

Week 1: Model Development using Azure ML Studio Designer

Agenda:

1. Setting up an ML pipeline using Designer (Azure ML)
 2. Executing the pipeline
 3. Scoring and Evaluating the model.

How to create a new resource group?

The screenshot shows the Microsoft Azure portal interface. At the top, there is a navigation bar with the Microsoft Azure logo, a search bar containing "Search resources, services, and docs (G+)", and various icons for notifications and account management. The user's email, "shambhavi.k.prasad@g...", and the default directory, "DEFAULT DIRECTORY (SHAMBH...)" are also visible.

The main content area is titled "Azure services". It features several quick access links: "Create a resource" (with a plus sign icon), "Resource groups" (with a hexagonal icon), "Subscriptions" (with a key icon), "Monitor" (with a gauge icon), "Quickstart Center" (with a rocket icon), "App Services" (with a globe icon), and "More services" (with a right-pointing arrow icon). A "Create" button within the "Resource groups" card is highlighted with a red box.

Below this, there is a section for "Free training from Microsoft" titled "Control and organize Azure resources with Azu..." which includes "8 units · 46 min".

The "Resources" section has tabs for "Recent" (selected) and "Favorite". Under "Recent", there are entries for "my-ml" (represented by a blue thumbs-up icon) and "Azure for Students" (represented by a yellow key icon). There is also a "See all" link.

The "Resource groups" card also contains a "Useful links" section with links to "Overview", "Get started", and "Documentation".

On the right side, there is a "Last Viewed" section showing "Machine Learning workspace" (last viewed 2 hours ago) and another entry (last viewed a day ago).

At the bottom left, there is a "Navigate" button.

How to create a new resource group?

Microsoft Azure Search resources, services, and docs (G+) ≡ ⋮ ? ? ? ? ? ? shambhavi.k.prasad@g...
DEFAULT DIRECTORY (SHAMBH...) 

Home > Resource groups >

Create a resource group ... X

[Basics](#) [Tags](#) [Review + create](#)

Resource group - A container that holds related resources for an Azure solution. The resource group can include all the resources for the solution, or only those resources that you want to manage as a group. You decide how you want to allocate resources to resource groups based on what makes the most sense for your organization. [Learn more](#)

Project details

Subscription * ⓘ ⌄

Resource group * ⓘ ✓

Resource details

Region * ⓘ ⌄

Review + create < Previous Next : Tags >

How to add a new resource to the resource group?

The screenshot shows the Microsoft Azure portal interface. The top navigation bar includes the Microsoft Azure logo, a search bar, and user information (shambhavi.k.prasad@g...). Below the navigation bar, the breadcrumb path indicates the current location: Home > Resource groups > my-first-resource-group.

The main content area displays the details of the 'my-first-resource-group'. On the left, there is a sidebar with navigation links: Create, Manage view, Filter for any field..., Name (sorted), my-first-resource-group (selected), Overview, Activity log, Access control (IAM), Tags, Resource visualizer, Events, Settings, Deployments, Security, Deployment stacks, Policies, and Properties.

The main pane shows the 'my-first-resource-group' blade. It features a 'Search' input field, a 'Create' button (which is highlighted with a red box), and various management options like 'Manage view', 'Delete resource group', 'Refresh', and 'Export to CSV'. The 'Essentials' section is expanded, showing the 'Resources' tab selected. It includes a 'Filter for any field...' input, a 'Type equals all' dropdown, an 'Add filter' button, and a 'More (1)' link. Below this, it shows 'Showing 0 to 0 of 0 records.' and filtering options for 'Show hidden types' and 'No grouping'. A table at the bottom lists resources with columns for Name, Type, and Location.

At the bottom left, there is a page navigation bar with icons for back, forward, and search, followed by the text 'Page 1 of 1'.

How to create a new ML resource?

Microsoft Azure Search resources, services, and docs (G+) shambhavi.k.prasad@g... DEFAULT DIRECTORY

Home > my-first-resource-group >

Marketplace

Get Started Service Providers Management Private Marketplace Private Offer Management My Marketplace Favorites My solutions Recently created Private plans Categories AI + Machine Learning (577) Analytics (286) IT & Management Tools (185)

azure machine learning Pricing : All Operating System : All Publisher Type : All Product Type : All Publisher name : All

Azure services only

Showing 1 to 20 of 913 results for 'azure machine learning'. [Clear search](#) Tile view

 Azure Machine Learning Microsoft Azure Service Enterprise-grade machine learning to build and deploy models faster Create 	 Data Science Virtual Machine - Ubuntu 20.04 Microsoft Virtual Machine Data Science Virtual Machine - Ubuntu 20.04 Create 	 KoçSistem Azure Machine Learning Management KoçSistem Bilgi ve İletişim Hizm... Managed Services Build, train and deploy machine learning models with Azure Machine Learning Services! Create 	 Data Science Virtual Machine - Windows 2019 Microsoft Virtual Machine Development and modeling tools for AI, data science and analytics Create 	 Azure CycleCloud microsoft-azure-cyclecloud - Mi... Virtual Machine Azure CycleCloud is a tool for orchestrating and managing HPC environments (Slurm, PBS, GridEngine, LSF, HPC Pack, etc.) on Azure. Create 
 Data Science Virtual Machine - Windows 2022 Machine Learning with Weka Create 	 Machine Learning with Weka Create 	 Weka Machine Learning Create 	 Jupyter Hub for Machine Learning using Python Create 	 Azure Synapse Analytics Create 

Is Marketplace helpful? 

How to create a new ML resource?

Microsoft Azure Search resources, services, and docs (G+/-) shambhavi.k.prasad@gmail.com DEFAULT DIRECTORY

Home > my-first-resource-group > Marketplace > Azure Machine Learning >

Azure Machine Learning

Create a machine learning workspace

Basics Networking Encryption Identity Tags Review + create

Resource details

Every workspace must be assigned to an Azure subscription, which is where billing happens. You use resource groups like folders to organize and manage resources, including the workspace you're about to create.

[Learn more about Azure resource groups](#)

Subscription * ①

Resource group * ① [Create new](#)

Workspace details

Configure your basic workspace settings like its storage connection, authentication, container, and more. [Learn more](#)

Name * ①

Region ①

Storage account * ① [Create new](#)

Key vault ① [Create new](#)

Review + create < Previous Next : Networking

How to create a new ML resource?

Microsoft Azure Search resources, services, and docs (G+) ≡ 1 ? ? shambhavi.k.prasad@g... DEFAULT DIRECTORY (SHAMBH...)

Dashboard > Resource groups > my-first-resource-group > Marketplace > Azure Machine Learning >

Azure Machine Learning ... X

Create a machine learning workspace

Workspace details

Configure your basic workspace settings like its storage connection, authentication, container, and more. [Learn more](#)

Name * my-ml-resource ✓

Region * Central India

Storage account * (new) mymlresource7031625296 ▼
[Create new](#)

Key vault * (new) mymlresource6277967018 ▼
[Create new](#)

Application insights * (new) mymlresource0771154178 ▼
[Create new](#)

Container registry None ▼
[Create new](#)

Review + create < Previous Next : Networking

How to create a new ML resource?

Microsoft Azure Search resources, services, and docs (G+ /) shambhavi.k.prasad@g... DEFAULT DIRECTORY

Home > Microsoft.MachineLearningServices | Overview X ...

 Microsoft.MachineLearningServices Deployment

 Search << Delete Cancel Redeploy Download Refresh

 Overview  Inputs  Outputs 

Deployment is in progress

Deployment name: Microsoft.MachineLearningServices
Subscription: Azure for Students
Resource group: my-first-resource-group

Start time: 2/9/2024, 1:29:02 PM
Correlation ID: 6afb29c8-6563-4ae2-96a2-686c07291f0f Copy

 Deployment details

Resource	Type	Status	Operation details
 myfirstmlresou9326381358	Microsoft.OperationalInsights/wor...	Created	Operation details

[Give feedback](#)  [Tell us about your experience with deployment](#) 

 Microsoft Defender for Cloud
Secure your apps and infrastructure [Go to Microsoft Defender for Cloud >](#)

[Free Microsoft tutorials](#)
[Start learning today >](#)

[Work with an expert](#)
Azure experts are service provider partners who can help manage your assets on Azure and be your first line of support. [Find an Azure expert >](#)

How to use the resource in a project?

Microsoft Azure Search resources, services, and docs (G/)

shambhavi.k.prasad@g...
DEFAULT DIRECTORY (SHAMBH...)

Home >

my-ml-resource ...

