

Tutorial 7: Time Series Forecasting in AutoML

Before proceeding with the tasks ahead, make sure you have created *resource*, and *workspace* in MS Azure Machine Learning. If you need guidance on this, please refer to Tutorial 4.

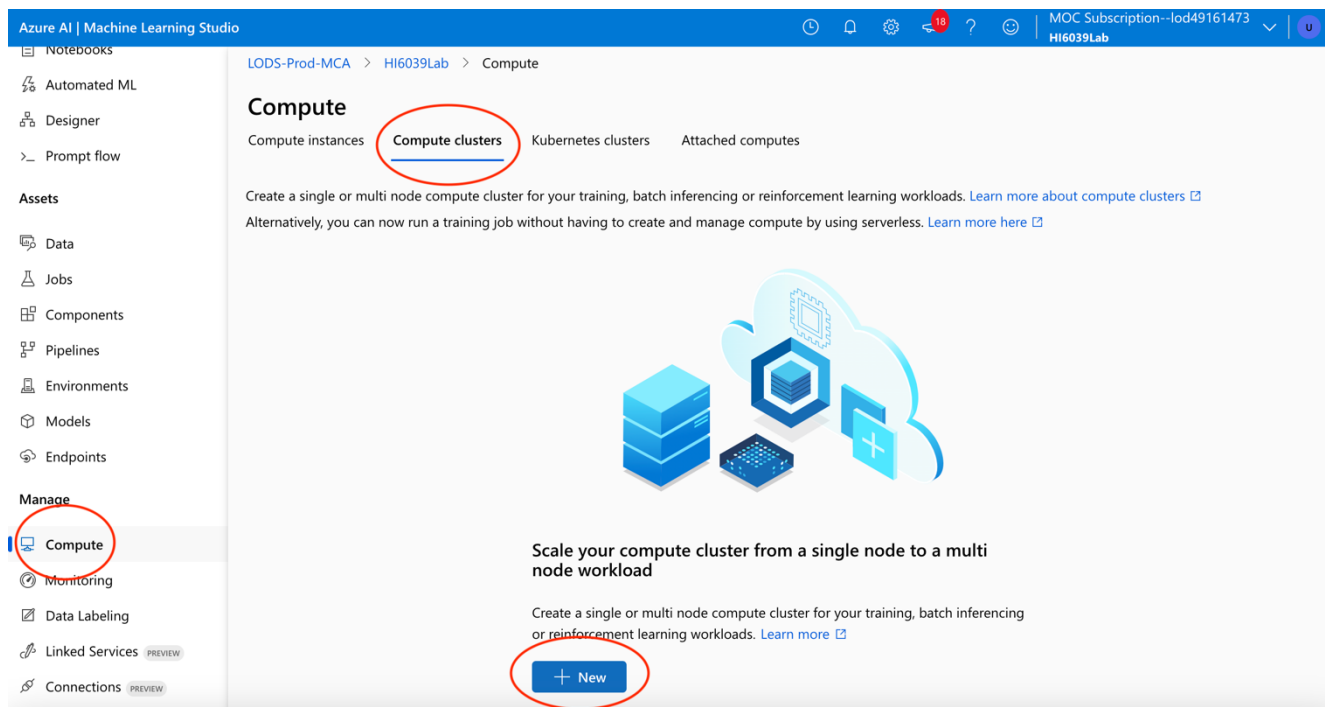
Download dataset

You need to download electric production dataset (Electric_Production.csv) from Blackboard or from the following URL: https://www.kaggle.com/datasets/shenba/electricity-production?select=Electric_Production.csv

You can also download this dataset on Blackboard directly.

Create compute

1. Click 'Compute', select 'Compute clusters', and click 'New'



2. Select a location, and click 'Next'

Azure AI | Machine Learning Studio

Create compute cluster

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

Location *
Australia Southeast

Virtual machine tier
☒ Dedicated ☐ Low priority

Virtual machine type
☒ CPU ☐ GPU

Virtual machine size
☒ Select from recommended options ☐ Select from all options

| | Name ↑ | Category | Workload types | Available quota | Cost |
|----------------------------------|---|------------------|--|-----------------|------------|
| <input type="radio"/> | Standard_DS11_v2 2 cores, 14GB RAM, 28GB storage | Memory optimized | Development on Notebooks (or other IDE) and light weight testing | 300 cores | \$0.20/... |
| <input checked="" type="radio"/> | Standard_DS3_v2 4 cores, 14GB RAM, 28GB storage | General purpose | Classical ML model training on small datasets | 300 cores | \$0.31/... |

