Nompumelelo Ngwenya

Deploying a machine learning model on Azure AI Platform.

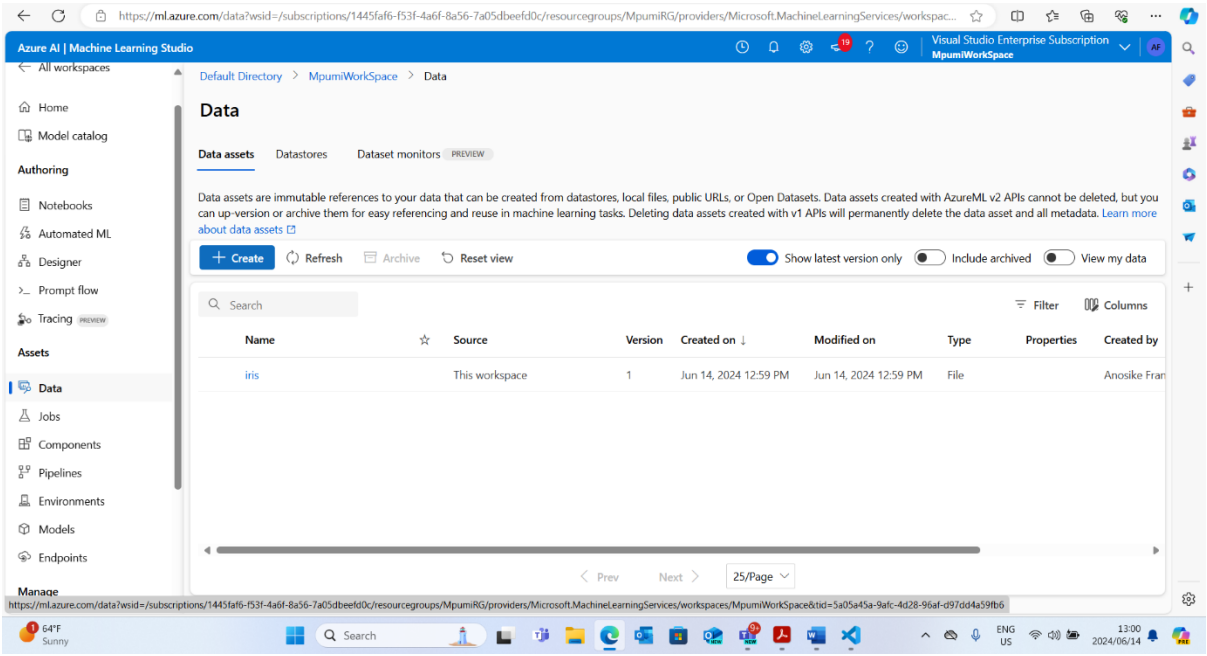
Introduction and Overview

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)]

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.

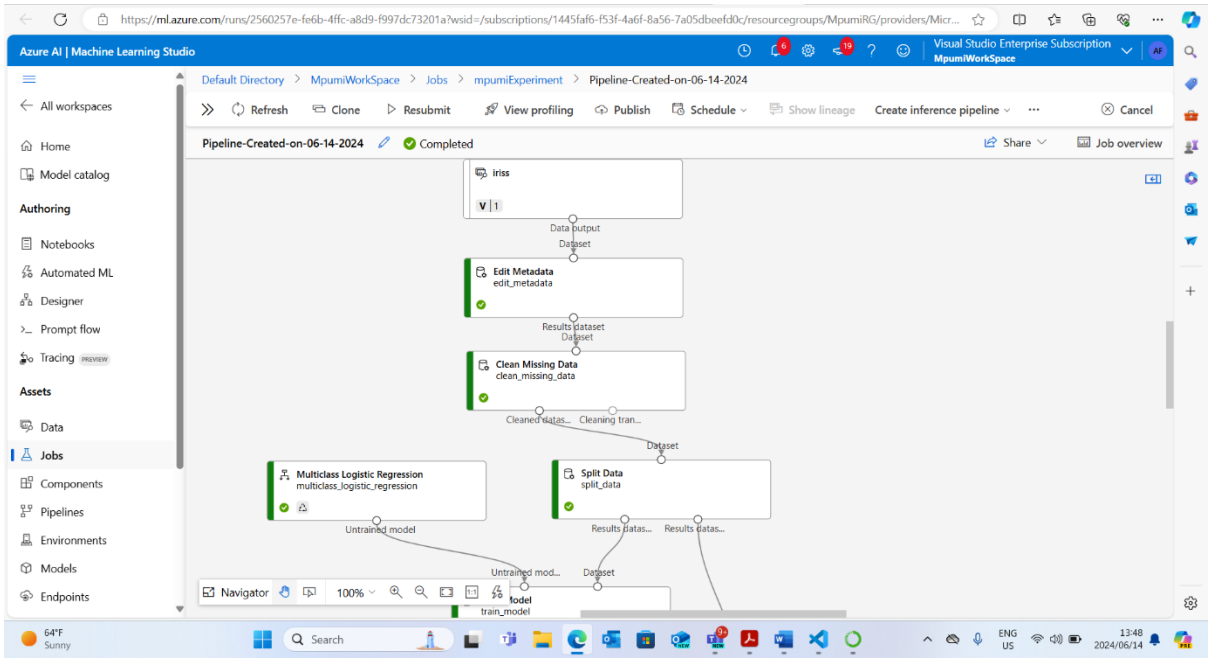
Our machine learning model should be able to classify whether an iris flower falls under Setosa, Versicolor or Virginica given the four features.

