

# Microsoft

## Exam Questions DP-100

Designing and Implementing a Data Science Solution on Azure



**NEW QUESTION 1**

- (Exam Topic 3)

Note: This question is part of a series of questions that present the same scenario. Each question in the series contains a unique solution that might meet the stated goals. Some question sets might have more than one correct solution, while others might not have a correct solution.

After you answer a question in this section, you will NOT be able to return to it. As a result, these questions will not appear in the review screen.

You create a model to forecast weather conditions based on historical data.

You need to create a pipeline that runs a processing script to load data from a datastore and pass the processed data to a machine learning model training script.

Solution: Run the following code:

```
datastore = ws.get_default_datastore()
data_output = PipelineData("processed_data", datastore=datastore)
process_step = PythonScriptStep(script_name="process.py",
    arguments=["--data_for_train", data_output],
    outputs=[data_output], compute_target=aml_compute,
    source_directory=process_directory)
pipeline = Pipeline(workspace=ws, steps=[process_step])
```

Does the solution meet the goal?

- A. Yes
- B. No

**Answer: B**

**Explanation:**

train\_step is missing. Reference:

<https://docs.microsoft.com/en-us/python/api/azureml-pipeline-core/azureml.pipeline.core.pipelinedata?view=azu>

**NEW QUESTION 2**

- (Exam Topic 3)

You have a dataset that includes confidential data. You use the dataset to train a model.

You must use a differential privacy parameter to keep the data of individuals safe and private. You need to reduce the effect of user data on aggregated results.

What should you do?

- A. Decrease the value of the epsilon parameter to reduce the amount of noise added to the data
- B. Increase the value of the epsilon parameter to decrease privacy and increase accuracy
- C. Decrease the value of the epsilon parameter to increase privacy and reduce accuracy
- D. Set the value of the epsilon parameter to 1 to ensure maximum privacy

**Answer: C**

**Explanation:**

Differential privacy tries to protect against the possibility that a user can produce an indefinite number of reports to eventually reveal sensitive data. A value known as epsilon measures how noisy, or private, a report is. Epsilon has an inverse relationship to noise or privacy. The lower the epsilon, the more noisy (and private) the data is.

Reference:

<https://docs.microsoft.com/en-us/azure/machine-learning/concept-differential-privacy>

**NEW QUESTION 3**

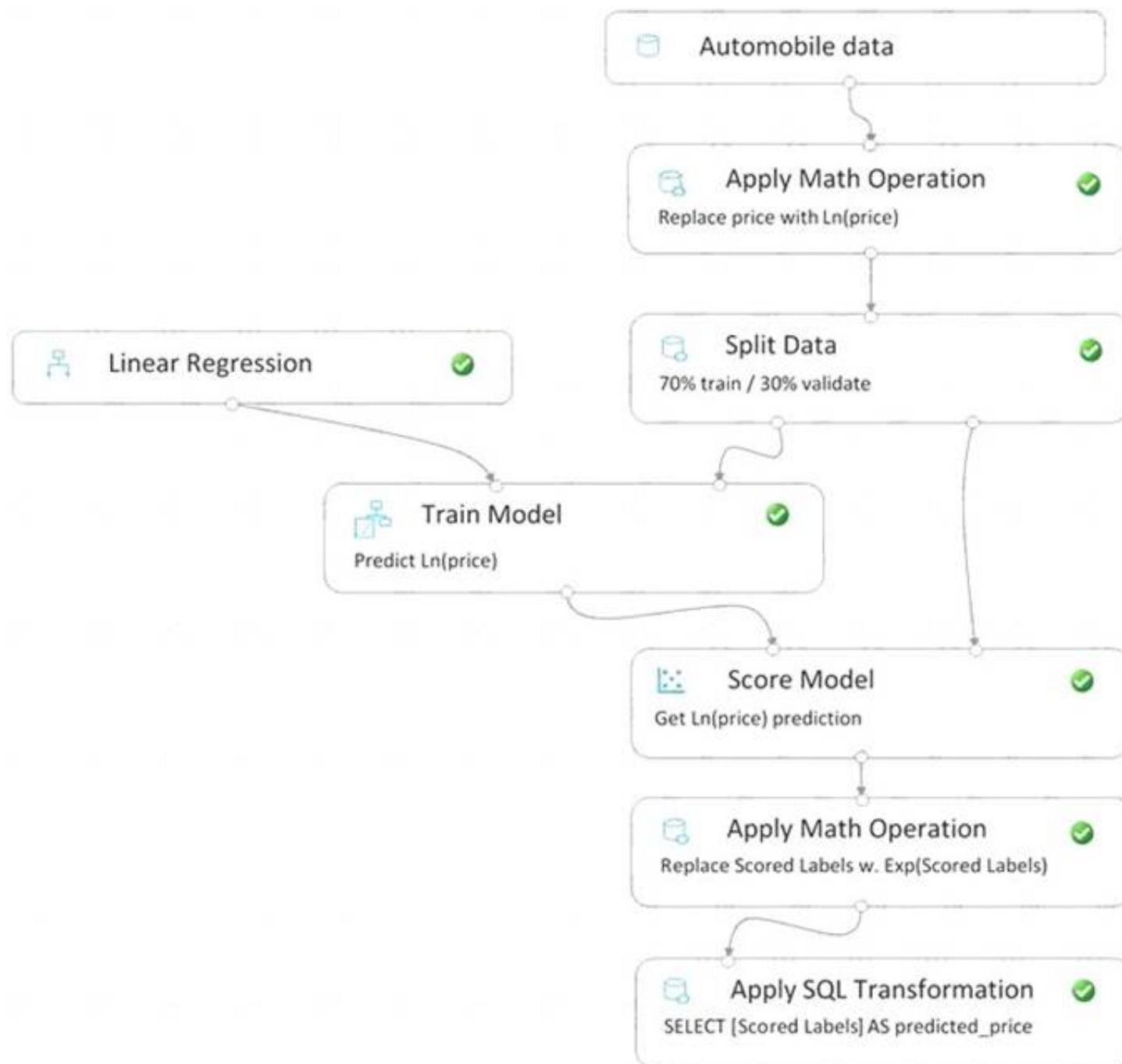
- (Exam Topic 3)

You create a pipeline in designer to train a model that predicts automobile prices.

Because of non-linear relationships in the data, the pipeline calculates the natural log (Ln) of the prices in the training data, trains a model to predict this natural log of price value, and then calculates the exponential of the scored label to get the predicted price.

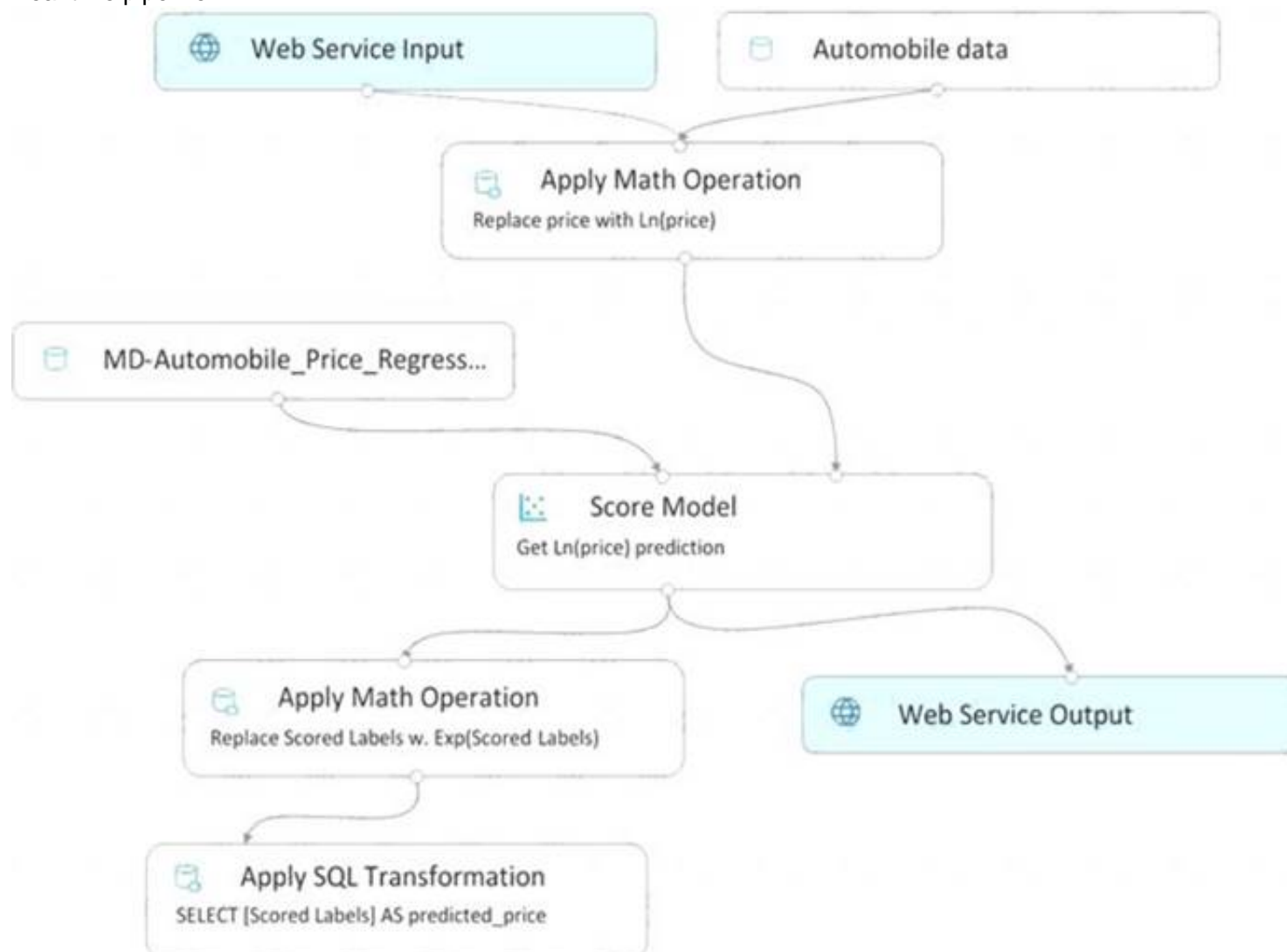
The training pipeline is shown in the exhibit. (Click the Training pipeline tab.)

Training pipeline



You create a real-time inference pipeline from the training pipeline, as shown in the exhibit. (Click the Real-time pipeline tab.)

Real-time pipeline



You need to modify the inference pipeline to ensure that the web service returns the exponential of the scored label as the predicted automobile price and that client applications are not required to include a price value in the input values.

Which three modifications must you make to the inference pipeline? Each correct answer presents part of the solution.

NOTE: Each correct selection is worth one point.

- A. Connect the output of the Apply SQL Transformation to the Web Service Output module.
- B. Replace the Web Service Input module with a data input that does not include the price column.
- C. Add a Select Columns module before the Score Model module to select all columns other than price.
- D. Replace the training dataset module with a data input that does not include the price column.
- E. Remove the Apply Math Operation module that replaces price with its natural log from the data flow.

F. Remove the Apply SQL Transformation module from the data flow.

**Answer:** ACE

#### NEW QUESTION 4

- (Exam Topic 3)

You are creating a classification model for a banking company to identify possible instances of credit card fraud. You plan to create the model in Azure Machine Learning by using automated machine learning.

The training dataset that you are using is highly unbalanced. You need to evaluate the classification model.

Which primary metric should you use?

- A. normalized\_mean\_absolute\_error
- B. [spearman\_correlation
- C. AUC.weighted
- D. accuracy
- E. normalized\_root\_mean\_squared\_error

**Answer:** C

#### Explanation:

AUC\_weighted is a Classification metric.

Note: AUC is the Area under the Receiver Operating Characteristic Curve. Weighted is the arithmetic mean of the score for each class, weighted by the number of true instances in each class.

Reference:

<https://docs.microsoft.com/en-us/azure/machine-learning/how-to-understand-automated-ml>

#### NEW QUESTION 5

- (Exam Topic 3)

You are developing a deep learning model by using TensorFlow. You plan to run the model training workload on an Azure Machine Learning Compute Instance.

You must use CUDA-based model training. You need to provision the Compute Instance.

Which two virtual machines sizes can you use? To answer, select the appropriate virtual machine sizes in the answer area.

NOTE: Each correct selection is worth one point.

#### Virtual machine size

Name ↑	vCPUs	GPUs	RAM	Resource disk
BASIC_A0	1		0.75 GB	20 GB
STANDARD_D3_V2	4		14 GB	200 GB
STANDARD_E64_V3	64		432 GB	1,600 GB
STANDARD_M64LS	64		512 GB	2,000 GB
STANDARD_NC12	12	2	112 GB	680 GB
STANDARD_NC24	24	4	224 GB	1,440 GB

- A. Mastered
- B. Not Mastered

**Answer:** A

#### Explanation:

CUDA is a parallel computing platform and programming model developed by Nvidia for general computing on its own GPUs (graphics processing units). CUDA enables developers to speed up compute-intensive applications by harnessing the power of GPUs for the parallelizable part of the computation.

Reference:

<https://www.infoworld.com/article/3299703/what-is-cuda-parallel-programming-for-gpus.html>

#### NEW QUESTION 6

- (Exam Topic 3)

Your Azure Machine Learning workspace has a dataset named real\_estate\_data. A sample of the data in the dataset follows.

postal_code	num_bedrooms	sq_feet	garage	price
12345	3	1300	0	23,9000
54321	1	950	0	11,0000
12346	2	1200	1	15,0000

You want to use automated machine learning to find the best regression model for predicting the price column. You need to configure an automated machine learning experiment using the Azure Machine Learning SDK. How should you complete the code? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.

