Nawatech

Crafting Experience Beyond Engineering

nawatech.co

Day 1

Prepared for: HAGI

Explore the Azure Machine Learning workspace resources and assets

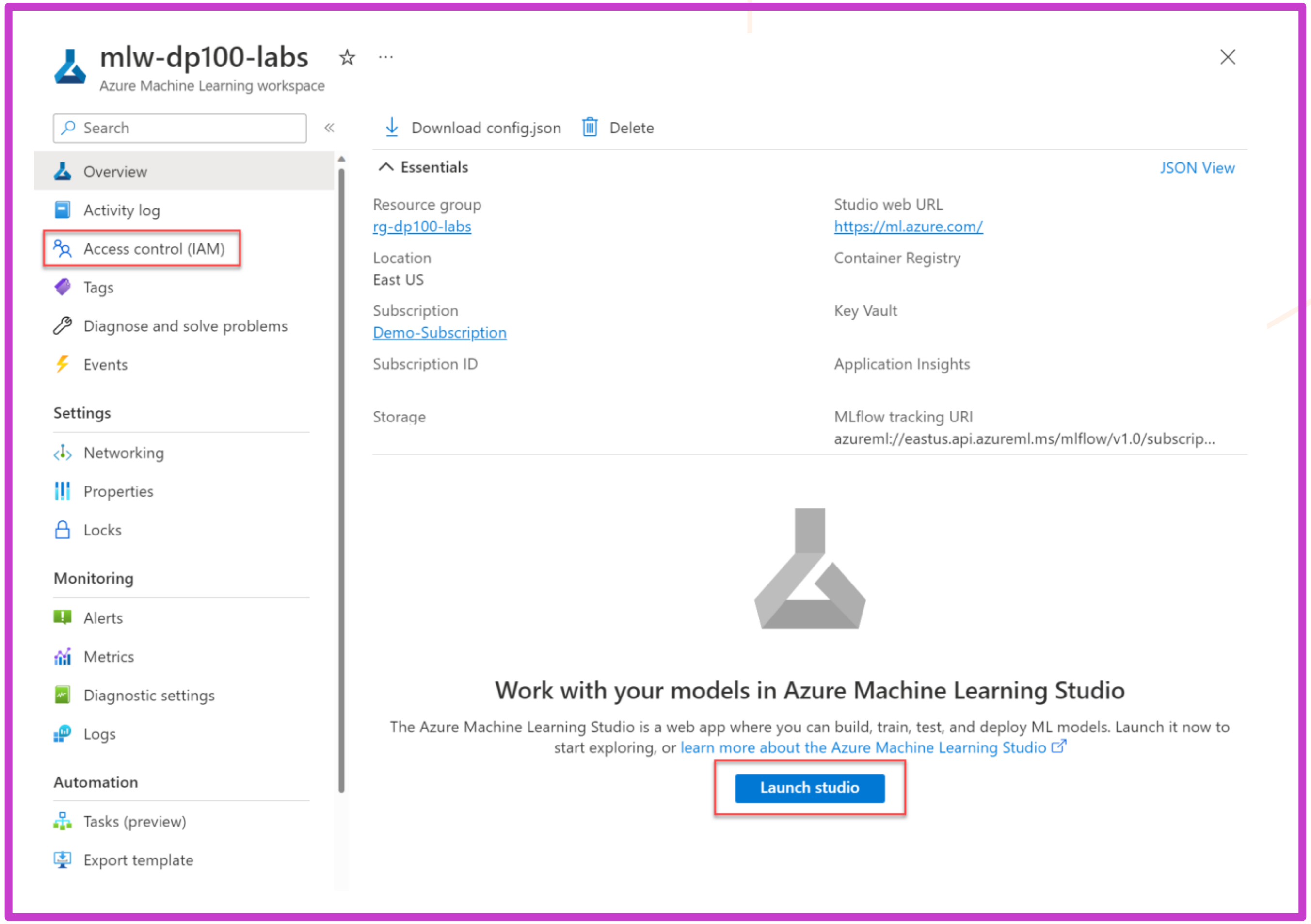
Introducing Azure Machine Learning

Azure Machine Learning provides a platform for data scientists to train, deploy, and manage their machine learning models on the Microsoft Azure platform.

It provides a comprehensive set of resources and assets to train and deploy effective machine learning models.

Explore the workspace in the Azure portal

- Give others access to the Azure Machine Learning workspace, using the **Access control**.
- Launch the **Azure Machine Learning studio**, an easy-to-use interface to create, manage, and use resources and assets in the workspace.



mlw-dp100-labs

Azure Machine Learning workspace

Essentials

Resource group
rg-dp100-labs

Location
East US

Subscription
Demo-Subscription

Subscription ID

Storage

Studio web URL
https://ml.azure.com/

Container Registry

Key Vault

Application Insights

MLflow tracking URI
azureml://eastus.api.azureml.ms/mlflow/v1.0/subscrip...

Work with your models in Azure Machine Learning Studio

The Azure Machine Learning Studio is a web app where you can build, train, test, and deploy ML models. Launch it now to start exploring, or learn more about the Azure Machine Learning Studio

Identify Azure Machine Learning resources

1

- The workspace – The top-level resource for Azure Machine Learning. The workspace keeps an overview of all logs, metrics, outputs, models, and snapshots of your code.

2

- Compute resources – There are five types of compute in the Azure Machine Learning workspace: compute instances, compute clusters, Kubernetes clusters, attached computes, and serverless compute.

3

- Datastores – All data is stored in datastores, which are references to Azure data services. Four datastores will exist by default.

Identify Azure Machine Learning assets

1

- Models – Save trained models in the workspace. A common way to store such models is to package the model as a Python pickle file (.pkl extension).

2

- Environments – Specify software packages, environment variables, and software settings to run scripts. An environment is stored as an image in the Azure Container Registry created with the workspace when it's used for the first time.