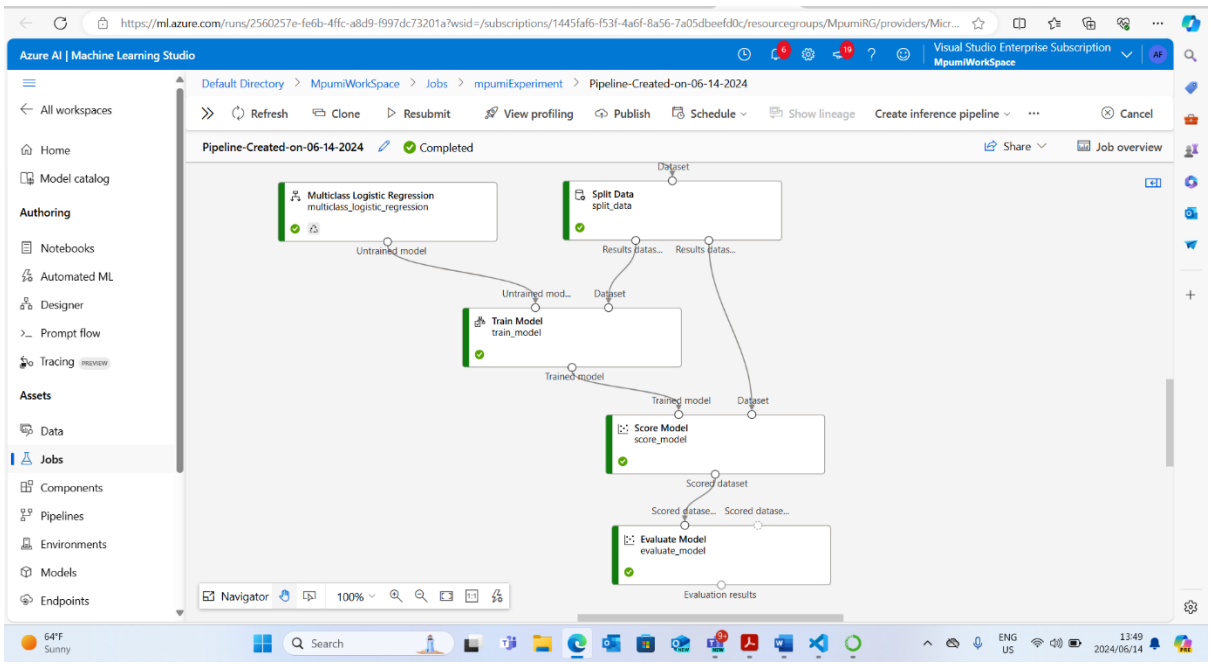


Data Preprocessing:

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.

