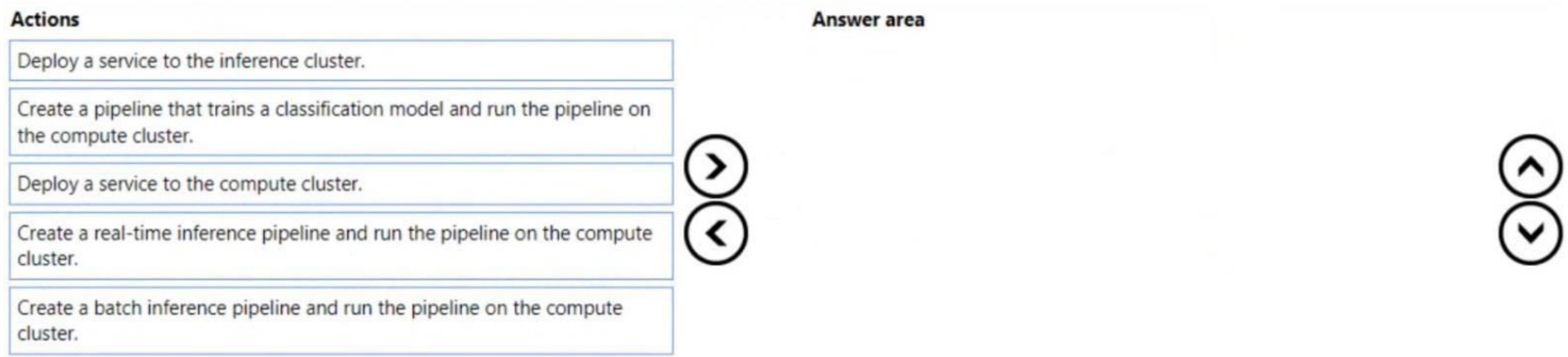
Microsoft.DP-100.v2026-06-26.q134

Exam Code:	DP-100
Exam Name:	Designing and Implementing a Data Science Solution on Azure
Certification Provider:	Microsoft
Free Question Number:	134
Version:	v2026-06-26
# of views:	127
# of Questions views:	1340
https://www.exam-tests.com/DP-100-exam/Microsoft.DP-100.v2026-06-26.q134.html	

NEW QUESTION: 1
Drag and Drop Question
You have an Azure Machine Learning workspace that contains a training cluster and an inference cluster.
You plan to create a classification model by using the Azure Machine Learning designer.
You need to ensure that client applications can submit data as HTTP requests and receive predictions as responses.
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.



Answer:

Actions

Deploy a service to the compute cluster.

Create a batch inference pipeline and run the pipeline on the compute cluster.

Answer area

Create a pipeline that trains a classification model and run the pipeline on the compute cluster.

Create a real-time inference pipeline and run the pipeline on the compute cluster.

Deploy a service to the inference cluster.

NEW QUESTION: 2

```
source_directory=scripts_folder,
entry_script="batch_pipeline.py",
mini_batch_size="5",
error_threshold=10,
output_action="append_row",
environment=batch_env,
compute_target=compute_target,
logging_level="DEBUG",
node_count=4)
```


You need to obtain the output from the pipeline execution. Where will you find the output?

A. a file named parallel_run_step.txt located in the output folder

B. the Inference Clusters tab in Machine Learning studio

C. the debug log

D. the Activity Log in the Azure portal for the Machine Learning workspace

E. the digitjdentification.py script

Answer: D (LEAVE A REPLY)

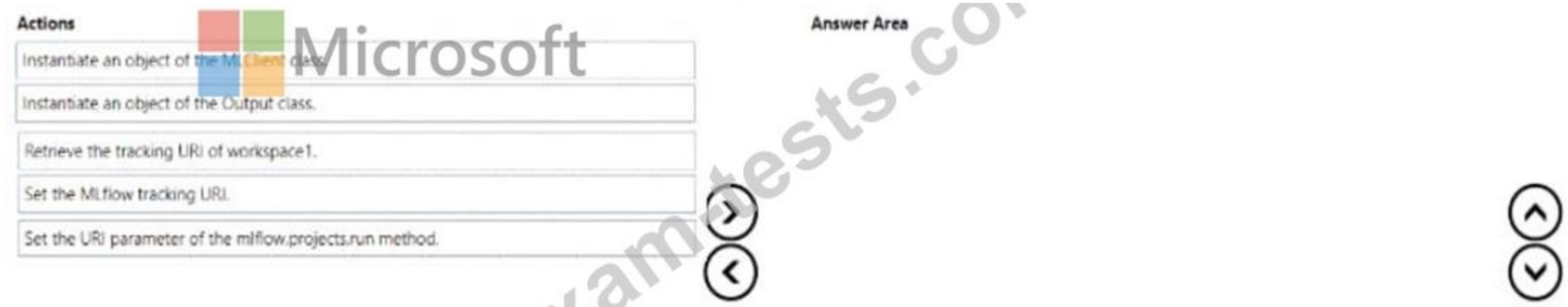
NEW QUESTION: 3

You manage an Azure Machine Learning workspace named workspace1 and a Data Science Virtual Machine (DSVM) named DSMV1.

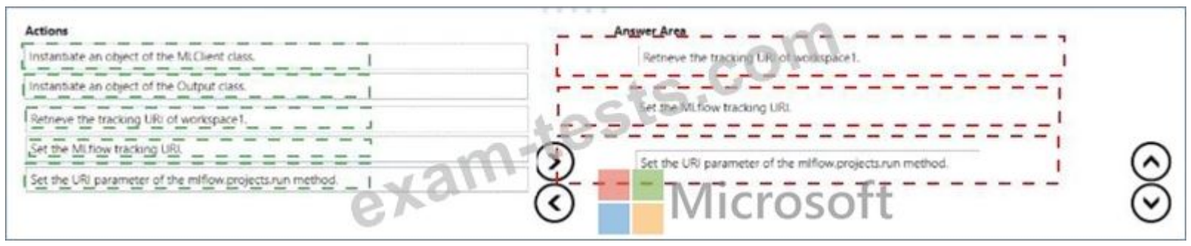
You must an experiment in DSMV1 by using a Jupiter notebook and Python SDK v2 code. You must store metrics and artifacts in workspace 1 You start by creating Python SCK v2 code to import ail required packages.

You need to implement the Python SOK v2 code to store metrics and article in workspace1.

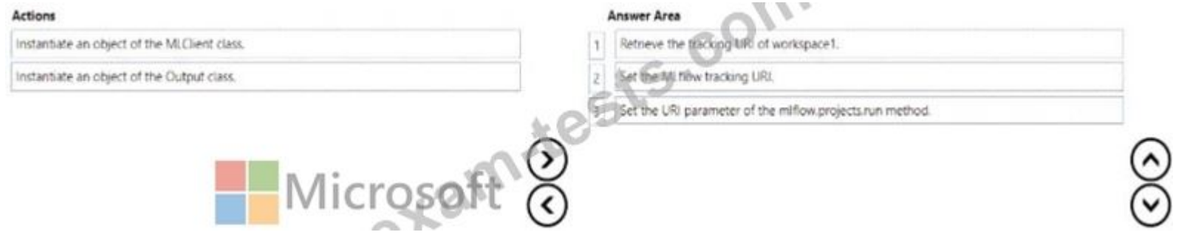
Which three actions should you perform in sequence? To answer, move the appropriate actions from the list of actions to the answer area and arrange them the correctly order.



Answer:



Explanation



NEW QUESTION: 4

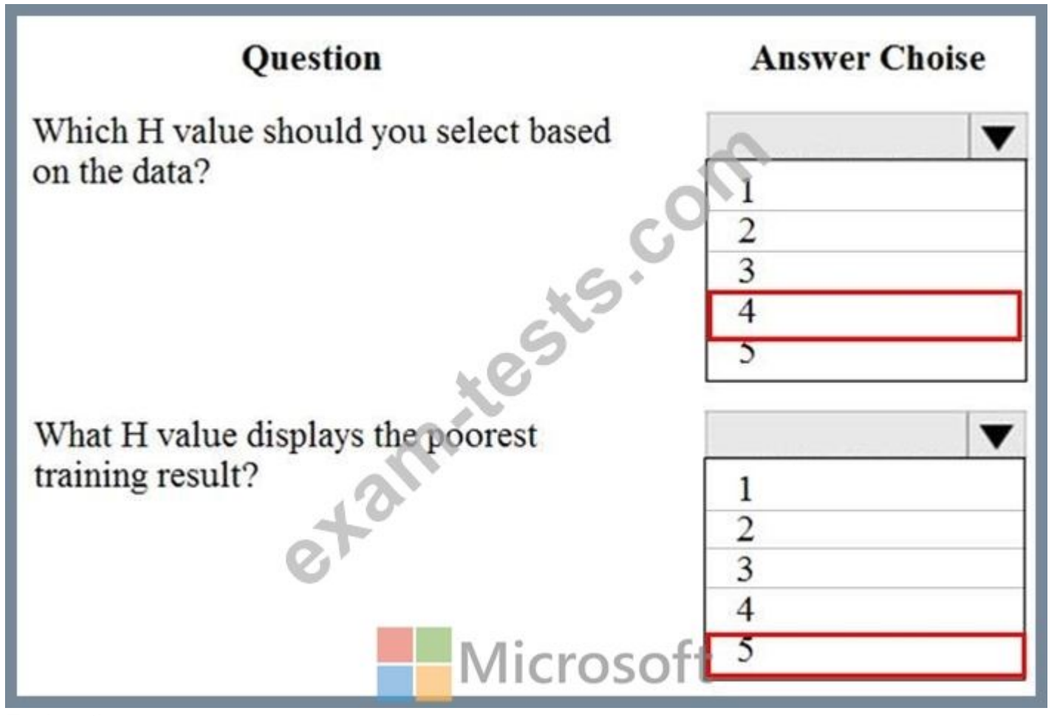
You are tuning a hyperparameter for an algorithm. The following table shows a data set with different hyperparameter, training error, and validation errors.

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

Answer:



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-2d2b85651bbd

NEW QUESTION: 5

You are preparing to use the Azure ML SDK to run an experiment and need to create compute. You run the following code:

```
from azureml.core.compute import ComputeTarget, AmlCompute
from azureml.core.compute_target import ComputeTargetException
ws = Workspace.from_config()
cluster_name = 'aml-cluster'
try:
  training_compute = ComputeTarget(workspace=ws, name=cluster_name)
except ComputeTargetException:
  compute_config = AmlCompute.provisioning_configuration(vm_size='STANDARD_D2_V2', vm_priority='lowpriority',
max_nodes=4)
  training_compute = ComputeTarget.create(ws, cluster_name, compute_config)
  training_compute.wait_for_completion(show_output=True)
```

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

	Yes	No
If a training cluster named aml-cluster already exists in the workspace, it will be deleted and replaced.	○	○
The `wait_for_completion()` method will not return until the aml-cluster compute has four active nodes.	○	○
If the code creates a new aml-cluster compute target, it may be preempted due to capacity constraints.	○	○
The aml-cluster compute target is deleted from the workspace after the training experiment completes.	○	○

Answer:

	Yes	No
If a training cluster named aml-cluster already exists in the workspace, it will be deleted and replaced.	○	●
The `wait_for_completion()` method will not return until the aml-cluster compute has four active nodes.	●	○
If the code creates a new aml-cluster compute target, it may be preempted due to capacity constraints.	●	○
The aml-cluster compute target is deleted from the workspace after the training experiment completes.	○	●

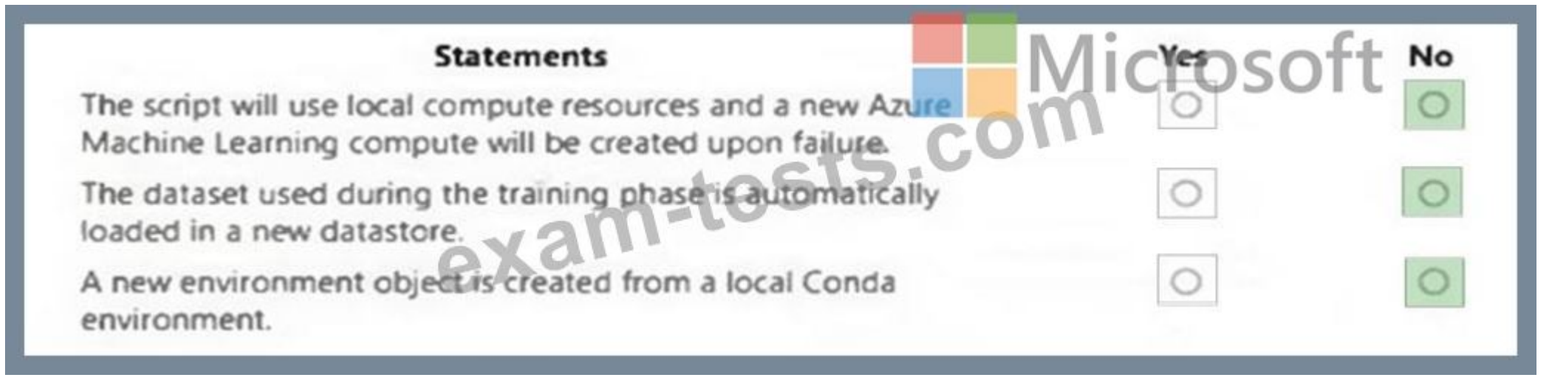
Explanation:
Box 1: No
If a training cluster already exists it will be used.
Box 2: Yes
The wait_for_completion method waits for the current provisioning operation to finish on the cluster.
Box 3: Yes
Low Priority VMs use Azure's excess capacity and are thus cheaper but risk your run being pre-empted.
Box 4: No
Need to use training_compute.delete() to deprovision and delete the AmlCompute target.
Reference:
https://notebooks.azure.com/azureml/projects/azureml-getting-started/html/how-to-use-azureml/training/train-on-amlcompute/train-on-amlcompute.ipynb
https://docs.microsoft.com/en-us/python/api/azureml-core/azureml.core.compute.computetarget

NEW QUESTION: 6
Hotspot Question
You create an Azure Machine Learning workspace and load a Python training script named train.py in the src subfolder. The dataset used to train your model is available locally.
You run the following script to train the model:

```
ws = Workspace.from_config()
experiment = Experiment(workspace=ws, name='nlp-experiment-train')

cpu_cluster_name = "cpu-cluster"
try:
    cpu_cluster = ComputeTarget(workspace=ws, name=cpu_cluster_name)
except ComputeTargetException:
    compute_config = AmlCompute.provisioning_configuration(vm_size='STANDARD_D2_V2', max_nodes=4)
    cpu_cluster = ComputeTarget.create(ws, cpu_cluster_name, compute_config)

cpu_cluster.wait_for_completion(show_output=True)

config = ScriptRunConfig(source_directory='./src',
                         script='train.py',
                         compute_target='cpu-cluster')

env = Environment.from_conda_specification(
    name='pytorch-env',
    file_path='./.azureml/pytorch-env.yml'
)
config.run_config.environment = env

run = experiment.submit(config)
```

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.

Answer:



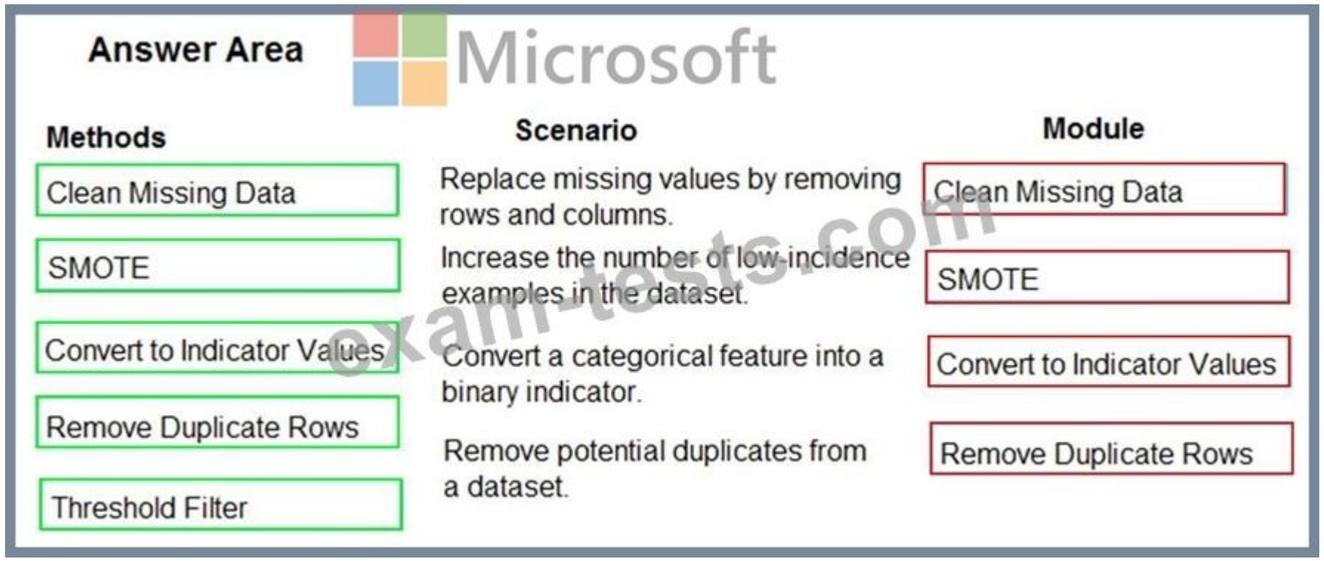
Explanation:
https://learn.microsoft.com/en-us/python/api/azureml-core/azureml.core.environment(class)?view=azure-ml-p

NEW QUESTION: 7
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.

Answer Area

Methods	Scenario	Module
Clean Missing Data	Replace missing values by removing rows and columns.	
SMOTE	Increase the number of low-incidence examples in the dataset.	
Convert to Indicator Values	Convert a categorical feature into a binary indicator.	
Remove Duplicate Rows	Remove potential duplicates from a dataset.	
Threshold Filter		

Answer:



Explanation:

Box 1: Clean Missing Data

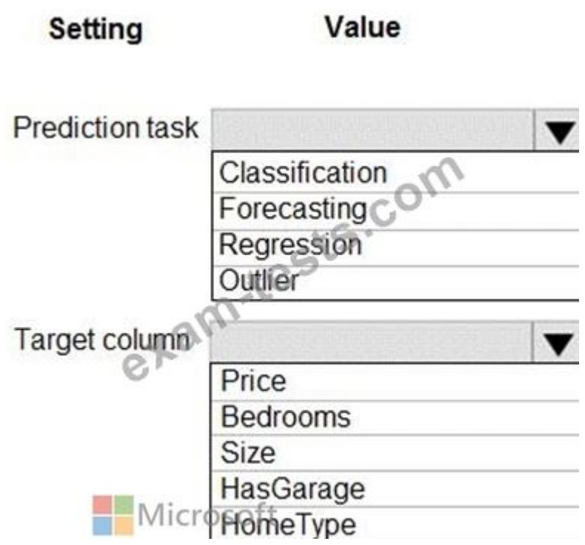
Box 2: SMOTE

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

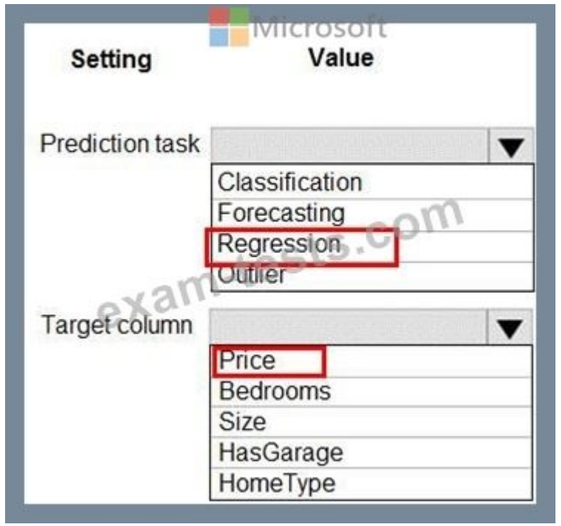
NEW QUESTION: 8
You have a dataset that includes home sales data for a city. The dataset includes the following columns.

Name	Description
Price	The sales price for the house.
Bedrooms	The number of bedrooms in the house.
Size	The size of the house in square feet.
HasGarage	A binary value indicating whether or not the house has a garage.
HomeType	The category of home, for example, apartment, townhouse, single-family home.

Each row in the dataset corresponds to an individual home sales transaction.
You need to use automated machine learning to generate the best model for predicting the sales price based on the features of the house.
Which values should you use? To answer, select the appropriate options in the answer area.
NOTE: Each correct selection is worth one point.



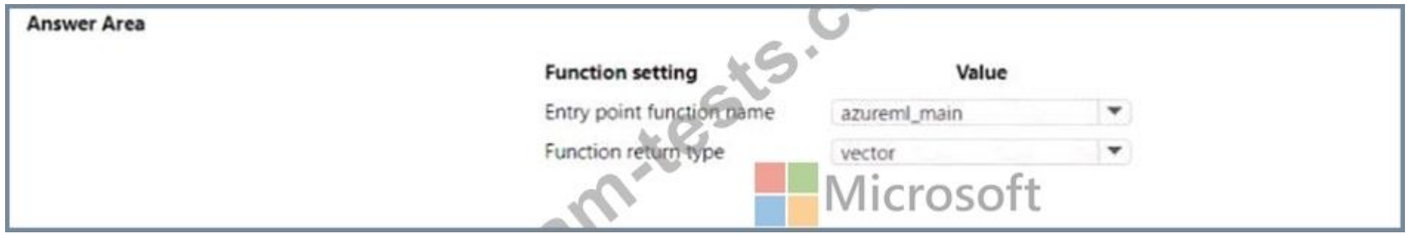
Answer:



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

NEW QUESTION: 9
You manage an Azure AI Foundry project.
You develop a Prompt flow that includes a large language model (LLM) node and an upstream node with a single output. You need to link the LLM node input with the output of the upstream node by using a YAML flow configuration. Which flow configuration should you use?
A. {{upstream.node.nameoutput})
B. (% upstream node_name,output%}
C. $(upstream_node_name.output)
D. <#upstream_node_nameoutput#>
Answer: D (LEAVE A REPLY)

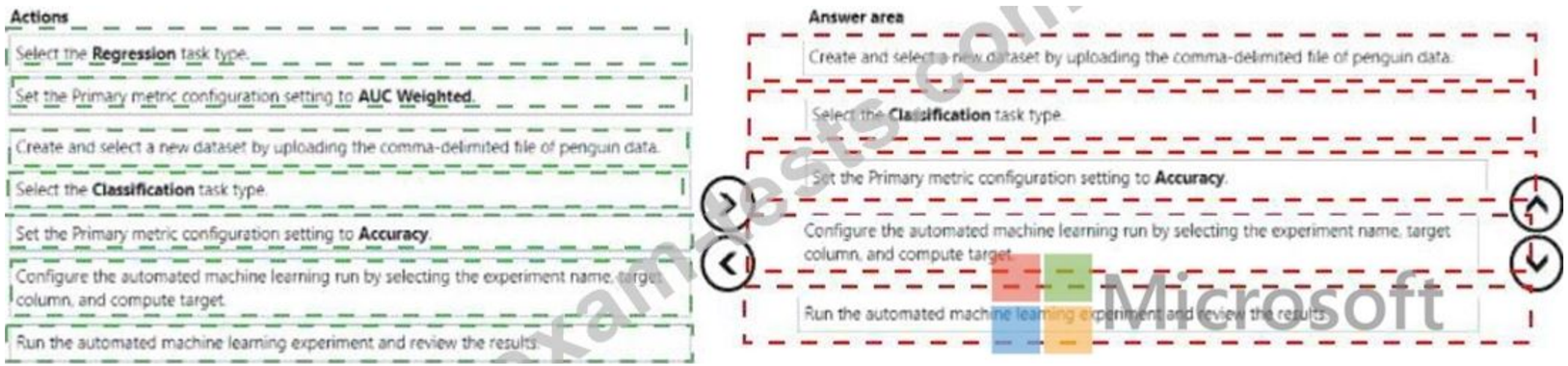
NEW QUESTION: 10
You create an Azure Machine Learning dataset containing automobile price data. The dataset includes 10.000 rows and 10 columns. You use the Azure Machine Learning designer to transform the dataset by using an Execute Python Script component and custom code.
The code must combine three columns to create a new column.
You need to configure the code function.
Which configurations should you use? To answer, select the appropriate options in the answer area.
NOTE: Each correct selection is worth one point.
Answer:



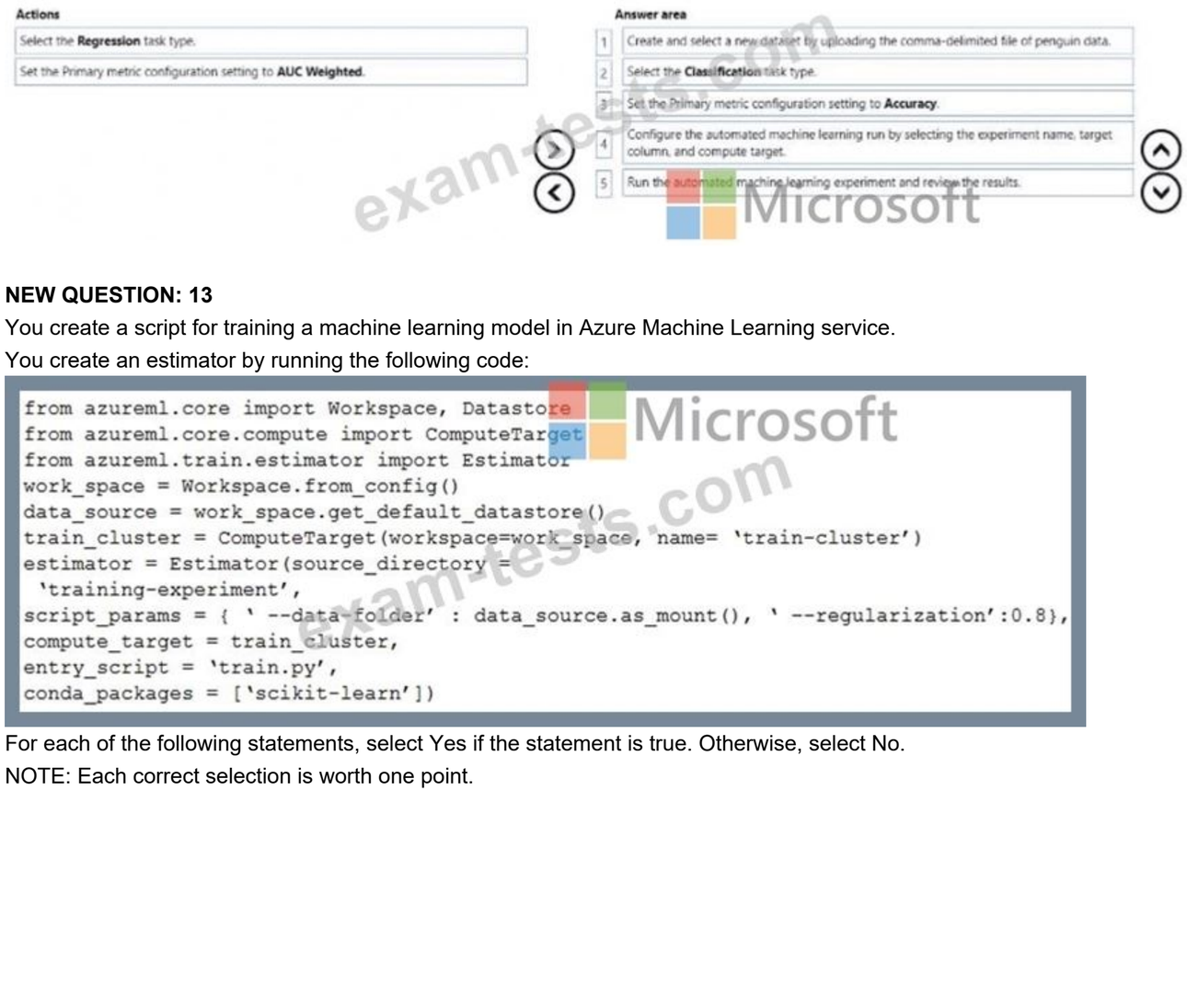
NEW QUESTION: 11
You manage an Azure Machine Learning workspace by using the Azure CLI ml extension v2. You need to define a YAML schema to create a compute cluster. Which schema should you use?
A. https://azuremlschemas.azureedge.net/latest/computdnstarKeichema.json
B. https://azuremlschemas.azureedge.net/latest/vmCompute.schema.json
C. https://azuremlschemas.azureedge.net/latest/amlCompute.schemajson
D. https://azuremlschemas.azureedge.net/latest/kubernetesCompute.schema.json
Answer: C (LEAVE A REPLY)

NEW QUESTION: 12
You are creating a machine learning model that can predict the species of a penguin from its measurements.
You have a file that contains measurements for free species of penguin in comma delimited format.
The model must be optimized for area under the received operating characteristic curve performance metric averaged for each class.
You need to use the Automated Machine Learning user interface in Azure Machine Learning studio to run an experiment and find the best performing model.
Which five 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 collect order.

Answer:



Explanation:



NEW QUESTION: 13

You create a script for training a machine learning model in Azure Machine Learning service.