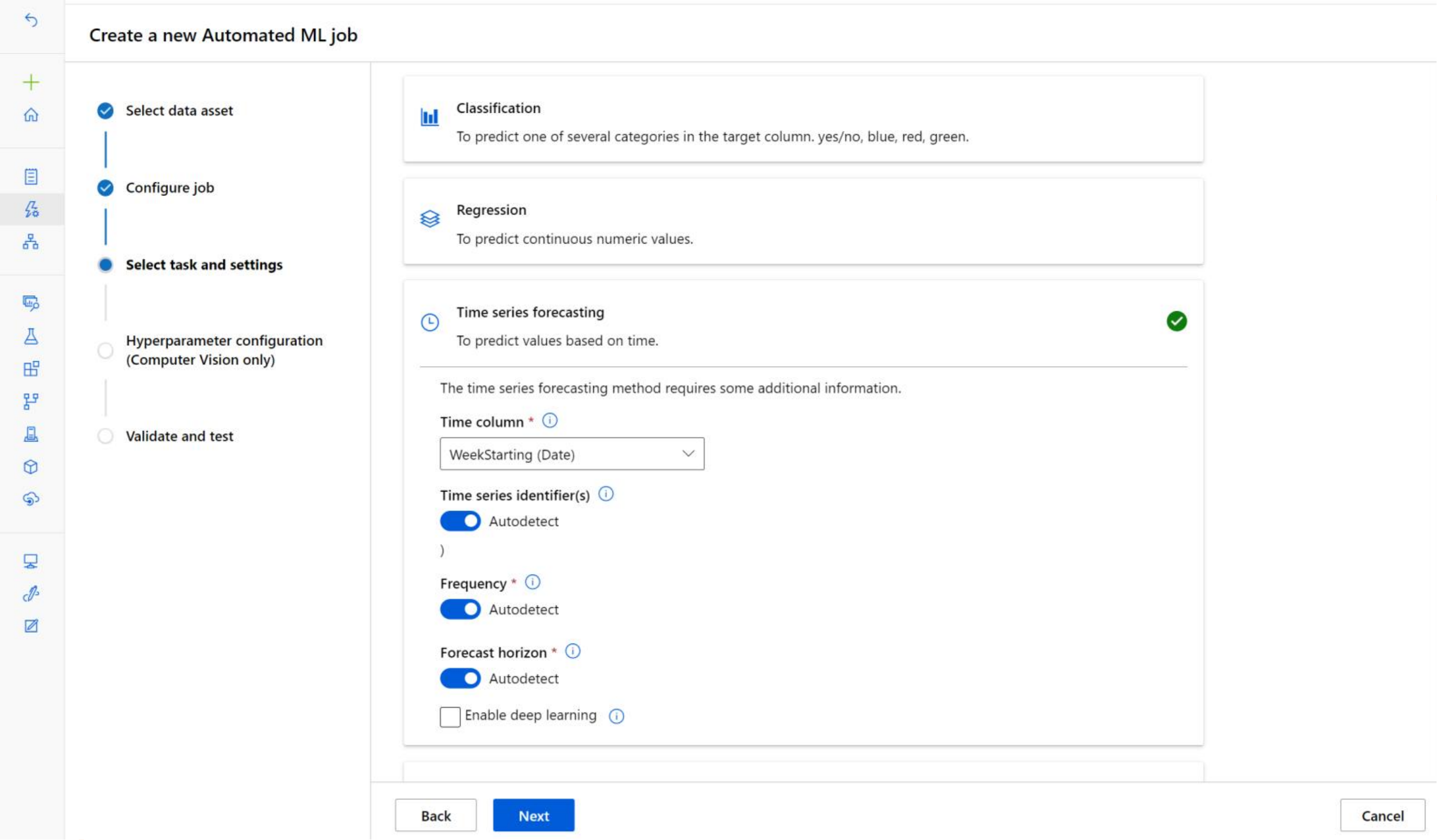
3

- Data – You can use data assets to easily access data every time, without having to provide authentication every time you want to access it.

4

- Components – Make it easier to share code with component in a workspace.

Explore algorithms and hyperparameter values with Automated Machine Learning



Create a new Automated ML job

- ✓ Select data asset
- ✓ Configure job
- ● **Select task and settings**
- ○ Hyperparameter configuration (Computer Vision only)
- ○ Validate and test

Classification
To predict one of several categories in the target column. yes/no, blue, red, green.

Regression
To predict continuous numeric values.

Time series forecasting ✓
To predict values based on time.

The time series forecasting method requires some additional information.

Time column *
WeekStarting (Date)

Time series identifier(s)
Autodetect

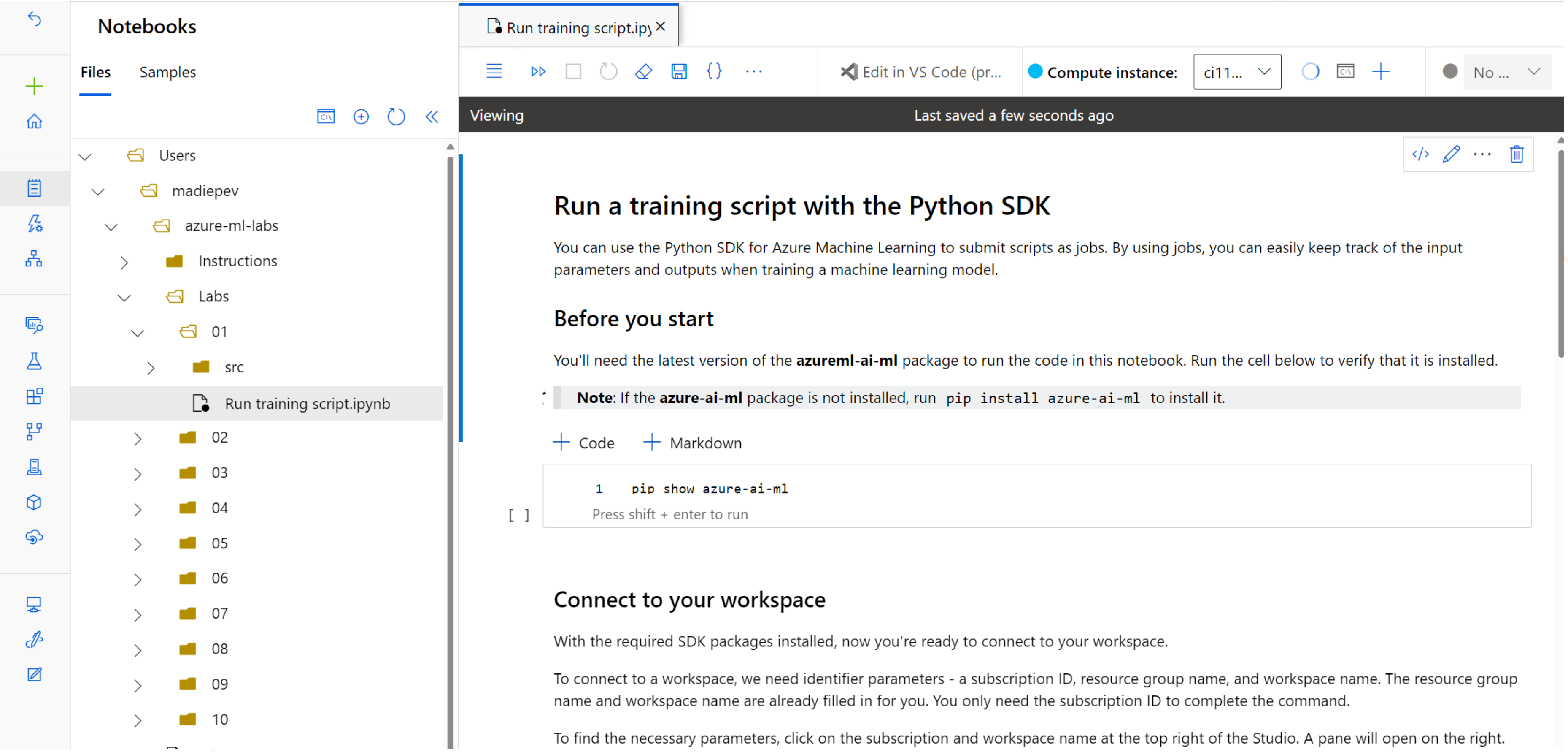
)

Frequency *
Autodetect

Forecast horizon *
Autodetect

□ Enable deep learning

Run a notebook



Run a training script with the Python SDK

You can use the Python SDK for Azure Machine Learning to submit scripts as jobs. By using jobs, you can easily keep track of the input parameters and outputs when training a machine learning model.

Before you start

You'll need the latest version of the **azureml-ai-ml** package to run the code in this notebook. Run the cell below to verify that it is installed.

> **Note**: If the **azure-ai-ml** package is not installed, run `pip install azure-ai-ml` to install it.

```
pip show azure-ai-ml
```

Press shift + enter to run

Connect to your workspace

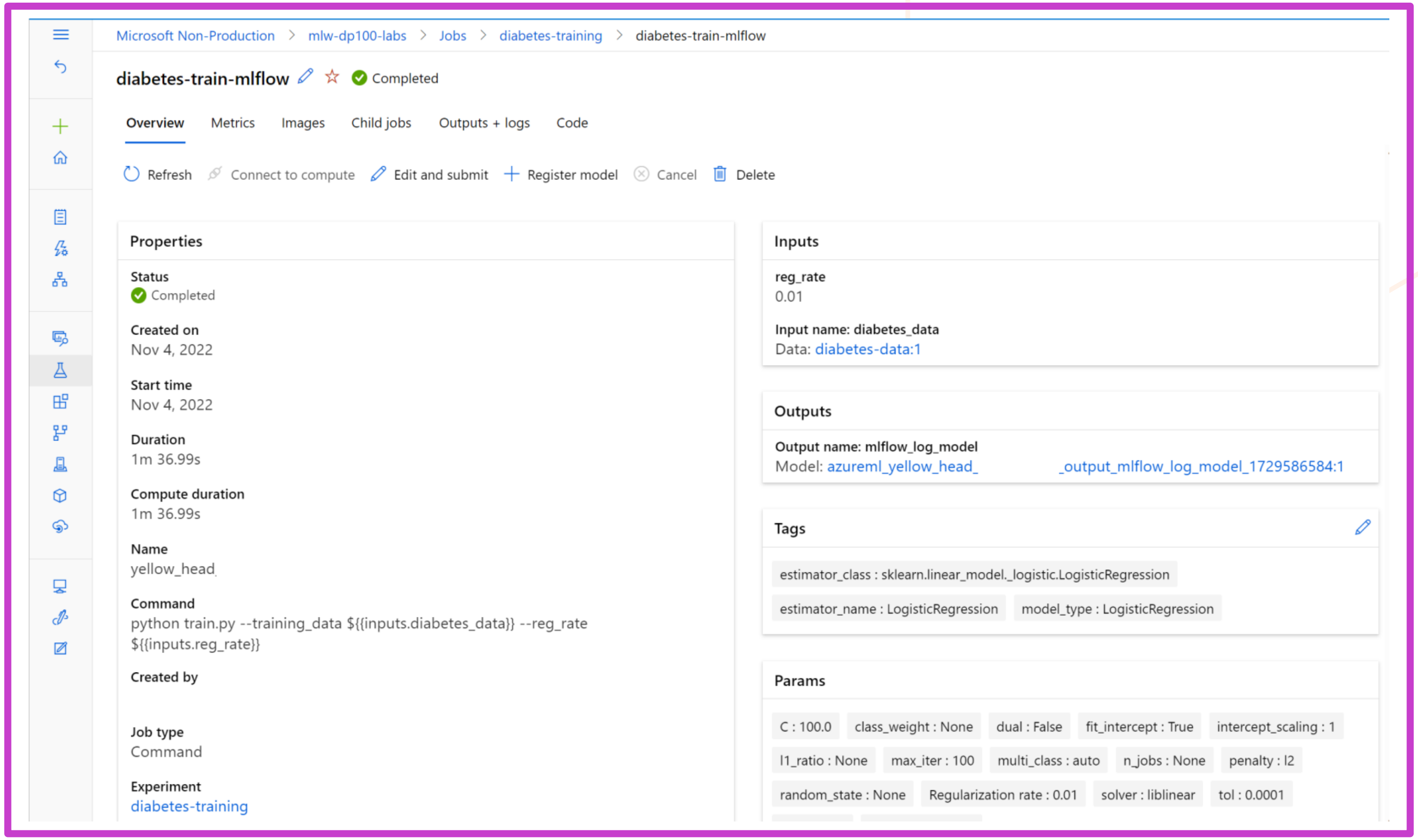
With the required SDK packages installed, now you're ready to connect to your workspace.