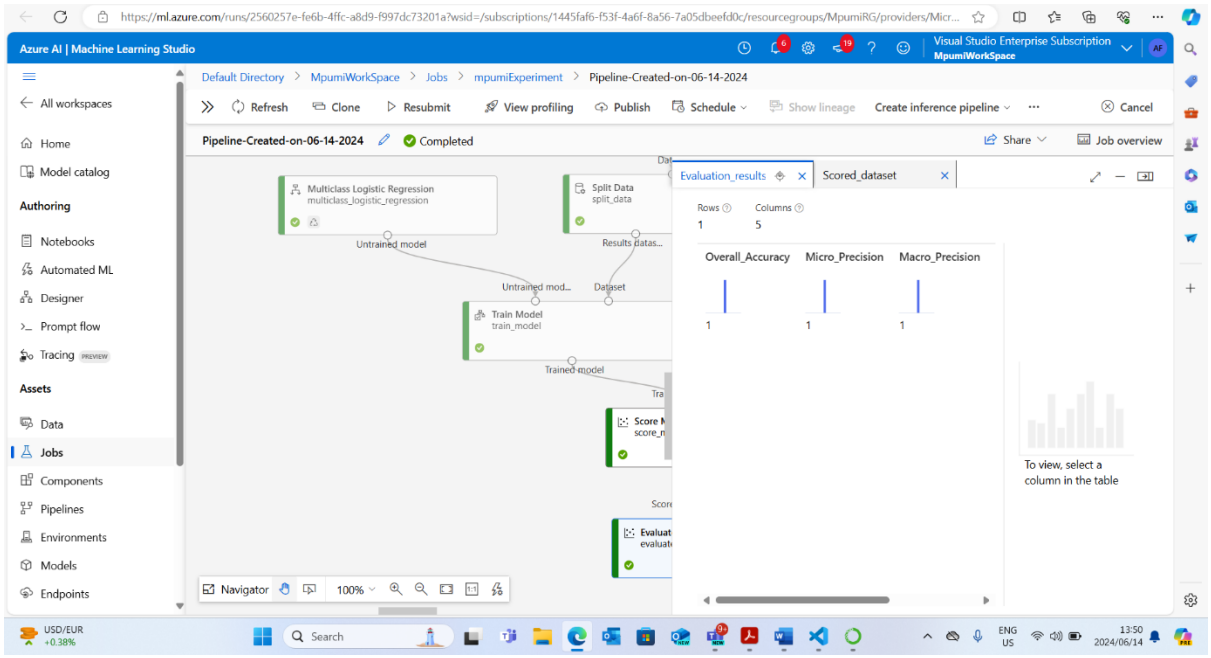
Model trained using logistic regression. I used logistic regression because it is a classification algorithm and here we want to be able to classify on whether which specie an Iris flower falls under.