You create an estimator by running the following code:

```
from azureml.core import Workspace, Datastore
from azureml.core.compute import ComputeTarget
from azureml.train.estimator import Estimator
work_space = Workspace.from_config()
data_source = work_space.get_default_datastore()
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.

	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.	○	○
The estimator can run Scikit-learn experiments.	○	○

Answer:

	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.	○	○
The estimator can run Scikit-learn experiments.	○	○

NEW QUESTION: 14
You are with a time series dataset in Azure Machine Learning Studio.
You need to split your dataset into training and testing subsets by using the Split Data module.
Which splitting mode should you use?
A. Split Rows with the Randomized split parameter set to true
B. Recommender Split
C. Relative Expression Split
D. Regular Expression Split
Answer: A (LEAVE A REPLY)
Topic 2, Case Study
Overview
You are a data scientist for Fabrikam Residences, a company specializing in quality private and commercial property in the United States. Fabrikam Residences is considering expanding into Europe and has asked you to investigate prices for private residences in major European cities. You use Azure Machine Learning Studio to measure the median value of properties. You produce a regression model to predict property prices by using the Linear Regression and Bayesian Linear Regression modules.
Datasets

There are two datasets in CSV format that contain property details for two cities, London and Paris, with the following columns:

Column heading	Description
CapitaCrimeRate	per capita crime rate by town
Zoned	proportion of residential land zoned for lots over 25.000 square feet
NonRetailAcres	proportion of retail business acres per town
NextToRiver	proximity of the property to the river
NitrogenOxideConcentration	nitric oxides concentration (parts per 10 million)
AvgRoomsPerHouse	average number of rooms per dwelling
Age	proportion of owner-occupied units built prior to 1940
DistanceToEmploymentCenter	weighted distances to employment centers
AccessibilityToHighway	index of accessibility to radial highways to a value of two decimal places
Tax	full value property tax rate per $10,000
PupilTeacherRatio	pupil to teacher ratio by town
ProfessionalClass	professional class percentage
LowerStatus	percentage lower status of the population
MedianValue	median value of owner-occupied homes in $1000s

The two datasets have been added to Azure Machine Learning Studio as separate datasets and included as the starting point of the experiment.

Dataset issues

The AccessibilityToHighway column in both datasets contains missing values. The missing data must be replaced with new data so that it is modeled conditionally using the other variables in the data before filling in the missing values.

Columns in each dataset contain missing and null values. The dataset also contains many outliers. The Age column has a high proportion of outliers. You need to remove the rows that have outliers in the Age column. 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.

Model fit

The model shows signs of overfitting. You need to produce a more refined regression model that reduces the overfitting.

Experiment Requirements

You must set up the experiment to cross-validate the Linear Regression and Bayesian Linear Regression modules to evaluate performance.

In each case, the predictor of the dataset is the column named MedianValue. An initial investigation showed 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. You must ensure that the datatype of the MedianValue column of the Paris dataset matches the structure of the London dataset.

You must prioritize the columns of data for predicting the outcome. You must use non-parameters statistics to measure the relationships.

You must use a feature selection algorithm to analyze the relationship between the MedianValue and AvgRoomsinHouse columns.

Model training

Given a trained model and a test dataset, you need to compute the permutation feature importance scores of feature variables. You need to set up the Permutation Feature Importance module to select the correct metric to investigate the model's accuracy and replicate the findings.

You want to configure hyperparameters in the model learning process to speed the learning phase by using hyperparameters. In addition, this configuration should cancel the lowest performing runs at each evaluation interval, thereby directing effort and resources towards models that are more likely to be successful.

You are concerned that the model might not efficiently use compute resources in hyperparameter tuning. You also are concerned that the model might prevent an increase in the overall tuning time. Therefore, you need to implement an early stopping criterion on models that provides savings without terminating promising jobs.

Testing

You must produce multiple partitions of a dataset based on sampling using the Partition and Sample module in Azure Machine Learning Studio. You must create three equal partitions for cross-validation. You must also configure the cross-validation process so that the rows in the test and training datasets are divided evenly by properties that are near each city's main river. The data that identifies that a property is near a river is held in the column named NextToRiver. You want to complete this task before the data goes through the sampling process.

When you train a Linear Regression module using a property dataset that shows data for property prices for a large city, you need to determine the best features to use in a model. You can choose standard metrics provided to measure performance before and after the feature importance process completes. You must ensure that the distribution of the features across multiple training models is consistent.

Data visualization

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.

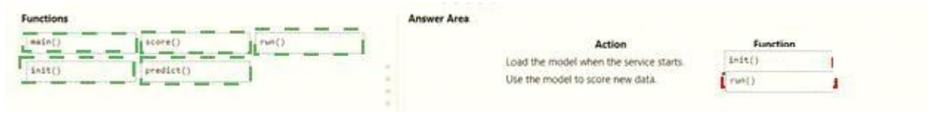
NEW QUESTION: 15

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.

Answer:



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()

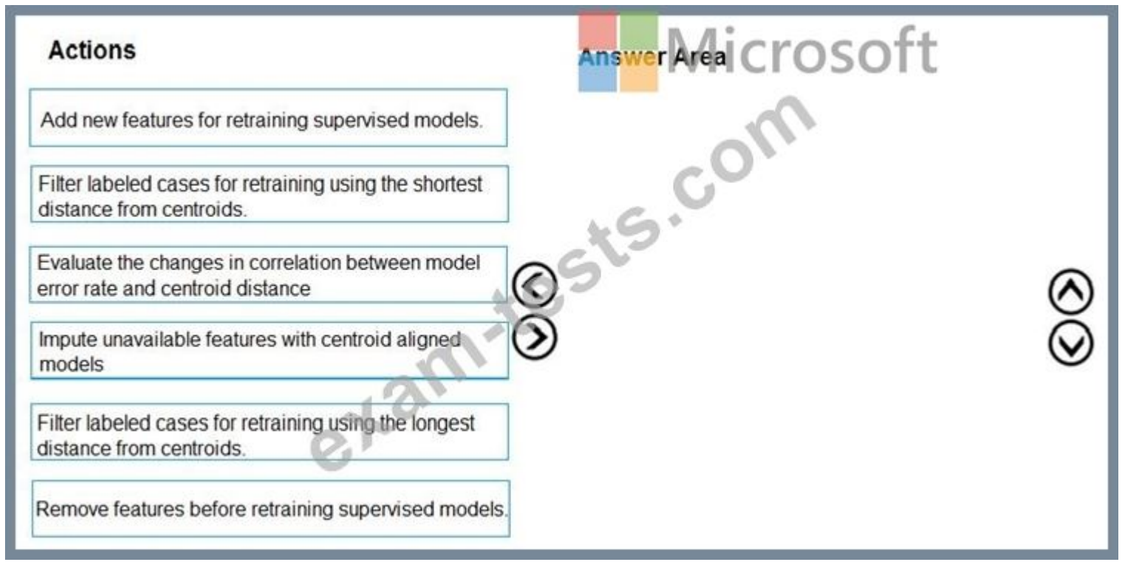
Reference:

https://docs.microsoft.com/en-us/azure/machine-learning/how-to-deploy-existing-model

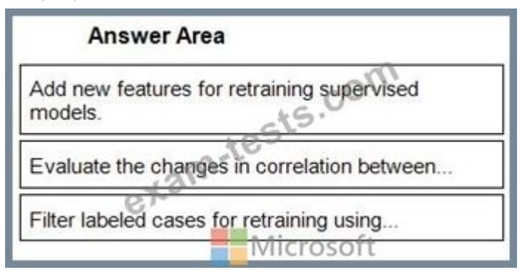
NEW QUESTION: 16

You need to define an evaluation strategy for the crowd sentiment models.

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.



Answer:



1 - Add new features for retraining supervised models.
2 - Evaluate the changes in correlation between...
3 - Filter labeled cases for retraining using...
Reference:
https://en.wikipedia.org/wiki/Nearest_centroid_classifier
https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/sweep-clustering

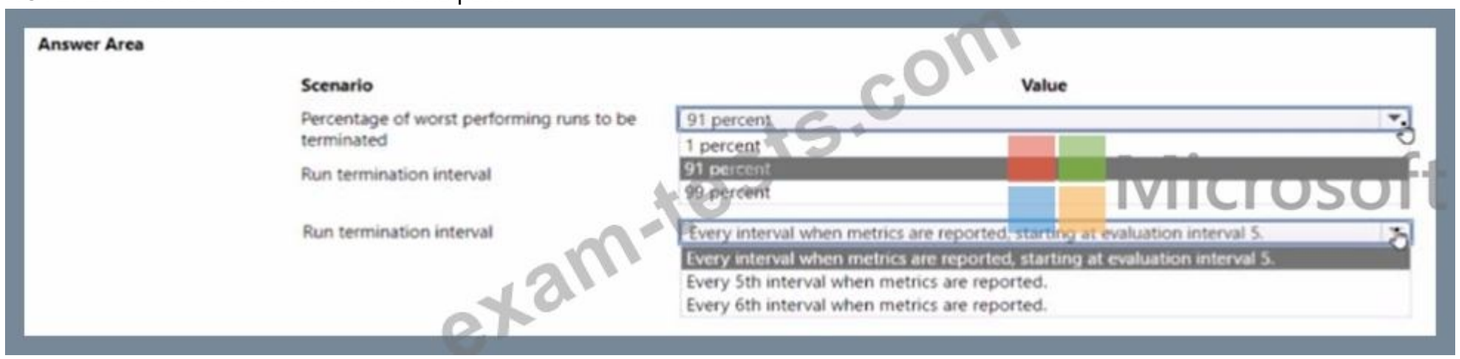
NEW QUESTION: 17
Note: This question is part of a series of questions that present the same scenario. Each question in the series contains a unique solution that might meet the stated goals. Some question sets might have more than one correct solution, while others might not have a correct solution.
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:

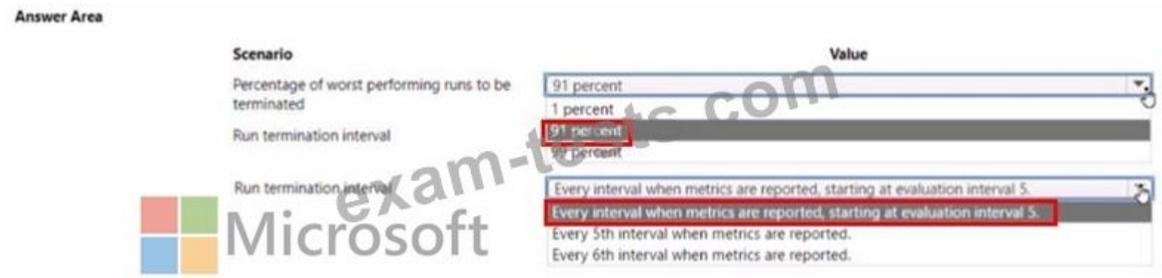
```
from azureml.train.sklearn import SKLearn
sk_est = SKLearn(source_directory='./scripts',
 compute_target=aml-compute,
 entry_script='train.py')
```

Does the solution meet the goal?
A. Yes
B. No
Answer: A (LEAVE A REPLY)
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: 18
You use Azure Machine Learning to implement hyperparameter tuning with a Bandit early termination policy.
The policy uses a slack_factor set to 01. an evaluation interval set to 1, and an evaluation delay set to b.
You need to evaluate the outcome of the early termination policy
What should you evaluate? To answer, select the appropriate options m the answer area.
NOTE: Each correct selection is worth one point.



Answer:



NEW QUESTION: 19

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:

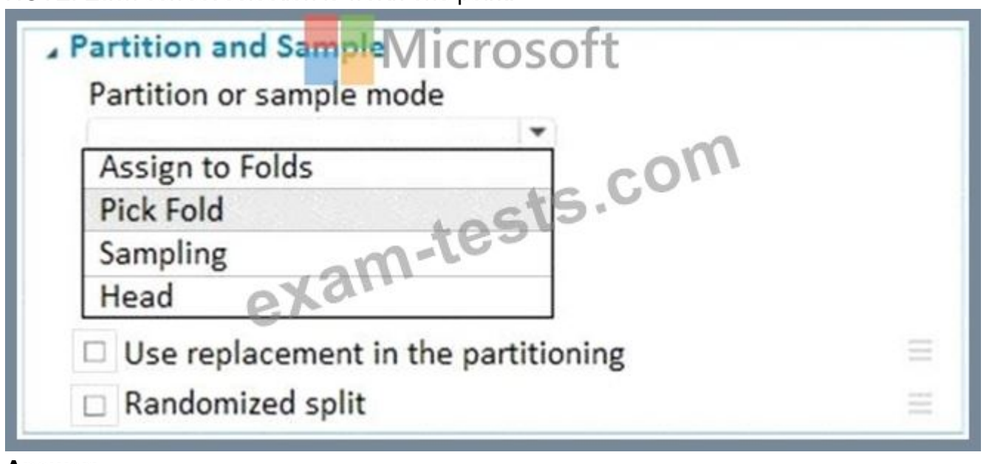
Divide the data into subsets

Assign the rows into folds using a round-robin method

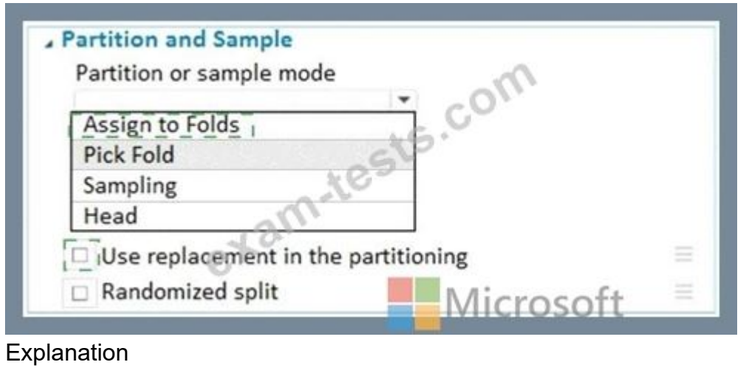
Allow rows in the dataset to be reused

How should you configure the module? To answer, select the appropriate options in the dialog box in the answer area.

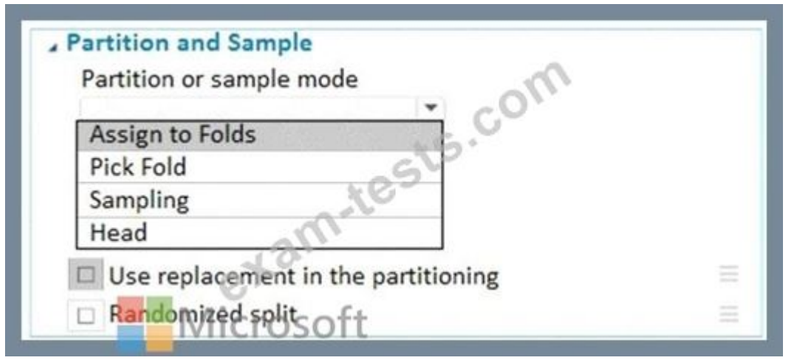
NOTE: Each correct selection is worth one point.



Answer:



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.

Add the Partition and Sample module to your experiment in Studio (classic), and connect the dataset.

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: 20

You create a multi-class image classification deep learning model that uses the PyTorch deep learning framework.

You must configure Azure Machine Learning Hyperdrive to optimize the hyperparameters for the classification model.

You need to define a primary metric to determine the hyperparameter values that result in the model with the best accuracy score.

Which three actions must you perform? Each correct answer presents part of the solution.

NOTE: Each correct selection is worth one point.

A. Set the primary_metric_goal of the estimator used to run the bird_classifier_train.py script to maximize.

B. Add code to the bird_classifier_train.py script to calculate the validation loss of the model and log it as a float value with the key loss.

C. Set the primary_metric_goal of the estimator used to run the bird_classifier_train.py script to minimize.

D. Set the primary_metric_name of the estimator used to run the bird_classifier_train.py script to accuracy.

E. Set the primary_metric_name of the estimator used to run the bird_classifier_train.py script to loss.

F. Add code to the bird_classifier_train.py script to calculate the validation accuracy of the model and log it as a float value with the key accuracy.

Answer: A,D,F (LEAVE A REPLY)

Explanation

AD:

primary_metric_name="accuracy",

primary_metric_goal=PrimaryMetricGoal.MAXIMIZE

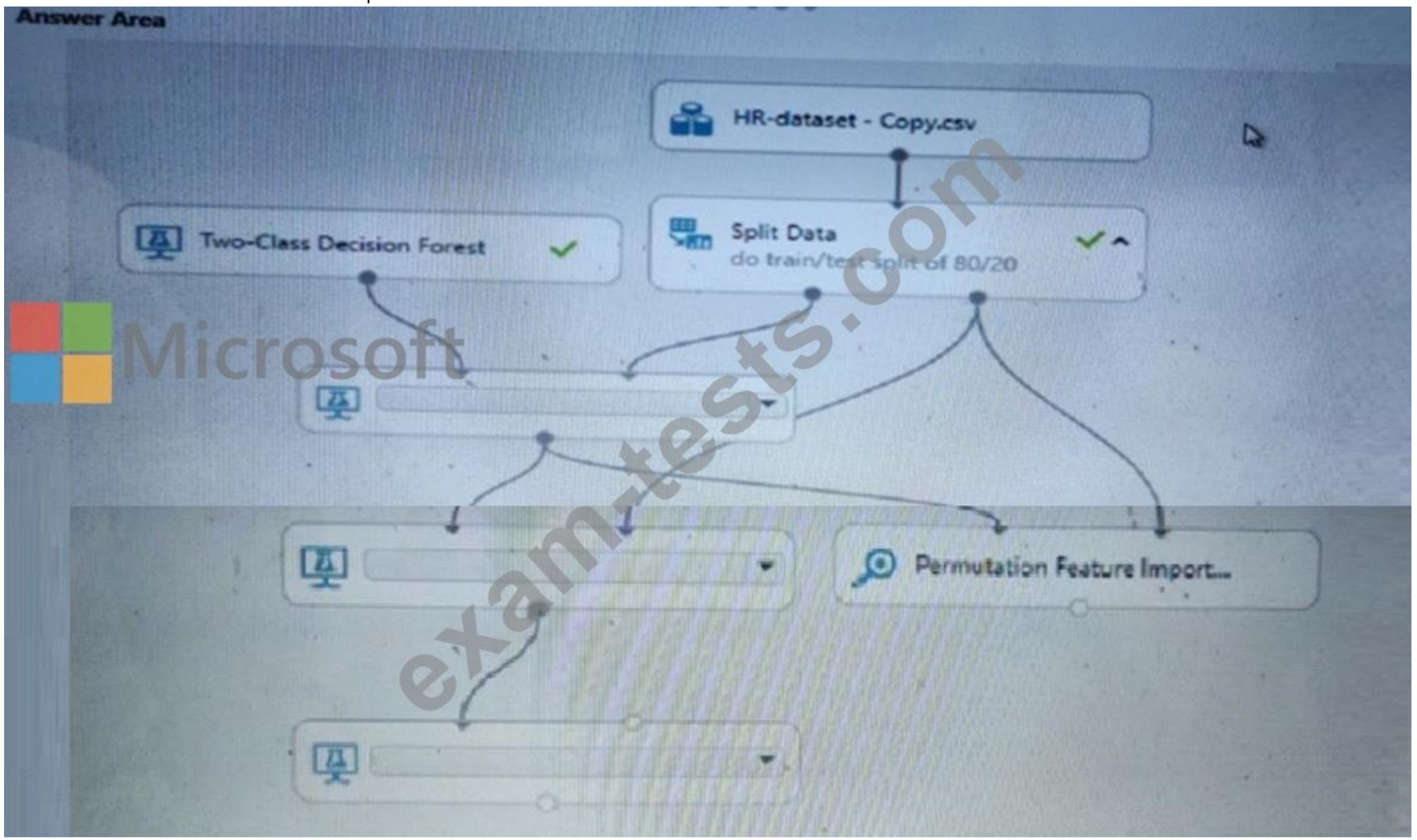
Optimize the runs to maximize "accuracy". Make sure to log this value in your training script.

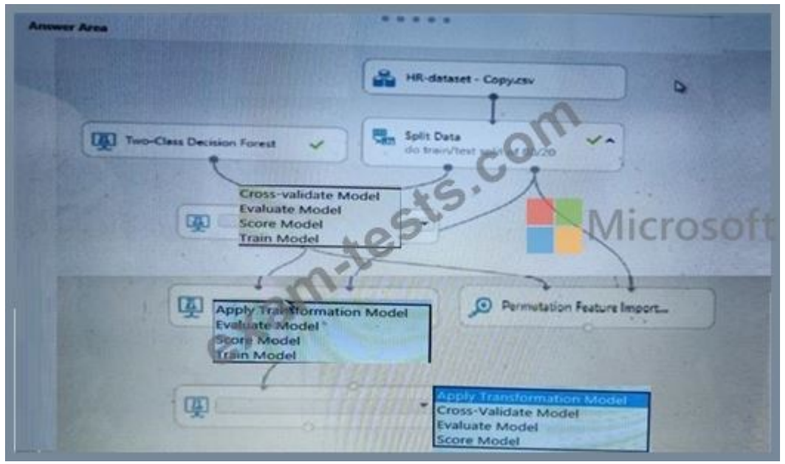
Note:

primary_metric_name: The name of the primary metric to optimize. The name of the primary metric needs to exactly match the name of the metric logged by the training script.

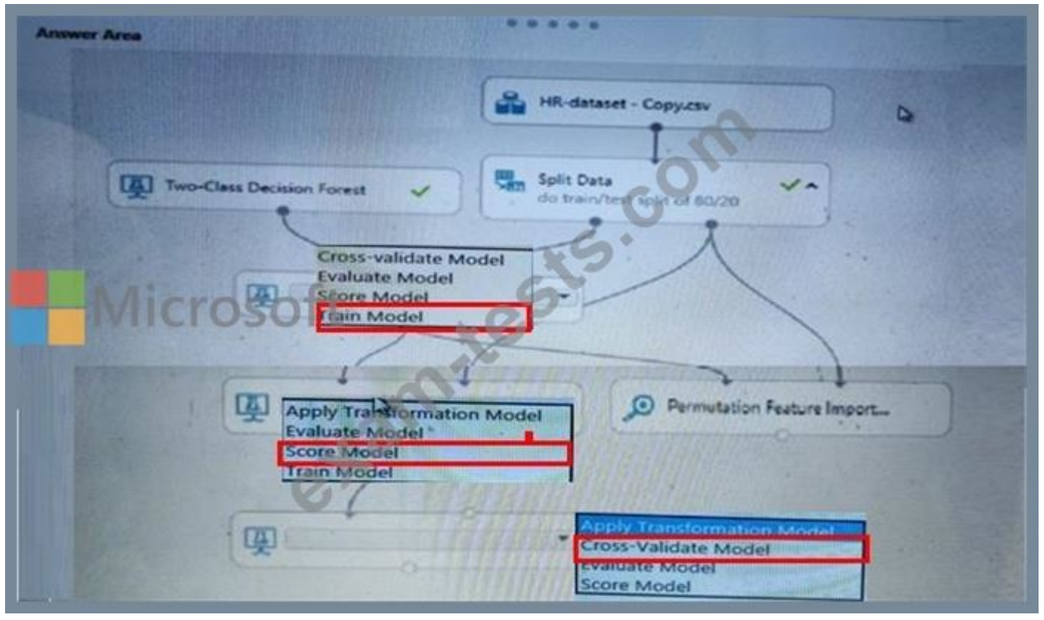
primary_metric_goal: It can be either PrimaryMetricGoal.MAXIMIZE or PrimaryMetricGoal.MINIMIZE and determines whether the primary metric will be maximized or minimized when evaluating the runs.
F: The training script calculates the val_accuracy and logs it as "accuracy", which is used as the primary metric.

NEW QUESTION: 21
You create a binary classification model using Azure Machine Learning Studio.
You must use a Receiver Operating Characteristic (RO C) curve and an F1 score to evaluate the model.
You need to create the required business metrics.
How should you complete the experiment? To answer, select the appropriate options in the dialog box in the answer area.
NOTE: Each correct selection is worth one point.





Answer:



NEW QUESTION: 22

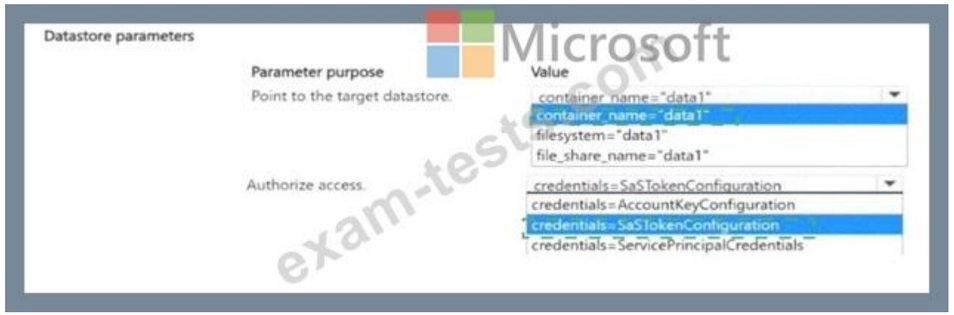
You manage an Azure Machine Learning workspace named Workspace1 and an Azure Blob Storage accessed by using the URL https://storage1.blob.core.wmdows.net/data1.

You plan to create an Azure Blob datastore in Workspace1. The datastore must target the Blob Storage by using Azure Machine Learning Python SDK v2. Access authorization to the datastore must be limited to a specific amount of time.

You need to select the parameters of the Azure Blob Datastore class that will point to the target datastore and authorize access to it.

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

Answer:



Explanation:

NEW QUESTION: 23

You are retrieving data from a large datastore by using Azure Machine Learning Studio.

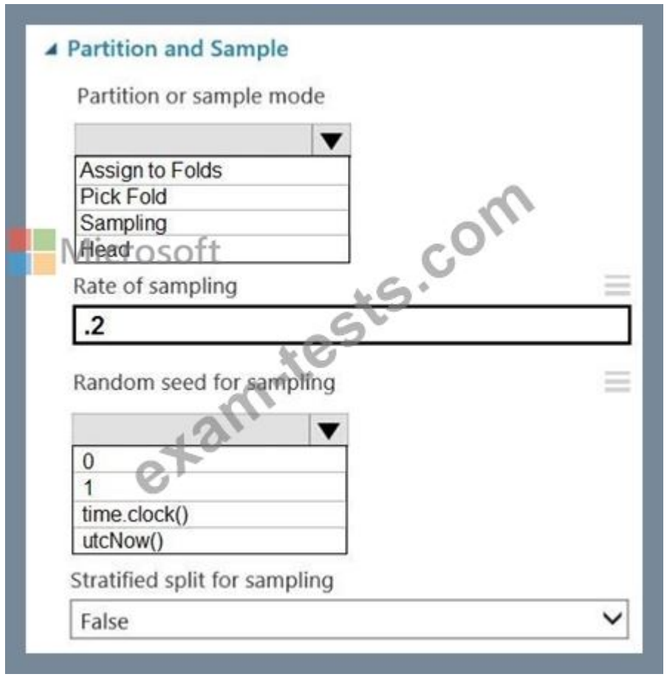
You must create a subset of the data for testing purposes using a random sampling seed based on the system clock.

You add the Partition and Sample module to your experiment.

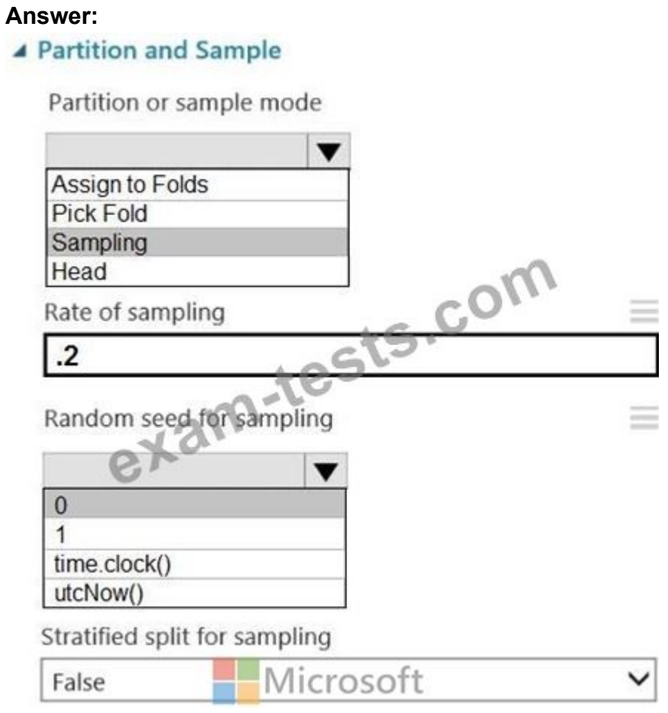
You need to select the properties for the module.

Which values should you select? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.



Answer:

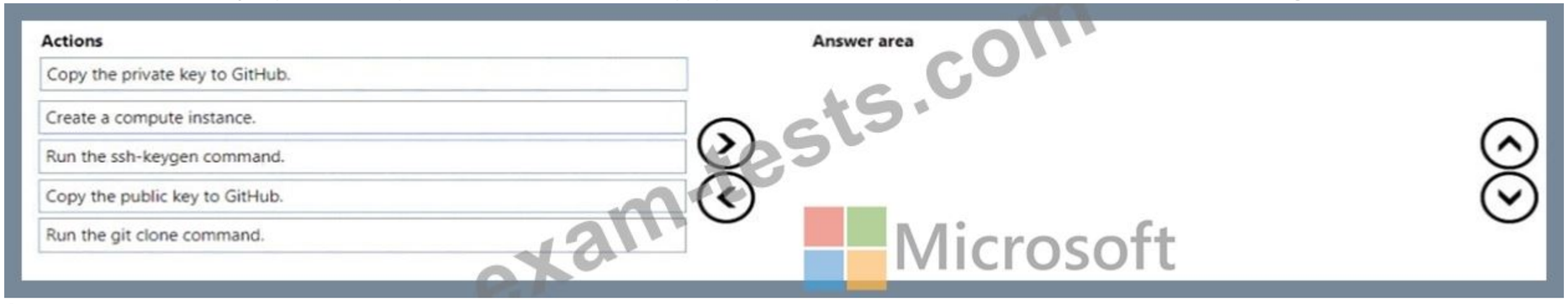


Box 1: Sampling

Create a sample of data

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

NEW QUESTION: 24
You create an Azure Machine Learning workspace.
You need to use the shared file system of the workspace to store a clone of a private Git repository.
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.



Answer:

Explanation

NEW QUESTION: 25

You are performing a classification task in Azure Machine Learning Studio.

You must prepare balanced testing and training samples based on a provided data set.

You need to split the data with a 0.75:0.25 ratio.

Which value should you use for each parameter? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.

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

Answer:

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

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

NEW QUESTION: 26
You register the following versions of a model.

Model name	Model version	Tags	Properties
healthcare_model	3	'Training context':'CPU Compute'	value:87.43
healthcare_model	2	'Training context':'CPU Compute'	value:54.98
healthcare_model	1	'Training context':'CPU Compute'	value:23.56

You use the Azure ML Python SDK to run a training experiment. You use a variable named run to reference the experiment run.
After the run has been submitted and completed, you run the following code:

```
run.register_model(model_path='outputs/model.pkl',
 model_name='healthcare_model',
 tags={'Training context':'CPU Compute'} )
```

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

	Yes	No
The code will cause a previous version of the saved model to be overwritten.	○	○
The version number will now be 4.	○	○
The latest version of the stored model will have a property of value: 87.43.	○	○

Answer:

	Yes	No
The code will cause a previous version of the saved model to be overwritten.	○	○
The version number will now be 4.	○	○
The latest version of the stored model will have a property of value: 87.43.	○	○

Explanation:

	Yes	No
The code will cause a previous version of the saved model to be overwritten.	○	○
The version number will now be 4.	○	○
The latest version of the stored model will have a property of value: 87.43.	○	○

Reference:
https://docs.microsoft.com/en-us/azure/machine-learning/how-to-deploy-and-where

NEW QUESTION: 27
Note: This question is part of a series of questions that present the same scenario. Each question in the series contains a unique solution that might meet the stated goals. Some question sets might have more than one correct solution, while others might not have a correct solution.
After you answer a question in this section, you will NOT be able to return to it. As a result, these questions will not appear in the review screen.
You create a model to forecast weather conditions based on historical data.
You need to create a pipeline that runs a processing script to load data from a datastore and pass the processed data to a machine learning model training script.

Solution: Run the following code:

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

Does the solution meet the goal?

A. Yes

B. No

Answer: B (LEAVE A REPLY)

train_step is missing.

Reference:

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

NEW QUESTION: 28

Drag and Drop Question

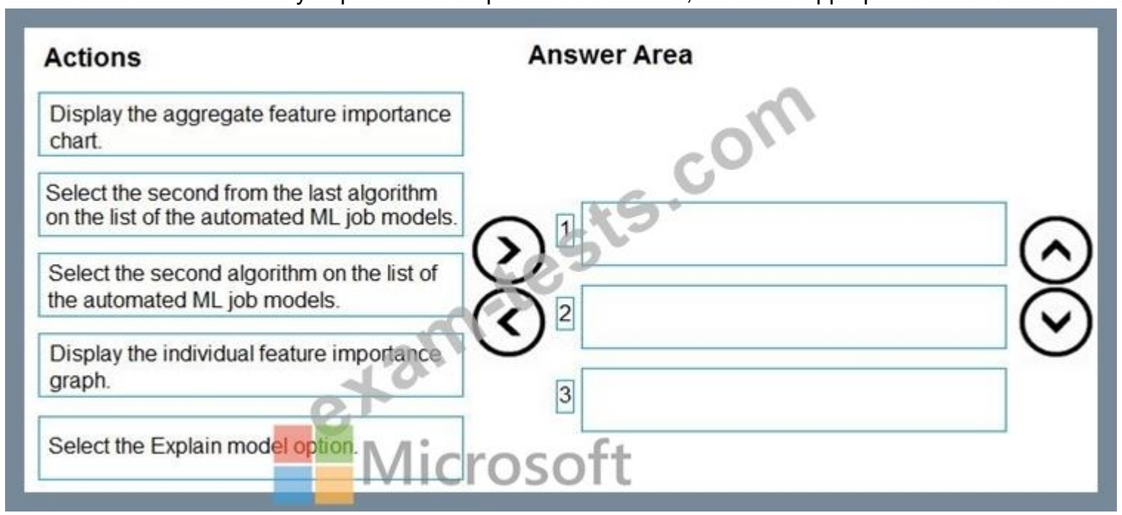
You create an Azure Machine Learning workspace. You are training a classification model with no-code AutoML in Azure Machine Learning studio.

The model must predict if a client of a financial institution will subscribe to a fixed-term deposit.

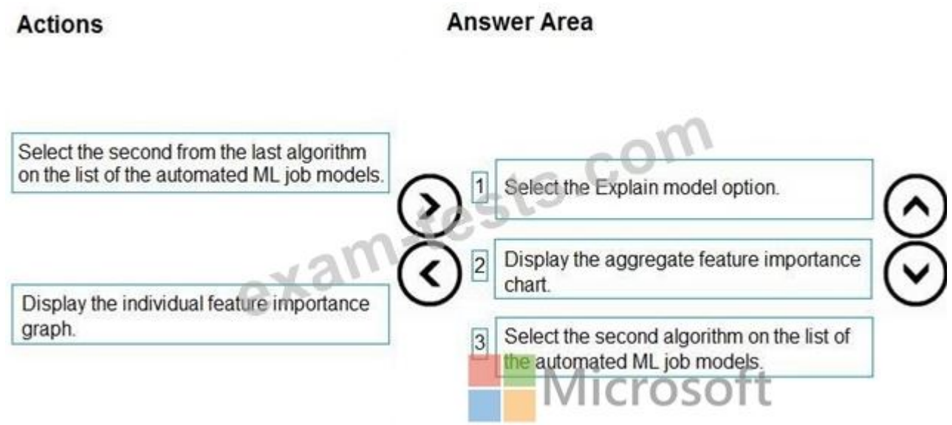
You must identify the feature that has the most influence on the predictions of the model for the second highest scoring algorithm. You must minimize the effort and time to identify the feature.

You need to complete the identification.

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.



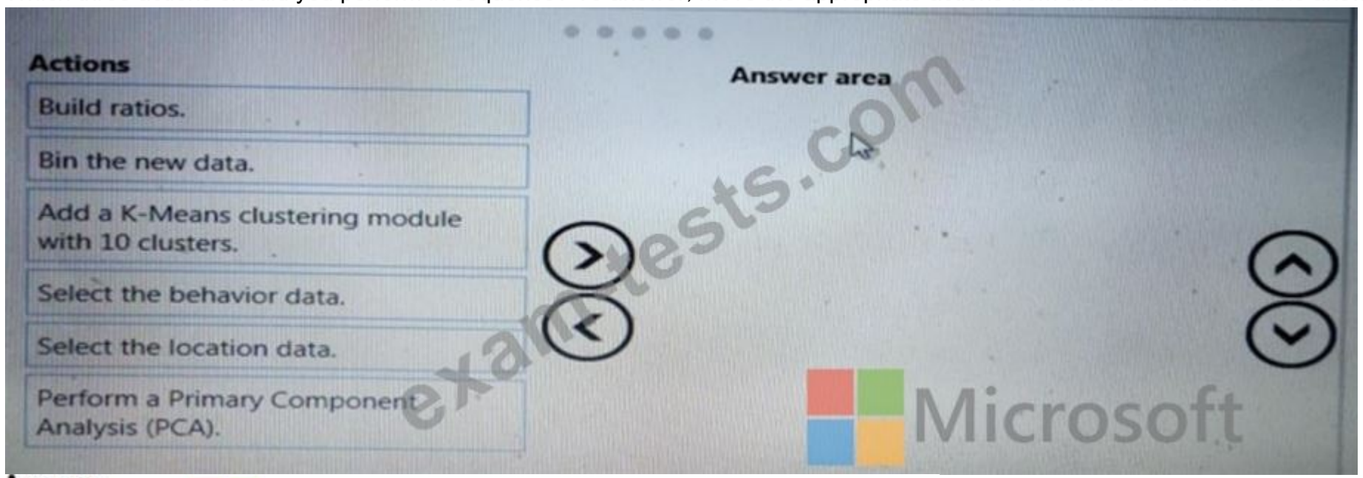
Answer:



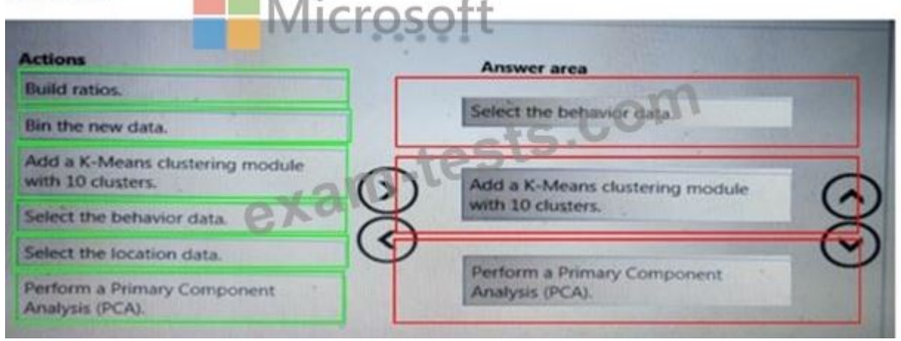
NEW QUESTION: 29

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.



Answer:



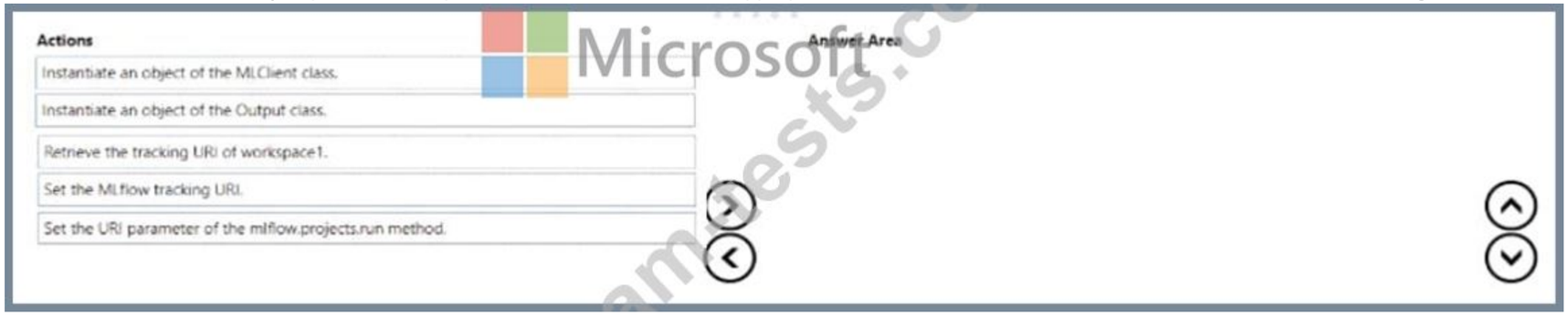
Answer:

Answer Area

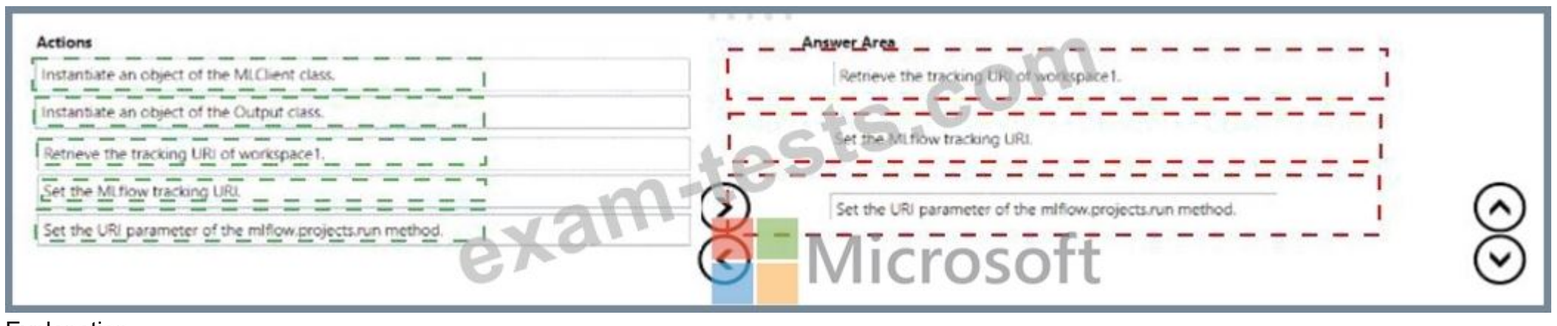
Select teh behavior data.
Add a K-Means clustering module with 10 clusters.
Perform a Primary Component Analysis (PCA).

1 - Select teh behavior data.
2 - Add a K-Means clustering module with 10 clusters.
3 - Perform a Primary Component Analysis (PCA).

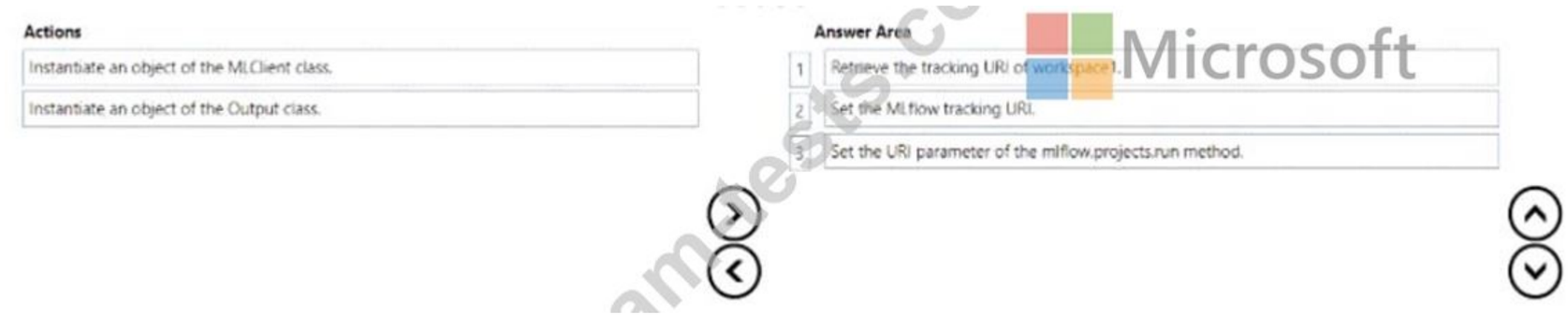
NEW QUESTION: 30
You manage an Azure Machine Learning workspace named workspace1 and a Data Science Virtual Machine (DSVM) named DSMV1.
You must an experiment in DSMV1 by using a Jupiter notebook and Python SDK v2 code. You must store metrics and artifacts in workspace 1 You start by creating Python SCK v2 code to import ail required packages.
You need to implement the Python SOK v2 code to store metrics and article workspace1.
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 the correctly order.



Answer:

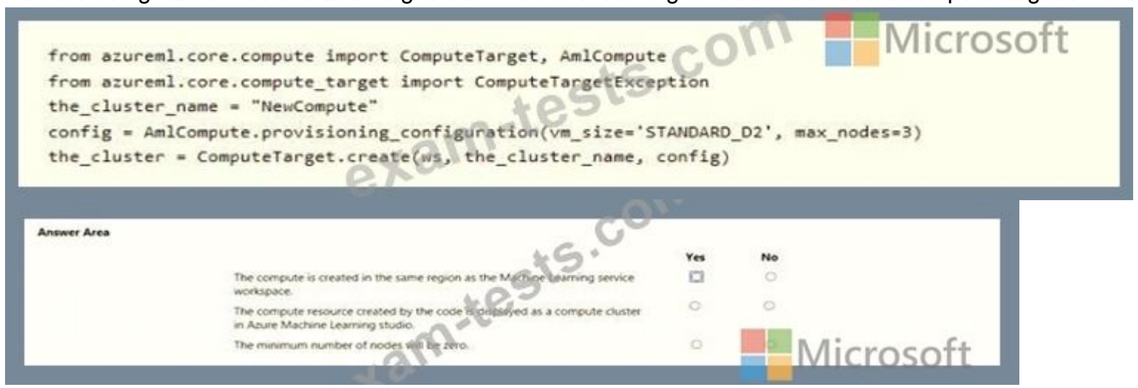


Explanation

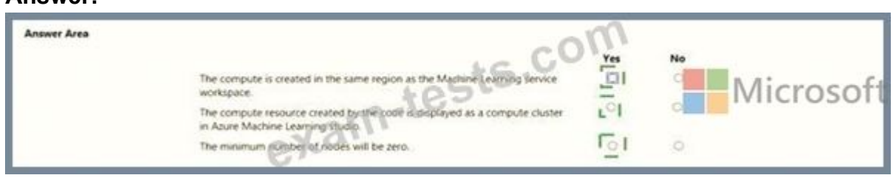


NEW QUESTION: 31

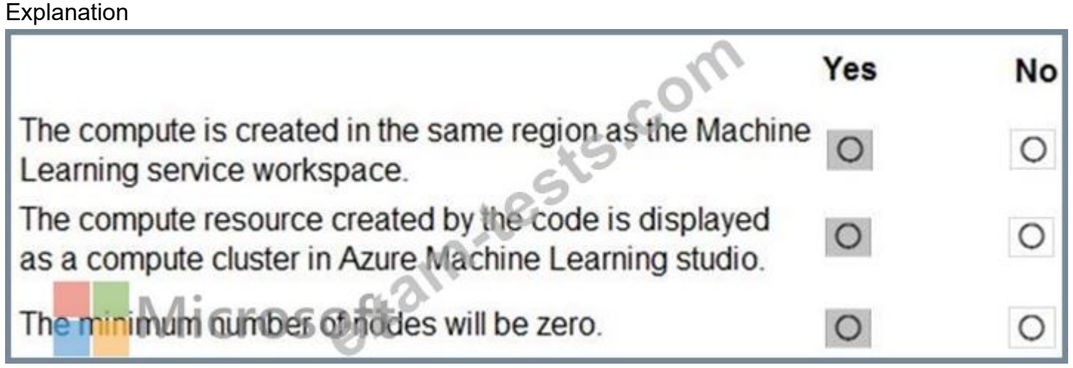
You are using Azure Machine Learning to train machine learning models. You need a compute target on which to remotely run the training script. You run the following Python code:



Answer:



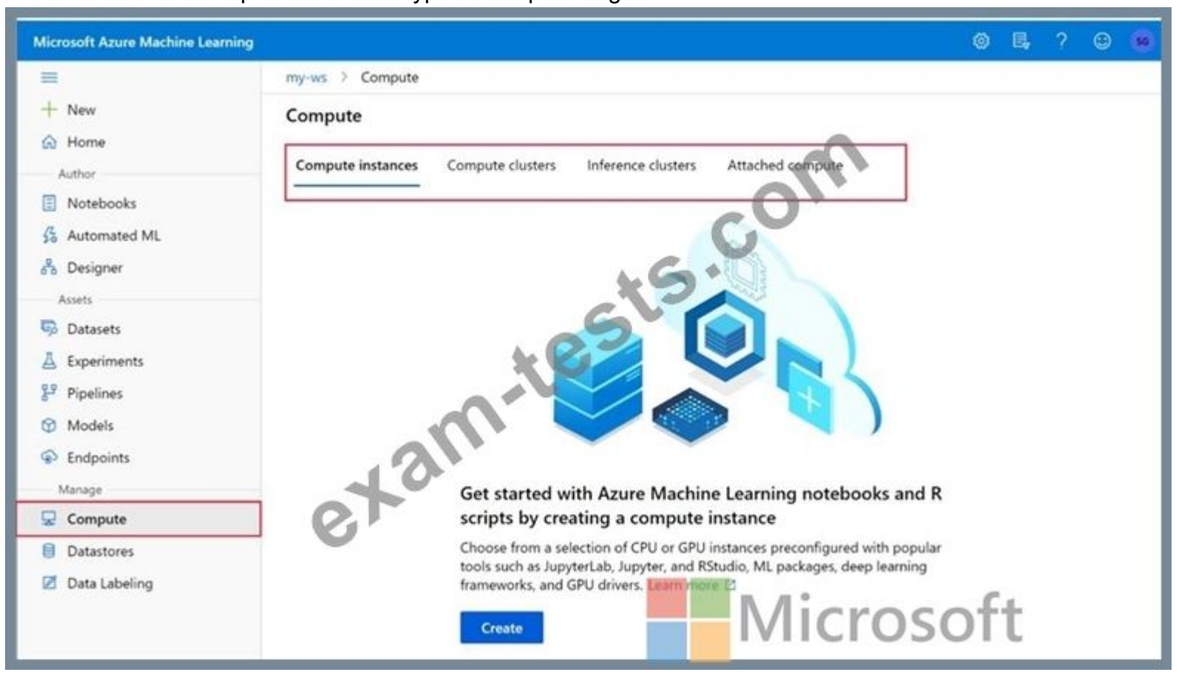
Explanation



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.amlcomputeprovis
https://docs.microsoft.com/en-us/azure/machine-learning/how-to-create-attach-compute-studio

NEW QUESTION: 32
Hotspot Question
You train a classification model by using a decision tree algorithm.

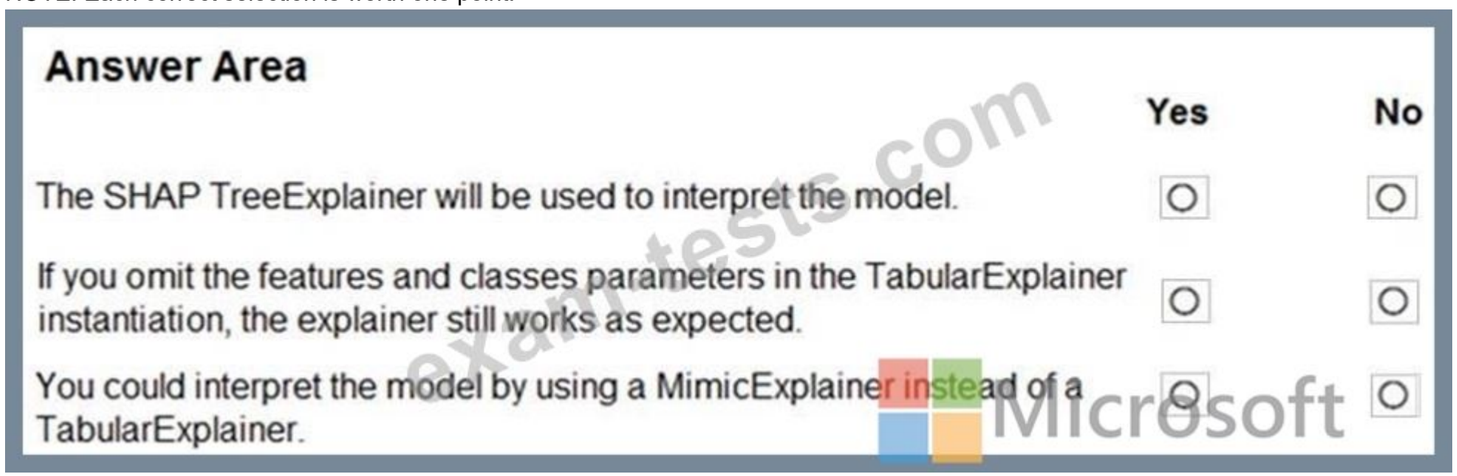
You create an estimator by running the following Python code. The variable feature_names is a list of all feature names, and class_names is a list of all class names.

