

**Explore Azure Machine Learning**

**Introduction**

In this lab, you will create a pipeline based on sample data using Azure Machine Learning Studio Designer.

The lab will begin with logging into Azure Machine Learning Studio – or signing up if you have not already done so. You will then create a Compute Instance and explore Azure Machine Learning Studio Designer by loading sample data and defining the features.

**Estimated Time**

60 minutes

**Objectives**

Upon completing this lab, you will be able to:

* Log into Azure Machine Learning Studio
* Create Azure ML Compute Instance
* Load Sample Data
* Define the features

Table of Contents

[Lab Title 1: Explore Azure Machine Learning 3](#_Toc87150893)

[Exercise 1: Access Azure Machine Learning Workspace 3](#_Toc87150894)

[Exercise 2: Create AzureML Compute Instance 4](#_Toc87150895)

[Exercise 3: Create a Pipeline using Designer 7](#_Toc87150896)

[Exercise 4: Load Data in the Pipeline 9](#_Toc87150897)

[Exercise 5: Define the Features and run the Pipeline 11](#_Toc87150898)

Lab Title 1: Explore Azure Machine Learning

In this lab, you will create a basic Azure Machine Learning Pipeline.

The lab will begin with logging into Azure Machine Learning Studio. You will then load a sample data and define features. The lab exercises must be completed in sequential order for this lab.

Exercise 1: Access Azure Machine Learning Workspace

This exercise shows how to log into Azure Machine Learning Studio.

Tasks

1. Open a web browser. (Microsoft Edge)
2. Go to <https://ml.azure.com/>
3. Signed in if asked

Exercise 1 has been completed

Exercise 2: Create AzureML Compute Instance

This exercise shows how to create AzureML Compute Instance to submit and run pipelines.

Tasks

1. On the main Azure ML Studio page, click Compute on the left, under the “Manage ” section:

Graphical user interface, application

Description automatically generated

1. To create a new compute instance, click “**+New**” under the **Compute instances** tab:

Graphical user interface, application

Description automatically generated

1. When you create a VM, follow these rules:

* Name is **required** and can't be empty.
* Name needs to be **unique** (in a case insensitive fashion) across all existing compute instances in the Azure region of the workspace/compute instance. You'll get an alert if the name you choose is not unique.
* Select the Virtual Machine size from the available choices. For this lab, the **default** **VM** is a good choice.

Graphical user interface, application

Description automatically generated

1. Then select **Create**. It can take approximately 5 minutes to set up your VM.

Exercise 2 has been completed

Exercise 3: Create a Pipeline using Designer

This exercise shows how to create a pipeline using designer. You will use pre=built components to create a sequence of actions.

Tasks

1. On the main Azure ML Studio page, click Designer on the left.

Graphical user interface, application

Description automatically generated

1. Create new blank pipeline

Graphical user interface, application

Description automatically generated

1. **Rename** your pipelineto **Intro to ML – Pipeline 1** by replacing the following text, “Pipeline-Created-on-<today’s date>”. See below.

Graphical user interface, text, application

Description automatically generated

Exercise 3 has been completed

Exercise 4: Load Data in the Pipeline

This exercise shows how to load data into the pipeline.

Tasks

1. Click arrow next to Sample Datasets
2. Drag and drop Automobile Price Data (RAW) to the middle of the pipeline canvas.

Graphical user interface, text, application, Word

Description automatically generated

1. Look at the data

Let’s look at the data. **Right click** **the circle on the bottom of the Auto price data (RAW) object** and choose **Preview Data** to see the data**.**

Graphical user interface, text, application

Description automatically generated

1. Examine the data

**Scroll to the right and examine the data**. Note the histogram for values in each column. Also note the total record count: **205**.

Graphical user interface, application

Description automatically generated

1. Close the window

Close the window for the visualization

Exercise 4 has been completed

Exercise 5: Define the Features and run the Pipeline

This exercise shows how to choose the features utilized for the pipelines prediction.

Tasks

1. Click the arrow sign next to Data Transformation.
2. Find Select Columns in Datasets Module and place it on the Pipeline Canvas under the Automobile price data (RAW).

Graphical user interface, application

Description automatically generated

1. Create an arrow starting from Automobile price data (RAW) to Select Columns in Dataset. Do this by clicking the circle on the bottom of the Auto price data (RAW) object and drag your mouse to the dot at the top of the Select Columns in Dataset object.

Diagram

Description automatically generated

1. Click the Select Columns in Dataset and click the Edit Columns menu item on the right-hand side of the pipeline canvas.

Graphical user interface, application

Description automatically generated

1. Change the selection mode from “With rules” to “By name”:

Graphical user interface, application

Description automatically generated

1. Select the following columns from the AVAILABLE COLUMNS section and click the plus sign for each of them to move them into the SELECTED COLUMNS section.

* Engine-type
* num-of-cylinders
* horsepower
* city-mpg
* highway-mpg
* price
* body-style
* engine-size

The columns are added to the right as shown below.

Graphical user interface, application, Word

Description automatically generated

1. Then click the Save button at the bottom of the Select columns window.
2. To Submit and run the pipeline, open settings at the top of the canvas. Then choose “Compute instance” in the “Select compute type” list box, and in the “Select Azure ML compute instance” list box, select the compute instance name you created previously in the “Create AzureML Compute Instance” exercise:

Graphical user interface

Description automatically generated

1. Click the Submit button at the top:

Graphical user interface, text, application, website

Description automatically generated

1. In the “Set up pipeline run” window, select “Create new” experiment and set the experiment name as Intro\_to\_ML\_Pipeline\_1. Then click Submit.

Graphical user interface, text, application, email

Description automatically generated

1. Let’s look at the data. Visualize the data in the Select Column in Dataset transformation. As you can see, only the data selected is shown.

Diagram

Description automatically generated

Graphical user interface

Description automatically generated with medium confidence

Now let’s split the data.

1. Click the arrow sign next to Data Transformation on the left side menu.
2. Find Split Data Module and Drag and Drop it to the Pipeline Canvas under the Select Columns in Dataset.

Graphical user interface

Description automatically generated

1. When training a Machine Learning model, it is common split your data set into 2 parts; 1 for training the algorithm, and 1 for testing. The ratio is somewhat subjective, but a good rule of thumb is to use 75% and 25% for training and testing the algorithm, respectively. Configure this in the Split Data tab of the Split Item object by changing the “Fraction of rows in the first dataset” from .5 to .75. This configures the Split Data step to send 75% of the data to Result dataset1. The remaining 25% is sent to Result dataset2.
2. Click the bottom circle on the Select Columns in Dataset object and drag it to the top circle of Dataset to Split Data object creating an arrow as shown below.

Graphical user interface

Description automatically generated

1. Click the Submit button. In the “Set up pipeline run” window, leave everything by default and click Submit:

Graphical user interface, text, application, email

Description automatically generated

You should see green “Completed” marks on each step once the Run has completed. This may take a minute or so.

Click the LEFT circle (Results dataset1) on the bottom of the Split Data step. Select “Preview Data”.

Diagram

Description automatically generated

1. Note that the number of rows is 154, which is approximately 75% of the total (205 rows).

Graphical user interface

Description automatically generated with low confidence

If you were to visualize the data from the RIGHT output (Results dataset2), you’d see the remaining 25% of the total rows from the data set.

Exercise 5 has been completed