To connect to a workspace, we need identifier parameters - a subscription ID, resource group name, and workspace name. The resource group name and workspace name are already filled in for you. You only need the subscription ID to complete the command.

To find the necessary parameters, click on the subscription and workspace name at the top right of the Studio. A pane will open on the right.

Run a script as a job

- When you submit a job to the workspace, all inputs and outputs will be stored in the workspace.
- There are different types of jobs:
  - **Command**: Execute a single script.
  - **Sweep**: Perform hyperparameter tuning when executing a single script.
  - **Pipeline**: Run a pipeline consisting of multiple scripts or components.



Properties

Status
Completed

Created on
Nov 4, 2022

Start time
Nov 4, 2022

Duration
1m 36.99s

Compute duration
1m 36.99s

Name
yellow_head

Command
python train.py --training_data ${{inputs.diabetes_data}} --reg_rate ${{inputs.reg_rate}}

Created by

Job type
Command

Experiment
diabetes-training

Inputs

reg_rate
0.01

Input name: diabetes_data
Data: diabetes-data:1

Outputs

Output name: mlflow_log_model
Model: azureml_yellow_head_ _output_mlflow_log_model_1729586584:1

Tags

estimator_class : sklearn.linear_model._logistic.LogisticRegression
estimator_name : LogisticRegression
model_type : LogisticRegression

Params

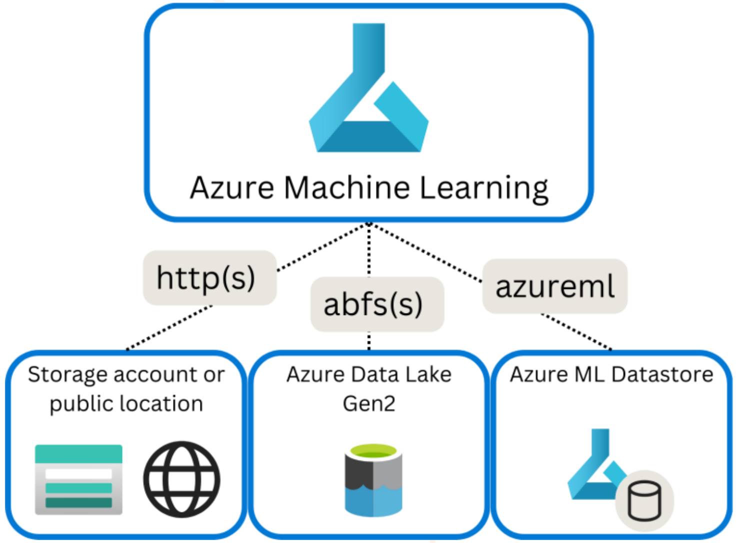
C : 100.0
class_weight : None
dual : False
fit_intercept : True
intercept_scaling : 1
l1_ratio : None
max_iter : 100
multi_class : auto
n_jobs : None
penalty : l2
random_state : None
Regularization rate : 0.01
solver : liblinear
tol : 0.0001

Make data available in Azure Machine Learning

Understand URIs

A URI references the **location of your data.**

For Azure Machine Learning to connect to your data directly, you need to prefix the URI with the appropriate protocol.

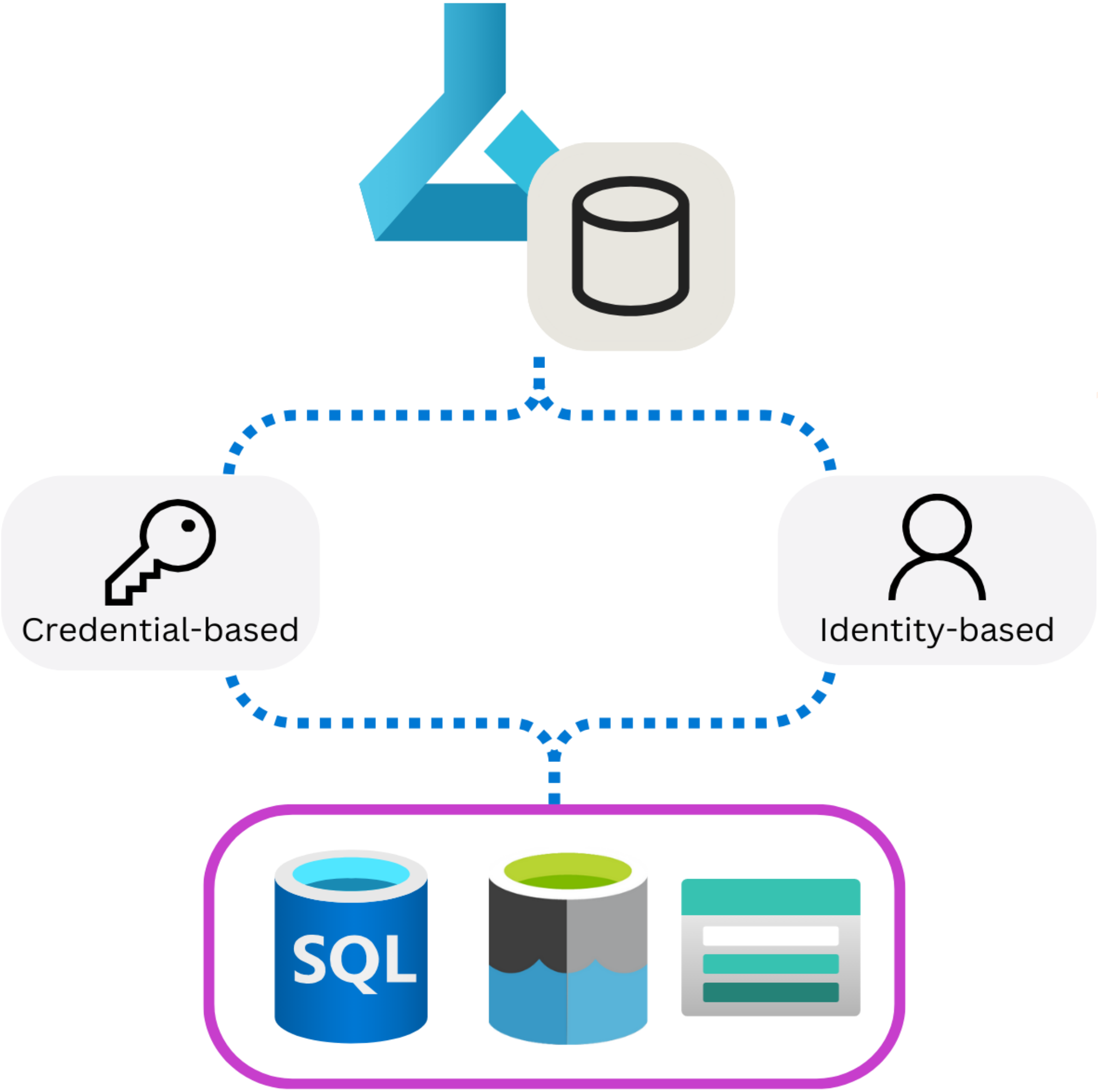


Understand datastores

Datastores are **abstractions** for cloud data sources, storing the **connection information**.

