

Nompumelelo Ngwenya

Deploying a machine learning model on Azure AI Platform.

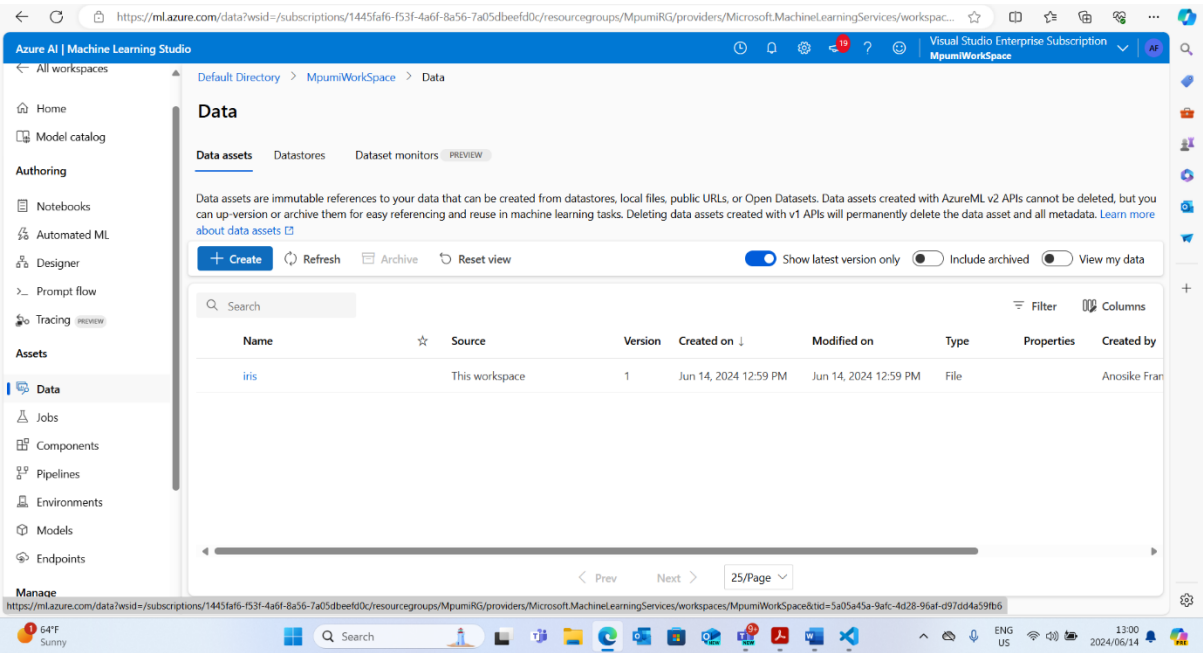
Introduction and Overview

This document serves as a guide for deploying a logistic regression model using the famous Iris dataset on the Azure AI Platform. Its primary purpose is to provide a comprehensive overview of the steps involved in deploying a machine learning model, specifically focusing on logistic regression.

Data Source: [Iris Flower Dataset \(kaggle.com\)](https://www.kaggle.com/datasets/iris-flower-dataset)

Below is our Iris flower dataset, which consists of 150 samples of iris flowers from three different species: Setosa, Versicolor, and Virginica where each sample includes four features: sepal length, sepal width, petal length, and petal width.

