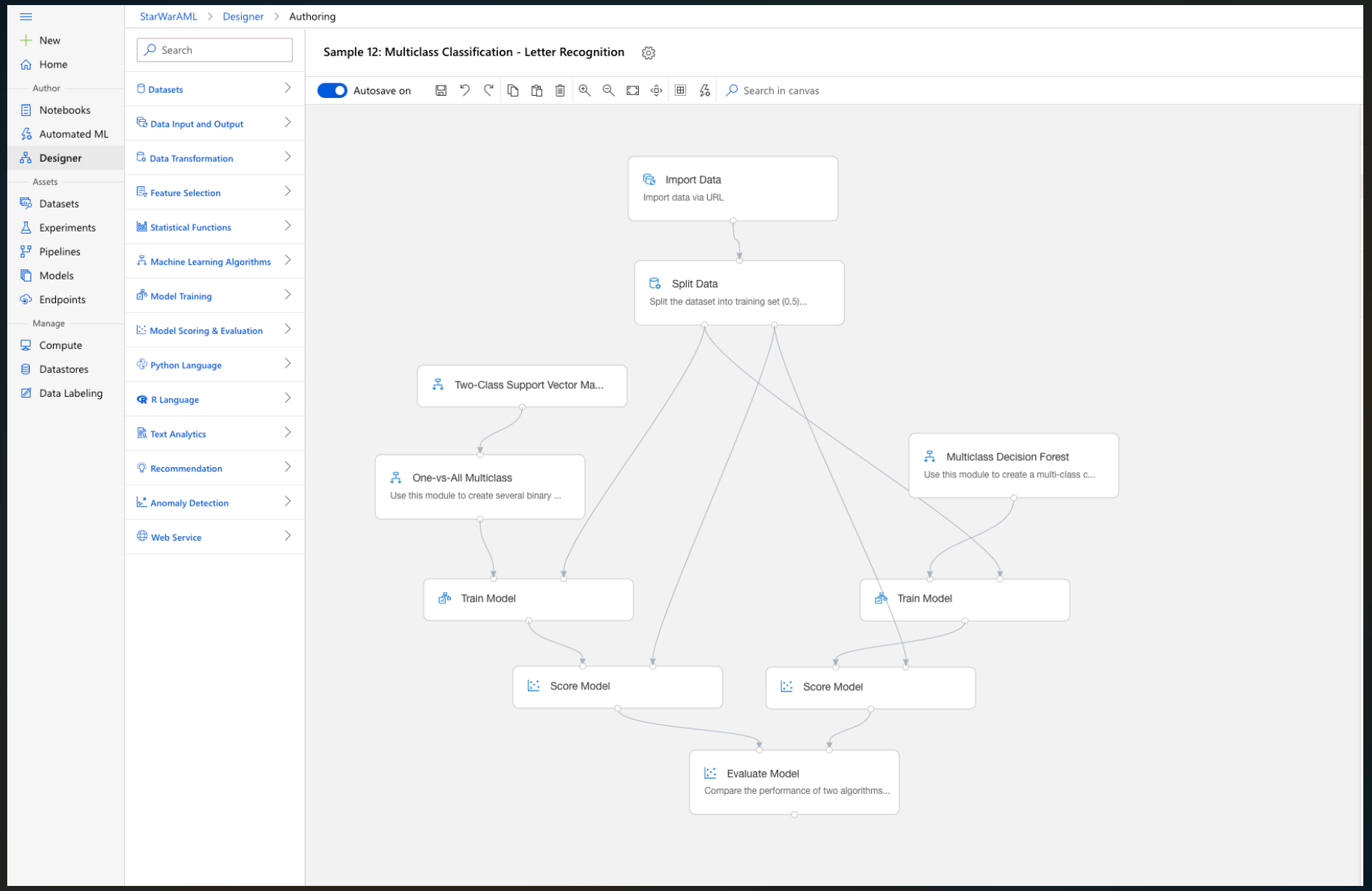


Model Registration

Azure Machine Learning Automated ML



Azure Machine Learning Designer



DevOps 过程

1 plan

2 Develop + Test

3 Release

4 Monitor + Learn

Development

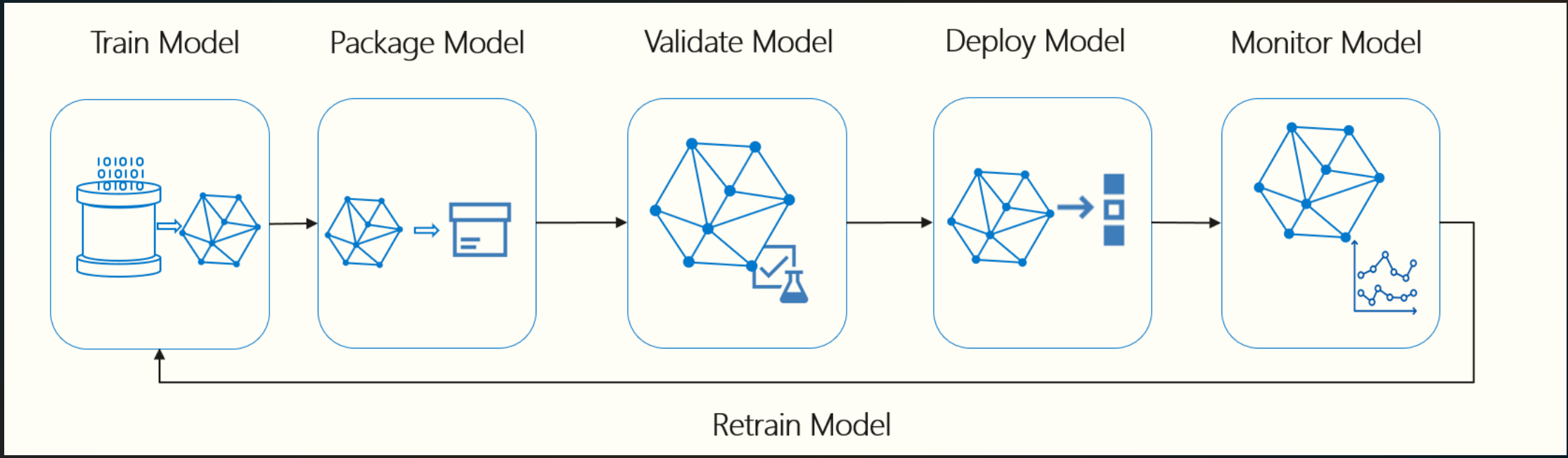
Production

DevOps Good Practices

- Source Control
- Continuous Integration
- Continuous Delivery

MLOps

- MLOps are communications between data scientists and operations or production teams. It is essentially deep collaboration, designed to