Back Next Cancel

3. Enter a compute name, select the 'Maximum number of nodes', and click 'Create'

Create compute cluster

Configure Settings
Configure compute cluster settings for your selected virtual machine size.

| Name | Category | Cores | Available quota | RAM | Storage | Cost/Node |
|-----------------|-----------------|-------|-----------------|-------|---------|-----------|
| Standard_DS3_v2 | General purpose | 4 | 300 cores | 14 GB | 28 GB | \$0.31/hr |

Compute name *
compute-cluster1

Minimum number of nodes *
0

Maximum number of nodes *
75

Idle seconds before scale down *
120

☐ Enable SSH access

Advanced settings

Back Create Download a template for automation. Cancel

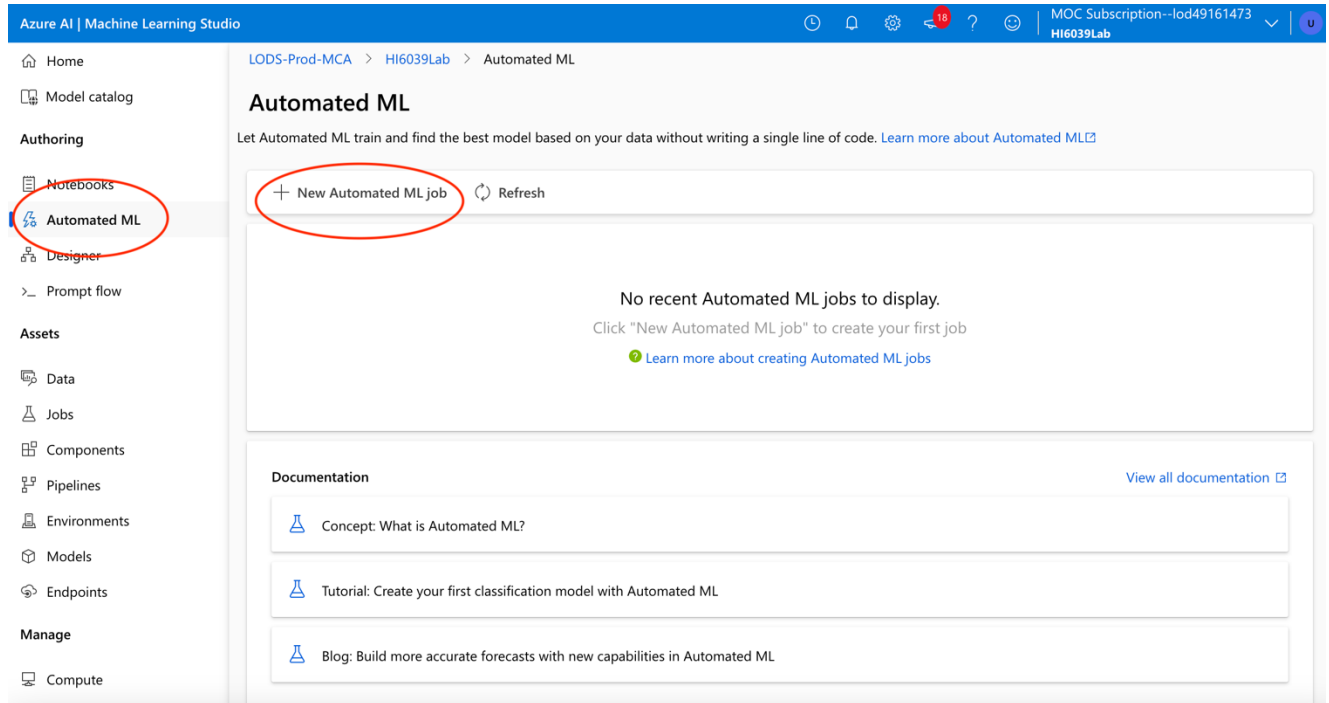
4. The compute will be created as follows:

The screenshot shows the Azure AI Machine Learning Studio interface. The left sidebar contains navigation options: Notebooks, Automated ML, Designer, Prompt flow, Assets, Data, Jobs, Components, Pipelines, Environments, Models, Endpoints, and Manage. The 'Manage' section is expanded, showing 'Compute', Monitoring, Data Labeling, Linked Services (PREVIEW), and Connections (PREVIEW). The main panel is titled 'Compute' and shows the 'Compute clusters' tab selected. A message at the top states: 'The "Kubernetes clusters" tab is now where you can access previous versions of "inference clusters" (also known as "AKS clusters") and "attached Kubernetes" compute types along with any previously created compute targets using those types. Learn more about Kubernetes clusters.' Below this, there are tabs for 'Compute instances', 'Compute clusters', 'Kubernetes clusters', and 'Attached computes'. The 'Compute clusters' tab is active, displaying instructions on how to create a cluster and links for more information. A table below lists the clusters. The first cluster, 'compute-cluster1', is circled in red and shows a 'Succeeded (0 nodes)' status. The table has columns for Name, State, Size, Location, Created on, and Active runs.

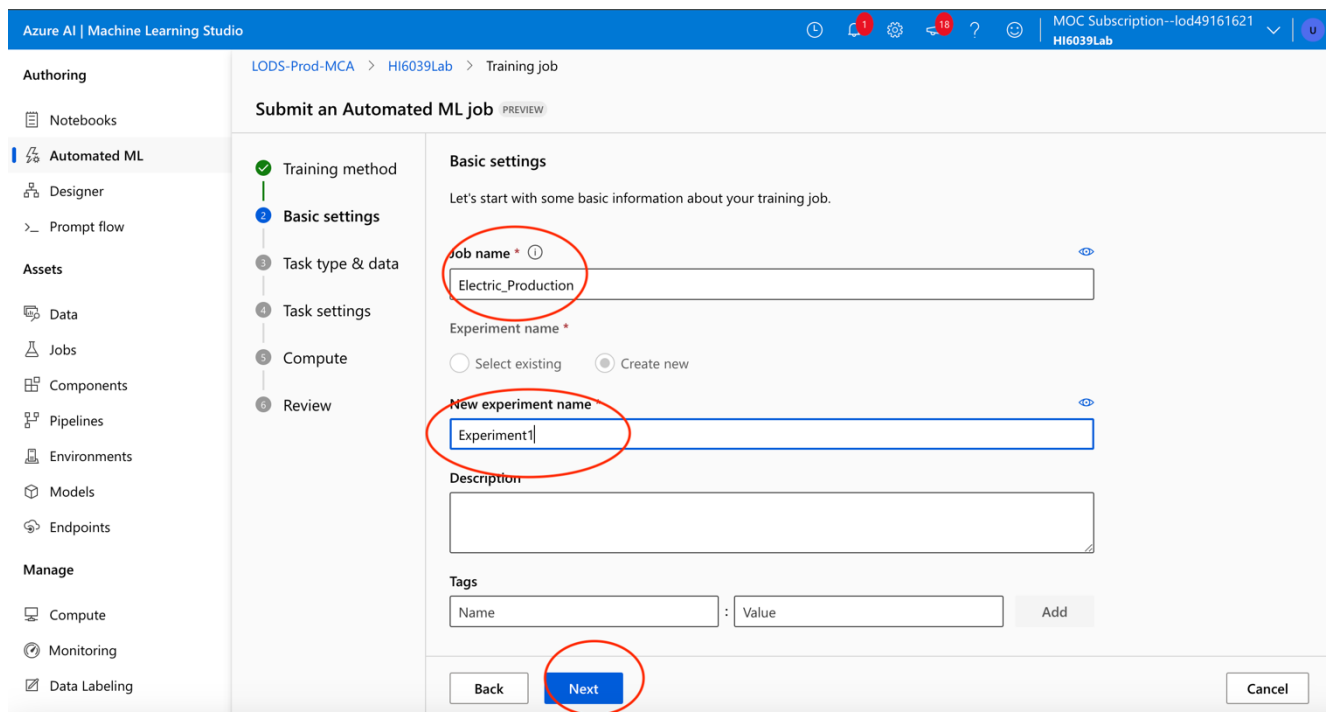
| Name | State | Size | Location | Created on | Active runs |
|------------------|---------------------|-----------------|--------------------|----------------------|-------------|
| compute-cluster1 | Succeeded (0 nodes) | Standard_DS3_v2 | australiasouthe... | May 15, 2024 9:28 AM | 0 |

