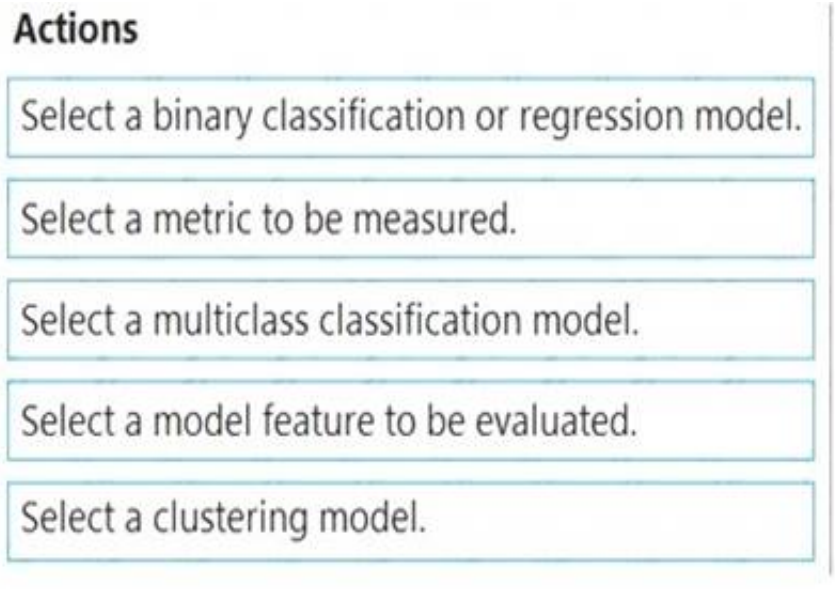
Exam Questions DP-100

Designing and Implementing a Data Science Solution on Azure

https://www.2passeasy.com/dumps/DP-100/

NEW QUESTION 1
- (Exam Topic 3)
You have several machine learning models registered in an Azure Machine Learning workspace. You must use the Fairlearn dashboard to assess fairness in a selected model.
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.



A. Mastered
B. Not Mastered

Answer: A

Explanation:
Graphical user interface, text, application Description automatically generated
Step 1: Select a model feature to be evaluated.
Step 2: Select a binary classification or regression model.
Register your models within Azure Machine Learning. For convenience, store the results in a dictionary,
which maps the id of the registered model (a string in name:version format) to the predictor itself. Example:
model_dict = {}
lr_reg_id = register_model("fairness_logistic_regression", lr_predictor) model_dict[lr_reg_id] = lr_predictor
svm_reg_id = register_model("fairness_svm", svm_predictor) model_dict[svm_reg_id] = svm_predictor
Step 3: Select a metric to be measured Precompute fairness metrics.
Create a dashboard dictionary using Fairlearn's metrics package. Reference:
https://docs.microsoft.com/en-us/azure/machine-learning/how-to-machine-learning-fairness-aml

NEW QUESTION 2
- (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.
An IT department creates the following Azure resource groups and resources:

Resource group	Resources
ml_resources	• an Azure Machine Learning workspace named amlworkspace • an Azure Storage account named amlworkspace12345 • an Application Insights instance named amlworkspace54321 • an Azure Key Vault named amlworkspace6789O • an Azure Container Registry named amlworkspace09876
general_compute	A virtual machine named mlvm with the following configuration: • Operating system: Ubuntu Linux • Software installed: Python 3.6 and Jupyter Notebooks • Size: NC6 (6 vCPUs, 1 vGPU, 56 Gb RAM)

The IT department creates an Azure Kubernetes Service (AKS)-based inference compute target named
aks-cluster in the Azure Machine Learning workspace. You have a Microsoft Surface Book computer with a GPU. Python 3.6 and Visual Studio Code are installed.
You need to run a script that trains a deep neural network (DNN) model and logs the loss and accuracy metrics.
Solution: Install the Azure ML SDK on the Surface Book. Run Python code to connect to the workspace. Run the training script as an experiment on the aks-cluster compute target.
Does the solution meet the goal?

A. Yes
B. No

Answer: B

Explanation:
Need to attach the mlvm virtual machine as a compute target in the Azure Machine Learning workspace.
Reference:
https://docs.microsoft.com/en-us/azure/machine-learning/concept-compute-target

NEW QUESTION 3

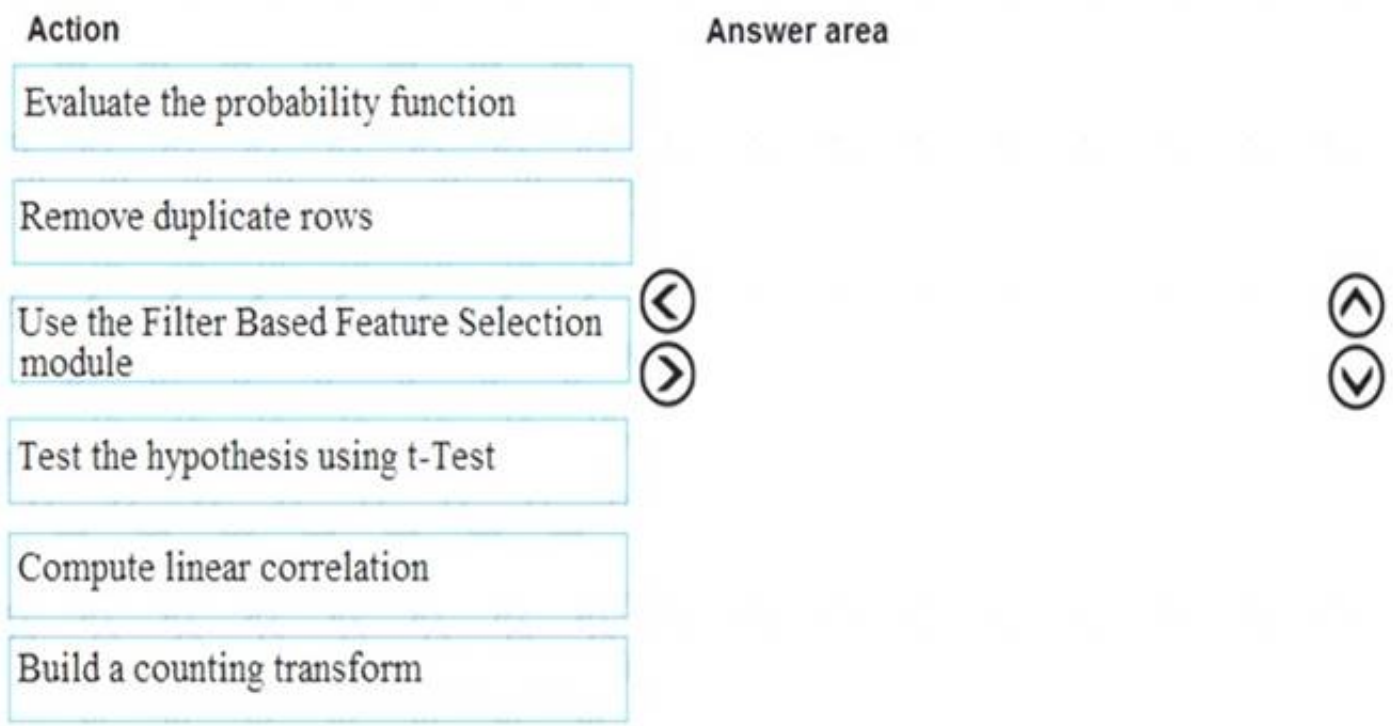
- (Exam Topic 3)
You train a machine learning model.
You must deploy the model as a real-time inference service for testing. The service requires low CPU utilization and less than 48 MB of RAM. The compute target for the deployed service must initialize automatically while minimizing cost and administrative overhead.
Which compute target should you use?

A. Azure Kubernetes Service (AKS) inference cluster
B. Azure Machine Learning compute cluster
C. Azure Container Instance (ACI)
D. attached Azure Databricks cluster

Answer: C

Explanation:
Azure Container Instances (ACI) are suitable only for small models less than 1 GB in size. Use it for low-scale CPU-based workloads that require less than 48 GB of RAM.
Note: Microsoft recommends using single-node Azure Kubernetes Service (AKS) clusters for dev-test of larger models.
Reference:
https://docs.microsoft.com/id-id/azure/machine-learning/how-to-deploy-and-where

NEW QUESTION 4
- (Exam Topic 3)
You are producing a multiple linear regression model in Azure Machine Learning Studio. Several independent variables are highly correlated.
You need to select appropriate methods for conducting effective feature engineering on all the data.
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.



A. Mastered
B. Not Mastered

Answer: A

Explanation:
Step 1: Use the Filter Based Feature Selection module
Filter Based Feature Selection identifies the features in a dataset with the greatest predictive power.
The module outputs a dataset that contains the best feature columns, as ranked by predictive power. It also outputs the names of the features and their scores from the selected metric.
Step 2: Build a counting transform
A counting transform creates a transformation that turns count tables into features, so that you can apply the transformation to multiple datasets.
Step 3: Test the hypothesis using t-Test References:
https://docs.microsoft.com/bs-latn-ba/azure/machine-learning/studio-module-reference/filter-based-feature-selec
https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/build-counting-transform

NEW QUESTION 5
- (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 6
- (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.dnn import TensorFlow
sk_est = TensorFlow(source_directory='./scripts',
 compute_target=aml-compute,
 entry_script='train.py')
```

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 7
- (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_list('Label Values', label_vals) Does the solution meet the goal?

A. Yes
B. No

Answer: A

Explanation:
run.log_list log a list of values to the run with the given name using log_list. Example: run.log_list("accuracies", [0.6, 0.7, 0.87])
Note:
Data= pd.read_csv('data.csv')
Data is read into a pandas.DataFrame, which is a two-dimensional, size-mutable, potentially heterogeneous tabular data.
label_vals =data['label'].unique

label_vals contains a list of unique label values. Reference:
https://www.element61.be/en/resource/azure-machine-learning-services-complete-toolbox-ai https://docs.microsoft.com/en-us/python/api/azureml-core/azureml.core.run(class) https://pandas.pydata.org/docs/reference/api/pandas.DataFrame.html

NEW QUESTION 8
- (Exam Topic 3)
You retrain an existing model.
You need to register the new version of a model while keeping the current version of the model in the registry.
What should you do?

A. Register a model with a different name from the existing model and a custom property named versionwith the value 2.
B. Register the model with the same name as the existing model.
C. Save the new model in the default datastore with the same name as the existing mode
D. Do not register the new model.
E. Delete the existing model and register the new one with the same name.

Answer: B

Explanation:
Model version: A version of a registered model. When a new model is added to the Model Registry, it is added as Version 1. Each model registered to the same model name increments the version number.
Reference:
https://docs.microsoft.com/en-us/azure/databricks/applications/mlflow/model-registry

NEW QUESTION 9
- (Exam Topic 3)
You are creating a machine learning model. You need to identify outliers in the data.
Which two visualizations can you use? Each correct answer presents a complete solution.
NOTE: Each correct selection is worth one point. NOTE: Each correct selection is worth one point.

A. box plot
B. scatter
C. random forest diagram
D. Venn diagram
E. ROC curve

Answer: AB

Explanation:
The box-plot algorithm can be used to display outliers.
One other way to quickly identify Outliers visually is to create scatter plots. References:
https://blogs.msdn.microsoft.com/azuredev/2017/05/27/data-cleansing-tools-in-azure-machine-learning/

NEW QUESTION 10
- (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.
You need to add logging to the script to allow Hyperdrive to optimize hyperparameters for the AUC metric. Solution: Run the following code:

```
import numpy as np
from sklearn.metrics import roc_auc_score
# code to train model omitted
auc = roc_auc_score(y_test, y_predicted)
print(np.float(auc))
```

Does the solution meet the goal?

A. Yes
B. No

Answer: B

Explanation:
Use a solution with logging.info(message) instead. Note: Python printing/logging example: logging.info(message)

Destination: Driver logs, Azure Machine Learning designer Reference:
https://docs.microsoft.com/en-us/azure/machine-learning/how-to-debug-pipelines

NEW QUESTION 10
- (Exam Topic 3)
You are evaluating a completed binary classification machine learning model. You need to use the precision as the valuation metric.
Which visualization should you use?

A. Binary classification confusion matrix
B. box plot
C. Gradient descent
D. coefficient of determination

Answer: A

Explanation:
References:
https://machinelearningknowledge.ai/confusion-matrix-and-performance-metrics-machine-learning/

NEW QUESTION 15
- (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 18
- (Exam Topic 3)
You create a datastore named training_data that references a blob container in an Azure Storage account. The blob container contains a folder named csv_files in which multiple comma-separated values (CSV) files are stored.
You have a script named train.py in a local folder named ./script that you plan to run as an experiment using an estimator. The script includes the following code to read data from the csv_files folder:

```
import os
import argparse
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LogisticRegression
from azureml.core import Run

run = Run.get_context()
parser = argparse.ArgumentParser()
parser.add_argument('--data-folder', type=str, dest='data_folder', help='data reference')
args = parser.parse_args()

data_folder = args.data_folder
csv_files = os.listdir(data_folder)
training_data = pd.concat((pd.read_csv(os.path.join(data_folder,csv_file)) for csv_file in csv_files))

# Code goes on to split the training data and train a logistic regression model
```

You have the following script.

```
from azureml.core import Workspace, Datastore, Experiment
from azureml.train.sklearn import SKLearn

ws = Workspace.from_config()
exp = Experiment(workspace=ws, name='csv_training')
ds = Datastore.get(ws, datastore_name='training_data')
data_ref = ds.path('csv_files')

# Code to define estimator goes here

run = exp.submit(config=estimator)
run.wait_for_completion(show_output=True)
```

You need to configure the estimator for the experiment so that the script can read the data from a data reference named data_ref that references the csv_files folder in the training_data datastore.
Which code should you use to configure the estimator?

A.
```
estimator = SKLearn(source_directory='./script',
 inputs=[data_ref.as_named_input('data-folder').to_pandas_dataframe()],
 compute_target='local',
 entry_script='train.py')
```

B.
```
script_params = {
    '--data-folder': data_ref.as_mount()
}
estimator = SKLearn(source_directory='./script',
 script_params=script_params,
 compute_target='local',
 entry_script='train.py'
```

C.
```
estimator = SKLearn(source_directory='./script',
 inputs=[data_ref.as_named_input('data-folder').as_mount()],
 compute_target='local',
 entry_script='train.py')
```

D.
```
script_params = {
    '--data-folder': data_ref.as_download(path_on_compute='csv_files')
}
estimator = SKLearn(source_directory='./script',
 script_params=script_params,
 compute_target='local',
 entry_script='train.py'
```

E.
```
estimator = SKLearn(source_directory='./script',
 inputs=[data_ref.as_named_input('data-folder').as_download(path_on_compute='csv_files')],
 compute_target='local',
 entry_script='train.py')
```

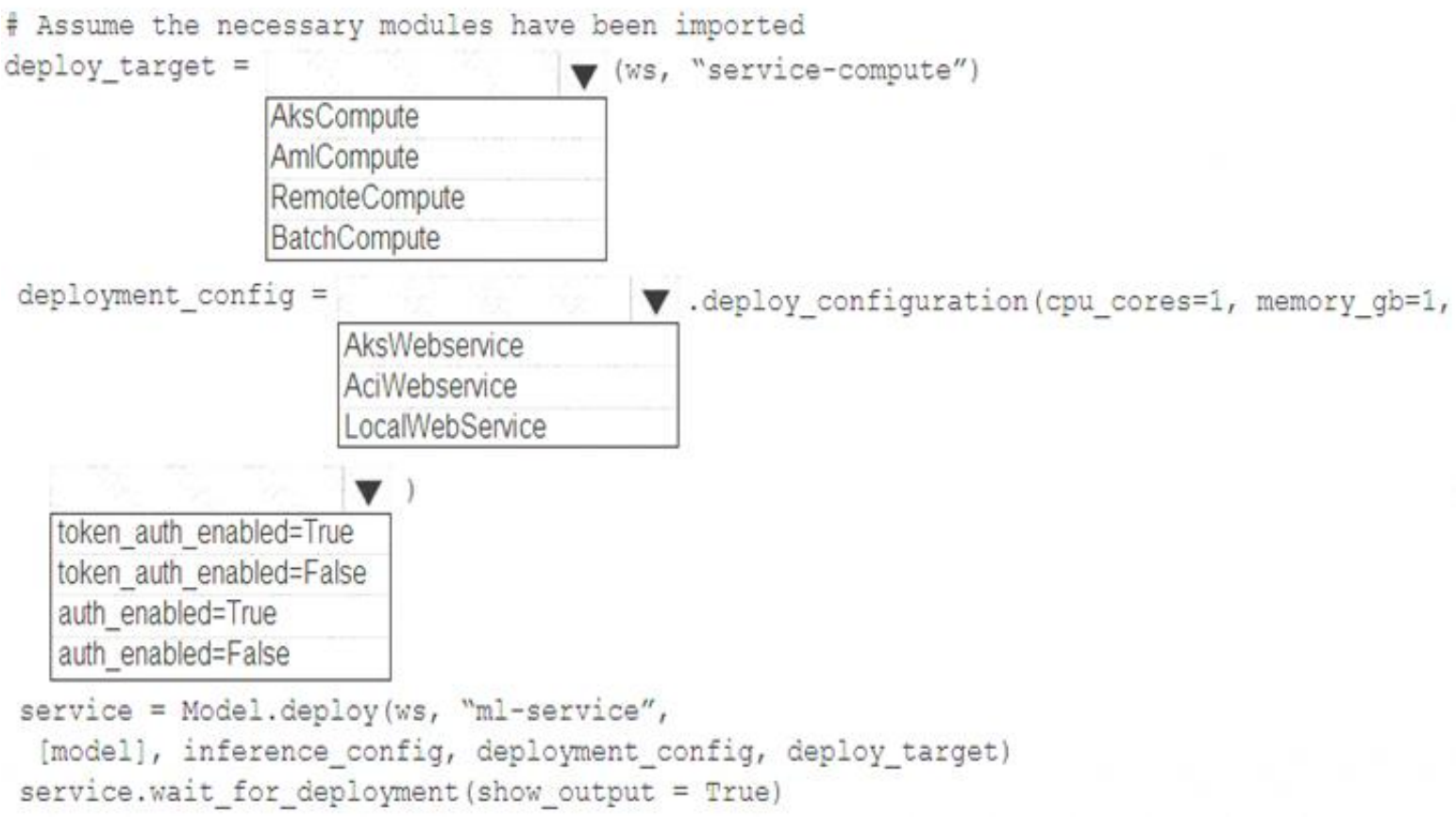
A. Option A
B. Option B
C. Option C
D. Option D
E. Option E

Answer: B

Explanation:
Besides passing the dataset through the inputs parameter in the estimator, you can also pass the dataset through script_params and get the data path (mounting point) in your training script via arguments. This way, you can keep your training script independent of azureml-sdk. In other words, you will be able use the same training script for local debugging and remote training on any cloud platform.
Example:
from azureml.train.sklearn import SKLearn script_params = {
# mount the dataset on the remote compute and pass the mounted path as an argument to the training script '--data-folder':
mnist_ds.as_named_input('mnist').as_mount(),
'--regularization': 0.5
}
est = SKLearn(source_directory=script_folder, script_params=script_params, compute_target=compute_target, environment_definition=env,
entry_script='train_mnist.py')
# Run the experiment
run = experiment.submit(est) run.wait_for_completion(show_output=True) Reference:
https://docs.microsoft.com/es-es/azure/machine-learning/how-to-train-with-datasets

NEW QUESTION 22
- (Exam Topic 3)
You use Azure Machine Learning to train and register a model.
You must deploy the model into production as a real-time web service to an inference cluster named service-compute that the IT department has created in the Azure Machine Learning workspace.
Client applications consuming the deployed web service must be authenticated based on their Azure Active Directory service principal.
You need to write a script that uses the Azure Machine Learning SDK to deploy the model. The necessary modules have been imported.

How should you complete the code? 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:
Box 1: AksCompute Example:
aks_target = AksCompute(ws,"myaks")
# If deploying to a cluster configured for dev/test, ensure that it was created with enough
# cores and memory to handle this deployment configuration. Note that memory is also used by
# things such as dependencies and AML components.
deployment_config = AksWebservice.deploy_configuration(cpu_cores = 1, memory_gb = 1)
service = Model.deploy(ws, "myservice", [model], inference_config, deployment_config, aks_target) Box 2: AksWebservice
Box 3: token_auth_enabled=Yes
Whether or not token auth is enabled for the Webservice.
Note: A Service principal defined in Azure Active Directory (Azure AD) can act as a principal on which authentication and authorization policies can be enforced in Azure Databricks.
The Azure Active Directory Authentication Library (ADAL) can be used to programmatically get an Azure AD access token for a user.
Reference:
https://docs.microsoft.com/en-us/azure/machine-learning/how-to-deploy-azure-kubernetes-service https://docs.microsoft.com/en-us/azure/databricks/dev-tools/api/latest/aad/service-prin-aad-token

NEW QUESTION 27
- (Exam Topic 3)
You use the following code to run a script as an experiment in Azure Machine Learning:

```
from azureml.core import Workspace, Experiment, Run
from azureml.core import RunConfig, ScriptRunConfig
ws = Workspace.from_config()
run_config = RunConfiguration()
run_config.target='local'
script_config = ScriptRunConfig(source_directory='./script', script='experiment.py', run_config=run_config)
experiment = Experiment(workspace=ws, name='script experiment')
run = experiment.submit(config=script_config)
run.wait_for_completion()
```

You must identify the output files that are generated by the experiment run. You need to add code to retrieve the output file names.
Which code segment should you add to the script?

A. files = run.get_properties()
B. files= run.get_file_names()
C. files = run.get_details_with_logs()
D. files = run.get_metrics()
E. files = run.get_details()

Answer: B

Explanation:
You can list all of the files that are associated with this run record by called run.get_file_names() Reference:
https://docs.microsoft.com/en-us/azure/machine-learning/how-to-track-experiments

NEW QUESTION 32
- (Exam Topic 3)
You plan to deliver a hands-on workshop to several students. The workshop will focus on creating data visualizations using Python. Each student will use a device that has internet access.
Student devices are not configured for Python development. Students do not have administrator access to install software on their devices. Azure subscriptions are not available for students.
You need to ensure that students can run Python-based data visualization code. Which Azure tool should you use?

A. Anaconda Data Science Platform
B. Azure BatchAI
C. Azure Notebooks
D. Azure Machine Learning Service

Answer: C

Explanation:
References:
https://notebooks.azure.com/

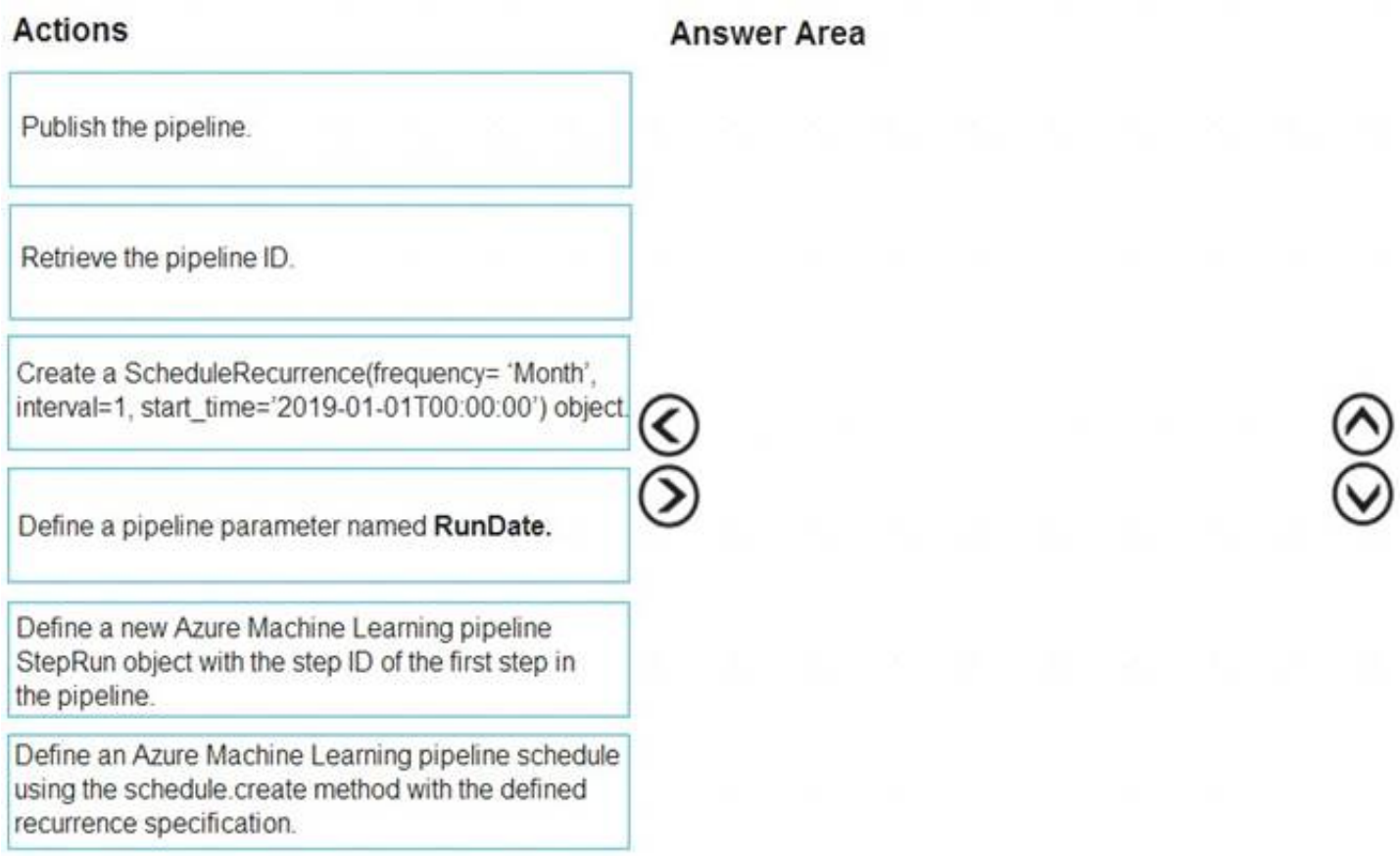
NEW QUESTION 35
- (Exam Topic 3)
Note: This question is part of a series of questions that present the same scenario. Each question in the series contains a unique solution that might meet the stated goals. Some question sets might have more than one correct solution, while others might not have a correct 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 analyzing a numerical dataset which contains missing values in several columns.
You must clean the missing values using an appropriate operation without affecting the dimensionality of the feature set.
You need to analyze a full dataset to include all values.
Solution: Replace each missing value using the Multiple Imputation by Chained Equations (MICE) method. Does the solution meet the goal?

A. Yes
B. NO

Answer: A

Explanation:
Replace using MICE: For each missing value, this option assigns a new value, which is calculated by using a method described in the statistical literature as "Multivariate Imputation using Chained Equations" or "Multiple Imputation by Chained Equations". With a multiple imputation method, each variable with missing data is modeled conditionally using the other variables in the data before filling in the missing values.
Note: Multivariate imputation by chained equations (MICE), sometimes called “fully conditional specification” or “sequential regression multiple imputation” has emerged in the statistical literature as one principled method of addressing missing data. Creating multiple imputations, as opposed to single imputations, accounts for the statistical uncertainty in the imputations. In addition, the chained equations approach is very flexible and can handle variables of varying types (e.g., continuous or binary) as well as complexities such as bounds or survey skip patterns.
References: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3074241/
https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/clean-missing-data

NEW QUESTION 36
- (Exam Topic 3)
You create a multi-class image classification deep learning model.
The model must be retrained monthly with the new image data fetched from a public web portal. You create an Azure Machine Learning pipeline to fetch new data, standardize the size of images, and retrain the model.
You need to use the Azure Machine Learning SDK to configure the schedule for the pipeline.
Which four 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.



A. Mastered
B. Not Mastered

Answer: A

Explanation:
Step 1: Publish the pipeline.
To schedule a pipeline, you'll need a reference to your workspace, the identifier of your published pipeline, and the name of the experiment in which you wish to create the schedule.
Step 2: Retrieve the pipeline ID. Needed for the schedule.
Step 3: Create a ScheduleRecurrence..
To run a pipeline on a recurring basis, you'll create a schedule. A Schedule associates a pipeline, an experiment, and a trigger.
First create a schedule. Example: Create a Schedule that begins a run every 15 minutes: recurrence = ScheduleRecurrence(frequency="Minute", interval=15)
Step 4: Define an Azure Machine Learning pipeline schedule.. Example, continued:
recurring_schedule = Schedule.create(ws, name="MyRecurringSchedule", description="Based on time",
pipeline_id=pipeline_id, experiment_name=experiment_name, recurrence=recurrence)
Reference:
https://docs.microsoft.com/en-us/azure/machine-learning/how-to-schedule-pipelines

NEW QUESTION 39
- (Exam Topic 3)
You plan to use Hyperdrive to optimize the hyperparameters selected when training a model. You create the following code to define options for the hyperparameter experiment

```
import azureml.train.hyperdrive.parameter_expressions as pe
from azureml.train.hyperdrive import GridParameterSampling, HyperDriveConfig

param_sampling = GridParameterSampling({
   "max_depth" : pe.choice(6, 7, 8, 9),
   "learning_rate" : pe.choice(0.05, 0.1, 0.15)
   })
hyperdrive_run_config = HyperDriveConfig(
   estimator = estimator,
   hyperparameter_sampling = param_sampling,
   policy = None,
   primary_metric_name = "auc",
   primary_metric_goal = PrimaryMetricGoal.MAXIMIZE,

 estimator = estimator,
 hyperparameter_sampling = param_sampling,
 policy = None,
 primary_metric_name = "auc",
 primary_metric_goal = PrimaryMetricGoal.MAXIMIZE,
 max_total_runs = 50,
 max_concurrent_runs = 4)
```

For each of the following statements, select Yes if the statement is true. Otherwise, select No. NOTE: Each correct selection is worth one point.

Answer Area

	Yes	No
There will be 50 runs for this hyperparameter tuning experiment.	○	○
You can use the policy parameter in the HyperDriveConfig class to specify a security policy.	○	○
The experiment will create a run for every possible value for the learning rate parameter between 0.05 and 0.15.	○	○

A. Mastered
B. Not Mastered

Answer: A

Explanation:
Box 1: No
max_total_runs (50 here)
The maximum total number of runs to create. This is the upper bound; there may be fewer runs when the sample space is smaller than this value.
Box 2: Yes
Policy EarlyTerminationPolicy
The early termination policy to use. If None - the default, no early termination policy will be used. Box 3: No
Discrete hyperparameters are specified as a choice among discrete values. choice can be: one or more comma-separated values
> a range object

any arbitrary list object Reference:
https://docs.microsoft.com/en-us/python/api/azureml-train-core/azureml.train.hyperdrive.hyperdriveconfig https://docs.microsoft.com/en-us/azure/machine-learning/how-to-tune-hyperparameters

NEW QUESTION 43
- (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.sklearn import SKLearn
sk_est = SKLearn(source_directory='./scripts',
 compute_target=aml-compute,
 entry_script='train.py')
```

Does the solution meet the goal?

A. Yes
B. No

Answer: A

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 44
- (Exam Topic 3)
You use the Azure Machine Learning Python SDK to define a pipeline to train a model.
The data used to train the model is read from a folder in a datastore.
You need to ensure the pipeline runs automatically whenever the data in the folder changes. What should you do?

A. Set the regenerate_outputs property of the pipeline to True
B. Create a ScheduleRecurrance object with a Frequency of aut
C. Use the object to create a Schedule for the pipeline
D. Create a PipelineParameter with a default value that references the location where the training data is stored
E. Create a Schedule for the pipelin
F. Specify the datastore in the datastore property, and the folder containing the training data in the path_on_datascore property

Answer: D

Explanation:
Reference:
https://docs.microsoft.com/en-us/azure/machine-learning/how-to-trigger-published-pipeline

NEW QUESTION 49
- (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 53
- (Exam Topic 3)
You write code to retrieve an experiment that is run from your Azure Machine Learning workspace.
The run used the model interpretation support in Azure Machine Learning to generate and upload a model explanation.
Business managers in your organization want to see the importance of the features in the model.
You need to print out the model features and their relative importance in an output that looks similar to the following.

Feature	Importance
0	1.5627435610083558
2	0.6077689312583112
4	0.5574002432900718
3	0.42858759955671777
1	0.3501361539771977

How should you complete the code? To answer, select the appropriate options in the answer area.
NOTE: Each correct selection is worth one point.

```
# Assume required modules are imported

ws = Workspace.from_config()
feature_importances = explanation.  [ from_run | list_model_explanations | from_run_id | download_model_explanation ]  ( workspace = ws,
                                    experiment_name='train_and_explain',
                                    run_id='train_and_explain_12345')
explanation = client.               [ upload_model_explanation | list_model_explanations | run | download_model_explanation ]  ()

feature_importances = explanation.  [ explanation | explanation_client | get_feature_important_dict | download_model_explanation ]  ()

for key, value in feature_importances.items():
            print(key, "\t", value)
```

A. Mastered
B. Not Mastered

Answer: A

Explanation:
Box 1: from_run_id
from_run_id(workspace, experiment_name, run_id) Create the client with factory method given a run ID. Returns an instance of the explanations Client.
Parameters
- Workspace Workspace An object that represents a workspace.
- experiment_name str The name of an experiment.
- run_id str A GUID that represents a run.

Box 2: list_model_explanations
list_model_explanations returns a dictionary of metadata for all model explanations available.
Returns
A dictionary of explanation metadata such as id, data type, explanation: method, model type, and upload time, sorted by upload time
Box 3: explanation:
Reference:
https://docs.microsoft.com/en-us/python/api/azureml-contrib-interpret/azureml.contrib.interpret.