eliminate waste through machine learning, automate as much as possible, and produce richer and more consistent insights.



Source Control

- Code and comments only (not Jupyter output)
- Plus every part of the pipeline
- And Infrastructure and dependencies
- And maybe a subset of data

Continuous Integration

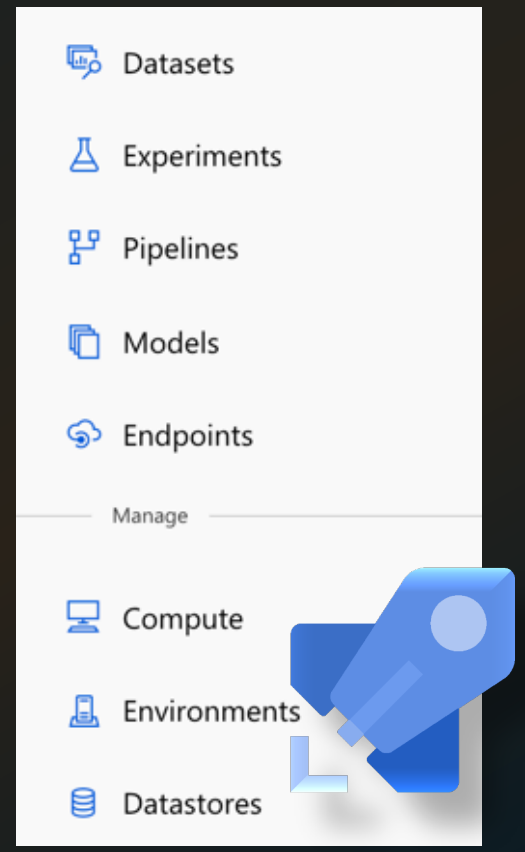
- Triggered on code change
- Refresh and execute AML Pipeline
- Code quality, linting, and unit testing
- Pull request process
- aka.ms/mitt/azuredevops

