

MLOPs Assignment Report

Problem Statement:

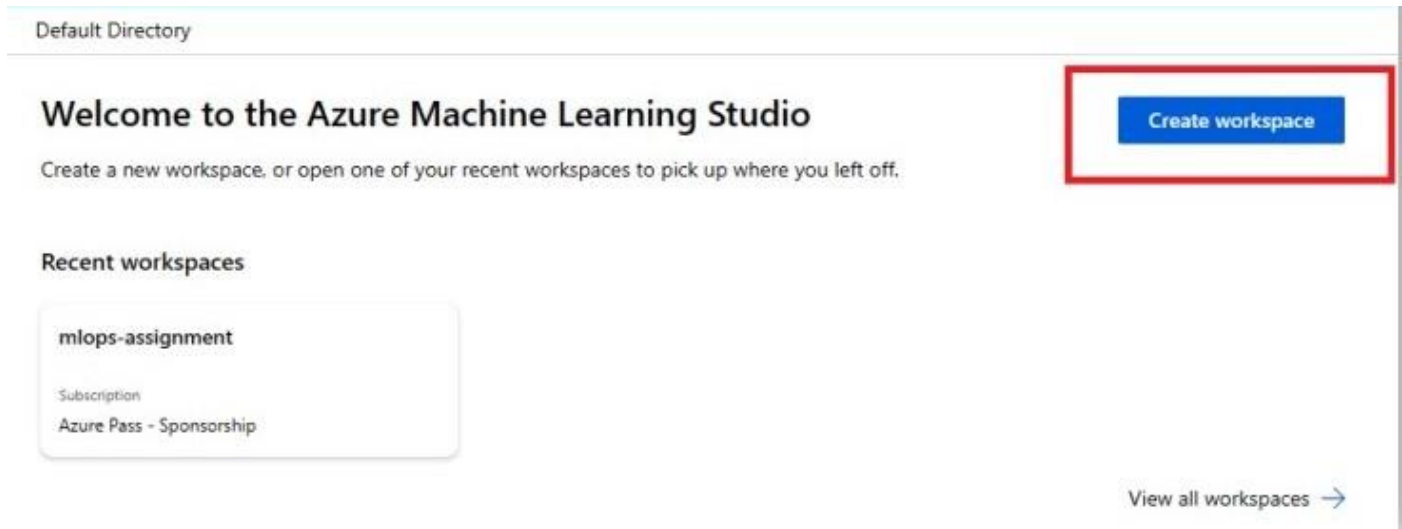
The company wants you to train this data set and deploy the best ML model using Azure ML tools.

Following are the tasks which are expected:

1. Go to Azure Portal. Create a Machine Learning Space to use Azure Machine Learning Model.
2. Upload Bank_Marketing data set (provided).
3. Train the uploaded data set using Azure Automated ML.
4. Finally, create a pipeline to run the model.

1. Go to Azure Portal. Create a Machine Learning Space to use Azure Machine Learning Model.

Have created a Machine Learning workspace by visiting, ml.azure.com, (See the attached screenshot)



Post, clicking on, “ Create Workspace ”, Workspace name, Resource group, and region selected.

The image shows a 'Create new workspace' modal window. It contains the following fields: 'Workspace name' with the value 'bankmarketing_mlops_assignment', 'Subscription' with a dropdown menu showing 'Azure Pass - Sponsorship', 'Resource group' with a dropdown menu showing 'er.prashantkrmali-rg', and 'Region' with a dropdown menu showing 'Central India'. There are also links for 'Refresh subscriptions' and 'Create new'. At the bottom right, there are two buttons: 'Create' (blue) and 'Cancel' (white with a grey border).

Workspace takes around 3-5 minutes to get finalized, however, to move forward, we have to do few things to create an experiment with shared Dataset (bankmarketing_train-2) using Azure Automated ML and pipeline – workflow.

- Create blob storage to upload the dataset
- Then, upload the Dataset,
- Create a compute instance
- Then, we can create an Azure Automated ML job with the uploaded Dataset
- We can also create a pipeline – workflow

2. Upload Bank_Marketing data-set (provided).

- a. Creating a data asset: By clicking on, Data asset under the **Asset section**

Create data asset

1 Data type

2 Data source

Set the name and type for your data asset

Name *

bankmarketing

Spelling correction

Open with the left-click (Alt+Down Arrow)

Data asset description

Type * i

Tabular

- b. Data Source: Data source selected as, **From local files**

Create data asset

✓ Data type

2 Data source

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 publicly available web location.

From Azure storage

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 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

- c. Select a datastore: I choose **Azure Blob Storage**

Create data asset

✓ Data type

✓ Data source

3 Storage type

4 File or folder selection

5 Settings

Select a datastore

Choose a storage type and a datastore from the list. You can also create a new datastore.

Datastore type *

Azure Blob Storage

Create new datastore

Search datastore

- d. Choose to upload files, by clicking **Upload >> Upload files >> Window pop-up**, to locate the file, then select

Create data asset

- ✓ Data type
- ✓ Data source
- ✓ Storage type
- 4 File or folder selection
- 5 Settings
- 6 Schema
- 7 Review

Choose a file or folder

Choose files or folders to upload from in a containing folder.

Upload path

azureml://subscriptions/e1f22667-dc9d-4b6e-99ed-4e22df4b30c8/resourcegroups

Upload

☐ Overwrite if already exists

Upload list

bankmarketing_train-2.csv	3.96 MB/3.96 MB
---------------------------	-----------------

- e. Preview of the uploaded Dataset appears and some setting related to it, I keep it, **Default**.

Settings

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

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

Column headers	Skip rows
All files have same headers	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.

- f. After **Schema**, I am at the **Review** part, where all the related details are visible, i.e., Data type, Data source with location, and related settings are shared.

After uploading the dataset, I created a compute instance to run the Automated ML job or to create pipe line work flow.

3. Train the uploaded data set using Azure Automated ML.

To create an Azure Automated ML job for the binary classification as the target column 'y' has 'yes' and 'no' to predict, I simply clicked on **Start Now** under the **Home tab** or try another way >> **Automated ML** under the **Authoring section** and select **New Automated ML job**.

However, I have done all the processes required and the job created and trained **59 Models** in a total of 1-hour duration. The final model is trained with 'VotingEnsemble'

I have attached all the relevant screenshots inside a folder named **MLOps_Assignment_Screenshots**.

Also, The Google drive link for the best model obtained with the Azure Automated ML:

<https://drive.google.com/drive/folders/1vHtm67OVdviubF2X-pms662hXN9o857Z?usp=sharing>

4. Finally, create a pipeline to run the model.

Using the **Designer**, I have created a designer pipeline – workflow to train the model for which screenshots are attached in the same folder.

- Then, I created, the Inference Pipeline >> Real-time pipeline and submitted it.
- After a few minutes, the pipeline was created and then, I deployed it.