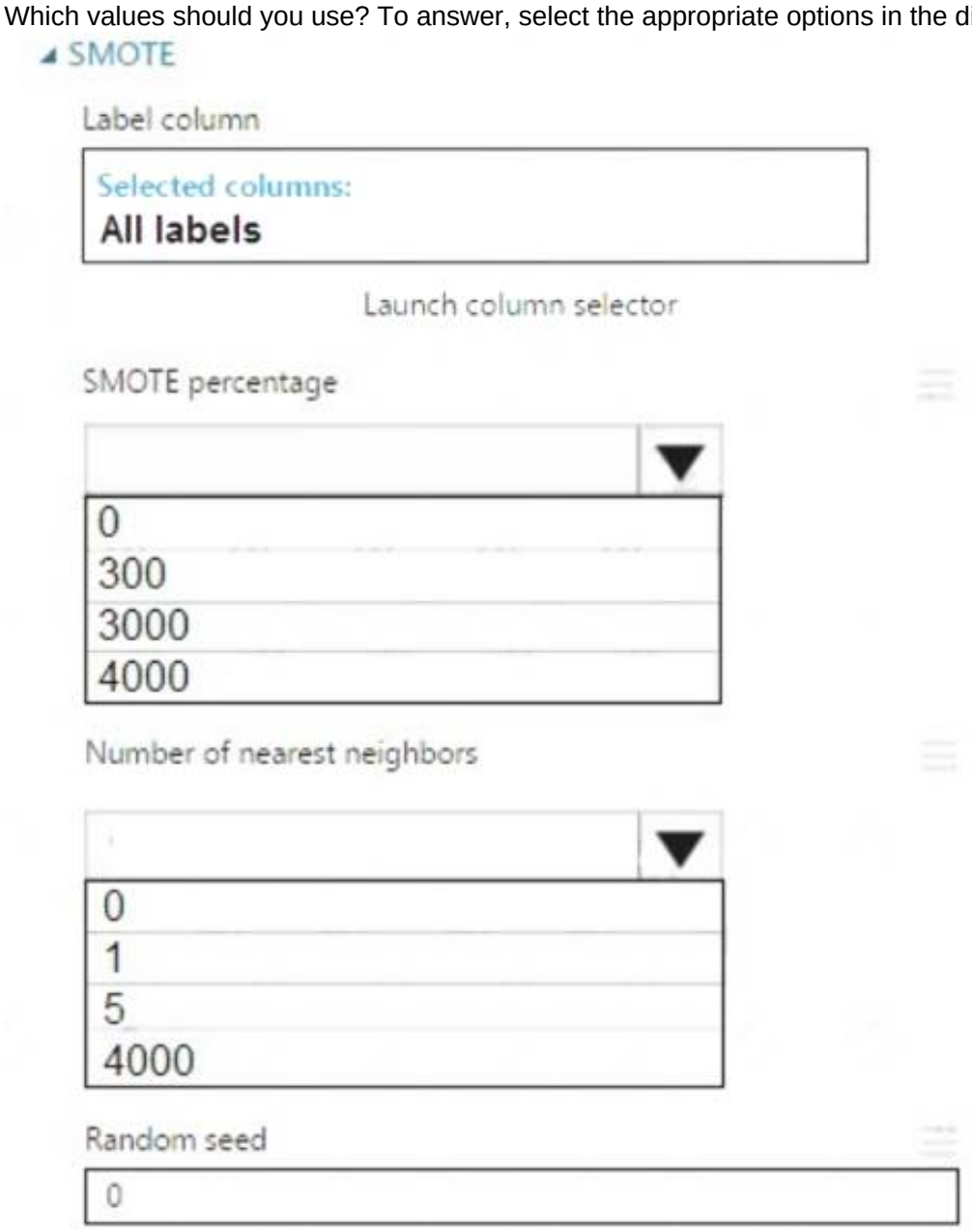
NEW QUESTION 56
- (Exam Topic 3)
You have a Python script that executes a pipeline. The script includes the following code: from azureml.core import Experiment
pipeline_run = Experiment(ws, 'pipeline_test').submit(pipeline) You want to test the pipeline before deploying the script.
You need to display 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 58
- (Exam Topic 3)
You create an experiment in Azure Machine Learning Studio. You add a training dataset that contains 10,000 rows. The first 9,000 rows represent class 0 (90 percent).
The remaining 1,000 rows represent class 1 (10 percent).
The training set is imbalances between two classes. You must increase the number of training examples for class 1 to 4,000 by using 5 data rows. You add the Synthetic Minority Oversampling Technique (SMOTE) module to the experiment.
You need to configure the module.
Which values should you use? To answer, select the appropriate options in the dialog box in the answer area. NOTE: Each correct selection is worth one point.



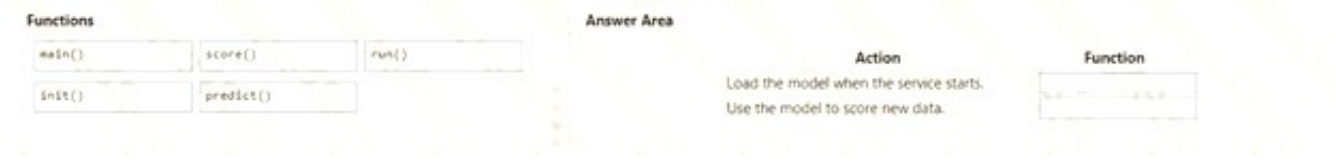
A. Mastered
B. Not Mastered

Answer: A

Explanation:
Box 1: 300
You type 300 (%), the module triples the percentage of minority cases (3000) compared to the original dataset (1000).
Box 2: 5
We should use 5 data rows.
Use the Number of nearest neighbors option to determine the size of the feature space that the SMOTE algorithm uses when in building new cases. A nearest neighbor is a row of data (a case) that is very similar to some target case. The distance between any two cases is measured by combining the weighted vectors of all features.
By increasing the number of nearest neighbors, you get features from more cases.
By keeping the number of nearest neighbors low, you use features that are more like those in the original sample.
References:
https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/smote

NEW QUESTION 61
- (Exam Topic 3)
You use Azure Machine Learning to deploy a model as a real-time web service.
You need to create an entry script for the service that ensures that the model is loaded when the service starts and is used to score new data as it is received.

Which functions should you include in the script? To answer, drag the appropriate functions to the correct actions. Each function 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.



A. Mastered
B. Not Mastered

Answer: A

Explanation:
Box 1: init()
The entry script has only two required functions, init() and run(data). These functions are used to initialize the service at startup and run the model using request data passed in by a client. The rest of the script handles loading and running the model(s).
Box 2: run() Reference:
https://docs.microsoft.com/en-us/azure/machine-learning/how-to-deploy-existing-model