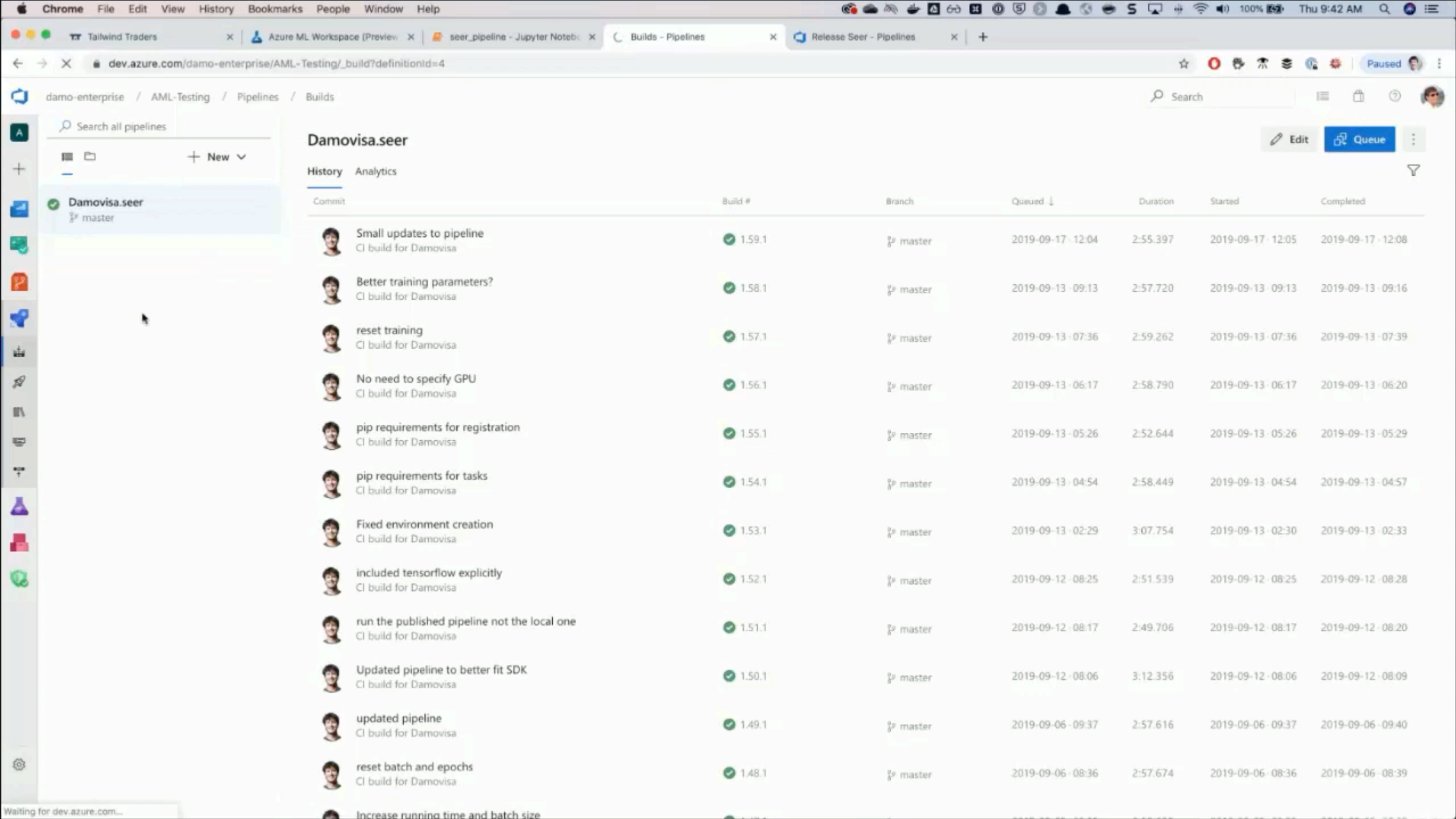
Continuous Delivery

- Trigger on model registration
- Deploy to test and staging environments
- Run integration and load tests
- Control: rollout, feature flags, A/B testing

Summary

- All code and infra in source control
- Known, shared data sources
- AML Pipelines and Azure Pipelines
- Have retraining strategies
- DevOps good practices for delivery