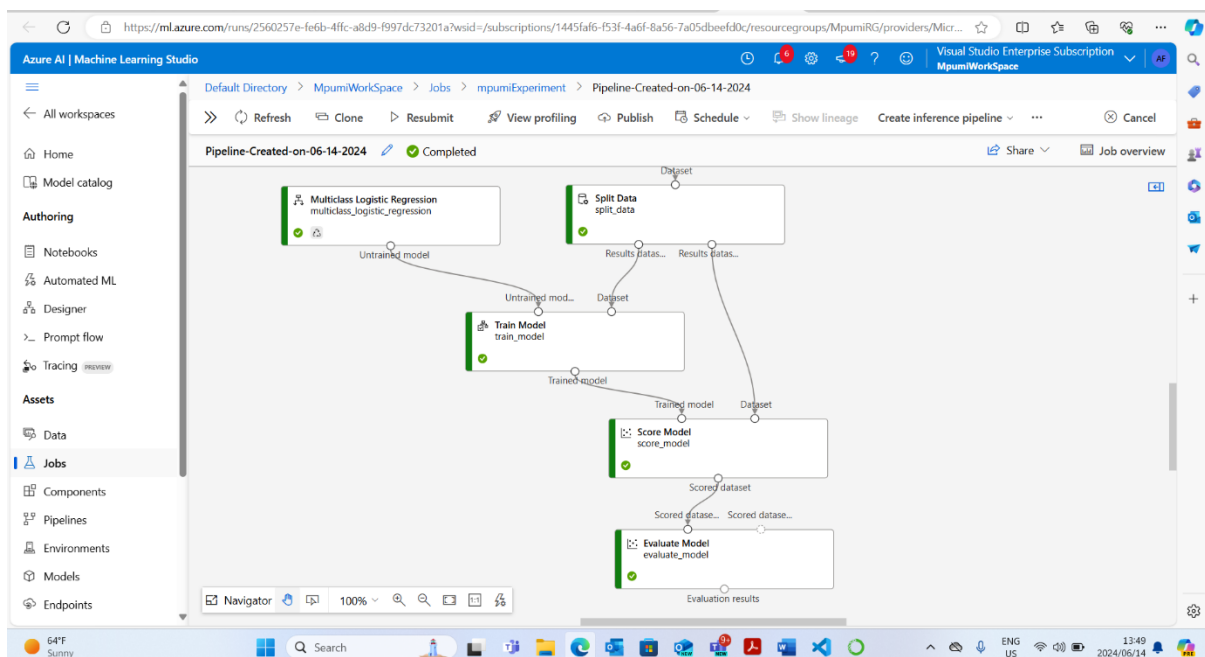
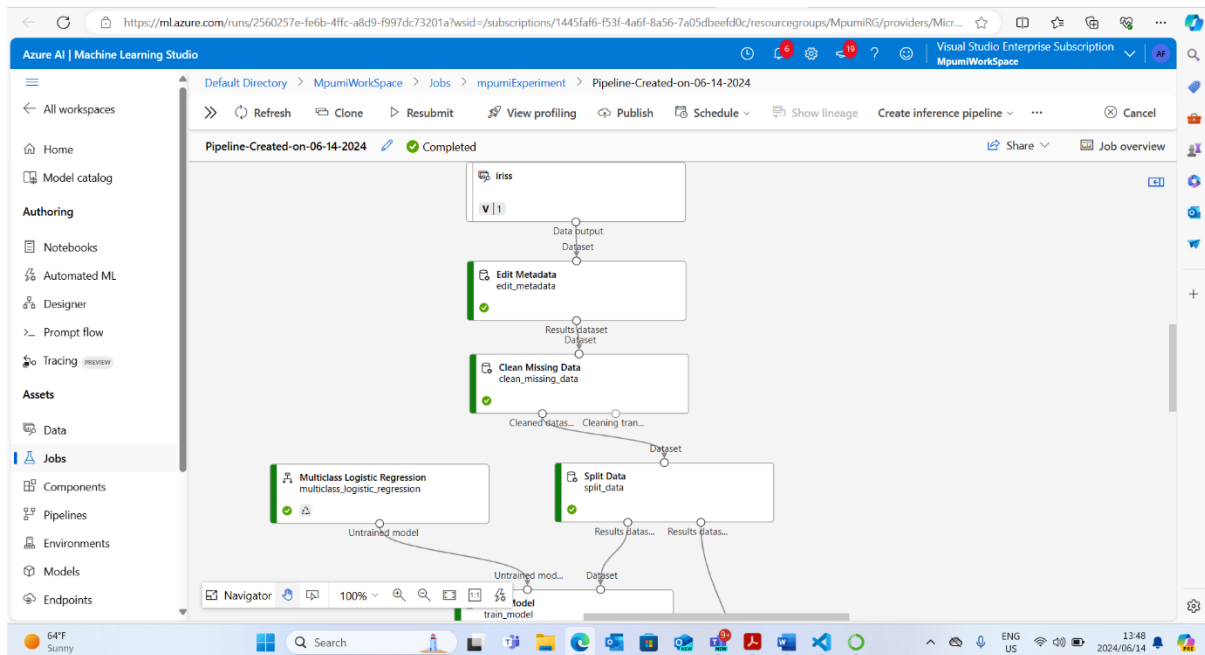
Our machine learning model should be able to classify whether an iris flower falls under Setosa, Versicolor or Virginica given the four features.