Time Series Forecasting in AutoML

1. Click 'Automated ML', and click 'New Automated ML job'.



2. Enter a job name, an experiment name; and click 'Next'.



3. Select task type 'Time series forecasting', and click 'Create' to create a dataset.

Azure AI | Machine Learning Studio

LODS-Prod-MCA > HI6039Lab > Training job

Submit an Automated ML job PREVIEW

Training method
Basic settings
Task type & data
Task settings
Compute
Review

Task type & data

Choose the type of task that you would like your model to perform and the data to use for training. [Learn more](#)

Select task type * ⓘ
Time series forecasting *

Select data
Make sure your data is preprocessed into a supported format.

+ Create Refresh Show supported data assets only Reset view

Search Filter Columns

| Name | Type | Created on ↓ | Modified on |
|------|------|--------------|-------------|
|------|------|--------------|-------------|

Back Next Cancel

4. Enter a name, select type 'Tabular', and click 'Next'.

Azure AI | Machine Learning Studio

Create data asset

1 Data type
2 Data source

Set the name and type for your data asset

Name *
dataset1 *

Description
Data asset description

Type * ⓘ
Tabular

Back Next Cancel

Use cases for data types

When should I use File type?

The File type is recommended in most scenarios when you are working with a single data file of any type (including tabular data). This type allows you to specify a file location by URI in a storage location on your local computer, an attached Datastore, blob/ADLS storage, or a publicly available http(s) location. There are many types of supported URIs. In the Azure Machine Learning CLI v2 or Python SDK v2, this data type is called `uri_file`. [Learn more about the uri_file type](#)

When should I use Folder type?

The Folder type has all the same capabilities and use cases as the File type, but is used when specifying a folder location. In the Azure Machine Learning CLI v2 or Python SDK v2, this data type is called `uri_folder`. [Learn more about the uri_folder type](#)

5. Select 'From local files', and click 'Next'.

Create data asset

Data type (selected)
Data source (selected)
Destination storage type
File or folder selection
Settings
Schema
Review

Choose a source for your data asset
Choose the data source you want to create your asset from. A data source can be from a local storage location on your computer, from an attached datastore, from Azure storage, or from a publicly available web location.

- From Azure storage**
Create a data asset from registered data storage services including Azure Blob Storage, Azure file share, and Azure Data Lake.
- From local files** (selected)
Create a data asset by uploading files from your local drive.
- From SQL databases**
Create a dataset from Azure SQL database and Azure PostgreSQL database.
- From web files**
Create a data asset from a single file located at a public web URL.
- From Azure Open Datasets**
Create a dataset with one-click from pre-made data sets. These data sets are created by the general public and published as Azure Open Datasets.

Back **Next** **Cancel**

6. Click 'Next'

Create data asset

Data type (selected)
Data source (selected)
Destination storage type (selected)
File or folder selection
Settings
Schema
Review

Select a datastore
Choose a storage type and a datastore to upload your data to in the next step. You can also create a new datastore for your data first.

Datastore type *
Azure Blob Storage (selected) [Create new datastore](#)