## Answer Area

```
from azureml.core import Workspace
from azureml.core.compute import ComputeTarget
from azureml.core.runconfig import RunConfiguration
from azureml.train.automl import AutoMLConfig

ws = Workspace.from_config()
training_cluster = ComputeTarget(workspace=ws, name= 'aml-cluster1')
real_estate_ds = ws.datasets.get('real_estate_data')
split1_ds, split2_ds = real_estate_ds.random_split(percentage=0.7, seed=123)
automl_run_config = RunConfiguration(framework= "python")
automl_config = AutoMLConfig(
    task= 'regression',
    compute_target= training_cluster,
    run_configuration=automl_run_config,
    primary_metric='r2_score',
     =split1_ds,
     =split2_ds,
     ='price')

```

X  
Y  
X\_valid  
Y\_valid  
training\_data

X  
Y  
X\_valid  
Y\_valid  
validation\_data  
training\_data

y  
y\_valid  
y\_max  
label\_column\_name  
exclude\_nan\_labels

- A. Mastered  
 B. Not Mastered

**Answer: A**

### Explanation:

Box 1: training\_data

The training data to be used within the experiment. It should contain both training features and a label column (optionally a sample weights column). If training\_data is specified, then the label\_column\_name parameter must also be specified.

Box 2: validation\_data

Provide validation data: In this case, you can either start with a single data file and split it into training and validation sets or you can provide a separate data file for the validation set. Either way, the validation\_data parameter in your AutoMLConfig object assigns which data to use as your validation set.

Example, the following code example explicitly defines which portion of the provided data in dataset to use for training and validation.

```
dataset = Dataset.Tabular.from_delimited_files(data)
```

```
training_data, validation_data = dataset.random_split(percentage=0.8, seed=1) automl_config = AutoMLConfig(compute_target = aml_remote_compute, task = 'classification',
```

```
primary_metric = 'AUC_weighted', training_data = training_data,
```

```
validation_data = validation_data, label_column_name = 'Class'
```

```
)
```

Box 3: label\_column\_name label\_column\_name:

The name of the label column. If the input data is from a pandas.DataFrame which doesn't have column names, column indices can be used instead, expressed as integers.

This parameter is applicable to training\_data and validation\_data parameters. Reference:

<https://docs.microsoft.com/en-us/python/api/azureml-train-automl-client/azureml.train.automl.automlconfig.auto>

### NEW QUESTION 7





- (Exam Topic 3)

You have a dataset that contains over 150 features. You use the dataset to train a Support Vector Machine (SVM) binary classifier.

You need to use the Permutation Feature Importance module in Azure Machine Learning Studio to compute a set of feature importance scores for the dataset.

In which order should you perform the actions? To answer, move all actions from the list of actions to the answer area and arrange them in the correct order.



Actions	Answer Area
Add a Two-Class Support Vector Machine module to initialize the SVM classifier.	
Set the Metric for measuring performance property to <b>Classification - Accuracy</b> and then run the experiment.	
Add a Permutation Feature Importance module and connect the trained model and test dataset.	 
	 
Add a dataset to the experiment.	
Add a Split Data module to create training and test datasets.	

- A. Mastered
- B. Not Mastered

**Answer:** A

**Explanation:**

Step 1: Add a Two-Class Support Vector Machine module to initialize the SVM classifier.

Step 2: Add a dataset to the experiment

Step 3: Add a Split Data module to create training and test dataset.

To generate a set of feature scores requires that you have an already trained model, as well as a test dataset. Step 4: Add a Permutation Feature Importance module and connect to the trained model and test dataset. Step 5: Set the Metric for measuring performance property to Classification - Accuracy and then run the experiment.

Reference:

<https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/two-class-support-vector-mac> <https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/permutation-feature-importan>

**NEW QUESTION 8**

- (Exam Topic 3)

You use the designer to create a training pipeline for a classification model. The pipeline uses a dataset that includes the features and labels required for model training.

You create a real-time inference pipeline from the training pipeline. You observe that the schema for the generated web service input is based on the dataset and includes the label column that the model predicts. Client applications that use the service must not be required to submit this value.

You need to modify the inference pipeline to meet the requirement. What should you do?

- A. Add a Select Columns in Dataset module to the inference pipeline after the dataset and use it to select all columns other than the label.
- B. Delete the dataset from the training pipeline and recreate the real-time inference pipeline.
- C. Delete the Web Service Input module from the inference pipeline.
- D. Replace the dataset in the inference pipeline with an Enter Data Manually module that includes data for the feature columns but not the label column.

**Answer:** A

**Explanation:**

By default, the Web Service Input will expect the same data schema as the module output data which connects to the same downstream port as it. You can remove the target variable column in the inference pipeline using Select Columns in Dataset module. Make sure that the output of Select Columns in Dataset removing target variable column is connected to the same port as the output of the Web Service Input module.

Reference:

<https://docs.microsoft.com/en-us/azure/machine-learning/tutorial-designer-automobile-price-deploy>

**NEW QUESTION 9**

- (Exam Topic 3)

You are running a training experiment on remote compute in Azure Machine Learning.

The experiment is configured to use a conda environment that includes the mlflow and azureml-contrib-run packages.

You must use MLflow as the logging package for tracking metrics generated in the experiment. You need to complete the script for the experiment.

How should you complete the code? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.

```
import numpy as np
# Import library to log metrics
```

```
from azureml.core import Run
import mlflow
import logging
```

```
# Start logging for this run
```

```
run = Run.get_context()
mlflow.start_run()
logger = logging.getLogger('Run')
reg_rate = 0.01
# Log the reg_rate metric
```

```
run.log('reg_rate', np.float(reg_rate))
mlflow.log_metric('reg_rate', np.float(reg_rate))
logger.info(np.float(reg_rate))
```

```
# Stop logging for this run
```

```
run.complete()
mlflow.end_run()
logger.setLevel(logging.INFO)
```

- A. Mastered
- B. Not Mastered

**Answer:** A

**Explanation:**

Box 1: import mlflow

Import the mlflow and Workspace classes to access MLflow's tracking URI and configure your workspace. Box 2: mlflow.start\_run()

Set the MLflow experiment name with set\_experiment() and start your training run with start\_run(). Box 3: mlflow.log\_metric('..')

Use log\_metric() to activate the MLflow logging API and begin logging your training run metrics. Box 4: mlflow.end\_run()

Close the run: run.endRun() Reference:

<https://docs.microsoft.com/en-us/azure/machine-learning/how-to-use-mlflow>

**NEW QUESTION 10**

- (Exam Topic 3)

You run an experiment that uses an AutoMLConfig class to define an automated machine learning task with a maximum of ten model training iterations. The task will attempt to find the best performing model based on a metric named accuracy.

You submit the experiment with the following code:

You need to create Python code that returns the best model that is generated by the automated machine learning task. Which code segment should you use?

A)

```
best_model = automl_run.get_details()
```

B)

```
best_model = automl_run.get_output()[1]
```

C)

```
best_model = automl_run.get_file_names()[1]
```

D)

```
best_model = automl_run.get_metrics()
```

- A. Option A
- B. Option B
- C. Option C
- D. Option D

**Answer:** B

**Explanation:**

The get\_output method returns the best run and the fitted model. Reference:

<https://notebooks.azure.com/azureml/projects/azureml-getting-started/html/how-to-use-azureml/automated-mach>

**NEW QUESTION 10**

- (Exam Topic 3)

You create an Azure Machine Learning workspace.

You must create a custom role named DataScientist that meets the following requirements:

- Role members must not be able to delete the workspace.
- Role members must not be able to create, update, or delete compute resource in the workspace.
- Role members must not be able to add new users to the workspace.

You need to create a JSON file for the DataScientist role in the Azure Machine Learning workspace. The custom role must enforce the restrictions specified by the IT Operations team.

Which JSON code segment should you use?

A)

```
{
  "Name": "DataScientist",
  "IsCustom": true,
  "Description": "Project Data Scientist role",
  "Actions": ["*"],
  "NotActions": [
    "Microsoft.MachineLearningServices/workspaces/*/delete",
    "Microsoft.MachineLearningServices/workspaces/computes/*/write",
    "Microsoft.MachineLearningServices/workspaces/computes/*/delete",
    "Microsoft.Authorization/*/write"
  ],
  "AssignableScopes": [
    "/subscriptions/<id>/resourceGroups/ml-rg/providers/Microsoft.MachineLearningServices/workspaces/ml-ws"
  ]
}
```

B)

```
{
  "Name": "DataScientist",
  "IsCustom": true,
  "Description": "Project Data Scientist role",
  "Actions": ["*"],
  "NotActions": [],
  "AssignableScopes": [
    "/subscriptions/<id>/resourceGroups/ml-rg/providers/Microsoft.MachineLearningServices/workspaces/ml-ws"
  ]
}
```

C)

```
{
  "Name": "DataScientist",
  "IsCustom": true,
  "Description": "Project Data Scientist role",
  "Actions": [
    "Microsoft.MachineLearningServices/workspaces/*/delete",
    "Microsoft.MachineLearningServices/workspaces/computes/*/write",
    "Microsoft.MachineLearningServices/workspaces/computes/*/delete",
    "Microsoft.Authorization/*/write"
  ],
  "NotActions": [],
  "AssignableScopes": [
    "/subscriptions/<id>/resourceGroups/ml-rg/providers/Microsoft.MachineLearningServices/workspaces/ml-ws"
  ]
}
```

D)

```
{
  "Name": "DataScientist",
  "IsCustom": true,
  "Description": "Project Data Scientist role",
  "Actions": [],
  "NotActions": ["*"],
  "AssignableScopes": [
    "/subscriptions/<id>/resourceGroups/ml-rg/providers/Microsoft.MachineLearningServices/workspaces/ml-ws"
  ]
}
```

A. Option A

B. Option B

C. Option C

D. Option D

**Answer: A**

**Explanation:**

The following custom role can do everything in the workspace except for the following actions:

- It can't create or update a compute resource.
- It can't delete a compute resource.
- It can't add, delete, or alter role assignments.



➤ It can't delete the workspace.

To create a custom role, first construct a role definition JSON file that specifies the permission and scope for the role. The following example defines a custom role named "Data Scientist Custom" scoped at a specific workspace level:

data\_scientist\_custom\_role.json :

```
{
  "Name": "Data Scientist Custom", "IsCustom": true,
  "Description": "Can run experiment but can't create or delete compute.", "Actions": ["*"],
  "NotActions": [
    "Microsoft.MachineLearningServices/workspaces/*/delete", "Microsoft.MachineLearningServices/workspaces/write",
    "Microsoft.MachineLearningServices/workspaces/computes/*/write", "Microsoft.MachineLearningServices/workspaces/computes/*/delete",
    "Microsoft.Authorization/*/write"
  ],
  "AssignableScopes": [ "/subscriptions/<subscription_id>/resourceGroups/<resource_group_name>/providers/Microsoft.MachineLearni"
]
```

Reference:

<https://docs.microsoft.com/en-us/azure/machine-learning/how-to-assign-roles>

### NEW QUESTION 15

- (Exam Topic 3)

You are moving a large dataset from Azure Machine Learning Studio to a Weka environment. You need to format the data for the Weka environment. Which module should you use?

- A. Convert to CSV
- B. Convert to Dataset
- C. Convert to ARFF
- D. Convert to SVMLight

**Answer:** C

#### Explanation:

Use the Convert to ARFF module in Azure Machine Learning Studio, to convert datasets and results in Azure Machine Learning to the attribute-relation file format used by the Weka toolset. This format is known as ARFF.

The ARFF data specification for Weka supports multiple machine learning tasks, including data preprocessing, classification, and feature selection. In this format, data is organized by entites and their attributes, and is contained in a single text file.

References:

<https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/convert-to-arff>

### NEW QUESTION 18

- (Exam Topic 3)

You plan to build a team data science environment. Data for training models in machine learning pipelines will be over 20 GB in size.

You have the following requirements:

- Models must be built using Caffe2 or Chainer frameworks.
- Data scientists must be able to use a data science environment to build the machine learning pipelines and train models on their personal devices in both connected and disconnected network environments.
- Personal devices must support updating machine learning pipelines when connected to a network. You need to select a data science environment.

Which environment should you use?

- A. Azure Machine Learning Service
- B. Azure Machine Learning Studio
- C. Azure Databricks
- D. Azure Kubernetes Service (AKS)

**Answer:** A

#### Explanation:

The Data Science Virtual Machine (DSVM) is a customized VM image on Microsoft's Azure cloud built specifically for doing data science. Caffe2 and Chainer are supported by DSVM.

DSVM integrates with Azure Machine Learning.

### NEW QUESTION 19

- (Exam Topic 3)

You are building a machine learning model for translating English language textual content into French language textual content.

You need to build and train the machine learning model to learn the sequence of the textual content. Which type of neural network should you use?

- A. Multilayer Perceptions (MLPs)
- B. Convolutional Neural Networks (CNNs)
- C. Recurrent Neural Networks (RNNs)
- D. Generative Adversarial Networks (GANs)

**Answer:** C

#### Explanation:

To translate a corpus of English text to French, we need to build a recurrent neural network (RNN).

Note: RNNs are designed to take sequences of text as inputs or return sequences of text as outputs, or both. They're called recurrent because the network's hidden layers have a loop in which the output and cell state from each time step become inputs at the next time step. This recurrence serves as a form of memory. It allows contextual information to flow through the network so that relevant outputs from previous time steps can be applied to network operations at the current time step.

References:

<https://towardsdatascience.com/language-translation-with-rnns-d84d43b40571>

#### NEW QUESTION 22

- (Exam Topic 3)

You deploy a real-time inference service for a trained model.

The deployed model supports a business-critical application, and it is important to be able to monitor the data submitted to the web service and the predictions the data generates.