Data Preprocessing:

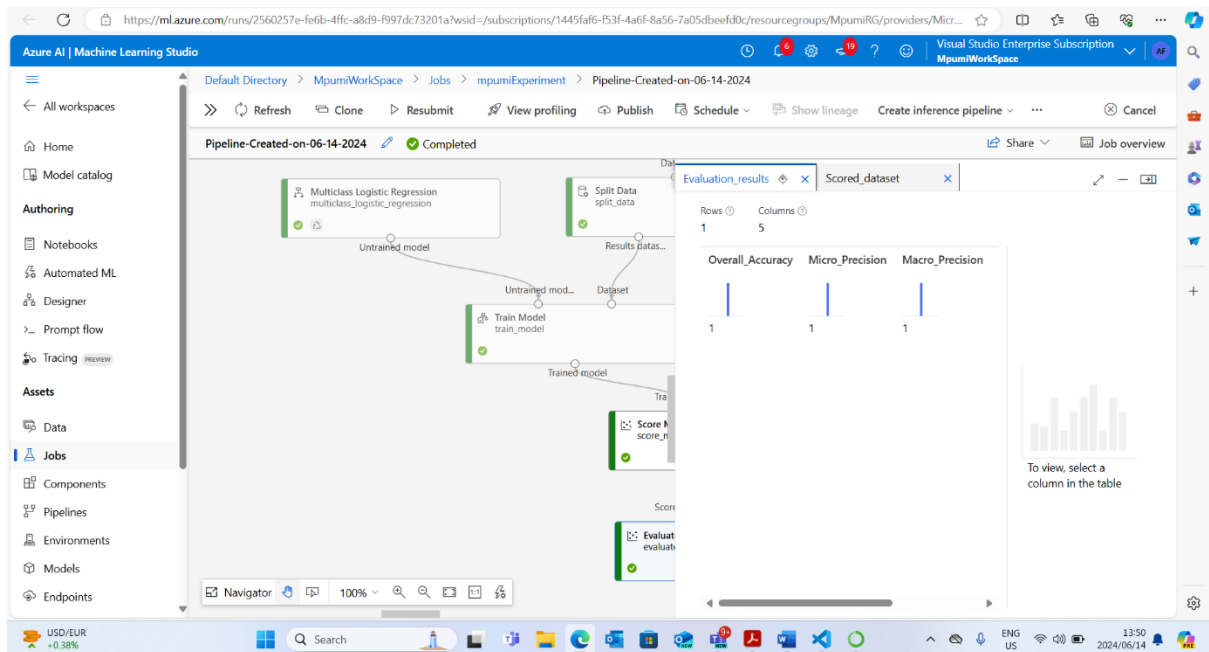
This step involves preparing the dataset for model training, including handling missing values, scaling features if necessary, and splitting the dataset into training and testing sets.

Model trained using logistic regression. I used logistic regression because it is a classification algorithm and here we want to be able to classify on whether which specie an Iris flower falls under.