The benefits of datastores:

- Provide easy-to-use URIs to your data storage.
- Facilitates data discovery within Azure Machine Learning.
- Securely stores connection information, without exposing secrets and keys to data scientists.

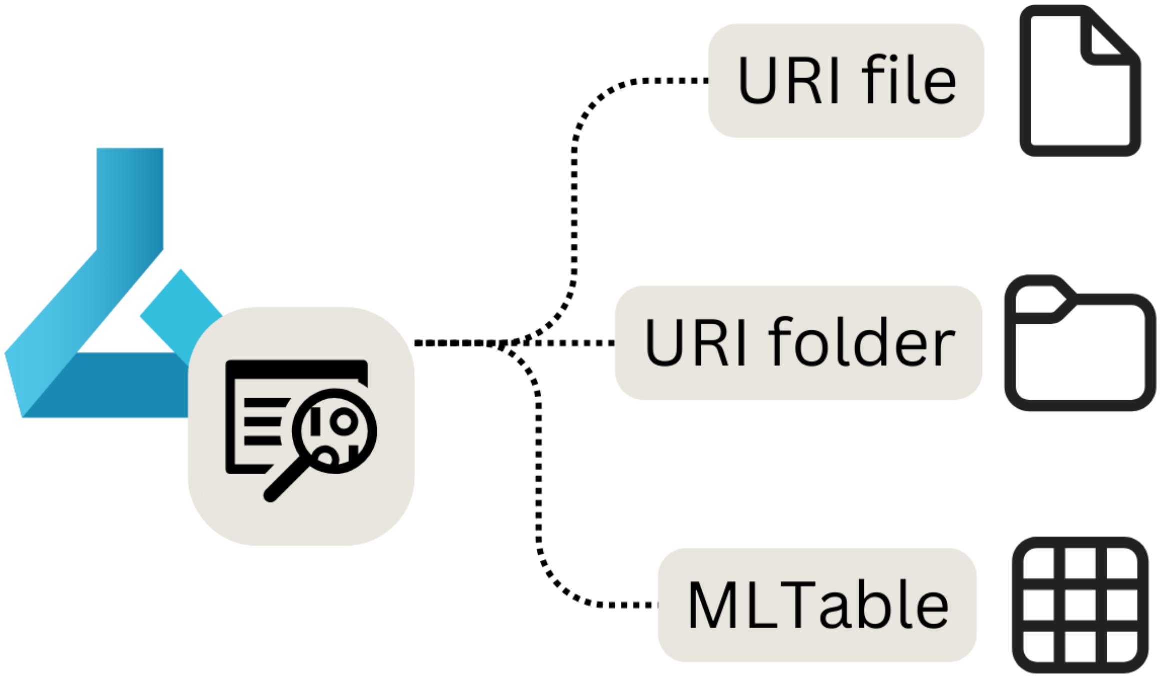


Understand data assets

Data assets are **references** to where the data is stored, how to get access, and any other relevant metadata.

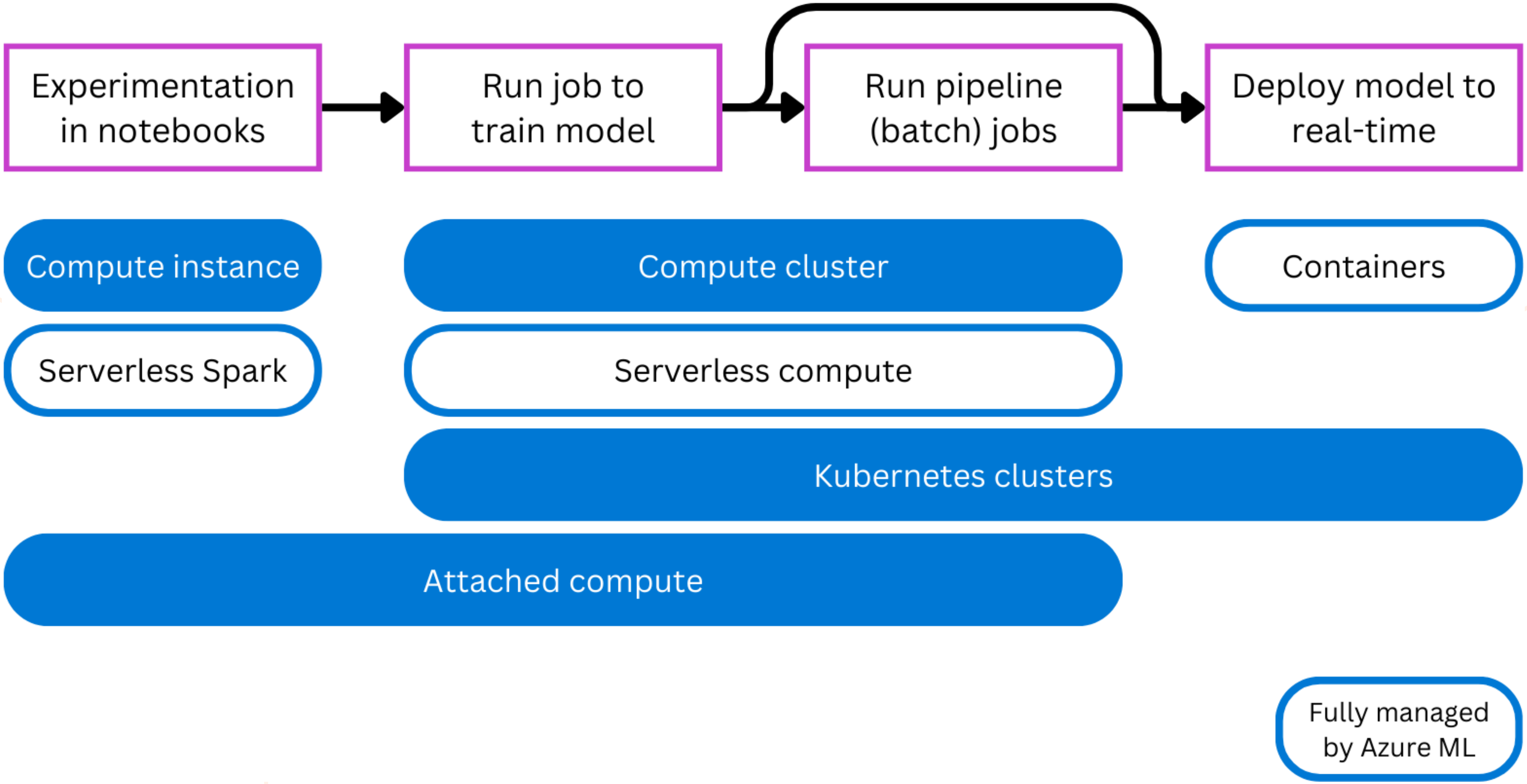
The benefits of data assets:

- Share and reuse data with other members.
- Seamlessly access data during model training (on any supported compute type) without worrying about connection strings or data paths.
- Version the metadata of the data asset.