You need to implement a monitoring solution for the deployed model using minimal administrative effort. What should you do?

- A. View the explanation for the registered model in Azure ML studio.
- B. Enable Azure Application Insights for the service endpoint and view logged data in the Azure portal.
- C. Create an ML Flow tracking URI that references the endpoint, and view the data logged by ML Flow.
- D. View the log files generated by the experiment used to train the model.

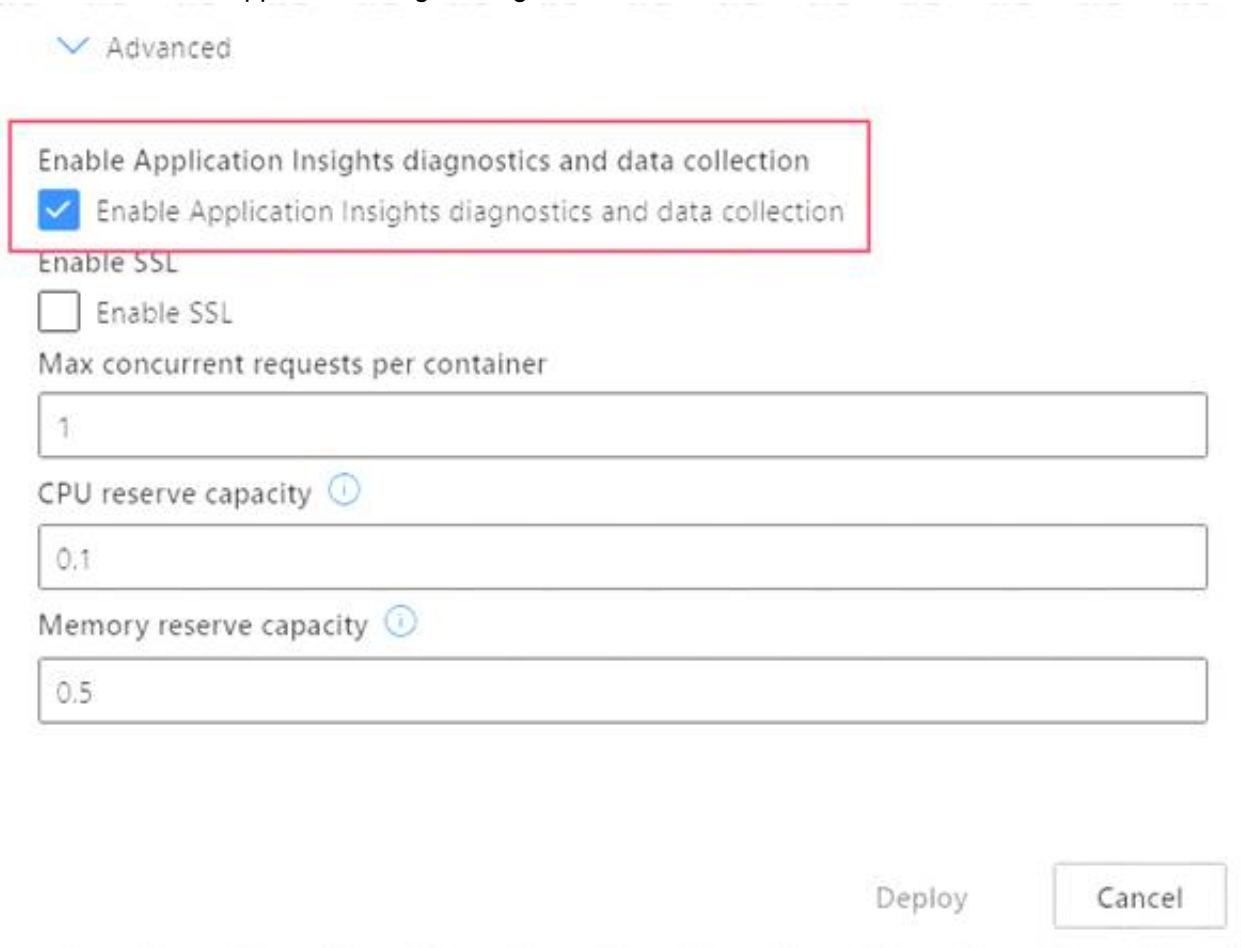
**Answer: B**

#### Explanation:

Configure logging with Azure Machine Learning studio

You can also enable Azure Application Insights from Azure Machine Learning studio. When you're ready to deploy your model as a web service, use the following steps to enable Application Insights:

- \* 1. Sign in to the studio at <https://ml.azure.com>.
- \* 2. Go to Models and select the model you want to deploy.
- \* 3. Select +Deploy.
- \* 4. Populate the Deploy model form.
- \* 5. Expand the Advanced menu.
- \* 6. Select Enable Application Insights diagnostics and data collection.



Advanced

Enable Application Insights diagnostics and data collection

☒ Enable Application Insights diagnostics and data collection

Enable SSL

☐ Enable SSL

Max concurrent requests per container

1

CPU reserve capacity ⓘ

0.1

Memory reserve capacity ⓘ

0.5

Deploy Cancel

Reference:

<https://docs.microsoft.com/en-us/azure/machine-learning/how-to-enable-app-insights>

#### NEW QUESTION 24

- (Exam Topic 3)

You create a multi-class image classification deep learning experiment by using the PyTorch framework. You plan to run the experiment on an Azure Compute cluster that has nodes with GPU's.

You need to define an Azure Machine Learning service pipeline to perform the monthly retraining of the image classification model. The pipeline must run with minimal cost and minimize the time required to train the model.

Which three pipeline steps should you run in sequence? To answer, move the appropriate actions from the list of actions to the answer area and arrange them in the correct order.

#### Actions

- Configure a DataTransferStep() to fetch new image data from public web portal, running on the cpu-compute compute target.
- Configure an EstimatorStep() to run an estimator that runs the bird\_classifier\_train.py model training script on the gpu\_compute compute target.
- Configure a PythonScriptStep() to run both image\_fetcher.py and image\_resize.py on the cpu-compute compute target.
- Configure an EstimatorStep() to run an estimator that runs the bird\_classifier\_train.py model training script on the cpu\_compute compute target.
- Configure a PythonScriptStep() to run image\_fetcher.py on the cpu-compute compute target.
- Configure a PythonScriptStep() to run image\_resize.py on the cpu-compute compute target.
- Configure a PythonScriptStep() to run bird\_classifier\_train.py on the cpu-compute compute target.
- Configure a PythonScriptStep() to run bird\_classifier\_train.py on the gpu-compute compute target.

#### Answer Area

- A. Mastered
- B. Not Mastered

**Answer:** A

#### Explanation:

Step 1: Configure a DataTransferStep() to fetch new image data...

Step 2: Configure a PythonScriptStep() to run image\_resize.y on the cpu-compute compute target. Step 3: Configure the EstimatorStep() to run training script on the gpu\_compute computer target.

The PyTorch estimator provides a simple way of launching a PyTorch training job on a compute target. Reference:

<https://docs.microsoft.com/en-us/azure/machine-learning/how-to-train-pytorch>

#### NEW QUESTION 29

- (Exam Topic 3)

Note: This question is part of a series of questions that present the same scenario. Each question in the series contains a unique solution that might meet the stated goals. Some question sets might have more than one correct solution, while others might not have a correct solution.

After you answer a question in this section, you will NOT be able to return to it. As a result, these questions will not appear in the review screen.

You are creating a new experiment in Azure Machine Learning Studio.

One class has a much smaller number of observations than tin- other classes in the training set. You need to select an appropriate data sampling strategy to compensate for the class imbalance. Solution: You use the Principal Components Analysis (PCA) sampling mode.

Does the solution meet the goal?

- A. Yes
- B. No

**Answer:** B

#### Explanation:

Instead use the Synthetic Minority Oversampling Technique (SMOTE) sampling mode.


Note: SMOTE is used to increase the number of underrepresented cases in a dataset used for machine learning. SMOTE is a better way of increasing the number of rare cases than simply duplicating existing cases.

#### NEW QUESTION 32

- (Exam Topic 3)

You are using a Git repository to track work in an Azure Machine Learning workspace. You need to authenticate a Git account by using SSH.

Which three actions should you perform in sequence? To answer, move the appropriate actions from the list of actions to the answer area and arrange them in the correct order.

Actions	Answer Area
Generate a public/private key pair	
Add the private key to the Git account	
Clone the Git repository by using an SSH repository URL	
Add the public key to the Git account	
Create a new Azure Key Vault resource	

- A. Mastered
- B. Not Mastered

**Answer:** A

**Explanation:**

Graphical user interface, text, application, chat or text message Description automatically generated  
 Authenticate your Git Account with SSH: Step 1: Generating a public/private key pair Generate a new SSH key  
 \* 1. Open the terminal window in the Azure Machine Learning Notebook Tab.  
 \* 2. Paste the text below, substituting in your email address. `ssh-keygen -t rsa -b 4096 -C "your_email@example.com"`  
 This creates a new ssh key, using the provided email as a label.  
 > Generating public/private rsa key pair.  
 Step 2: Add the public key to the Git Account  
 In your terminal window, copy the contents of your public key file. Step 3: Clone the Git repository by using an SSH repository URL  
 \* 1. Copy the SSH Git clone URL from the Git repo.  
 \* 2. Paste the url into the git clone command below, to use your SSH Git repo URL. This will look something like:  
`git clone git@example.com:GitUser/azureml-example.git` Cloning into 'azureml-example'.  
 Reference:  
<https://docs.microsoft.com/en-us/azure/machine-learning/concept-train-model-git-integration>

**NEW QUESTION 34**

- (Exam Topic 3)

You are performing clustering by using the K-means algorithm. You need to define the possible termination conditions.  
 Which three conditions can you use? Each correct answer presents a complete solution. NOTE: Each correct selection is worth one point.

- A. A fixed number of iterations is executed.
- B. The residual sum of squares (RSS) rises above a threshold.
- C. The sum of distances between centroids reaches a maximum.
- D. The residual sum of squares (RSS) falls below a threshold.
- E. Centroids do not change between iterations.

**Answer:** ADE

**Explanation:**

References:  
<https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/k-means-clustering> <https://nlp.stanford.edu/IR-book/html/htmledition/k-means-1.html>

**NEW QUESTION 38**

- (Exam Topic 3)

You plan to create a speech recognition deep learning model. The model must support the latest version of Python.  
 You need to recommend a deep learning framework for speech recognition to include in the Data Science Virtual Machine (DSVM).  
 What should you recommend?

- A. Apache Drill
- B. Tensorflow
- C. Rattle
- D. Weka

**Answer:** B

**Explanation:**

TensorFlow is an open source library for numerical computation and large-scale machine learning. It uses Python to provide a convenient front-end API for building applications with the framework  
 TensorFlow can train and run deep neural networks for handwritten digit classification, image recognition, word embeddings, recurrent neural networks, sequence-to-sequence models for machine translation, natural language processing, and PDE (partial differential equation) based simulations.  
 References:  
<https://www.infoworld.com/article/3278008/what-is-tensorflow-the-machine-learning-library-explained.html>



**NEW QUESTION 42**

- (Exam Topic 3)

You plan to run a Python script as an Azure Machine Learning experiment. The script must read files from a hierarchy of folders. The files will be passed to the script as a dataset argument. You must specify an appropriate mode for the dataset argument. Which two modes can you use? Each correct answer presents a complete solution. NOTE: Each correct selection is worth one point.

- A. to\_pandas\_dataframe ()
- B. as\_download()
- C. as\_upload()
- D. as mount ()

**Answer: B**

**Explanation:**

Reference:  
<https://docs.microsoft.com/en-us/python/api/azureml-core/azureml.data.filedataset?view=azure-ml-py>

**NEW QUESTION 46**

- (Exam Topic 3)

You have a feature set containing the following numerical features: X, Y, and Z. The Poisson correlation coefficient (r-value) of X, Y, and Z features is shown in the following image:

	X	Y	Z
X	1	0.149676	-0.106276
Y	0.149676	1	0.859122
Z	-0.106276	0.859122	1

Use the drop-down menus to select the answer choice that answers each question based on the information presented in the graphic. NOTE: Each correct selection is worth one point.

What is the r-value for the correlation of Y to Z?

▼

-0.106276  
 0.149676  
 0.859122  
 1

Which type of relationship exists between Z and Y in the feature set?

▼

a positive linear relationship  
 a negative linear relationship  
 no linear relationship

- A. Mastered
- B. Not Mastered

**Answer: A**

**Explanation:**

Box 1: 0.859122  
 Box 2: a positively linear relationship  
 +1 indicates a strong positive linear relationship  
 -1 indicates a strong negative linear correlation  
 0 denotes no linear relationship between the two variables. References:  
<https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/compute-linear-correlation>

**NEW QUESTION 48**

- (Exam Topic 3)

You previously deployed a model that was trained using a tabular dataset named training-dataset, which is based on a folder of CSV files. Over time, you have collected the features and predicted labels generated by the model in a folder containing a CSV file for each month. You have created two tabular datasets based on the folder containing the inference data: one named predictions-dataset with a schema that matches the training data exactly, including the predicted label; and another named features-dataset with a schema containing all of the feature columns and a timestamp column based on the filename, which includes the day, month, and year. You need to create a data drift monitor to identify any changing trends in the feature data since the model was trained. To accomplish this, you must define the required datasets for the data drift monitor. Which datasets should you use to configure the data drift monitor? To answer, drag the appropriate datasets to the correct data drift monitor options. Each source may be used once, more than once, or not at all. You may need to drag the split bar between panes or scroll to view content. NOTE: Each correct selection is worth one point.