## Model Evaluation:

Once trained, the model's performance is evaluated using metrics such as accuracy, precision, recall, and F1-score on the test dataset to assess its effectiveness in making predictions.



https://tsemoeti242.westus3.instances.azureml.ms/lab/tree/Untitled.ipynb

File Edit View Run Kernel Tabs Settings Help

Hi There CPU: 0% Mem: 1.17 / 13.65 GB

Filter files by name

Name Last Modified

- AutoML\_16... 25 minutes ago
- AutoML\_37... 2 hours ago
- AutoML\_b6... 2 hours ago
- explanation 2 hours ago
- inference an hour ago
- Logs 2 hours ago
- Users 2 hours ago
- automl-chil... 8 minutes ago
- automl.log 25 minutes ago
- iris.csv 2 hours ago
- myenv.yml 6 minutes ago
- Untitled.ipynb... 2 minutes ago

```
[8]: experiment = Experiment(ws, 'IrisExpe')
[9]: run = experiment.submit(automl_config, show_outputs=True)
```

WARNING:root:Received unrecognized parameter debug\_logs  
No run\_configuration provided, running on local with default configuration  
Running in the active local environment.

Experiment	Id	Type	Status	Details Page	Docs Page
IrisExpe	AutoML_16c6bae-c232-4c73-90f0-42be2947b4c	automl	Preparing	<a href="#">Link to Azure Machine Learning studio</a>	<a href="#">Link to Documentation</a>

Current status: DatasetEvaluation. Gathering dataset statistics.  
Current status: FeaturesGeneration. Generating features for the dataset.  
Current status: DatasetFeaturization. Beginning to fit featureizers and featurize the dataset.  
Current status: DatasetFeaturizationCompleted. Completed fit featureizers and featurizing the dataset.  
Current status: DatasetCrossValidationSplit. Generating individually featurized CV splits.

2024/06/14 15:48:53 WARNING mlflow.sklearn: Model was missing function: predict. Not logging python\_function flavor!

=====

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

=====

Simple 0 1 Python 3.8 - AzureML | Idle Mem: 1.16 / 13.65 GB Mode: Command Ln 1, Col 1 Untitled.ipynb

62°F Partly cloudy Search 1813 2024/06/14

https://tsemoeti242.westus3.instances.azureml.ms/lab/tree/Untitled.ipynb

File Edit View Run Kernel Tabs Settings Help

Hi There CPU: 0% Mem: 1.18 / 13.65 GB

Filter files by name

Name Last Modified

- AutoML\_16... 25 minutes ago
- AutoML\_37... 2 hours ago
- AutoML\_b6... 2 hours ago
- explanation 2 hours ago
- inference an hour ago
- Logs 2 hours ago
- Users 2 hours ago
- automl-chil... 9 minutes ago
- automl.log 25 minutes ago
- iris.csv 2 hours ago
- myenv.yml 6 minutes ago
- Untitled.ipynb... 2 minutes ago

```
ITER: 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.
```

ITER	PIPELINE	DURATION	METRIC	BEST
0	MaxAbsScaler LightGBM	0:00:43	1.0000	1.0000
1	MaxAbsScaler XGBoostClassifier	0:01:09	0.9901	1.0000
2	MaxAbsScaler ExtremeRandomTrees	0:00:44	1.0000	1.0000
3	MaxAbsScaler RandomForest	0:00:43	1.0000	1.0000
4	StandardScalerWrapper LightGBM	0:00:44	1.0000	1.0000
5	StandardScalerWrapper KNN	0:00:42	1.0000	1.0000
6	SparseNormalizer XGBoostClassifier	0:01:01	0.9856	1.0000
7	SparseNormalizer RandomForest	0:00:45	0.9859	1.0000
8	RobustScaler KNN	0:00:44	1.0000	1.0000
9	MinMaxScaler RandomForest	0:00:45	1.0000	1.0000
10	StandardScalerWrapper LogisticRegression	0:00:44	1.0000	1.0000
11	StandardScalerWrapper SVM	0:00:43	1.0000	1.0000
12	StandardScalerWrapper XGBoostClassifier	0:01:01	1.0000	1.0000
13	SparseNormalizer KNN	0:00:45	0.9918	1.0000
14	RobustScaler ExtremeRandomTrees	0:00:44	1.0000	1.0000
15	SparseNormalizer XGBoostClassifier	0:01:01	0.9719	1.0000
16	MinMaxScaler ExtremeRandomTrees	0:00:44	1.0000	1.0000
17	MinMaxScaler ExtremeRandomTrees	0:00:44	1.0000	1.0000
18	VotingEnsemble	0:00:46	1.0000	1.0000
19	StackEnsemble	0:00:45	1.0000	1.0000

Stopping criteria reached at iteration 20. Ending experiment.