NEW QUESTION 66
- (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 70
- (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])
```

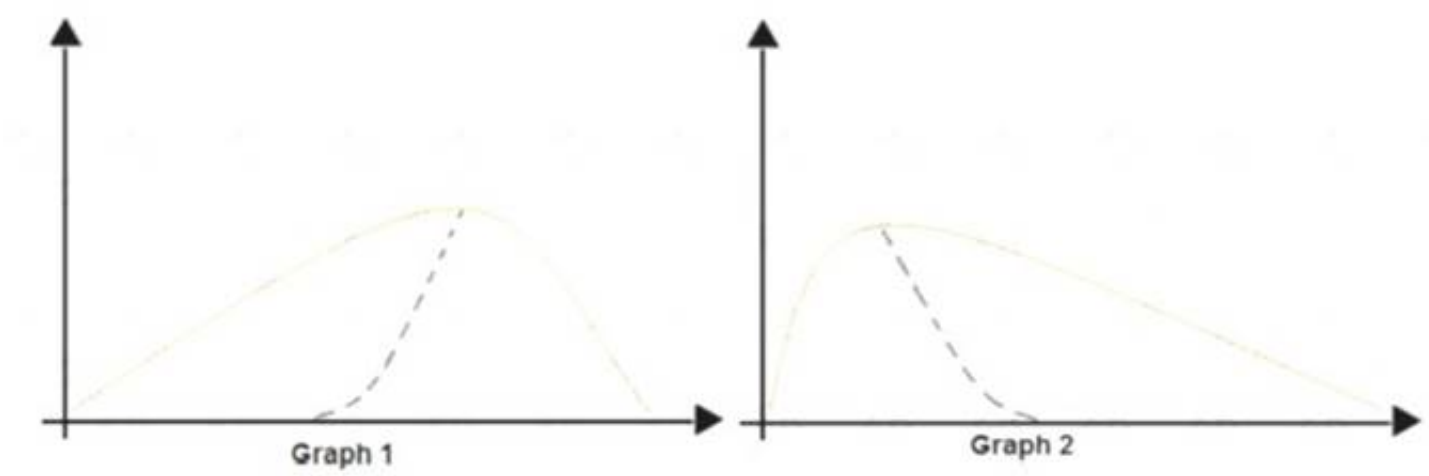
Does the solution meet the goal?

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 71
- (Exam Topic 3)
You are analyzing the asymmetry in a statistical distribution.
The following image contains two density curves that show the probability distribution of two datasets.



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.

Question	Answer choice
Which type of distribution is shown for the dataset density curve of Graph 1?	Negative skew Positive skew Normal distribution Bimodal distribution
Which type of distribution is shown for the dataset density curve of Graph 2?	Negative skew Positive skew Normal distribution Bimodal distribution

A. Mastered
B. Not Mastered

Answer: A

Explanation:
Box 1: Positive skew
Positive skew values means the distribution is skewed to the right. Box 2: Negative skew
Negative skewness values mean the distribution is skewed to the left. References:
https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/compute-elementary-statistic

NEW QUESTION 74
- (Exam Topic 3)
You arc 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 76
- (Exam Topic 3)
You develop and train a machine learning model to predict fraudulent transactions for a hotel booking website. Traffic to the site varies considerably. The site experiences heavy traffic on Monday and Friday and much lower traffic on other days. Holidays are also high web traffic days. You need to deploy the model as an Azure Machine Learning real-time web service endpoint on compute that can dynamically scale up and down to support demand. Which deployment compute option should you use?

A. attached Azure Databricks cluster
B. Azure Container Instance (ACI)
C. Azure Kubernetes Service (AKS) inference cluster
D. Azure Machine Learning Compute Instance
E. attached virtual machine in a different region

Answer: D

Explanation:
Azure Machine Learning compute cluster is a managed-compute infrastructure that allows you to easily create a single or multi-node compute. The compute is created within your workspace region as a resource that can be shared with other users in your workspace. The compute scales up automatically when a job is submitted, and can be put in an Azure Virtual Network.
Reference:
https://docs.microsoft.com/en-us/azure/machine-learning/how-to-create-attach-compute-sdk

NEW QUESTION 81
- (Exam Topic 3)
You are tuning a hyperparameter for an algorithm. The following table shows a data set with different hyperparameter, training error, and validation errors.

Hyperparameter (H)	Training error (TE)	Validation error (VE)
1	105	95
2	200	85
3	250	100
4	105	100
5	400	50

Use the drop-down menus to select the answer choice that answers each question based on the information presented in the graphic.

Question	Answer Choise
Which H value should you select based on the data?	1 2 3 4 5
What H value displays the poorest training result?	1 2 3 4 5

A. Mastered
B. Not Mastered

Answer: A

Explanation:
Box 1: 4
Choose the one which has lower training and validation error and also the closest match. Minimize variance (difference between validation error and train error).
Box 2: 5
Minimize variance (difference between validation error and train error). Reference:
https://medium.com/comet-ml/organizing-machine-learning-projects-project-management-guidelines-2d2b8565

NEW QUESTION 84
- (Exam Topic 3)
You plan to use the Hyperdrive feature of Azure Machine Learning to determine the optimal hyperparameter values when training a model.
You must use Hyperdrive to try combinations of the following hyperparameter values:
• learning_rate: any value between 0.001 and 0.1
• batch_size: 16, 32, or 64
You need to configure the search space for the Hyperdrive experiment.
Which two parameter expressions should you use? Each correct answer presents part of the solution. NOTE: Each correct selection is worth one point.

A. a choice expression for learning_rate
B. a uniform expression for learning_rate
C. a normal expression for batch_size
D. a choice expression for batch_size
E. a uniform expression for batch_size

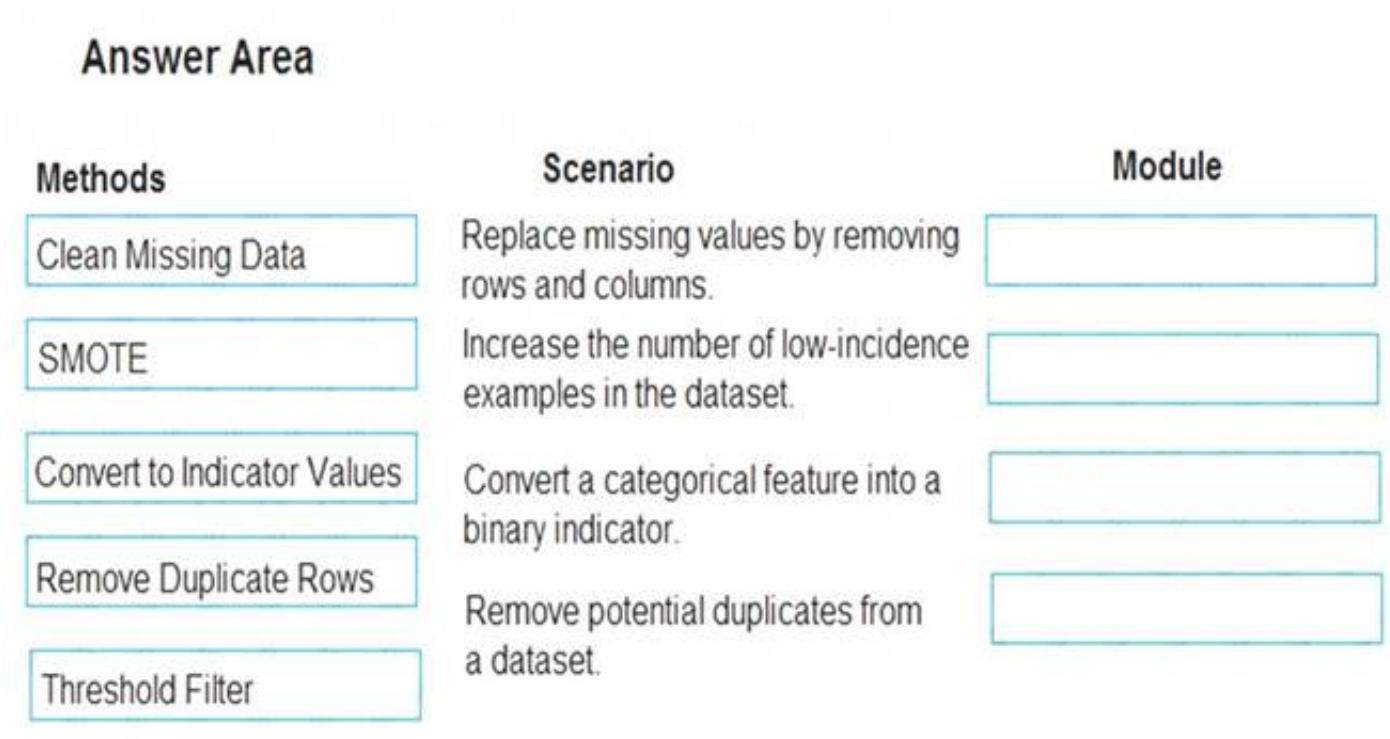
Answer: BD

Explanation:
B: Continuous hyperparameters are specified as a distribution over a continuous range of values. Supported distributions include:
- uniform(low, high) - Returns a value uniformly distributed between low and high

D: Discrete hyperparameters are specified as a choice among discrete values. choice can be:
- one or more comma-separated values
- a range object
- any arbitrary list object Reference:

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

NEW QUESTION 89
- (Exam Topic 3)
You are analyzing a raw dataset that requires cleaning.
You must perform transformations and manipulations by using Azure Machine Learning Studio. You need to identify the correct modules to perform the transformations.
Which modules should you choose? To answer, drag the appropriate modules to the correct scenarios. Each module 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.



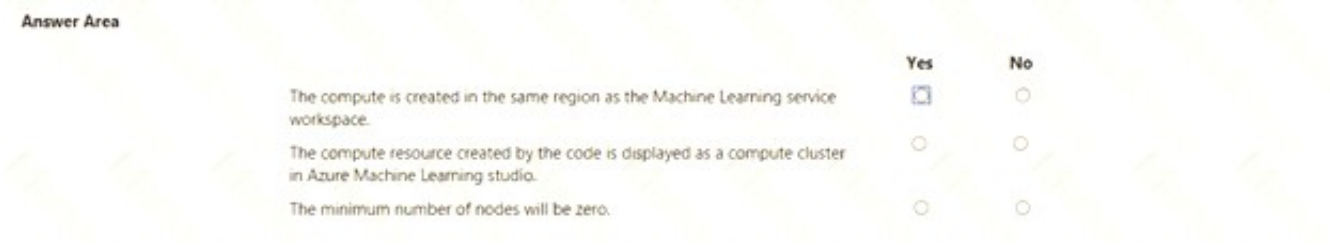
A. Mastered
B. Not Mastered

Answer: A

Explanation:
Box 1: Clean Missing Data Box 2: SMOTE
Use the SMOTE module in Azure Machine Learning Studio to increase the number of underepresented 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.
Box 3: Convert to Indicator Values
Use the Convert to Indicator Values module in Azure Machine Learning Studio. The purpose of this module is to convert columns that contain categorical values into a series of binary indicator columns that can more easily be used as features in a machine learning model.
Box 4: Remove Duplicate Rows References:
https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/smote https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/convert-to-indicator-values

NEW QUESTION 91
- (Exam Topic 3)
You are using Azure Machine Learning to train machine learning models. You need a compute target on which to remotely run the training script. You run the following Python code:

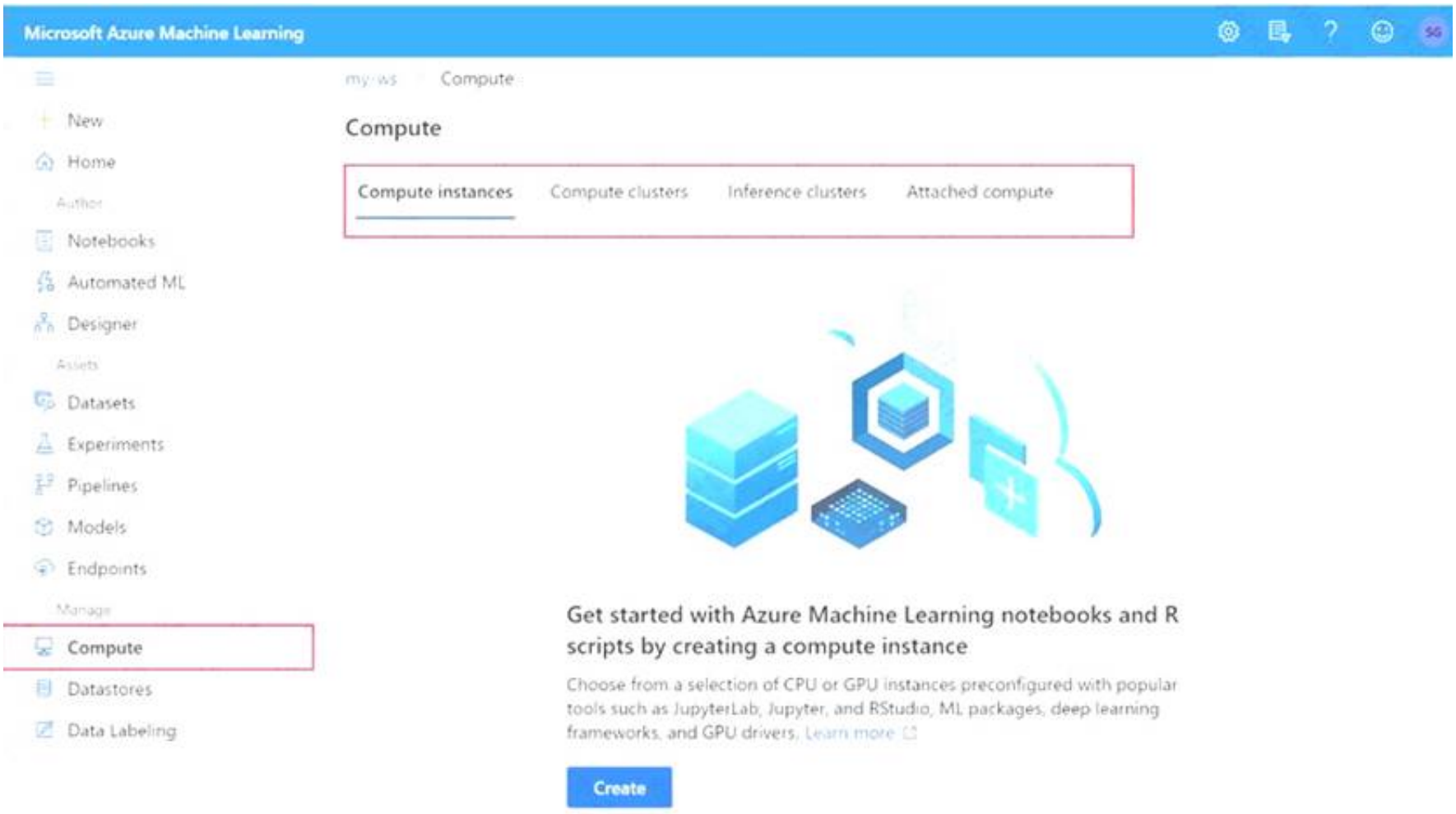
```
from azureml.core.compute import ComputeTarget, AmlCompute
from azureml.core.compute_target import ComputeTargetException
the_cluster_name = "NewCompute"
config = AmlCompute.provisioning_configuration(vm_size='STANDARD_D2', max_nodes=3)
the_cluster = ComputeTarget.create(ws, the_cluster_name, config)
```



A. Mastered
B. Not Mastered

Answer: A

Explanation:
Box 1: Yes
The compute is created within your workspace region as a resource that can be shared with other users. Box 2: Yes