Azure Machine Learning workspace

Search Download config.json Delete

Storage mymlresource0564443470

Application insights mymlresource4543644848

MLflow tracking URI
<azureml://centralindia.api.azureml.ms/mlflow/v1.0/subscriptions/1f3dbf...>

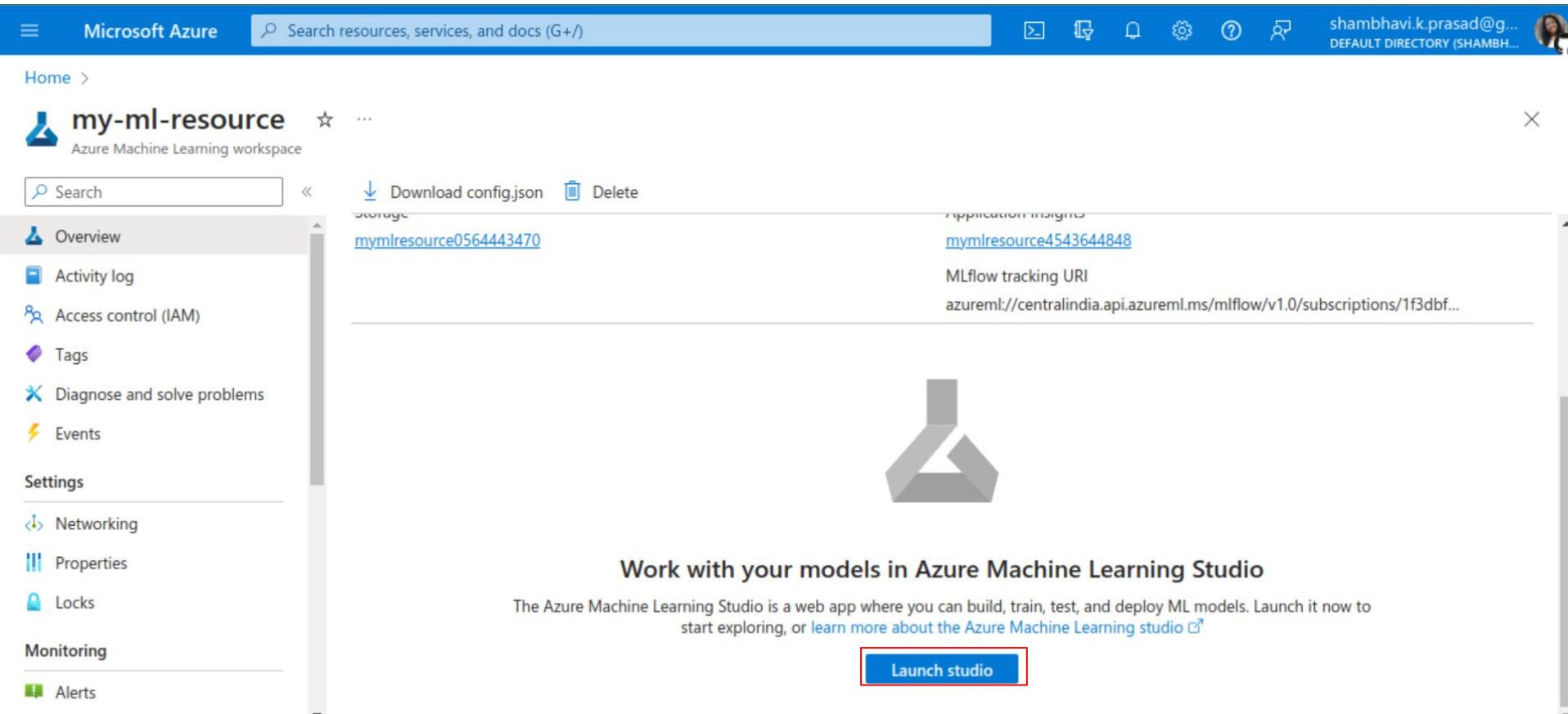
Overview Activity log Access control (IAM) Tags Diagnose and solve problems Events

Networking Properties Locks

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](#)

Launch studio



How to create a new ML job?

Azure AI | Machine Learning Studio

my-ml-resource

Generative AI with Prompt flow

Multi-Round Q&A on Your Data

Create a chatbot that uses LLM and data from your own indexed files to ground multi-round question and answering capabilities in enterprise chat scenarios.

Start Clone

Q&A on Your Data

Use LLM and data from your own indexed files to ground multi-round question and answering capabilities.

Start Clone

Web Classification

Use LLM to classify URLs into multiple categories.

Start Clone

Chat with Wikipedia

Create a chatbot that leverages Wikipedia data to ground the responses.

Start Clone

Use GPT

Learn how to extend the external data

Start

Generative AI models

mistralai-Mistral-7B-v01

Text generation

mistralai-Mixtral-8x7B-v01

Text generation

tiiuae-falcon-40b

Text generation

mistralai-Mistral-7B-Instru...

Chat completion

tiiuae

Notebook samples

View all

Customize view

+ New

View prompt flow

Authoring

- Notebooks
- Automated ML
- Designer
- Prompt flow

Assets

- Data
- Jobs
- Components
- Pipelines
- Environments
- Models
- Endpoints

View all

[View all](https://ml.azure.com/automl/welcome?wsid=/subscriptions/1f3dbf80-6d2b-44ea-abaa-172a959a26e0/resourcegroups/my-first-resource-group/providers/Microsoft.MachineLearningServices/workspaces/my-ml-resource&tid=6d0f22cb-a705-4de4-aa2e-31ffd2e16194)

How to create a new ML job in Designer?

Azure AI | Machine Learning Studio

Default Directory > my-resource > Designer

Designer

New pipeline

Classic prebuilt Custom Show more samples ▾

This low-code option uses existing prebuilt components and earlier dataset types (tabular, file), and is best suited for data processing and traditional machine learning tasks like regression and classification. This option continues to be supported but will not have any new components added.

Create a new pipeline using classic prebuilt components ⓘ Image Classification using DenseNet ⓘ Binary Classification using Vowpal Wabbit Model - ... ⓘ Wide & Deep based Recommendation - Rest... ⓘ Regression - Automobile Price Prediction (Basic) ⓘ

Pipelines

Pipeline drafts Pipeline jobs

Refresh Delete View options ▾

Name	Pipeline type	Updated on ↓	Created by

Search Filter Columns

All workspaces Home Model catalog Notebooks Automated ML Designer Prompt flow Data Jobs Components Pipelines Environments Models Endpoints

Azure for Students my-resource SP

<https://ml.azure.com/visualinterface?wsid=/subscriptions/1f3dbf80-6d2b-44ea-aba2-172a959a26e0/resourcegroups/my-first-resource-group/providers/Microsoft.MachineLearningServices/workspaces/my-resource&tid=6d0f22cb-a705-4de4-aa2e-31ffdd2e16194>

How to create a new ML job in Designer?

Azure AI | Machine Learning Studio

Default Directory > my-resource > Designer

Designer

New pipeline

Classic prebuilt Custom Show more samples ▾

This low-code option uses existing prebuilt components and earlier dataset types (tabular, file), and is best suited for data processing and traditional machine learning tasks like regression and classification. This option continues to be supported but will not have any new components added.



Create a new pipeline using classic prebuilt components ⓘ Image Classification using DenseNet ⓘ Binary Classification using Vowpal Wabbit Model - ... ⓘ Wide & Deep based Recommendation - Rest... ⓘ Regression - Automobile Price Prediction (Basic) ⓘ

Pipelines

Pipeline drafts Pipeline jobs Refresh Delete View options ▾

Name	Pipeline type	Updated on ↓	Created by

Search Filter Columns

How to add a new dataset?

Azure AI | Machine Learning Studio

Azure for Students
my-resource

Create data asset

1 Data type 2 Data source

Set the name and type for your data asset

Name * 

Description

Type * 

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 URLs. 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](#)

When should I use Table type?

The Table type is most useful for advanced scenarios where you might need to abstract the schema definition for easier sharing.

Back  Next 

How to add a new dataset?

Azure AI | Machine Learning Studio

Azure for Students
my-resource

Create data asset

1 Data type
2 Data source
3 Destination storage type
4 File or folder selection
5 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
Create a data asset by uploading files from your local drive. From local files

From web files
Create a data asset from a single file located at a public web URL.

Back Next Cancel

How to add a new dataset?

Azure AI | Machine Learning Studio

Azure for Students my-resource SP

Create data asset

1 Data type
2 Data source
3 Destination storage type
4 File or folder selection
5 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 Create new datastore

Search datastore

Name ↓	Storage name	Created on
workspaceblobstore	myresource5660592970	Feb 9, 2024 2:54 PM
workspaceartifactstore	myresource5660592970	Feb 9, 2024 2:54 PM

Filter Columns

Back Next Cancel

How to add a new dataset?