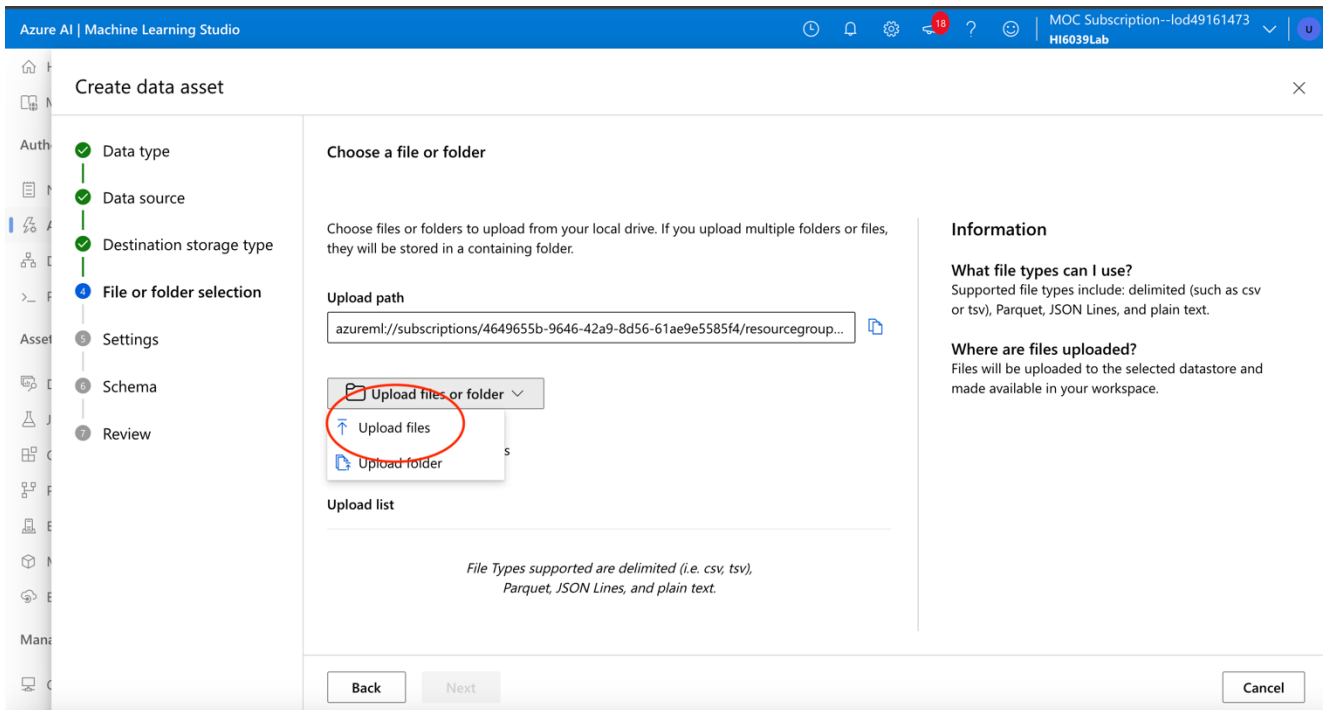
Filter **Columns**

| Name ↓ | Storage name | Created on |
|--|---------------------|----------------------|
| <input checked="" type="checkbox"/> workspaceblobstore | hi6039lab4765069976 | May 15, 2024 9:24 AM |
| workspaceartifactstore | hi6039lab4765069976 | May 15, 2024 9:24 AM |

Back **Next** **Cancel**

Page 1 of 1 25/Page

7. Select 'Upload files'.



Create data asset

Choose a file or folder

Choose files or folders to upload from your local drive. If you upload multiple folders or files, they will be stored in a containing folder.

Upload path

azureml://subscriptions/4649655b-9646-42a9-8d56-61ae9e5585f4/resourcegroup...

Upload files or folder

- Upload files
- Upload folder

Upload list

File Types supported are delimited (i.e. csv, tsv), Parquet, JSON Lines, and plain text.