WARNING:root:Received unrecognized parameter debug\_logs  
WARNING:root:Received unrecognized parameter debug\_logs

Current status: BestRunExplainModel. Best run model explanations started  
INFO: interpret\_community.common.explanation\_utils:Using default datastore for uploads  
Current status: ModelExplanationDataSetSetup. Model explanations data setup completed

Simple 0 1 Python 3.8 - AzureML | Idle Mem: 1.16 / 13.65 GB Mode: Command Ln 1, Col 1 Untitled.ipynb

62°F Partly cloudy Search 1813 2024/06/14

https://tsemoeti242.westus3.instances.azureml.ms/lab/tree/Untitled.ipynb

File Edit View Run Kernel Tabs Settings Help

Hi There CPU: 0% Mem: 1.17 / 13.65 GB

Filter files by name

Untitled.ipynb

Python 3.8 - AzureML

```
Stopping criteria reached at iteration 20. Ending experiment.
WARNING:root:Received unrecognized parameter debug_logs
WARNING:root:Received unrecognized parameter debug_logs
Current status: BestRunExplainModel. Best run model explanations started
INFO:interpret_community.common.explanation_utils:Using default datastore for uploads
Current status: ModelExplanationDataSetSetup. Model explanations data setup completed
Current status: PickSurrogateModel. Choosing LightGBM as the surrogate model for explanations
Current status: EngineeredFeatureExplanations. Computation of engineered features started
Current status: EngineeredFeatureExplanations. Computation of engineered features completed
Current status: RawFeaturesExplanations. Computation of raw features started
Current status: RawFeaturesExplanations. Computation of raw features completed
Current status: BestRunExplainModel. Best run model explanations completed
```

[10]: best\_run, model = run.get\_output()

WARNING:root:Received unrecognized parameter debug\_logs  
WARNING:root:Received unrecognized parameter debug\_logs  
WARNING:root:Received unrecognized parameter debug\_logs

[11]: RunDetails(run).show()

AutoML\_16c6bae-c232-4c73-90f0-42be29f47b4c:  
Status: Completed

Status -

Iteration Pipeline

Iteration	Pipeline	Iteration metric	Best metric	Status	Duration	Started	Run Id
0							
2							
4							
6							
8							
10							
12							
14							
16							
18							

Simple 0 1 Python 3.8 - AzureML Idle Mem: 1.16 / 13.65 GB

Saving started Mode: Command Ln 1, Col 1

62°F Partly cloudy

https://tsemoeti242.westus3.instances.azureml.ms/lab/tree/Untitled.ipynb

File Edit View Run Kernel Tabs Settings Help

Hi There CPU: 0% Mem: 1.17 / 13.65 GB

Filter files by name

Untitled.ipynb

Python 3.8 - AzureML

Iteration	Pipeline	Iteration metric	Best metric	Status	Duration	Started	Run Id
4	StandardScalerWrapper, LightGBM	1	1	Completed	0:00:43	Jun 14, 2024 5:52 PM	
5	StandardScalerWrapper, KNN	1	1	Completed	0:00:42	Jun 14, 2024 5:53 PM	