Create data asset

- Data type
- Data source
- Destination storage type
- 4** File or folder selection
- 5 Review

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/1f3dbf80-6d2b-44ea-aba2-172a959a26e0/resourcegroups/my-first-resource...



Upload files or folder

Overwrite if already exists

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

How to add a new dataset?

Azure AI | Machine Learning Studio

Azure for Students
my-resource

Create data asset

1 Data type
2 Data source
3 Destination storage type
4 File or folder selection
5 Review

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/1f3dbf80-6d2b-44ea-aba2-172a959a26e0/resourcegroups/my-first-resource...

Upload files or folder

Overwrite if already exists

Upload list

Admission_Predict.csv 11.46 KB/11.46 KB ...

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

How to add a new dataset?

Create data asset

- Data type
- Data source
- Destination storage type
- File or folder selection
- 5 Review

Review

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

Data type

Name

admission-predict

Description

Predict the admission of students in Masters program

Type

file

Data source

Type

Local

File selection

Upload path

azureml://subscriptions/1f3dbf80-6d2b-44ea-aba2-172a959a26e0/resourcegroups/my-first-resource-group/workspaces/my-resource/datastores/workspaceblobstore/paths/1f/2024-02-09_093634_UTC/Admission_Predict.csv

Storage

Datastore type

AzureBlob

Datastore name

workspaceblobstore

Back

Create

Cancel

How to add a new dataset?

Azure AI | Machine Learning Studio

DataOutput X

Number of files: 1 (sampled)

Path	File Na...	Modifie...	Created...	File Size	File For...
/UI/2024...	2024-02...	2024-02...	2024-02...	11.46 KiB	.csv

» Preview X

Display as grid With column headers

Serial No.	GRE Score	TOEFL Score	University ...	SOP	LOR	CGPA	Research	Chance of ...
1	337	118	4	4.5	4.5	9.65	1	1
2	324	107	4	4	4.5	8.87	1	1
3	316	104	3	3	3.5	8	1	0
4	322	110	3	3.5	2.5	8.67	1	1
5	314	103	2	2	3	8.21	0	0
6	330	115	5	4.5	3	9.34	1	1
7	321	109	3	3	4	8.2	1	1
8	308	101	2	3	4	7.9	0	0
9	302	102	1	2	1.5	8	0	0
10	323	108	3	3.5	3	8.6	0	0
11	325	106	3	3.5	4	8.4	1	0
12	327	111	4	4	4.5	9	1	1
13	328	112	4	4	4.5	9.1	1	1
14	307	109	3	4	3	8	1	0

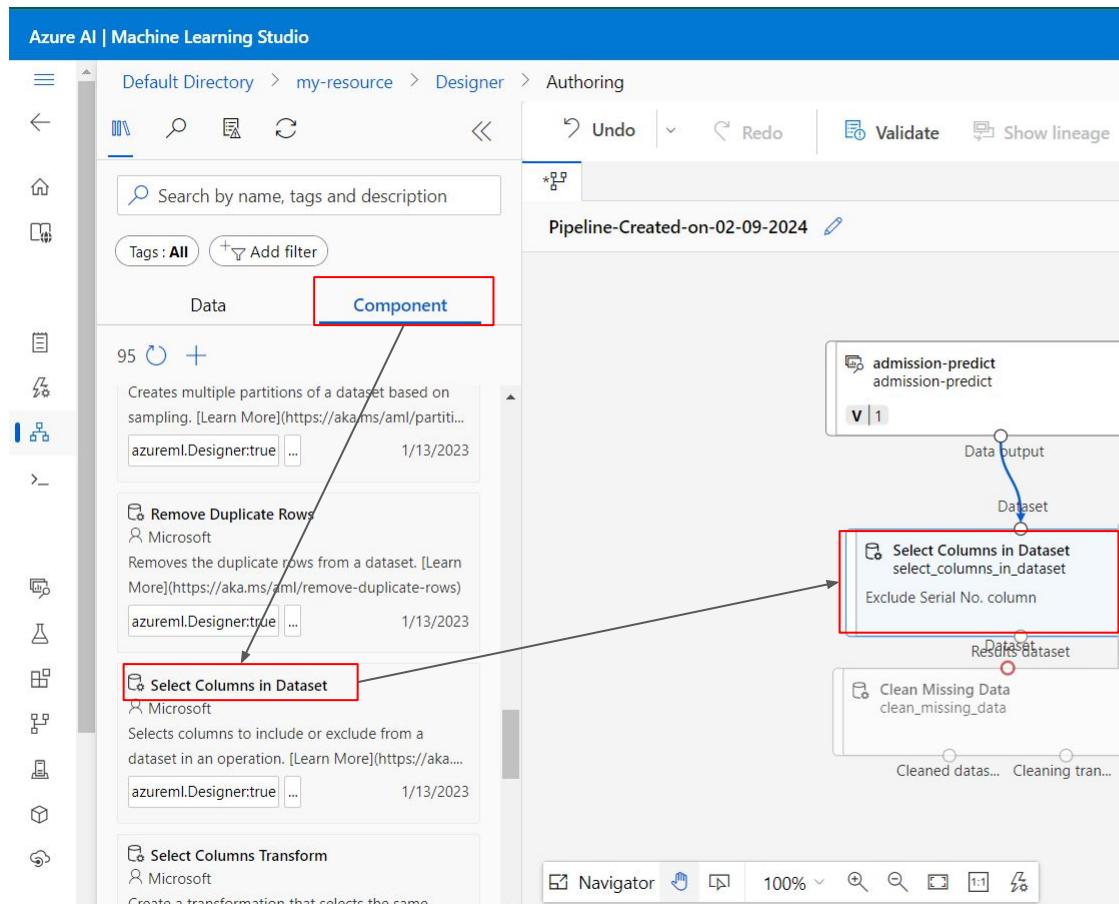
Azure for Students my-resource SP

How to build a model pipeline?

1. A simple pipeline could be as follows:
 - a. Data ingestion.
 - b. Selecting the columns to be used for model-building by removing redundant columns, like Serial Number.
 - c. Train-test split.
 - d. Data pre-processing (say, cleaning missing data.)
 - e. Training the algorithm on train dataset (say, decision tree classifier.)
 - f. Scoring the model by feeding the test dataset.
 - g. Evaluating the model.
2. Proceed to build the pipeline as shown. Populate different stages of the pipeline by selecting the different components from the left panel, labelled ‘Component’.

How to build a model pipeline?

1. After the dataset has been uploaded, create the pipeline using the Designer tab under Authoring section.
2. Add different components like dataset, select columns and clean missing data, by selecting the respective components from the 'Component' tab on the left of the Panel.
3. Proceed to build the pipeline as shown.



How to build a model pipeline?

The screenshot shows the Azure AI | Machine Learning Studio interface. In the center, a modal dialog titled "Select columns" is open. It has two tabs: "Select columns" (selected) and "With rules" (radio button selected). Below these are two dropdown menus: "Include" (set to "All columns") and "Exclude" (set to "Column names"). A single column name, "Serial No.", is listed under the exclude dropdown. The entire "Exclude" section is highlighted with a red box. At the bottom right of the dialog are "Save" and "Cancel" buttons.

1. If you have any redundant columns, you can use the “select columns” component to specify the same.

2. You can enhance your model pipeline by adding as many different components as needed.

How to build a model pipeline?

Azure AI | Machine Learning Studio Azure for Students my-resource

Default Directory > my-resource > Designer > Authoring

Pipeline-Created-on-02-09-2024

Configure & Submit

Search by name, tags and description

Tags : All Add filter

Data Component

Creates multiple partitions of a dataset based on sampling. [Learn More](https://aka.ms/aml/partition-dataset)

azureml.Designer:true ... 1/13/2023

Remove Duplicate Rows

Microsoft

Removes the duplicate rows from a dataset. [Learn More](https://aka.ms/aml/remove-duplicate-rows)

azureml.Designer:true ... 1/13/2023

Select Columns in Dataset

Microsoft

Selects columns to include or exclude from a dataset in an operation. [Learn More](https://aka.ms/aml/select-columns-in-dataset)

azureml.Designer:true ... 1/13/2023

Select Columns Transform

Microsoft

Create a transformation that selects the same

Undo Redo Validate Show lineage Clone AutoSave

Save Pipeline interface

Configure & Submit

Pipeline-Created-on-02-09-2024

admission-predict admission-predict

v | 1

Data output

Dataset

Select Columns in Dataset select_columns_in_dataset Exclude Serial No. column

Dataset

Clean Missing Data clean_missing_data

Cleaned data... Cleaning tran...