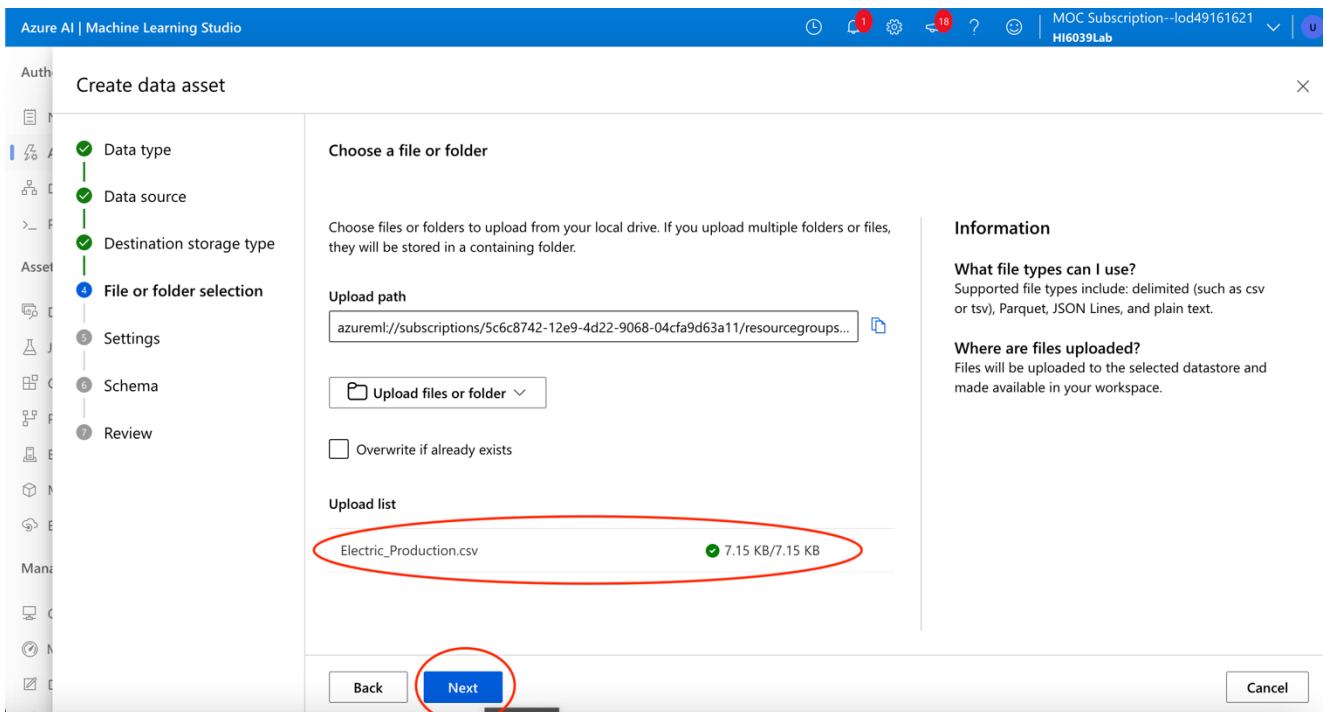
Information

What file types can I use?
Supported file types include: delimited (such as csv or tsv), Parquet, JSON Lines, and plain text.

Where are files uploaded?
Files will be uploaded to the selected datastore and made available in your workspace.

Back **Next** **Cancel**

8. Upload the dataset file you have already downloaded, and click 'Next'.



Create data asset

Choose a file or folder

Choose files or folders to upload from your local drive. If you upload multiple folders or files, they will be stored in a containing folder.

Upload path

azureml://subscriptions/5c6c8742-12e9-4d22-9068-04cfa9d63a11/resourcegroups...

Upload files or folder

☐ Overwrite if already exists

Upload list

| | |
|-------------------------|-----------------|
| Electric_Production.csv | 7.15 KB/7.15 KB |
|-------------------------|-----------------|

Back **Next** **Cancel**

9. Click 'Next'.

Azure AI | Machine Learning Studio

Create data asset

Settings

These settings determine how the data is parsed. The initial settings are automatically detected; you can change them as needed to reparse the data.

File format: Delimited | Delimiter: Comma | Example: Field1,Field2,Field3 | Encoding: UTF-8

Column headers: All files have same headers | Skip rows: None

☐ Dataset contains multi-line data

Note: Processing tabular files with multi-line data is slower because multiple CPU cores cannot be used to ingest the data in parallel. Checking this option may result in slower processing times.

Data preview

| DATE | IPG2211A2N |
|---------------------|------------|
| 1985-01-01 00:00:00 | 72.505 |
| 1985-02-01 00:00:00 | 70.672 |
| 1985-03-01 00:00:00 | 62.45 |
| 1985-04-01 00:00:00 | 57.471 |
| 1985-05-01 00:00:00 | 55.315 |

Back Next Review Cancel

10. Click 'Next'.

Azure AI | Machine Learning Studio

Create data asset

Schema

Column types are auto-detected based on the initial subset of the data and can be updated here. Values not aligning with the specified column type will fail conversion and would be either null-filled or replaced with error value. Any conversions preview errors are non-blocking and you can proceed.

Search column name

| Include | Column name | Type | Example values | Date format | Properties |
|-------------------------------------|-------------|-------------------|---------------------------|------------------------|---------------------|
| <input type="checkbox"/> | Path | String | | Not applicable to s... | Not applicable t... |
| <input checked="" type="checkbox"/> | DATE | Date | 1985-01-01 00:00:00, 1... | %m/%d/%Y | None |
| <input checked="" type="checkbox"/> | IPG2211A2N | Decimal (dot '.') | 72.5052, 70.672, 62.4502 | Not applicable to s... | Not applicable t... |

Back Next Cancel

11. Click 'Create'.

The screenshot shows the 'Create data asset' dialog in Azure Machine Learning Studio. The 'Review' step is active, showing details for 'dataset1'. The 'Create' button is circled in red.