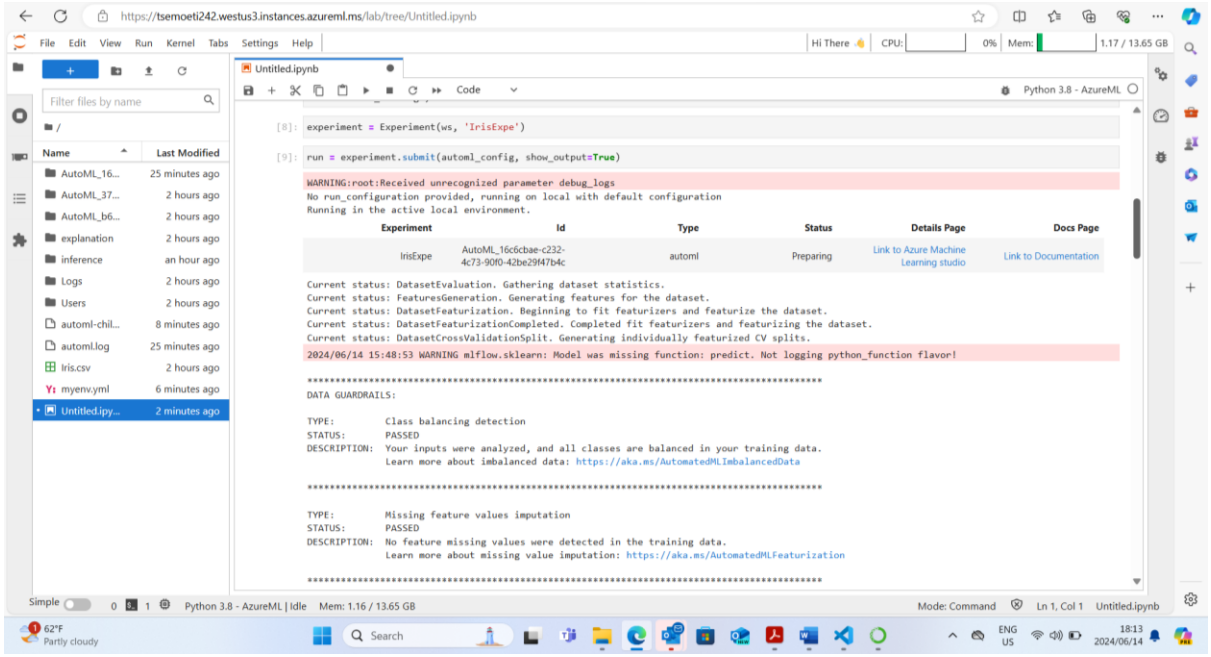


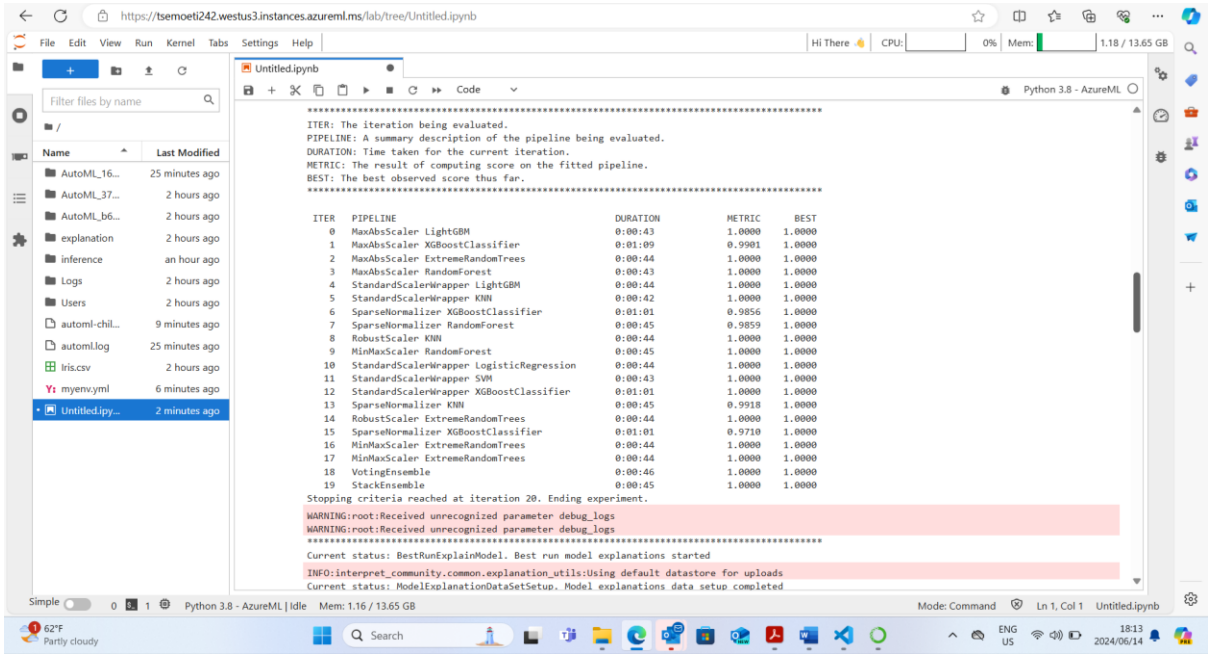


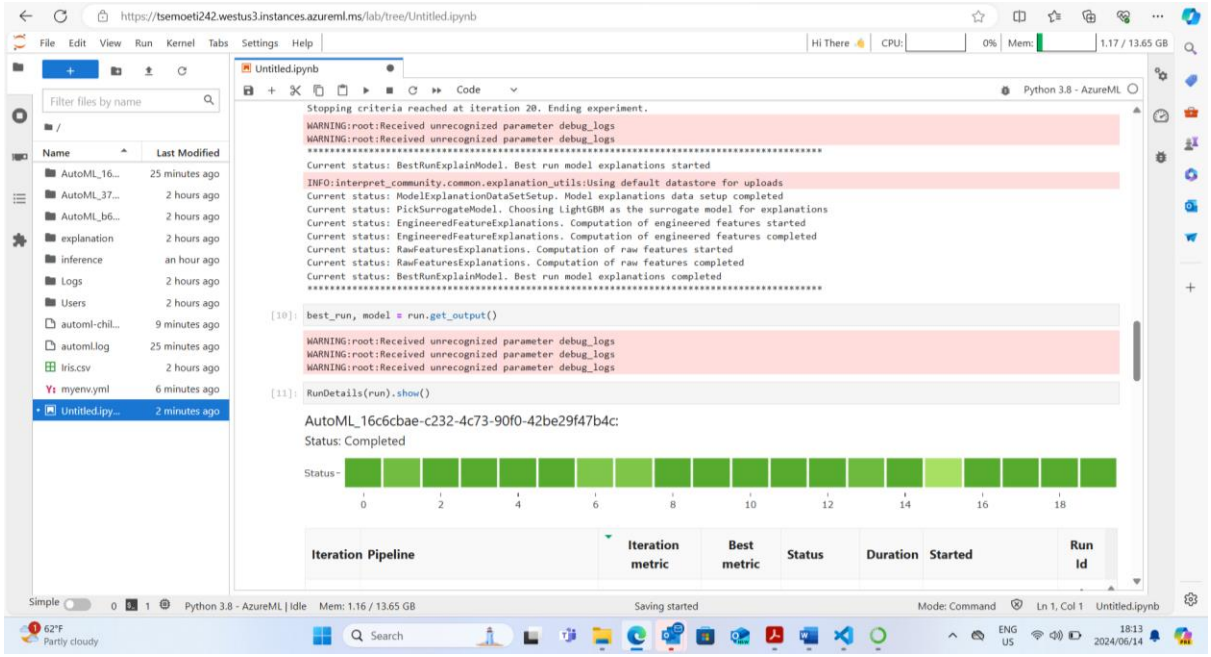
Model Evaluation:

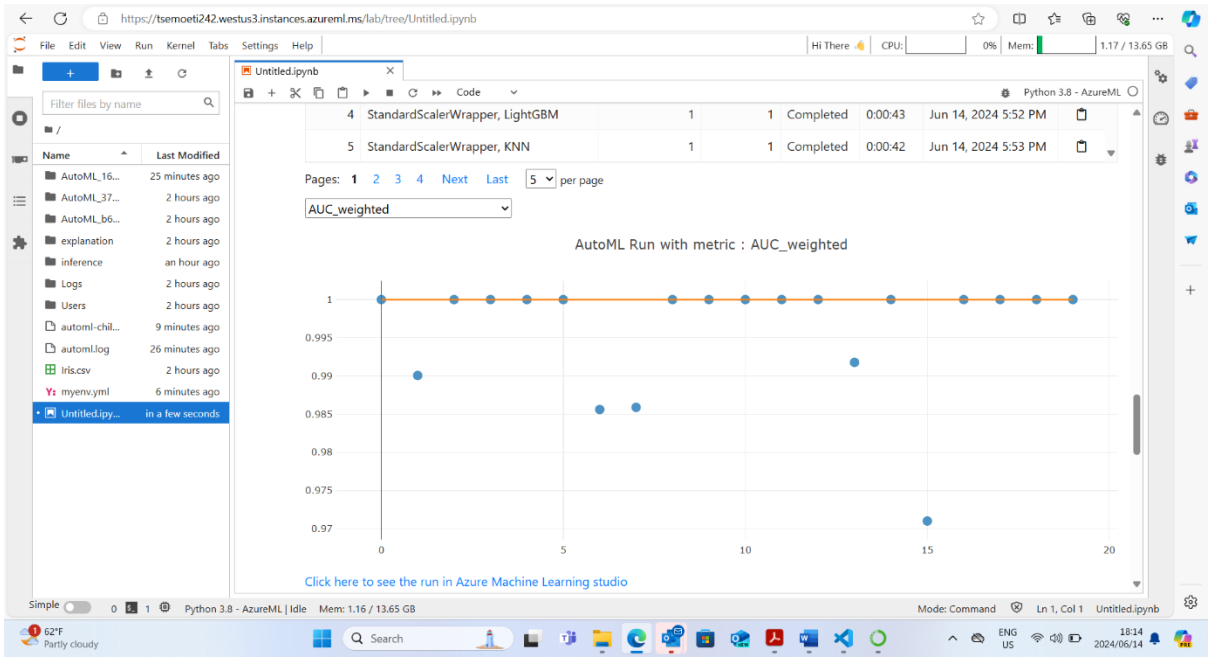
Once trained, the model's performance is evaluated using metrics such as accuracy, precision, recall, and F1-score on the test dataset to assess its effectiveness in making predictions.