Navigator 100% 1:1

Configure & Submit

Select Columns in Dataset

All columns
Exclude column names: Serial No.

Output settings

Input settings

Run settings

Node information

Node name

select_columns_in_dataset

Comment

Exclude Serial No. column

Component information

For each component in the pipeline, you can add more information related to it, like a small description. It is not necessary, but good to have, for future reference.

How to build a model pipeline?

Azure AI | Machine Learning Studio Azure for Students my-resource

Default Directory > my-resource > Designer > Authoring

Undo Redo Validate Show lineage Clone AutoSave

Pipeline-Created-on-02-09-2024

Configure & Submit

Save Pipeline interface

```
graph TD; A[admission-predict] --> B[Select Columns in Dataset]; B --> C[Split Data]; C --> D[Clean Missing Data];
```

The screenshot shows the Azure Machine Learning Studio Designer interface. On the left, there's a sidebar with various icons. The main area displays a pipeline flow. At the top is a 'Pipeline-Created-on-02-09-2024' header. Below it, a red box highlights a 'Two-Class Boosted Decision Tree' component labeled 'two_class_boosted_decision_tree'. This component has a 'Data output' port connected to a 'Dataset' node. This dataset is then processed by a 'Select Columns in Dataset' component, which has an 'Exclude Serial No. column' setting. The output of this step is a 'Results dataset', which is then split into '75% train and 25% test datasets' by a 'Split Data' component. Finally, the training data is cleaned by a 'Clean Missing Data' component. The pipeline interface includes standard buttons like Undo, Redo, Validate, Show lineage, Clone, AutoSave, Configure & Submit, Save, and Pipeline interface.

How to build a model pipeline?

The screenshot shows the Azure AI | Machine Learning Studio interface. The top navigation bar includes 'Azure for Students' and 'my-resource'. The main area shows a file structure: Default Directory > my-resource > Designer > Authoring. The 'Component' tab is selected. A 'Label column' dialog box is open, with the title 'Label column' highlighted by a red box. Inside the dialog, there is a search bar 'Search by name, tags and description' and a 'Tags : All' button. Below it, a dropdown menu says 'Select a single column' and 'Column names'. The text 'Chance of Admit' is entered in this field and is also highlighted by a red box. At the bottom of the dialog are 'Save' and 'Cancel' buttons. To the right of the dialog, there is a 'Pipeline interface' section with a 'Save' button and an 'Edit column' link. A note on the left side of the screen states: '1. In the component specifying the training algorithm, please specify the target column.' and '2. As we have a classification problem here, we need to add a label column as the target column for our decision tree classifier.' Below this note is a bolded 'Note' section: 'Note: The column name is case-sensitive, so please pay attention while updating this field.'

Default Directory > my-resource > Designer > Authoring

Validate Show lineage Clone AutoSave

Configure & Submit

Save Pipeline interface

Label column

Search by name, tags and description

Tags : All Add filter

Select a single column Column names

Chance of Admit

Save Cancel

1. In the component specifying the **training algorithm**, please specify the target column.

2. As we have a classification problem here, we need to add a label column as the target column for our decision tree classifier.

Note: The column name is case-sensitive, so please pay attention while updating this field.

Component information

Model Pipeline

Azure AI | Machine Learning Studio Azure for Students my-resource

Default Directory > my-resource > Designer > Authoring

Undo Redo Validate Show lineage Clone AutoSave

Pipeline-Created-on-02-09-2024

Configure & Submit

Save Pipeline interface

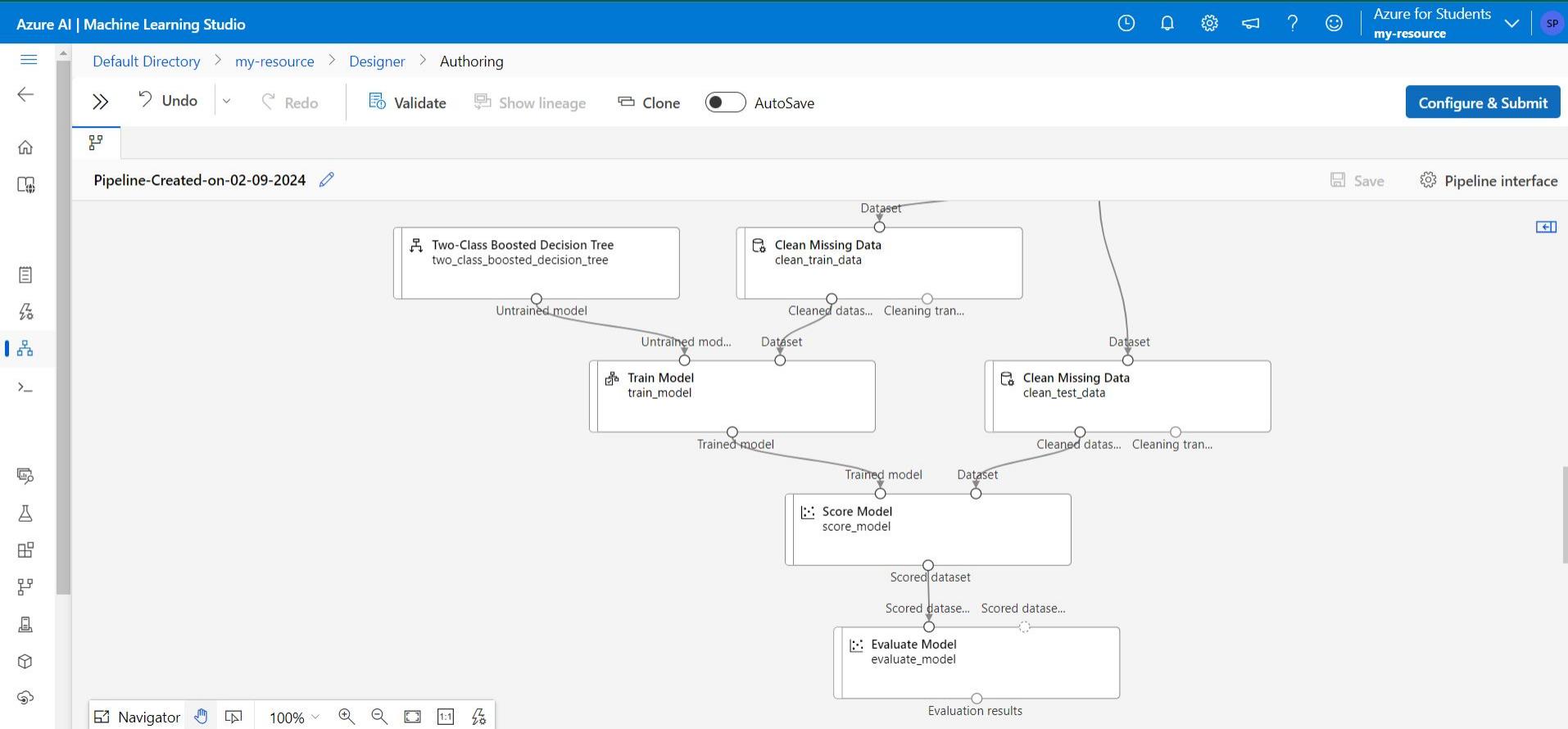
The screenshot shows a machine learning pipeline in the Azure Machine Learning Studio. The pipeline consists of the following steps:

- admission-predict**: A dataset input block.
- Select Columns in Dataset**: A step that excludes the "Serial No." column from the dataset.
- Split Data**: A step that splits the dataset into 75% train and 25% test datasets.
- Clean Missing Data**: A step that performs data cleaning on the training set.
- Two-Class Boosted Decision Tree**: A model training step.

The pipeline interface includes a sidebar with various icons for data management, a bottom navigation bar with links like Navigator, and a status bar at the bottom.

```
graph TD; A[admission-predict] --> B[Select Columns in Dataset]; B --> C[Split Data]; C --> D[Clean Missing Data]; D --> E[Two-Class Boosted Decision Tree]
```

Model Pipeline



How to run the model pipeline?

Azure AI | Machine Learning Studio Azure for Students my-resource

Default Directory > my-resource > Designer > Authoring

Configure & Submit

Pipeline-Created-on-02-09-2024

Save Pipeline interface

The screenshot shows a machine learning pipeline in the Azure AI Machine Learning Studio. The pipeline consists of the following steps:

- Two-Class Boosted Decision Tree** (two_class_boosted_decision_tree): An untrained model.
- Clean Missing Data** (clean_train_data): Takes a dataset and produces "Cleaned data..." and "Cleaning tran...".
- Train Model** (train_model): Takes an untrained model and a dataset, producing a "Trained model".
- Clean Missing Data** (clean_test_data): Takes a dataset and produces "Cleaned data..." and "Cleaning tran...".
- Score Model** (score_model): Takes a trained model and a dataset, producing a "Scored dataset".
- Evaluate Model** (evaluate_model): Takes a scored dataset and produces "Scored database..." and "Evaluation results".