Create data asset

Review
Review the settings for your data asset and make any changes as needed.

Data type

Name: dataset1
Description: --
Type: tabular

Data source

Type: Local

File selection

Upload path: azureml://subscriptions/5c6c8742-12e9-4d22-9068-04cfa9d63a11/resourceGroups/hi6039rg/workspaces/Hi6039Lab/datastores/workspaceblobstore/paths/UI/2024-05-15_004015_UTC/Electric_Production.csv

Schema

| DATE | Date |
|------------|---------|
| IPG2211A2N | Decimal |

Buttons: Back, **Create** (circled in red), Cancel

12. Select the dataset, and click 'Next'.

The screenshot shows the 'Submit an Automated ML job' dialog in Azure Machine Learning Studio. The 'Task type & data' step is active. The 'dataset1' dataset is selected in the list, and the 'Next' button is circled in red.

Submit an Automated ML job PREVIEW

Task type & data

Success: dataset1 data asset created successfully. It may take a few seconds for lists to be updated. [Click here to go to this data asset](#)

Choose the type of task that you would like your model to perform and the data to use for training. [Learn more](#)

Select task type *

Time series forecasting

Select data

Make sure your data is preprocessed into a supported format.

Buttons: + Create, Refresh, Show supported data assets only, Reset view

| Name | Type | Created on ↓ | Modified on |
|--|-------|-----------------------|------------------------|
| <input checked="" type="checkbox"/> dataset1 | Table | May 15, 2024 10:41 AM | May 15, 2024 10:41 ... |

Buttons: Back, **Next** (circled in red), Cancel

13. Select target column, time column; and click 'Next'.

The screenshot shows the 'Task settings' step in the 'Submit an Automated ML job' wizard. The left sidebar shows the 'Task settings' step is selected. The main content area shows the 'Task settings' section with the following options:

- Task type:** Time series forecasting
- Data:** dataset1 (View data)
- Target column *:** IPG221A2N (Decimal)
- Forecasting settings:**
 - Time column *:** DATE (Date)
 - ☒ Autodetect time series identifier(s)
 - ☒ Autodetect frequency *
 - ☒ Autodetect forecast horizon *

The 'Next' button is highlighted with a red circle.

14. Select compute type 'Compute cluster', select the compute you have already created, and click 'Next'.

The screenshot shows the 'Compute' step in the 'Submit an Automated ML job' wizard. The left sidebar shows the 'Compute' step is selected. The main content area shows the 'Compute' section with the following options:

- Compute:** Select and configure the compute resource for executing your training job.
- Select compute type:** Compute cluster
- Select Azure ML compute cluster:** compute-cluster1

The 'Next' button is highlighted with a red circle.

15. Click 'Submit training job'.

Azure AI | Machine Learning Studio

LODS-Prod-MCA > HI6039Lab > Training job

Submit an Automated ML job PREVIEW

Review

Review or make changes to your job before submission.

Basic settings

Name: Electric_Production

Experiment name: Experiment1

Description: --

Timeout (hours): --

Tags: --

Task type & data

Task type: Time series forecasting

Data: dataset1

Task setting

Target column: IPG2211A2N

Time column: DATE

Autodetect time series: Yes

Autodetect frequency: Yes

Autodetect forecast horizon: Yes

Enable deep learning: No

[Back](#) [Submit training job](#) [Cancel](#)

16. Click 'refresh' to check status.

Azure AI | Machine Learning Studio

LODS-Prod-MCA > HI6039Lab > Automated ML > Experiment1 > Electric_Production

Electric_Production Running

[Refresh](#) [Edit and submit \(preview\)](#) [+ Register model](#) [X Cancel](#) [Delete](#) [Compare \(preview\)](#)

Properties

Status: Running

Model training

Created on: May 15, 2024 10:42 AM

Start time: May 15, 2024 10:42 AM

Compute target: [compute-cluster1](#)

Name: electric_production

Script name: --

Inputs

Input name: training_data

Data asset: [dataset1:1](#)

Asset URI: [azureml:dataset1:1](#)

Outputs

Output name: full_training_dataset

Dataset: [a2512d6e-b5ff-48b5-a3db-987b946faac0](#)

Best model summary

No data

Run summary

Task name