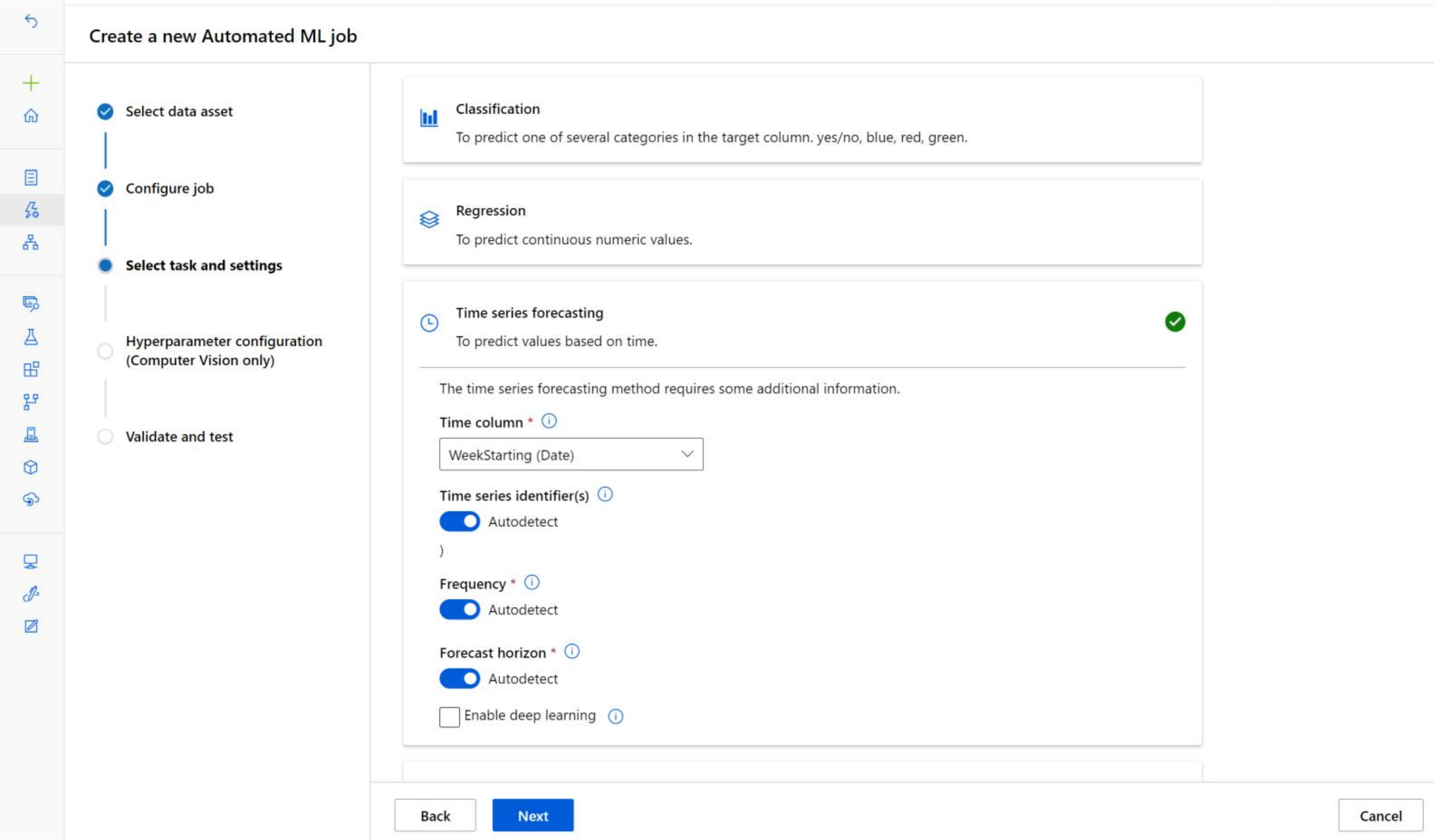
Work with compute resources in Azure Machine Learning

Choose the appropriate compute target



Automated Machine Learning

Explore algorithms and hyperparameter values with Automated Machine Learning



Create a new Automated ML job

- Select data asset
- Configure job
- Select task and settings
- Hyperparameter configuration (Computer Vision only)
- Validate and test

Classification

To predict one of several categories in the target column. yes/no, blue, red, green.

Regression

To predict continuous numeric values.

Time series forecasting

To predict values based on time.

The time series forecasting method requires some additional information.

Time column *

WeekStarting (Date)

Time series identifier(s)

Autodetect

)