```
from interpret.ext.blackbox import TabularExplainer
explainer = TabularExplainer(model,
                             x_train,
                             features=feature_names,
                             classes=classe_names)
```

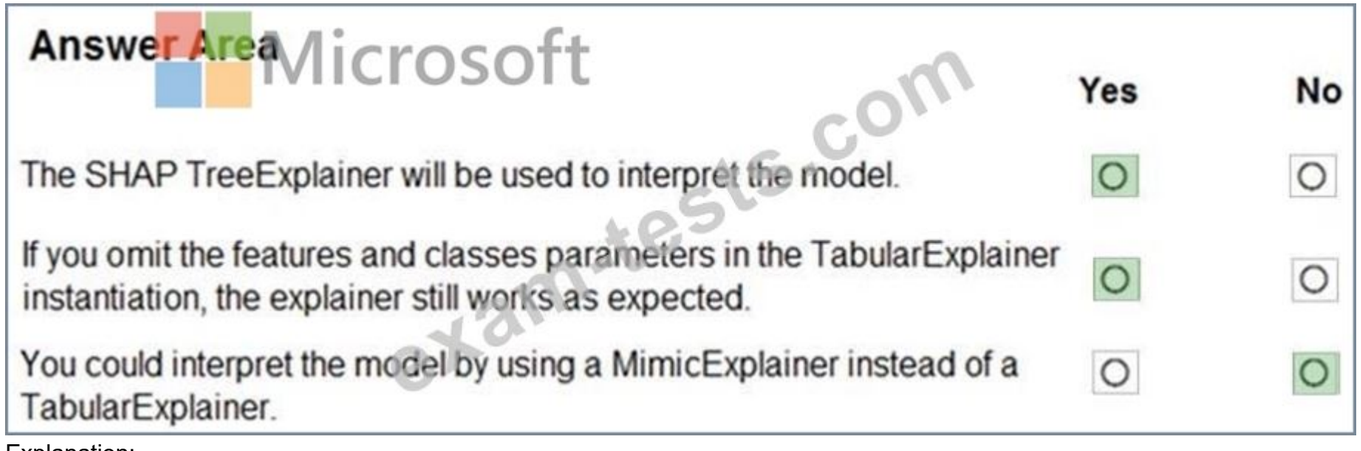
You need to explain the predictions made by the model for all classes by determining the importance of all features.

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:



Explanation:

Box 1: Yes

TabularExplainer calls one of the three SHAP explainers underneath (TreeExplainer, DeepExplainer, or KernelExplainer).

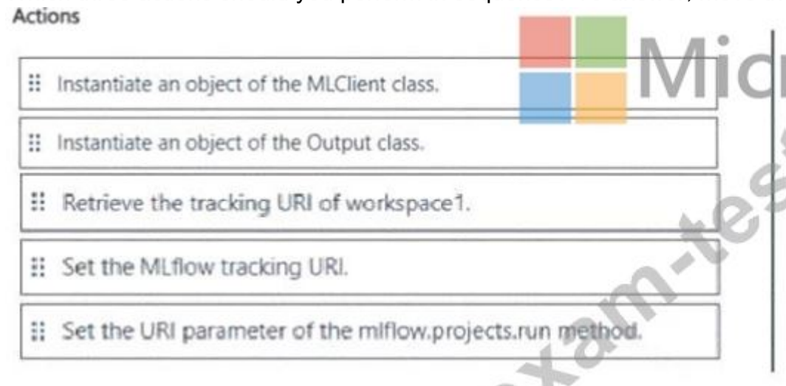
Box 2: Yes

To make your explanations and visualizations more informative, you can choose to pass in feature names and output class names if doing classification.

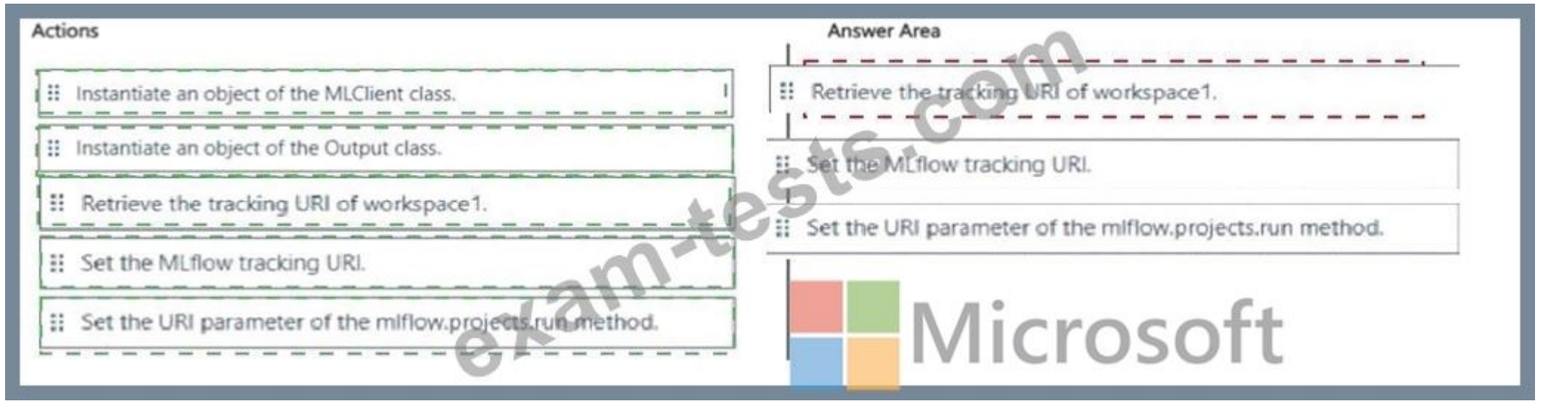
Box 3: No

TabularExplainer automatically selects the most appropriate one for your use case, but you can call each of its three underlying explainers underneath (TreeExplainer, DeepExplainer, or KernelExplainer) directly.
Reference:
https://docs.microsoft.com/en-us/azure/machine-learning/how-to-machine-learning- interpretability-aml

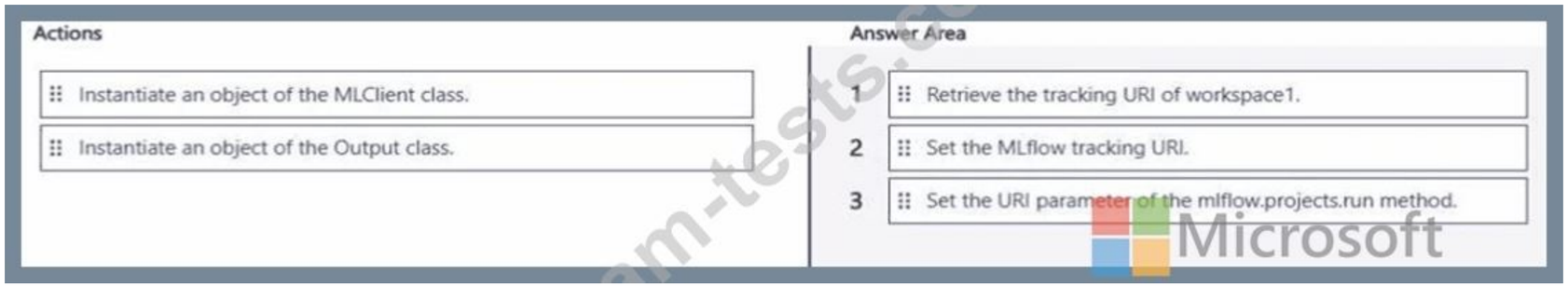
NEW QUESTION: 33
You manage an Azure Machine Learning workspace named workspace 1 and a Data Science Virtual Machine (DSVM) named DSMV1.
You must run an experiment on DSMV1 by using a Jupyter notebook and Python SDK v2 code. You must store metrics and artifacts in workspace1. You start by creating Python SDK v2 code to import all required packages.
You need to implement the Python SDK v2 code to store metrics and artifacts in workspace1.
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.



Answer:



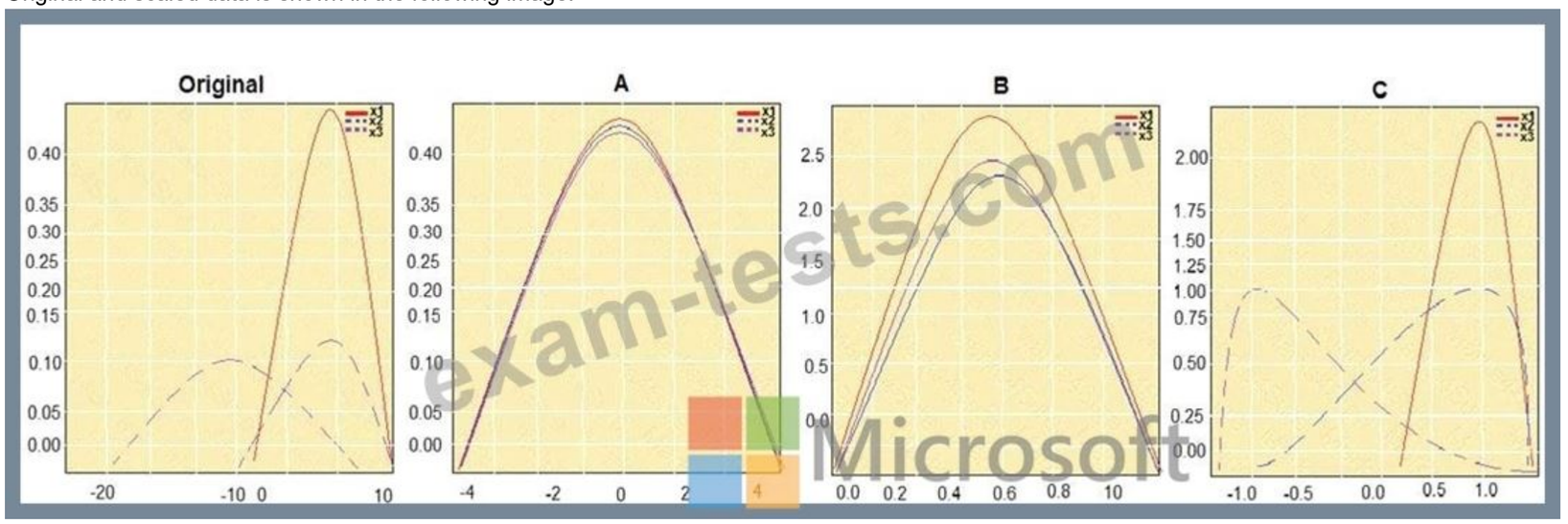
Explanation:



NEW QUESTION: 34

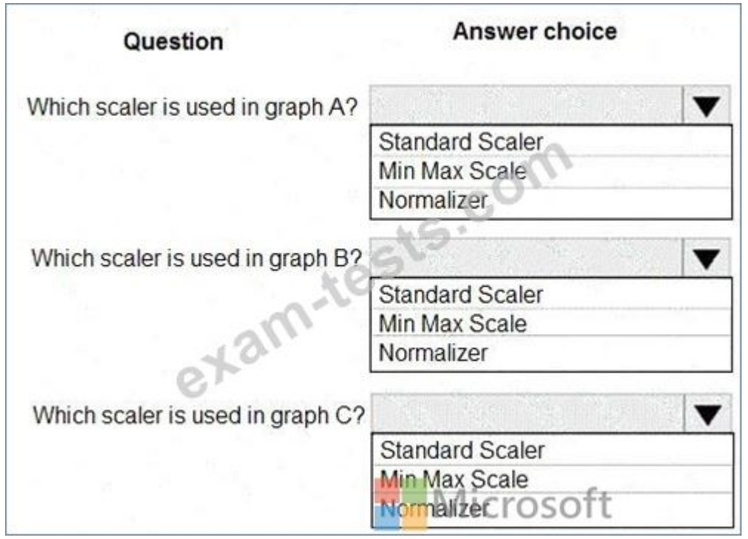
You are performing feature scaling by using the scikit-learn Python library for x.1 x2, and x3 features.

Original and scaled data is shown in the following image.

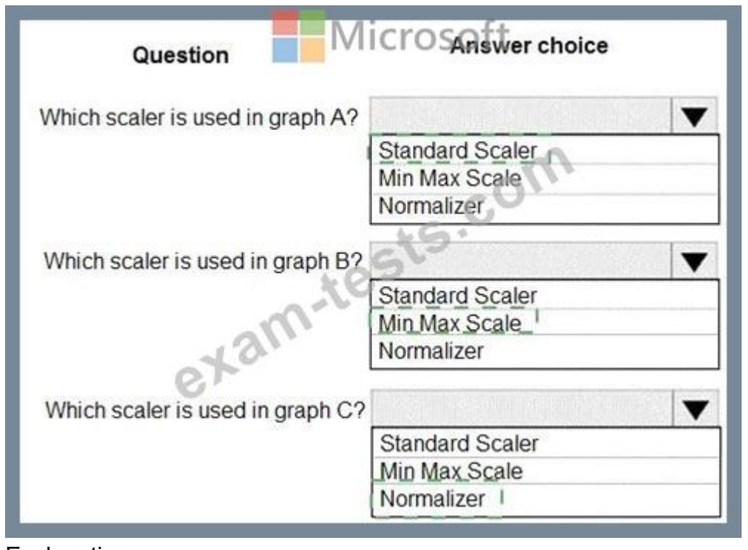


Use the drop-down menus to select the answer choice that answers each question based on the information presented in the graphic.

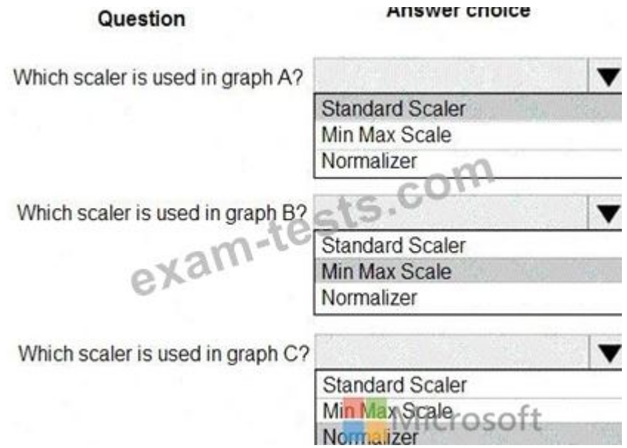
NOTE: Each correct selection is worth one point.



Answer:



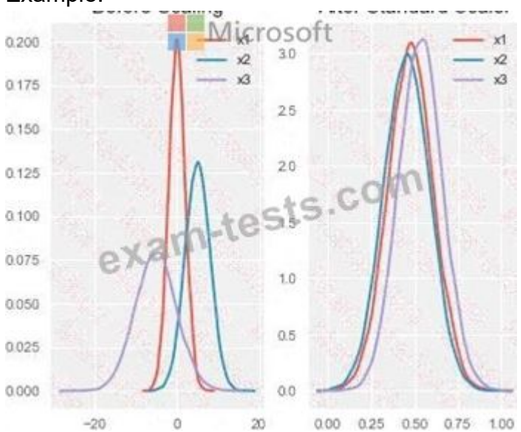
Explanation:



Box 1: StandardScaler

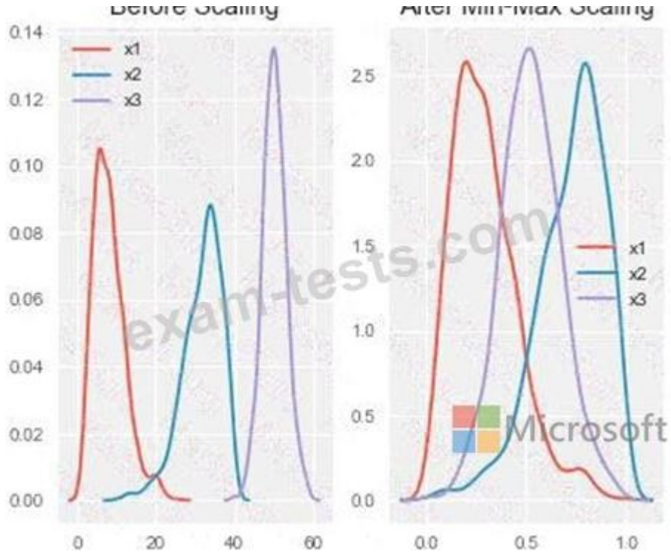
The StandardScaler assumes your data is normally distributed within each feature and will scale them such that the distribution is now centred around 0, with a standard deviation of 1.

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/

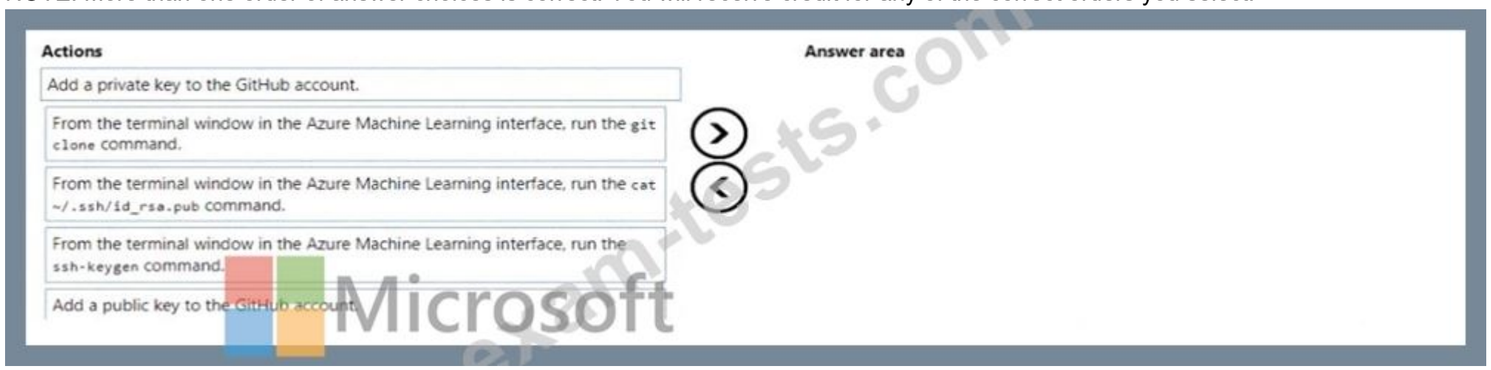
NEW QUESTION: 35

You have an existing GitHub repository containing Azure Machine Learning project files.

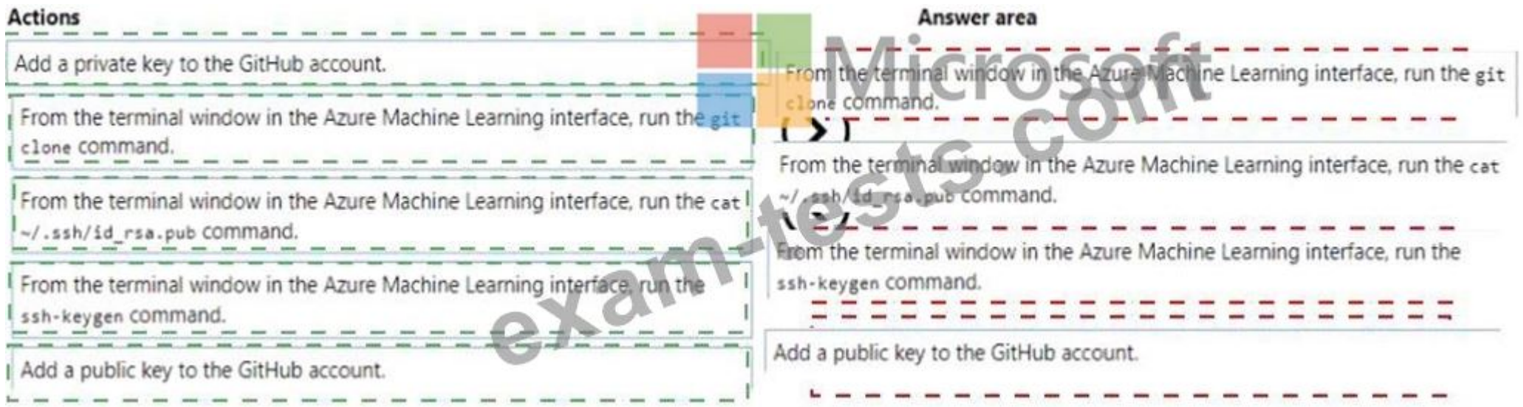
You need to clone the repository to your Azure Machine Learning shared workspace file system.

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.

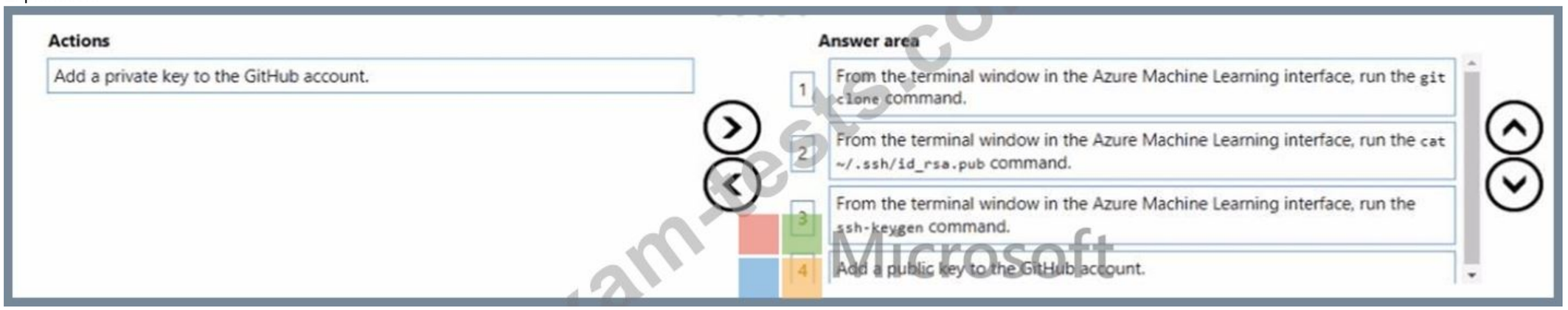
NOTE: More than one order of answer choices is correct. You will receive credit for any of the correct orders you select.



Answer:



Explanation:



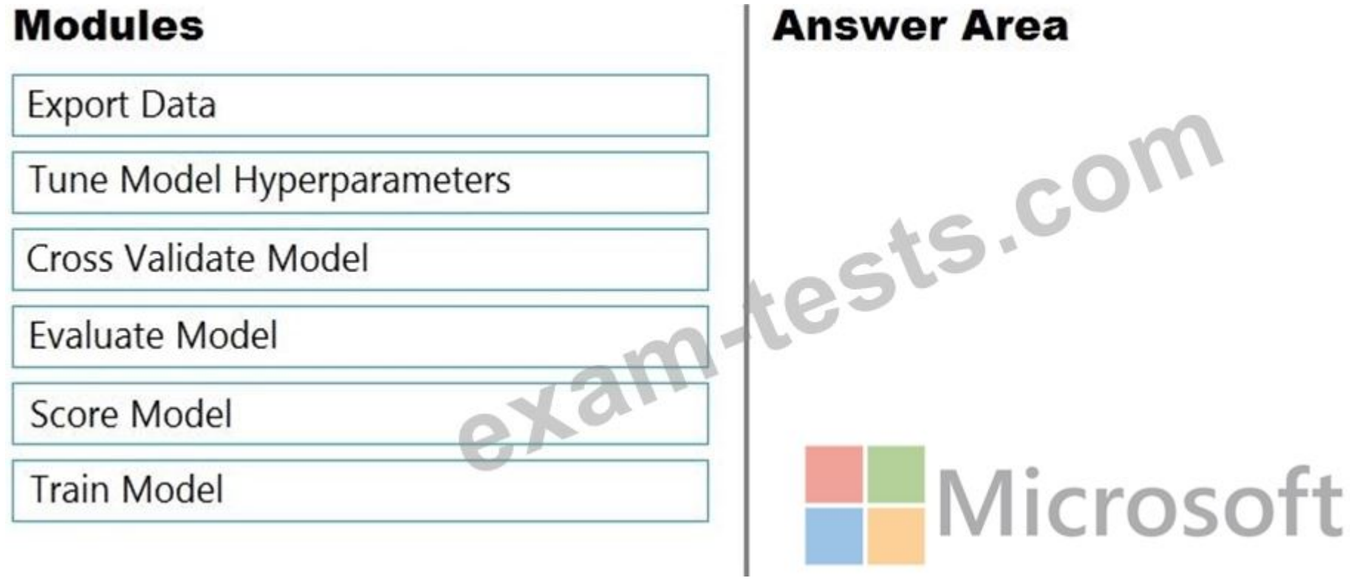
NEW QUESTION: 36

You are building an experiment using the Azure Machine Learning designer.

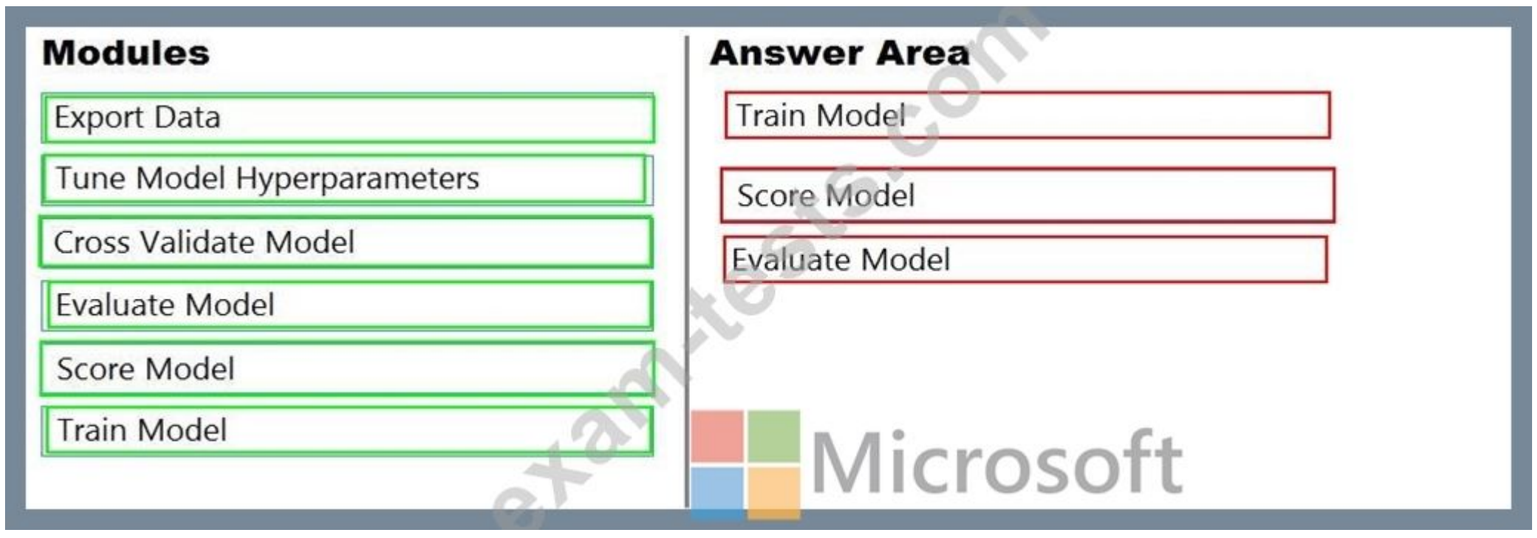
You split a dataset into training and testing sets. You select the Two-Class Boosted Decision Tree as the algorithm.

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.



Answer:



NEW QUESTION: 37

You run an automated machine learning experiment in an Azure Machine Learning workspace. Information about the run is listed in the table below:

Experiment	Run ID	Status	Created on	Duration
auto_ml_classification	AutoML_1234567890-123	Completed	11/11/2019 11:00:00 AM	00:27:11

You need to write a script that uses the Azure Machine Learning SDK to retrieve the best iteration of the experiment run. Which Python code segment should you use?

A.
```
from azureml.core import Workspace
from azureml.train.automl.run import AutoMLRun
ws = Workspace.from_config()
automl_ex = ws.experiments.get('auto_ml_classification')
automl_run = AutoMLRun(automl_ex, 'AutoML_1234567890-123')
best_iter = automl_run.get_output()[0]
```

B.
```
from azureml.core import Workspace
from azureml.train.automl.run import AutoMLRun
ws = Workspace.from_config()
automl_ex = ws.experiments.get('auto_ml_classification')
best_iter = list(automl_ex.get_runs())[0]
```

C.
```
from azureml.core import Workspace
from azureml.train.automl.run import AutoMLRun
ws = Workspace.from_config()
automl_ex = ws.experiments.get('auto_ml_classification')
best_iter = list(automl_ex.get_runs())[0]
```

D.
```
from azureml.core import Workspace
```

Answer: A (LEAVE A REPLY)

The get_output method on automl_classifier returns the best run and the fitted model for the last invocation.

Overloads on get_output allow you to retrieve the best run and fitted model for any logged metric or for a particular iteration.

In [ ]:

best_run, fitted_model = local_run.get_output()

Reference:

https://notebooks.azure.com/azureml/projects/azureml-getting-started/html/how-to-use-azureml/automated- machine-learning/classification-with-deployment/auto-ml-classification-with-deployment.ipynb

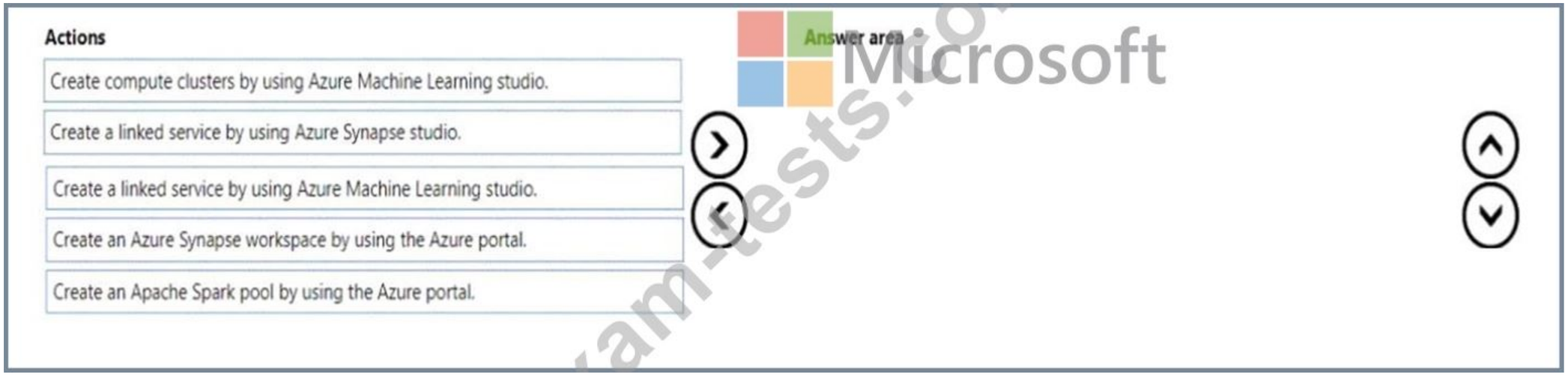
NEW QUESTION: 38

You create an Azure Machine Learning workspace.

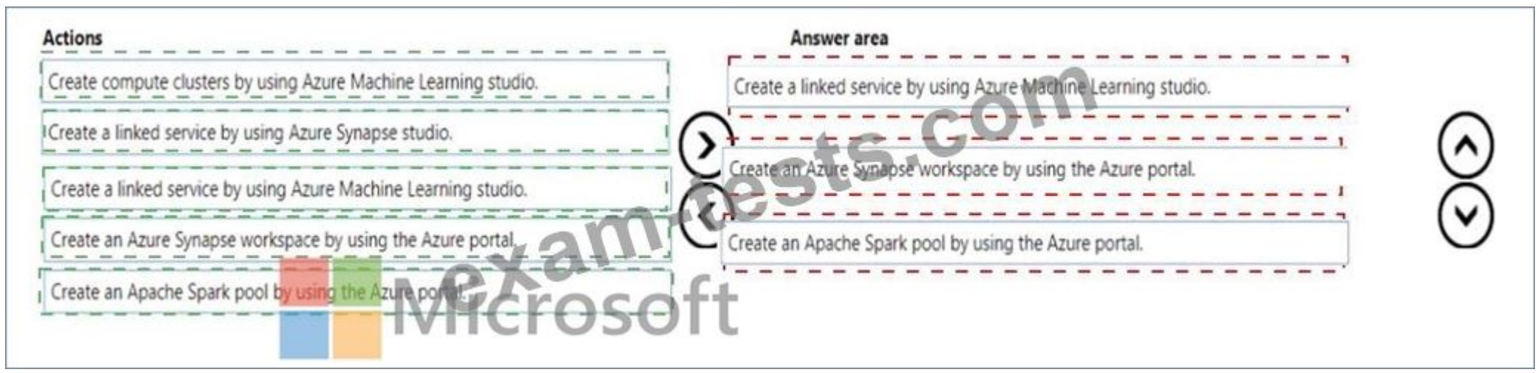
You must implement dedicated compute for model training in the workspace by using Azure Synapse compute resources. The solution must attach the dedicated compute and start an Azure Synapse session.

You need to implement the compute resources.

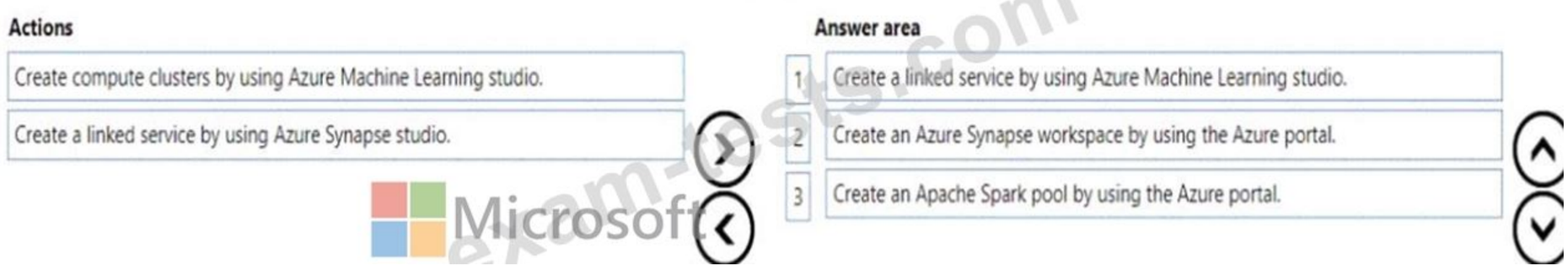
Which three actions should you perform in sequence? To answer, move the appropriate actions from the list of actions to the answer area and arrange them in the correct order.



Answer:



Explanation



NEW QUESTION: 39

You plan to use automated machine learning by using Azure Machine Learning Python SDK v2 to train a regression model. You have data that has features with missing values, and categorical features with few distinct values.

You need to control whether automated machine learning automatically imputes missing values and encode categorical features as part of the training task.

Which enum of the automl package should you use?

A. ForecastHorizonMode
B. RegressionModels
C. FeaturizationMode
D. RegressionPrimaryMetrics
Answer: C (LEAVE A REPLY)
Constructor:
FeaturizationMode(value, names=None, *, module=None, qualname=None, type=None, start=1, boundary=None) Fields:
AUTO, CUSTOM, OFF
https://learn.microsoft.com/en-us/python/api/azure-ai-ml/azure.ai.ml.automl.featurizationmode?view=azure-python

NEW QUESTION: 40
You manage an Azure AI Foundry project in your subscription. You deploy a gpt-4o model. You must test the model before you use it in an existing front-end application. You need to adjust the parameters to get more creative responses.
Solution: Increase Temperature.
Does the solution meet the goal?
A. Yes
B. No
Answer: (SHOW ANSWER)

NEW QUESTION: 41
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.

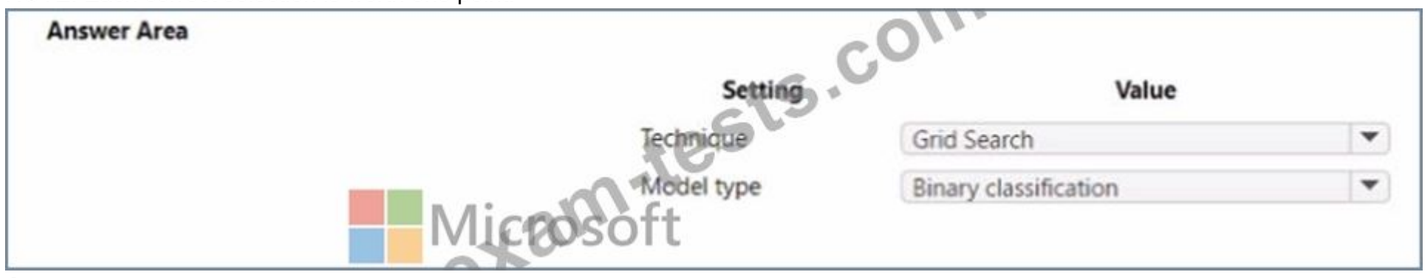
Environment name	Compute type
nb_server	Compute Instance
aks_cluster	Azure Kubernetes Service
mlc_cluster	Machine Learning Compute

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.

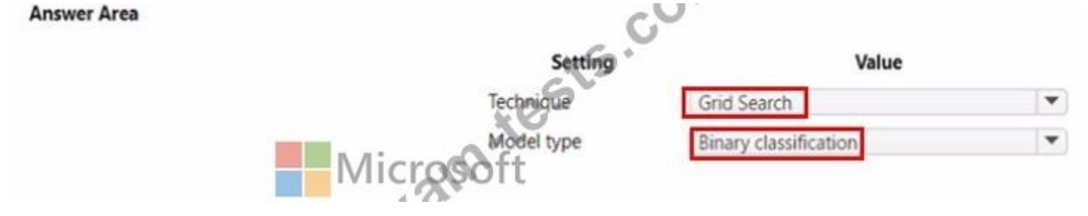
Answer:

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: 42
You have machine learning models produce unfair predictions across sensitive features.
You must use a post-processing technique to apply a constraint to the models to mitigate their unfairness.
You need to select a post-processing technique and model type.
What should you use? To answer, select the appropriate options in the answer area.
NOTE: Each correct selection is worth one point.



Answer:



NEW QUESTION: 43
You are running a training experiment on remote compute in Azure Machine Learning.
The experiment is configured to use a conda environment that includes the mlflow and azureml-contrib-run packages.
You must use MLflow as the logging package for tracking metrics generated in the experiment.
You need to complete the script for the experiment.
How should you complete the code? To answer, select the appropriate options in the answer area.
NOTE: Each correct selection is worth one point.

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

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

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

- run = Run.get_context()
- mlflow.start_run()
- logger = logging.getLogger('Run')

```
reg_rate = 0.01
# Log the reg_rate metric
```

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

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

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

Answer:

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

- from azureml.core import Run
- **import mlflow**
- import logging

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

- run = Run.get_context()
- **mlflow.start_run()**
- logger = logging.getLogger('Run')