It is displayed as a compute cluster. View compute targets
* 1. To see all compute targets for your workspace, use the following steps:
* 2. Navigate to Azure Machine Learning studio.
* 3. Under Manage, select Compute.
* 4. Select tabs at the top to show each type of compute target.



Box 3: Yes
min_nodes is not specified, so it defaults to 0. Reference:
https://docs.microsoft.com/en-us/python/api/azureml-core/azureml.core.compute.amlcompute.amlcomputeprovi https://docs.microsoft.com/en-us/azure/machine-learning/how-to-create-attach-compute-studio

NEW QUESTION 94
- (Exam Topic 3)
An organization uses Azure Machine Learning service and wants to expand their use of machine learning. You have the following compute environments. The organization does not want to create another compute
environment.

Environment name	Compute type
nb_server	Compute Instance
aks_cluster	Azure Kubernetes Service
mlc_cluster	Machine Learning Compute

You need to determine which compute environment to use for the following scenarios.
Which compute types should you use? To answer, drag the appropriate compute environments to the correct scenarios. Each compute environment 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.

A. Mastered
B. Not Mastered

Answer: A

Explanation:
Box 1: nb_server

Training targets	Automated ML	ML pipelines	Azure Machine Learning designer
Local computer	yes		
Azure Machine Learning compute cluster	yes & hyperparameter tuning	yes	yes
Azure Machine Learning compute instance	yes & hyperparameter tuning	yes	yes
Remote VM	yes & hyperparameter tuning	yes	
Azure Databricks	yes (SDK local mode only)	yes	
Azure Data Lake Analytics		yes	
Azure HDInsight		yes	
Azure Batch		yes	

Box 2: mlc_cluster
With Azure Machine Learning, you can train your model on a variety of resources or environments, collectively referred to as compute targets. A compute target can be a local machine or a cloud resource, such as an Azure Machine Learning Compute, Azure HDInsight or a remote virtual machine.
Reference:
https://docs.microsoft.com/en-us/azure/machine-learning/concept-compute-target https://docs.microsoft.com/en-us/azure/machine-learning/how-to-set-up-training-targets

NEW QUESTION 95
- (Exam Topic 3)
You plan to use the Hyperdrive feature of Azure Machine Learning to determine the optimal hyperparameter values when training a model.
You must use Hyperdrive to try combinations of the following hyperparameter values. You must not apply an early termination policy.
learning_rate: any value between 0.001 and 0.1
• batch_size: 16, 32, or 64
You need to configure the sampling method for the Hyperdrive experiment
Which two sampling methods can you use? Each correct answer is a complete solution. NOTE: Each correct selection is worth one point.

A. Grid sampling
B. No sampling
C. Bayesian sampling
D. Random sampling

Answer: CD

Explanation:
C: Bayesian sampling is based on the Bayesian optimization algorithm and makes intelligent choices on the hyperparameter values to sample next. It picks the sample based on how the previous samples performed, such that the new sample improves the reported primary metric.
Bayesian sampling does not support any early termination policy Example:
from azureml.train.hyperdrive import BayesianParameterSampling from azureml.train.hyperdrive import uniform, choice param_sampling =
BayesianParameterSampling({ "learning_rate": uniform(0.05, 0.1),
"batch_size": choice(16, 32, 64, 128)
}
)
D: In random sampling, hyperparameter values are randomly selected from the defined search space. Random sampling allows the search space to include both discrete and continuous hyperparameters.
Reference:
https://docs.microsoft.com/en-us/azure/machine-learning/how-to-tune-hyperparameters

NEW QUESTION 99
- (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?

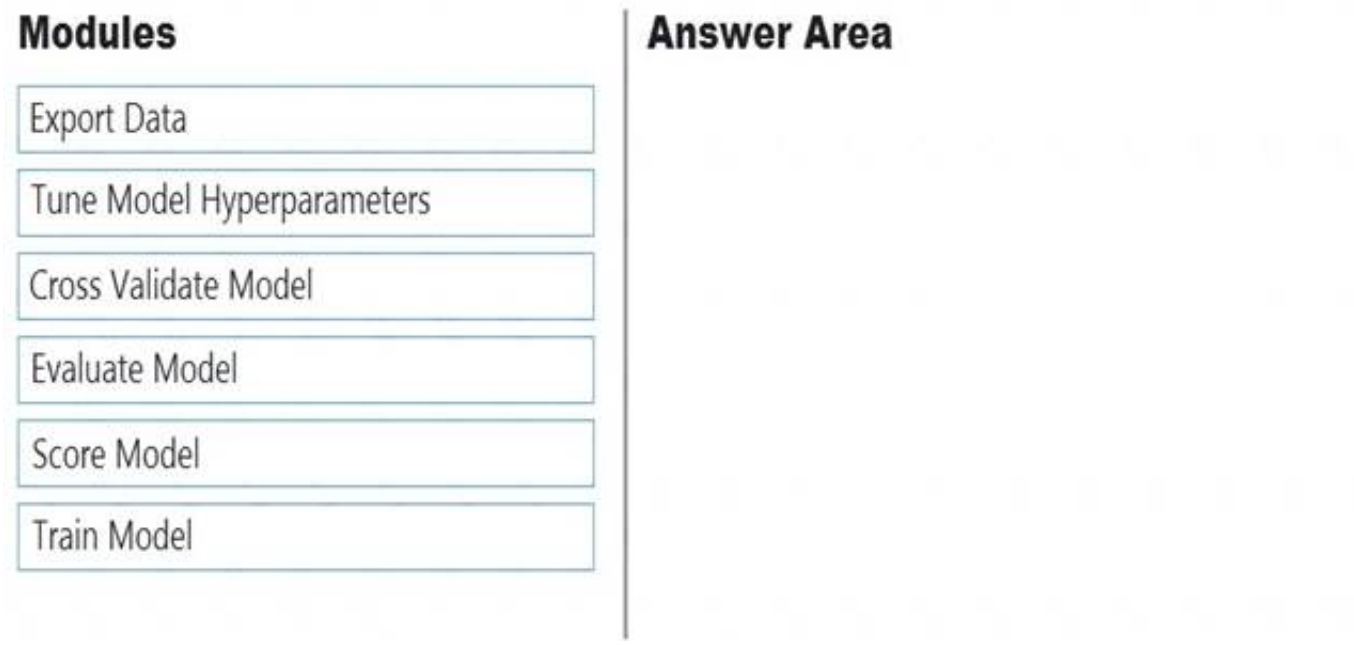
A. scatter plot
B. coefficient of determination
C. Receiver Operating Characteristic CROC) curve
D. Gradient descent

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

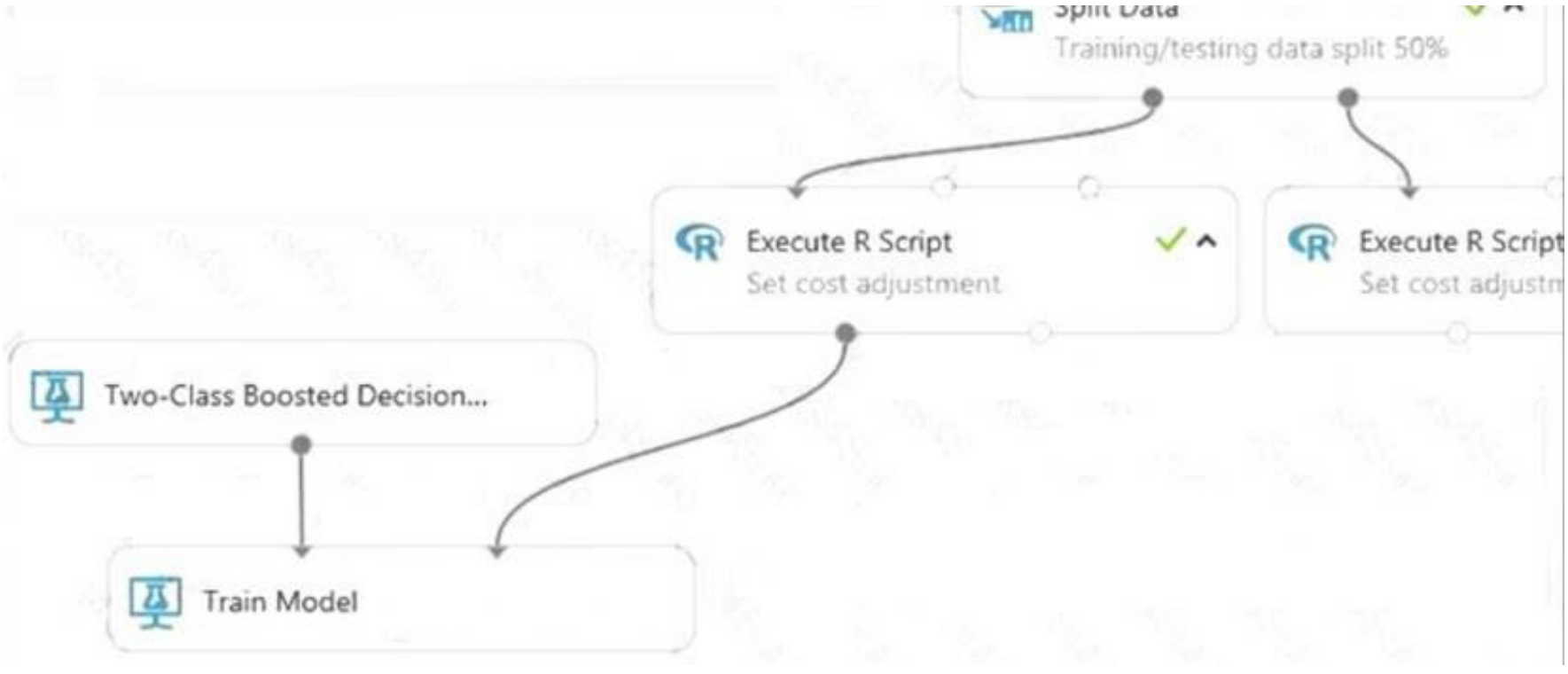
NEW QUESTION 100
- (Exam Topic 3)
You are building an experiment using the Azure Machine Learning designer.
You split a dataset into training and testing sets. You select the Two-Class Boosted Decision Tree as the algorithm.
You need to determine the Area Under the Curve (AUC) of the model.
Which three 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.



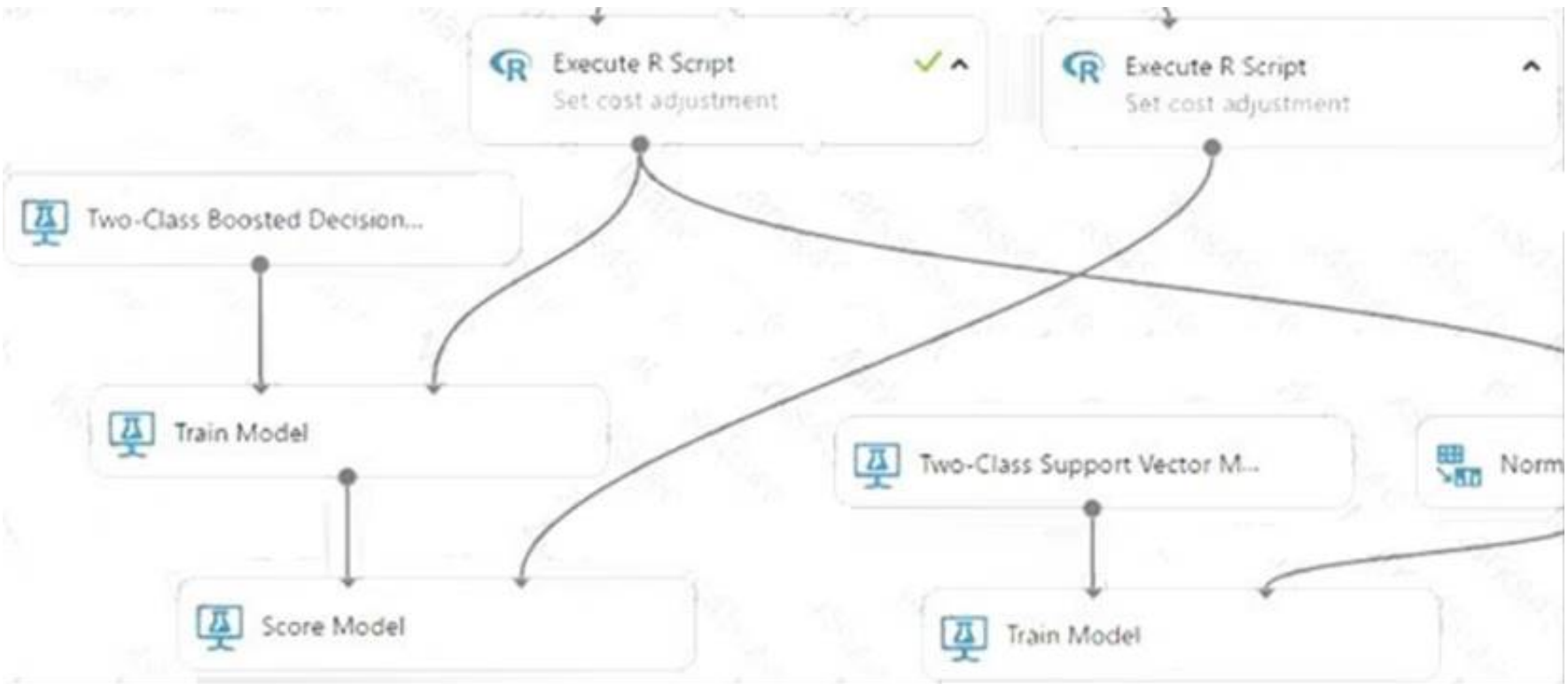
A. Mastered
B. Not Mastered

Answer: A

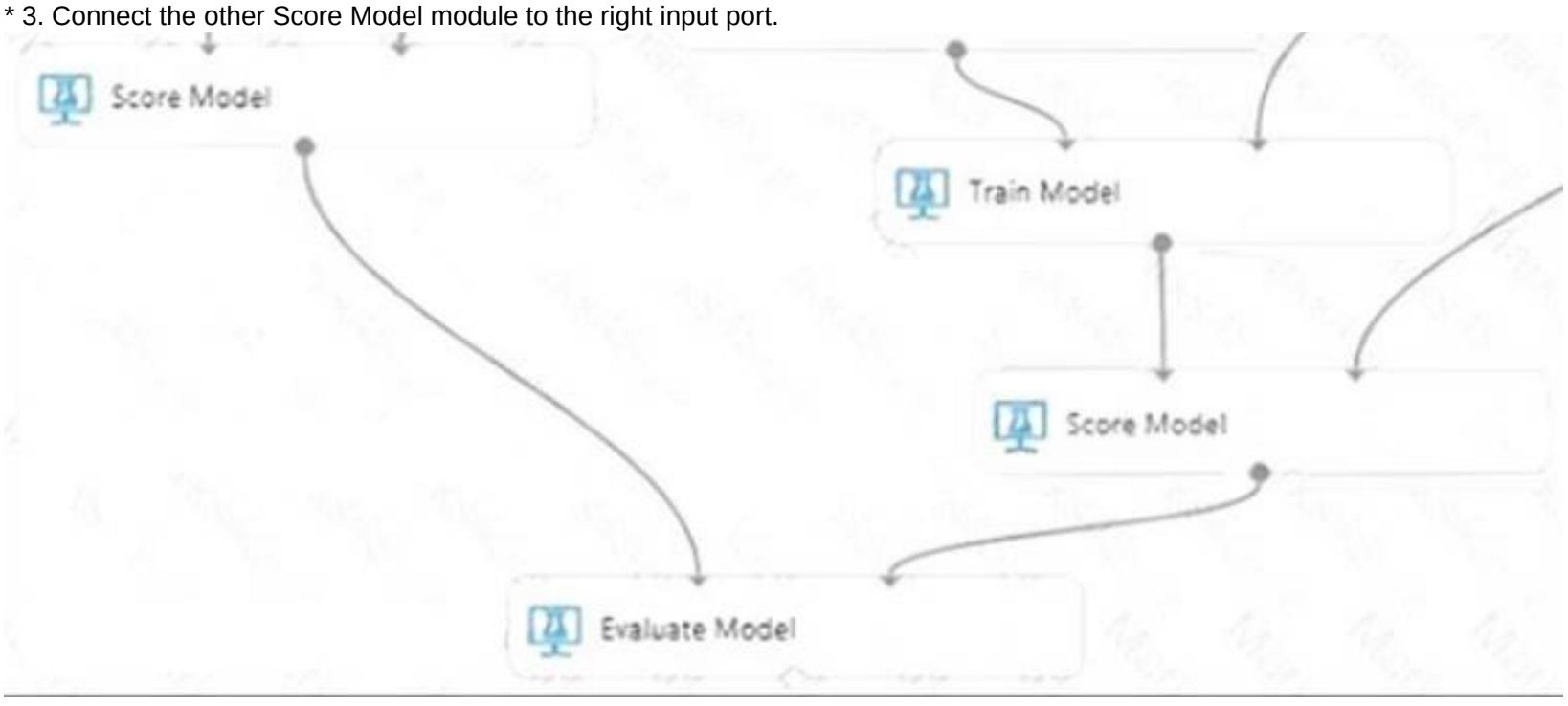
Explanation:
Step 1: Train Model
Two-Class Boosted Decision Tree
First, set up the boosted decision tree model.
* 1. Find the Two-Class Boosted Decision Tree module in the module palette and drag it onto the canvas.
* 2. Find the Train Model module, drag it onto the canvas, and then connect the output of the Two-Class Boosted Decision Tree module to the left input port of the Train Model module.
The Two-Class Boosted Decision Tree module initializes the generic model, and Train Model uses training data to train the model.
* 3. Connect the left output of the left Execute R Script module to the right input port of the Train Model
module (in this tutorial you used the data coming from the left side of the Split Data module for training). This portion of the experiment now looks something like this:



Step 2: Score Model
Score and evaluate the models
You use the testing data that was separated out by the Split Data module to score our trained models. You can then compare the results of the two models to see which generated better results.
Add the Score Model modules
* 1. Find the Score Model module and drag it onto the canvas.
* 2. Connect the Train Model module that's connected to the Two-Class Boosted Decision Tree module to the left input port of the Score Model module.
* 3. Connect the right Execute R Script module (our testing data) to the right input port of the Score Model module.



Step 3: Evaluate Model
To evaluate the two scoring results and compare them, you use an Evaluate Model module.
* 1. Find the Evaluate Model module and drag it onto the canvas.
* 2. Connect the output port of the Score Model module associated with the boosted decision tree model to the left input port of the Evaluate Model module.
* 3. Connect the other Score Model module to the right input port.



NEW QUESTION 104
- (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 AksWebservice instance.
Set the value of the auth_enabled property to False.
Set the value of the token_auth_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 106
- (Exam Topic 3)