Frequency *

Autodetect

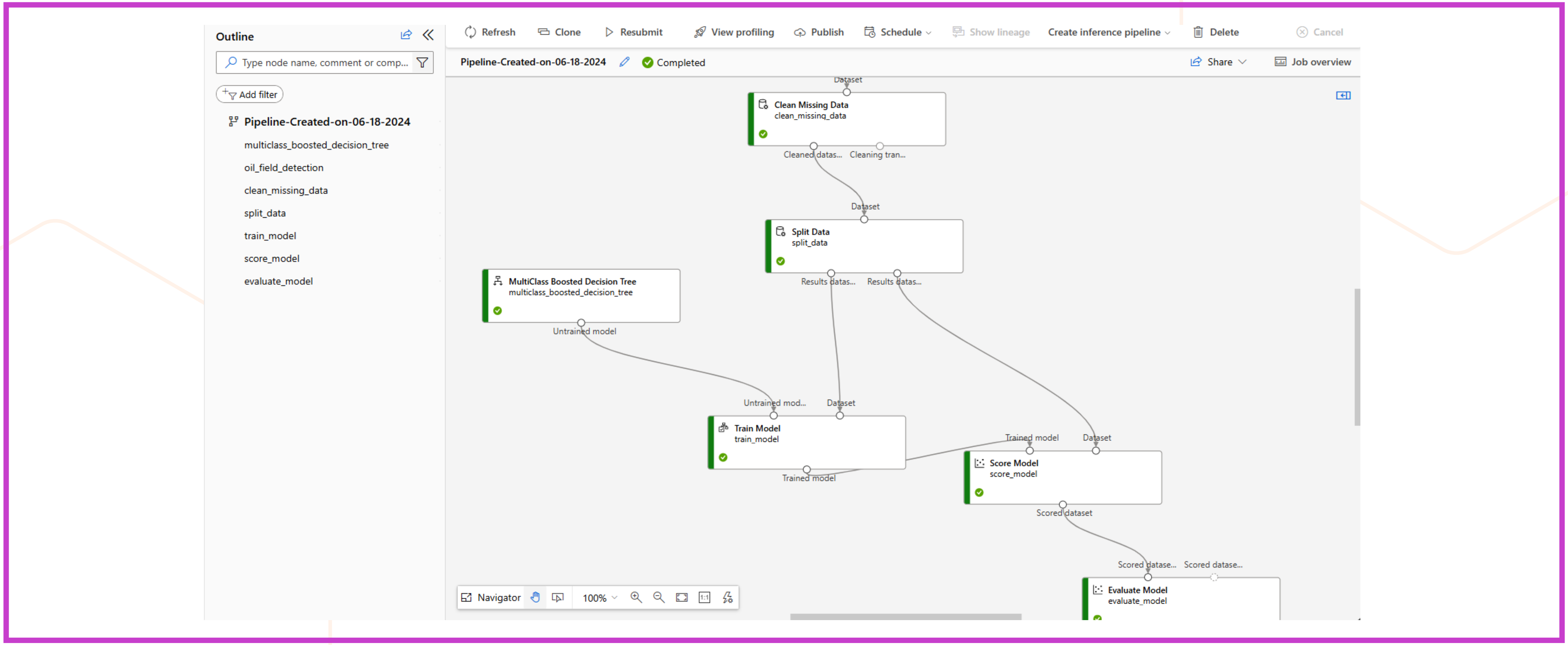
Forecast horizon *

Autodetect

□ Enable deep learning

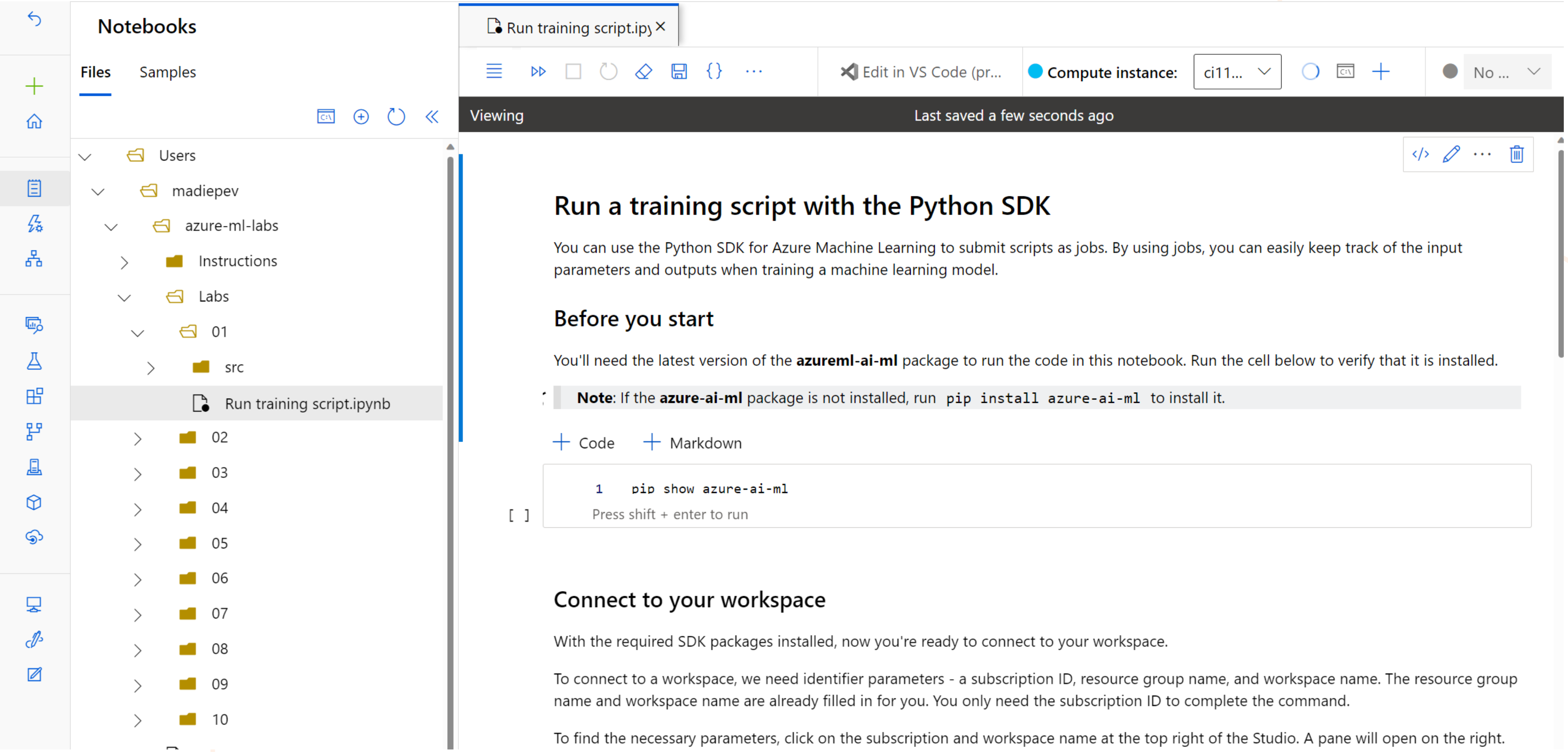
Azure Machine Learning Designer

Create pipeline designer



Azure Machine Learning Notebook

Run a notebook



Run a training script with the Python SDK

You can use the Python SDK for Azure Machine Learning to submit scripts as jobs. By using jobs, you can easily keep track of the input parameters and outputs when training a machine learning model.

Before you start

You'll need the latest version of the **azureml-ai-ml** package to run the code in this notebook. Run the cell below to verify that it is installed.

> **Note**: If the **azure-ai-ml** package is not installed, run `pip install azure-ai-ml` to install it.

```
[ ]  pip show azure-ai-ml
```

Press shift + enter to run

Connect to your workspace

With the required SDK packages installed, now you're ready to connect to your workspace.

To connect to a workspace, we need identifier parameters - a subscription ID, resource group name, and workspace name. The resource group name and workspace name are already filled in for you. You only need the subscription ID to complete the command.

To find the necessary parameters, click on the subscription and workspace name at the top right of the Studio. A pane will open on the right.

Environment Setup: Azure Machine Learning

Environment Setup

#1 Compute

Prerequisites: Make sure your account already has access and can sign in to the Azure Machine Learning workspace.

Navigate to Compute Instance and Compute Cluster.

Create Compute Instance/Cluster.

Set configuration compute.

#2 Kernel, ML version

Prerequisites: Make sure to having compute instance and some notebook.

Start compute and open a notebook.

Choose kernel based on use case.

Exercise:
Azure Machine Learning
Hands On

USECASE

Geothermal Gradient Prediction

- Objectives :

This Project is aimed at predicting geothermal characteristics for Colombia, specifically the geothermal gradient, based on available geological and geophysical data. We employ machine learning techniques to make these predictions. This project focuses on predicting the Apparent Geothermal Gradient (°C/Km) as an essential factor in geothermal exploration. The code and the results are in Model.ipynb.

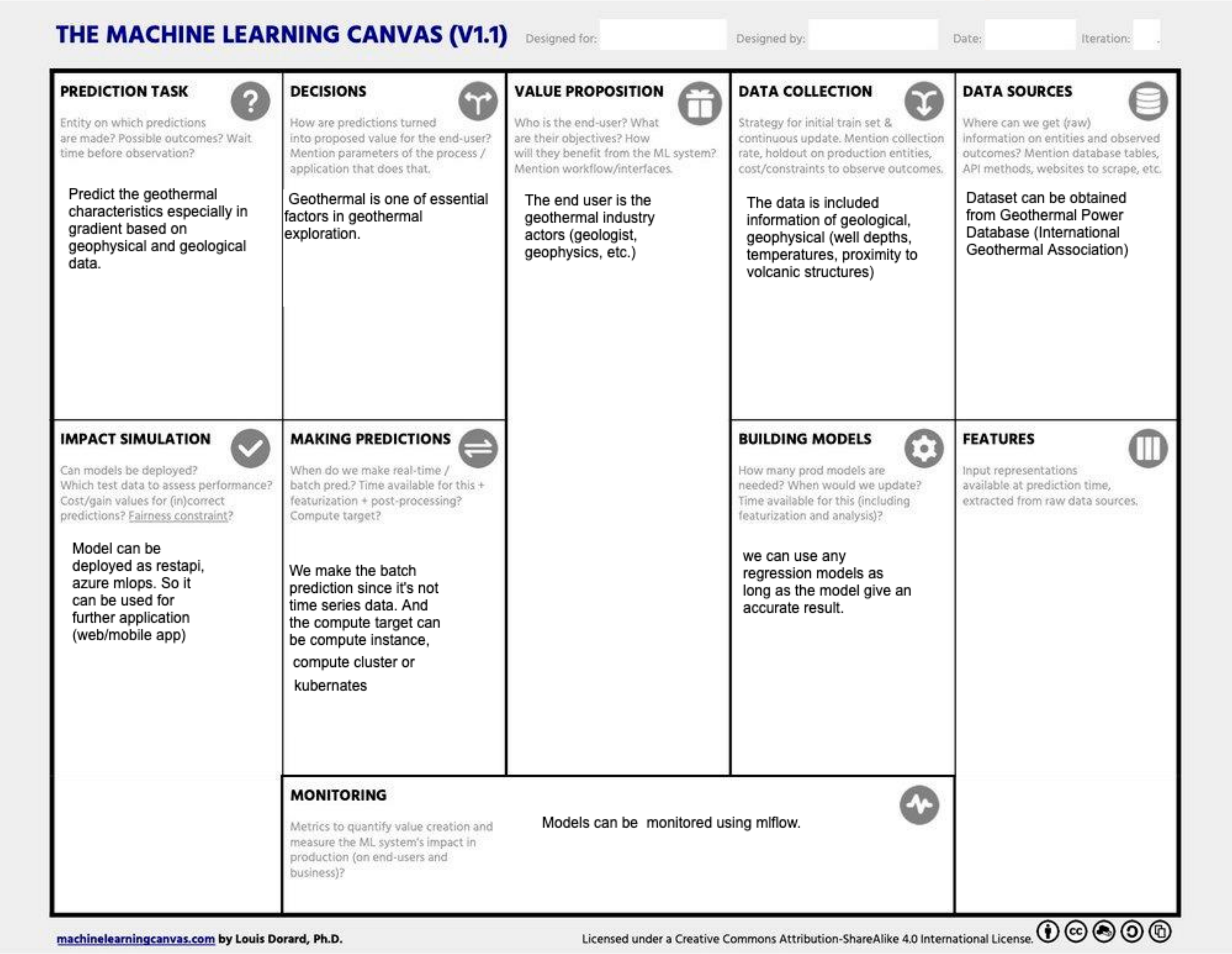
- Methodology :

The project utilizes geospatial data, geophysical information, and geothermal measurements. These datasets are located in the data folder of this repository. The data includes information on well depths, temperatures, geological features, and proximity to volcanic structures.

USECASE

Geothermal Gradient Prediction

- Machine Learning Canvas



THE MACHINE LEARNING CANVAS (V1.1)

Designed for:
Designed by:
Date:
Iteration:

PREDICTION TASK

Entity on which predictions are made? Possible outcomes? Wait time before observation?

Predict the geothermal characteristics especially in gradient based on geophysical and geological data.

DECISIONS

How are predictions turned into proposed value for the end-user? Mention parameters of the process / application that does that.

Geothermal is one of essential factors in geothermal exploration.

VALUE PROPOSITION

Who is the end-user? What are their objectives? How will they benefit from the ML system? Mention workflow/interfaces.