The pipeline interface includes a toolbar with Undo, Redo, Validate, Show lineage, Clone, AutoSave, and a Configure & Submit button. The left sidebar contains various icons for managing datasets, models, and experiments. The bottom navigation bar includes links for Navigator, Search, and zoom controls.

How to run a model pipeline?

Azure AI | Machine Learning Studio

Default Directory > my-resource > Designer > Authoring

Pipeline-Created-on-02-09-2024

Two-Class Boosted Decision Tree
two_class_boosted_decision_tree

Untrained model

Train Model
train_model

Trained model

Set up pipeline job

Basics

Inputs & outputs

Runtime settings

Review + Submit

Runtime settings

Default compute

Please select a default compute to run a pipeline.

Select compute type

Compute instance

Compute instance

Compute cluster

Kubernetes compute

Attached compute

Select datastore *

workspaceblobstore

Advanced settings

Continue on step failure

Review + Submit

Back

Next

Close

How to run a model pipeline?

Azure AI | Machine Learning Studio Default Directory > my-resource > Designer > Authoring Undo Redo Validate Show lineage Clone ...

Set up pipeline job

Basics
Inputs & outputs
Runtime settings
Review + Submit

Runtime settings

Default compute ⓘ
Please select a default compute to run a pipeline.

Select compute type

Compute instance

Compute instance
Compute cluster
Kubernetes compute
Attached compute

Select datastore *
workspaceblobstore

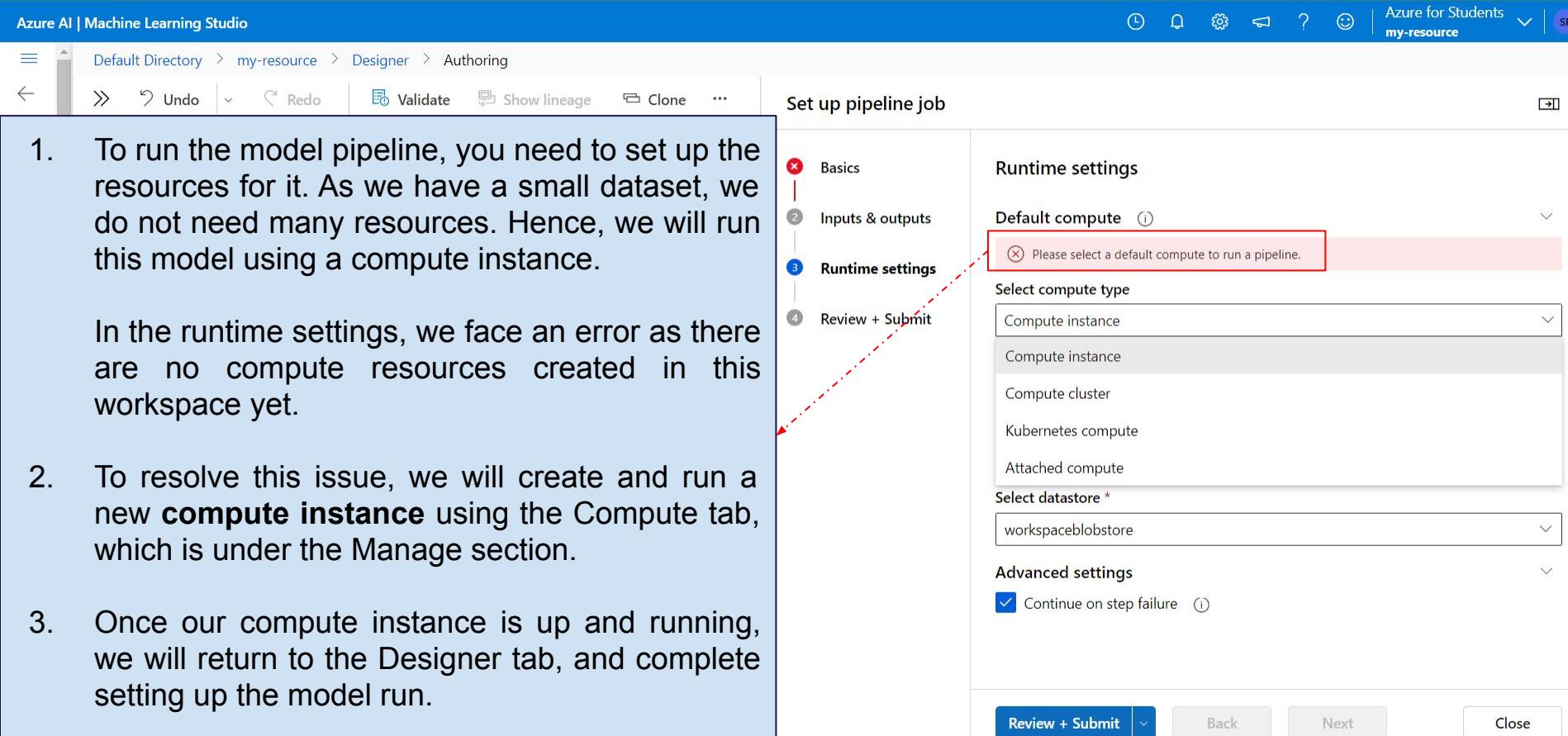
Advanced settings

Continue on step failure ⓘ

Review + Submit Back Next Close

1. To run the model pipeline, you need to set up the resources for it. As we have a small dataset, we do not need many resources. Hence, we will run this model using a compute instance.

In the runtime settings, we face an error as there are no compute resources created in this workspace yet.
2. To resolve this issue, we will create and run a new **compute instance** using the Compute tab, which is under the Manage section.
3. Once our compute instance is up and running, we will return to the Designer tab, and complete setting up the model run.



How to set up the compute resource?

Azure AI | Machine Learning Studio

Authoring Default Directory > my-resource > Compute

Compute

Choose from a selection of CPU or GPU instances preconfigured with popular tools such as VS Code, JupyterLab, Jupyter, and RStudio, ML packages, deep learning frameworks, and GPU drivers. [Learn more about compute instances](#)



Get started with Azure Machine Learning notebooks and R scripts by creating a compute instance

Choose from a selection of CPU or GPU instances preconfigured with popular tools such as VS Code, JupyterLab, Jupyter, and RStudio, ML packages, deep learning frameworks, and GPU drivers. [Learn more](#)

+ New

Assets Data, Jobs, Components, Pipelines, Environments, Models, Endpoints

Manage Compute (highlighted with a red box), Monitoring, Data Labeling, Linked Services (PREVIEW)

Azure for Students my-resource SP

How to set up the compute resource?

Azure AI | Machine Learning Studio

Auth

Create compute instance

Required settings

Scheduling optional

Security optional

Applications optional

Tags optional

Review

compute-instance-one

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	6 cores	\$0.19/hr
<input checked="" type="radio"/> Standard_DS3_v2 4 cores, 14GB RAM, 28GB storage	General purpose	Classical ML model training on small datasets	6 cores	\$0.34/hr
<input type="radio"/> Standard_E4ds_v4 4 cores, 32GB RAM, 150GB storage	Memory optimized	Data manipulation and training on medium-sized datasets (1-10GB)	4 cores	\$0.30/hr
<input type="radio"/> Standard_F4s_v2 4 cores, 8GB RAM, 32GB storage	Compute optimiz...	Data manipulation and training on large datasets (>10 GB)	16 cores	\$0.17/hr

Review + Create Back Next Cancel

How to set up the compute resource?

Azure AI | Machine Learning Studio

Auth

Create compute instance

Required settings

Scheduling optional

Security optional

Applications optional

Tags optional

Review

Review or make changes to your job before submission. Download a template for automation.

Required settings

Compute name: compute-instance-one

Virtual machine: Standard_DS3_v2

Virtual machine type: CPU

Scheduling

Auto shutdown enabled by default

Auto shutdown: After 60 minutes of inactivity

Start up and shutdown schedule: --

Security

Enable SSH: no

Enable managed identity: no

Create

Back

Cancel

Azure for Students
my-resource

SP

How to set up the compute resource?

The screenshot shows the Azure AI | Machine Learning Studio interface. The left sidebar has sections for Authoring, Assets, and Manage. Under Manage, the 'Compute' option is selected, indicated by a blue bar at the top. The main content area shows the 'Compute' page for the 'my-resource' directory. The 'Compute instances' tab is selected. At the top, there is a note about the 'Kubernetes clusters' tab. Below it are buttons for '+ New', 'Refresh' (which is highlighted with a red box), 'Start', 'Stop', 'Restart', 'Schedule and idle shutdown', 'Delete', 'View options', and 'View quota'. A search bar and filter/column options are also present. A table lists a single compute instance named 'compute-instance-one' with a status of 'Creating'. The bottom navigation bar includes 'PREVIEW'.