```
reg_rate = 0.01
# Log the reg_rate metric
```

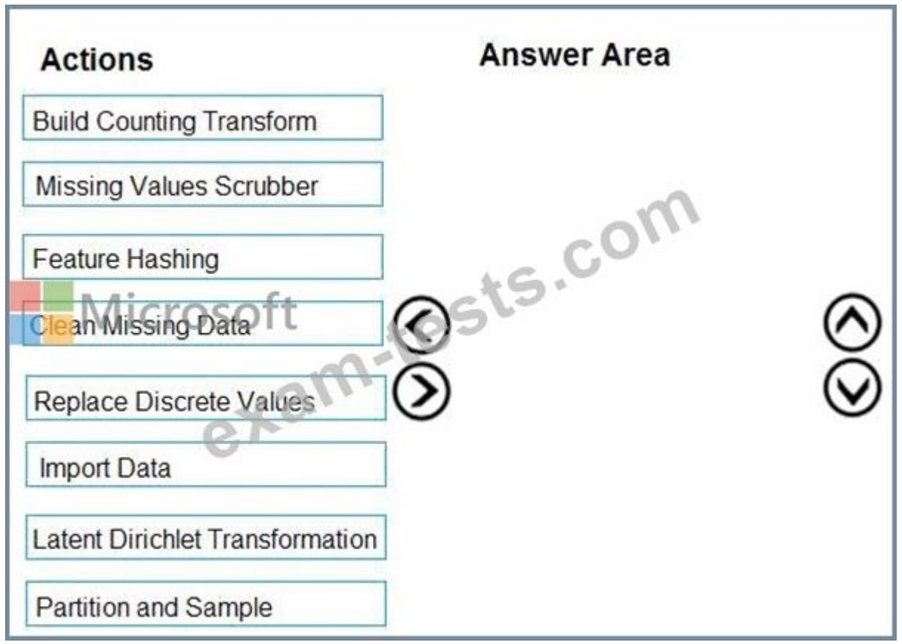
- run.log('reg_rate', np.float(reg_rate))
- **mlflow.log_metric('reg_rate', np.float(reg_rate))**
- logger.info(np.float(reg_rate))

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

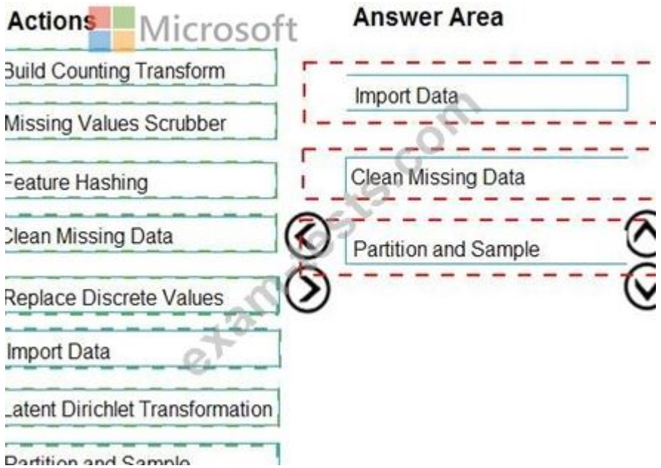
- run.complete()
- **mlflow.end_run()**
- logger.setLevel(logging.INFO)

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

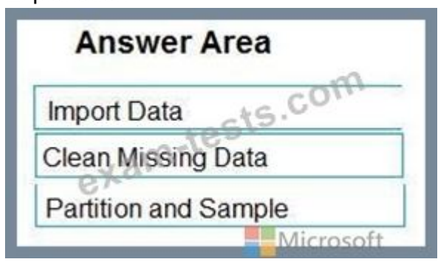
NEW QUESTION: 44
You are creating an experiment by using Azure Machine Learning Studio.
You must divide the data into four subsets for evaluation. There is a high degree of missing values in the data.
You must prepare the data for analysis.
You need to select appropriate methods for producing the experiment.
Which three modules should you run in sequence? To answer, move the appropriate actions from the list of actions to the answer area and arrange them in the correct order.
NOTE: More than one order of answer choices is correct. You will receive credit for any of the correct orders you select.



Answer:



Explanation



The Clean Missing Data module in Azure Machine Learning Studio, to remove, replace, or infer missing values.

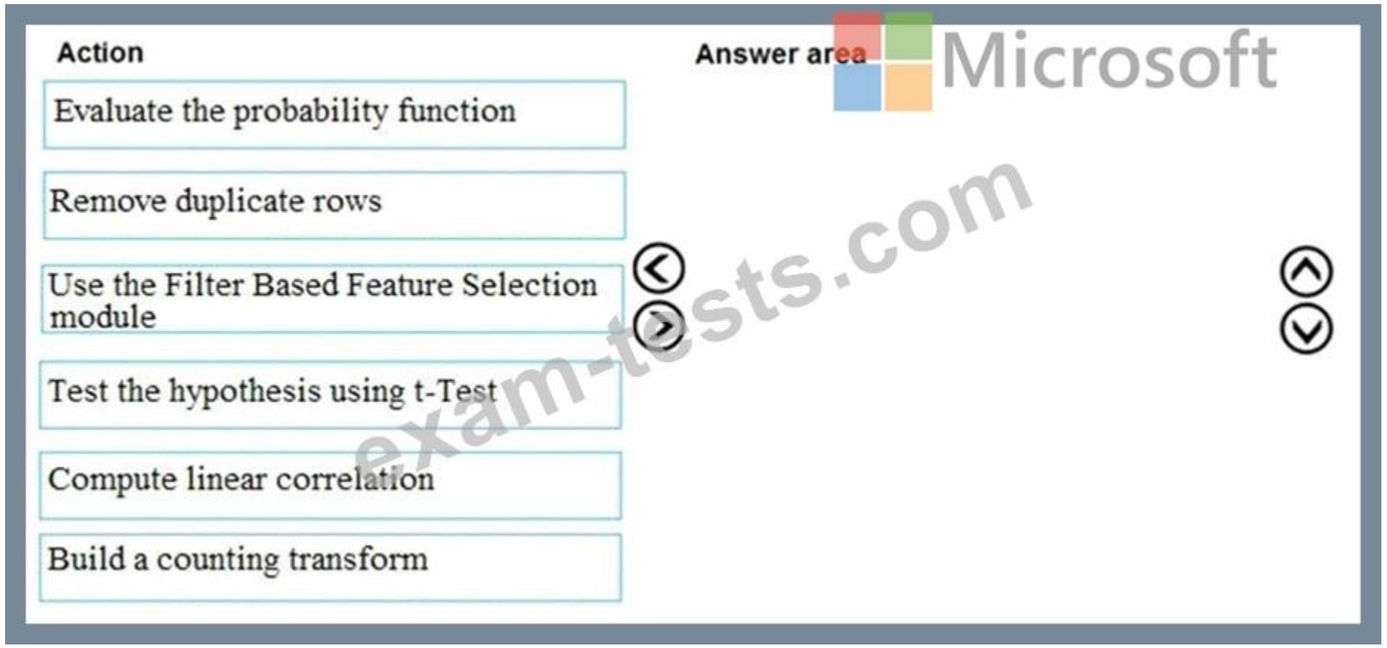
NEW QUESTION: 45

You are producing a multiple linear regression model in Azure Machine Learning Studio.

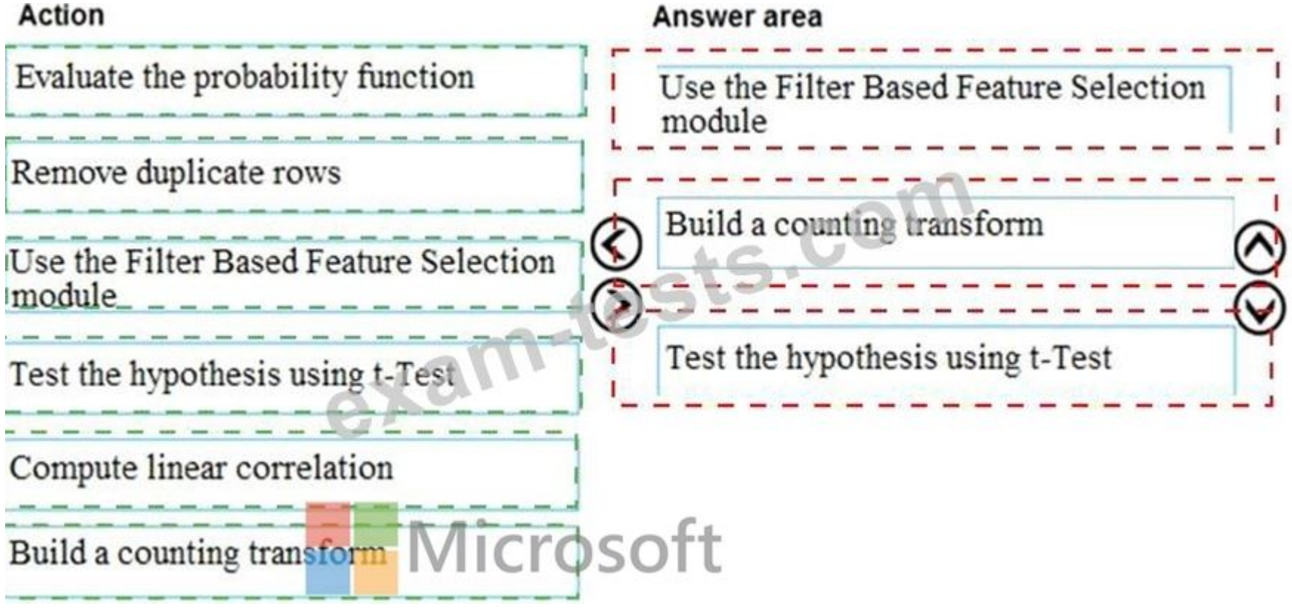
Several independent variables are highly correlated.

You need to select appropriate methods for conducting effective feature engineering on all the data.

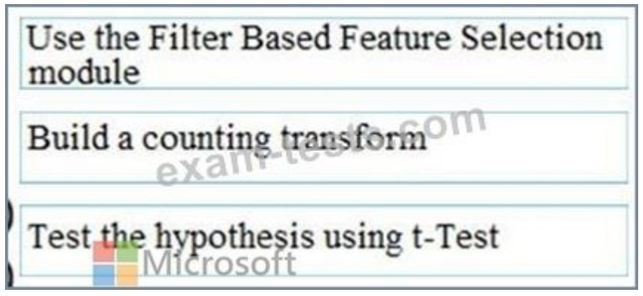
Which three actions should you perform in sequence? To answer, move the appropriate actions from the list of actions to the answer area and arrange them in the correct order.



Answer:



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: 46

You create an Azure Machine Learning workspace and a dataset. The dataset includes age values for a large group of diabetes patients. You use the dp.mean function from the SmartNoise library to calculate the mean of the age value. You store the value in a variable named age.mean.

You must output the value of the interval range of released mean values that will be returned 95 percent of the time.

You need to complete the code.

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

Answer:

NEW QUESTION: 47

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 dat a. 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 json, os
from sklearn.metrics import roc_auc_score
# code to train model omitted
auc = roc_auc_score(y_test, y_predicted)
os.makedirs("outputs", exist_ok = True)
with open("outputs/AUC.txt", "w") as file_cur:
   file_cur.write(auc)
```

Does the solution meet the goal?

A. Yes

B. No

Answer: B (LEAVE A REPLY)

Use a solution with logging.info(message) instead.

Note: Python printing/logging example:

logging.info(message)

Destination: Driver logs, Azure Machine Learning designer

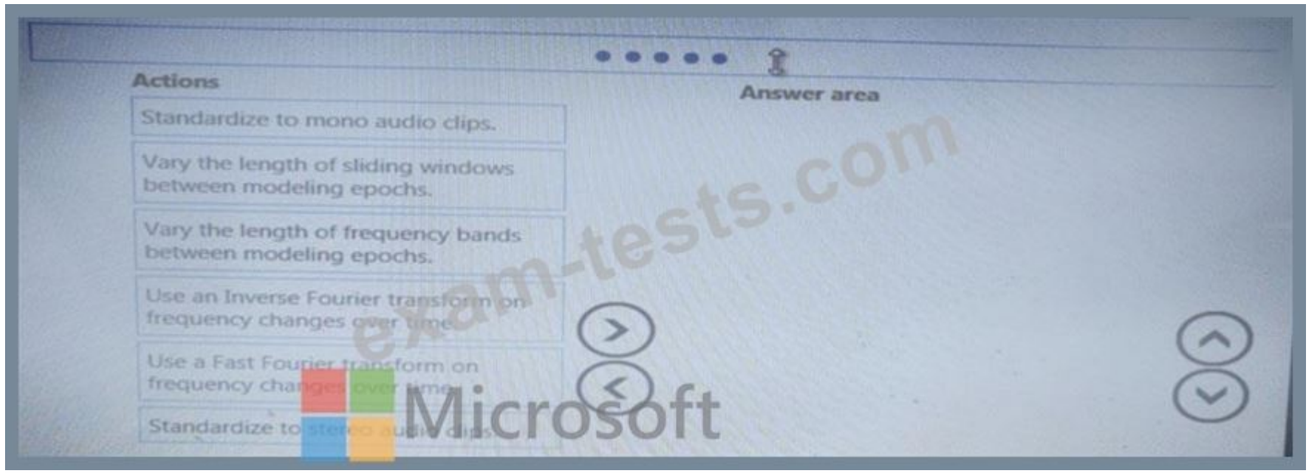
Reference:

https://docs.microsoft.com/en-us/azure/machine-learning/how-to-debug-pipelines

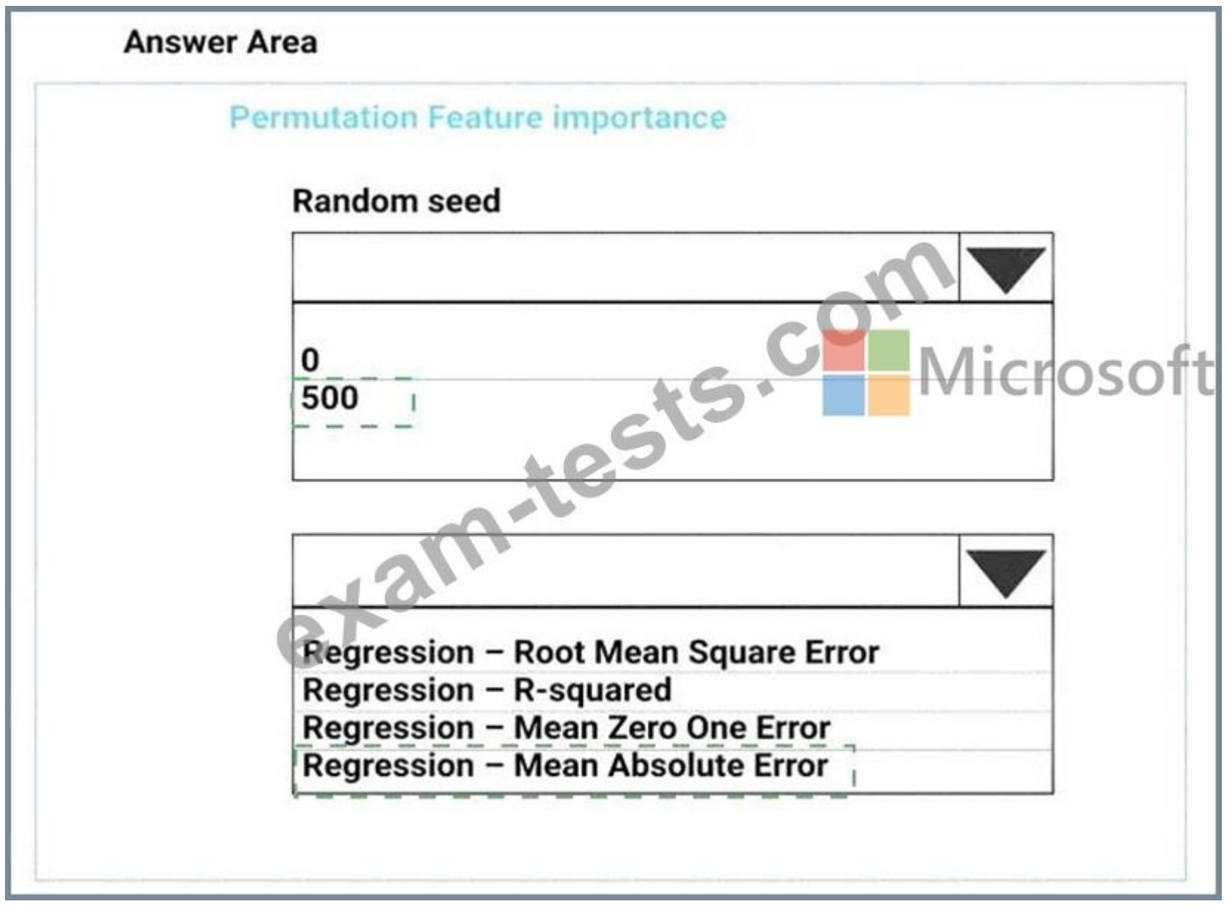
NEW QUESTION: 48

You need to define a process for penalty event detection.

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


Actions
Standardize to mono audio clips.
Vary the length of sliding windows between modeling epochs.
Vary the length of frequency bands between modeling epochs.
Use an Inverse Fourier transform on frequency changes over time.
Use a Fast Fourier transform on frequency changes over time.
Standardize to stereo audio clips.
Answer area
exam-tests.com
Microsoft

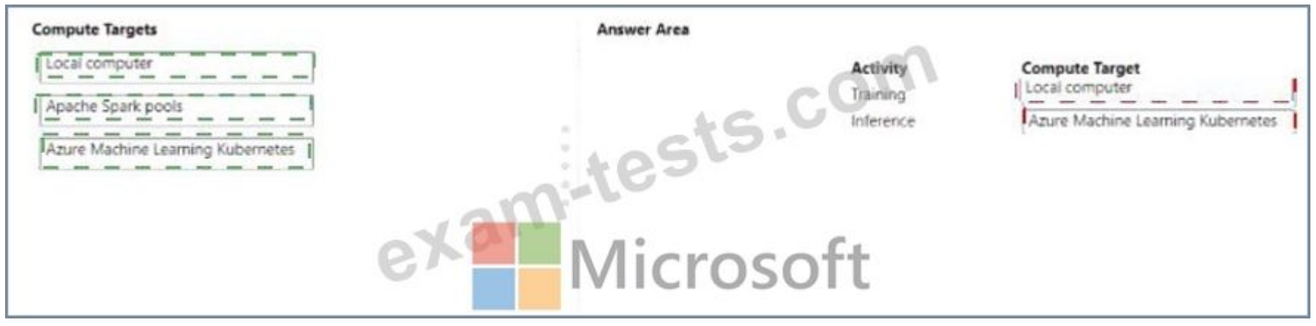
Answer:


Answer Area
Permutation Feature importance
Random seed
0
500
Regression – Root Mean Square Error
Regression – R-squared
Regression – Mean Zero One Error
Regression – Mean Absolute Error
exam-tests.com
Microsoft

NEW QUESTION: 49
You create an MLflow model
You must deploy the model to Azure Machine Learning for batch inference.
You need to create the batch deployment.
Which two components should you use? Each correct answer presents a complete solution.
NOTE: Each correct selection is worth one point
A. Environment
B. Model files
C. Kubernetes online endpoint
D. Online endpoint
E. Compute target
Answer: (SHOW ANSWER)

NEW QUESTION: 50
You are designing an Azure Machine Leaning solution by using the Python SDK v2.
You must train and deploy the solution by using a compute target. The compute target must meet the following requirements:
* Enable the use of on-premises compute resources.
* Support autoscalling.
You need to configure a compute target for training and inference.
Which compute target t should you configure?
To answer select the appropriate options in the answer area.
NOTE: Each correct selection is worth one point.

Answer:



Explanation

NEW QUESTION: 51

You have an Azure Machine Learning workspace.

You plan to set up logging and tracking experiments by using MLflow Tracking.

You need to log the accuracy as a numerical value and the training loss as a plot.

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

NOTE: Each correct selection is worth one point.

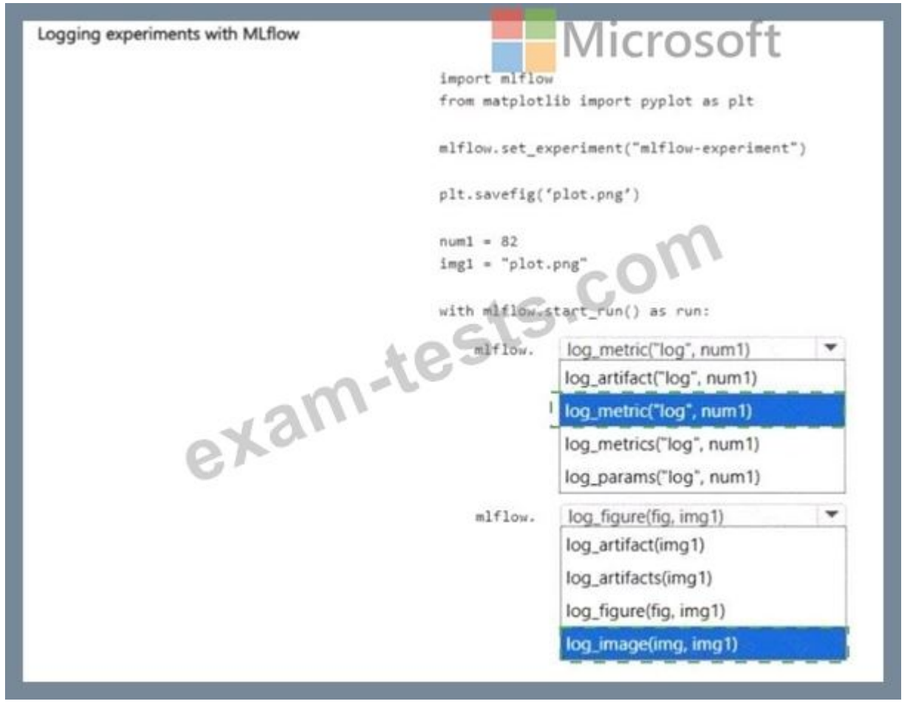
Logging experiments with MLflow

Microsoft

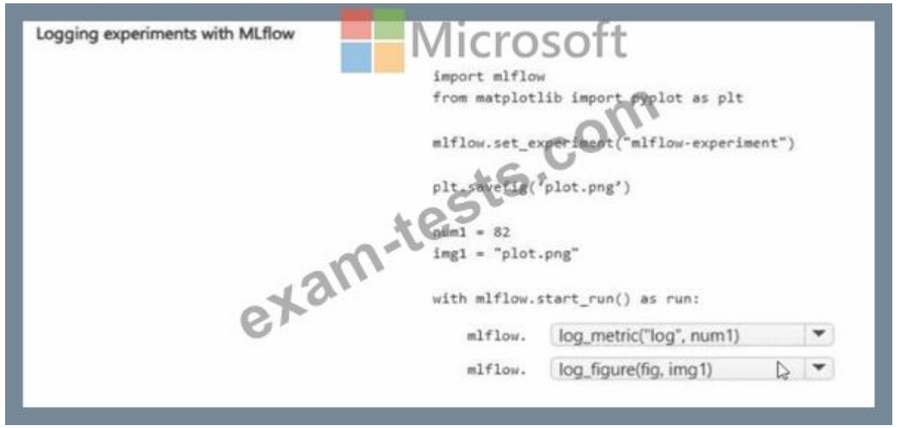
```
import mlflow
from matplotlib import pyplot as plt

mlflow.set_experiment("mlflow-experiment")

plt.savefig('plot.png')

num1 = 82
img1 = "plot.png"

with mlflow.start_run() as run:
    mlflow.[log_metric("log", num1) ▼]
    mlflow.[log_figure(fig, img1) ▼]
```

- log_metric("log", num1)
  - log_artifact("log", num1)
  - log_metric("log", num1)
  - log_metrics("log", num1)
  - log_params("log", num1)
- log_figure(fig, img1)
  - log_artifact(img1)
  - log_artifacts(img1)
  - log_figure(fig, img1)
  - log_image(img, img1)

Answer:



Explanation:



NEW QUESTION: 52

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

You must use CUDA-based model training.

You need to provision the Compute Instance.

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

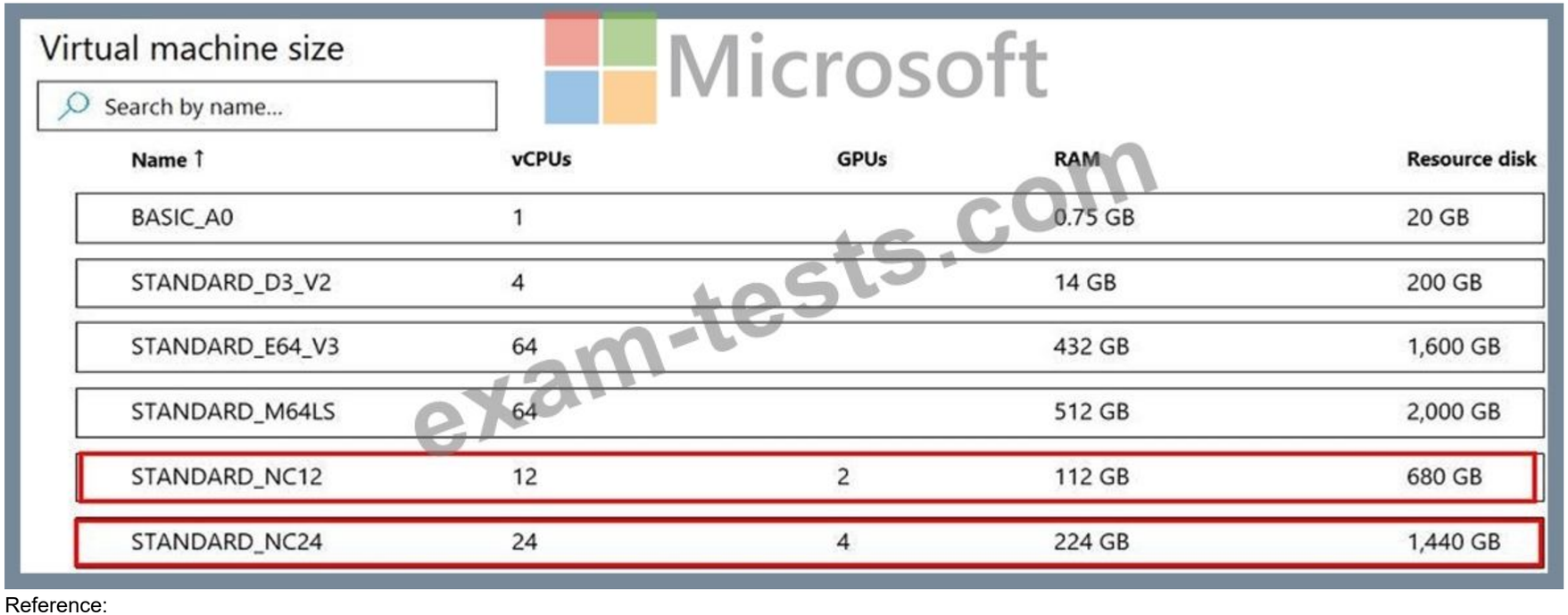
NOTE: Each correct selection is worth one point.

Virtual machine size

Search by name...

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

Answer:



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

NEW QUESTION: 53

You are building an intelligent solution using machine learning models.

The environment must support the following requirements:

Data scientists must build notebooks in a cloud environment

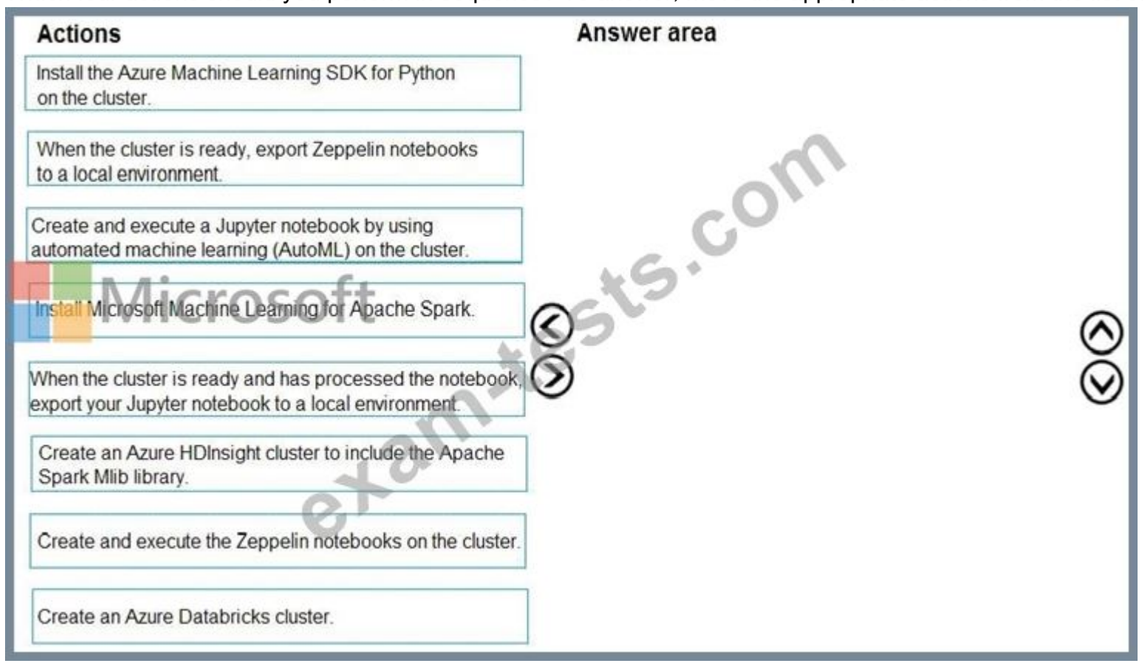
Data scientists must use automatic feature engineering and model building in machine learning pipelines.

Notebooks must be deployed to retrain using Spark instances with dynamic worker allocation.

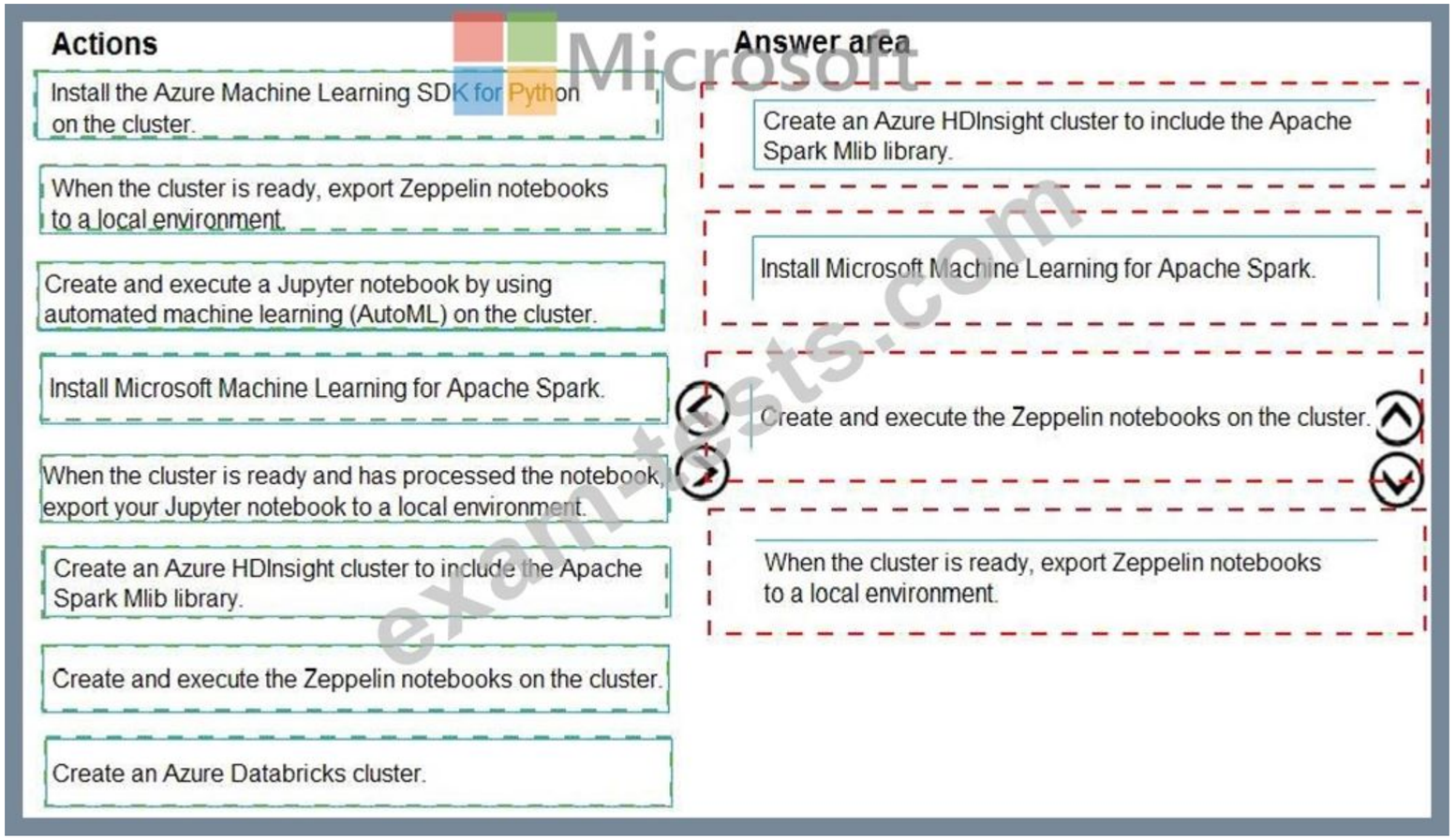
Notebooks must be exportable to be version controlled locally.

You need to create the environment.

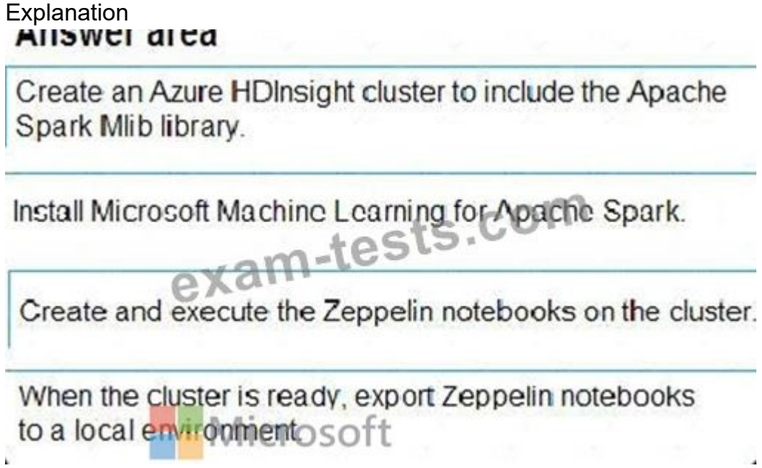
Which four actions should you perform in sequence? To answer, move the appropriate actions from the list of actions to the answer area and arrange them in the correct order.



Answer:



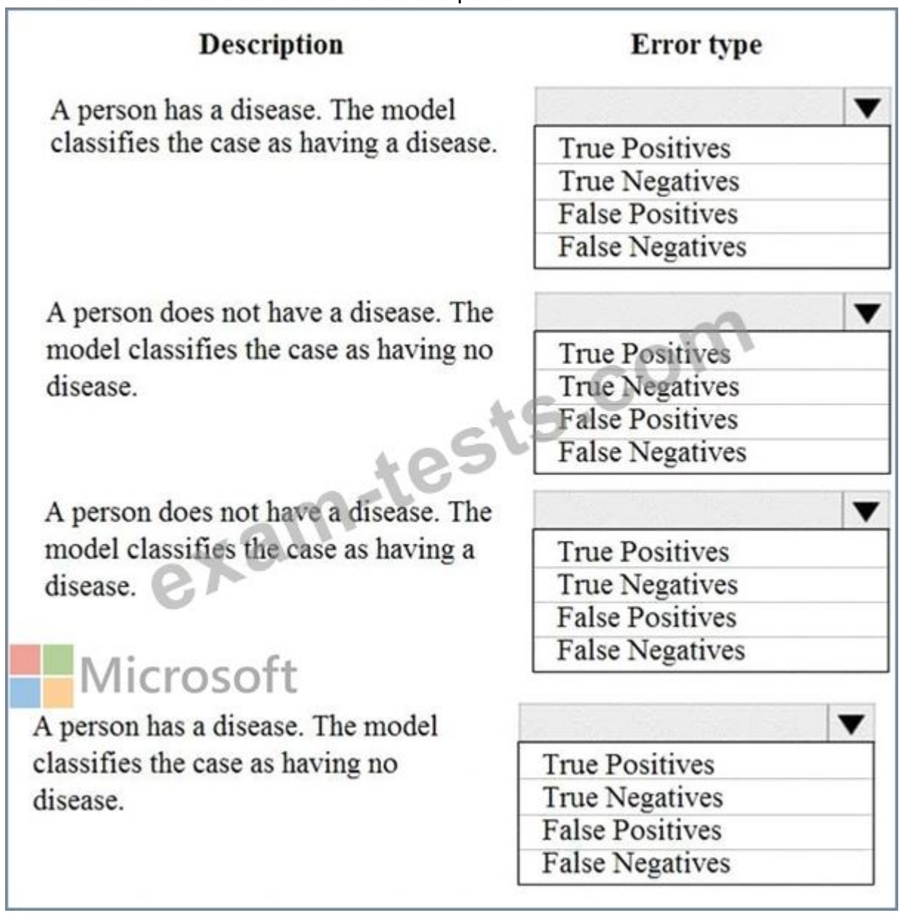
Explanation



Step 1: Create an Azure HDInsight cluster to include the Apache Spark Mlib library Step 2: Install Microsot Machine Learning for Apache Spark You install AzureML on your Azure HDInsight cluster.

Microsoft Machine Learning for Apache Spark (MMLSpark) provides a number of deep learning and data science tools for Apache Spark, including seamless integration of Spark Machine Learning pipelines with Microsoft Cognitive Toolkit (CNTK) and OpenCV, enabling you to quickly create powerful, highly-scalable predictive and analytical models for large image and text datasets.
Step 3: Create and execute the Zeppelin notebooks on the cluster
Step 4: When the cluster is ready, export Zeppelin notebooks to a local environment.
Notebooks must be exportable to be version controlled locally.
References:
https://docs.microsoft.com/en-us/azure/hdinsight/spark/apache-spark-zeppelin-notebook
https://azuremlbuild.blob.core.windows.net/pysparkapi/intro.html

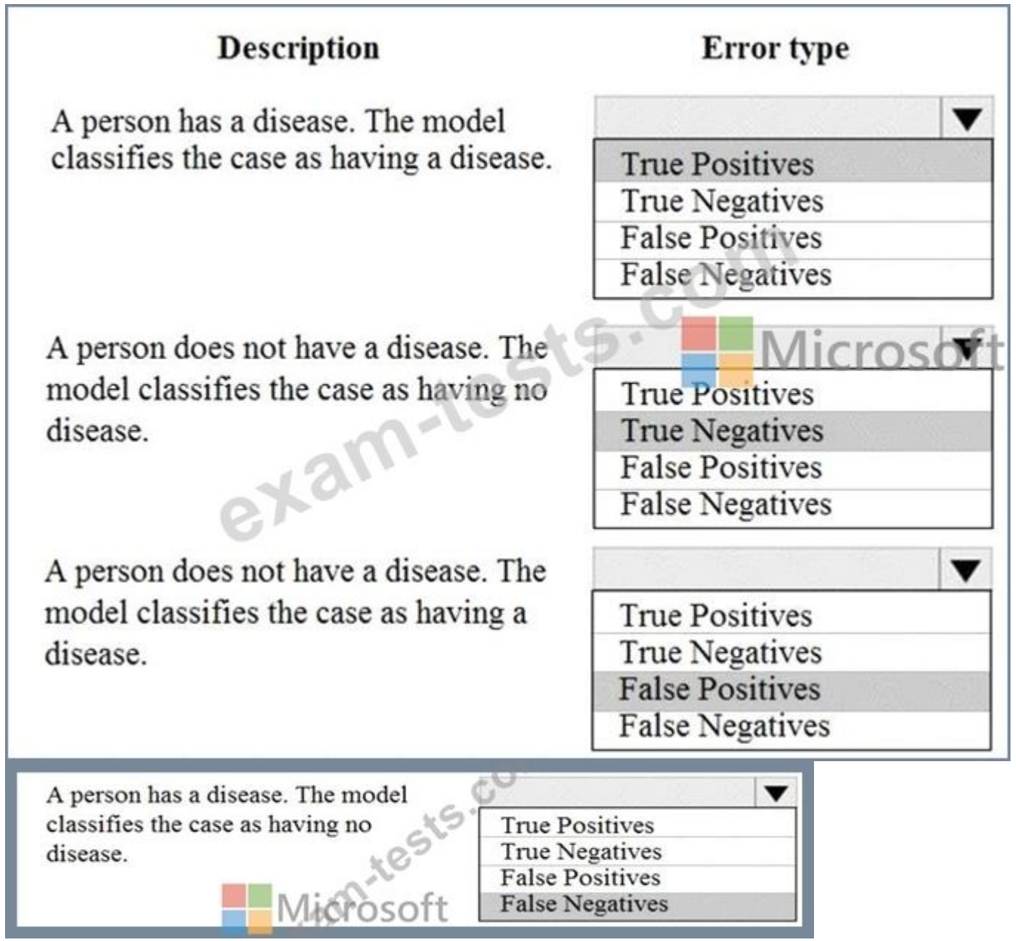
NEW QUESTION: 54
You create a binary classification model to predict whether a person has a disease.
You need to detect possible classification errors.
Which error type should you choose for each description? To answer, select the appropriate options in the answer area.
NOTE: Each correct selection is worth one point.



Answer:

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

Explanation



Box 1: True Positive

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

Box 3: False Positive

A false positive is an outcome where the model incorrectly predicts the positive class.

Box 4: False Negative

A false negative is an outcome where the model incorrectly predicts the negative class.

Note: Let's make the following definitions:

"Wolf" is a positive class.

"No wolf" is a negative class.

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

Reference:

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

NEW QUESTION: 55

You manage an Azure Machine Learning workspace named workspace1 by using the Python SDK v2.
The default datastore of workspace1 contains a folder named sample_dat
a. The folder structure contains the following content:

```
├── sample_data
    ├── MLTable
    ├── file1.txt
    ├── file2.txt
    ├── file3.txt
```

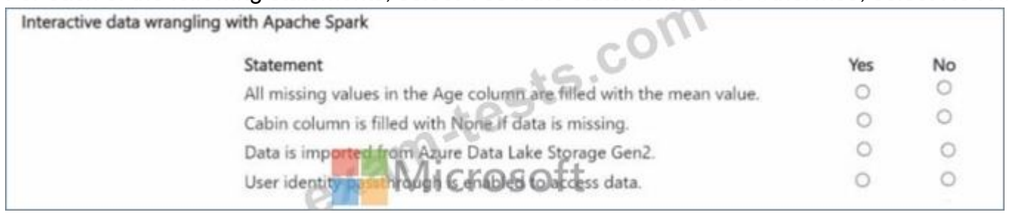
You write Python SDK v2 code to materialize the data from the files in the sample.data folder into a Pandas data frame. You need to complete the Python SDK v2 code to use the MLTaWe folder as the materialization blueprint. How should you complete the code? To answer, select the appropriate options in the answer area.
NOTE: Each correct selection is worth one point.

Answer:

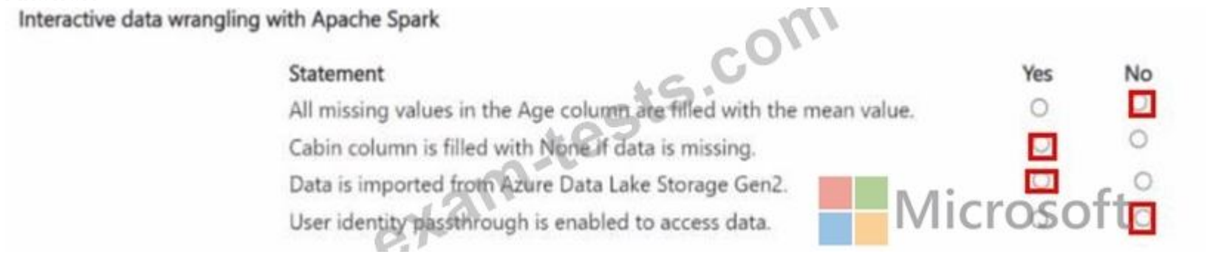
NEW QUESTION: 56
You manage an Azure Machine Learning workspace. The development environment is configured with a Serverless Spark compute in Azure Machine Learning Notebooks.
You perform interactive data wrangling to clean up the Titanic dataset and store it as a new dataset (Line numbers are used for reference only.)

```
01 import pyspark.pandas as pd
02 from pyspark.ml.feature import Imputer
03
04 df = pd.read_csv("abfss://<FILE_SYSTEM_NAME>@<STORAGE_ACCOUNT_NAME>.dfs.core.windows.net/data/titanic.csv",
index_col="PassengerId")
05
06 imputer = Imputer(inputCols=["Age"], outputCol="Age").setStrategy("mean")
07 df.fillna(value={"Cabin": "None"}, inplace=False)
08 df.dropna(inplace=True)
09
10 df.to_csv( "abfss://<FILE_SYSTEM_NAME>@<STORAGE_ACCOUNT_NAME>.dfs.core.windows.net/data/wrangled",
index_col="PassengerId")
```

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



Answer:



NEW QUESTION: 57

You are retrieving data from a large datastore by using Azure Machine Learning Studio.

You must create a subset of the data for testing purposes using a random sampling seed based on the system clock.