Pages: 1 2 3 4 Next Last 5 per page

AUC\_weighted

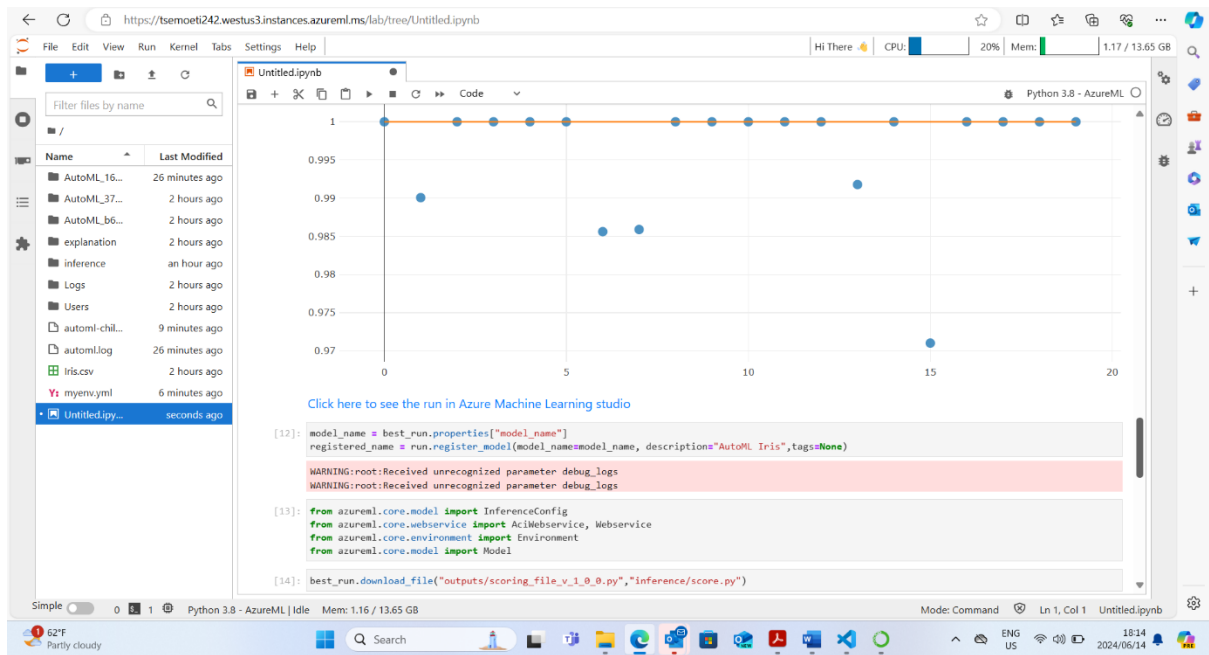
AutoML Run with metric : AUC\_weighted

Click here to see the run in Azure Machine Learning studio

Simple 0 1 Python 3.8 - AzureML Idle Mem: 1.16 / 13.65 GB

Mode: Command Ln 1, Col 1

62°F Partly cloudy



## Deployment:

After training of the model was done, I then went to deploy the model, I started with Registering the model and then deployed it so that the developed model is available in a production environment, enabling it to generate real-world predictions.