```
train_cluster = ComputeTarget(workspace=work_space, name='train-cluster')
estimator = Estimator(source_directory =
 'training-experiment',
 script_params = {'--data-folder': data_source.as_mount(), '--regularization': 0.8},
 compute_target = train_cluster,
 entry_script = 'train.py',
 conda_packages = ['scikit-learn'])
```

For each of the following statements, select Yes if the statement is true. Otherwise, select No. NOTE: Each correct selection is worth one point.

Answer Area

	Yes	No
The estimator will look for the files it needs to run an experiment in the training-experiment directory of the local compute environment.	☐	○
The estimator will mount the local data-folder folder and make it available to the script through a parameter.	○	○
The train.py script file will be created if it does not exist.	○	○

A. Mastered
B. Not Mastered

Answer: A

Explanation:

Answer Area

	Yes	No
The estimator will look for the files it needs to run an experiment in the training-experiment directory of the local compute environment.	☐	○
The estimator will mount the local data-folder folder and make it available to the script through a parameter.	☐	○
The train.py script file will be created if it does not exist.	○	☐

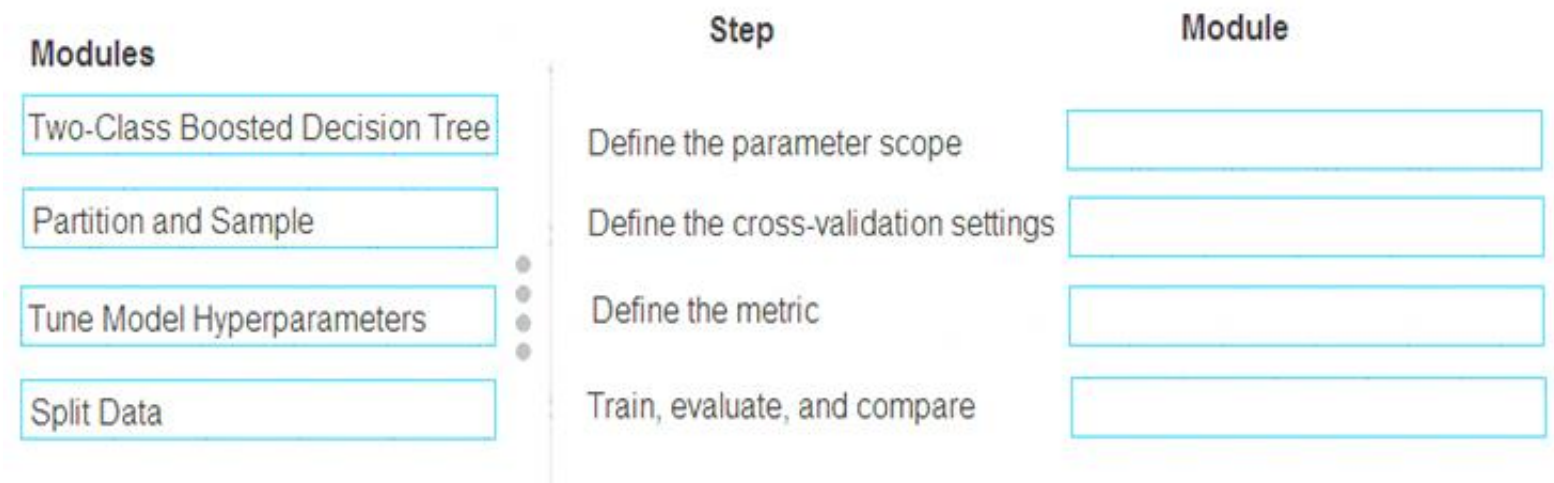
NEW QUESTION 108
- (Exam Topic 3)
You are a data scientist building a deep convolutional neural network (CNN) for image classification. The CNN model you built shows signs of overfitting.
You need to reduce overfitting and converge the model to an optimal fit.
Which two actions should you perform? Each correct answer presents a complete solution.
NOTE: Each correct selection is worth one point.

A. Reduce the amount of training data.
B. Add an additional dense layer with 64 input units
C. Add L1/L2 regularization.
D. Use training data augmentation
E. Add an additional dense layer with 512 input units.

Answer: AC

Explanation:
References:
https://machinelearningmastery.com/how-to-reduce-overfitting-in-deep-learning-with-weight-regularization/
https://en.wikipedia.org/wiki/Convolutional_neural_network

NEW QUESTION 113
- (Exam Topic 3)
You have a model with a large difference between the training and validation error values. You must create a new model and perform cross-validation.
You need to identify a parameter set for the new model using Azure Machine Learning Studio.
Which module you should use for each step? To answer, drag the appropriate modules to the correct steps. Each module may be used once or 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.

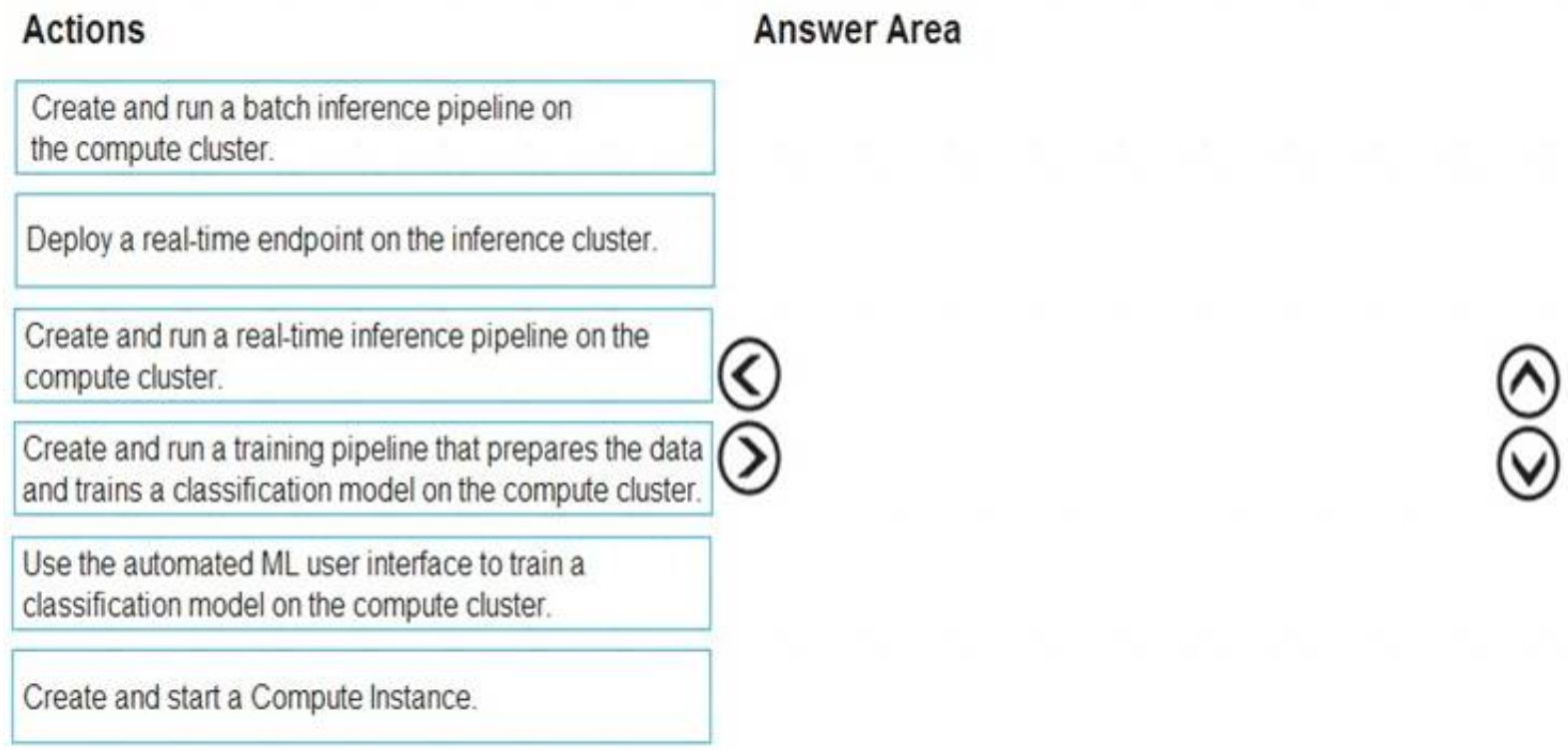


A. Mastered
B. Not Mastered

Answer: A

Explanation:
Box 1: Split data
Box 2: Partition and Sample
Box 3: Two-Class Boosted Decision Tree Box 4: Tune Model Hyperparameters
Integrated train and tune: You configure a set of parameters to use, and then let the module iterate over multiple combinations, measuring accuracy until it finds a "best" model. With most learner modules, you can choose which parameters should be changed during the training process, and which should remain fixed.
We recommend that you use Cross-Validate Model to establish the goodness of the model given the specified
parameters. Use Tune Model Hyperparameters to identify the optimal parameters. References:
https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/partition-and-sample

NEW QUESTION 117
- (Exam Topic 3)
You have an Azure Machine Learning workspace that contains a CPU-based compute cluster and an Azure Kubernetes Services (AKS) inference cluster. You create a tabular dataset containing data that you plan to use to create a classification model.
You need to use the Azure Machine Learning designer to create a web service through which client applications can consume the classification model by submitting new data and getting an immediate prediction as a response.
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.



A. Mastered
B. Not Mastered

Answer: A

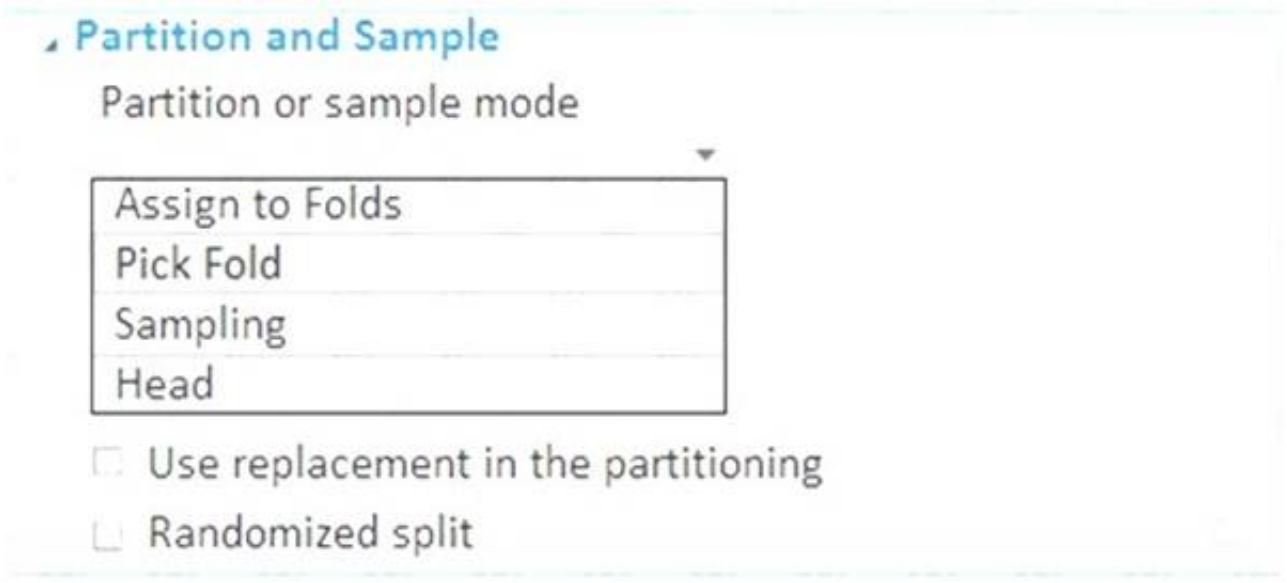
Explanation:
Step 1: Create and start a Compute Instance
To train and deploy models using Azure Machine Learning designer, you need compute on which to run the training process, test the model, and host the model in a deployed service.
There are four kinds of compute resource you can create:
Compute Instances: Development workstations that data scientists can use to work with data and models. Compute Clusters: Scalable clusters of virtual machines for on-demand processing of experiment code. Inference Clusters: Deployment targets for predictive services that use your trained models.
Attached Compute: Links to existing Azure compute resources, such as Virtual Machines or Azure Databricks clusters.
Step 2: Create and run a training pipeline..
After you've used data transformations to prepare the data, you can use it to train a machine learning model. Create and run a training pipeline
Step 3: Create and run a real-time inference pipeline
After creating and running a pipeline to train the model, you need a second pipeline that performs the same data transformations for new data, and then uses the trained model to inference (in other words, predict) label values based on its features. This pipeline will form the basis for a predictive service that you can publish for applications to use.

Reference:
https://docs.microsoft.com/en-us/learn/modules/create-classification-model-azure-machine-learning-designer/

NEW QUESTION 120
- (Exam Topic 3)
You have a dataset that contains 2,000 rows. You are building a machine learning classification model by using Azure Learning Studio. You add a Partition and Sample module to the experiment.
You need to configure the module. You must meet the following requirements:

- Divide the data into subsets
- Assign the rows into folds using a round-robin method
- Allow rows in the dataset to be reused

How should you configure the module? To answer, select the appropriate options in the dialog box in the answer area.
NOTE: Each correct selection is worth one point.



A. Mastered
B. Not Mastered

Answer: A

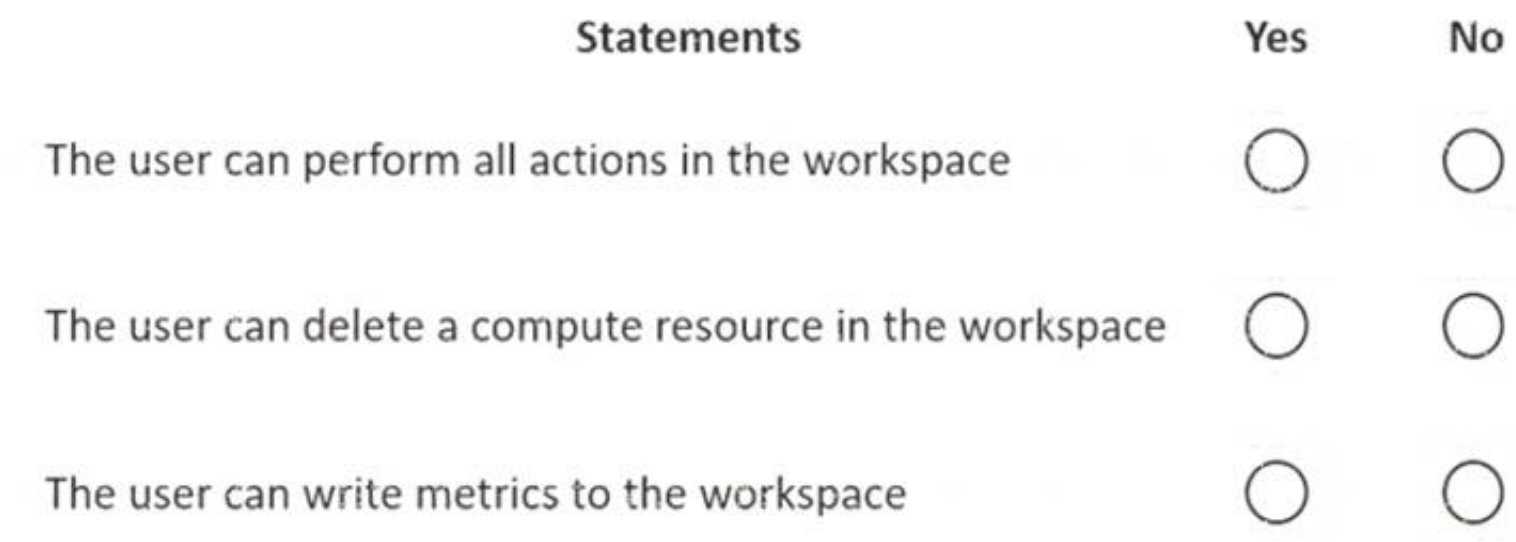
Explanation:
Use the Split data into partitions option when you want to divide the dataset into subsets of the data. This option is also useful when you want to create a custom number of folds for cross-validation, or to split rows into several groups.

- For Partition or sample mode, select Assign to Folds.
- Use replacement in the partitioning: Select this option if you want the sampled row to be put back into the pool of rows for potential reuse. As a result, the same row might be assigned to several folds.
- If you do not use replacement (the default option), the sampled row is not put back into the pool of rows for potential reuse. As a result, each row can be assigned to only one fold.
- Randomized split: Select this option if you want rows to be randomly assigned to folds. If you do not select this option, rows are assigned to folds using the round-robin method. References:
https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/partition-and-sample

NEW QUESTION 124
- (Exam Topic 3)
You create an Azure Machine Learning workspace named workspace1. You assign a custom role to a user of workspace1.
The custom role has the following JSON definition:

```
{
 "Name": "MyRole",
 "IsCustom": true,
 "Description": "New custom role description.",
 "Actions": ["*"],
 "NotActions": [
   "Microsoft.MachineLearningServices/workspaces/write",
   "Microsoft.MachineLearningServices/workspaces/computes/*/write",
   "Microsoft.MachineLearningServices/workspaces/computes/*/delete",
   "Microsoft. Authorization/*/write"
 ],
 "AssignableScopes": [
   "/subscriptions/<subscription_id>/resourceGroups/resourcegroup1/providers/
   Microsoft.MachineLearningServices/workspaces/workspace1"
 ]
}
```

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.

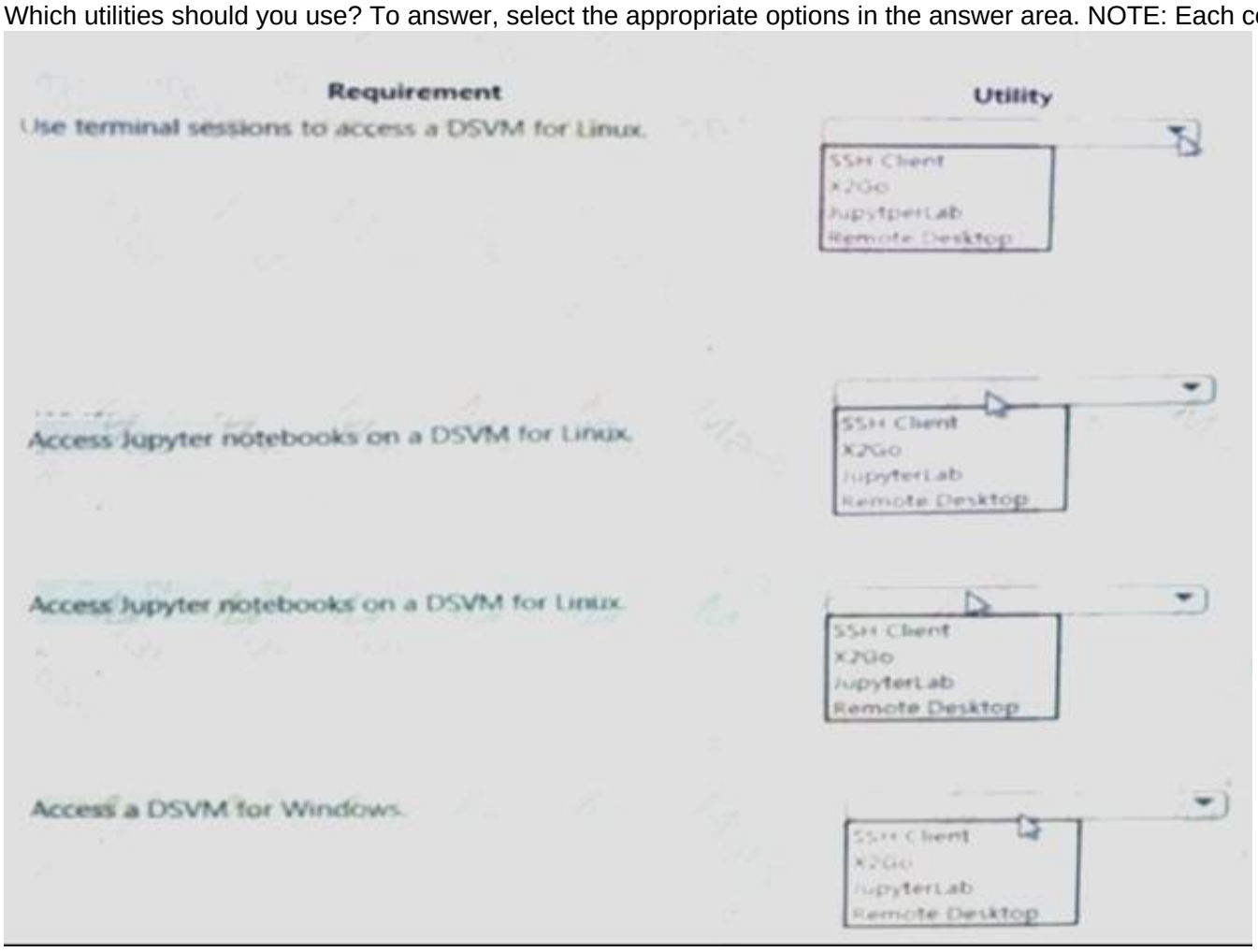


A.

Answer:

Explanation:
Graphical user interface, text, application, email Description automatically generated
Box 1: No
The actions listed in NotActions are prohibited.
If the roles include Actions that have a wildcard (*), the effective permissions are computed by subtracting the NotActions from the allowed Actions.
Box 2: No
Deleting compute resources in the workspace is in the NotActions list. Box 3: Yes
Writing metrics is not listed in NotActions. Reference:
https://docs.microsoft.com/en-us/azure/role-based-access-control/overview#how-azure-rbac-determines-if-a-use

NEW QUESTION 129
- (Exam Topic 3)
You use Data Science Virtual Machines (DSVMs) for Windows and Linux in Azure. You need to access the DSVMs.
Which utilities 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 131
- (Exam Topic 3)
You have a Jupyter Notebook that contains Python code that is used to train a model.
You must create a Python script for the production deployment. The solution must minimize code maintenance.
Which two actions should you perform? Each correct answer presents part of the solution.
NOTE: Each correct selection is worth one point.

A. Refactor the Jupyter Notebook code into functions
B. Save each function to a separate Python file
C. Define a main() function in the Python script
D. Remove all comments and functions from the Python script

Answer: AC

Explanation:
Reference:
https://www.guru99.com/learn-python-main-function-with-examples-understand-main.html https://towardsdatascience.com/from-jupyter-notebook-to-deployment-a-straightforward-example-1838c203a43

NEW QUESTION 134
- (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 137
- (Exam Topic 3)
You have a comma-separated values (CSV) file containing data from which you want to train a classification model.
You are using the Automated Machine Learning interface in Azure Machine Learning studio to train the classification model. You set the task type to Classification.
You need to ensure that the Automated Machine Learning process evaluates only linear models. What should you do?

A. Add all algorithms other than linear ones to the blocked algorithms list.
B. Set the Exit criterion option to a metric score threshold.
C. Clear the option to perform automatic featurization.
D. Clear the option to enable deep learning.
E. Set the task type to Regression.

Answer: C

Explanation:
Automatic featurization can fit non-linear models. Reference: https://econml.azurewebsites.net/spec/estimation/dml.html
https://docs.microsoft.com/en-us/azure/machine-learning/how-to-use-automated-ml-for-ml-models

NEW QUESTION 141
- (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 a classification model by using a logistic regression algorithm.
You must be able to explain the model's predictions by calculating the importance of each feature, both as an overall global relative importance value and as a measure of local importance for a specific set of predictions.
You need to create an explainer that you can use to retrieve the required global and local feature importance values.
Solution: Create a PFIExplainer. Does the solution meet the goal?

A. Yes
B. No

Answer: A

Explanation:
Permutation Feature Importance Explainer (PFI): Permutation Feature Importance is a technique used to explain classification and regression models. At a high level, the way it works is by randomly shuffling data one feature at a time for the entire dataset and calculating how much the performance metric of interest changes. The larger the change, the more important that feature is. PFI can explain the overall behavior of any underlying model but does not explain individual predictions.
Reference:
https://docs.microsoft.com/en-us/azure/machine-learning/how-to-machine-learning-interpretability

NEW QUESTION 144
- (Exam Topic 3)
You are creating a new Azure Machine Learning pipeline using the designer.
The pipeline must train a model using data in a comma-separated values (CSV) file that is published on a website. You have not created a dataset for this file.
You need to ingest the data from the CSV file into the designer pipeline using the minimal administrative effort.
Which module should you add to the pipeline in Designer?

A. Convert to CSV
B. Enter Data Manually D
C. Import Data
D. Dataset

Answer: D

Explanation:
The preferred way to provide data to a pipeline is a Dataset object. The Dataset object points to data that lives in or is accessible from a datastore or at a Web URL. The Dataset class is abstract, so you will create an instance of either a FileDataset (referring to one or more files) or a TabularDataset that's created by from one or more files with delimited columns of data.
Example:
from azureml.core import Dataset
iris_tabular_dataset = Dataset.Tabular.from_delimited_files([(def_blob_store, 'train-dataset/iris.csv')]) Reference:
https://docs.microsoft.com/en-us/azure/machine-learning/how-to-create-your-first-pipeline

NEW QUESTION 148
- (Exam Topic 3)
You create an Azure Machine Learning compute resource to train models. The compute resource is configured as follows:
- Minimum nodes: 2
- Maximum nodes: 4

You must decrease the minimum number of nodes and increase the maximum number of nodes to the following values:
- Minimum nodes: 0
- Maximum nodes: 8

You need to reconfigure the compute resource.
What are three possible ways to achieve this goal? Each correct answer presents a complete solution.
NOTE: Each correct selection is worth one point.

A. Use the Azure Machine Learning studio.
B. Run the update method of the AmlCompute class in the Python SDK.
C. Use the Azure portal.
D. Use the Azure Machine Learning designer.
E. Run the refresh_state() method of the BatchCompute class in the Python SDK

Answer: ABC

Explanation:
Reference:
https://docs.microsoft.com/en-us/python/api/azureml-core/azureml.core.compute.amlcompute(class)

NEW QUESTION 149
- (Exam Topic 3)

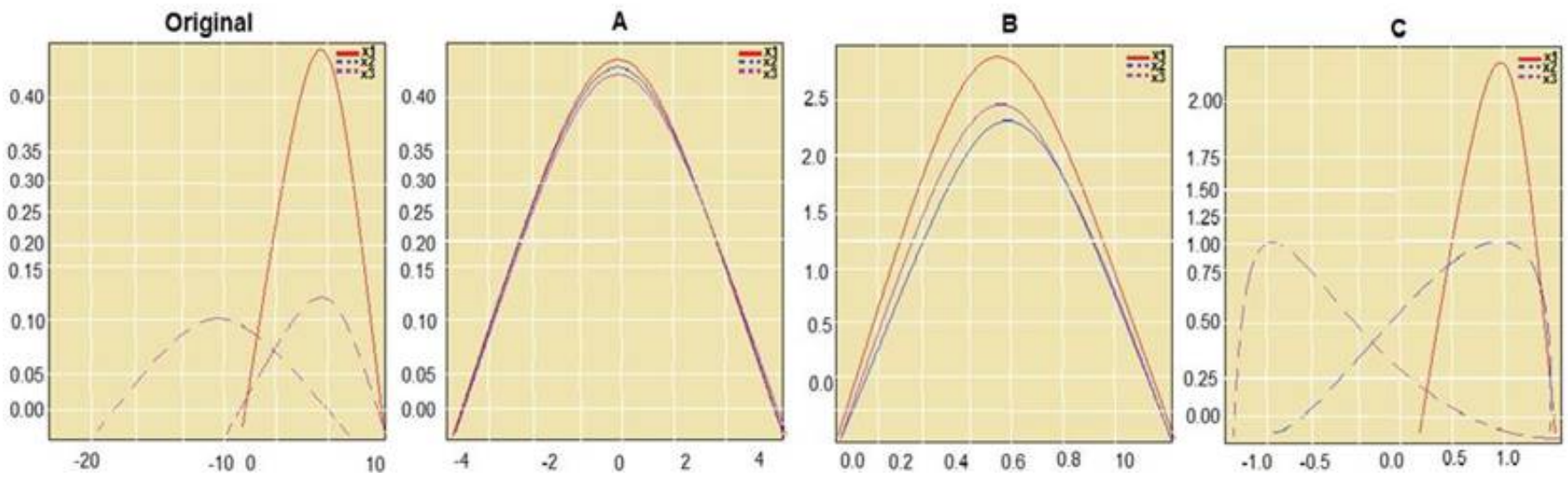
You create a deep learning model for image recognition on Azure Machine Learning service using GPU-based training.
You must deploy the model to a context that allows for real-time GPU-based inferencing. You need to configure compute resources for model inferencing.
Which compute type should you use?

A. Azure Container Instance
B. Azure Kubernetes Service
C. Field Programmable Gate Array
D. Machine Learning Compute

Answer: B

Explanation:
You can use Azure Machine Learning to deploy a GPU-enabled model as a web service. Deploying a model on Azure Kubernetes Service (AKS) is one option. The AKS cluster provides a GPU resource that is used by the model for inference.
Inference, or model scoring, is the phase where the deployed model is used to make predictions. Using GPUs instead of CPUs offers performance advantages on highly parallelizable computation.
Reference:
https://docs.microsoft.com/en-us/azure/machine-learning/how-to-deploy-inferencing-gpus

NEW QUESTION 152
- (Exam Topic 3)
You are performing feature scaling by using the scikit-learn Python library for x.1 x2, and x3 features. Original and scaled data is shown in the following image.