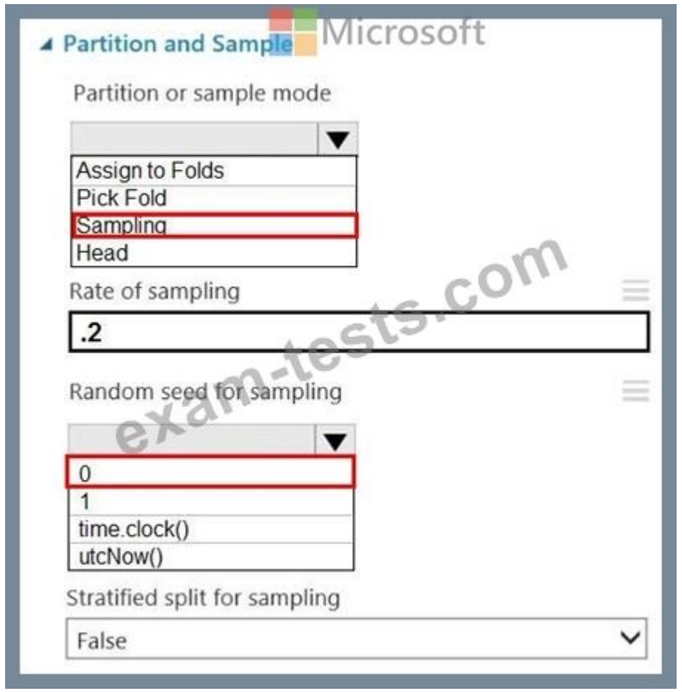
You add the Partition and Sample module to your experiment.

You need to select the properties for the module.

Which values should you select? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.

Answer:



Explanation:

Box 1: Sampling

Create a sample of data

This option supports simple random sampling or stratified random sampling. This is useful if you want to create a smaller representative sample dataset for testing.

1. Add the Partition and Sample module to your experiment in Studio, and connect the dataset.
2. Partition or sample mode: Set this to Sampling.
3. Rate of sampling. See box 2 below.

Box 2: 0

3. Rate of sampling. Random seed for sampling: Optionally, type an integer to use as a seed value.

This option is important if you want the rows to be divided the same way every time. The default value is 0, meaning that a starting seed is generated based on the system clock. This can lead to slightly different results each time you run the experiment.

References:

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

NEW QUESTION: 58

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

age,city,income,home_owner

21,Chicago,50000,0

35,Seattle,120000,1

23,Seattle,65000,0

45,Seattle,130000,1

18,Chicago,48000,0

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

the number of observations in the dataset

a box plot of income by home_owner

a dictionary containing the city names and the average income for each city You need to use the appropriate logging methods of the experiment's run object to log the required information.

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

NOTE: Each correct selection is worth one point.

Code segments

- log
- log_list
- log_row
- log_table
- log_image

Answer Area

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


Answer:

Code segments

- log
- log_list
- log_row
- log_table
- log_image

Answer Area

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


Explanation:

Box 1: log

The number of observations in the dataset.

run.log(name, value, description='')

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

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

Box 2: log_image

A box plot of income by home_owner.

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

Example: run.log_image("ROC", plot=plt)

Box 3: log_table

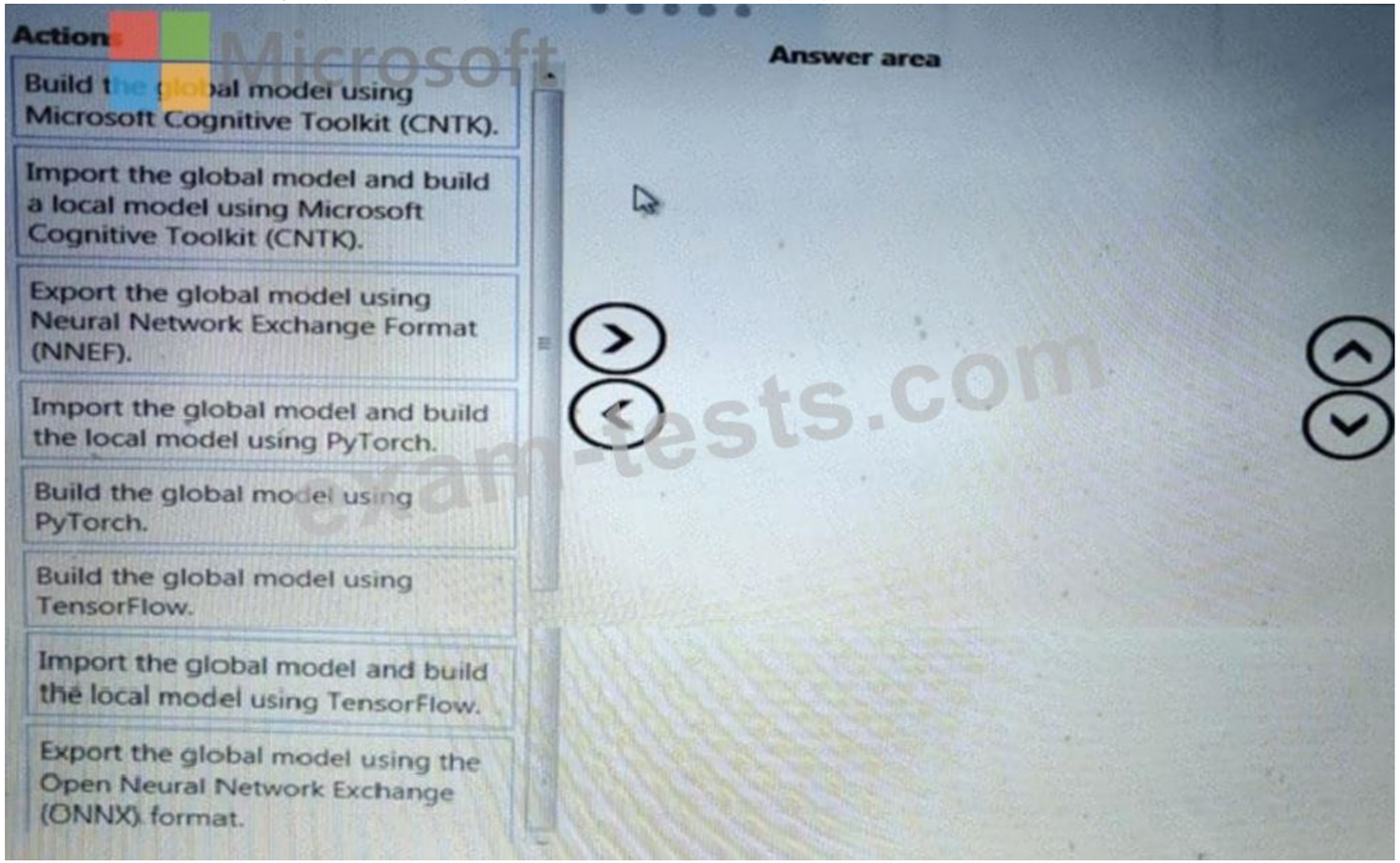
A dictionary containing the city names and the average income for each city.

log_table: Log a dictionary object to the run with the given name.

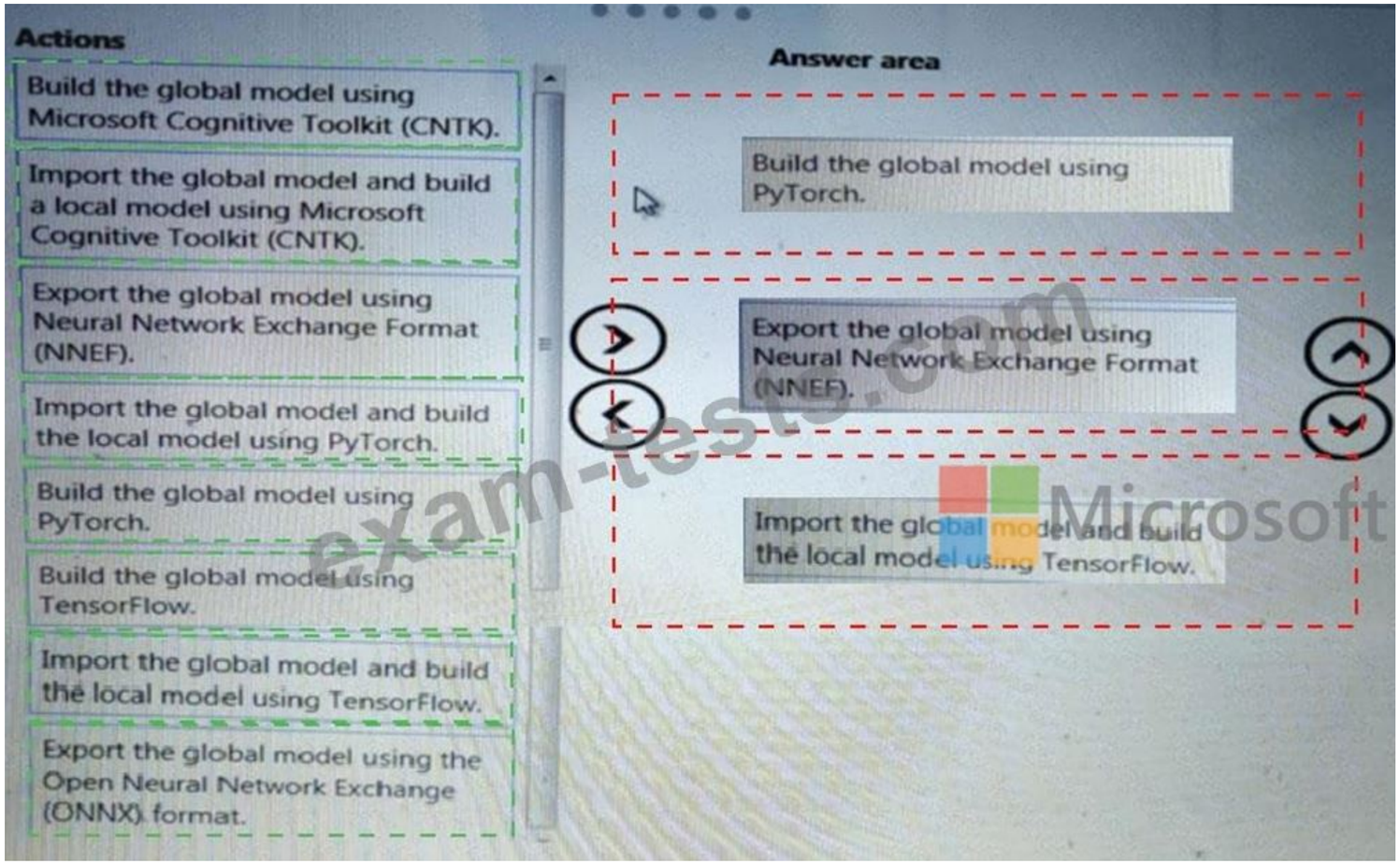
NEW QUESTION: 59

You need to define a process for penalty event detection.

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.



Answer:



NEW QUESTION: 60

You need to configure the Permutation Feature Importance module for the model training requirements.

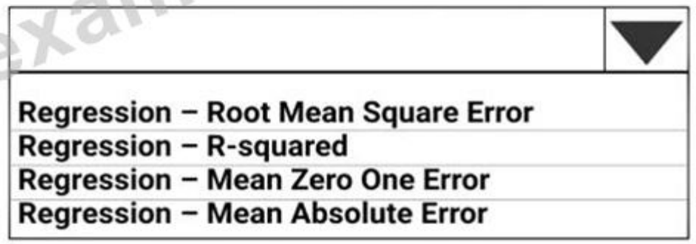
What should you do? To answer, select the appropriate options in the dialog box in the answer area.

NOTE: Each correct selection is worth one point.

Answer Area

Permutation Feature importance

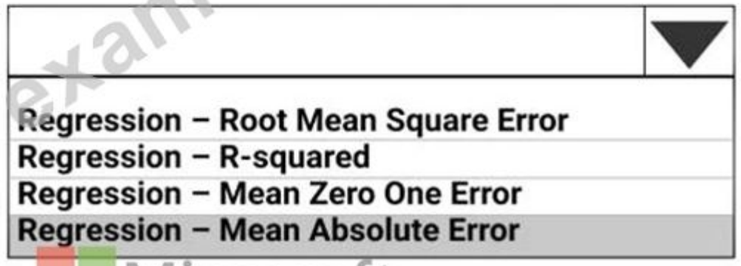
Random seed



Answer Area

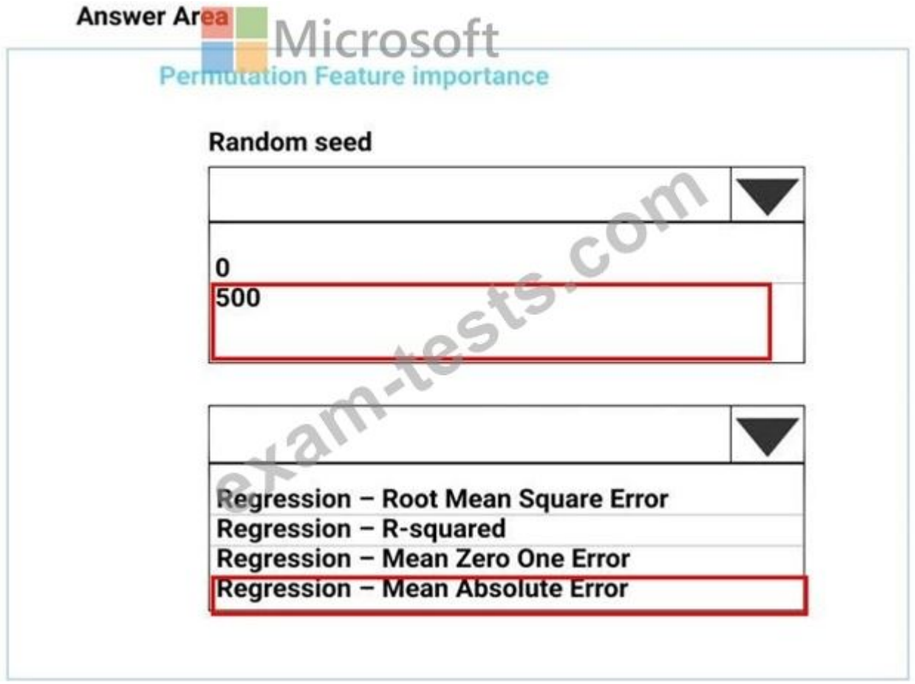
Permutation Feature importance

Random seed



Answer:

Answer Area



Explanation:
Box 1: 500
For Random seed, type a value to use as seed for randomization. If you specify 0 (the default), a number is generated based on the system clock.
A seed value is optional, but you should provide a value if you want reproducibility across runs of the same experiment.
Here we must replicate the findings.
Box 2: Mean Absolute Error
Scenario: Given a trained model and a test dataset, you must compute the Permutation Feature Importance scores of feature variables. You need to set up the Permutation Feature Importance module to select the correct metric to investigate the model's accuracy and replicate the findings.
Regression. Choose one of the following: Precision, Recall, Mean Absolute Error , Root Mean Squared Error, Relative Absolute Error, Relative Squared Error, Coefficient of Determination References:
https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/permutation-feature-importance

NEW QUESTION: 61
You are using the Azure Machine Learning SDK to run a training experiment that trains a classification model and calculates its accuracy metric.
The model will be retrained each month as new data is available.
You must register the model for use in a batch inference pipeline.
You need to register the model by using mlflow and ensure that the models created by subsequent retraining experiments are registered only if their accuracy is higher than the currently registered model.
What should you do?
A. Register the model with the same name each time regardless of accuracy, and always use the latest version of the model in the batch inferencing pipeline.
B. Register a metric named accuracy with the accuracy metric as a value when registering the model, and only register subsequent models if their accuracy is higher than the accuracy tag value of the currently registered model.
C. Specify the model framework version when registering the model, and only register subsequent models if this value is higher.
D. Specify a different name for the model each time you register it.

Answer: C (LEAVE A REPLY)

NEW QUESTION: 62

You manage an Azure Machine Learning workspace. You configure an automated machine learning regression training job by using the Azure Machine Learning Python SDK v2. You configure the regression job by using the following script:

```
regression_job.set_limits(
    timeout_minutes = 60,
    max_concurrent_trials = 5,
    enable_early_termination = True
)
```

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

Answer Area

Statements	Yes	No
The job is terminated if the score is not improving in a specific number of iterations.	○	○
A maximum of five AutoML trials are run in parallel during the regression job.	○	○
One AutoML trial can run for 60 minutes before it is terminated.	○	○
The AutoML trial run can take up to 1 month before it terminates.	○	○

Answer:

Answer Area

Statements	Yes	No
The job is terminated if the score is not improving in a specific number of iterations.	○	○
A maximum of five AutoML trials are run in parallel during the regression job.	○	○
One AutoML trial can run for 60 minutes before it is terminated.	○	○
The AutoML trial run can take up to 1 month before it terminates.	○	○

Explanation:

Answer Area

Statements	Yes	No
The job is terminated if the score is not improving in a specific number of iterations.	○	●
A maximum of five AutoML trials are run in parallel during the regression job.	●	○
One AutoML trial can run for 60 minutes before it is terminated.	●	○
The AutoML trial run can take up to 1 month before it terminates.	○	●

NEW QUESTION: 63
You are using a Git repository to track work in an Azure Machine Learning workspace.
You need to authenticate a Git account by using SSH.
Which three actions should you perform in sequence? To answer, move the appropriate actions from the list of actions to the answer area and arrange them in the correct order.

Actions

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

Answer Area

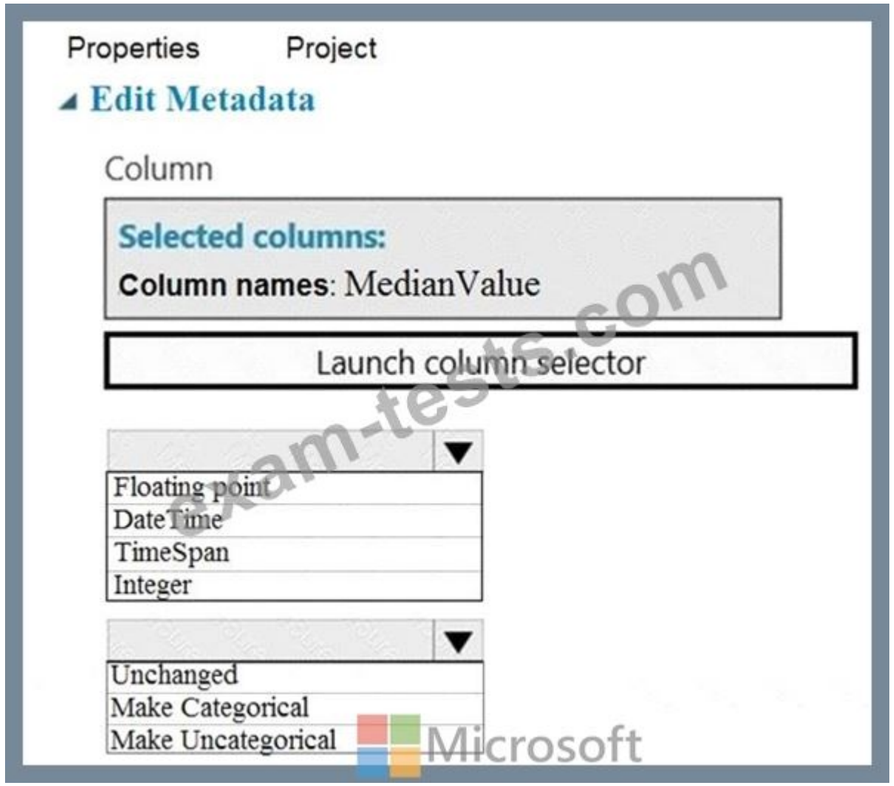
Answer:

Answer Area

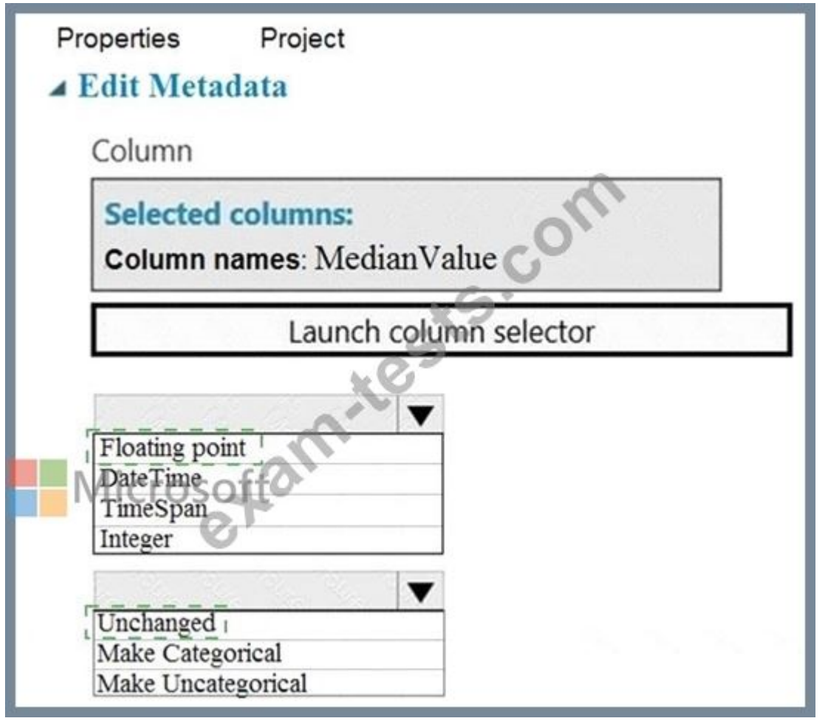
1. Generating a public/private key pair
2. Add the public key to the Git Account
3. Clone the Git repository by using an SSH repository URL

1 - Generating a public/private key pair
2 - Add the public key to the Git Account
3 - Clone the Git repository by using an SSH repository URL
Reference:
https://docs.microsoft.com/en-us/azure/machine-learning/concept-train-model-git-integration

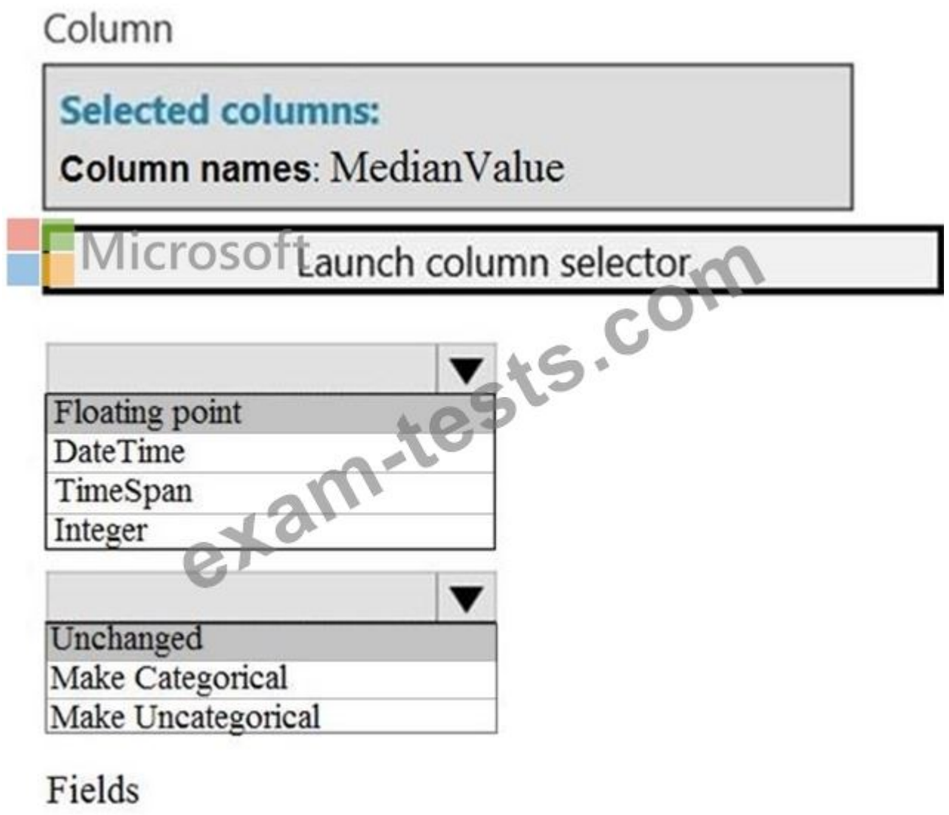
NEW QUESTION: 64
You need to configure the Edit Metadata module so that the structure of the datasets match.
Which configuration options should you select? To answer, select the appropriate options in the answer area.
NOTE: Each correct selection is worth one point.



Answer:



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: 65
You create a new Azure Databricks workspace.
You configure a new cluster for long-running tasks with mixed loads on the compute cluster as shown in the image below.

Microsoft Azure

Microsoft

Azure Databricks

Home

Workspace

Recents

Data

Clusters

Jobs

Models

Search

Create Cluster

New Cluster

Cancel

Create Cluster

2-8 Workers: 28.0-112.0 GB Memory, 8-32 Cores, 1.5-6 DBU
1 Driver: 14.0 GB Memory, 4 Cores, 0.75 DBU

Cluster Name

mysparkcluster

Cluster Mode

Standard

Pool

None

Databricks Runtime Version

Learn more

Runtime: 6.4 (Scala 2.11, Spark 2.4.5)

New This Runtime version supports only Python 3.

Autopilot Options

Enable autoscaling

Terminate after 120 minutes of inactivity

Worker Type

Min Workers

Max Workers

Standard_DS3_v2 14.0 GB Memory, 4 Cores, 0.75 DBU

2

8

Driver Type

Same as worker 14.0 GB Memory, 4 Cores, 0.75 DBU

Advanced Options

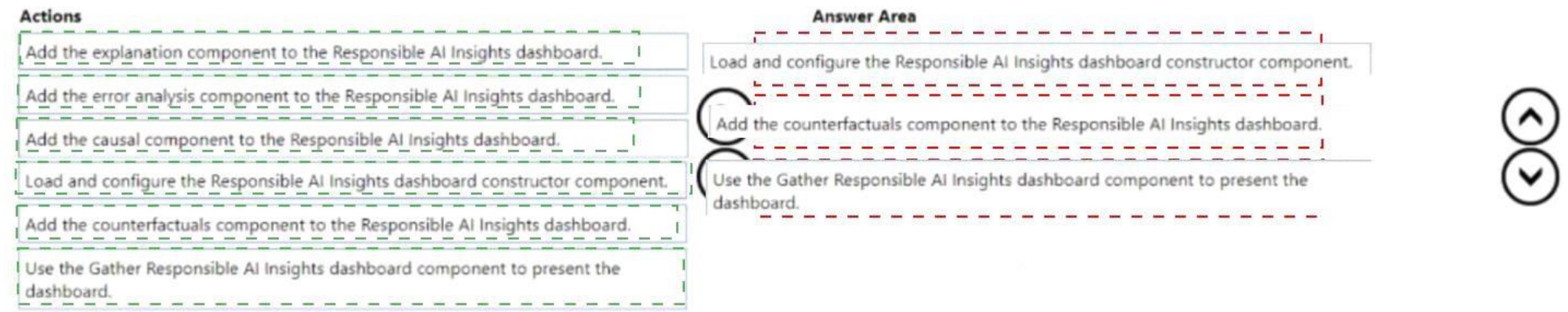
Use the drop-down menus to select the answer choice that completes each statement based on the information presented in the graphic.
NOTE: Each correct selection is worth one point.

Answer:

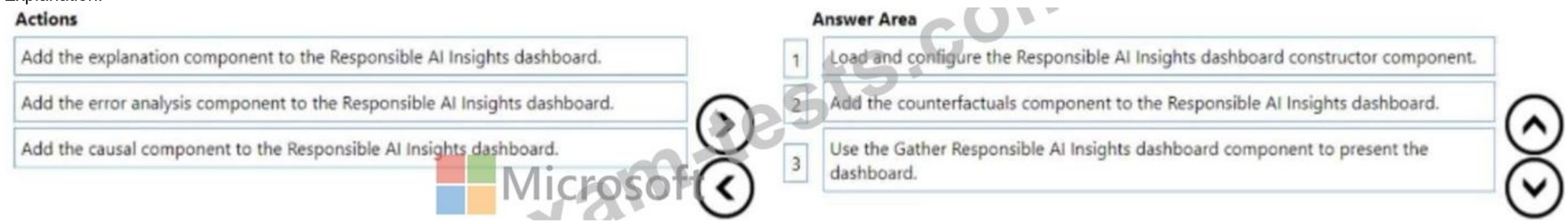
Explanation
Box 1: No
Running user code in separate processes is not possible in Scala.
Box 2: No
Autoscaling is enabled. Minimum 2 workers, Maximum 8 workers.
Reference:
https://docs.databricks.com/clusters/configure.html

NEW QUESTION: 66
You manage an Azure Machine Learning workspace. You train a model named model1.
You must identify the features to modify for a differing model prediction result.
You need to configure the Responsible AI (RAI) dashboard for model1.
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.

Answer:



Explanation:



NEW QUESTION: 67

You are developing deep learning models to analyze semi-structured, unstructured, and structured data types.

You have the following data available for model building:

Video recordings of sporting events

Transcripts of radio commentary about events

Logs from related social media feeds captured during sporting events

You need to select an environment for creating the model.

Which environment should you use?
A. Azure Cognitive Services
B. Azure Data Lake Analytics
C. Azure HDInsight with Spark MLib
D. Azure Machine Learning Studio
Answer: A (LEAVE A REPLY)
Azure Cognitive Services expand on Microsoft's evolving portfolio of machine learning APIs and enable developers to easily add cognitive features - such as emotion and video detection; facial, speech, and vision recognition; and speech and language understanding - into their applications. The goal of Azure Cognitive Services is to help developers create applications that can see, hear, speak, understand, and even begin to reason. The catalog of services within Azure Cognitive Services can be categorized into five main pillars - Vision, Speech, Language, Search, and Knowledge.
References:
https://docs.microsoft.com/en-us/azure/cognitive-services/welcome

NEW QUESTION: 68
Hotspot Question
You are implementing hyperparameter tuning for a model training from a notebook. The notebook is in an Azure Machine Learning workspace. You add code that imports all relevant Python libraries.
You must configure Bayesian sampling over the search space for the num_hidden_layers and batch_size hyperparameters.
You need to complete the following Python code to configure Bayesian sampling.
Which code segments should you use? To answer, select the appropriate options in the answer area.
NOTE: Each correct selection is worth one point.

Answer Area

```
param_sampling = BayesianParameterSampling( {
     "learning_rate": uniform(0.05, 0.1),
     "batch_size":  [choice | range | loguniform] ( [lognormal | range | uniform] (16, 128, 16))
     }
}
```


Answer:

Answer Area

```
param_sampling = BayesianParameterSampling( {
     "learning_rate": uniform(0.05, 0.1),
     "batch_size": [choice | range | loguniform] ( [lognormal | range | uniform] (16, 128, 16)
     }
)
```


NEW QUESTION: 69

You have the following Azure subscriptions and Azure Machine Learning service workspaces:

Subscription	Workspace	Comment
385bdfe5-4cef-4ad4-b977-3f86d92727c9	ml-default	This is the default subscription.
5a5891d1-557a-4234-9b83-2e90412b1068	ml-project	The information required to uniquely identify this workspace is stored in the file config.json in the same folder as the Python script.

You need to obtain a reference to the ml-project workspace.

Solution: Run the following Python code:

```
from azureml.core import Workspace
ws = Workspace(workspace_name="ml-project")
```

Does the solution meet the goal?

A. No

B. Yes

Answer: A (LEAVE A REPLY)

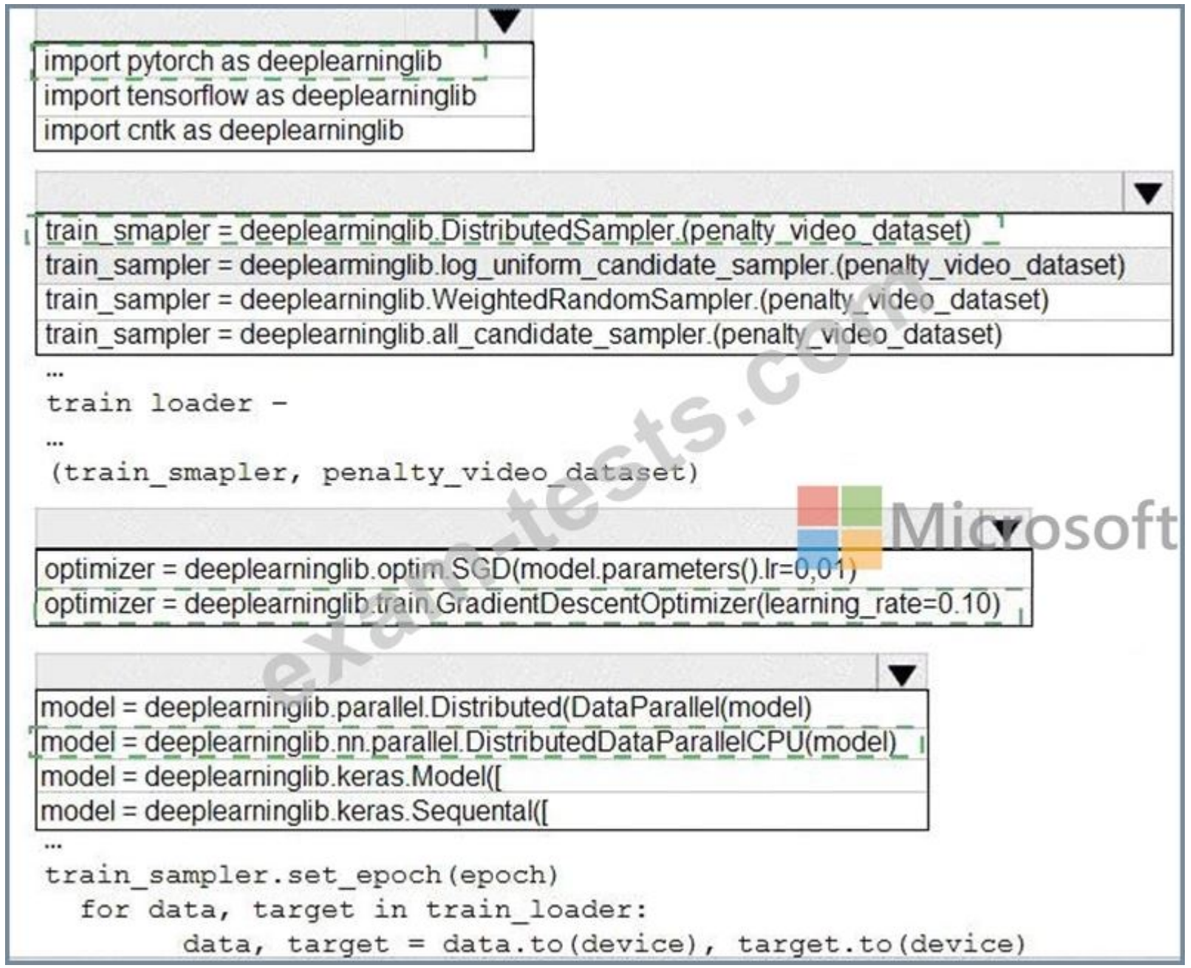
NEW QUESTION: 70

You need to use the Python language to build a sampling strategy for the global penalty detection models.

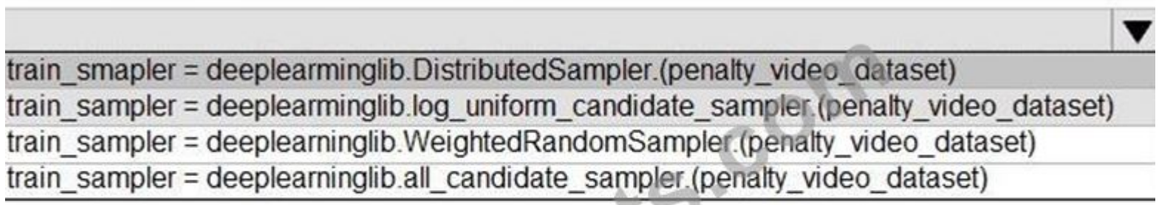
How should you complete the code segment? To answer, select the appropriate options in the answer area.

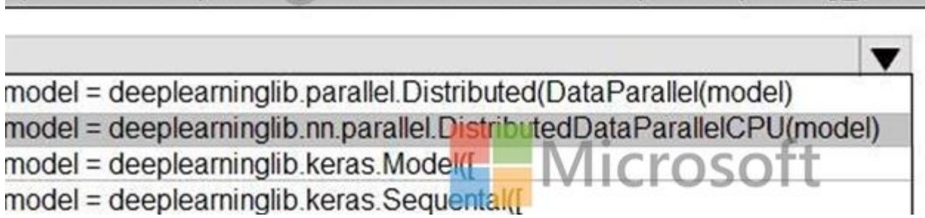
NOTE: Each correct selection is worth one point.

Answer:



Explanation:





Box 1: import pytorch as deeplearninglib

Box 2: ..DistributedSampler(Sampler)..

DistributedSampler(Sampler):

Sampler that restricts data loading to a subset of the dataset.

It is especially useful in conjunction with class:`torch.nn.parallel.DistributedDataParallel`. In such case, each process can pass a DistributedSampler instance as a DataLoader sampler, and load a subset of the original dataset that is exclusive to it.

Scenario: Sampling must guarantee mutual and collective exclusively between local and global segmentation models that share the same features.

Box 3: optimizer = deeplearninglib.train. GradientDescentOptimizer(learning_rate=0.10)

Topic 2, Case Study 2

Case study

Overview

You are a data scientist for Fabrikam Residences, a company specializing in quality private and commercial property in the United States. Fabrikam Residences is considering expanding into Europe and has asked you to investigate prices for private residences in major European cities. You use Azure Machine Learning Studio to measure the median value of properties. You produce a regression model to predict property prices by using the Linear Regression and Bayesian Linear Regression modules.

Datasets

There are two datasets in CSV format that contain property details for two cities, London and Paris, with the following columns:

Column heading	Description
CapitaCrimeRate	per capita crime rate by town
Zoned	proportion of residential land zoned for lots over 25.000 square feet
NonRetailAcres	proportion of retail business acres per town
NextToRiver	proximity of the property to the river
NitrogenOxideConcentration	nitric oxides concentration (parts per 10 million)
AvgRoomsPerHouse	average number of rooms per dwelling
Age	proportion of owner-occupied units built prior to 1940
DistanceToEmploymentCenter	weighted distances to employment centers
AccessibilityToHighway	index of accessibility to radial highways to a value of two decimal places
Tax	full value property tax rate per $10,000
PupilTeacherRatio	pupil to teacher ratio by town
ProfessionalClass	professional class percentage
LowerStatus	percentage lower status of the population
MedianValue	median value of owner-occupied homes in $1000s

The two datasets have been added to Azure Machine Learning Studio as separate datasets and included as the starting point of the experiment.

Dataset issues

The AccessibilityToHighway column in both datasets contains missing values. The missing data must be replaced with new data so that it is modeled conditionally using the other variables in the data before filling in the missing values.

Columns in each dataset contain missing and null values. The dataset also contains many outliers. The Age column has a high proportion of outliers. You need to remove the rows that have outliers in the Age column.

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.

Model fit

The model shows signs of overfitting. You need to produce a more refined regression model that reduces the overfitting.

Experiment requirements

You must set up the experiment to cross-validate the Linear Regression and Bayesian Linear Regression modules to evaluate performance.

In each case, the predictor of the dataset is the column named MedianValue. An initial investigation showed 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. You must ensure that the datatype of the MedianValue column of the Paris dataset matches the structure of the London dataset.

You must prioritize the columns of data for predicting the outcome. You must use non-parameters statistics to measure the relationships.

You must use a feature selection algorithm to analyze the relationship between the MedianValue and AvgRoomsinHouse columns.

Model training

Given a trained model and a test dataset, you need to compute the permutation feature importance scores of feature variables. You need to set up the Permutation Feature Importance module to select the correct metric to investigate the model's accuracy and replicate the findings.

You want to configure hyperparameters in the model learning process to speed the learning phase by using hyperparameters. In addition, this configuration should cancel the lowest performing runs at each evaluation interval, thereby directing effort and resources towards models that are more likely to be successful.

You are concerned that the model might not efficiently use compute resources in hyperparameter tuning. You also are concerned that the model might prevent an increase in the overall tuning time. Therefore, you need to implement an early stopping criterion on models that provides savings without terminating promising jobs.

Testing

You must produce multiple partitions of a dataset based on sampling using the Partition and Sample module in Azure Machine Learning Studio. You must create three equal partitions for cross-validation. You must also configure the cross-validation process so that the rows in the test and training datasets are divided evenly by properties that are near each city's main river. The data that identifies that a property is near a river is held in the column named NextToRiver. You want to complete this task before the data goes through the sampling process.

When you train a Linear Regression module using a property dataset that shows data for property prices for a large city, you need to determine the best features to use in a model. You can choose standard metrics provided to measure performance before and after the feature importance process completes. You must ensure that the distribution of the features across multiple training models is consistent.

Data visualization

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.

NEW QUESTION: 71

You manage an Azure Machine Learning workspace. You submit a training job with the Azure Machine Learning Python SDK v2. You must use MLflow to log metrics, model parameters, and mode! artifacts automatically when training a model.

You start by writing the following code segment:

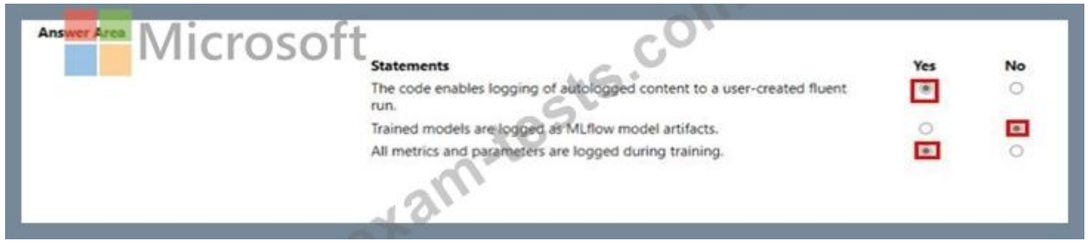
```
import mlflow
mlflow.autolog(log_models=False, exclusive=True)
```

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

Answer Area

Statements	Yes	No
The code enables logging of autologged content to a user-created fluent run.	◉	○
Trained models are logged as MLflow model artifacts.	○	◉
All metrics and parameters are logged during training.	◉	○

Answer:



NEW QUESTION: 72

You have a Python data frame named salesData in the following format:

	shop	2017	2018
0	Shop X	34	25
1	Shop Y	65	76
2	Shop Z	48	55

The data frame must be unpivoted to a long data format as follows:

	shop	year	value
0	Shop X	2017	34
1	Shop Y	2017	65
2	Shop Z	2017	48
3	Shop X	2018	25
4	Shop Y	2018	76
5	Shop Z	2018	55

You need to use the pandas.melt() function in Python to perform the transformation.