Target datasets	Answer Area
training-dataset	Baseline dataset
predictions-dataset	Target dataset
features-dataset	Target dataset

- A. Mastered  
 B. Not Mastered

**Answer:** A

**Explanation:**

Text Description automatically generated with medium confidence

Box 1: training-dataset

Baseline dataset - usually the training dataset for a model. Box 2: predictions-dataset

Target dataset - usually model input data - is compared over time to your baseline dataset. This comparison means that your target dataset must have a timestamp column specified.

The monitor will compare the baseline and target datasets. Reference:

<https://docs.microsoft.com/en-us/azure/machine-learning/how-to-monitor-datasets>

**NEW QUESTION 49**

- (Exam Topic 3)

You are using a decision tree algorithm. You have trained a model that generalizes well at a tree depth equal to 10.

You need to select the bias and variance properties of the model with varying tree depth values.

Which properties should you select for each tree depth? To answer, select the appropriate options in the answer area.

**Answer Area**

Tree Depth	Bias	Variance
5	<div>▼</div> <div>High</div> <div>Low</div> <div>Identical</div>	<div>▼</div> <div>High</div> <div>Low</div> <div>Identical</div>
15	<div>▼</div> <div>High</div> <div>Low</div> <div>Identical</div>	<div>▼</div> <div>High</div> <div>Low</div> <div>Identical</div>

- A. Mastered  
 B. Not Mastered

**Answer:** A

**Explanation:**

In decision trees, the depth of the tree determines the variance. A complicated decision tree (e.g. deep) has low bias and high variance.

Note: In statistics and machine learning, the bias–variance tradeoff is the property of a set of predictive models whereby models with a lower bias in parameter estimation have a higher variance of the parameter estimates across samples, and vice versa. Increasing the bias will decrease the variance. Increasing the variance will decrease the bias.

References:

<https://machinelearningmastery.com/gentle-introduction-to-the-bias-variance-trade-off-in-machine-learning/>

**NEW QUESTION 50**

- (Exam Topic 3)

You use an Azure Machine Learning workspace.

You have a trained model that must be deployed as a web service. Users must authenticate by using Azure Active Directory.

What should you do?

- A. Deploy the model to Azure Kubernetes Service (AKS). During deployment, set the token\_auth\_enabled parameter of the target configuration object to true  
 B. Deploy the model to Azure Container Instance  
 C. During deployment, set the auth\_enabled parameter of the target configuration object to true  
 D. Deploy the model to Azure Container Instance  
 E. During deployment, set the token\_auth\_enabled parameter of the target configuration object to true  
 F. Deploy the model to Azure Kubernetes Service (AKS). During deployment, set the auth\_enabled parameter of the target configuration object to true  
 G. enabled parameter of the target configuration object to true

**Answer:** A

**Explanation:**

To control token authentication, use the token\_auth\_enabled parameter when you create or update a deployment

Token authentication is disabled by default when you deploy to Azure Kubernetes Service.

Note: The model deployments created by Azure Machine Learning can be configured to use one of two authentication methods:

key-based: A static key is used to authenticate to the web service.

token-based: A temporary token must be obtained from the Azure Machine Learning workspace (using Azure Active Directory) and used to authenticate to the web service.

Reference:

<https://docs.microsoft.com/en-us/azure/machine-learning/how-to-authenticate-web-service>

**NEW QUESTION 52**

- (Exam Topic 3)

Note: This question is part of a series of questions that present the same scenario. Each question in the series contains a unique solution that might meet the stated goals. Some question sets might have more than one correct solution, while others might not have a correct solution.

After you answer a question in this section, you will NOT be able to return to it. As a result, these questions will not appear in the review screen.

You have a Python script named train.py in a local folder named scripts. The script trains a regression model by using scikit-learn. The script includes code to load a training data file which is also located in the scripts folder.

You must run the script as an Azure ML experiment on a compute cluster named aml-compute.

You need to configure the run to ensure that the environment includes the required packages for model training. You have instantiated a variable named aml-compute that references the target compute cluster.

Solution: Run the following code:

```
from azureml.train.estimator import Estimator
sk_est = Estimator(source_directory='./scripts',
    compute_target=aml-compute,
    entry_script='train.py',
    conda_packages=['scikit-learn'])
```

Does the solution meet the goal?

- A. Yes
- B. No

**Answer: B**

**Explanation:**

The scikit-learn estimator provides a simple way of launching a scikit-learn training job on a compute target. It is implemented through the SKLearn class, which can be used to support single-node CPU training.

Example:

```
from azureml.train.sklearn import SKLearn
}
estimator = SKLearn(source_directory=project_folder, compute_target=compute_target, entry_script='train_iris.py'
)
```

Reference:

<https://docs.microsoft.com/en-us/azure/machine-learning/how-to-train-scikit-learn>

**NEW QUESTION 57**

- (Exam Topic 3)

You register a file dataset named csvfolder that references a folder. The folder includes multiple com

ma-separated values (CSV) files in an Azure storage blob container. You plan to use the following code to run a script that loads data from the file dataset. You create and instantiate the following variables:

Variable	Description
remote_cluster	References the Azure Machine Learning compute cluster
ws	References the Azure Machine Learning workspace

You have the following code:

```
from azureml.train.estimator import Estimator
file_dataset = ws.datasets.get('csv_folder')
estimator = Estimator(source_directory=script_folder,

    compute_target = remote_cluster,
    entry_script='script.py')
run = experiment.submit(config=estimator)
run.wait_for_completion(show_output=True)
```

You need to pass the dataset to ensure that the script can read the files it references. Which code segment should you insert to replace the code comment?

A)

```
inputs=[file_dataset.as_named_input('training_files').to_pandas_dataframe()],
```

B)

```
inputs=[file_dataset.as_named_input('training_files').as_mount()],
```

C)

```
script_params={'--training_files': file_dataset},
```



D)

```
inputs=[file_dataset.as_named_input('training_files')],
```

- A. Option A
- B. Option B
- C. Option C
- D. Option D

**Answer:** D

**Explanation:**

Example:

```
from azureml.train.estimator import Estimator
script_params = {
    # to mount files referenced by mnist dataset
    '--data-folder': mnist_file_dataset.as_named_input('mnist_opendataset').as_mount(),
    '--regularization': 0.5
}
est = Estimator(source_directory=script_folder, script_params=script_params, compute_target=compute_target, environment_definition=env,
entry_script='train.py')
```

Reference:

<https://docs.microsoft.com/en-us/azure/machine-learning/tutorial-train-models-with-aml>

**NEW QUESTION 61**

- (Exam Topic 3)

You create a machine learning model by using the Azure Machine Learning designer. You publish the model as a real-time service on an Azure Kubernetes Service (AKS) inference compute cluster. You make no changes to the deployed endpoint configuration.

You need to provide application developers with the information they need to consume the endpoint.

Which two values should you provide to application developers? Each correct answer presents part of the solution.

NOTE: Each correct selection is worth one point.

- A. The name of the AKS cluster where the endpoint is hosted.
- B. The name of the inference pipeline for the endpoint.
- C. The URL of the endpoint.
- D. The run ID of the inference pipeline experiment for the endpoint.
- E. The key for the endpoint.

**Answer:** CE

**Explanation:**

Deploying an Azure Machine Learning model as a web service creates a REST API endpoint. You can send data to this endpoint and receive the prediction returned by the model.

You create a web service when you deploy a model to your local environment, Azure Container Instances, Azure Kubernetes Service, or field-programmable gate arrays (FPGA). You retrieve the URI used to access the web service by using the Azure Machine Learning SDK. If authentication is enabled, you can also use the SDK to get the authentication keys or tokens.

Example:

```
# URL for the web service
```

```
scoring_uri = '<your web service URI>'
```

```
# If the service is authenticated, set the key or token key = '<your key or token>'
```

Reference:

<https://docs.microsoft.com/en-us/azure/machine-learning/how-to-consume-web-service>

**NEW QUESTION 66**

- (Exam Topic 3)

You create an Azure Databricks workspace and a linked Azure Machine Learning workspace. You have the following Python code segment in the Azure Machine Learning workspace:

```
import mlflow
import mlflow.azureml
import azureml.mlflow
import azureml.core
from azureml.core import Workspace
subscription_id = 'subscription_id'
resource_group = 'resource_group_name'
workspace_name = 'workspace_name'
ws = Workspace.get(name=workspace_name, subscription_id=subscription_id, resource_group=resource_group)
experimentName = "/Users/{user_name}/{experiment_folder}/{experiment_name}"
mlflow.set_experiment(experimentName)
uri = ws.get_mlflow_tracking_uri()
mlflow.set_tracking_uri(uri)
```

Instructions: For each of the following statements, select Yes if the statement is true. Otherwise, select No.

NOTE: Each correct selection is worth one point.

	Yes	No
A resource group and Azure Machine Learning workspace will be created.	<input type="radio"/>	<input type="radio"/>
An Azure Databricks experiment will be tracked only in the Azure Machine Learning workspace.	<input type="radio"/>	<input type="radio"/>
The epoch loss metric is set to be tracked.	<input type="radio"/>	<input type="radio"/>

- A. Mastered
- B. Not Mastered



**Answer:** A

**Explanation:**

A screenshot of a computer Description automatically generated with medium confidence

Box 1: No

The Workspace.get method loads an existing workspace without using configuration files. ws = Workspace.get(name="myworkspace", subscription\_id='<azure-subscription-id>', resource\_group='myresourcegroup')

Box 2: Yes

MLflow Tracking with Azure Machine Learning lets you store the logged metrics and artifacts from your local runs into your Azure Machine Learning workspace. The get\_mlflow\_tracking\_uri() method assigns a unique tracking URI address to the workspace, ws, and set\_tracking\_uri() points the MLflow tracking URI to that address.

Box 3: Yes

Note: In Deep Learning, epoch means the total dataset is passed forward and backward in a neural network once.

Reference:

<https://docs.microsoft.com/en-us/python/api/azureml-core/azureml.core.workspace.workspace> <https://docs.microsoft.com/en-us/azure/machine-learning/how-to-use-mlflow>

**NEW QUESTION 67**

- (Exam Topic 3)

You plan to explore demographic data for home ownership in various cities. The data is in a CSV file with the following format:

age,city,income,home\_owner 21,Chicago,50000,0 35,Seattle,120000,1 23,Seattle,65000,0 45,Seattle,130000,1 18,Chicago,48000,0

You need to run an experiment in your Azure Machine Learning workspace to explore the data and log the results. The experiment must log the following information:

- the number of observations in the dataset
- a box plot of income by home\_owner
- a dictionary containing the city names and the average income for each city

You need to use the appropriate logging methods of the experiment's run object to log the required information.

How should you complete the code? To answer, drag the appropriate code segments to the correct locations. Each code segment may be used once, more than once, or not at all. You may need to drag the split bar between panes or scroll to view content.

NOTE: Each correct selection is worth one point.

**Code segments**

log  
log\_list  
log\_row  
log\_table  
log\_image

**Answer Area**

```
from azureml.core import Experiment, Run
import pandas as pd
import matplotlib.pyplot as plt
# Create an Azure ML experiment in workspace
experiment = Experiment(workspace = ws, name = "demo-experiment")
# Start logging data from the experiment
run = experiment.start_logging()
# load the dataset
data = pd.read_csv('research/demographics.csv')
# Log the number of observations
row_count = (len(data))
run. Segment ("observations", row_count)
# Log box plot for income by home_owner
fig = plt.figure(figsize=(9, 6))
ax = fig.gca()
data.boxplot(column = 'income', by = "home_owner", ax = ax)
ax.set_title('income by home_owner')
ax.set_ylabel('income')
run. Segment (name = 'income_by_home_owner', plot = fig)
# Create a dataframe of mean income per city
mean_inc_df = data.groupby('city')['income'].agg(np.mean).to_frame().reset_index()
# Convert to a dictionary
mean_inc_dict = mean_inc_df.to_dict('dict')
# Log city names and average income dictionary
run. Segment (name="mean_income_by_city", value= mean_inc_dict)
# Complete tracking and get link to details
run.complete()
```

- A. Mastered  
B. Not Mastered

**Answer:** A

**Explanation:**

Box 1: log

The number of observations in the dataset. run.log(name, value, description="")

Scalar values: Log a numerical or string value to the run with the given name. Logging a metric to a run causes that metric to be stored in the run record in the experiment. You can log the same metric multiple times within a run, the result being considered a vector of that metric.

Example: run.log("accuracy", 0.95)

Box 2: log\_image

A box plot of income by home\_owner.

log\_image Log an image to the run record. Use log\_image to log a .PNG image file or a matplotlib plot to the run. These images will be visible and comparable in the run record.

Example: run.log\_image("ROC", plot=plt) Box 3: log\_table

A dictionary containing the city names and the average income for each city. log\_table: Log a dictionary object to the run with the given name.

#### NEW QUESTION 69

- (Exam Topic 3)

You deploy a model as an Azure Machine Learning real-time web service using the following code.

```
# ws, model, inference_config, and deployment_config defined previously
service = Model.deploy(ws, 'classification-service', [model], inference_config, deployment_config)
service.wait_for_deployment(True)
```

The deployment fails.

You need to troubleshoot the deployment failure by determining the actions that were performed during deployment and identifying the specific action that failed. Which code segment should you run?

- A. service.get\_logs()
- B. service.state
- C. service.serialize()
- D. service.update\_deployment\_state()

**Answer:** A

#### Explanation:

You can print out detailed Docker engine log messages from the service object. You can view the log for ACI, AKS, and Local deployments. The following example demonstrates how to print the logs.

```
# if you already have the service object handy print(service.get_logs())
```

```
# if you only know the name of the service (note there might be multiple services with the same name but different version number)
```

```
print(ws.webservices['mysvc'].get_logs())
```

 Reference:

<https://docs.microsoft.com/en-us/azure/machine-learning/how-to-troubleshoot-deployment>

#### NEW QUESTION 71

- (Exam Topic 3)

HOTSPOT

You have an Azure blob container that contains a set of TSV files. The Azure blob container is registered as a datastore for an Azure Machine Learning service workspace. Each TSV file uses the same data schema.