Azure AI | Machine Learning Studio

Authoring

- Notebooks
- Automated ML
- Designer
- Prompt flow

Assets

- Data
- Jobs
- Components
- Pipelines
- Environments
- Models
- Endpoints

Manage

- Compute
- Monitoring
- Data Labeling
- Linked Services

PREVIEW

Default Directory > my-resource > Compute

Compute

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.](#)

Compute instances Compute clusters Kubernetes clusters Attached computes

Choose from a selection of CPU or GPU instances preconfigured with popular tools such as VS Code, JupyterLab, Jupyter, and RStudio, ML packages, deep learning frameworks, and GPU drivers. [Learn more about compute instances](#)

+ New Refresh Start Stop Restart Schedule and idle shutdown Delete View options View quota

Name	State	Idle shutdown	Applications	Size	Created
compute-instance-one	Creating	--		STANDARD_DS3_V2	Feb 9,

How to set up the compute resource?

The screenshot shows the Azure AI | Machine Learning Studio interface. The top navigation bar includes 'Azure for Students' and 'my-resource'. The left sidebar has sections for 'Authoring', 'Assets', and 'Manage'. Under 'Manage', the 'Compute' section is selected, indicated by a blue bar. The main content area is titled 'Compute' and shows the 'Compute instances' tab selected. A note says: '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 are tabs for 'Compute instances', 'Compute clusters', 'Kubernetes clusters', and 'Attached computes'. A message encourages choosing from preconfigured CPU or GPU instances. The 'Compute instances' table lists one entry: 'compute-instance-one' (Status: Running). The 'Refresh' button in the toolbar is highlighted with a red box.

Azure AI | Machine Learning Studio

Authoring

Default Directory > my-resource > Compute

Compute

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.

Compute instances Compute clusters Kubernetes clusters Attached computes

Choose from a selection of CPU or GPU instances preconfigured with popular tools such as VS Code, JupyterLab, Jupyter, and RStudio, ML packages, deep learning frameworks, and GPU drivers. Learn more about compute instances

+ New Refresh Start Stop Restart Schedule and idle shutdown Delete View options View quota

Name	State	Idle shutdown	Applications	Size	Created
compute-instance-one	Running	1 hour	JupyterLab Jupyter VS Code (Web) PREVIEW ...	Standard_DS3_v2	Feb 9,

Search Filter Columns

Manage

Compute

Monitoring

Data Labeling

Linked Services PREVIEW

How to run a model pipeline?

Azure AI | Machine Learning Studio

Default Directory > my-resource > Designer > Authoring

Graph validation

1 errors detected

Select compute target in submission wizard.

Basics

Inputs & outputs

Runtime settings

Review + Submit

Set up pipeline job

Runtime settings

Default compute

Select compute type

Compute instance

Select Azure ML compute instance

compute-instance-one - Running

Create Azure ML compute instance

Refresh Compute

Default datastore

Select datastore *

workspaceblobstore

Advanced settings

Continue on step failure

Review + Submit

Back

Next

Close

Navigator

100%

undo

redo

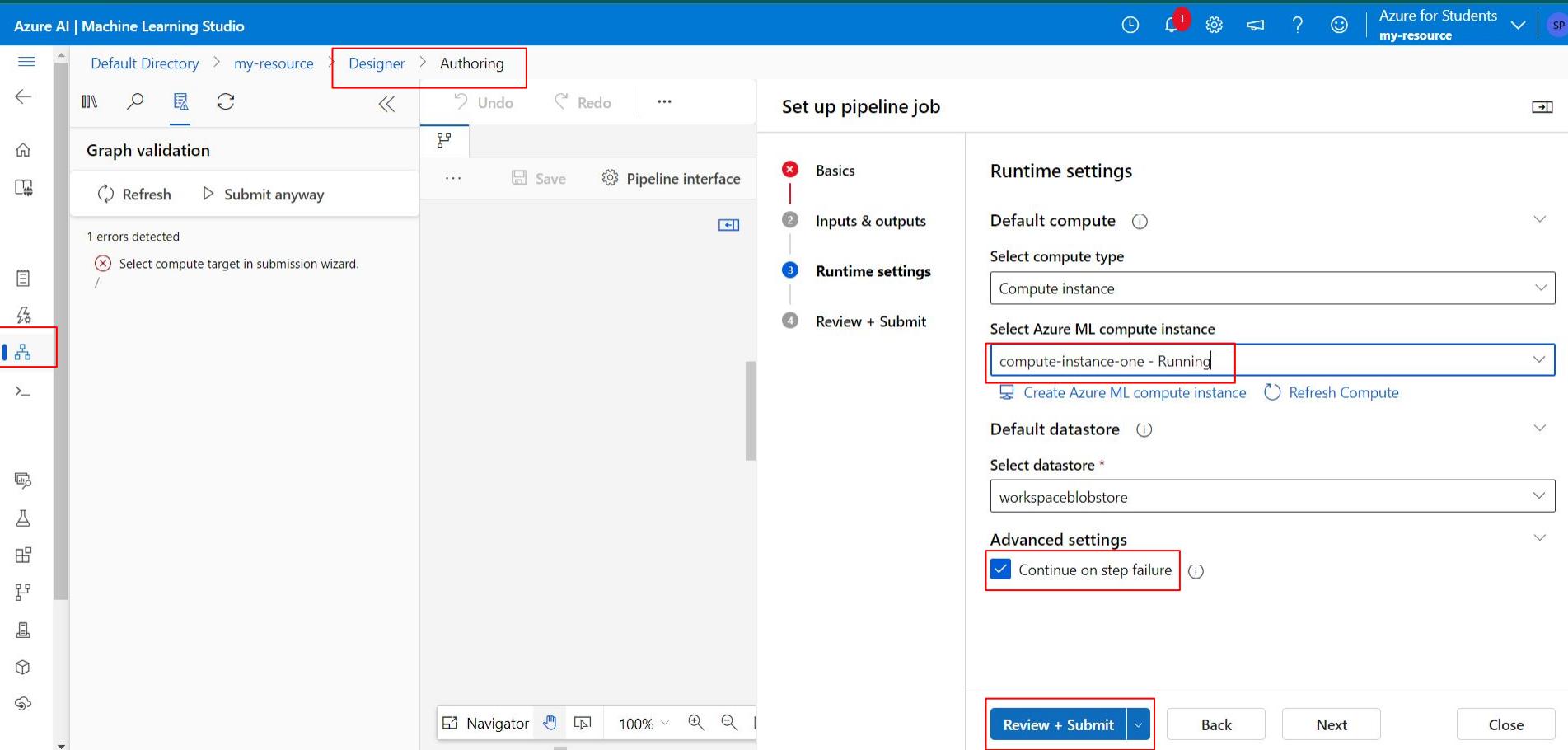
Graph validation icon (highlighted)

Authoring tab (highlighted)

Compute instance dropdown (highlighted)

Continue on step failure checkbox (highlighted)

Review + Submit button (highlighted)



How to run a model pipeline?

Azure AI | Machine Learning Studio

Default Directory > my-resource > Designer > Authoring

Graph validation

Refresh

Graph validation

Undo Redo ... Save Pipeline interface

The pipeline looks good, you can submit or publish now

Set up pipeline job

Basics
Inputs & outputs
Runtime settings
Review + Submit

Review + Submit

Basics

Job display name: Pipeline-Created-on-02-09-2024

Job description: Pipeline created on 20240209

Experiment: admission-prediction

Tags: No tags

Inputs & outputs

Inputs: AdmissionPredict

Data asset name:

Submit Back Next Close

The pipeline looks good, you can submit or publish now

Basics

Inputs & outputs

Runtime settings

Review + Submit

Job display name

Pipeline-Created-on-02-09-2024

Job description

Pipeline created on 20240209

Experiment

admission-prediction

Tags

No tags

Inputs

AdmissionPredict

Data asset name

Submit

Back

Next

Close

How to run a model pipeline?

Azure AI | Machine Learning Studio Azure for Students my-resource

Default Directory > my-resource > Designer > Authoring

Success: Pipeline job has been submitted. Loading...