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

NOTE: Each correct selection is worth one point.

Answer Area

```
import pandas as pd
salesData = pd.melt( [ ▼ ] ,id_vars=' [ ▼ ] ',value_vars= [ ▼ ] )
```

- dataFrame
- pandas
- salesData
- year

- shop
- year
- value
- Shop X, Shop Y, Shop Z

- 'shop'
- 'year'
- ['year']
- ['2017', '2018]

Answer:

Answer Area

```
import pandas as pd
salesData = pd.melt( [dataFrame ▼] ,id_vars=' [shop ▼] ',value_vars= [['2017', '2018] ▼] )
```

Explanation:

Box 1: dataFrame

Syntax: pandas.melt(frame, id_vars=None, value_vars=None, var_name=None, value_name='value', col_level=None)[source] Where frame is a DataFrame Box 2: shop Paramter id_vars id_vars : tuple, list, or ndarray, optional Column(s) to use as identifier variables.

Box 3: ['2017','2018']

value_vars : tuple, list, or ndarray, optional

Column(s) to unpivot. If not specified, uses all columns that are not set as id_vars.

Example:

df = pd.DataFrame({'A': {0: 'a', 1: 'b', 2: 'c'},

... 'B': {0: 1, 1: 3, 2: 5},

... 'C': {0: 2, 1: 4, 2: 6}})

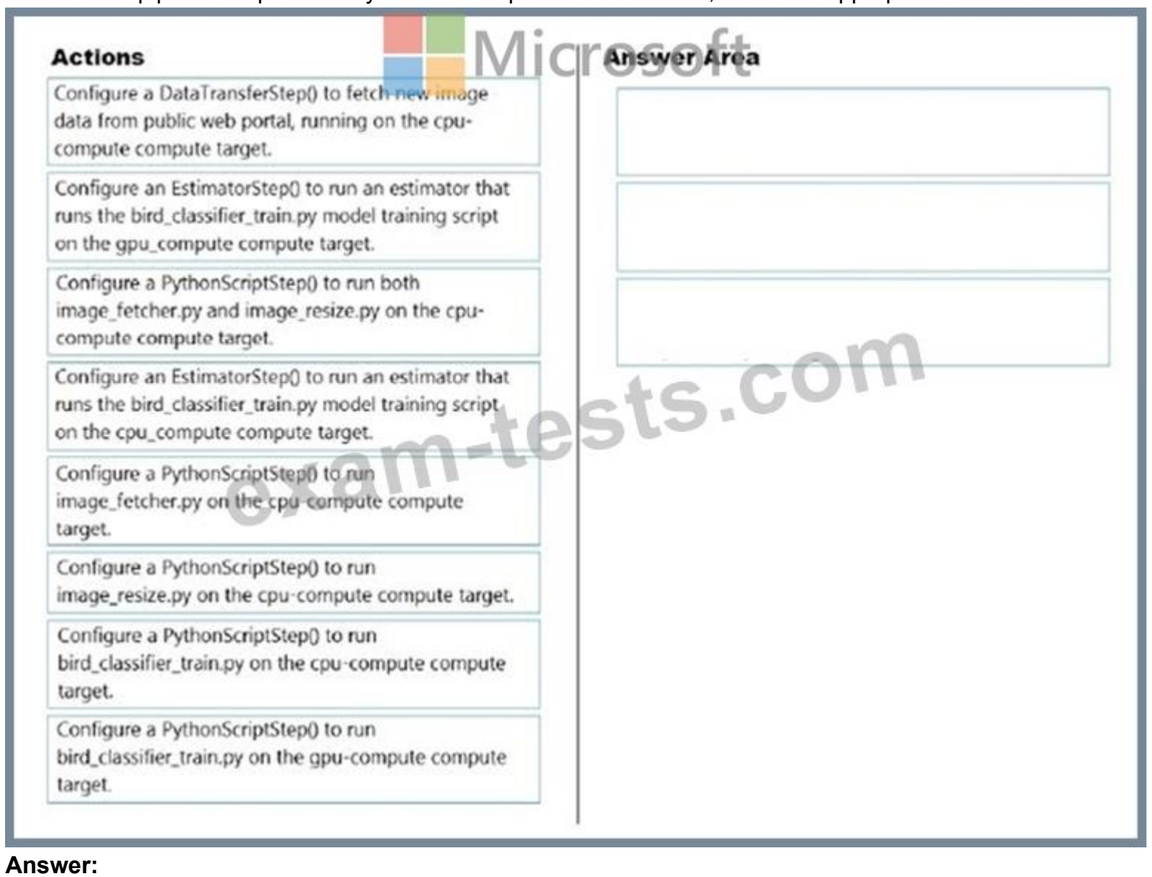
pd.melt(df, id_vars=['A'], value_vars=['B', 'C'])

A variable value

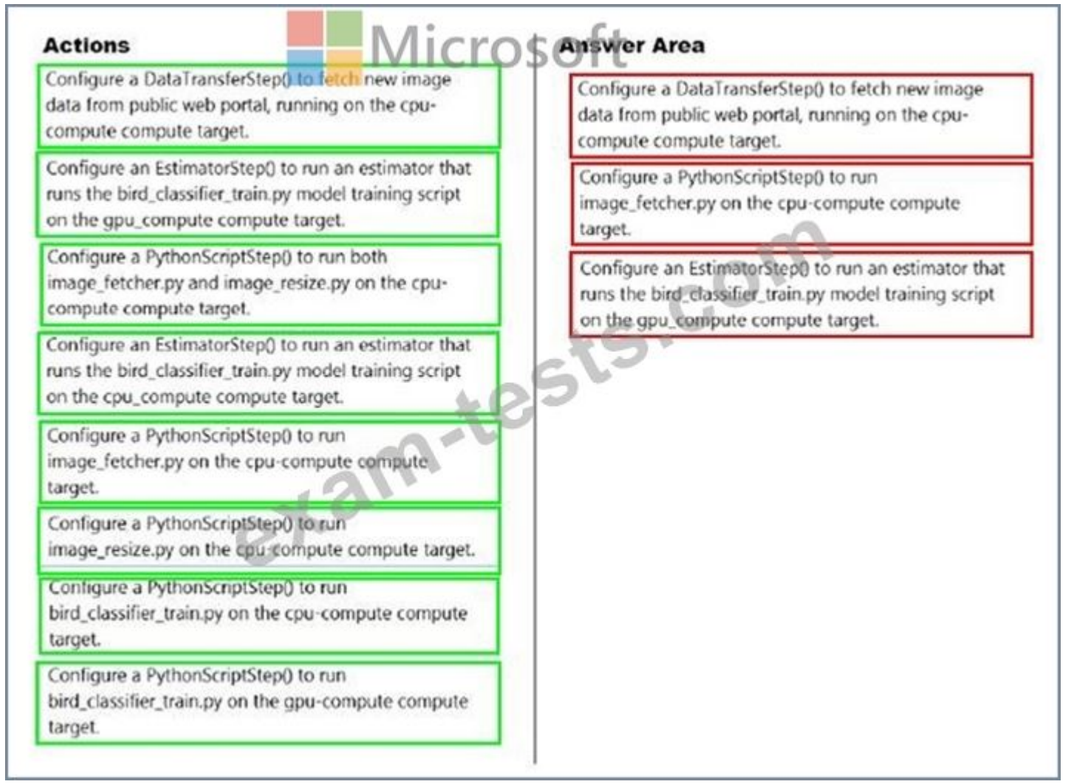
0 a B 1

1 b B 3
2 c B 5
3 a C 2
4 b C 4
5 c C 6
References:
https://pandas.pydata.org/pandas-docs/stable/reference/api/pandas.melt.html

NEW QUESTION: 73
You create a multi-class image classification deep learning experiment by using the PyTorch framework. You plan to run the experiment on an Azure Compute cluster that has nodes with GPU's.
You need to define an Azure Machine Learning service pipeline to perform the monthly retraining of the image classification model. The pipeline must run with minimal cost and minimize the time required to train the model.
Which three pipeline steps should you run in sequence? To answer, move the appropriate actions from the list of actions to the answer area and arrange them in the correct order.



Answer:



Explanation:

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

Step 2: Configure a PythonScriptStep() to run image_resize.y on the cpu-compute compute target.

Step 3: Configure the EstimatorStep() to run training script on the gpu_compute computer target.

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

Reference:

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

NEW QUESTION: 74

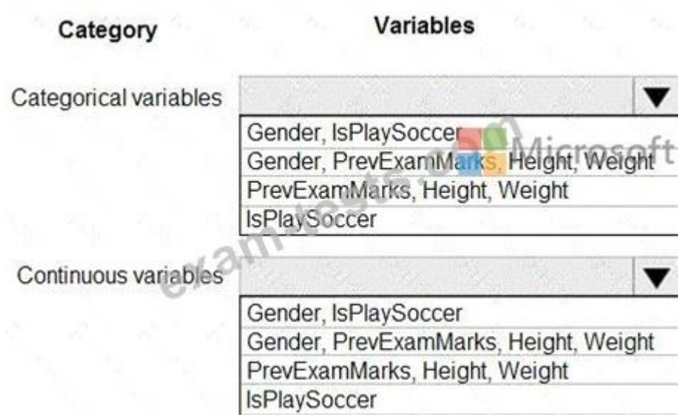
You are working on a classification task. You have a dataset indicating whether a student would like to play soccer and associated attributes. The dataset includes the following columns:

Name	Description
IsPlaySoccer	Values can be 1 and 0.
Gender	Values can be M or F.
PrevExamMarks	Stores values from 0 to 100
Height	Stores values in centimeters
Weight	Stores values in kilograms

You need to classify variables by type.

Which variable should you add to each category? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.



Answer:

Explanation:

References:

https://www.edureka.co/blog/classification-algorithms/

NEW QUESTION: 75
You plan to explore demographic data for home ownership in various cities. The data is in a CSV file with the following format:
age,city,income,home_owner
21,Chicago,50000,0
35,Seattle,120000,1
23,Seattle,65000,0
45,Seattle,130000,1
18,Chicago,48000,0
You need to run an experiment in your Azure Machine Learning workspace to explore the data and log the results. The experiment must log the following information:
the number of observations in the dataset
a box plot of income by home_owner
a dictionary containing the city names and the average income for each city You need to use the appropriate logging methods of the experiment's run object to log the required information.
How should you complete the code? To answer, drag the appropriate code segments to the correct locations. Each code segment may be used once, more than once, or not at all. You may need to drag the split bar between panes or scroll to view content.
NOTE: Each correct selection is worth one point.

Code segments

- log
- log_list
- log_row
- log_table
- log_image

Answer Area

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


Answer:

Code segments

- log
- log_list
- log_row
- log_table
- log_image

Answer Area

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

NEW QUESTION: 76

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 amlworkspace67890 • 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.

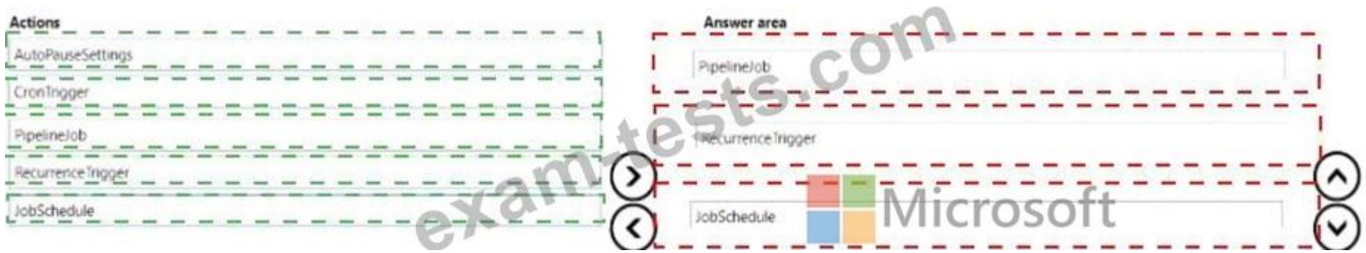
You need to run a script that trains a deep neural network (DNN) model and logs the loss and accuracy metrics.

Solution: Install the Azure ML SDK on the Surface Book. Run Python code to connect to the workspace and then run the training script as an experiment on local compute.

Does the solution meet the goal?
A. Yes
B. No
Answer: B (LEAVE A REPLY)
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: 77
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 Python SEX v2 to configure the schedule for the pipeline. The schedule should be defined by using the frequency and interval properties with frequency set to month' and interval set to "1:
Which three classes should you instantiate in sequence"' To answer, move the appropriate actions from the list of actions to the answer area and arrange them in the correct order.

Answer:



Explanation:

NEW QUESTION: 78
You are creating a new experiment in Azure Machine Learning Studio. You have a small dataset that has missing values in many columns.
The data does not require the application of predictors for each column. You plan to use the Clean Missing Data module to handle the missing data.
You need to select a data cleaning method.
Which method should you use?
A. Replace using; Probabilistic PCA
B. Normalization
C. Synthetic Minority Oversampling Technique (SMOTE)
D. Replace using MICE
Answer: C (LEAVE A REPLY)

NEW QUESTION: 79
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 an Azure Machine Learning workspace that includes an AmlCompute cluster and a batch endpoint.
You clone a repository that contains an MLflow model to your local computer.
You need to ensure that you can deploy the model to the batch endpoint.
Solution: Create a datastore in the workspace.
Does the solution meet the goal?
A. No
B. Yes
Answer: A (LEAVE A REPLY)

NEW QUESTION: 80
You need to use the Python language to build a sampling strategy for the global penalty detection models.
How should you complete the code segment? To answer, select the appropriate options in the answer area.
NOTE: Each correct selection is worth one point.

import pytorch as deeplearninglib
import tensorflow as deeplearninglib
import cntk as deeplearninglib

train_smapler = deeplearminglib.DistributedSampler.(penalty_video_dataset)
train_sampler = deeplearminglib.log_uniform_candidate_sampler.(penalty_video_dataset)
train_sampler = deeplearninglib.WeightedRandomSampler.(penalty_video_dataset)
train_sampler = deeplearninglib.all_candidate_sampler.(penalty_video_dataset)

```
...
train loader -
...
(train_smapler, penalty_video_dataset)
```

optimizer = deeplearninglib.optim.SGD(model.parameters().lr=0,01)
optimizer = deeplearninglib.train.GradientDescentOptimizer(learning_rate=0.10)

model = deeplearninglib.parallel.Distributed(DataParallel(model)
model = deeplearninglib.nn.parallel.DistributedDataParallelCPU(model)
model = deeplearninglib.keras.Model([
model = deeplearninglib.keras.Sequental([

```
...
train_sampler.set_epoch(epoch)
  for data, target in train_loader:
        data, target = data.to(device), target.to(device)
```

Answer:

Explanation:

Box 1: import pytorch as deeplearninglib

Box 2: ..DistributedSampler(Sampler)..

DistributedSampler(Sampler):

Sampler that restricts data loading to a subset of the dataset.

It is especially useful in conjunction with class:`torch.nn.parallel.DistributedDataParallel`. In such case, each process can pass a DistributedSampler instance as a DataLoader sampler, and load a subset of the original dataset that is exclusive to it.

Scenario: Sampling must guarantee mutual and collective exclusively between local and global segmentation models that share the same features.

Box 3: optimizer = deeplearninglib.train. GradientDescentOptimizer(learning_rate=0.10) Incorrect Answers: ..SGD..

Scenario: All penalty detection models show inference phases using a Stochastic Gradient Descent (SGD) are running too slow.

Box 4: .. nn.parallel.DistributedDataParallel..

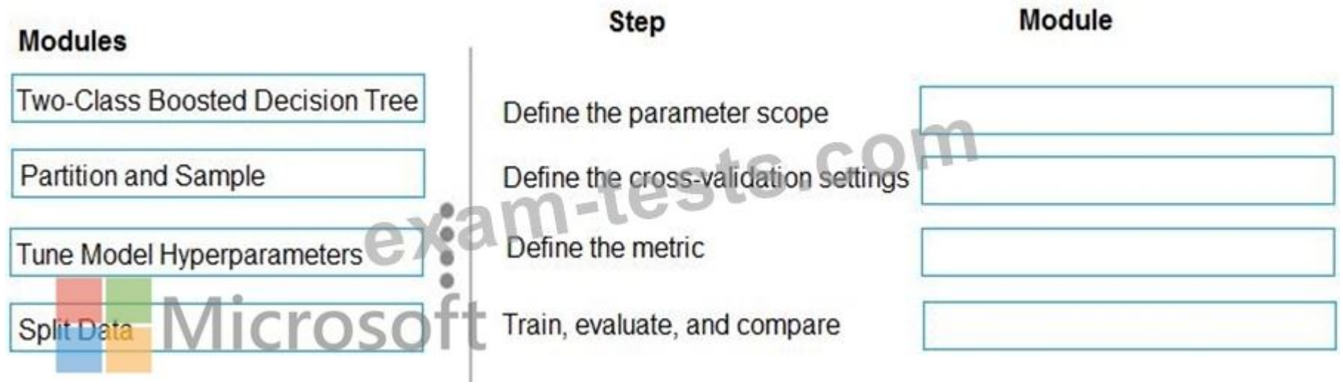
DistributedSampler(Sampler): The sampler that restricts data loading to a subset of the dataset.

It is especially useful in conjunction with :class:`torch.nn.parallel.DistributedDataParallel`.

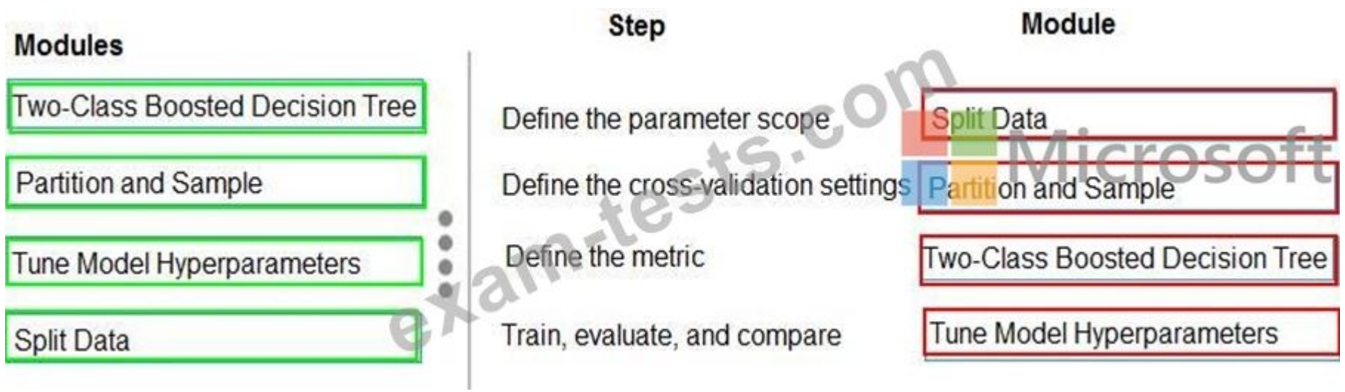
References:

https://github.com/pytorch/pytorch/blob/master/torch/utils/data/distributed.py

NEW QUESTION: 81
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.



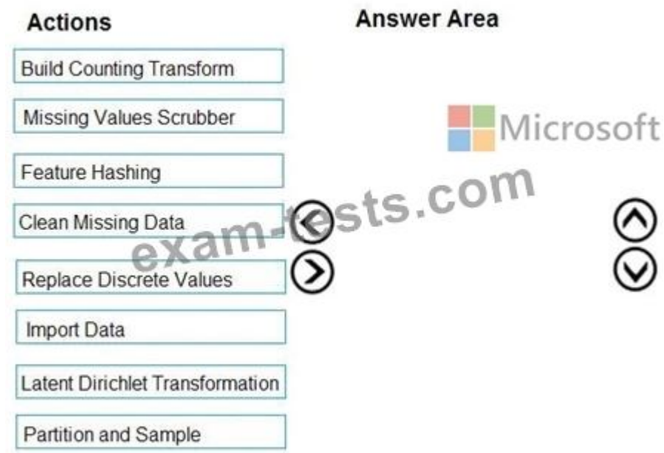
Answer:



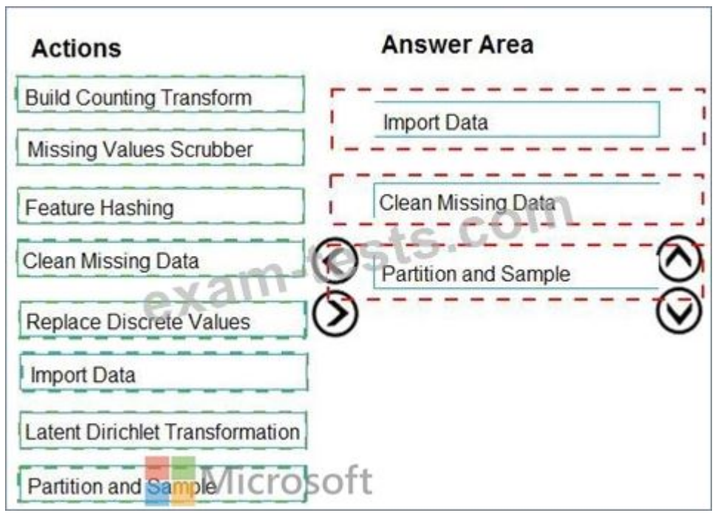
Reference:
https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/partition-and-sample

NEW QUESTION: 82
You are creating an experiment by using Azure Machine Learning Studio.
You must divide the data into four subsets for evaluation. There is a high degree of missing values in the data.
You must prepare the data for analysis.
You need to select appropriate methods for producing the experiment.
Which three modules should you run in sequence? To answer, move the appropriate actions from the list of actions to the answer area and arrange them in the correct order.

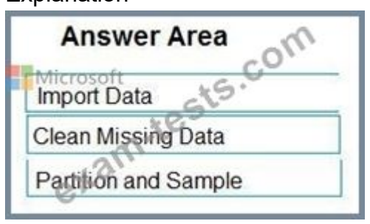
NOTE: More than one order of answer choices is correct. You will receive credit for any of the correct orders you select.



Answer:



Explanation



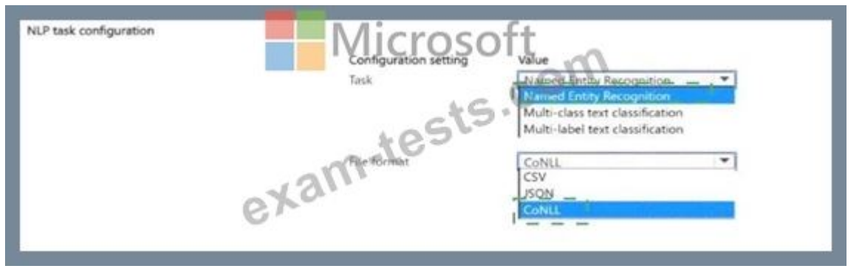
The Clean Missing Data module in Azure Machine Learning Studio, to remove, replace, or infer missing values.

NEW QUESTION: 83

You manage an Azure Machine Learning workspace.

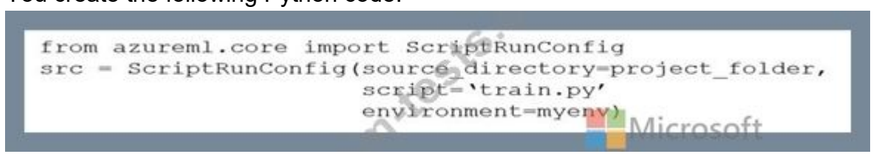
You plan to train a natural language processing (NLP) model that will assign labels ' or designated tokens in unstructured text You need to configure the NLP task by using automated machine learning.
Which configuration values should you use? To answer, select the appropriate options in the answer area.
NOTE Each correct selection is worth one point.

Answer:



Explanation:

NEW QUESTION: 84
You use an Azure Machine Learning workspace.
You create the following Python code:



```
from azureml.core import ScriptRunConfig
src = ScriptRunConfig(source_directory=project_folder,
                      script='train.py'
                      environment=myenv)
```

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

Statements	Yes	No
The default environment will be created	○	○
The training script will run on local compute	○	○
A script run configuration runs a training script named `train.py` located in a directory defined by the `project_folder` variable	○	○

Answer:

Statements	Yes	No
The default environment will be created	○	○
The training script will run on local compute	○	○
A script run configuration runs a training script named `train.py` located in a directory defined by the `project_folder` variable	○	○

Reference:
https://docs.microsoft.com/en-us/python/api/azureml-core/azureml.core.scriptrunconfig
https://docs.microsoft.com/en-us/python/api/azureml-core/azureml.core.environment.environment

NEW QUESTION: 85
You create a batch inference pipeline by using the Azure ML SDK. You run the pipeline by using the following code:

```
from azureml.pipeline.core import Pipeline
from azureml.core.experiment import Experiment
pipeline = Pipeline(workspace=ws, steps=[parallelrun_step])
pipeline_run = Experiment(ws, 'batch_pipeline').submit(pipeline)
```

You need to monitor the progress of the pipeline execution.
What are two possible ways to achieve this goal? Each correct answer presents a complete solution.
NOTE: Each correct selection is worth one point.

A. Run the following code in a notebook:

```
from azureml.contrib.interpret.explanation.explanation_client import ExplanationClient
client = ExplanationClient.from_run(pipeline_run)
explanation = client.download_model_explanation()
explanation = client.download_model_explanation(top_k=4)
global_importance_values = explanation.get_ranked_global_values()
global_importance_names = explanation.get_ranked_global_names()
print('global importance values: {}'.format(global_importance_values))
print('global importance names: {}'.format(global_importance_names))
```

B. Use the Inference Clusters tab in Machine Learning Studio.

C. Use the Activity log in the Azure portal for the Machine Learning workspace.

D. Run the following code in a notebook:

```
from azureml.widgets import RunDetails
RunDetails(pipeline_run).show()
```

E. Run the following code and monitor the console output from the PipelineRun object:

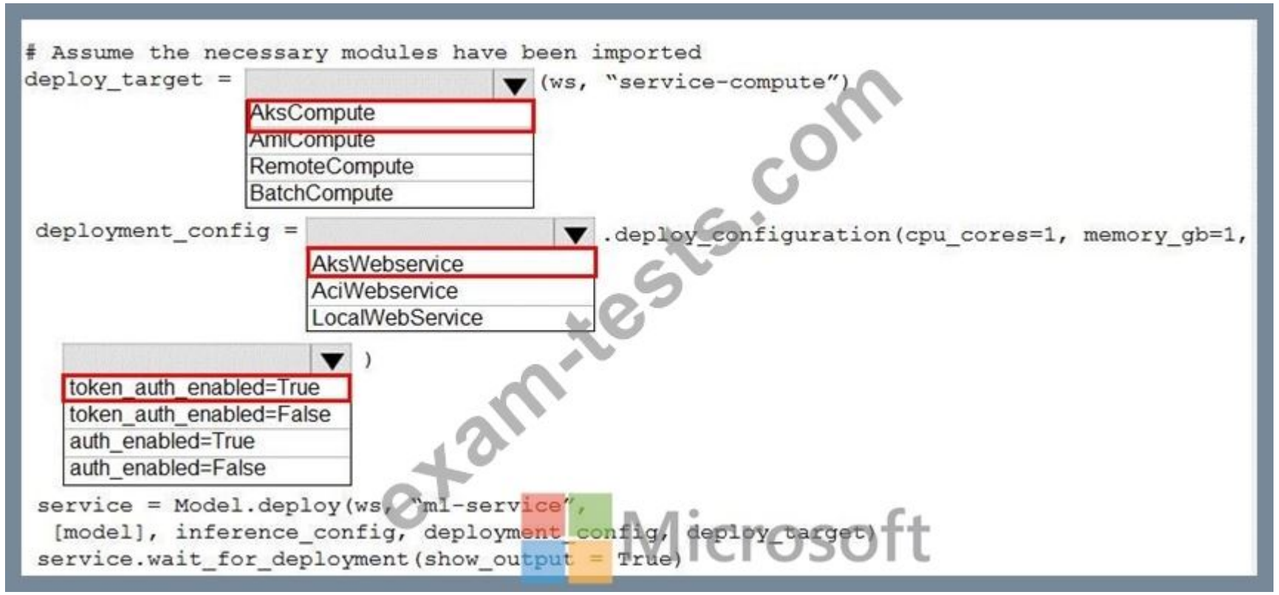
```
pipeline_run.wait_for_completion(show_output=True)
```

A. Option A
B. Option B
C. Option C
D. Option D
E. Option E
Answer: D,E (LEAVE A REPLY)
Explanation
A batch inference job can take a long time to finish. This example monitors progress by using a Jupyter widget. You can also manage the job's progress by using:
Azure Machine Learning Studio.
Console output from the PipelineRun object.
from azureml.widgets import RunDetails
RunDetails(pipeline_run).show()
pipeline_run.wait_for_completion(show_output=True)
Reference:
https://docs.microsoft.com/en-us/azure/machine-learning/how-to-use-parallel-run-step#monitor-the-parallel-run-

NEW QUESTION: 86
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.

```
 Assume the necessary modules have been imported
eploy_target =                                  (ws, "service-compute")
                AksCompute
                AmlCompute
                RemoteCompute
                BatchCompute
deployment_config =                             .deploy_configuration(cpu_cores=1, memory_gb=1,
                AksWebservice
                AciWebservice
                LocalWebService
                                                )
   token_auth_enabled=True
   token_auth_enabled=False
   auth_enabled=True
   auth_enabled=False
service = Model.deploy(ws, "ml-service",
  [model], inference_config, deployment_config, deploy_target)
service.wait_for_deployment(show_output = True)
```

Answer:



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.
Incorrect Answers:
auth_enabled (bool): Whether or not to enable key auth for this Webservice. Defaults to True.
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: 87
You create a script for training a machine learning model in Azure Machine Learning service.
You create an estimator by running the following code:

```
from azureml.core import Workspace, Datastore
from azureml.core.compute import ComputeTarget
from azureml.train.estimator import Estimator
work_space = Workspace.from_config()
data_source = work_space.get_default_datastore()
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.

	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.	○	○
The estimator can run Scikit-learn experiments.	○	○

Answer:

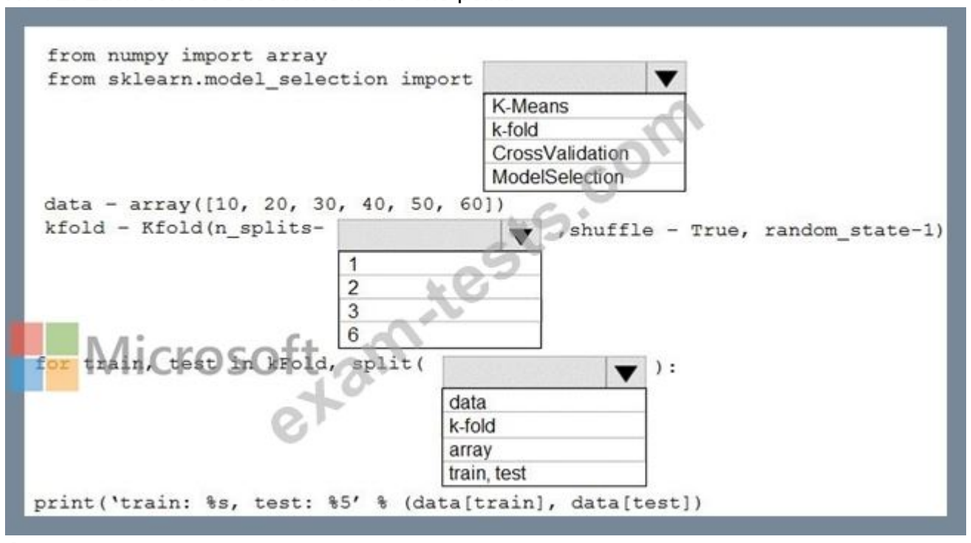
	Yes	No
The estimator will look for the files it needs to run an experiment in the training-experiment directory of the local compute environment.	○	○
The estimator will mount the local data-folder folder and make it available to the script through a parameter.	○	○
The train.py script file will be created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

Explanation:

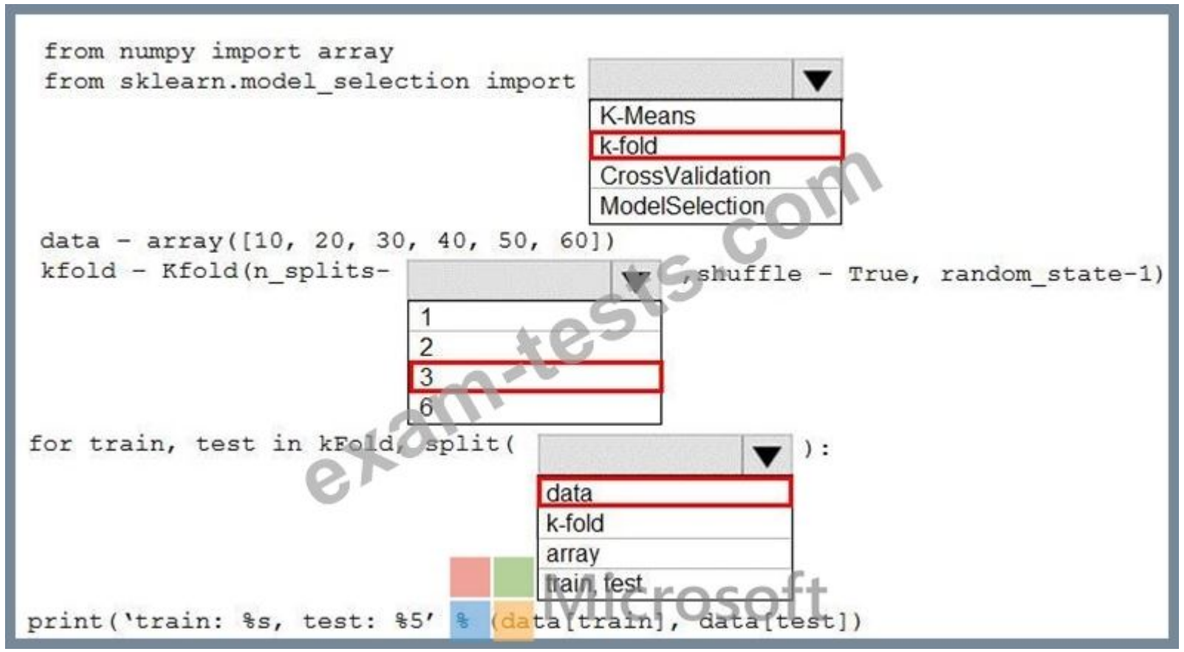
	Yes	No
The estimator will look for the files it needs to run an experiment in the training-experiment directory of the local compute environment.	○	○
The estimator will mount the local data-folder folder and make it available to the script through a parameter.	○	○
The train.py script file will be created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

Box 1: Yes
Parameter source_directory is a local directory containing experiment configuration and code files needed for a training job.
Box 2: Yes
script_params is a dictionary of command-line arguments to pass to the training script specified in entry_script.
Box 3: No
Box 4: Yes
The conda_packages parameter is a list of strings representing conda packages to be added to the Python environment for the experiment.

NEW QUESTION: 88
You are evaluating a Python NumPy array that contains six data points defined as follows:
data = [10, 20, 30, 40, 50, 60]
You must generate the following output by using the k-fold algorithm implantation in the Python Scikit-learn machine learning library:
train: [10 40 50 60], test: [20 30]
train: [20 30 40 60], test: [10 50]
train: [10 20 30 50], test: [40 60]
You need to implement a cross-validation to generate the output.
How should you complete the code segment? To answer, select the appropriate code segment in the dialog box in the answer area.
NOTE: Each correct selection is worth one point.



Answer:



Explanation:

Box 1: k-fold

Box 2: 3

K-Folds cross-validator provides train/test indices to split data in train/test sets. Split dataset into k consecutive folds (without shuffling by default).

The parameter n_splits (int, default=3) is the number of folds. Must be at least 2.

Box 3: data

Example: Example:

```
>>>
>>> from sklearn.model_selection import KFold
>>> X = np.array([[1, 2], [3, 4], [1, 2], [3, 4]])
>>> y = np.array([1, 2, 3, 4])
>>> kf = KFold(n_splits=2)
>>> kf.get_n_splits(X)
2
>>> print(kf)
KFold(n_splits=2, random_state=None, shuffle=False)
>>> for train_index, test_index in kf.split(X):
... print("TRAIN:", train_index, "TEST:", test_index)
... X_train, X_test = X[train_index], X[test_index]
... y_train, y_test = y[train_index], y[test_index]
TRAIN: [2 3] TEST: [0 1]
TRAIN: [0 1] TEST: [2 3]
```

References:

https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.KFold.html

NEW QUESTION: 89
You use an Azure Machine Learning workspace.
You create the following Python code:

```
from azureml.core import ScriptRunConfig
src = ScriptRunConfig(source_directory=project_folder,
                      script='train.py'
                      environment=myenv)
```

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

Statements	Yes	No
The default environment will be created	○	○
The training script will run on local compute	○	○
A script run configuration runs a training script named `train.py` located in a directory defined by the `project_folder` variable	○	○

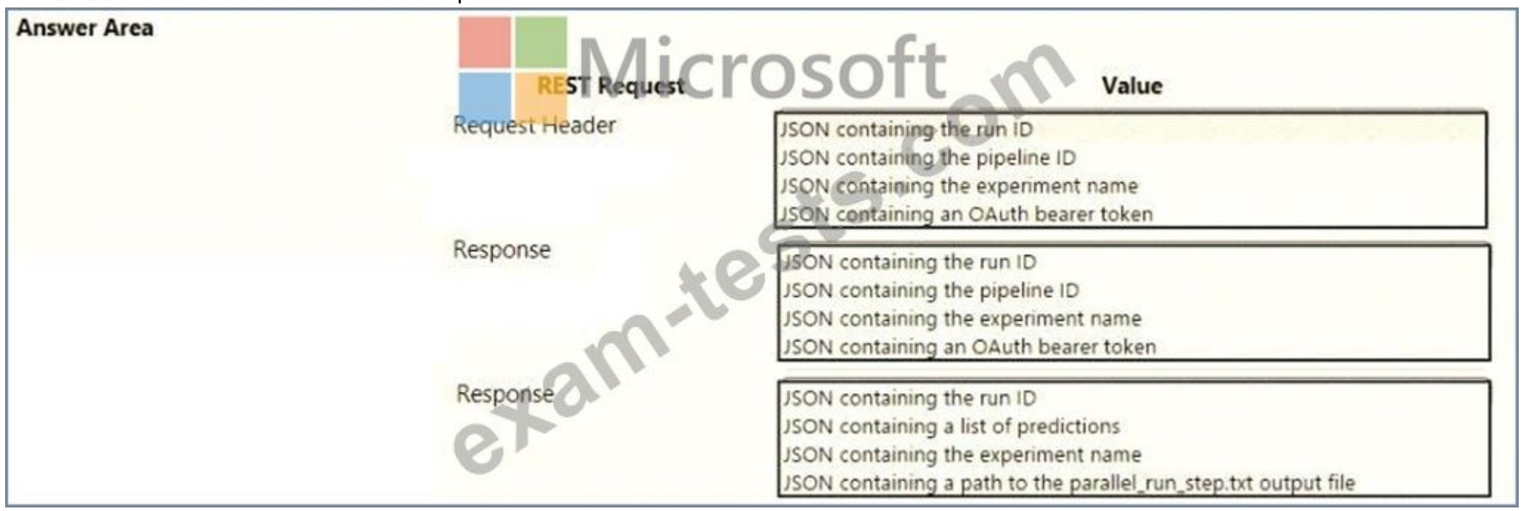
Answer:

Statements	Yes	No
The default environment will be created	○	●
The training script will run on local compute	●	○
A script run configuration runs a training script named `train.py` located in a directory defined by the `project_folder` variable	●	○

Reference:
https://docs.microsoft.com/en-us/python/api/azureml-core/azureml.core.scriptrunconfig
https://docs.microsoft.com/en-us/python/api/azureml-core/azureml.core.environment.environment

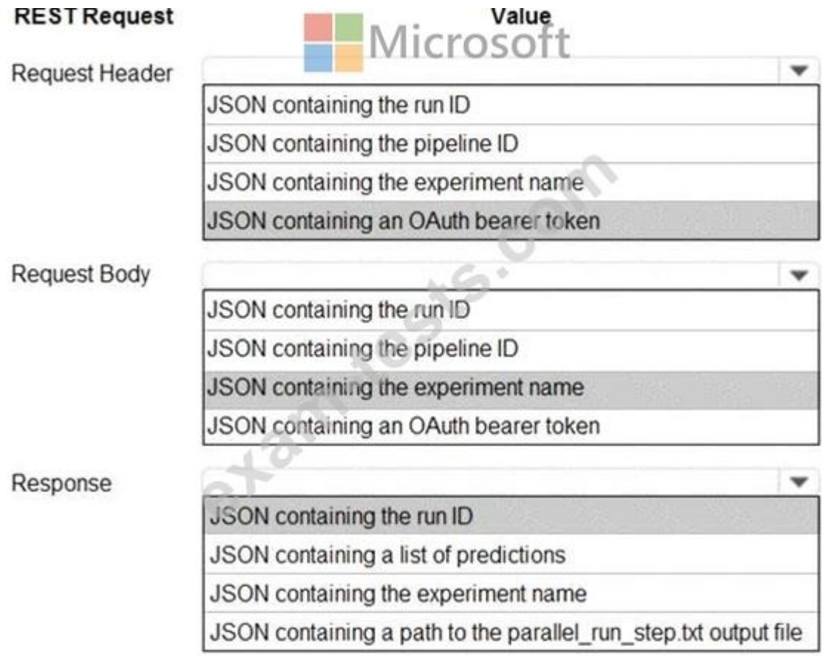
NEW QUESTION: 90
You publish a batch inferencing pipeline that will be used by a business application.
The application developers need to know which information should be submitted to and returned by the REST interface for the published pipeline.

You need to identify the information required in the REST request and returned as a response from the published pipeline.
Which values should you use in the REST request and to expect in the response? To answer, select the appropriate options in the answer area.
NOTE: Each correct selection is worth one point.



Answer:

Explanation



Box 1: JSON containing an OAuth bearer token

Specify your authentication header in the request.

To run the pipeline from the REST endpoint, you need an OAuth2 Bearer-type authentication header.

Box 2: JSON containing the experiment name

Add a JSON payload object that has the experiment name.

Example:

```
rest_endpoint = published_pipeline.endpoint
response = requests.post(rest_endpoint,
headers=auth_header,
json={"ExperimentName": "batch_scoring",
"ParameterAssignments": {"process_count_per_node": 6}})
run_id = response.json()["Id"]
```

Box 3: JSON containing the run ID

Make the request to trigger the run. Include code to access the Id key from the response dictionary to get the value of the run ID.

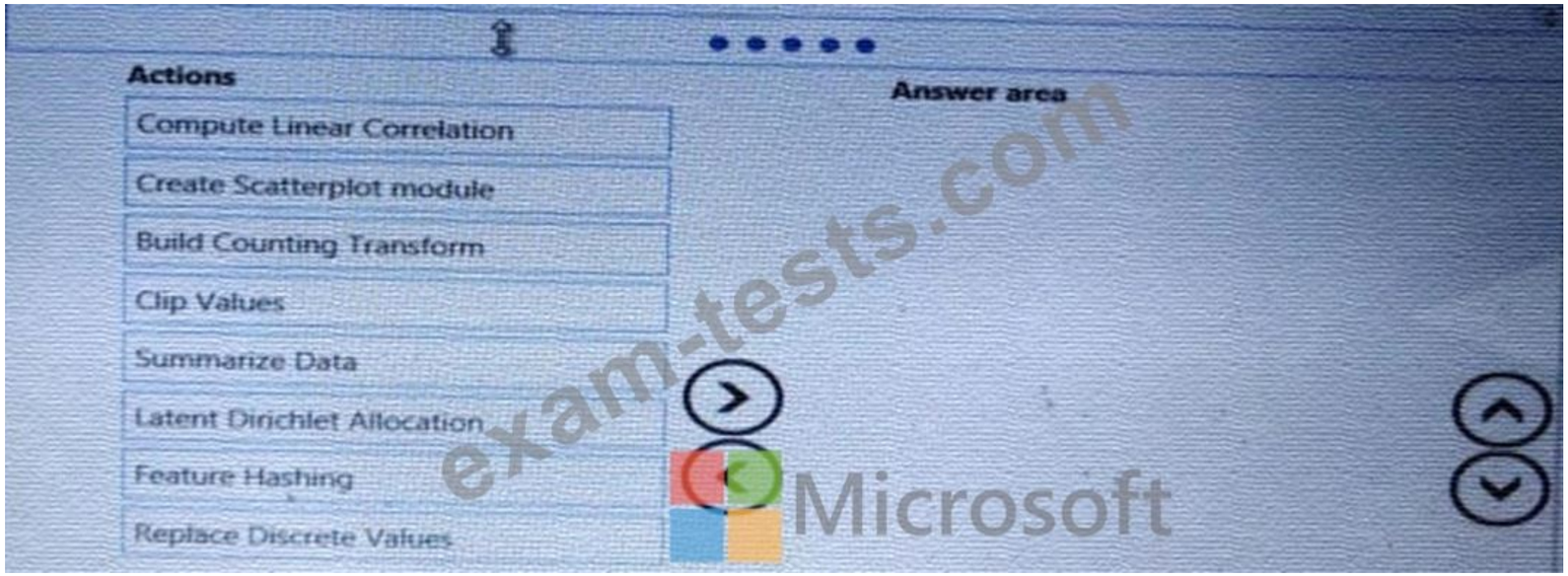
Reference:

https://docs.microsoft.com/en-us/azure/machine-learning/tutorial-pipeline-batch-scoring-classification

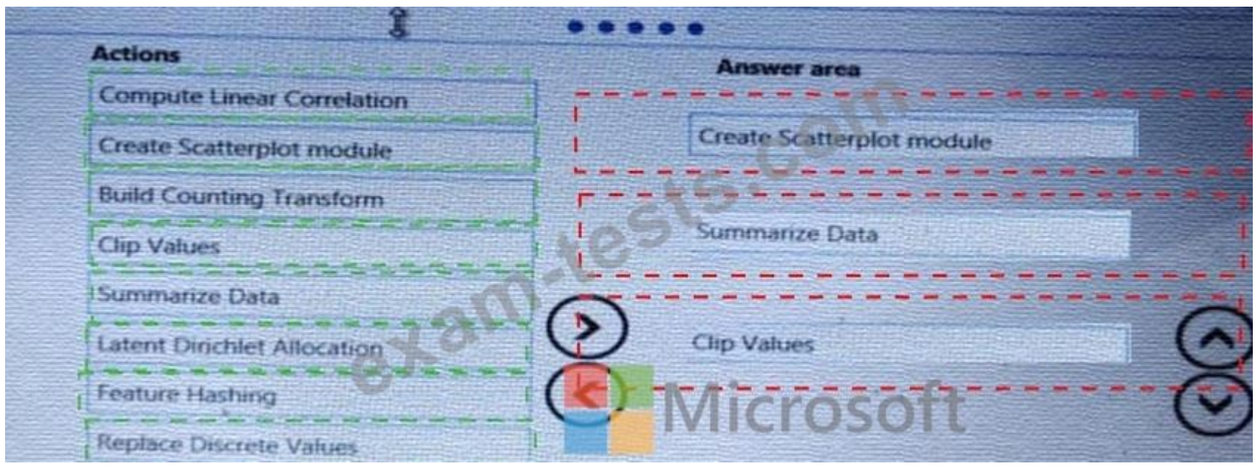
NEW QUESTION: 91

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

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



Answer:



Explanation:

Create Scatterplot

Summarize Data

Clip Values

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

References:

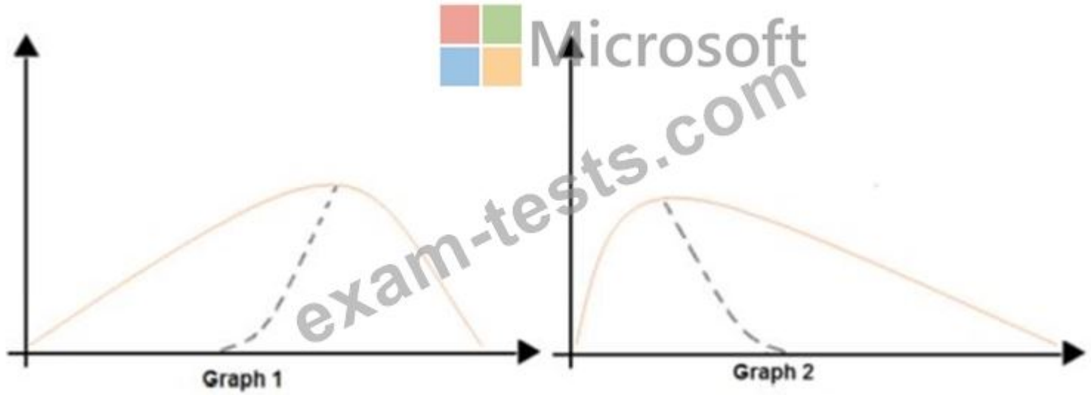
https://blogs.msdn.microsoft.com/azuredev/2017/05/27/data-cleansing-tools-in-azure-machine-learning/

https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/clip-values

NEW QUESTION: 92

You are analyzing the asymmetry in a statistical distribution.

The following image contains two density curves that show the probability distribution of two datasets.

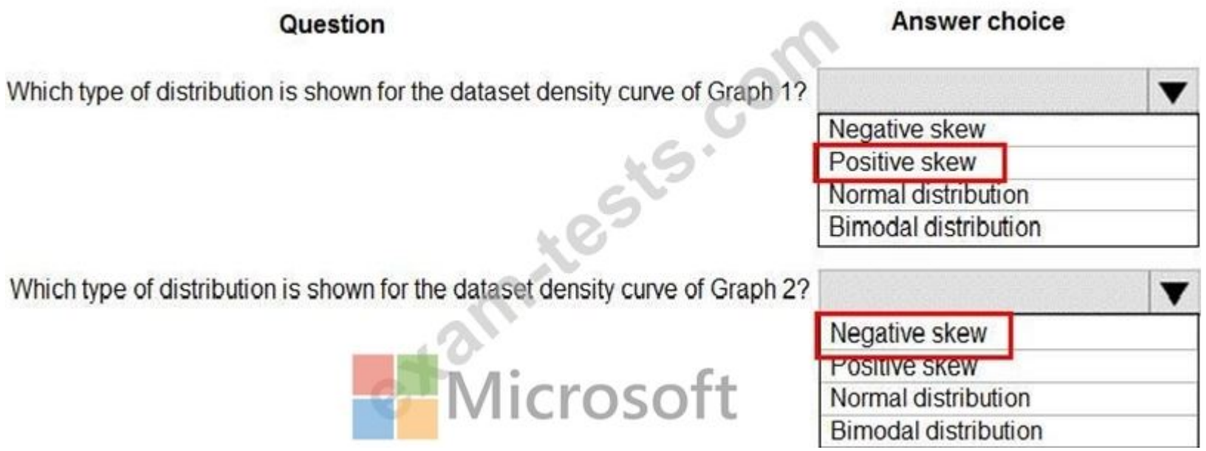


Use the drop-down menus to select the answer choice that answers each question based on the information presented in the graphic.

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

Answer:



Box 1: Positive skew

Reference:

https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/compute-elementary-statistics

NEW QUESTION: 93

You manage an Azure Machine Learning workspace. You plan to import data from Azure Data Lake Storage Gen2. You need to build a URI that represents the storage location. Which protocol should you use?

A. adl

B. https

C. abfss

D. wasbs

Answer: C (LEAVE A REPLY)

NEW QUESTION: 94

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

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

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

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

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

You need to configure the automated machine learning run.

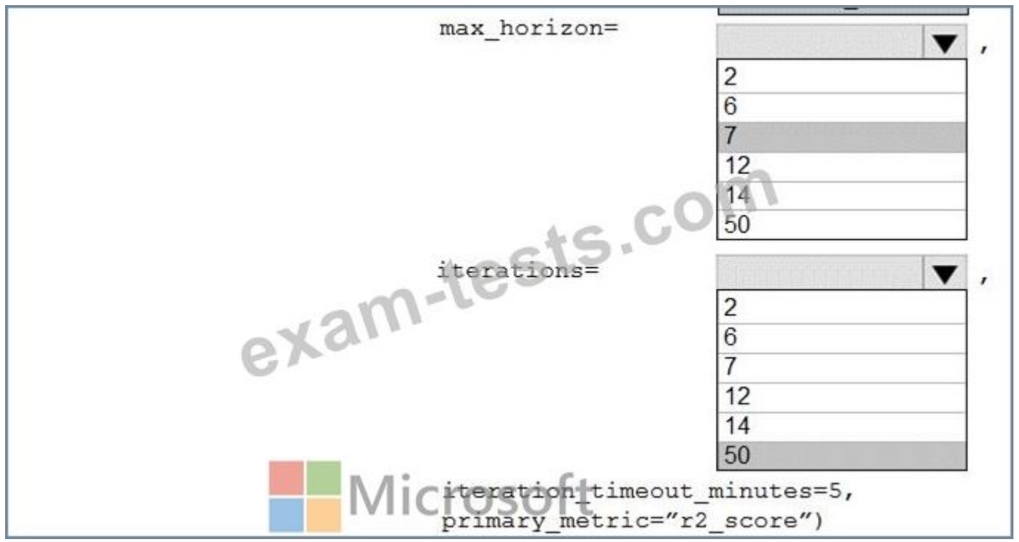
How should you complete the AutoMLConfig definition? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.

```
automl_config = AutoMLConfig(task="  [regression / forecasting / classification / deep learning]  ",
                             training_data=weather_df,
                             label_column_name="  [humidity / pressure / visibility / temperature / days_since_last / observation_time]  ",
                             time_column_name="  [humidity / pressure / visibility / temperature / days_since_last / observation_time]  ",
                             max_horizon=  [2 / 6 / 7 / 12 / 14 / 50]  ,
                             iterations=  [2 / 6 / 7 / 12 / 14 / 50]  ,
                             iteration_timeout_minutes=5,
                             primary_metric="r2_score")
```


Answer:

```
utoml_config = AutoMLConfig(task="  [regression | forecasting | classification | deep learning]  "
                            training_data=weather_df,
                            label_column_name="  [humidity | pressure | visibility | temperature | days_since_last | observation_time]  ",
                            time_column_name="  [humidity | pressure | visibility | temperature | days_since_last | observation_time]  ",
                            max_horizon=  [2 | 6 | 7 | 12 | 14 | 50]  ,
                            iterations=  [2 | 6 | 7 | 12 | 14 | 50]  ,
                            iteration_timeout_minutes=5,
                            primary_metric="r2_score")
```



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

NEW QUESTION: 95
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

Answer:

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

).

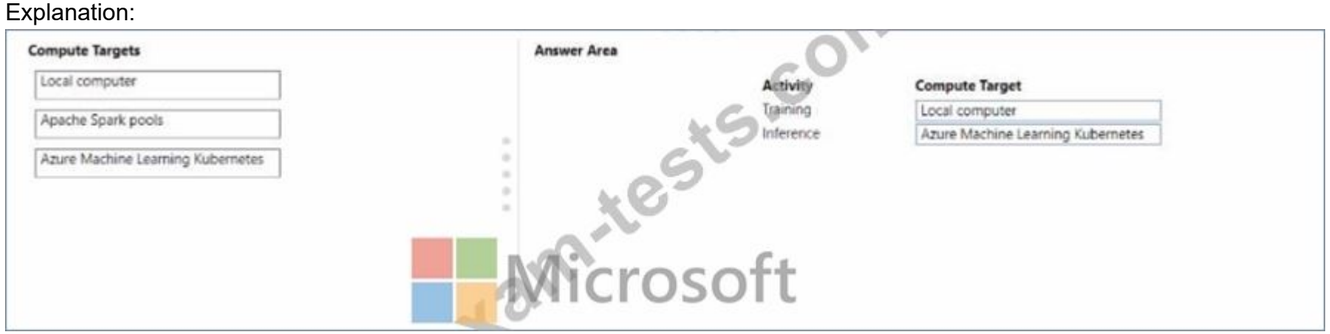
Reference:

https://medium.com/comet-ml/organizing-machine-learning-projects-project-management-guidelines-2d2b85651bbd

NEW QUESTION: 96
You are designing an Azure Machine Leaning solution by using the Python SDK v2.
You must train and deploy the solution by using a compute target. The compute target must meet the following requirements:
* Enable the use of on-premises compute resources.
* Support autoscalling.
You need to configure a compute target for training and inference.
Which compute target t should you configure?
To answer select the appropriate options in the answer area.
NOTE: Each correct selection is worth one point.

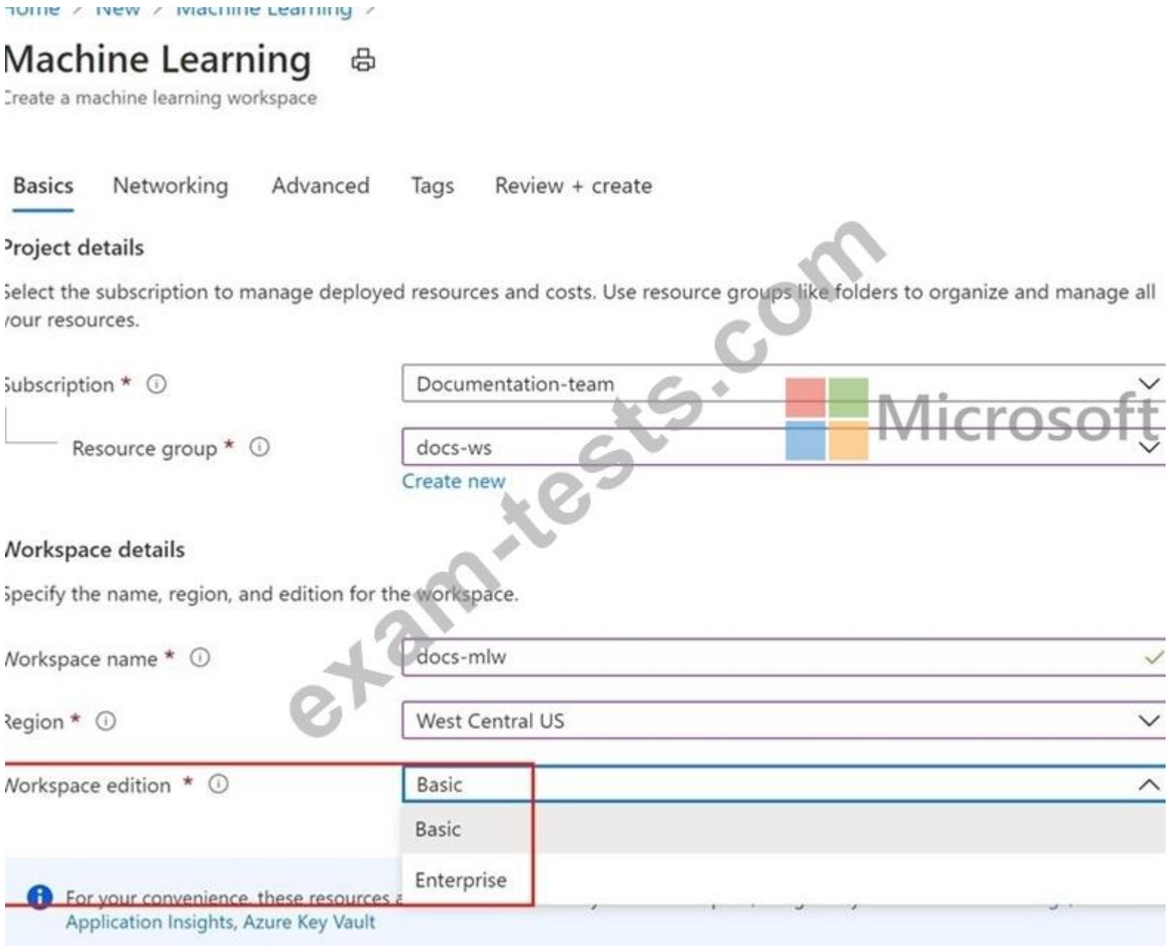
Answer:

Explanation:



NEW QUESTION: 97
You create a new Azure subscription. No resources are provisioned in the subscription.

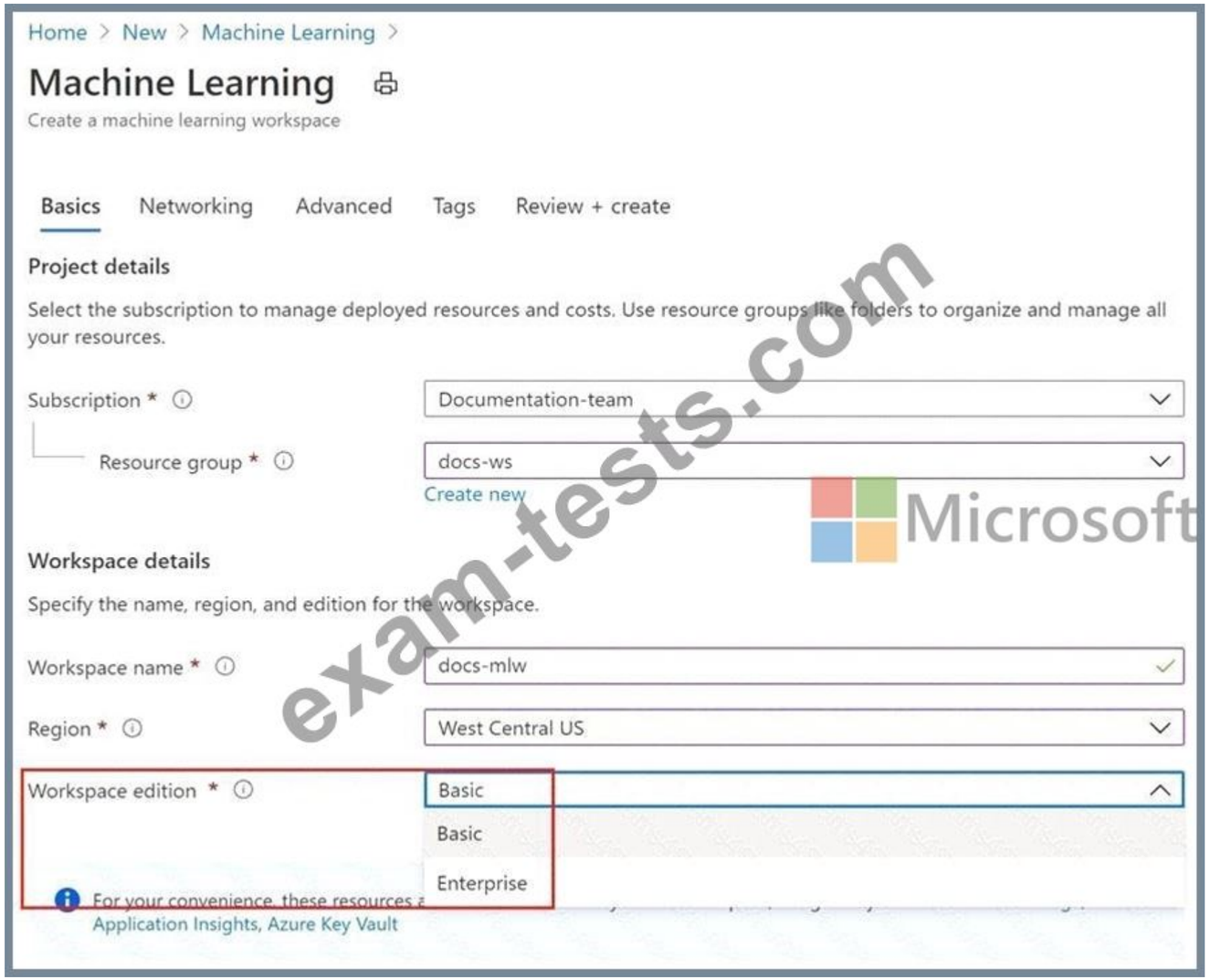
You need to create an Azure Machine Learning workspace.
What are three possible ways to achieve this goal? Each correct answer presents a complete solution.
NOTE: Each correct selection is worth one point.
A. Run Python code that uses the Azure ML SDK library and calls the Workspace.create method with name, subscription_id, resource_group, and location parameters.
B. Use an Azure Resource Management template that includes a Microsoft.MachineLearningServices/ workspaces resource and its dependencies.
C. Use the Azure Command Line Interface (CLI) with the Azure Machine Learning extension to call the az group create function with --name and --location parameters, and then the az ml workspace create function, specifying -w and -g parameters for the workspace name and resource group.
D. Navigate to Azure Machine Learning studio and create a workspace.
E. Run Python code that uses the Azure ML SDK library and calls the Workspace.get method with name, subscription_id, and resource_group parameters.
Answer: B,C,D (LEAVE A REPLY)
B: You can use an Azure Resource Manager template to create a workspace for Azure Machine Learning.
Example:
{"type": "Microsoft.MachineLearningServices/workspaces",
...
C: You can create a workspace for Azure Machine Learning with Azure CLI Install the machine learning extension.
Create a resource group: az group create --name <resource-group-name> --location <location> To create a new workspace where the services are automatically created, use the following command: az ml workspace create -w <workspace-name> -g <resource-group-name> D: You can create and manage Azure Machine Learning workspaces in the Azure portal.
Sign in to the Azure portal by using the credentials for your Azure subscription.
In the upper-left corner of Azure portal, select + Create a resource.
Use the search bar to find Machine Learning.
Select Machine Learning.
In the Machine Learning pane, select Create to begin.



Reference:
https://docs.microsoft.com/en-us/azure/machine-learning/how-to-create-workspace-template
https://docs.microsoft.com/en-us/azure/machine-learning/how-to-manage-workspace-cli
https://docs.microsoft.com/en-us/azure/machine-learning/how-to-manage-workspace

NEW QUESTION: 98
You create a new Azure subscription. No resources are provisioned in the subscription.
You need to create an Azure Machine Learning workspace.
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. Run Python code that uses the Azure ML SDK library and calls the Workspace.create method with name, subscription_id, resource_group, and location parameters.
B. Use an Azure Resource Management template that includes a Microsoft.MachineLearningServices/ workspaces resource and its dependencies.
C. Use the Azure Command Line Interface (CLI) with the Azure Machine Learning extension to call the az group create function with --name and --location parameters, and then the az ml workspace create function, specifying -w and -g parameters for the workspace name and resource group.

D. Navigate to Azure Machine Learning studio and create a workspace.
E. Run Python code that uses the Azure ML SDK library and calls the Workspace.get method with name, subscription_id, and resource_group parameters.
Answer: B,C,D (LEAVE A REPLY)
Explanation
B: You can use an Azure Resource Manager template to create a workspace for Azure Machine Learning.
Example:
{"type": "Microsoft.MachineLearningServices/workspaces",
...
C: You can create a workspace for Azure Machine Learning with Azure CLI Install the machine learning extension.
Create a resource group: az group create --name <resource-group-name> --location <location> To create a new workspace where the services are automatically created, use the following command: az ml workspace create -w <workspace-name> -g <resource-group-name> D: You can create and manage Azure Machine Learning workspaces in the Azure portal.
Sign in to the Azure portal by using the credentials for your Azure subscription.
In the upper-left corner of Azure portal, select + Create a resource.
Use the search bar to find Machine Learning.
Select Machine Learning.
In the Machine Learning pane, select Create to begin.



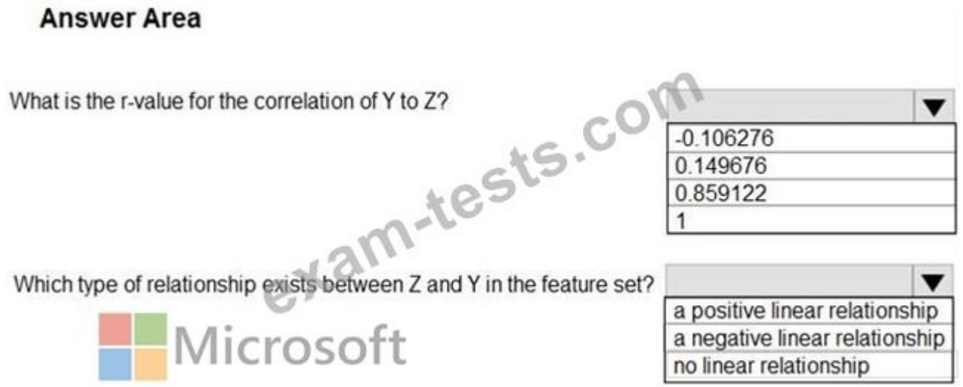
Reference:
https://docs.microsoft.com/en-us/azure/machine-learning/how-to-create-workspace-template
https://docs.microsoft.com/en-us/azure/machine-learning/how-to-manage-workspace-cli
https://docs.microsoft.com/en-us/azure/machine-learning/how-to-manage-workspace

NEW QUESTION: 99
Hotspot Question
You have a feature set containing the following numerical features: X, Y, and Z.
The Poisson correlation coefficient (r-value) of X, Y, and Z features is shown in the following image:

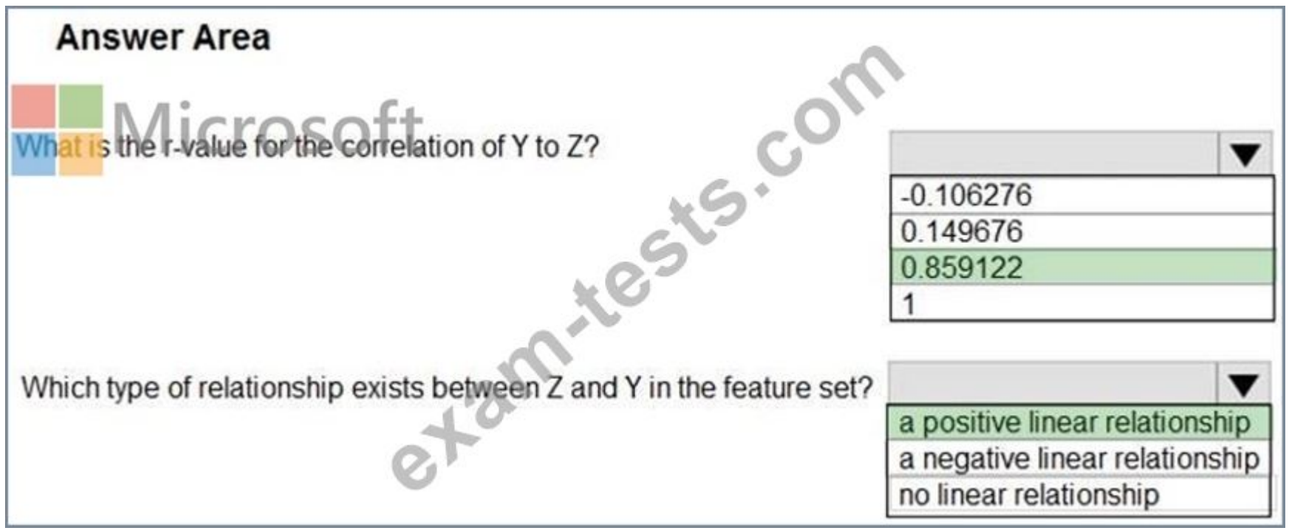
	X	Y	Z
X	1	0.149676	-0.106276
Y	0.149676	1	0.859122
Z	-0.106276	0.859122	1

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

NOTE: Each correct selection is worth one point.



Answer:



Explanation:

Box 1: 0.859122

Box 2: a positively linear relationship

+1 indicates a strong positive linear relationship

-1 indicates a strong negative linear correlation

0 denotes no linear relationship between the two variables.

References:

https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/compute- linear-correlation

NEW QUESTION: 100

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 (LEAVE A REPLY)
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)
Incorrect Answers:

A: Pandas DataFrame not used.
Reference:
https://docs.microsoft.com/es-es/azure/machine-learning/how-to-train-with-datasets

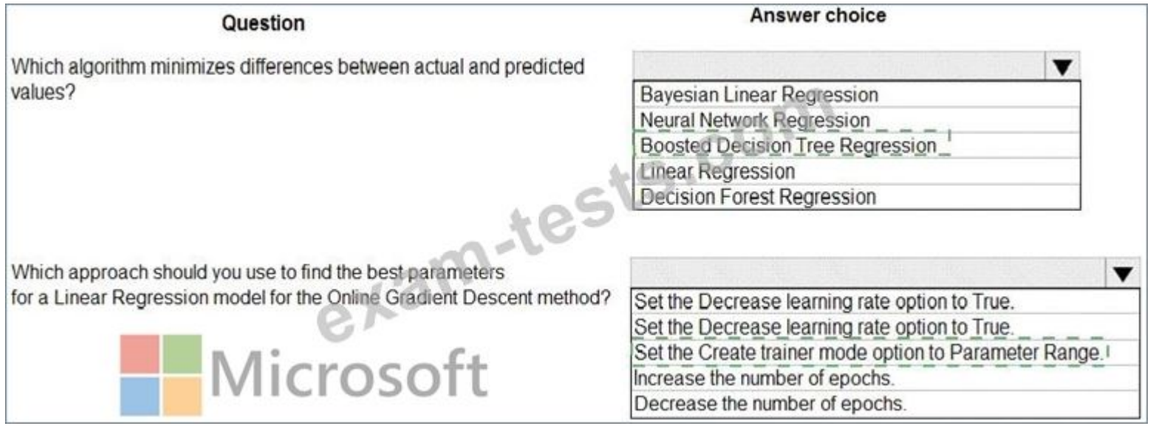
NEW QUESTION: 101
You are developing a linear regression model in Azure Machine Learning Studio. You run an experiment to compare different algorithms.
The following image displays the results dataset output:

Algorithm	Mean Absolute Error	Root Mean Squared Error	Relative Absolute Error	Relative Squared Error
Bayesian Liner	3.276025	4.655442	0.511436	0.282138
Neural Network	2.676538	3.621476	0.417847	0.17073
Boosted Decision Tree	2.168847	2.878077	0.338589	0.107831
Linear	6.350005	8.720718	0.99133	0.99002
Decision Forest	2.390206	3.315 164	0.373146	0.14307

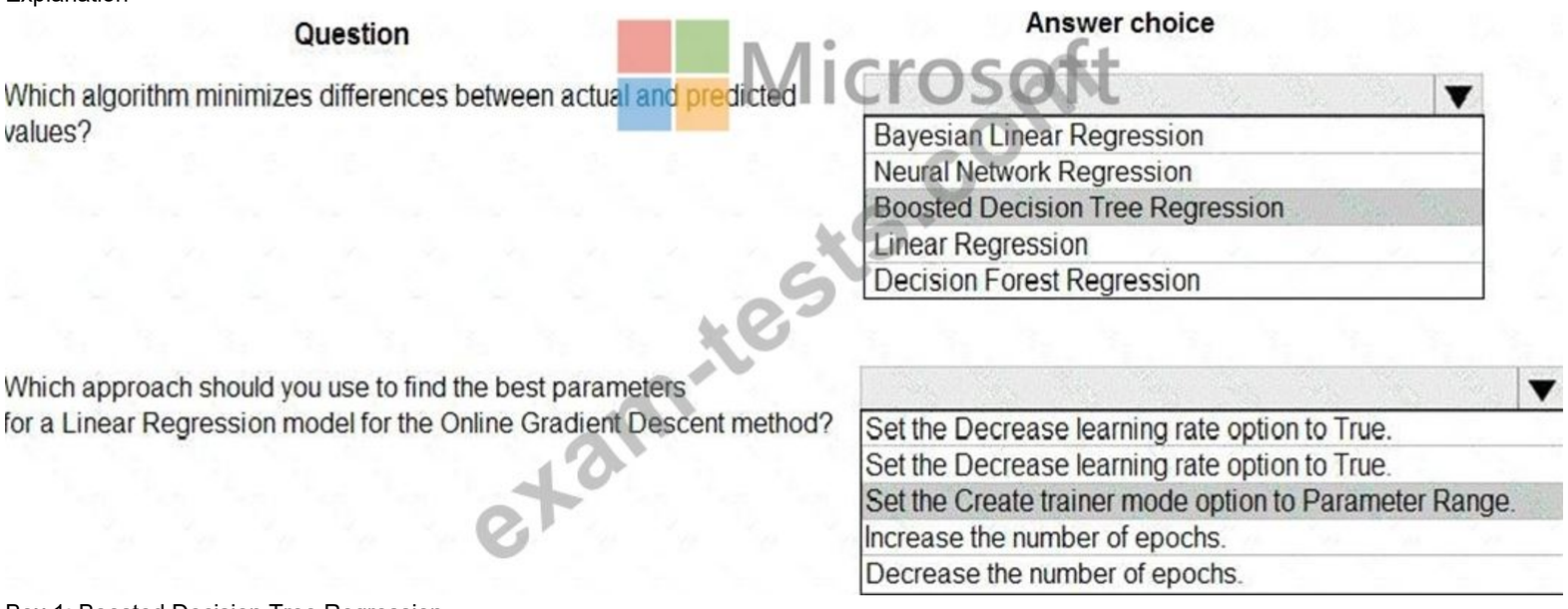
Use the drop-down menus to select the answer choice that answers each question based on the information presented in the image.
NOTE: Each correct selection is worth one point.

Question	Answer choice
Which algorithm minimizes differences between actual and predicted values?	Bayesian Linear Regression Neural Network Regression Boosted Decision Tree Regression Linear Regression Decision Forest Regression
Which approach should you use to find the best parameters for a Linear Regression model for the Online Gradient Descent method?	Set the Decrease learning rate option to True. Set the Decrease learning rate option to True. Set the Create trainer mode option to Parameter Range. Increase the number of epochs. Decrease the number of epochs.

Answer:



Explanation



Box 1: Boosted Decision Tree Regression

Mean absolute error (MAE) measures how close the predictions are to the actual outcomes; thus, a lower score is better.

Box 2:

Online Gradient Descent: If you want the algorithm to find the best parameters for you, set Create trainer mode option to Parameter Range. You can then specify multiple values for the algorithm to try.

References:

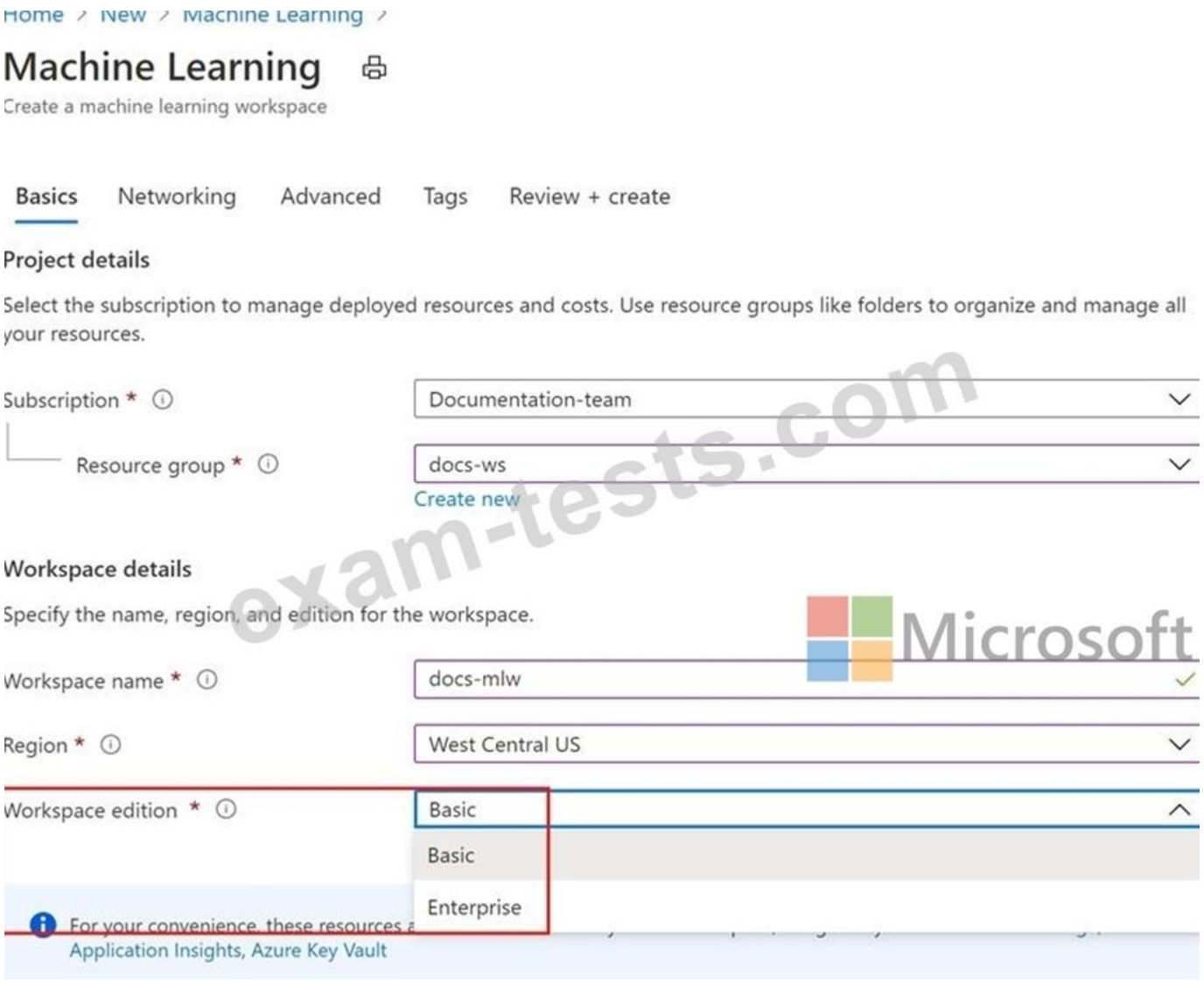
https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/evaluate-model

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

NEW QUESTION: 102

You create a new Azure subscription. No resources are provisioned in the subscription.

You need to create an Azure Machine Learning workspace.
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. Run Python code that uses the Azure ML SDK library and calls the Workspace.create method with name, subscription_id, resource_group, and location parameters.
B. Use an Azure Resource Management template that includes a Microsoft.MachineLearningServices
/workspaces resource and its dependencies.
C. Use the Azure Command Line Interface (CLI) with the Azure Machine Learning extension to call the azgroup create function with --name and --location parameters, and then the az ml workspace createfunction, specifying -w and -g parameters for the workspace name and resource group.
D. Navigate to Azure Machine Learning studio and create a workspace.
E. Run Python code that uses the Azure ML SDK library and calls the Workspace.get method with name, subscription_id, and resource_group parameters.
Answer: (SHOW ANSWER)
B: You can use an Azure Resource Manager template to create a workspace for Azure Machine Learning.
Example:
{"type": "Microsoft.MachineLearningServices/workspaces",
...
C: You can create a workspace for Azure Machine Learning with Azure CLI Install the machine learning extension.
Create a resource group: az group create --name <resource-group-name> --location <location> To create a new workspace where the services are automatically created, use the following command: az ml workspace create -w <workspace-name> -g <resource-group-name> D: You can create and manage Azure Machine Learning workspaces in the Azure portal.
Sign in to the Azure portal by using the credentials for your Azure subscription.
In the upper-left corner of Azure portal, select + Create a resource.
Use the search bar to find Machine Learning.
Select Machine Learning.
In the Machine Learning pane, select Create to begin.



Reference:
https://docs.microsoft.com/en-us/azure/machine-learning/how-to-create-workspace-template
https://docs.microsoft.com/en-us/azure/machine-learning/how-to-manage-workspace-cli
https://docs.microsoft.com/en-us/azure/machine-learning/how-to-manage-workspace

NEW QUESTION: 103
You create a batch inference pipeline by using the Azure ML SDK. You run the pipeline by using the following code:
from azureml.pipeline.core import Pipeline
from azureml.core.experiment import Experiment
pipeline = Pipeline(workspace=ws, steps=[parallelrun_step])

pipeline_run = Experiment(ws, 'batch_pipeline').submit(pipeline)

You need to monitor the progress of the pipeline execution.

What are two possible ways to achieve this goal? Each correct answer presents a complete solution.

NOTE: Each correct selection is worth one point.

A. Run the following code in a notebook:

```
from azureml.contrib.interpret.explanation.explanation_client import ExplanationClient
client = ExplanationClient.from_run(pipeline_run)
explanation = client.download_model_explanation()
explanation = client.download_model_explanation(top_k=4)
global_importance_values = explanation.get_ranked_global_values()
global_importance_names = explanation.get_ranked_global_names()
print('global importance values: {}'.format(global_importance_values))
print('global importance names: {}'.format(global_importance_names))
```

B. Use the Inference Clusters tab in Machine Learning Studio.

C. Use the Activity log in the Azure portal for the Machine Learning workspace.

D. Run the following code in a notebook:

```
from azureml.widgets import RunDetails
RunDetails(pipeline_run).show()
```

E. Run the following code and monitor the console output from the PipelineRun object:

```
pipeline_run.wait_for_completion(show_output=True)
```

A. Option A

B. Option B

C. Option C

D. Option D

E. Option E

Answer: D,E (LEAVE A REPLY)

A batch inference job can take a long time to finish. This example monitors progress by using a Jupyter widget. You can also manage the job's progress by using:

Azure Machine Learning Studio.

Console output from the PipelineRun object.

from azureml.widgets import RunDetails

RunDetails(pipeline_run).show()

pipeline_run.wait_for_completion(show_output=True)

Reference:

https://docs.microsoft.com/en-us/azure/machine-learning/how-to-use-parallel-run-step#monitor-the-parallel- run-job

NEW QUESTION: 104

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

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

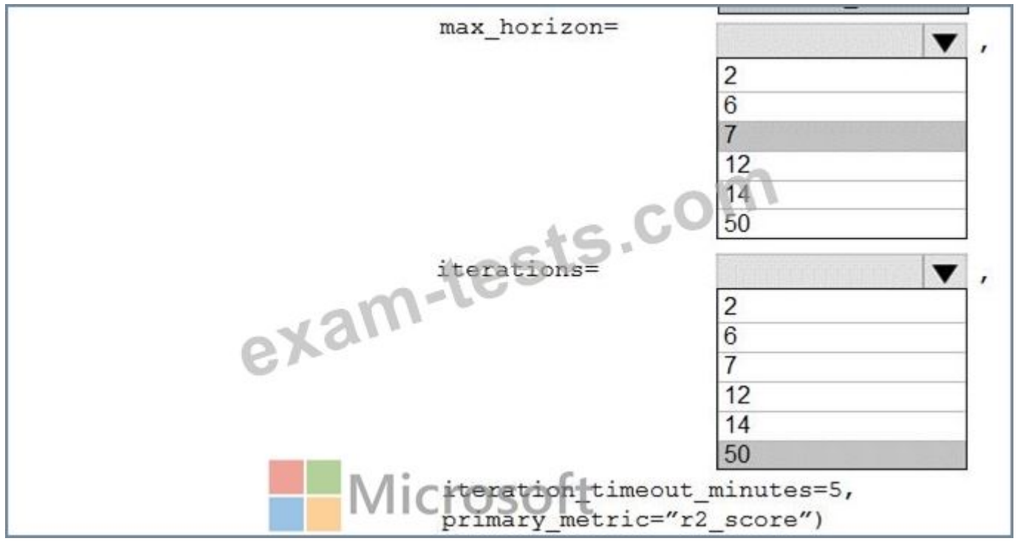
The data is collected every 12 hours: noon and midnight.

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

You must use the Azure Machine Learning SDK to run an automated machine learning experiment to train these models.
You need to configure the automated machine learning run.
How should you complete the AutoMLConfig definition? To answer, select the appropriate options in the answer area.
NOTE: Each correct selection is worth one point.