You plan to aggregate data for all of the TSV files together and then register the aggregated data as a dataset in an Azure Machine Learning workspace by using the Azure Machine Learning SDK for Python.

You run the following code.

```
from azureml.core.workspace import Workspace
from azureml.core.datastore import Datastore
from azureml.core.dataset import Dataset
import pandas as pd
datastore_paths = (datastore, './data/*.tsv')
myDataset_1 = Dataset.File.from_files(path=datastore_paths)
myDataset_2 = Dataset.Tabular.from_delimited_files(path=datastore_paths, separator='\t')
```

For each of the following statements, select Yes if the statement is true. Otherwise, select No.

NOTE: Each correct selection is worth one point.

	Yes	No
The myDataset_1 dataset can be converted into a pandas dataframe by using the following method: using myDataset_1.to_pandas_dataframe()	<input type="radio"/>	<input type="radio"/>
The myDataset_1.to_path() method returns an array of file paths for all of the TSV files in the dataset.	<input type="radio"/>	<input type="radio"/>
The myDataset_2 dataset can be converted into a pandas dataframe by using the following method: myDataset_2.to_pandas_dataframe()	<input type="radio"/>	<input type="radio"/>

- A. Mastered
- B. Not Mastered

**Answer:** A

#### Explanation:

Box 1: No

FileDataset references single or multiple files in datastores or from public URLs. The TSV files need to be parsed.

Box 2: Yes

to\_path() gets a list of file paths for each file stream defined by the dataset. Box 3: Yes

TabularDataset.to\_pandas\_dataframe loads all records from the dataset into a pandas DataFrame. TabularDataset represents data in a tabular format created by parsing the provided file or list of files.

Note: TSV is a file extension for a tab-delimited file used with spreadsheet software. TSV stands for Tab Separated Values. TSV files are used for raw data and can be imported into and exported from spreadsheet software. TSV files are essentially text files, and the raw data can be viewed by text editors, though they are often used when moving raw data between spreadsheets.

Reference:



<https://docs.microsoft.com/en-us/python/api/azureml-core/azureml.data.tabulardataset>

### NEW QUESTION 73

- (Exam Topic 3)

You plan to provision an Azure Machine Learning Basic edition workspace for a data science project. You need to identify the tasks you will be able to perform in the workspace.

Which three tasks will you be able to perform? Each correct answer presents a complete solution.

NOTE: Each correct selection is worth one point. D

- A. Create a Compute Instance and use it to run code in Jupyter notebooks.
- B. Create an Azure Kubernetes Service (AKS) inference cluster.
- C. Use the designer to train a model by dragging and dropping pre-defined modules.
- D. Create a tabular dataset that supports versioning.
- E. Use the Automated Machine Learning user interface to train a model.

**Answer:** ABD

#### Explanation:

Reference:

<https://azure.microsoft.com/en-us/pricing/details/machine-learning/>

### NEW QUESTION 77

- (Exam Topic 3)

You are developing a hands-on workshop to introduce Docker for Windows to attendees. You need to ensure that workshop attendees can install Docker on their devices.

Which two prerequisite components should attendees install on the devices? Each correct answer presents part of the solution.

NOTE: Each correct selection is worth one point.

- A. Microsoft Hardware-Assisted Virtualization Detection Tool
- B. Kitematic
- C. BIOS-enabled virtualization
- D. VirtualBox
- E. Windows 10 64-bit Professional

**Answer:** CE

#### Explanation:

C: Make sure your Windows system supports Hardware Virtualization Technology and that virtualization is enabled.

Ensure that hardware virtualization support is turned on in the BIOS settings. For example:



E: To run Docker, your machine must have a 64-bit operating system running Windows 7 or higher. References:

[https://docs.docker.com/toolbox/toolbox\\_install\\_windows/](https://docs.docker.com/toolbox/toolbox_install_windows/) <https://blogs.technet.microsoft.com/canitpro/2015/09/08/step-by-step-enabling-hyper-v-for-use-on-windows-10/>

### NEW QUESTION 81

- (Exam Topic 3)

You have a Python script that executes a pipeline. The script includes the following code:

```
from azureml.core import Experiment
```

```
pipeline_run = Experiment(ws, 'pipeline_test').submit(pipeline)
```

You want to test the pipeline before deploying the script.

You need to display the pipeline run details written to the STDOUT output when the pipeline completes. Which code segment should you add to the test script?

- A. `pipeline_run.get.metrics()`
- B. `pipeline_run.wait_for_completion(show_output=True)`
- C. `pipeline_param = PipelineParameter(name="stdout", default_value="console")`
- D. `pipeline_run.get_status()`

**Answer:** B

**Explanation:**

wait\_for\_completion: Wait for the completion of this run. Returns the status object after the wait. Syntax: wait\_for\_completion(show\_output=False, wait\_post\_processing=False, raise\_on\_error=True) Parameter: show\_output  
 Indicates whether to show the run output on sys.stdout.

**NEW QUESTION 85**

- (Exam Topic 3)

You have a Python script that executes a pipeline. The script includes the following code: from azureml.core import Experiment  
 pipeline\_run = Experiment(ws, 'pipeline\_test').submit(pipeline) You want to test the pipeline before deploying the script.

You need to display the pipeline run details written to the STDOUT output when the pipeline completes. Which code segment should you add to the test script?

- A. pipeline\_run.get.metrics()
- B. pipeline\_run.wait\_for\_completion(show\_output=True)
- C. pipeline\_param = PipelineParameter(name="stdout", default\_value="console")
- D. pipeline\_run.get\_status()

**Answer: B**

**Explanation:**

wait\_for\_completion: Wait for the completion of this run. Returns the status object after the wait. Syntax: wait\_for\_completion(show\_output=False, wait\_post\_processing=False, raise\_on\_error=True) Parameter: show\_output  
 Indicates whether to show the run output on sys.stdout.

**NEW QUESTION 86**

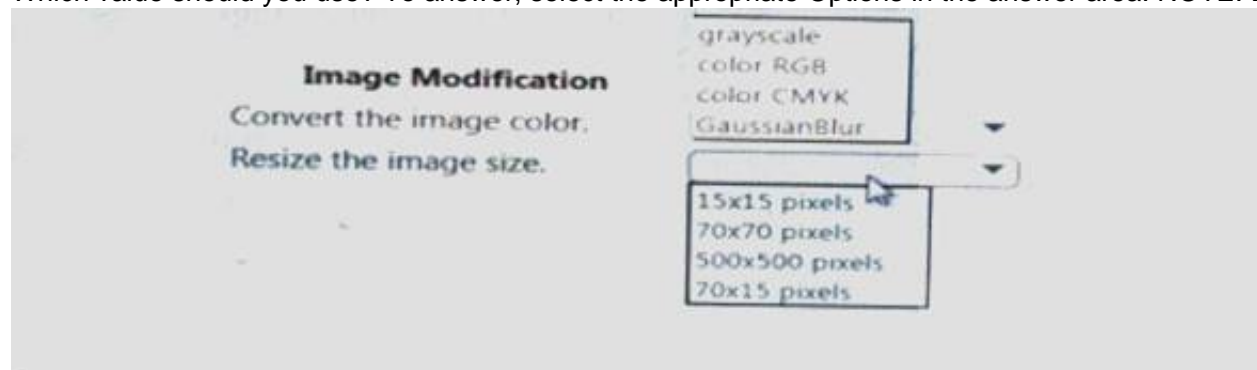
- (Exam Topic 3)

You are training a deep learning model to identify cats and dogs. You have 25,000 color images. You must meet the following requirements:

- Reduce the number of training epochs.
- Reduce the size of the neural network.
- Reduce over-fitting of the neural network.

You need to select the image modification values.

Which value should you use? To answer, select the appropriate Options in the answer area. NOTE: Each correct selection is worth one point.



- A. Mastered
- B. Not Mastered

**Answer: A**

**Explanation:**



**NEW QUESTION 91**

- (Exam Topic 3)

You define a datastore named ml-data for an Azure Storage blob container. In the container, you have a folder named train that contains a file named data.csv. You plan to use the file to train a model by using the Azure Machine Learning SDK.

You plan to train the model by using the Azure Machine Learning SDK to run an experiment on local compute.

You define a DataReference object by running the following code:



```
from azureml.core import Workspace, Datastore, Environment
from azureml.train.estimator import Estimator
ws = Workspace.from_config()
ml_data = Datastore.get(ws, datastore_name='ml-data')
data_ref = ml_data.path('train').as_download(path_on_compute='train_data')
estimator = Estimator(source_directory='experiment_folder',
    script_params={'--data-folder': data_ref},
    compute_target = 'local',
    entry_script='training.py')
run = experiment.submit(config=estimator)
run.wait_for_completion(show_output=True)
```

You need to load the training data. Which code segment should you use?

- A. 

```
import os
import argparse
import pandas as pd

parser = argparse.ArgumentParser()
parser.add_argument('--data-folder', type=str, dest='data_folder')
data_folder = args.data_folder
data = pd.read_csv(os.path.join(data_folder, 'ml-data', 'train_data', 'data.csv'))
```
- B. 

```
import os
import argparse
import pandas as pd

parser = argparse.ArgumentParser()
parser.add_argument('--data-folder', type=str, dest='data_folder')
data_folder = args.data_folder
data = pd.read_csv(os.path.join(data_folder, 'train', 'data.csv'))
```
- C. 

```
import pandas as pd

data = pd.read_csv('./data.csv')
```
- D. 

```
import os
import argparse
import pandas as pd

parser = argparse.ArgumentParser()
parser.add_argument('--data-folder', type=str, dest='data_folder')
data_folder = args.data_folder
data = pd.read_csv(os.path.join('ml_data', data_folder, 'data.csv'))
```
- E. 

```
import os
import argparse
import pandas as pd

parser = argparse.ArgumentParser()
parser.add_argument('--data-folder', type=str, dest='data_folder')
data_folder = args.data_folder
data = pd.read_csv(os.path.join(data_folder, 'data.csv'))
```

- A. Option A
- B. Option B
- C. Option C
- D. Option D
- E. Option E

**Answer:** E

**Explanation:**

Example:  
 data\_folder = args.data\_folder  
 # Load Train and Test data  
 train\_data = pd.read\_csv(os.path.join(data\_folder, 'data.csv')) Reference:  
<https://www.element61.be/en/resource/azure-machine-learning-services-complete-toolbox-ai>

**NEW QUESTION 95**

- (Exam Topic 3)

You use the Azure Machine Learning SDK to run a training experiment that trains a classification model and calculates its accuracy metric.

The model will be retrained each month as new data is available. You must register the model for use in a batch inference pipeline.

You need to register the model and ensure that the models created by subsequent retraining experiments are registered only if their accuracy is higher than the currently registered model.

What are two possible ways to achieve this goal? Each correct answer presents a complete solution. NOTE: Each correct selection is worth one point.

- A. Specify a different name for the model each time you register it.
- B. Register the model with the same name each time regardless of accuracy, and always use the latest version of the model in the batch inferencing pipeline.
- C. Specify the model framework version when registering the model, and only register subsequent models if this value is higher.
- D. Specify a property named accuracy with the accuracy metric as a value when registering the model, and only register subsequent models if their accuracy is higher than the accuracy property value of the currently registered model.
- E. Specify a tag named accuracy with the accuracy metric as a value when registering the model, and only register subsequent models if their accuracy is higher than the accuracy tag value of the currently registered mode

**Answer:** CE

**Explanation:**

E: Using tags, you can track useful information such as the name and version of the machine learning library used to train the model. Note that tags must be alphanumeric.

Reference:

<https://notebooks.azure.com/xavierheriat/projects/azureml-getting-started/html/how-to-use-azureml/deployment/>

**NEW QUESTION 99**

- (Exam Topic 3)

Note: This question is part of a series of questions that present the same scenario. Each question in the series contains a unique solution that might meet the stated goals. Some question sets might have more than one correct solution, while others might not have a correct solution.

After you answer a question in this section, you will NOT be able to return to it. As a result, these questions will not appear in the review screen.

You create a model to forecast weather conditions based on historical data.

You need to create a pipeline that runs a processing script to load data from a datastore and pass the processed data to a machine learning model training script.

Solution: Run the following code:

```
datastore = ws.get_default_datastore()
data_output = pd.read_csv("traindata.csv")
process_step = PythonScriptStep(script_name="process.py",
    arguments=["--data_for_train", data_output],
    outputs=[data_output],compute_target=aml_compute,
    source_directory=process_directory)
train_step = PythonScriptStep(script_name="train.py",
    arguments=["--data_for_train", data_output],
    inputs=[data_output],compute_target=aml_compute,
    source_directory=train_directory)
pipeline = Pipeline(workspace=ws, steps=[process_step, train_step])
```

Does the solution meet the goal?

A. Yes

B. No

**Answer:** A

**Explanation:**

The two steps are present: process\_step and train\_step Note:

Data used in pipeline can be produced by one step and consumed in another step by providing a PipelineData object as an output of one step and an input of one or more subsequent steps.

PipelineData objects are also used when constructing Pipelines to describe step dependencies. To specify that a step requires the output of another step as input, use a PipelineData object in the constructor of both steps.