Configure & Submit

Pipeline-Created-on-02-09-2024

The screenshot shows the Azure Machine Learning Studio Designer interface. At the top, there's a success message: "Success: Pipeline job has been submitted." Below the header, there are standard toolbar buttons for Undo, Redo, Validate, Show lineage, Clone, and AutoSave. On the right, there's a "Configure & Submit" button. The main workspace displays a pipeline flow starting with a dataset named "admission-predict". This dataset is processed by a "Select Columns in Dataset" component, which excludes the "Serial No." column. The output of this step is a "Results dataset", which is then split into training and test datasets by a "Split Data" component. The "Split Data" component splits the data into 75% train and 25% test datasets. Finally, the training dataset is passed to a "Clean Missing Data" component. The pipeline is saved with the name "Pipeline-Created-on-02-09-2024".

```
graph TD; admission_predict[admission-predict] --> select_columns[Select Columns in Dataset]; select_columns --> results_dataset[Results dataset]; results_dataset --> split_data[Split Data]; split_data --> clean_missing_data[Clean Missing Data];
```

Two-Class Boosted Decision Tree
two_class_boosted_decision_tree

Navigator 100% Cleaned data... Cleaning tran...

How to check the status of model run?

1. After you have started running the model, you can check its status by going to the Jobs tab under the Assets section.
2. Here, you will see all the jobs that have been submitted for execution, along with the respective experiment that they are a part of. In this case, the name of the experiment is ‘admission-prediction’, and the name of the job is the pipeline name that we have just created.
3. Under the Jobs tab, you can select your desired job, to view its progress. You can also see the related metrics of the model post its execution.
4. By clicking on the job name, it will redirect you to the Models tab, and show the component being run currently.

How to view progress of model run?

Azure AI | Machine Learning Studio

Default Directory > my-resource > Jobs

Jobs

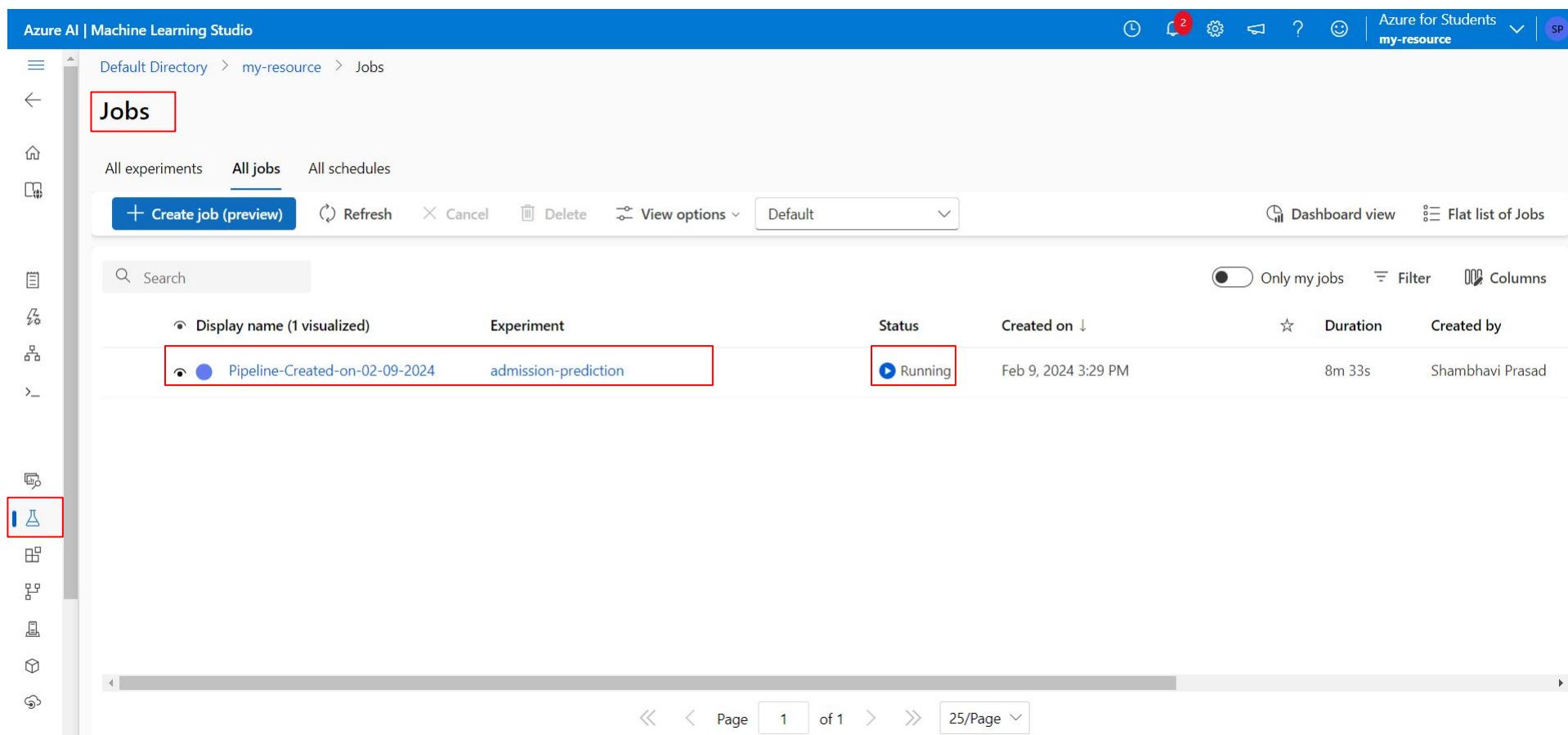
All experiments All jobs All schedules

+ Create job (preview) Refresh Cancel Delete View options Default Dashboard view Flat list of Jobs

Search Only my jobs Filter Columns

Display name (1 visualized)	Experiment	Status	Created on	Duration	Created by
Pipeline-Created-on-02-09-2024	admission-prediction	Running	Feb 9, 2024 3:29 PM	8m 33s	Shambhavi Prasad

Page 1 of 1 25/Page



How to view progress of model run?

Azure AI | Machine Learning Studio Azure for Students
Default Directory > my-resource > Jobs > admission-prediction > Pipeline-Created-on-02-09-2024

Refresh Clone Resubmit View profiling Publish Schedule Show lineage Show compare list Delete Share Add to compare Job overview

Pipeline-Created-on-02-09-2024 Running

The screenshot shows a machine learning pipeline named "admission-predict". It starts with a "Select Columns in Dataset" step, followed by a "Split Data" step, and then a "Clean Missing Data" step. The final step is a "Two-Class Boosted Decision Tree" component, which is highlighted with a red border. The status of the pipeline is "Running".

1. In the model pipeline, the solid green line indicates that the particular component has completed execution. A dashed line indicates that the process is still running.
2. At the end of a successful run, all the model components should have a solid green line.
3. You can refresh the model status at any time by clicking at the refresh button at the top-left.

How to view progress of model run?

Azure AI | Machine Learning Studio Azure for Students
Default Directory > my-resource > Jobs > admission-prediction > Pipeline-Created-on-02-09-2024

Pipeline-Created-on-02-09-2024 Running

```
graph TD; A[two_class_boosted_decision_tree] -- Untrained-model --> B[Train Model train_model]; B -- Trained model --> C[Score Model score_model]; C -- Scored dataset --> D[Evaluate Model evaluate_model]; E[clean_train_data] -- Dataset --> B; F[clean_missing_data clean_test_data] -- Dataset --> C;
```

1. Please note, the `score_model` component in this model takes two inputs: the trained model and the test dataset. We had cleaned the test dataset separately from the train dataset.

2. After scoring the model, we will evaluate it. This step completes the model development and execution process.

3. Please note, we will close this process here, as model deployment will be covered later, and is out of scope for the current case study.

How to view progress of model run?

Azure AI | Machine Learning Studio

Default Directory > my-resource > Jobs > admission-prediction > Pipeline-Created-on-02-09-2024

Pipeline-Created-on-02-09-2024 (Running)

two_class_boosted_decision_tree

clean_train_data

Train Model train_model

Clean Missing Data clean_test_data

Score Model score_model

Evaluate Model evaluate_model

Untrained model

Cleaned data... Cleaning tran...

Dataset

Trained model

Cleaned data... Cleaning tran...

Dataset

Scored dataset

Scored database... Scored database...

Scored dataset

Evaluation results

Refresh Clone Resubmit View profiling Publish Schedule Show lineage Show compare list Delete Share Add to compare Job overview

The screenshot shows a machine learning pipeline named "Pipeline-Created-on-02-09-2024" running in "my-resource". The pipeline consists of several steps: 1. An input dataset "two_class_boosted_decision_tree" is processed by "clean_train_data" to produce "Cleaned data... Cleaning tran...". 2. This leads to "Train Model train_model", which produces a "Trained model". 3. The "Trained model" is used by "Score Model score_model" to produce a "Scored dataset". 4. This "Scored dataset" is then evaluated by "Evaluate Model evaluate_model" to produce "Evaluation results". 5. A separate step "Clean Missing Data clean_test_data" processes a dataset to produce "Cleaned data... Cleaning tran...", which is then used by "Score Model score_model". The pipeline interface includes a left sidebar with various icons, a top navigation bar with "Azure for Students" and "my-resource" options, and a bottom toolbar with "Navigator", "Search", "Zoom", and "Help" buttons.