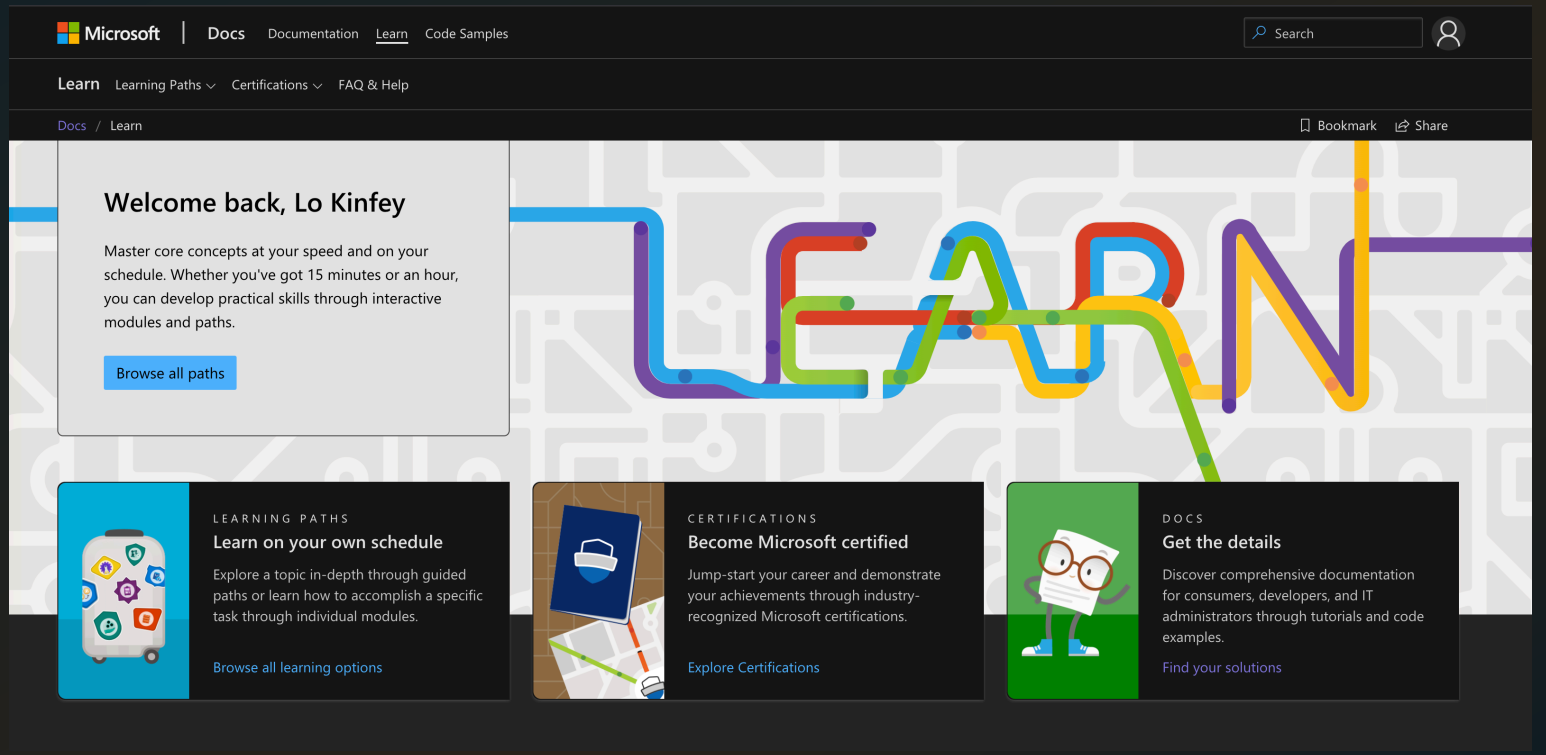



Chrome File Edit View History Bookmarks People Window Help
Thu 9:42 AM
Tailwind Traders
Azure ML Workspace (Preview
seer_pipeline - Jupyter Noteb
Builds - Pipelines
Release Seer - Pipelines
dev.azure.com/damo-enterprise/AML-Testing/_build?definitionId=4
Paused
damo-enterprise / AML-Testing / Pipelines / Builds
Search
Search all pipelines
New
Damovisa.seer
master
Damovisa.seer
Edit
Queue
History
Analytics
Commit
Build #
Branch
Queued
Duration
Started
Completed
Small updates to pipeline
CI build for Damovisa
1.59.1
master
2019-09-17 12:04
2:55.397
2019-09-17 12:05
2019-09-17 12:08
Better training parameters?
CI build for Damovisa
1.58.1
master
2019-09-13 09:13
2:57.720
2019-09-13 09:13
2019-09-13 09:16
reset training
CI build for Damovisa
1.57.1
master
2019-09-13 07:36
2:59.262
2019-09-13 07:36
2019-09-13 07:39
No need to specify GPU
CI build for Damovisa
1.56.1
master
2019-09-13 06:17
2:58.790
2019-09-13 06:17
2019-09-13 06:20
pip requirements for registration
CI build for Damovisa
1.55.1
master
2019-09-13 05:26
2:52.644
2019-09-13 05:26
2019-09-13 05:29
pip requirements for tasks
CI build for Damovisa
1.54.1
master
2019-09-13 04:54
2:58.449