For example, the pipeline train step depends on the process\_step\_output output of the pipeline process step: from azureml.pipeline.core import Pipeline, PipelineData

```
from azureml.pipeline.steps import PythonScriptStep
datastore = ws.get_default_datastore()
process_step_output = PipelineData("processed_data", datastore=datastore)
process_step = PythonScriptStep(script_name="process.py",
    arguments=["--data_for_train", process_step_output], outputs=[process_step_output], compute_target=aml_compute, source_directory=process_directory)
train_step = PythonScriptStep(script_name="train.py", arguments=["--data_for_train", process_step_output], inputs=[process_step_output],
    compute_target=aml_compute, source_directory=train_directory)
pipeline = Pipeline(workspace=ws, steps=[process_step, train_step])
```

Reference:

<https://docs.microsoft.com/en-us/python/api/azureml-pipeline-core/azureml.pipeline.core.pipelinedata?view=azu>

**NEW QUESTION 104**

- (Exam Topic 3)

You are creating a new experiment in Azure Machine Learning Studio. You have a small dataset that has missing values in many columns. The data does not require the application of predictors for each column. You plan to use the Clean Missing Data module to handle the missing data.

You need to select a data cleaning method. Which method should you use?

A. Synthetic Minority

B. Replace using Probabilistic PAC

C. Replace using MICE

D. Normalization

**Answer:** B

**NEW QUESTION 108**

- (Exam Topic 3)

HOTSPOT

You collect data from a nearby weather station. You have a pandas dataframe named weather\_df that includes the following data:

Temperature	Observation_time	Humidity	Pressure	Visibility	Days_since_last observation
74	2019/10/2 00:00	0.62	29.87	3	0.5
89	2019/10/2 12:00	0.70	28.88	10	0.5
72	2019/10/3 00:00	0.64	30.00	8	0.5
80	2019/10/3 12:00	0.66	29.75	7	0.5



The data is collected every 12 hours: noon and midnight.

You plan to use automated machine learning to create a time-series model that predicts temperature over the next seven days. For the initial round of training, you want to train a maximum of 50 different models.

You must use the Azure Machine Learning SDK to run an automated machine learning experiment to train these models.

You need to configure the automated machine learning run.

How should you complete the AutoMLConfig definition? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.

```

automl_config = AutoMLConfig(task="
                                regression
                                forecasting
                                classification
                                deep learning

                                training_data=weather_df,
                                label_column_name="
                                humidity
                                pressure
                                visibility
                                temperature
                                days_since_last
                                observation_time

                                time_column_name="
                                humidity
                                pressure
                                visibility
                                temperature
                                days_since_last
                                observation_time

                                max_horizon=
                                2
                                6
                                7
                                12
                                14
                                50

                                iterations=
                                2
                                6
                                7
                                12
                                14
                                50

                                iteration_timeout_minutes=5,
                                primary_metric="r2_score")
    
```

- A. Mastered
- B. Not Mastered

**Answer: A**

**Explanation:**

Box 1: forecasting

Task: The type of task to run. Values can be 'classification', 'regression', or 'forecasting' depending on the type of automated ML problem to solve.

Box 2: temperature

The training data to be used within the experiment. It should contain both training features and a label column (optionally a sample weights column).

Box 3: observation\_time

time\_column\_name: The name of the time column. This parameter is required when forecasting to specify the datetime column in the input data used for building the time series and inferring its frequency. This setting is being deprecated. Please use forecasting\_parameters instead.

Box 4: 7

"predicts temperature over the next seven days"

max\_horizon: The desired maximum forecast horizon in units of time-series frequency. The default value is 1. Units are based on the time interval of your training data, e.g., monthly, weekly that the forecaster should predict out. When task type is forecasting, this parameter is required.

Box 5: 50

"For the initial round of training, you want to train a maximum of 50 different models."

Iterations: The total number of different algorithm and parameter combinations to test during an automated

ML experiment. Reference:

<https://docs.microsoft.com/en-us/python/api/azureml-train-automl-client/azureml.train.automl.automlconfig.auto>

#### NEW QUESTION 111

- (Exam Topic 3)

You are solving a classification task.

You must evaluate your model on a limited data sample by using k-fold cross-validation. You start by configuring a k parameter as the number of splits.

You need to configure the k parameter for the cross-validation. Which value should you use?

- A. k=1
- B. k=10
- C. k=0.5
- D. k=0.9

**Answer: B**

#### Explanation:

Leave One Out (LOO) cross-validation

Setting  $K = n$  (the number of observations) yields n-fold and is called leave-one out cross-validation (LOO), a special case of the K-fold approach.

LOO CV is sometimes useful but typically doesn't shake up the data enough. The estimates from each fold are highly correlated and hence their average can have high variance.

This is why the usual choice is  $K=5$  or 10. It provides a good compromise for the bias-variance tradeoff.

#### NEW QUESTION 115

- (Exam Topic 3)

Note: This question is part of a series of questions that present the same scenario. Each question in the series contains a unique solution that might meet the stated goals. Some question sets might have more than one correct solution, while others might not have a correct solution.

After you answer a question in this section, you will NOT be able to return to it. As a result, these questions will not appear in the review screen.

You plan to use a Python script to run an Azure Machine Learning experiment. The script creates a reference to the experiment run context, loads data from a file, identifies the set of unique values for the label column, and completes the experiment run:

```
from azureml.core import Run
import pandas as pd

run = Run.get_context()
data = pd.read_csv('data.csv')
label_vals = data['label'].unique()
# Add code to record metrics here
run.complete()
```

The experiment must record the unique labels in the data as metrics for the run that can be reviewed later.

You must add code to the script to record the unique label values as run metrics at the point indicated by the comment.

Solution: Replace the comment with the following code: for label\_val in label\_vals:

run.log('Label Values', label\_val) Does the solution meet the goal?

- A. Yes
- B. No

**Answer: A**

#### Explanation:

The run\_log function is used to log the contents in label\_vals: for label\_val in label\_vals:

run.log('Label Values', label\_val) Reference:

<https://www.element61.be/en/resource/azure-machine-learning-services-complete-toolbox-ai>

#### NEW QUESTION 120

- (Exam Topic 3)

You are creating a new experiment in Azure Machine Learning Studio. You have a small dataset that has missing values in many columns. The data does not require the application of predictors for each column. You plan to use the Clean Missing Data module to handle the missing data.

You need to select a data cleaning method. Which method should you use?

- A. Synthetic Minority Oversampling Technique (SMOTE)
- B. Replace using MICE
- C. Replace using; Probabilistic PCA
- D. Normalization

**Answer: C**

#### Explanation:

Replace using Probabilistic PCA: Compared to other options, such as Multiple Imputation using Chained Equations (MICE), this option has the advantage of not requiring the application of predictors for each column. Instead, it approximates the covariance for the full dataset. Therefore, it might offer better performance for datasets that have missing values in many columns.

References:

<https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/clean-missing-data>

#### NEW QUESTION 123

- (Exam Topic 3)

You are evaluating a completed binary classification machine. You need to use the precision as the evaluation metric.

Which visualization should you use?

- A. scatter plot
- B. coefficient of determination
- C. Receiver Operating Characteristic CROC) curve
- D. Gradient descent



**Answer:** C

**Explanation:**

Receiver operating characteristic (or ROC) is a plot of the correctly classified labels vs. the incorrectly classified labels for a particular model.

References:

<https://docs.microsoft.com/en-us/azure/machine-learning/how-to-understand-automated-ml#confusion-matrix>

**NEW QUESTION 124**

- (Exam Topic 3)

You configure a Deep Learning Virtual Machine for Windows.

You need to recommend tools and frameworks to perform the following:

- Build deep neural network (DNN) models
- Perform interactive data exploration and visualization

Which tools and frameworks should you recommend? To answer, drag the appropriate tools to the correct tasks. Each tool may be used once, more than once, or not at all. You may need to drag the split bar between panes or scroll to view content.

NOTE: Each correct selection is worth one point.

Tools	Answer Area						
	<table border="1"><thead><tr><th>Task</th><th>Tool</th></tr></thead><tbody><tr><td>Build DNN models</td><td>Tool</td></tr><tr><td>Enable interactive data exploration and visualization</td><td>Tool</td></tr></tbody></table>	Task	Tool	Build DNN models	Tool	Enable interactive data exploration and visualization	Tool
Task	Tool						
Build DNN models	Tool						
Enable interactive data exploration and visualization	Tool						
Vowpal Wabbit							
PowerBI Desktop							
Azure Data Factory							
Microsoft Cognitive Toolkit							

- A. Mastered
- B. Not Mastered

**Answer:** A

**Explanation:**

Box 1: Vowpal Wabbit

Use the Train Vowpal Wabbit Version 8 module in Azure Machine Learning Studio (classic), to create a machine learning model by using Vowpal Wabbit.

Box 2: PowerBI Desktop

Power BI Desktop is a powerful visual data exploration and interactive reporting tool

BI is a name given to a modern approach to business decision making in which users are empowered to find, explore, and share insights from data across the enterprise.

References:

[https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/train-vowpal-wabbit-version-](https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/train-vowpal-wabbit-version-8) <https://docs.microsoft.com/en-us/azure/architecture/data-guide/scenarios/interactive-data-exploration>

**NEW QUESTION 129**

- (Exam Topic 3)

You create a classification model with a dataset that contains 100 samples with Class A and 10,000 samples with Class B

The variation of Class B is very high. You need to resolve imbalances. Which method should you use?

- A. Partition and Sample
- B. Cluster Centroids
- C. Tomek links
- D. Synthetic Minority Oversampling Technique (SMOTE)

**Answer:** D

**NEW QUESTION 134**

- (Exam Topic 3)

You use Azure Machine Learning Studio to build a machine learning experiment. You need to divide data into two distinct datasets.

Which module should you use?

- A. Partition and Sample
- B. Assign Data to Clusters
- C. Group Data into Bins
- D. Test Hypothesis Using t-Test

**Answer:** A

**Explanation:**

Partition and Sample with the Stratified split option outputs multiple datasets, partitioned using the rules you specified.

References:

<https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/partition-and-sample>

**NEW QUESTION 139**

- (Exam Topic 3)  
 You create machine learning models by using Azure Machine Learning.  
 You plan to train and score models by using a variety of compute contexts. You also plan to create a new compute resource in Azure Machine Learning studio.  
 You need to select the appropriate compute types.  
 Which compute types should you select? To answer, drag the appropriate compute types to the correct requirements. Each compute type may be used once, more than once, or not at all. You may need to drag the split bar between panes or scroll to view content.  
 NOTE: Each correct selection is worth one point.

**Compute types**

Attached compute

Inference cluster

Training cluster

**Answer Area**

Requirement	Compute type
Train models by using the Azure Machine Learning designer.	<div>Compute type</div>
Score new data through a trained model published as a real-time web service.	<div>Compute type</div>
Train models by using an Azure Databricks cluster.	<div>Compute type</div>
Deploy models by using the Azure Machine Learning designer.	<div>Compute type</div>

- A. Mastered
- B. Not Mastered

**Answer:** A

**Explanation:**

Box 1: Attached compute

Training targets	Automated ML	ML pipelines	Azure Machine Learning designer
Local computer	yes		
Azure Machine Learning compute cluster	yes & hyperparameter tuning	yes	yes
<div>Azure Machine Learning compute instance</div>	yes & hyperparameter tuning	yes	<div>yes</div>

Box 2: Inference cluster Box 3: Training cluster Box 4: Attached compute

**NEW QUESTION 141**

- (Exam Topic 3)  
 You have the following code. The code prepares an experiment to run a script:

```

from azureml.core import Workspace, Experiment, Run, ScriptRunConfig

ws = Workspace.from_config()
script_config = ScriptRunConfig(source_directory='experiment_files',
                                script='experiment.py')

script_experiment = Experiment(workspace=ws, name='script-experiment')
```

The experiment must be run on local computer using the default environment. You need to add code to start the experiment and run the script.  
 Which code segment should you use?

- A. run = script\_experiment.start\_logging()
- B. run = Run(experiment=script\_experiment)
- C. ws.get\_run(run\_id=experiment.id)
- D. run = script\_experiment.submit(config=script\_config)

**Answer:** D

**Explanation:**

The experiment class submit method submits an experiment and return the active created run.  
 Syntax: submit(config, tags=None, \*\*kwargs) Reference:  
<https://docs.microsoft.com/en-us/python/api/azureml-core/azureml.core.experiment.experiment>

**NEW QUESTION 146**

- (Exam Topic 3)  
 You are performing a classification task in Azure Machine Learning Studio.  
 You must prepare balanced testing and training samples based on a provided data set. You need to split the data with a 0.75:0.25 ratio.  
 Which value should you use for each parameter? To answer, select the appropriate options in the answer area.  
 NOTE: Each correct selection is worth one point.

Parameter	Value
Splitting mode	<div>▼</div> <div>Split rows Recommender Split Regular Expression Split Relative Expression Split</div>
Fraction of rows in the first output dataset	<div>▼</div> <div>0.75 0.25 0.5 1</div>
Randomized split	<div>▼</div> <div>True False</div>
Stratified split	<div>▼</div> <div>True False</div>

- A. Mastered  
B. Not Mastered

Answer: A

**Explanation:**

Box 1: Split rows

Use the Split Rows option if you just want to divide the data into two parts. You can specify the percentage of data to put in each split, but by default, the data is divided 50-50.

You can also randomize the selection of rows in each group, and use stratified sampling. In stratified sampling, you must select a single column of data for which you want values to be apportioned equally among the two result datasets.

Box 2: 0.75

If you specify a number as a percentage, or if you use a string that contains the "%" character, the value is interpreted as a percentage. All percentage values must be within the range (0, 100), not including the values 0 and 100.

Box 3: Yes

To ensure splits are balanced.

Box 4: No

If you use the option for a stratified split, the output datasets can be further divided by subgroups, by selecting a strata column.

Reference:

<https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/split-data>

**NEW QUESTION 148**

- (Exam Topic 3)

Note: This question is part of a series of questions that present the same scenario. Each question in the series contains a unique solution that might meet the stated goals. Some question sets might have more than one correct solution, while others might not have a correct solution.

After you answer a question in this section, you will NOT be able to return to it. As a result, these questions will not appear in the review screen.