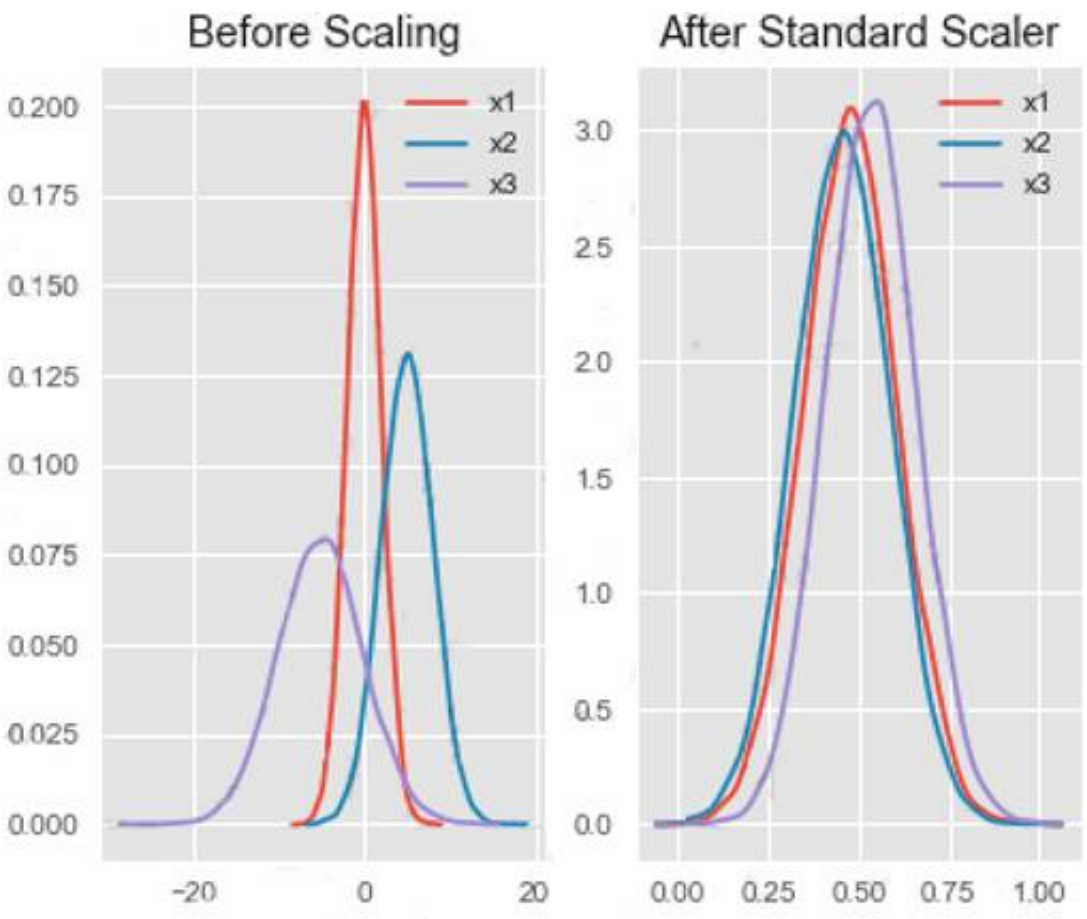
Use the drop-down menus to select the answer choice that answers each question based on the information presented in the graphic.
NOTE: Each correct selection is worth one point.

Question	Answer choice
Which scaler is used in graph A?	Standard Scaler / Min Max Scale / Normalizer
Which scaler is used in graph B?	Standard Scaler / Min Max Scale / Normalizer
Which scaler is used in graph C?	Standard Scaler / Min Max Scale / Normalizer

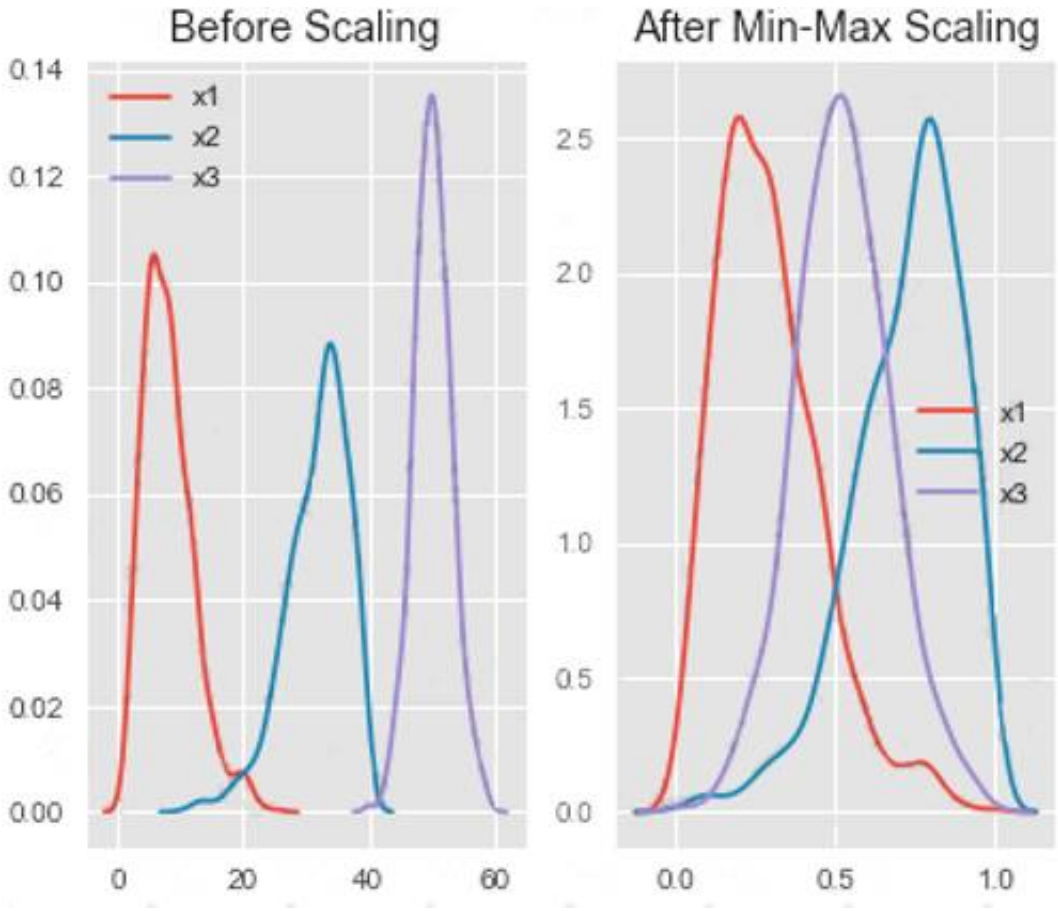
A. Mastered
B. Not Mastered

Answer: A

Explanation:
Box 1: StandardScaler
The StandardScaler assumes your data is normally distributed within each feature and will scale them such that the distribution is now centred around 0, with a standard deviation of 1.
Example:



All features are now on the same scale relative to one another. Box 2: Min Max Scaler



Notice that the skewness of the distribution is maintained but the 3 distributions are brought into the same scale so that they overlap.
Box 3: Normalizer
References:
http://benalexkeen.com/feature-scaling-with-scikit-learn/

NEW QUESTION 156
- (Exam Topic 3)
You use Azure Machine Learning to train a model based on a dataset named dataset1. You define a dataset monitor and create a dataset named dataset2 that contains new data.
You need to compare dataset1 and dataset2 by using the Azure Machine Learning SDK for Python. Which method of the DataDriftDetector class should you use?

A. run
B. get
C. backfill
D. update

Answer: C

Explanation:
A backfill run is used to see how data changes over time. Reference:
https://docs.microsoft.com/en-us/python/api/azureml-datadrift/azureml.datadrift.datadriftdetector.datadriftdetect

NEW QUESTION 161
- (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.

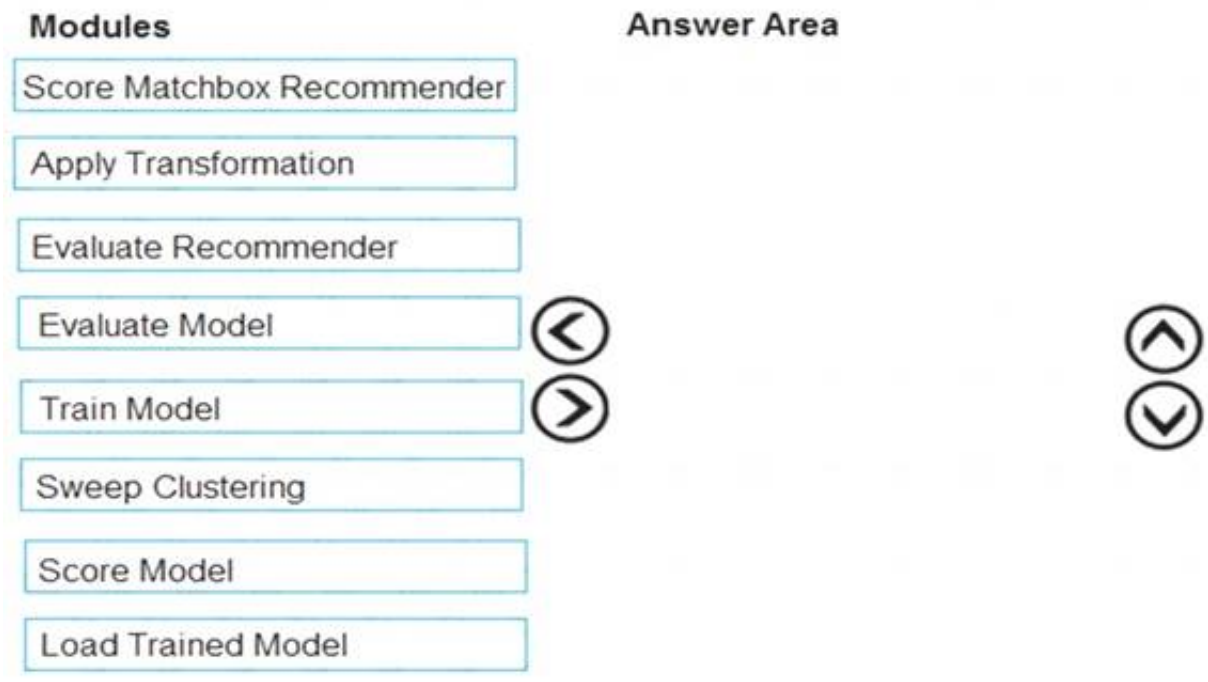
You need to load the model fairness metrics into Azure Machine Learning studio. What should you do?

A. Implement the download_dashboard_by_upload_id function
B. Implement the creace_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 162
- (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.

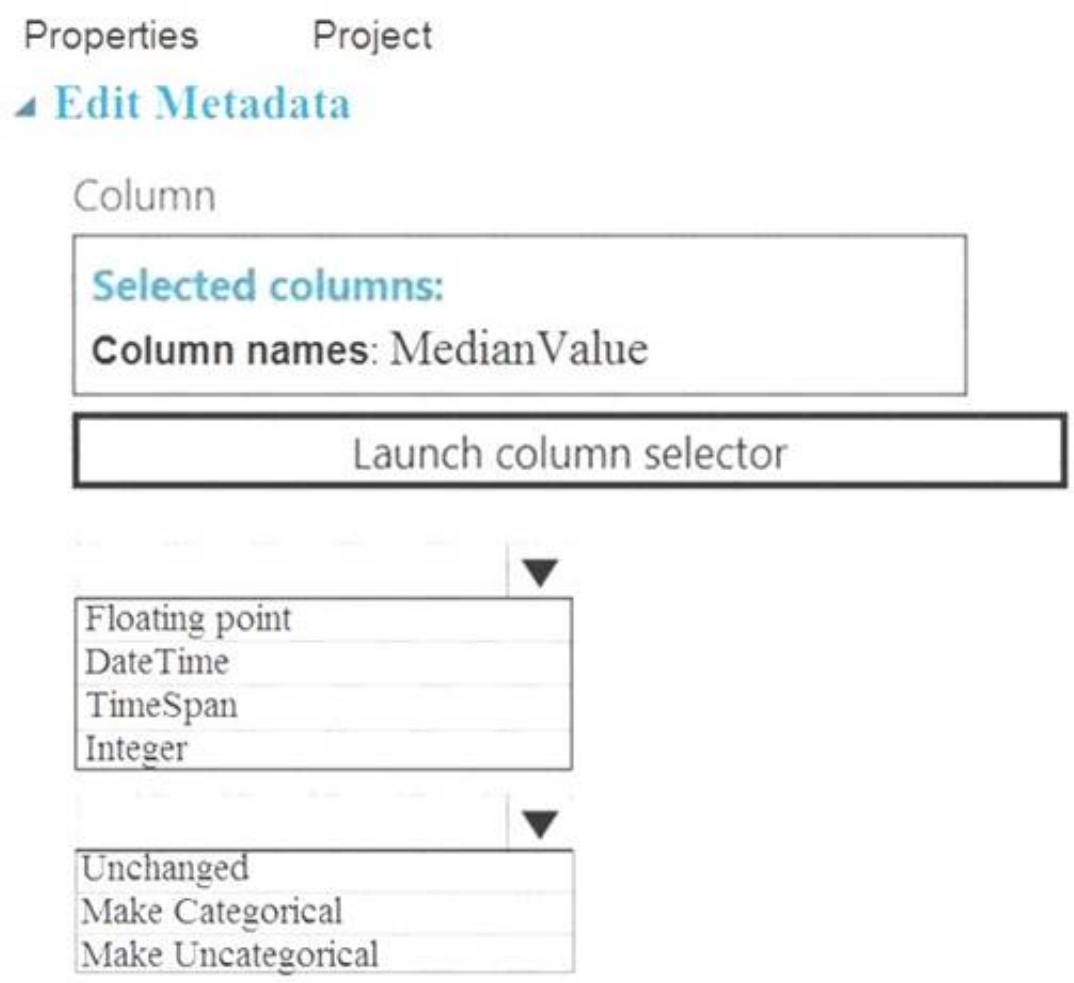


A. Mastered
B. Not Mastered

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 164
- (Exam Topic 2)
You need to configure the Edit Metadata module so that the structure of the datasets match.
Which configuration options should you select? 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:
Box 1: Floating point
Need floating point for Median values.
Scenario: An initial investigation shows that the datasets are identical in structure apart from the MedianValue column. The smaller Paris dataset contains the MedianValue in text format, whereas the larger London dataset contains the MedianValue in numerical format.
Box 2: Unchanged
Note: Select the Categorical option to specify that the values in the selected columns should be treated as categories.
For example, you might have a column that contains the numbers 0,1 and 2, but know that the numbers actually mean "Smoker", "Non smoker" and "Unknown". In that case, by flagging the column as categorical you can ensure that the values are not used in numeric calculations, only to group data.

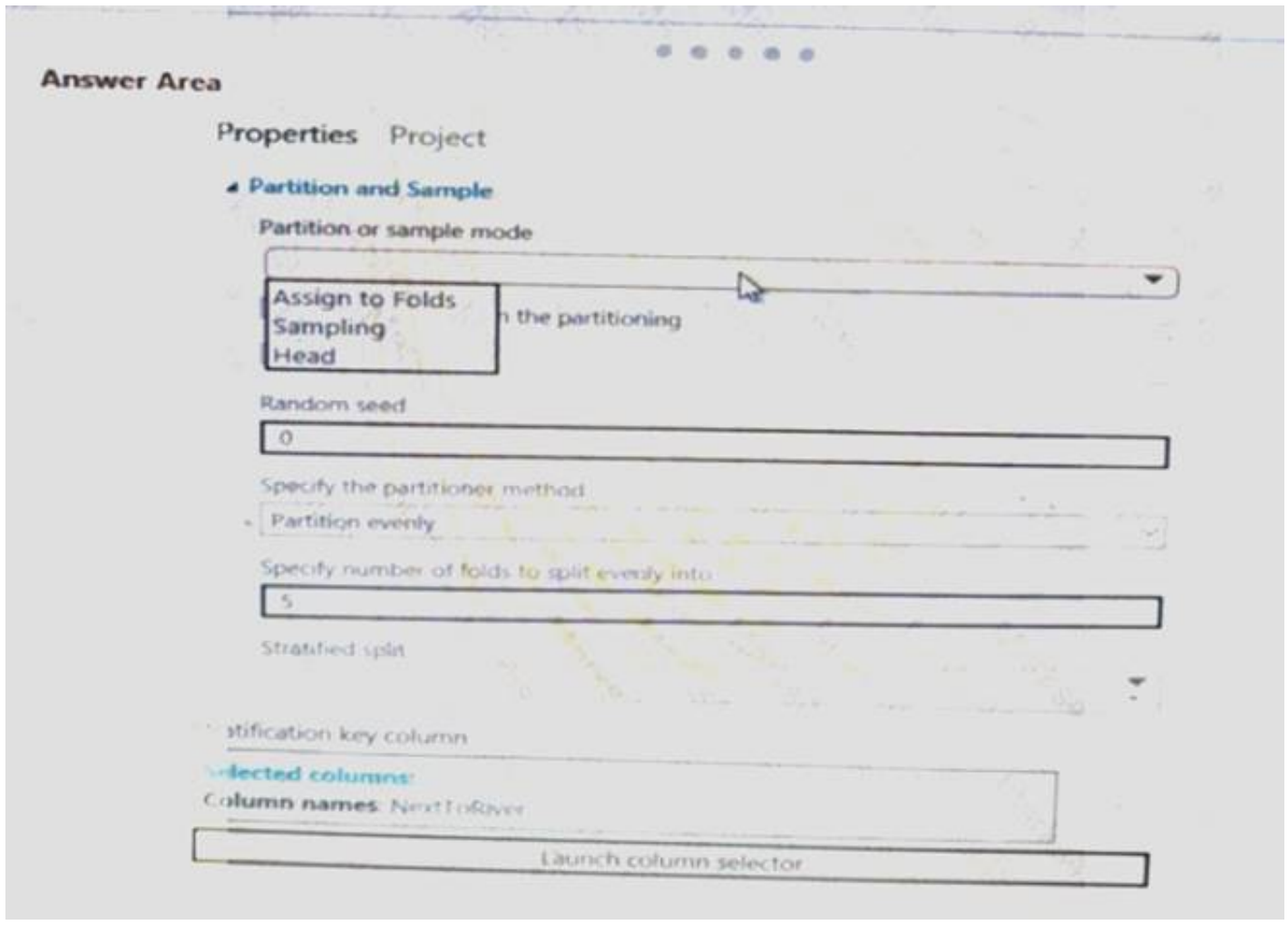
NEW QUESTION 167
- (Exam Topic 2)
You need to select a feature extraction method. Which method should you use?

A. Spearman correlation
B. Mutual information
C. Mann-Whitney test
D. Pearson’s correlation

Answer: A

Explanation:
Spearman's rank correlation coefficient assesses how well the relationship between two variables can be described using a monotonic function.
Note: Both Spearman's and Kendall's can be formulated as special cases of a more general correlation coefficient, and they are both appropriate in this scenario.
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/feature-selection-modules

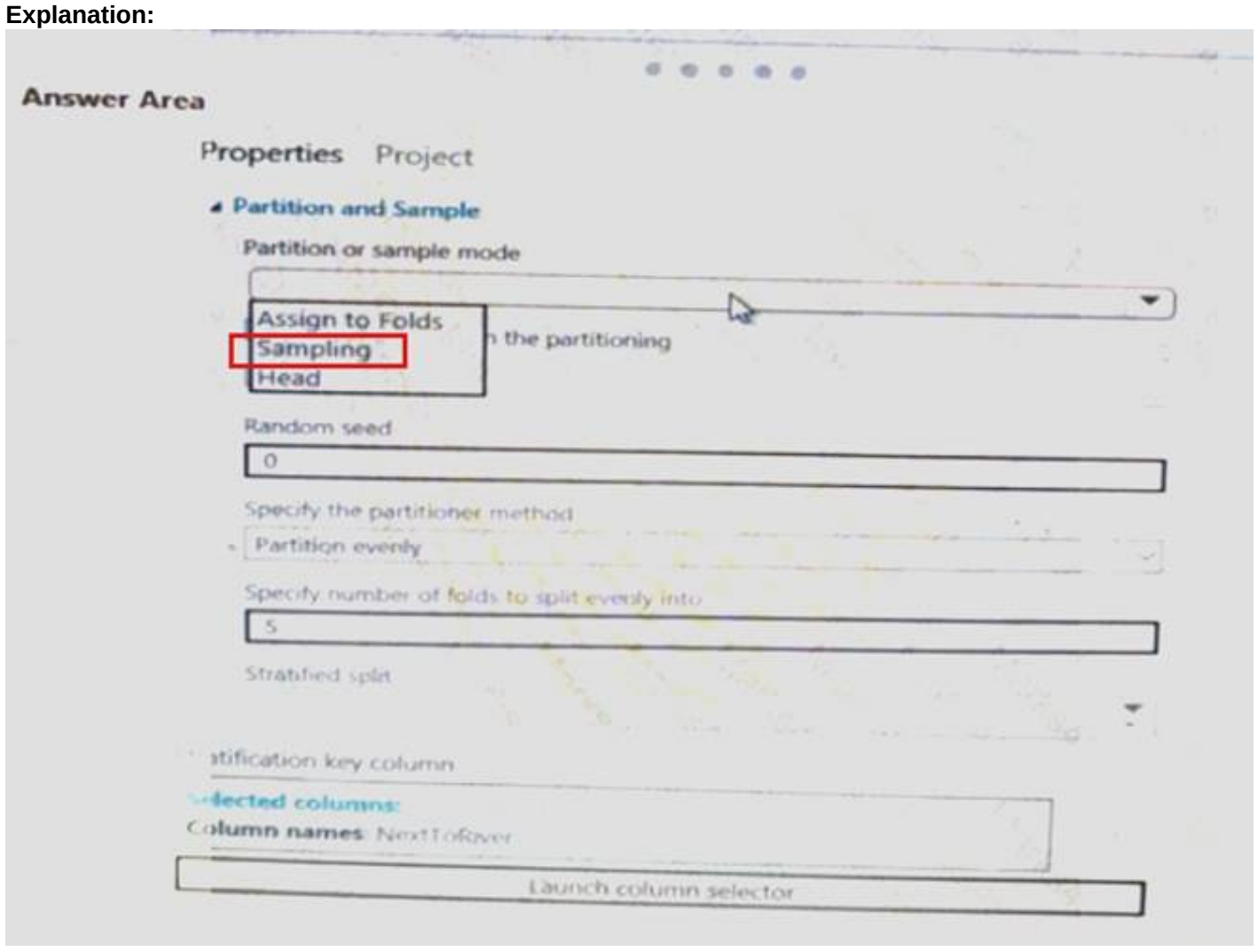
NEW QUESTION 172
- (Exam Topic 2)
You need to identify the methods for dividing the data according, to the testing requirements.
Which properties should you select? To answer, select the appropriate option-, m the answer area. NOTE: Each correct selection is worth one point.



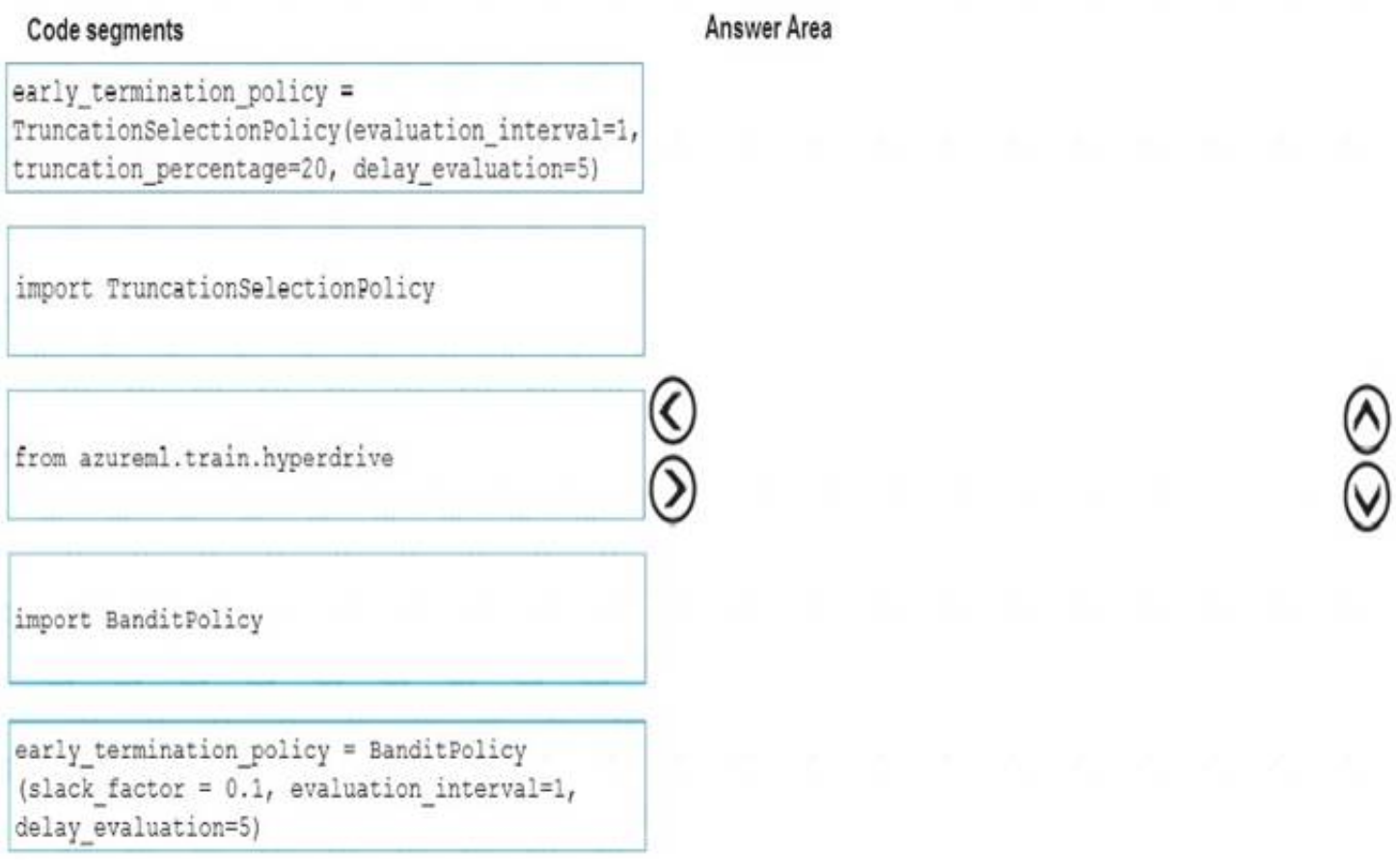
A. Mastered
B. Not Mastered

Answer: A

Explanation:



NEW QUESTION 176
- (Exam Topic 2)
You need to implement early stopping criteria as suited in the model training requirements.
Which three code segments should you use to develop the solution? To answer, move the appropriate code segments from the list of code segments to the answer area and arrange them in the correct order.
NOTE: More than one order of answer choices is correct. You will receive credit for any of the correct orders you select.



A. Mastered
B. Not Mastered

Answer: A

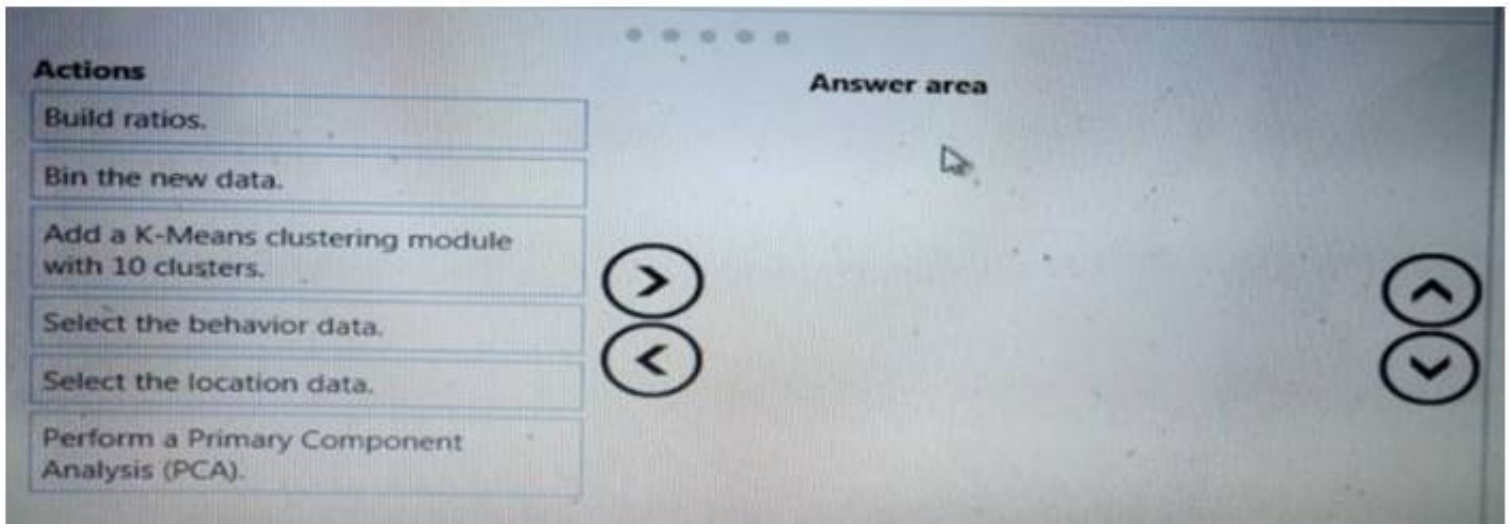
Explanation:
You need to implement an early stopping criterion on models that provides savings without terminating promising jobs.
Truncation selection cancels a given percentage of lowest performing runs at each evaluation interval. Runs are compared based on their performance on the primary metric and the lowest X% are terminated. Example:
from azureml.train.hyperdrive import TruncationSelectionPolicy
early_termination_policy = TruncationSelectionPolicy(evaluation_interval=1, truncation_percentage=20, delay_evaluation=5)

NEW QUESTION 179
- (Exam Topic 1)
You need to select an environment that will meet the business and data requirements. Which environment should you use?

A. Azure HDInsight with Spark MLlib
B. Azure Cognitive Services
C. Azure Machine Learning Studio
D. Microsoft Machine Learning Server

Answer: D

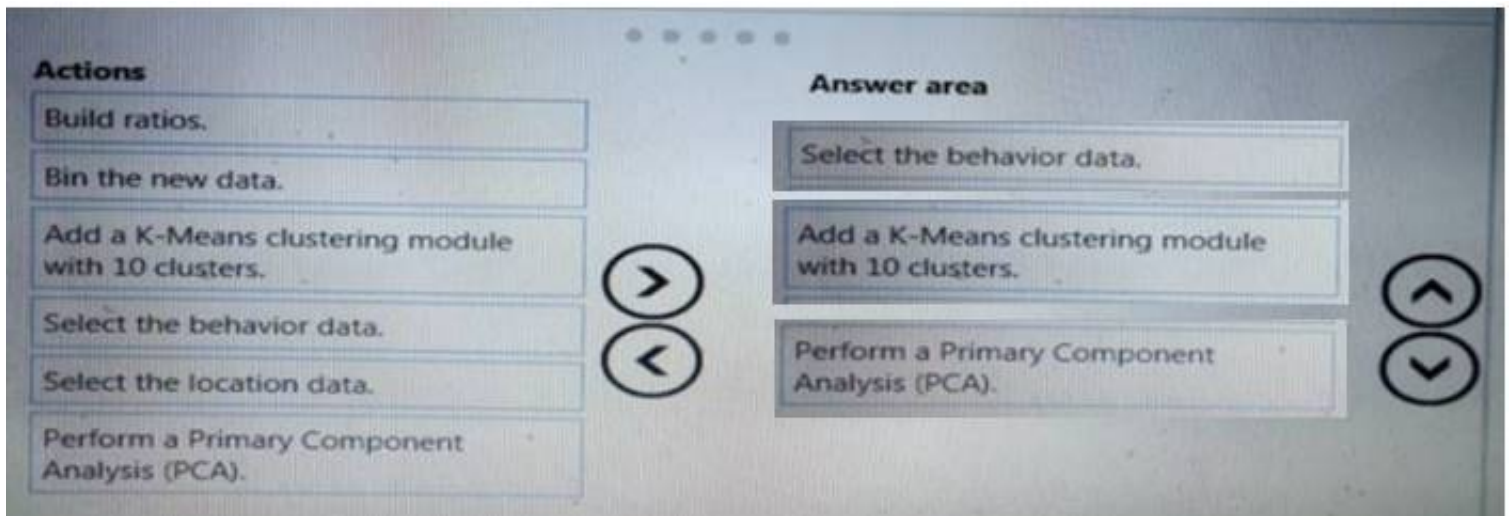
NEW QUESTION 181
- (Exam Topic 1)
You need to modify the inputs for the global penalty event model to address the bias and variance issue. Which three actions should you perform in sequence? To answer, move the appropriate actions from the list of actions to the answer area and arrange them in the correct order.



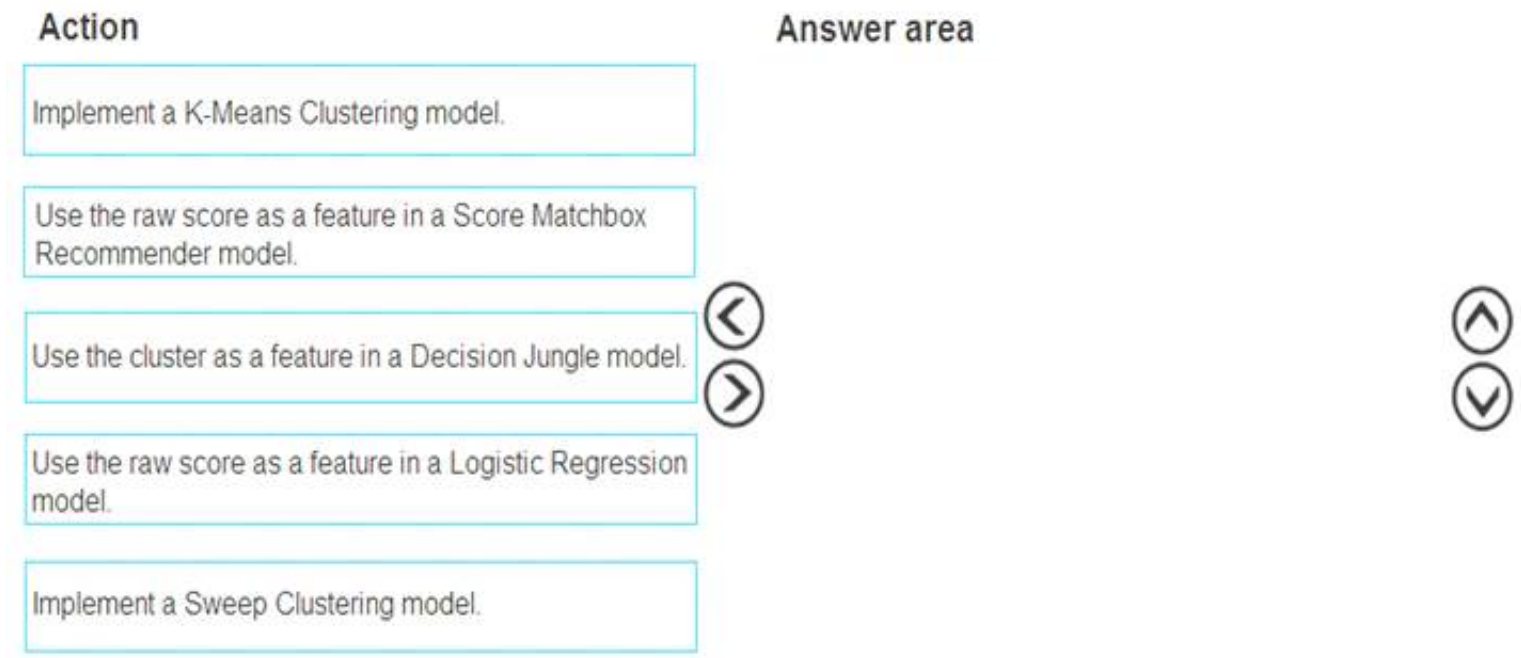
A. Mastered
B. Not Mastered

Answer: A

Explanation:



NEW QUESTION 185
- (Exam Topic 1)
You need to define a modeling strategy for ad response.
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.



A. Mastered
B. Not Mastered

Answer: A

Explanation:
Step 1: Implement a K-Means Clustering model
Step 2: Use the cluster as a feature in a Decision jungle model.
Decision jungles are non-parametric models, which can represent non-linear decision boundaries. Step 3: Use the raw score as a feature in a Score Matchbox Recommender model
The goal of creating a recommendation system is to recommend one or more "items" to "users" of the system. Examples of an item could be a movie, restaurant, book, or song. A user could be a person, group of persons, or other entity with item preferences.
Scenario:
Ad response rated declined.
Ad response models must be trained at the beginning of each event and applied during the sporting event. Market segmentation models must optimize for similar ad response history.
Ad response models must support non-linear boundaries of features. References:
https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/multiclass-decision-jungle https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/score-matchbox-recommende

NEW QUESTION 190
.....

THANKS FOR TRYING THE DEMO OF OUR PRODUCT

Visit Our Site to Purchase the Full Set of Actual DP-100 Exam Questions With Answers.

We Also Provide Practice Exam Software That Simulates Real Exam Environment And Has Many Self-Assessment Features. Order the DP-100 Product From:

https://www.2passeasy.com/dumps/DP-100/

Money Back Guarantee

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