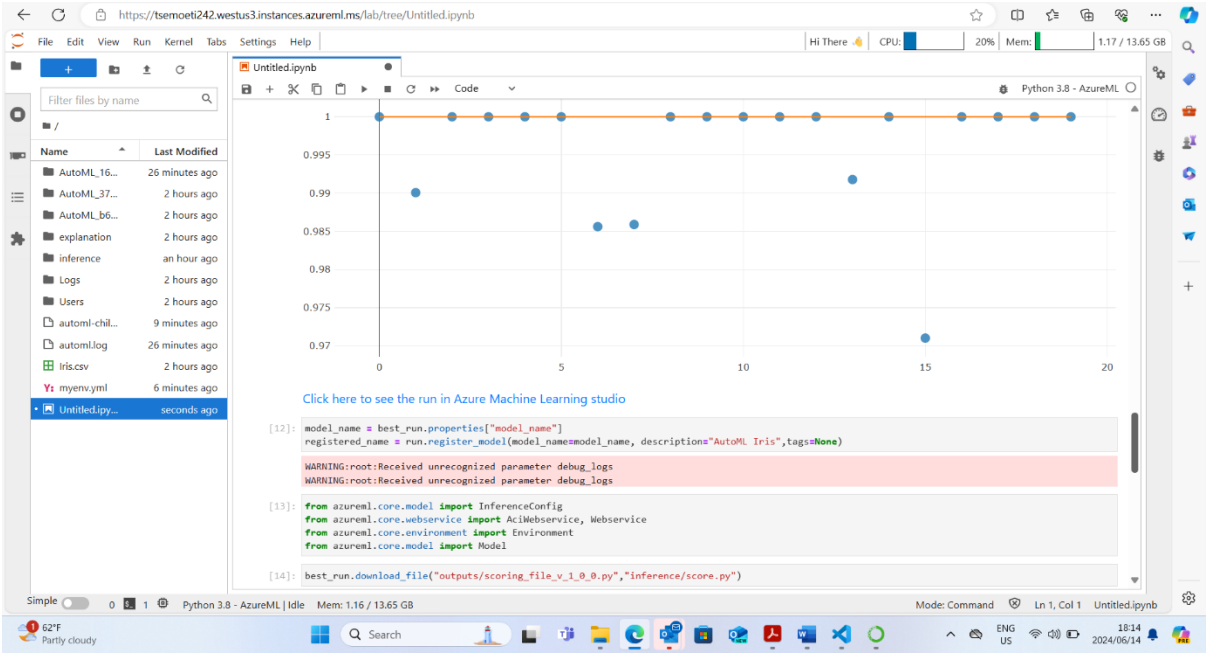



https://tsemoeti242.westus3.instances.azureml.ms/lab/tree/Untitled.ipynb
File Edit View Run Kernel Tabs Settings Help
Hi There
CPU: 0%
Mem: 1.17 / 13.65 GB
Untitled.ipynb
Code
Python 3.8 - AzureML
Filter files by name
Name
Last Modified
AutoML_16... 25 minutes ago
AutoML_37... 2 hours ago
AutoML_b6... 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.ipy... 2 minutes ago
[8]: experiment = Experiment(ws, 'IrisExpe')
[9]: run = experiment.submit(automl_config, show_output=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_16c6cbae-c232-4c73-90f0-42be29f47b4c
automl
Preparing
Link to Azure Machine Learning studio
Link to Documentation
Current status: DatasetEvaluation. Gathering dataset statistics.
Current status: FeaturesGeneration. Generating features for the dataset.
Current status: DatasetFeaturization. Beginning to fit featurizers and featurize the dataset.
Current status: DatasetFeaturizationCompleted. Completed fit featurizers 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
Python 3.8 - AzureML | Idle
Mem: 1.16 / 13.65 GB
Mode: Command
Ln 1, Col 1
Untitled.ipynb
62°F
Partly cloudy
Search
ENG
US
18:13
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
Untitled.ipynb
Code
Python 3.8 - AzureML
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.ipy... 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 0.9710 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
Python 3.8 - AzureML | Idle
Mem: 1.16 / 13.65 GB
Mode: Command
Ln 1, Col 1
Untitled.ipynb
62°F
Partly cloudy
Search
ENG
US
18:13
2024/06/14