You train and register a machine learning model.

You plan to deploy the model as a real-time web service. Applications must use key-based authentication to use the model.

You need to deploy the web service. Solution:

Create an AciWebservice instance.

Set the value of the ssl\_enabled property to True.

Deploy the model to the service. Does the solution meet the goal?

- A. Yes  
B. No

Answer: B

**Explanation:**

Instead use only auth\_enabled = TRUE Note: Key-based authentication.

Web services deployed on AKS have key-based auth enabled by default. ACI-deployed services have

key-based auth disabled by default, but you can enable it by setting auth\_enabled = TRUE when creating the ACI web service. The following is an example of creating an ACI deployment configuration with key-based auth enabled.

```
deployment_config <- aci_webservice_deployment_config(cpu_cores = 1,
```

```
memory_gb = 1, auth_enabled = TRUE)
```

Reference:

<https://azure.github.io/azureml-sdk-for-r/articles/deploying-models.html>

NEW QUESTION 153

- (Exam Topic 3)

You are retrieving data from a large datastore by using Azure Machine Learning Studio.  
You must create a subset of the data for testing purposes using a random sampling seed based on the system clock.  
You add the Partition and Sample module to your experiment. You need to select the properties for the module.  
Which values should you select? To answer, select the appropriate options in the answer area.  
NOTE: Each correct selection is worth one point.

Partition and Sample

Partition or sample mode

Assign to Folds  
Pick Fold  
Sampling  
Head

Rate of sampling

.2

Random seed for sampling

0  
1  
time.clock()  
utcNow()

Stratified split for sampling

False

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

Box 1: Sampling Create a sample of data  
This option supports simple random sampling or stratified random sampling. This is useful if you want to create a smaller representative sample dataset for testing.  
\* 1. Add the Partition and Sample module to your experiment in Studio, and connect the dataset.  
\* 2. Partition or sample mode: Set this to Sampling.  
\* 3. Rate of sampling.  
See box 2 below.  
Box 2: 0  
\* 3. Rate of sampling. Random seed for sampling: Optionally, type an integer to use as a seed value.  
This option is important if you want the rows to be divided the same way every time. The default value is 0, meaning that a starting seed is generated based on the system clock. This can lead to slightly different results each time you run the experiment.  
References:  
<https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/partition-and-sample>

NEW QUESTION 157

- (Exam Topic 3)

You create an Azure Machine Learning workspace and set up a development environment. You plan to train a deep neural network (DNN) by using the Tensorflow framework and by using estimators to submit training scripts.  
You must optimize computation speed for training runs.  
You need to choose the appropriate estimator to use as well as the appropriate training compute target configuration.  
Which values should you use? To answer, select the appropriate options in the answer area. NOTE: Each correct selection is worth one point.

Answer Area

Parameter	Value
Estimator	<div>Estimator SkLearn PyTorch Tensorflow Chainer</div>
Training compute	<div>12 vCPU, 48 GB memory, 96 GB SSD 12 vCPU, 112 GB memory, 680 GB SSD, 2 GPU, 24 GB GPU memory 16 vCPU, 128 GB memory, 160 GB HDD, 80 GB NVME disk (4000 MBps) 44 vCPU, 352 GB memory, 3.4 GHz CPU frequency all cores</div>

Use the selected estimator for training compute.

- A. Mastered
- B. Not Mastered



**Answer:** A

**Explanation:**

Box 1: Tensorflow

TensorFlow represents an estimator for training in TensorFlow experiments. Box 2: 12 vCPU, 112 GB memory...,2 GPU,..

Use GPUs for the deep neural network. Reference:

<https://docs.microsoft.com/en-us/python/api/azureml-train-core/azureml.train.dnn>

**NEW QUESTION 159**

- (Exam Topic 3)

You use Azure Machine Learning designer to create a real-time service endpoint. You have a single Azure Machine Learning service compute resource. You train the model and prepare the real-time pipeline for deployment You need to publish the inference pipeline as a web service. Which compute type should you use?

- A. HDInsight
- B. Azure Databricks
- C. Azure Kubernetes Services
- D. the existing Machine Learning Compute resource
- E. a new Machine Learning Compute resource

**Answer:** C

**Explanation:**

Azure Kubernetes Service (AKS) can be used real-time inference. Reference:

<https://docs.microsoft.com/en-us/azure/machine-learning/concept-compute-target>

**NEW QUESTION 160**

- (Exam Topic 3)

You are creating a machine learning model. You have a dataset that contains null rows.

You need to use the Clean Missing Data module in Azure Machine Learning Studio to identify and resolve the null and missing data in the dataset.

Which parameter should you use?

- A. Replace with mean
- B. Remove entire column
- C. Remove entire row
- D. Hot Deck

**Answer:** B

**Explanation:**

Remove entire row: Completely removes any row in the dataset that has one or more missing values. This is useful if the missing value can be considered randomly missing.

References:

<https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/clean-missing-data>

**NEW QUESTION 163**

- (Exam Topic 3)

You create a binary classification model to predict whether a person has a disease. You need to detect possible classification errors.

Which error type should you choose for each description? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.

Description	Error type
A person has a disease. The model classifies the case as having a disease.	<div>▼</div> <div>True Positives</div> <div>True Negatives</div> <div>False Positives</div> <div>False Negatives</div>
A person does not have a disease. The model classifies the case as having no disease.	<div>▼</div> <div>True Positives</div> <div>True Negatives</div> <div>False Positives</div> <div>False Negatives</div>
A person does not have a disease. The model classifies the case as having a disease.	<div>▼</div> <div>True Positives</div> <div>True Negatives</div> <div>False Positives</div> <div>False Negatives</div>
A person has a disease. The model classifies the case as having no disease.	<div>▼</div> <div>True Positives</div> <div>True Negatives</div> <div>False Positives</div> <div>False Negatives</div>

- A. Mastered  
 B. Not Mastered

Answer: A

#### Explanation:

Box 1: True Positive

A true positive is an outcome where the model correctly predicts the positive class Box 2: True Negative

A true negative is an outcome where the model correctly predicts the negative class. Box 3: False Positive

A false positive is an outcome where the model incorrectly predicts the positive class. Box 4: False Negative

A false negative is an outcome where the model incorrectly predicts the negative class. Note: Let's make the following definitions:

"Wolf" is a positive class. "No wolf" is a negative class.

We can summarize our "wolf-prediction" model using a 2x2 confusion matrix that depicts all four possible outcomes:

Reference:

<https://developers.google.com/machine-learning/crash-course/classification/true-false-positive-negative>

#### NEW QUESTION 165

- (Exam Topic 3)

You create a binary classification model. You need to evaluate the model performance.

Which two metrics can you use? Each correct answer presents a complete solution.

NOTE: Each correct selection is worth one point.

- A. relative absolute error  
 B. precision  
 C. accuracy  
 D. mean absolute error  
 E. coefficient of determination

Answer: BC

#### Explanation:

The evaluation metrics available for binary classification models are: Accuracy, Precision, Recall, F1 Score, and AUC.

Note: A very natural question is: 'Out of the individuals whom the model, how many were classified correctly (TP)?'

This question can be answered by looking at the Precision of the model, which is the proportion of positives that are classified correctly.

References:

<https://docs.microsoft.com/en-us/azure/machine-learning/studio/evaluate-model-performance>

#### NEW QUESTION 169

- (Exam Topic 3)

You are building a regression model tot estimating the number of calls during an event.

You need to determine whether the feature values achieve the conditions to build a Poisson regression model. Which two conditions must the feature set contain?

Each correct answer presents part of the solution. NOTE:

Each correct selection is worth one point.

- A. The label data must be a negative value.  
 B. The label data can be positive or negative,

- C. The label data must be a positive value
- D. The label data must be non discrete.
- E. The data must be whole numbers.

**Answer:** CE

**Explanation:**

Poisson regression is intended for use in regression models that are used to predict numeric values, typically counts. Therefore, you should use this module to create your regression model only if the values you are trying to predict fit the following conditions:

- The response variable has a Poisson distribution.
- Counts cannot be negative. The method will fail outright if you attempt to use it with negative labels.
- A Poisson distribution is a discrete distribution; therefore, it is not meaningful to use this method with non-whole numbers.

References:

<https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/poisson-regression>

**NEW QUESTION 172**

- (Exam Topic 3)

You are performing a filter based feature selection for a dataset 10 build a multi class classifies by using Azure Machine Learning Studio.

The dataset contains categorical features that are highly correlated to the output label column.

You need to select the appropriate feature scoring statistical method to identify the key predictors. Which method should you use?

- A. Chi-squared
- B. Spearman correlation
- C. Kendall correlation
- D. Person correlation

**Answer:** D

**Explanation:**

Pearson's correlation statistic, or Pearson's correlation coefficient, is also known in statistical models as the r value. For any two variables, it returns a value that indicates the strength of the correlation

Pearson's correlation coefficient is the test statistics that measures the statistical relationship, or association, between two continuous variables. It is known as the best method of measuring the association between variables of interest because it is based on the method of covariance. It gives information about the magnitude of the association, or correlation, as well as the direction of the relationship.

Reference:

<https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/filter-based-feature-selection> <https://www.statisticssolutions.com/pearsons-correlation-coefficient/>

**NEW QUESTION 176**

- (Exam Topic 3)

Note: This question is part of a series of questions that present the same scenario. Each question in the series contains a unique solution that might meet the stated goals. Some question sets might have more than one correct solution, while others might not have a correct solution.

After you answer a question in this section, you will NOT be able to return to it. As a result, these questions will not appear in the review screen.

You plan to use a Python script to run an Azure Machine Learning experiment. The script creates a reference to the experiment run context, loads data from a file, identifies the set of unique values for the label column, and completes the experiment run:

```
from azureml.core import Run
import pandas as pd

run = Run.get_context()
data = pd.read_csv('data.csv')
label_vals = data['label'].unique()
# Add code to record metrics here
run.complete()
```

The experiment must record the unique labels in the data as metrics for the run that can be reviewed later.

You must add code to the script to record the unique label values as run metrics at the point indicated by the comment.

Solution: Replace the comment with the following code: `run.log_table('Label Values', label_vals)`

Does the solution meet the goal?

- A. Yes
- B. No

**Answer:** B

**Explanation:**

Instead use the `run_log` function to log the contents in `label_vals`: `for label_val in label_vals:`

`run.log('Label Values', label_val)` Reference:

<https://www.element61.be/en/resource/azure-machine-learning-services-complete-toolbox-ai>

**NEW QUESTION 178**

- (Exam Topic 3)

Note: This question is part of a series of questions that present the same scenario. Each question in the series contains a unique solution that might meet the stated goals. Some question sets might have more than one correct solution, while others might not have a correct solution.

After you answer a question in this section, you will NOT be able to return to it. As a result, these questions will not appear in the review screen.

You are a data scientist using Azure Machine Learning Studio.

You need to normalize values to produce an output column into bins to predict a target column. Solution: Apply a Quantiles binning mode with a PQuantile normalization.

Does the solution meet the goal?

- A. Yes
- B. No

**Answer:** B

**Explanation:**

Use the Entropy MDL binning mode which has a target column. References:

<https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/group-data-into-bins>

**NEW QUESTION 180**

- (Exam Topic 3)

You create a script that trains a convolutional neural network model over multiple epochs and logs the validation loss after each epoch. The script includes arguments for batch size and learning rate.

You identify a set of batch size and learning rate values that you want to try.

You need to use Azure Machine Learning to find the combination of batch size and learning rate that results in the model with the lowest validation loss.

What should you do?

- A. Run the script in an experiment based on an AutoMLConfig object
- B. Create a PythonScriptStep object for the script and run it in a pipeline
- C. Use the Automated Machine Learning interface in Azure Machine Learning studio
- D. Run the script in an experiment based on a ScriptRunConfig object
- E. Run the script in an experiment based on a HyperDriveConfig object

**Answer:** E

**Explanation:**

Reference:

<https://docs.microsoft.com/en-us/azure/machine-learning/how-to-tune-hyperparameters>

**NEW QUESTION 182**

- (Exam Topic 3)

You are a lead data scientist for a project that tracks the health and migration of birds. You create a multi-class image classification deep learning model that uses a set of labeled bird photographs collected by experts.

You have 100,000 photographs of birds. All photographs use the JPG format and are stored in an Azure blob container in an Azure subscription.

You need to access the bird photograph files in the Azure blob container from the Azure Machine Learning service workspace that will be used for deep learning model training. You must minimize data movement.

What should you do?

- A. Create an Azure Data Lake store and move the bird photographs to the store.
- B. Create an Azure Cosmos DB database and attach the Azure Blob containing bird photographs storage to the database.
- C. Create and register a dataset by using TabularDataset class that references the Azure blob storage containing bird photographs.
- D. Register the Azure blob storage containing the bird photographs as a datastore in Azure Machine Learning service.
- E. Copy the bird photographs to the blob datastore that was created with your Azure Machine Learning service workspace.

**Answer:** D

**Explanation:**

We recommend creating a datastore for an Azure Blob container. When you create a workspace, an Azure blob container and an Azure file share are automatically registered to the workspace.

Reference:

<https://docs.microsoft.com/en-us/azure/machine-learning/how-to-access-data>

**NEW QUESTION 185**

- (Exam Topic 3)

You create a multi-class image classification deep learning model. You train the model by using PyTorch version 1.2.

You need to ensure that the correct version of PyTorch can be identified for the inferencing environment when the model is deployed.

What should you do?

- A. Save the model locally as a .pt file, and deploy the model as a local web service.
- B. Deploy the model on computer that is configured to use the default Azure Machine Learning conda environment.
- C. Register the model with a .pt file extension and the default version property.
- D. Register the model, specifying the model\_framework and model\_framework\_version properties.