The screenshot shows the Azure ML Studio interface with a Jupyter notebook titled 'Untitled.ipynb'. The notebook contains the following Python code:

```
[12]: model_name = best_run.properties["model_name"]
      registered_name = run.register_model(model_name=model_name, description="AutoML Iris", tags=None)

      WARNING:root:Received unrecognized parameter debug_logs
      WARNING:root:Received unrecognized parameter debug_logs

[13]: from azureml.core.model import InferenceConfig
      from azureml.webservice import AciWebservice, Webservice
      from azureml.core.environment import Environment
      from azureml.core.model import Model

[14]: best_run.download_file("outputs/scoring_file_v1_0_0.py", "inference/score.py")

[16]: from azureml.automl.core.shared import constants

      best_run.download_file(constants.CONDA_ENV_FILE_PATH, "myenv.yml")
      env = Environment.from_conda_specification(name="myenv", file_path="myenv.yml")

      inference_config = InferenceConfig(entry_script="inference/score.py", environment=env)
      aci_config = AciWebservice.deploy_configuration(cpu_cores=1, memory_gb=1, description="Iris Classification")
      service = Model.deploy(ws, "automl-iriss", [registered_name], inference_config, aci_config)

      service.wait_for_deployment(True)

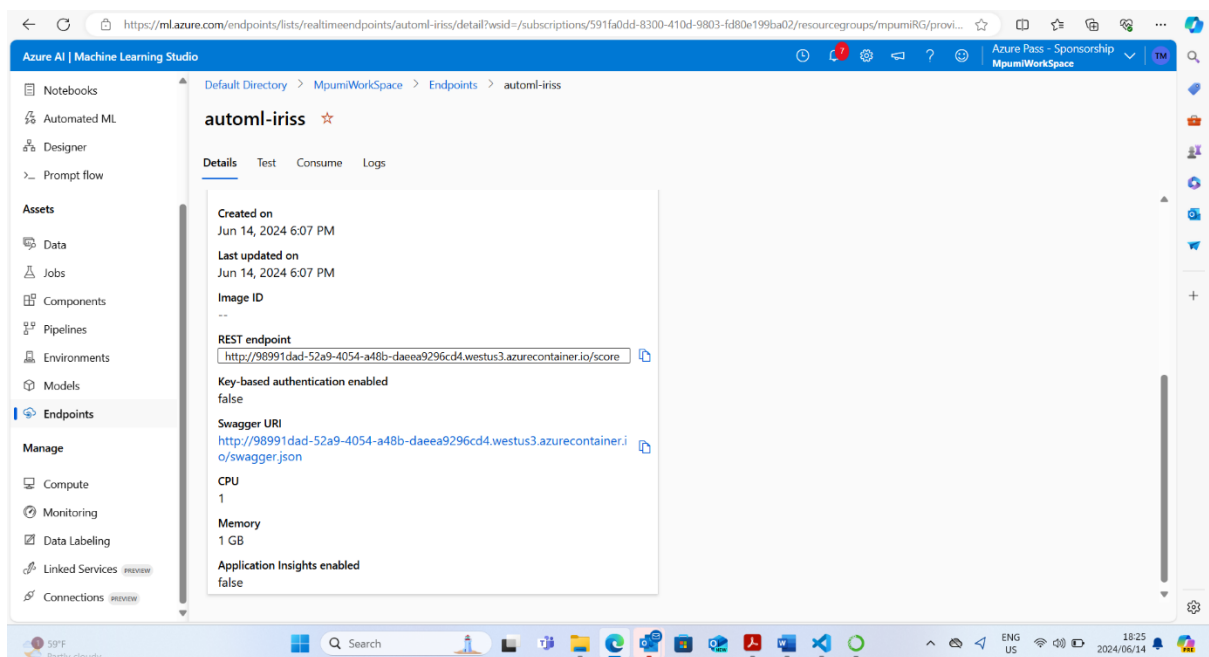
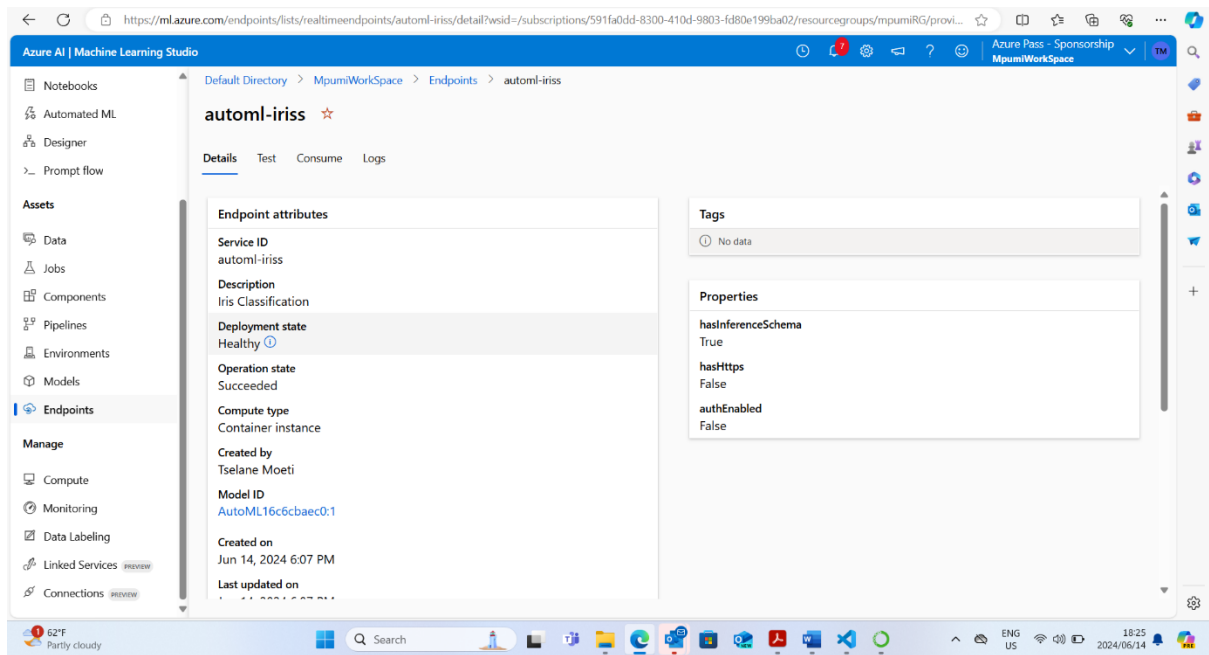
      Tips: You can try get_logs(): https://aka.ms/debugimage#dockerlog or local deployment: https://aka.ms/debugimage#debug-locally to debug if deployment takes longer than 10 minutes.
      Running
      2024-06-14 16:07:57+00:00 Creating Container Registry if not exists.
      2024-06-14 16:07:58+00:00 Use the existing image.
      2024-06-14 16:07:58+00:00 Generating deployment configuration.
      2024-06-14 16:07:59+00:00 Submitting deployment to compute.
      2024-06-14 16:08:08+00:00 Checking the status of deployment automl-iriss..
      2024-06-14 16:10:06+00:00 Checking the status of inference endpoint automl-iriss.
      Succeeded
      ACI service creation operation finished, operation "Succeeded"
```