The end user is the geothermal industry actors (geologist, geophysics, etc.)

DATA COLLECTION

Strategy for initial train set & continuous update. Mention collection rate, holdout on production entities, cost/constraints to observe outcomes.

The data is included information of geological, geophysical (well depths, temperatures, proximity to volcanic structures)

DATA SOURCES

Where can we get (raw) information on entities and observed outcomes? Mention database tables, API methods, websites to scrape, etc.

Dataset can be obtained from Geothermal Power Database (International Geothermal Association)

IMPACT SIMULATION

Can models be deployed? Which test data to assess performance? Cost/gain values for (in)correct predictions? Fairness constraint?

Model can be deployed as restapi, azure mlops. So it can be used for further application (web/mobile app)

MAKING PREDICTIONS

When do we make real-time / batch pred.? Time available for this + featurization + post-processing? Compute target?

We make the batch prediction since it's not time series data. And the compute target can be compute instance, compute cluster or kubernates

BUILDING MODELS

How many prod models are needed? When would we update? Time available for this (including featurization and analysis)?

we can use any regression models as long as the model give an accurate result.

FEATURES

Input representations available at prediction time, extracted from raw data sources.

MONITORING

Metrics to quantify value creation and measure the ML system's impact in production (on end-users and business)?

Models can be monitored using mlflow.

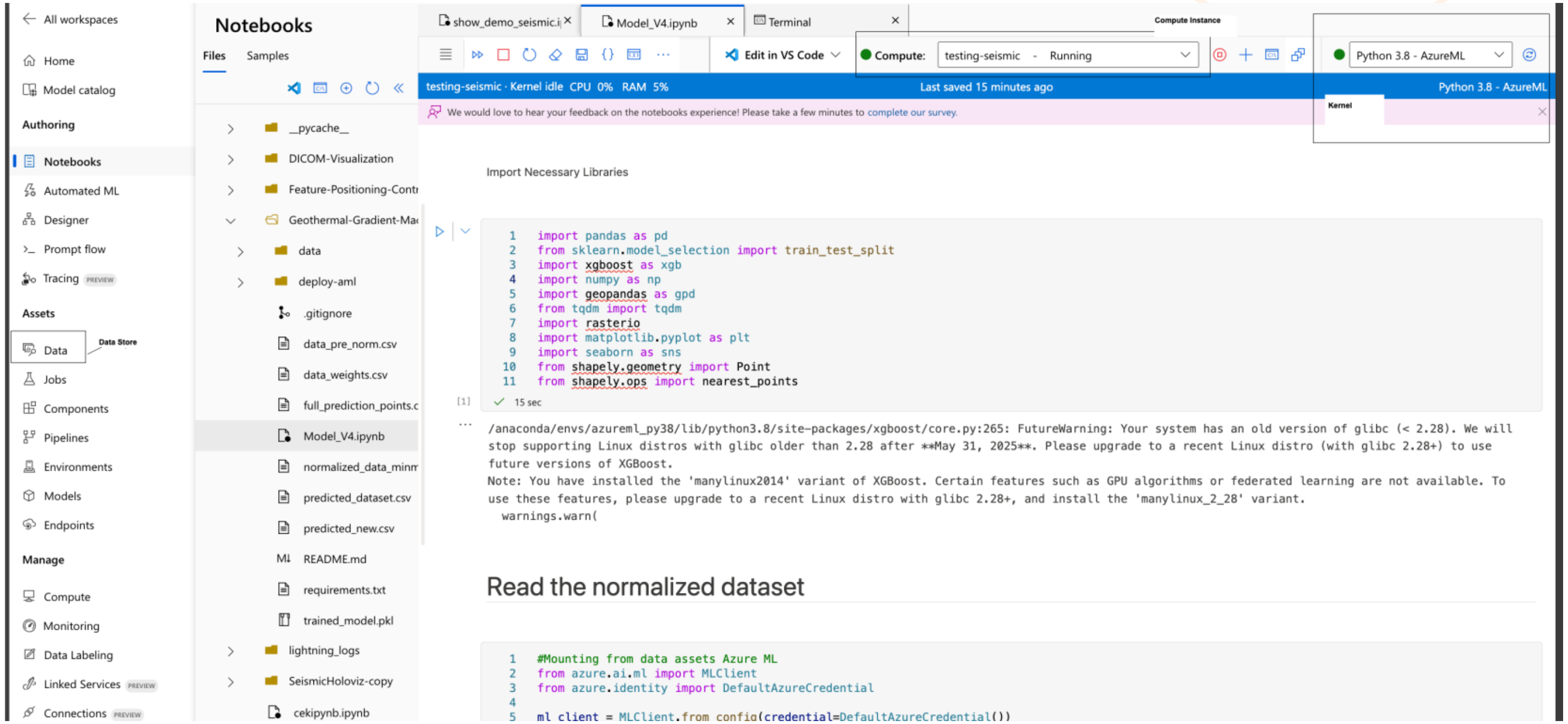
machinelearningcanvas.com by Louis Dorard, Ph.D.

USECASE

Geothermal Gradient Prediction using Notebook

Method :

- Open Azure ML Notebooks
- Connect to active compute instance and kernel
- Install all library required
- Connect your notebooks to data assets



Compute: testing-seismic - Running

Python 3.8 - AzureML

testing-seismic · Kernel idle CPU 0% RAM 5%

Last saved 15 minutes ago

Python 3.8 - AzureML

We would love to hear your feedback on the notebooks experience! Please take a few minutes to complete our survey.

Import Necessary Libraries

```
import pandas as pd
from sklearn.model_selection import train_test_split
import xgboost as xgb
import numpy as np
import geopandas as gpd
from tqdm import tqdm
import rasterio
import matplotlib.pyplot as plt
import seaborn as sns
from shapely.geometry import Point
from shapely.ops import nearest_points
```

[1] ✓ 15 sec

```
/anaconda/envs/azureml_py38/lib/python3.8/site-packages/xgboost/core.py:265: FutureWarning: Your system has an old version of glibc (< 2.28). We will
stop supporting Linux distros with glibc older than 2.28 after **May 31, 2025**. Please upgrade to a recent Linux distro (with glibc 2.28+) to use
future versions of XGBoost.
Note: You have installed the 'manylinux2014' variant of XGBoost. Certain features such as GPU algorithms or federated learning are not available. To
use these features, please upgrade to a recent Linux distro with glibc 2.28+, and install the 'manylinux_2_28' variant.
  warnings.warn(
```

Read the normalized dataset

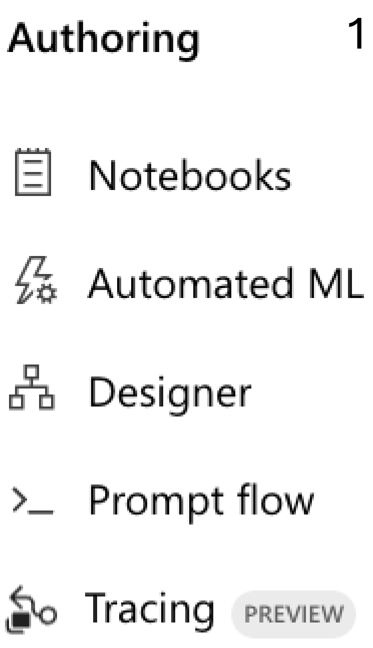
```
#Mounting from data assets Azure ML
from azure.ai.ml import MLClient
from azure.identity import DefaultAzureCredential

ml_client = MLClient.from_config(credential=DefaultAzureCredential())
```

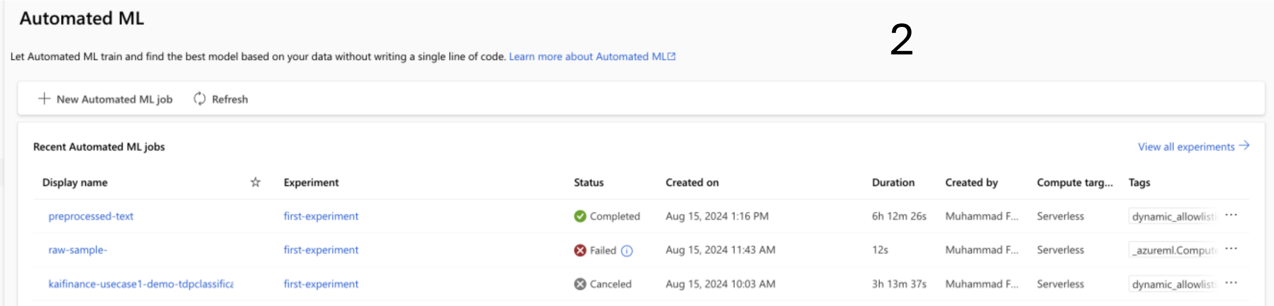
USECASE

Geothermal Gradient Prediction using Auto ML

Method :