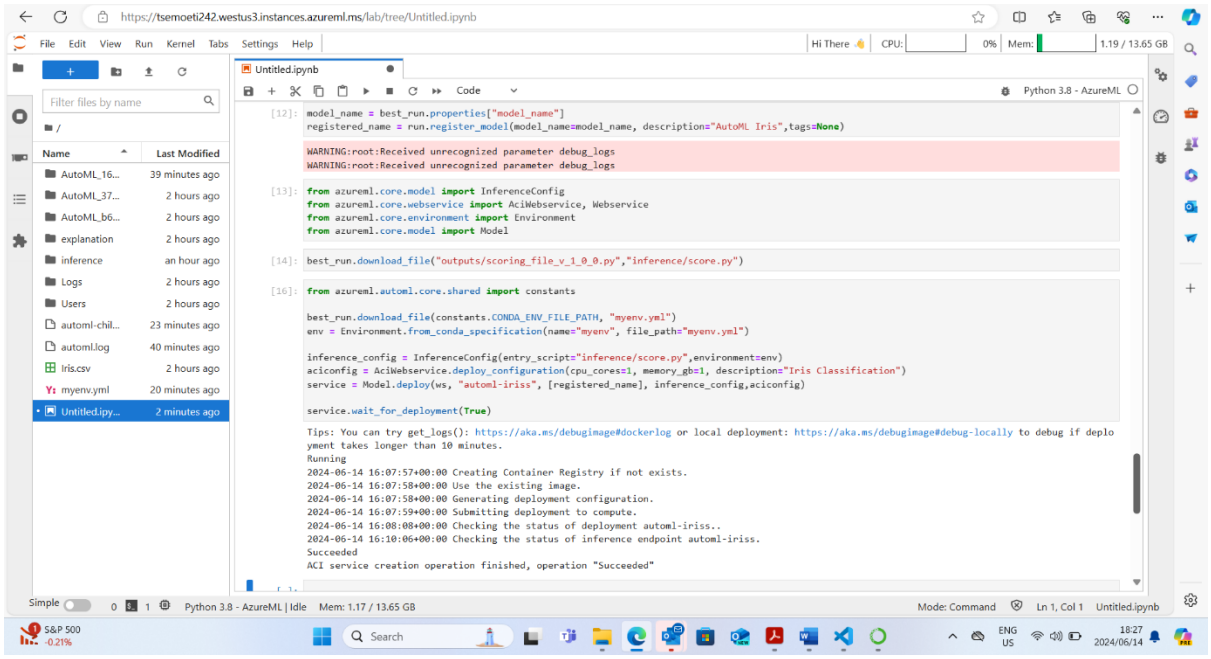

https://tsemoeti242.westus3.instances.azureml.ms/lab/tree/Untitled.ipynb
File Edit View Run Kernel Tabs Settings Help
Hi There
CPU:
0%
Mem:
1.17 / 13.65 GB
Untitled.ipynb
Code
Python 3.8 - AzureML
Filter files by name
Name
Last Modified
AutoML_16... 25 minutes ago
AutoML_37... 2 hours ago
AutoML_b6... 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.ipy... 2 minutes ago
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_16c6cbae-c232-4c73-90f0-42be29f47b4c:
Status: Completed
Status
0 2 4 6 8 10 12 14 16 18
Iteration Pipeline
Iteration metric
Best metric
Status
Duration
Started
Run Id
Simple
Python 3.8 - AzureML | Idle
Mem: 1.16 / 13.65 GB
Saving started
Mode: Command
Ln 1, Col 1
Untitled.ipynb
62°F
Partly cloudy
Search
ENG
US
18:13
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
Untitled.ipynb
Code
Python 3.8 - AzureML
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 26 minutes ago
Iris.csv 2 hours ago
myenv.yml 6 minutes ago
Untitled.ipy... in a few seconds
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
1
0.995
0.99
0.985
0.98
0.975
0.97
0 5 10 15 20
Click here to see the run in Azure Machine Learning studio
Simple
Python 3.8 - AzureML | Idle
Mem: 1.16 / 13.65 GB
Mode: Command
Ln 1, Col 1
Untitled.ipynb
62°F
Partly cloudy
Search
ENG
US
18:14
2024/06/14



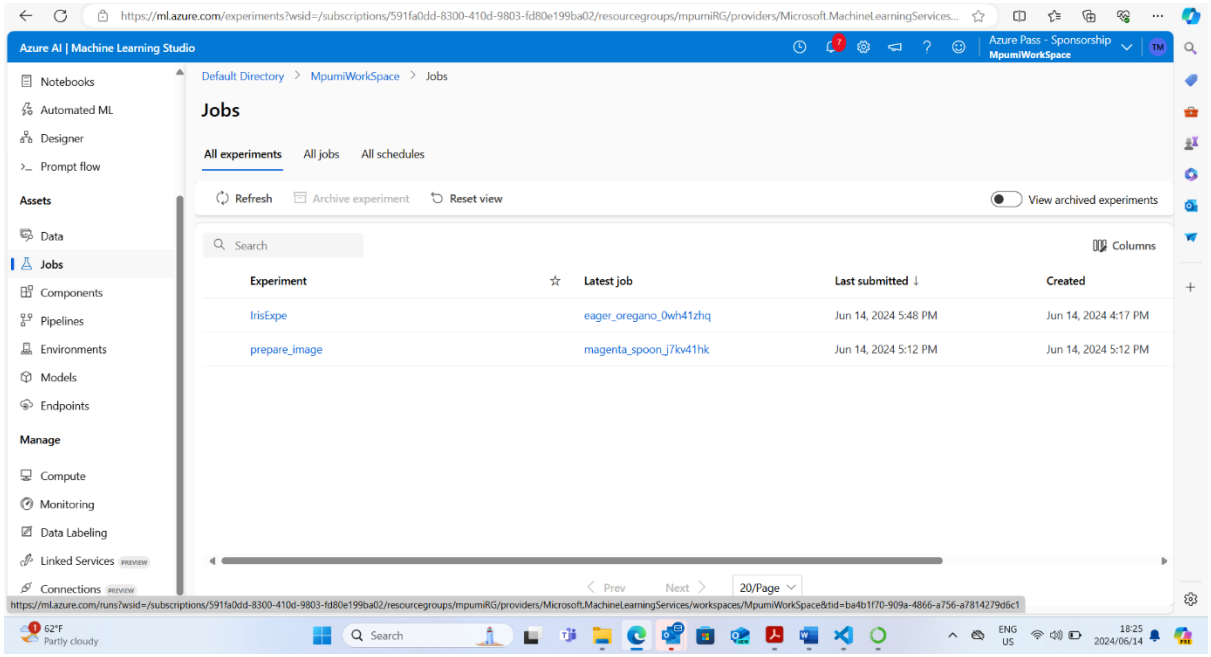
Deployment:

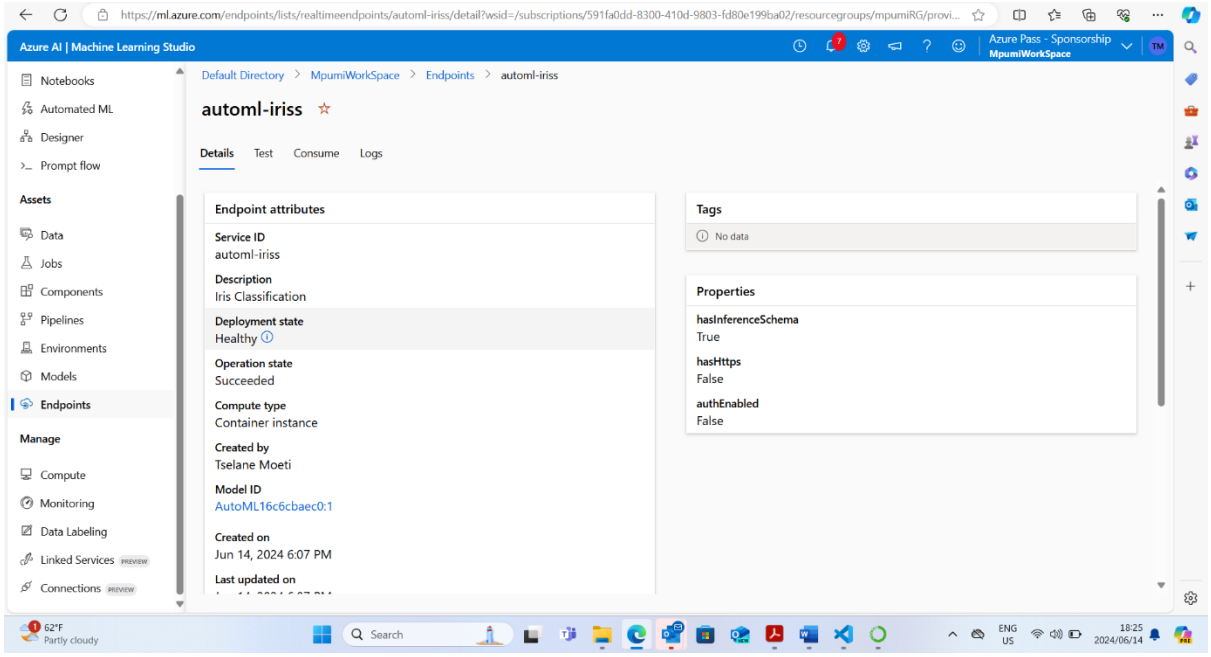
After training of the model was done, I then went to deploy the model, I started with Registering the model and then deployed it so that the developed model is available in a production environment, enabling it to generate real-world predictions.

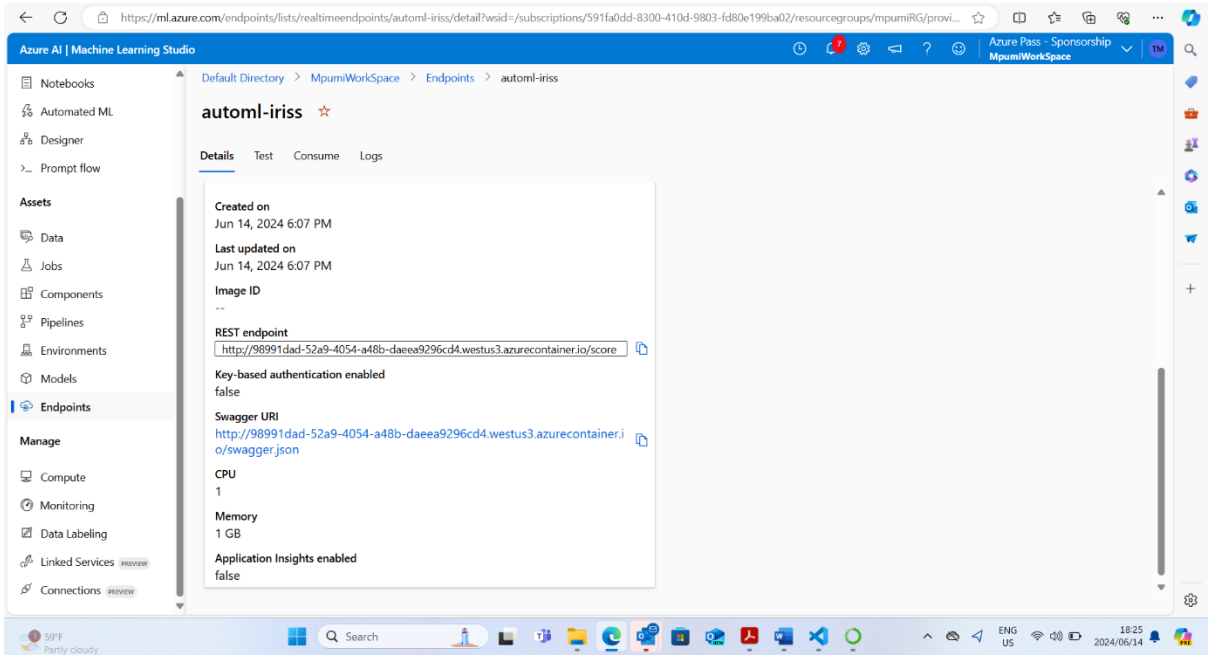


The predicted results:

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







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.