The predicted results:

Endpoint link: <http://98991dad-52a9-4054-a48b-daeea9296cd4.westus3.azurecontainer.io/score>

The screenshot shows the Azure AI Machine Learning Studio Jobs page. The table displays the following experiments:

Experiment	Latest job	Last submitted	Created
IrisExpe	eager_oregano_0wh41zhq	Jun 14, 2024 5:48 PM	Jun 14, 2024 4:17 PM
prepare_image	magenta_spoon_J7kv41hk	Jun 14, 2024 5:12 PM	Jun 14, 2024 5:12 PM



## Monitoring and Maintenance:

Once deployed, the model's performance needs to be monitored regularly to ensure that it continues to provide accurate predictions over time. This may involve updating the model with new data or retraining it periodically to maintain its accuracy.

Deployment Environment used for the deployment of the model:



## Frameworks and Libraries:

NumPy: For numerical computations and array manipulations.

Pandas: For data manipulation and analysis, particularly useful for handling datasets like the Iris dataset.

Scikit-learn: For machine learning algorithms and model training. It includes logistic regression and utilities for model evaluation.

## Development Tools:

IDEs: Such as PyCharm, VS Code, or Jupyter for coding and testing.

Version Control: Git for managing code versions.

Package Management: pip or conda for managing Python packages and dependencies.

## Security Considerations:

Access Control: Restrict access to the deployed model and its endpoints. Implement role-based access control (RBAC) to ensure only authorized personnel can interact with the model.

Encryption: Use encryption mechanisms (e.g., HTTPS/TLS) to secure data transmission between clients and the deployed model, preventing eavesdropping and data tampering.

Input Validation: Validate input data to prevent injection attacks and ensure that only expected and sanitized data is processed by the model.