How to view the model predictions?

Azure AI | Machine Learning Studio

Default Directory > my-resource > Jobs > admission-prediction > Pipeline-Created-on-02-09-2024

Pipeline-Created-on-02-09-2024 (Completed)

Untrained-model → Cleaned dataset → Scored dataset → Evaluation results

Two-Class Boosted Decision Tree → Train Model → Score Model → Evaluate Model

Preview data (highlighted with a red box) → Scored dataset

```
graph TD; A[Two-Class Boosted Decision Tree] --> B[Train Model]; B --> C[Score Model]; C --> D[Evaluate Model]; E[Clean Missing Data] --> F[Dataset]; F --> G[Scored dataset]; G --> H[Evaluation results];
```

View log
Locate in outline
Add to compare
Show parameters
Copy shareable link to node
Preview data > Scored dataset
Access data >
Register data >

Navigator Refresh Clone Resubmit View profiling Publish Schedule Show lineage Show compare list Create inference pipeline Delete Cancel Share Add to compare Job overview

How to view the model predictions?

Azure AI | Machine Learning Studio

Scored_dataset

Rows 100 Columns 10

GRE Score	TOEFL Score	University Rating	SOP	LOR	CGPA	Research	Chance of Admit	Scored Labels	Scored Probabilities
317	107	2	3.5	3	8.28	0	0	0	0.000806
334	119	5	4.5	4.5	9.48	1	1	1	1
304	100	4	1.5	2.5	7.84	0	0	0	0.000001
325	114	3	3.5	3	9.04	1	1	0	0.259595
336	118	5	4.5	5	9.53	1	1	1	1
298	99	2	4	2	7.6	0	0	0	0.000001
334	120	5	4	5	9.87	1	1	1	1
310	106	2	3.5	2.5	8.33	0	0	0	0.000194
305	102	2	2	2.5	8.18	0	0	0	0
328	115	4	4.5	4	9.16	1	1	1	1
310	103	2	2.5	2.5	8.24	0	0	0	0.000024
309	108	3	2.5	3	8.12	0	0	0	0.000004
328	110	4	5	4	9.14	1	1	1	1
325	111	4	4	4.5	9.11	1	1	1	1
310	102	3	3.5	4	8.02	1	0	0	0.000638
300	102	2	1.5	2	7.87	0	0	0	0
324	112	4	4	2.5	8.1	1	0	1	0.624024
339	116	4	4	3.5	9.8	1	1	1	0.999999
319	112	3	2.5	2	8.71	1	1	0	0.000915

Scored Probabilities

Statistics

Mean	0.4521
Median	0.1658
Min	0
Max	1
Standard deviation	0.4735
Unique values	98
Missing values	0
Feature type	Numeric Score

Visualizations

A histogram titled "Scored Probabilities" showing the frequency distribution. The x-axis is labeled "Scored Probabilities" and ranges from 0 to 1. The y-axis is labeled "Frequency" and ranges from 0 to 40. There are two main peaks: one at approximately 0.15 with a frequency of about 45, and another at approximately 0.9 with a frequency of about 40.

How to view the model metrics?

Azure AI | Machine Learning Studio

Default Directory > my-resource > Jobs > admission-prediction > Pipeline-Created-on-02-09-2024

Pipeline-Created-on-02-09-2024 (Running)

two_class_boosted_decision_tree

clean_train_data

Train Model train_model

Clean Missing Data clean_test_data

Score Model score_model

Evaluate Model evaluate_model

Untrained model

Cleaned data... Cleaning tran...

Dataset

Trained model

Cleaned data... Cleaning tran...

Dataset

Scored dataset

Scored dataset... Scored dataset...

Evaluation results

Refresh Clone Resubmit View profiling Publish Schedule Show lineage Show compare list Delete Share Add to compare Job overview

```
graph TD; A[two_class_boosted_decision_tree] --> B[Train Model train_model]; A --> C[Clean Missing Data clean_test_data]; B --> D[Score Model score_model]; C --> D; D --> E[Evaluate Model evaluate_model]; E --> F[Evaluation results]
```

How to view the model metrics?

Azure AI | Machine Learning Studio

Default Directory > my-resource > Jobs > admission-prediction > Pipeline-Created-on-02-09-2024

Refresh Clone Resubmit View profiling Publish Schedule Show lineage Show compare list Create inference pipeline Delete Cancel

Pipeline-Created-on-02-09-2024 Completed

Share Add to compare Job overview

Evaluate Model

Metrics

Select metrics	Accuracy	AUC	F1 Score
	0.82	0.9184253	0.7954545
	Precision	Recall	0.7954545
	Confusion matrix		

How to view the model metrics?

Azure AI | Machine Learning Studio

Metrics

Evaluate Model

Overview Parameters Outputs + logs Metrics Child jobs Images Code Explanations (preview) Fairness (preview) Monitoring

Refresh Create custom chart View as... Current view: Local Edit view

Accuracy: 0.82

AUC: 0.9184253

F1 Score: 0.7954545

Precision: 0.7954545

Recall: 0.7954545

Confusion matrix: Chart visualization not available for non-numeric values.

Lift curve: Lift curve. Number of true positive vs Lift curve. Positive rate

The lift curve plot shows the cumulative distribution of true positives. The x-axis is labeled "Lift curve. Positive rate" and ranges from 0 to 1. The y-axis is labeled "Lift curve. Number of true positive" and ranges from 0 to 40. A blue line represents the lift curve, starting at (0,0) and rising to approximately (1, 42). A diagonal line from (0,0) to (1,1) represents a random classifier.

How to view the model metrics?

Azure AI | Machine Learning Studio

Evaluate Model

Metrics

Overview Parameters Outputs + logs Metrics Child jobs Images Code Explanations (preview) Fairness (preview) Monitoring

Refresh Create custom chart View as... Current view: Local Edit view

Lift curve.Positive rate

Precision-recall curve

ROC curve

Scored bins

3

Azure for Students my-resource SP

The screenshot shows the 'Evaluate Model' interface in Azure AI | Machine Learning Studio. The 'Metrics' tab is selected, highlighted with a red box. The left sidebar has a 'Select metrics' section with various icons. The main area displays three charts:

- Lift curve.** Positive rate vs. Lift curve.Number of true positive.
- Precision-recall curve.** Precision vs. Precision-recall curve.Precision.
- ROC curve.** True positive rate vs. ROC curve.False positive rate.

Each chart includes a legend and edit icons.

Closing the project - resource deletion

Microsoft Azure Search resources, services, and docs (G+) shambhavi.k.prasad@g... DEFAULT DIRECTORY

Home > my-first-resource-group

Search

Overview Activity log Access control (IAM) Tags Resource visualizer Events

Deployments Security Deployment stacks Policies Properties Locks

Cost analysis Cost alerts (preview) Budgets

Create Manage view Delete resource group Refresh Export to CSV Open quick

my-first-resource-group Resource group

Essentials Resources Recommendations

Filter for any field... Type equals all Location equals all Add filter

Showing 1 to 7 of 7 records. Show hidden types

Name	Type	Action
Application Insights Smart Detection	Action group	Remove
Failure Anomalies	Smart detection	Remove
my-resource	Azure Monitor	Remove
myresource4701204726	Key vault	Remove
myresource5660592970	Storage	Remove
myresource6049307251	Log Analytics workspace	Remove
myresource8567026174	Application Insights	Remove
myresource8567026174	Application Insights	Remove

< Previous Page 1 of 1 Next >

Delete Resources

The selected resources along with their related resources and contents will be permanently deleted. If you are unsure of the selected resource dependencies, navigate to the individual resource page to perform the delete operation. More details of the resource dependencies are available in the manage experience.

Resources to be deleted (7)

Name	Resource type	Action
Application Insights Smart Detection	Action group	Remove
Failure Anomalies	Smart detection	Remove
my-resource	Azure Monitor	Remove
myresource4701204726	Key vault	Remove
myresource5660592970	Storage	Remove
myresource6049307251	Log Analytics workspace	Remove
myresource8567026174	Application Insights	Remove

Delete confirmation

Deleting the selected resources and their internal data is a permanent action and cannot be undone.

Enter "delete" to confirm deletion *