**Answer:** D

**Explanation:**

framework\_version: The PyTorch version to be used for executing training code. Reference:

<https://docs.microsoft.com/en-us/python/api/azureml-train-core/azureml.train.dnn.pytorch?view=azure-ml-py>

**NEW QUESTION 190**

- (Exam Topic 3)

You create a binary classification model. The model is registered in an Azure Machine Learning workspace. You use the Azure Machine Learning Fairness SDK to assess the model fairness.

You develop a training script for the model on a local machine.

You need to load the model fairness metrics into Azure Machine Learning studio. What should you do?

- A. Implement the download\_dashboard\_by\_upload\_id function
- B. Implement the create\_group\_metric\_sec function
- C. Implement the upload\_dashboard\_dictionary function
- D. Upload the training script

**Answer:** C

**Explanation:**



import azureml.contrib.fairness package to perform the upload:  
from azureml.contrib.fairness import upload\_dashboard\_dictionary, download\_dashboard\_by\_upload\_id Reference:  
<https://docs.microsoft.com/en-us/azure/machine-learning/how-to-machine-learning-fairness-aml>

**NEW QUESTION 193**

- (Exam Topic 3)

You create the following config.json file.

```
{  
  "workspace_name" : "ml-workspace"  
}
```

You must use the Azure Machine Learning SDK to interact with data and experiments in the workspace. You need to configure the config.json file to connect to the workspace from the Python environment. Which two additional parameters must you add to the config.json file in order to connect to the workspace?

Each correct answer presents part of the solution.

NOTE: Each correct selection is worth one point.

- A. subscription\_id
- B. Key
- C. resource\_group
- D. region
- E. Login

**Answer:** AC

**Explanation:**

To use the same workspace in multiple environments, create a JSON configuration file. The configuration file saves your subscription (subscription\_id), resource (resource\_group), and workspace name so that it can be easily loaded.

The following sample shows how to create a workspace. from azureml.core import Workspace

```
ws = Workspace.create(name='myworkspace', subscription_id='<azure-subscription-id>', resource_group='myresourcegroup', create_resource_group=True,  
location='eastus2'  
)
```

Reference:

<https://docs.microsoft.com/en-us/python/api/azureml-core/azureml.core.workspace.workspace>

**NEW QUESTION 197**

- (Exam Topic 3)

You run a script as an experiment in Azure Machine Learning.

You have a Run object named run that references the experiment run. You must review the log files that were generated during the experiment run.

You need to download the log files to a local folder for review.

Which two code segments can you run to achieve this goal? Each correct answer presents a complete solution.

NOTE: Each correct selection is worth one point.

- A. run.get\_details()
- B. run.get\_file\_names()
- C. run.get\_metrics()
- D. run.download\_files(output\_directory='./runfiles')
- E. run.get\_all\_logs(destination='./runlogs')

**Answer:** AE

**Explanation:**

The run Class get\_all\_logs method downloads all logs for the run to a directory.

The run Class get\_details gets the definition, status information, current log files, and other details of the run. Reference:

[https://docs.microsoft.com/en-us/python/api/azureml-core/azureml.core.run\(class\)](https://docs.microsoft.com/en-us/python/api/azureml-core/azureml.core.run(class))

**NEW QUESTION 198**

- (Exam Topic 3)

Note: This question is part of a series of questions that present the same scenario. Each question in the series contains a unique solution that might meet the stated goals. Some question sets might have more than one correct solution, while others might not have a correct solution.

After you answer a question in this section, you will NOT be able to return to it. As a result, these questions will not appear in the review screen.

You are using Azure Machine Learning to run an experiment that trains a classification model.

You want to use Hyperdrive to find parameters that optimize the AUC metric for the model. You configure a HyperDriveConfig for the experiment by running the following code:

```
hyperdrive = HyperDriveConfig(estimator=your_estimator,  
hyperparameter_sampling=your_params,  
policy=policy,  
primary_metric_name='AUC',  
primary_metric_goal=PrimaryMetricGoal.MAXIMIZE,  
max_total_runs=6,  
max_concurrent_runs=4)
```

You plan to use this configuration to run a script that trains a random forest model and then tests it with validation data. The label values for the validation data are stored in a variable named y\_test variable, and the predicted probabilities from the model are stored in a variable named y\_predicted.

Solution: Run the following code:

```
import numpy as np
from sklearn.metrics import roc_auc_score
from azureml.core.run import Run
run = Run.get_context()
# code to train model omitted
auc = roc_auc_score(y_test, y_predicted)
run.log("AUC", np.float(auc))
```

Does the solution meet the goal?

- A. Yes
- B. No

**Answer: A**

#### NEW QUESTION 199

- (Exam Topic 3)

You register a model that you plan to use in a batch inference pipeline.

The batch inference pipeline must use a ParallelRunStep step to process files in a file dataset. The script has the ParallelRunStep step runs must process six input files each time the inferencing function is called.

You need to configure the pipeline.

Which configuration setting should you specify in the ParallelRunConfig object for the ParallelRunStep step?

- A. process\_count\_per\_node= "6"
- B. node\_count= "6"
- C. mini\_batch\_size= "6"
- D. error\_threshold= "6"

**Answer: B**

#### Explanation:

node\_count is the number of nodes in the compute target used for running the ParallelRunStep. Reference:

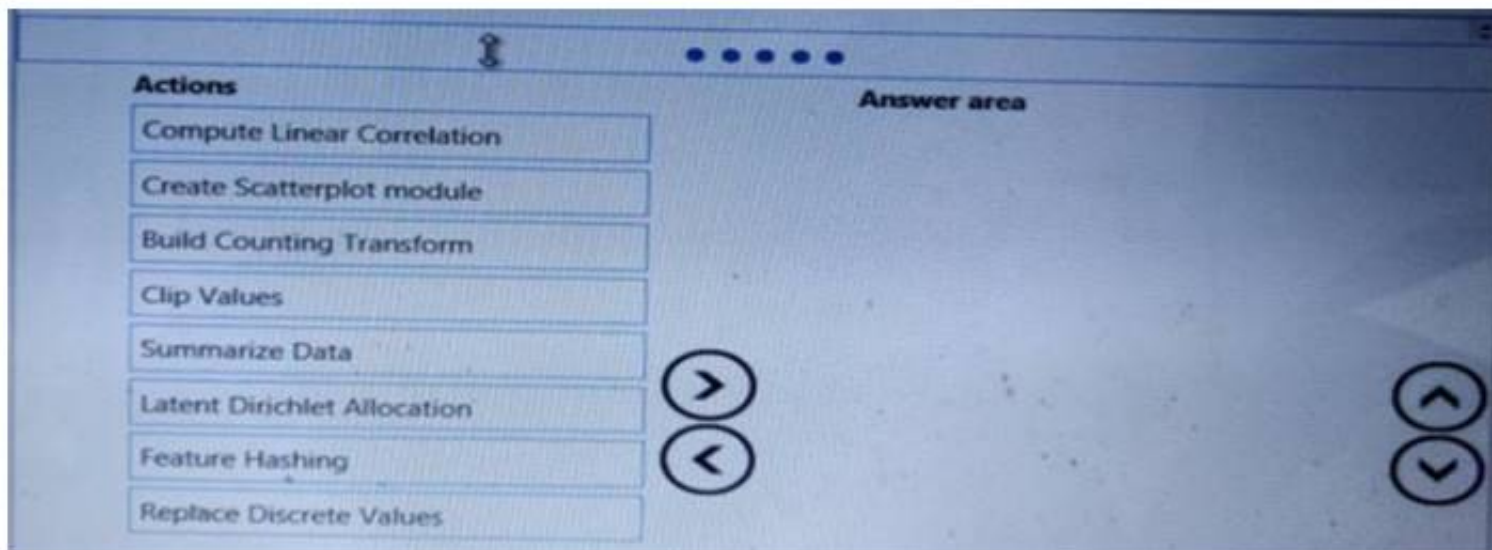
<https://docs.microsoft.com/en-us/python/api/azureml-contrib-pipeline-steps/azureml.contrib.pipeline.steps.parall>

#### NEW QUESTION 202

- (Exam Topic 2)

You need to visually identify whether outliers exist in the Age column and quantify the outliers before the outliers are removed.

Which three Azure Machine Learning Studio modules should you use in sequence? To answer, move the appropriate modules from the list of modules to the answer area and arrange them in the correct order.



- A. Mastered
- B. Not Mastered

**Answer: A**

#### Explanation:

Create Scatterplot Summarize Data Clip Values

You can use the Clip Values module in Azure Machine Learning Studio, to identify and optionally replace data values that are above or below a specified threshold. This is useful when you want to remove outliers or replace them with a mean, a constant, or other substitute value.

References:

<https://blogs.msdn.microsoft.com/azuredev/2017/05/27/data-cleansing-tools-in-azure-machine-learning/> <https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/clip-values>

#### NEW QUESTION 203

- (Exam Topic 2)

You need to produce a visualization for the diagnostic test evaluation according to the data visualization requirements.

Which three modules should you recommend be used in sequence? To answer, move the appropriate modules from the list of modules to the answer area and arrange them in the correct order.

Modules

Score Matchbox Recommender

Apply Transformation

Evaluate Recommender

Evaluate Model

Train Model

Sweep Clustering

Score Model

Load Trained Model

Answer Area

⏪

⏩

⏴

⏵

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

Step 1: Sweep Clustering  
Start by using the "Tune Model Hyperparameters" module to select the best sets of parameters for each of the models we're considering. One of the interesting things about the "Tune Model Hyperparameters" module is that it not only outputs the results from the Tuning, it also outputs the Trained Model.  
Step 2: Train Model Step 3: Evaluate Model  
Scenario: You need to provide the test results to the Fabrikam Residences team. You create data visualizations to aid in presenting the results. You must produce a Receiver Operating Characteristic (ROC) curve to conduct a diagnostic test evaluation of the model. You need to select appropriate methods for producing the ROC curve in Azure Machine Learning Studio to compare the Two-Class Decision Forest and the Two-Class Decision Jungle modules with one another.  
References:  
<http://breaking-bi.blogspot.com/2017/01/azure-machine-learning-model-evaluation.html>

NEW QUESTION 206

- (Exam Topic 2)  
You need to configure the Permutation Feature Importance module for the model training requirements. What should you do? To answer, select the appropriate options in the dialog box in the answer area. NOTE: Each correct selection is worth one point.

Answer Area

Permutation Feature importance

Random seed

0

500

Regression – Root Mean Square Error

Regression – R-squared

Regression – Mean Zero One Error

Regression – Mean Absolute Error

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

Box 1: 500  
For Random seed, type a value to use as seed for randomization. If you specify 0 (the default), a number is generated based on the system clock. A seed value is optional, but you should provide a value if you want reproducibility across runs of the same experiment. Here we must replicate the findings. Box 2: Mean Absolute Error  
Scenario: Given a trained model and a test dataset, you must compute the Permutation Feature Importance scores of feature variables. You need to set up the Permutation Feature Importance module to select the correct metric to investigate the model's accuracy and replicate the findings.

Regression. Choose one of the following: Precision, Recall, Mean Absolute Error , Root Mean Squared Error, Relative Absolute Error, Relative Squared Error, Coefficient of Determination  
References:  
<https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/permutation-feature-importance>

**NEW QUESTION 207**

- (Exam Topic 2)  
You need to configure the Feature Based Feature Selection module based on the experiment requirements and datasets.  
How should you configure the module properties? To answer, select the appropriate options in the dialog box in the answer area.  
NOTE: Each correct selection is worth one point.

Filter Based Feature Selection

Feature scoring method

Fisher Score  
Chi-squared  
Mutual information  
Counts

☒ Operate on feature columns only

Target column

MedianValue  
AvgRooms/nHouse

Launch column selector

Number of desired features

1

- A. Mastered
- B. Not Mastered

Answer: A

**Explanation:**  
Box 1: Mutual Information.  
The mutual information score is particularly useful in feature selection because it maximizes the mutual information between the joint distribution and target variables in datasets with many dimensions.  
Box 2: MedianValue  
MedianValue is the feature column, , it is the predictor of the dataset.  
Scenario: The MedianValue and AvgRoomsinHouse columns both hold data in numeric format. You need to select a feature selection algorithm to analyze the relationship between the two columns in more detail.  
References:  
<https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/filter-based-feature-selection>

**NEW QUESTION 211**

- (Exam Topic 1)  
You need to modify the inputs for the global penalty event model to address the bias and variance issue. Which three actions should you perform in sequence? To answer, move the appropriate actions from the list of actions to the answer area and arrange them in the correct order.

Actions

Build ratios.

Bin the new data.

Add a K-Means clustering module with 10 clusters.

Select the behavior data.

Select the location data.

Perform a Primary Component Analysis (PCA).

>

<

Answer area

<

>

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:



Actions

Build ratios.

Bin the new data.

Add a K-Means clustering module with 10 clusters.

Select the behavior data.

Select the location data.

Perform a Primary Component Analysis (PCA).

Answer area

Select the behavior data.

Add a K-Means clustering module with 10 clusters.

Perform a Primary Component Analysis (PCA).

NEW QUESTION 216

.....

## Thank You for Trying Our Product

### We offer two products:

1st - We have Practice Tests Software with Actual Exam Questions

2nd - Questions and Answers in PDF Format

### DP-100 Practice Exam Features:

- \* DP-100 Questions and Answers Updated Frequently
- \* DP-100 Practice Questions Verified by Expert Senior Certified Staff
- \* DP-100 Most Realistic Questions that Guarantee you a Pass on Your FirstTry
- \* DP-100 Practice Test Questions in Multiple Choice Formats and Updatesfor 1 Year

**100% Actual & Verified — Instant Download, Please Click**  
**[Order The DP-100 Practice Test Here](#)**