- **Open Azure Automated ML**
- **Create new project or pick existing project and pick ML type (classification, regression, etc)**
- **Connect to your dataset.**
- **Train your model.**


1

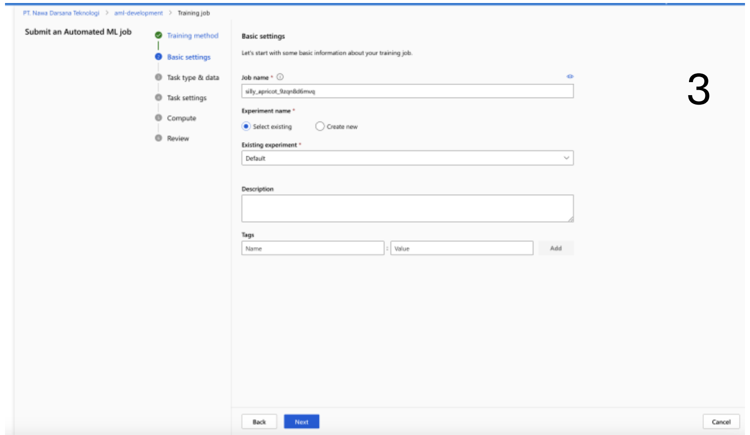

2

Automated ML

Let Automated ML train and find the best model based on your data without writing a single line of code. Learn more about Automated ML

Recent Automated ML jobs

Display name	Experiment	Status	Created on	Duration	Created by	Compute targ...	Tags
preprocessed-text	first-experiment	Completed	Aug 15, 2024 1:16 PM	6h 12m 26s	Muhammad F...	Serverless	dynamic_allowlist
raw-sample-	first-experiment	Failed	Aug 15, 2024 11:43 AM	12s	Muhammad F...	Serverless	_azureml.Comput
kaifinance-usecase1-demo-tdpclassifica	first-experiment	Canceled	Aug 15, 2024 10:03 AM	3h 13m 37s	Muhammad F...	Serverless	dynamic_allowlist


3

PT. Nawa Darsana Teknologi > aml-development > Training job

Submit an Automated ML job

Basic settings

Let's start with some basic information about your training job.

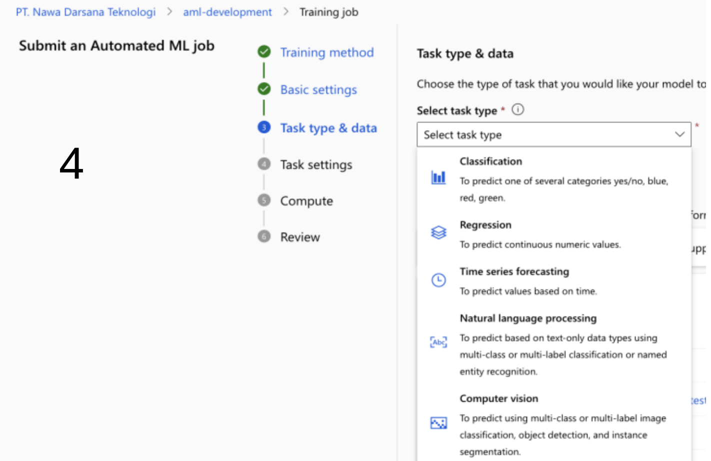
Job name *: silly_apricot_9zqn8d6mwq

Experiment name *: ● Select existing ○ Create new

Existing experiment *: Default

Description:

Tags: Name : Value


4

PT. Nawa Darsana Teknologi > aml-development > Training job

Submit an Automated ML job

Task type & data

Choose the type of task that you would like your model to

Select task type *: Select task type

Classification
To predict one of several categories yes/no, blue, red, green.

Regression
To predict continuous numeric values.

Time series forecasting
To predict values based on time.

Natural language processing
To predict based on text-only data types using multi-class or multi-label classification or named entity recognition.

Computer vision
To predict using multi-class or multi-label image classification, object detection, and instance segmentation.

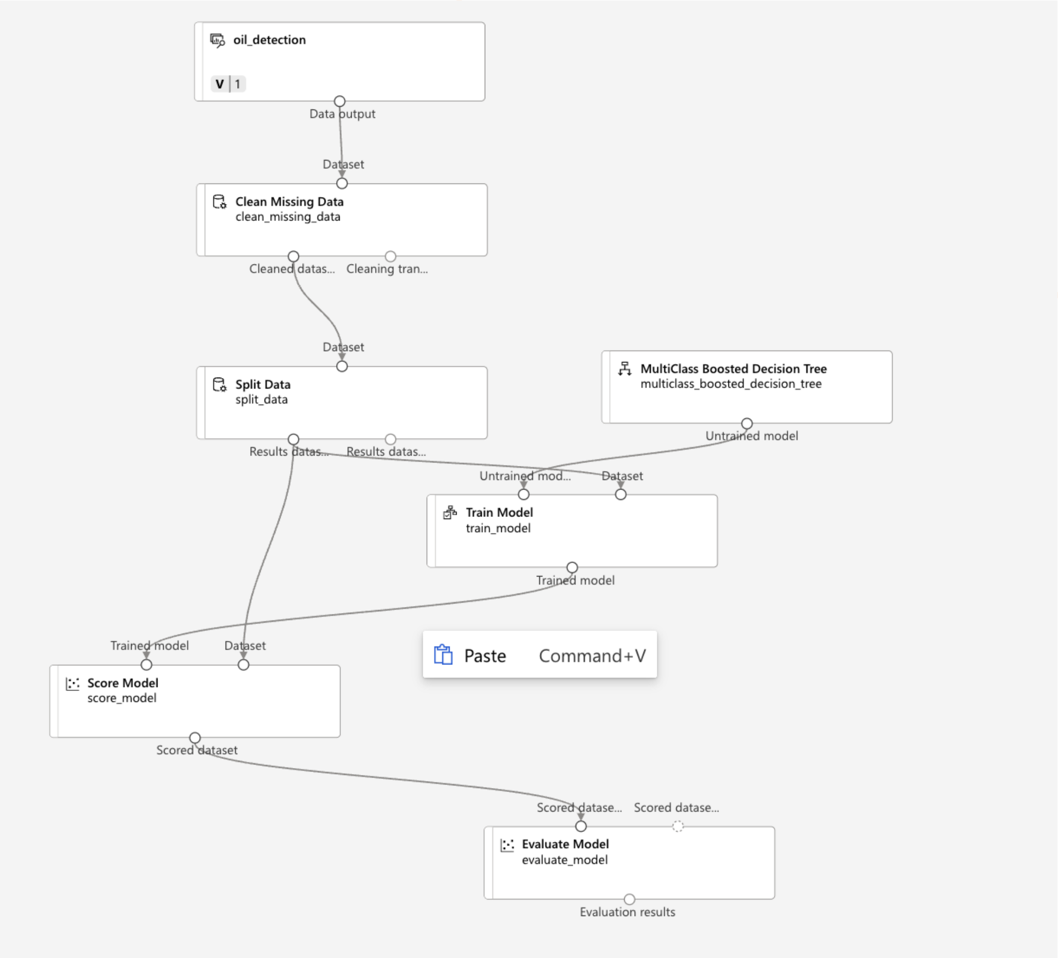
USECASE

Geothermal Gradient Prediction using AML Designer

Technically the steps we do in Azure Machine Learning Designer resembles the steps on Azure Machine Learning Notebooks / python file. But we transform it to be low code mode, using some flow boxes.

Methods :

- Select AML Designer menu.
- Create new pipeline/designer projects
- Follow the steps as shown on figure



Nawatech

Thank you

Let’s discuss our collaborations

PT Nawa Darsana Teknologi
Gedung Office 8, 18th Floor, Jl. Jend Sudirman Kav 52-53
SCBD Lot 8, Senopati Jakarta Selatan 12190 DKI Jakarta, Indonesia
Website : www.nawatech.co
Phone : +62 21 29552754

Contact
Innes Sabrina Husa
Senior Business Development Manager
Email : windy@nawatech.co
Phone : +62 812 8033 2457