2019-09-13 04:54
2019-09-13 04:57
Fixed environment creation
CI build for Damovisa
1.53.1
master
2019-09-13 02:29
3:07.754
2019-09-13 02:30
2019-09-13 02:33
included tensorflow explicitly
CI build for Damovisa
1.52.1
master
2019-09-12 08:25
2:51.539
2019-09-12 08:25
2019-09-12 08:28
run the published pipeline not the local one
CI build for Damovisa
1.51.1
master
2019-09-12 08:17
2:49.706
2019-09-12 08:17
2019-09-12 08:20
Updated pipeline to better fit SDK
CI build for Damovisa
1.50.1
master
2019-09-12 08:06
3:12.356
2019-09-12 08:06
2019-09-12 08:09
updated pipeline
CI build for Damovisa
1.49.1
master
2019-09-06 09:37
2:57.616
2019-09-06 09:37
2019-09-06 09:40
reset batch and epochs
CI build for Damovisa
1.48.1
master
2019-09-06 08:36
2:57.674
2019-09-06 08:36
2019-09-06 08:39
Increase running time and batch size
Waiting for dev.azure.com...

Some Tips for Learning Artificial Intelligence

/MS Learn

- Complete interactive learning exercises, watch videos, and practice and apply new skills learned.



Learning Path

Build AI solutions with Azure Machine Learning

https://aka.ms/pyhk/AIML

Introduction to machine learning with Python and Azure Notebooks

https://aka.ms/pyhk/MLpython

感謝