```
automl_config = AutoMLConfig(task="  [ ▼ ]  ",
                             training_data=weather_df,
                             label_column_name="  [ ▼ ]  ",
                             time_column_name="  [ ▼ ]  ",
                             max_horizon=  [ ▼ ] ,
                             iterations=  [ ▼ ] ,
                             iteration_timeout_minutes=5,
                             primary_metric="r2_score")
```

- task: regression, forecasting, classification, deep learning
- label_column_name: humidity, pressure, visibility, temperature, days_since_last, observation_time
- time_column_name: humidity, pressure, visibility, temperature, days_since_last, observation_time
- max_horizon: 2, 6, 7, 12, 14, 50
- iterations: 2, 6, 7, 12, 14, 50


max_horizon=
,
2
6
7
12
14
50
iterations=
,
2
6
7
12
14
50
iteration_timeout_minutes=5,
primary_metric="r2_score")

Answer:

```
automl_config = AutoMLConfig(task="  [regression / forecasting / classification / deep learning]  "
                             training_data=weather_df,
                             label_column_name="  [humidity / pressure / visibility / temperature / days_since_last / observation_time]  ",
                             time_column_name="  [humidity / pressure / visibility / temperature / days_since_last / observation_time]  ",
                             max_horizon=  [2 / 6 / 7 / 12 / 14 / 50]  ,
                             iterations=  [2 / 6 / 7 / 12 / 14 / 50]  ,
                             iteration_timeout_minutes=5,
                             primary_metric="r2_score")
```

Explanation:
Box 1: forcasting
Task: The type of task to run. Values can be 'classification', 'regression', or 'forecasting' depending on the type of automated ML problem to solve.
Box 2: temperature
The training data to be used within the experiment. It should contain both training features and a label column (optionally a sample weights column).
Box 3: observation_time
time_column_name: The name of the time column. This parameter is required when forecasting to specify the datetime column in the input data used for building the time series and inferring its frequency. This setting is being deprecated. Please use forecasting_parameters instead.
Box 4: 7
"predicts temperature over the next seven days"
max_horizon: The desired maximum forecast horizon in units of time-series frequency. The default value is 1.
Units are based on the time interval of your training data, e.g., monthly, weekly that the forecaster should predict out. When task type is forecasting, this parameter is required.
Box 5: 50
"For the initial round of training, you want to train a maximum of 50 different models." Iterations: The total number of different algorithm and parameter combinations to test during an automated ML experiment.
Reference:
https://docs.microsoft.com/en-us/python/api/azureml-train-automl-client/azureml.train.automl.automlconfig.automlconfig

NEW QUESTION: 105
You are preparing to use the Azure ML SDK to run an experiment and need to create compute. You run the following code:

```
from azureml.core.compute import ComputeTarget, AmlCompute
from azureml.core.compute_target import ComputeTargetException
ws = Workspace.from_config()
cluster_name = 'aml-cluster'
try:
  training_compute = ComputeTarget(workspace=ws, name=cluster_name)
except ComputeTargetException:
  compute_config = AmlCompute.provisioning_configuration(vm_size='STANDARD_D2_V2', vm_priority='lowpriority',
max_nodes=4)
  training_compute = ComputeTarget.create(ws, cluster_name, compute_config)
  training_compute.wait_for_completion(show_output=True)
```


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

	Yes	No
If a training cluster named aml-cluster already exists in the workspace, it will be deleted and replaced.	○	○
The `wait_for_completion()` method will not return until the aml-cluster compute has four active nodes.	○	○
If the code creates a new aml-cluster compute target, it may be preempted due to capacity constraints.	○	○
The aml-cluster compute target is deleted from the workspace after the training experiment completes.	○	○

Answer:

	Yes	No
If a training cluster named aml-cluster already exists in the workspace, it will be deleted and replaced.	○	●
The `wait_for_completion()` method will not return until the aml-cluster compute has four active nodes.	●	○
If the code creates a new aml-cluster compute target, it may be preempted due to capacity constraints.	●	○
The aml-cluster compute target is deleted from the workspace after the training experiment completes.	○	●

Explanation:

Box 1: No

If a training cluster already exists it will be used.

Box 2: Yes

The wait_for_completion method waits for the current provisioning operation to finish on the cluster.

Box 3: Yes

Low Priority VMs use Azure's excess capacity and are thus cheaper but risk your run being pre-empted.

Box 4: No

Need to use training_compute.delete() to deprovision and delete the AmlCompute target.

Reference:

https://notebooks.azure.com/azureml/projects/azureml-getting-started/html/how-to-use-azureml/training/train-on-amlcompute/train-on-amlcompute.ipynb

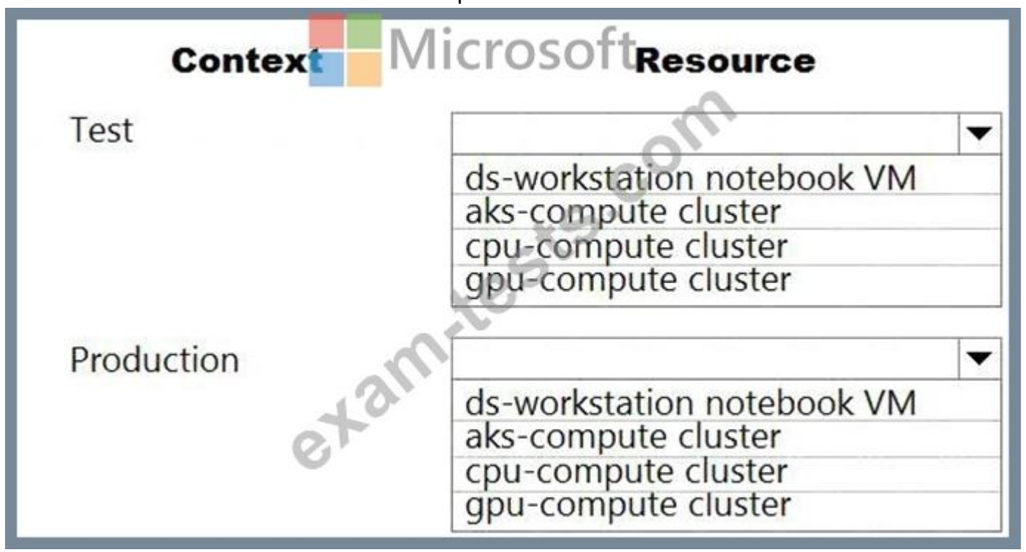
https://docs.microsoft.com/en-us/python/api/azureml-core/azureml.core.compute.computetarget

NEW QUESTION: 106

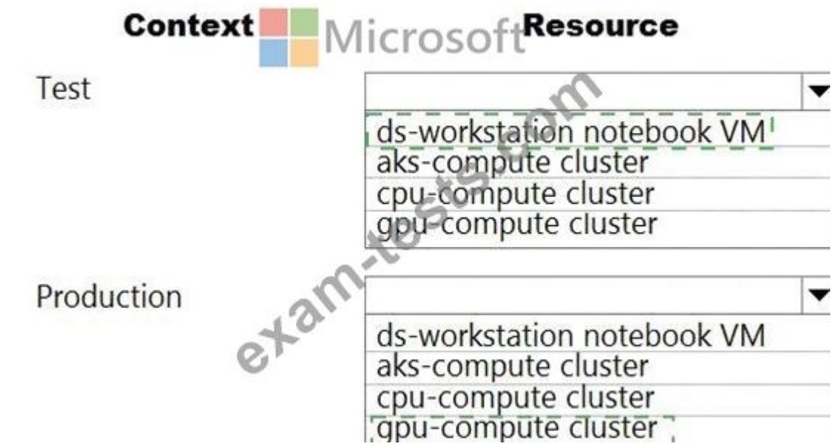
You are a lead data scientist for a project that tracks the health and migration of birds. You create a multi-image classification deep learning model that uses a set of labeled bird photos collected by experts. You plan to use the model to develop a cross-platform mobile app that predicts the species of bird captured by app users.

You must test and deploy the trained model as a web service. The deployed model must meet the following requirements:

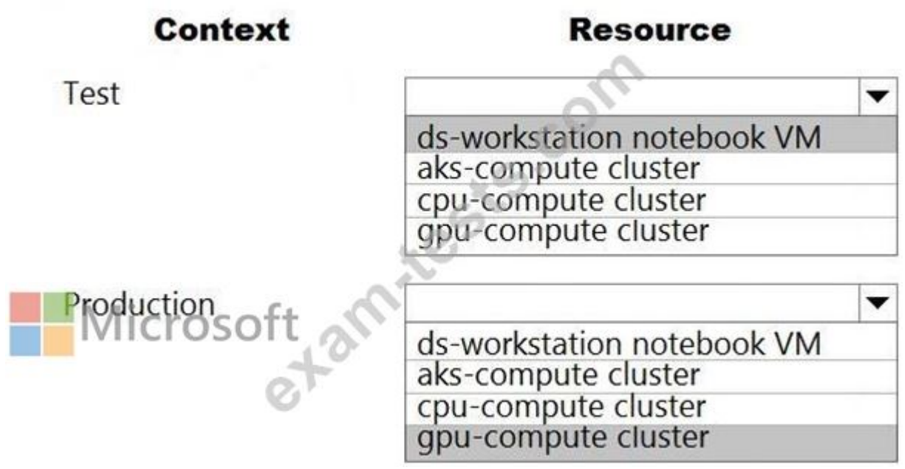
An authenticated connection must not be required for testing.
The deployed model must perform with low latency during inferencing.
The REST endpoints must be scalable and should have a capacity to handle large number of requests when multiple end users are using the mobile application.
You need to verify that the web service returns predictions in the expected JSON format when a valid REST request is submitted.
Which compute resources should you use? To answer, select the appropriate options in the answer area.
NOTE: Each correct selection is worth one point.



Answer:



Explanation



Box 1: ds-workstation notebook VM

An authenticated connection must not be required for testing.

On a Microsoft Azure virtual machine (VM), including a Data Science Virtual Machine (DSVM), you create local user accounts while provisioning the VM. Users then authenticate to the VM by using these credentials.

Box 2: gpu-compute cluster

Image classification is well suited for GPU compute clusters

Reference:

https://docs.microsoft.com/en-us/azure/machine-learning/data-science-virtual-machine/dsvm-common-identity

https://docs.microsoft.com/en-us/azure/architecture/reference-architectures/ai/training-deep-learning

NEW QUESTION: 107

Your team is building a data engineering and data science development environment.

The environment must support the following requirements:

* support Python and Scala
* compose data storage, movement, and processing services into automated data pipelines
* the same tool should be used for the orchestration of both data engineering and data science
* support workload isolation and interactive workloads
* enable scaling across a cluster of machines

You need to create the environment.

What should you do?

A. Build the environment in Apache Hive for HDInsight and use Azure Data Factory for orchestration.

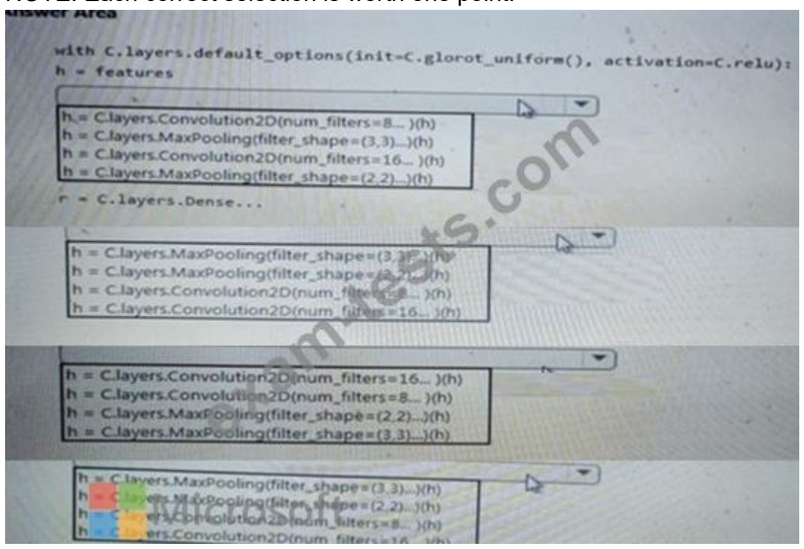
B. Build the environment in Azure Databricks and use Azure Data Factory for orchestration.

C. Build the environment in Apache Spark for HDInsight and use Azure Container Instances for orchestration.

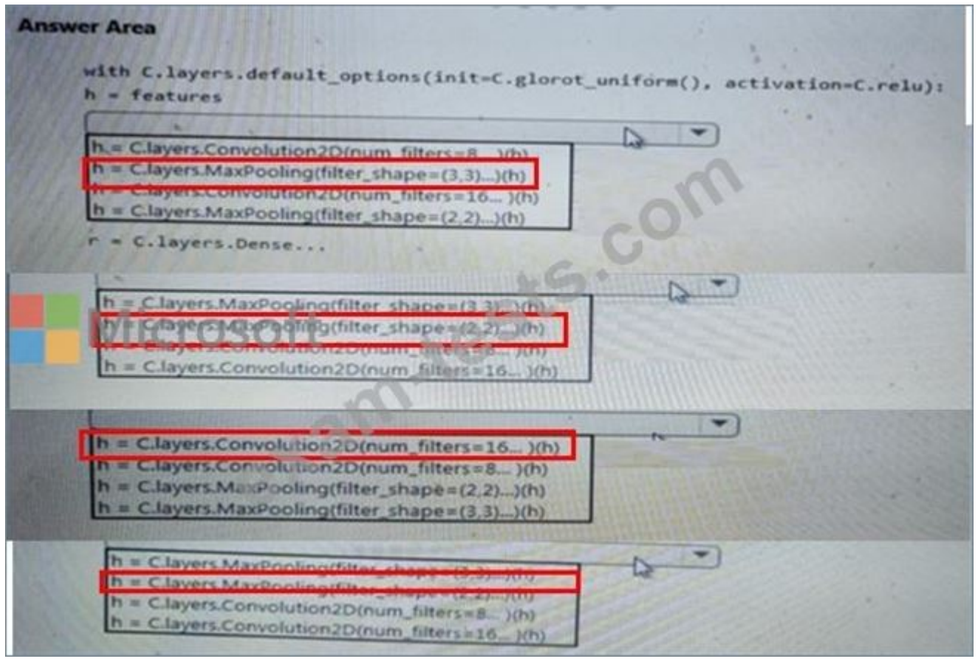
D. Build the environment in Azure Databricks and use Azure Container Instances for orchestration.
Answer: B (LEAVE A REPLY)
In Azure Databricks, we can create two different types of clusters.
* Standard, these are the default clusters and can be used with Python, R, Scala and SQL
* High-concurrency
Azure Databricks is fully integrated with Azure Data Factory.
Incorrect Answers:
D: Azure Container Instances is good for development or testing. Not suitable for production workloads.
References:
https://docs.microsoft.com/en-us/azure/architecture/data-guide/technology-choices/data-science-and-machinelearning

NEW QUESTION: 108
You build a data pipeline in an Azure Machine Learning workspace by using the Azure Machine Learning SDK for Python.
You need to run a Python script as a pipeline step.
Which two classes could you use? Each correct answer presents a complete solution.
NOTE: Each correct selection is worth one point.
A. PythonScriptStep
B. AutoMLStep
C. StepRun
D. CommandStep
Answer: A,D (LEAVE A REPLY)

NEW QUESTION: 109
You need to build a feature extraction strategy for the local models.
How should you complete the code segment? To answer, select the appropriate options in the answer area.
NOTE: Each correct selection is worth one point.



Answer:



NEW QUESTION: 110

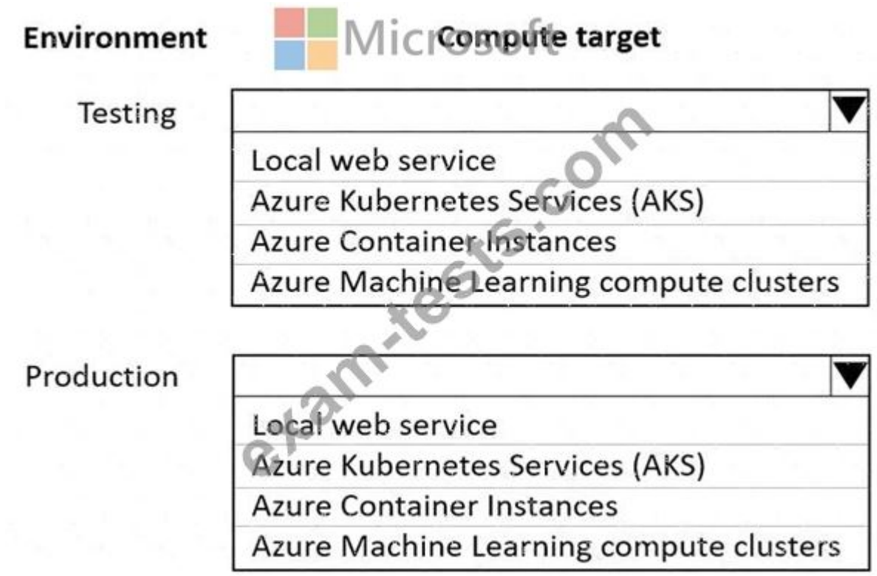
You are using an Azure Machine Learning workspace. You set up an environment for model testing and an environment for production.

The compute target for testing must minimize cost and deployment efforts. The compute target for production must provide fast response time, autoscaling of the deployed service, and support real-time inferencing.

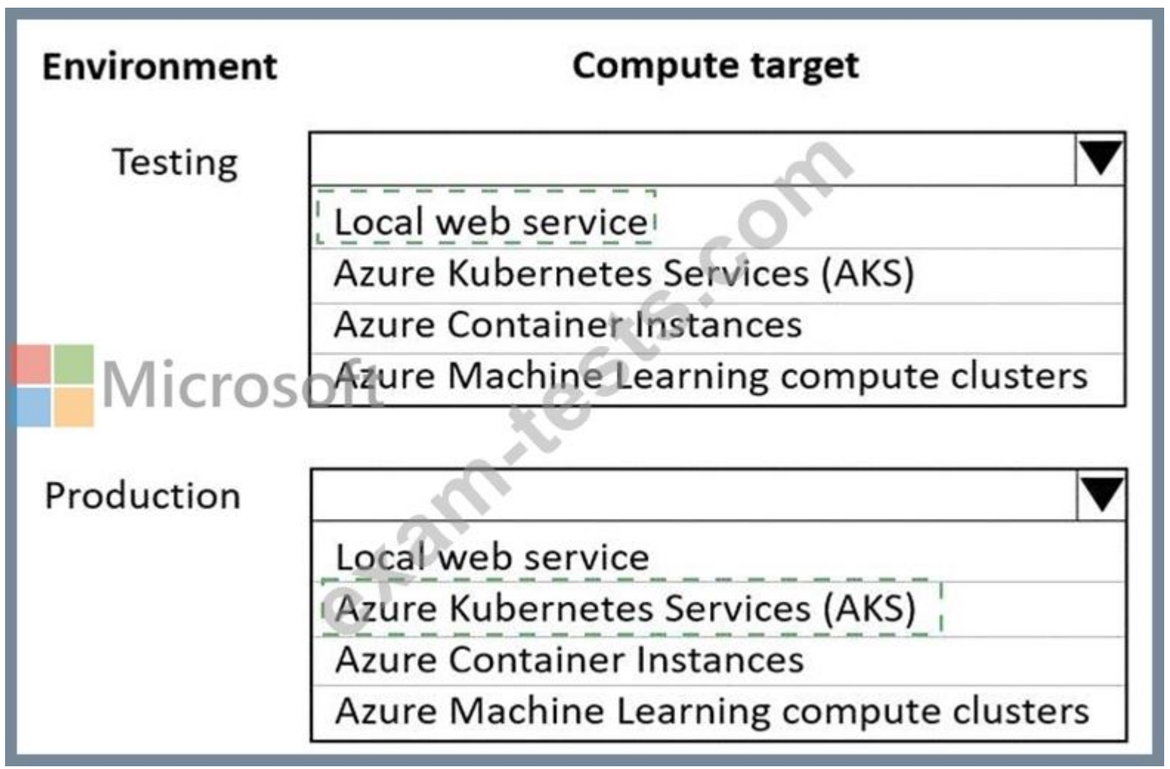
You need to configure compute targets for model testing and production.

Which compute targets should you use? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.

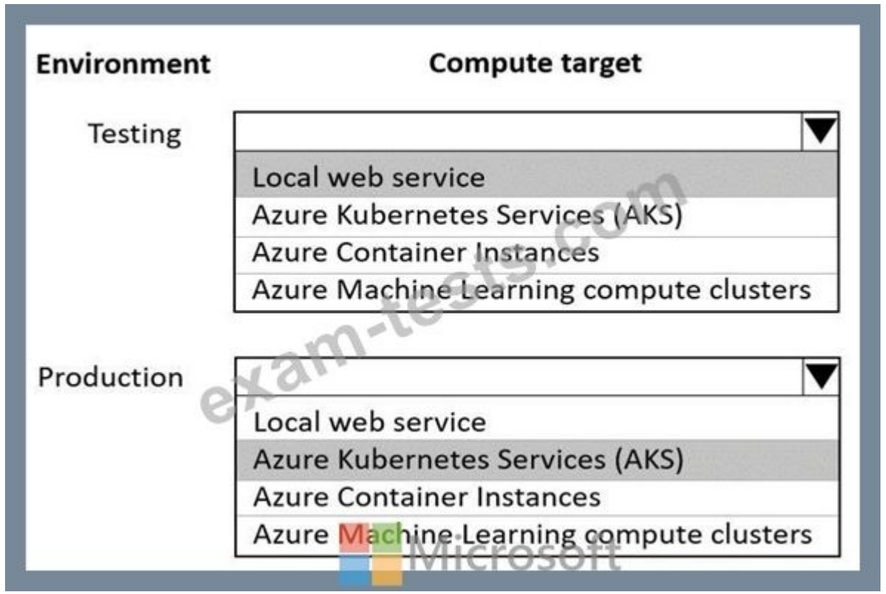


Answer:



Explanation

Text, application Description automatically generated



Box 1: Local web service

The Local web service compute target is used for testing/debugging. Use it for limited testing and troubleshooting. Hardware acceleration depends on use of libraries in the local system.

Box 2: Azure Kubernetes Service (AKS)

Azure Kubernetes Service (AKS) is used for Real-time inference.

Recommended for production workloads.

Use it for high-scale production deployments. Provides fast response time and autoscaling of the deployed service Reference:

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

NEW QUESTION: 111

You manage an Azure Machine Learning workspace. The development environment tor managing the workspace is configured to use Python SDK v2 in Azure Machine Learning Notebooks A Synapse Spark Compute is currently attached and uses system-assigned identity You need to use Python code to update the Synapse Spark Compute 10 use a user-assigned identity.

Solution: Configure the IdentityConfiguration class with the appropriate identity type.

Does the solution meet the goal?

A. Yes

B. No

Answer: (SHOW ANSWER)

NEW QUESTION: 112

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 (LEAVE A REPLY)

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: 113

You create a Python script that runs a training experiment in Azure Machine Learning. The script uses the Azure Machine Learning SDK for Python.

You must add a statement that retrieves the names of the logs and outputs generated by the script.

You need to reference a Python class object from the SDK for the statement.

Which class object should you use?

A. Run

B. ScripcRunConfig

C. Workspace

D. Experiment

Answer: A (LEAVE A REPLY)

A run represents a single trial of an experiment. Runs are used to monitor the asynchronous execution of a trial, log metrics and store output of the trial, and to analyze results and access artifacts generated by the trial.

The run Class get_all_logs method downloads all logs for the run to a directory.

Reference:

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

NEW QUESTION: 114

You have the following Azure subscriptions and Azure Machine Learning service workspaces:

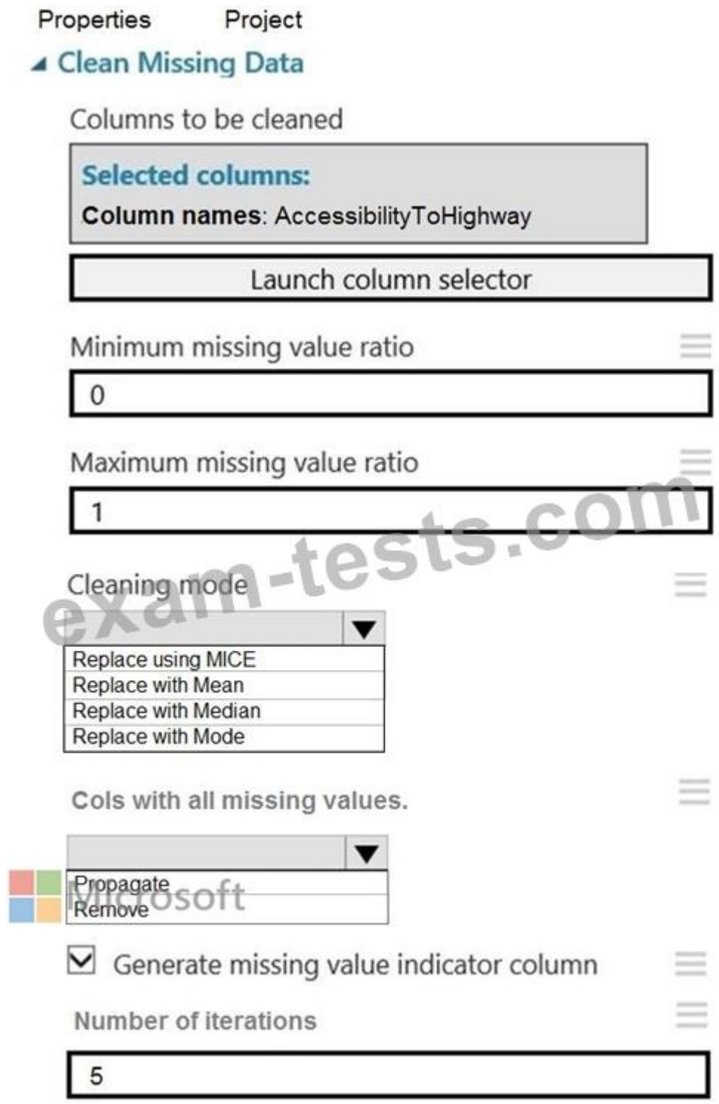
Subscription	Workspace	Comment
385bdfe5-4cef-4ad4-b977-3f86d92727c9	ml-default	This is the default subscription.
5a5891d1-557a-4234-9b83-2e90412b1068	ml-project	The information required to uniquely identify this workspace is stored in the file config.json in the same folder as the Python script.

You need to obtain a reference to the ml-project workspace.
Solution: Run the following Python code:

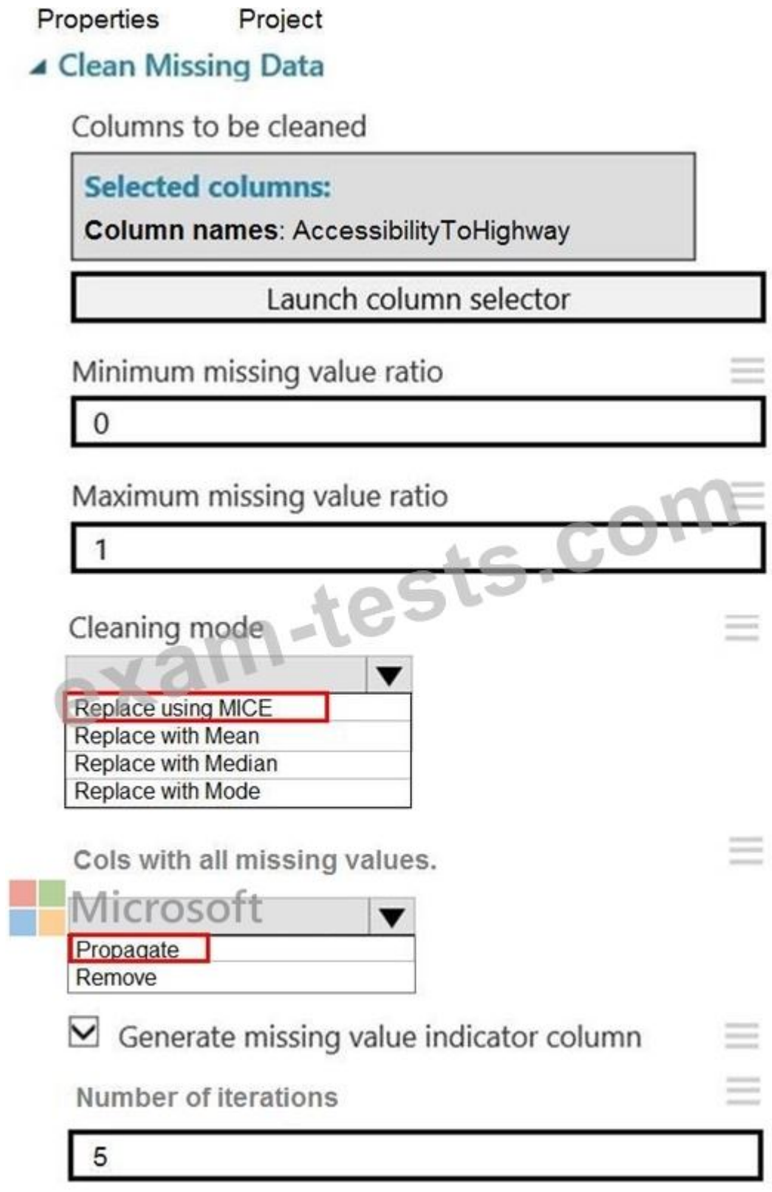
```
from azureml.core import Workspace
ws = Workspace(workspace_name="ml-project")
```

Does the solution meet the goal?
A. Yes
B. No
Answer: A (LEAVE A REPLY)

NEW QUESTION: 115
You need to replace the missing data in the AccessibilityToHighway columns.
How should you configure the Clean Missing Data module? To answer, select the appropriate options in the answer area.
NOTE: Each correct selection is worth one point.



Answer:



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

NEW QUESTION: 116
Your Azure Machine Learning workspace has a dataset named real_estate_data. A sample of the data in the dataset follows.

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

You want to use automated machine learning to find the best regression model for predicting the price column.

You need to configure an automated machine learning experiment using the Azure Machine Learning SDK.

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

NOTE: Each correct selection is worth one point.

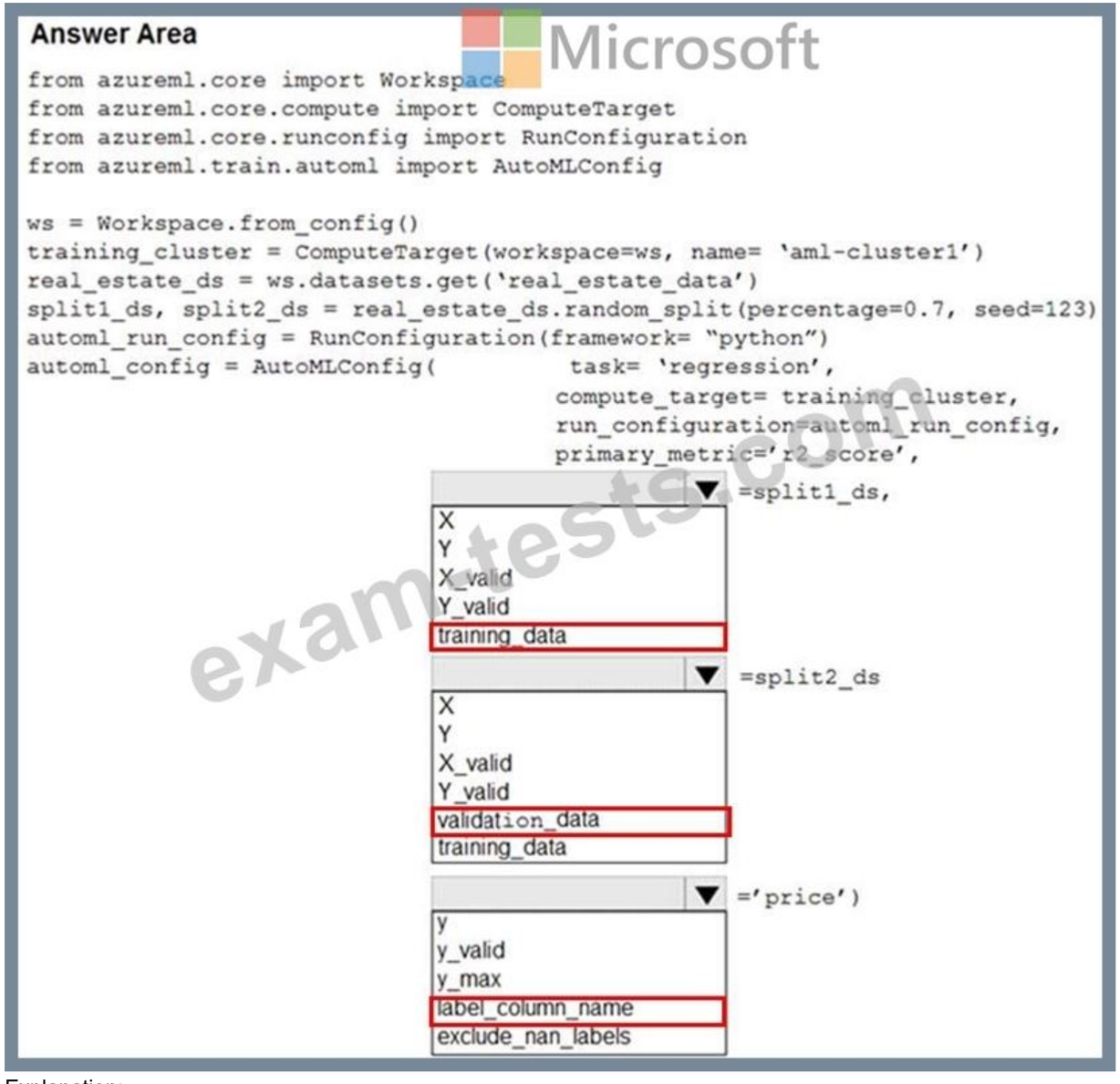
Answer Area

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

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

- split1_ds: X, Y, X_valid, Y_valid, training_data
- split2_ds: X, Y, X_valid, Y_valid, validation_data, training_data
- price: y, y_valid, y_max, label_column_name, exclude_nan_labels

Answer:

Answer Area



Explanation:

Box 1: training_data

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

Box 2: validation_data

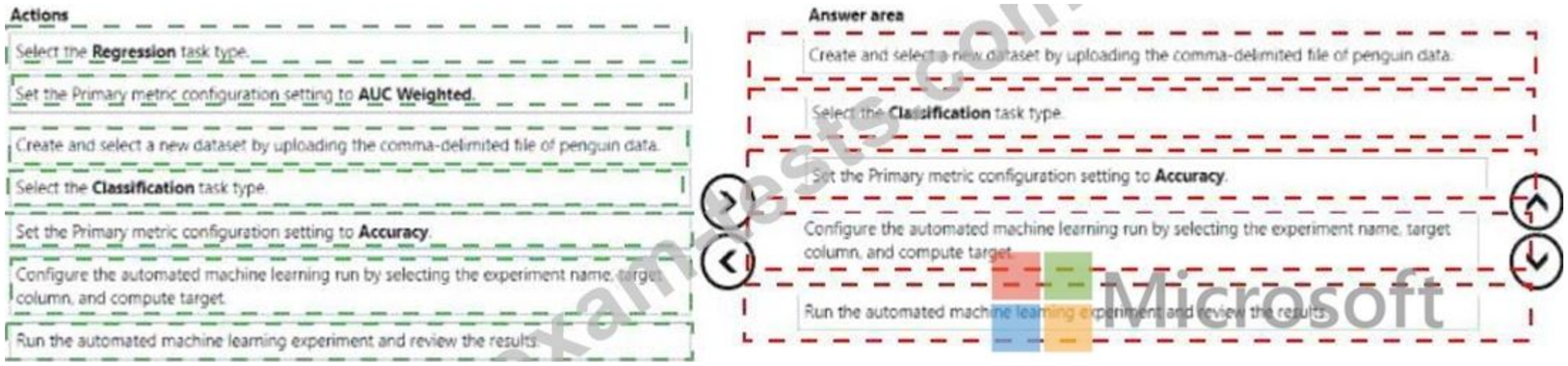
Provide validation data: In this case, you can either start with a single data file and split it into training and validation sets or you can provide a separate data file for the validation set. Either way, the validation_data parameter in your AutoMLConfig object assigns which data to use as your validation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

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

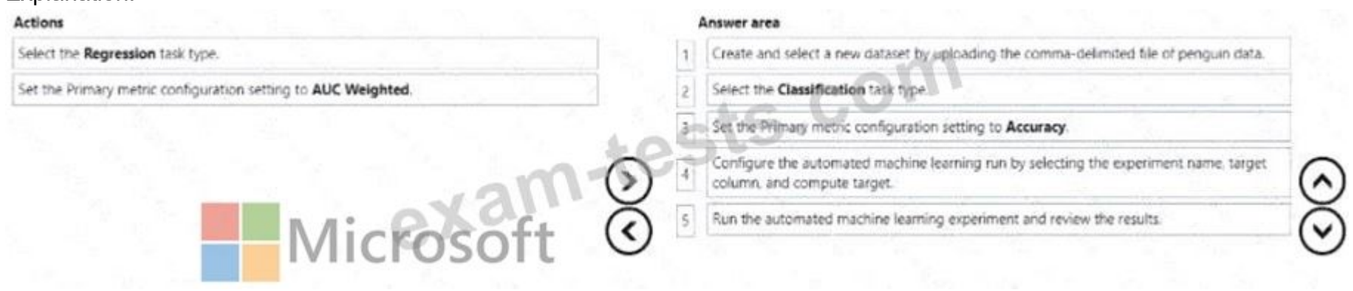
dataset = Dataset.Tabular.from_delimited_files(data)
training_data, validation_data = dataset.random_split(percentage=0.8, seed=1) automl_config = AutoMLConfig(compute_target = aml_remote_compute, task = 'classification', primary_metric = 'AUC_weighted', training_data = training_data, validation_data = validation_data, label_column_name = 'Class') Box 3: label_column_name label_column_name:
The name of the label column. If the input data is from a pandas.DataFrame which doesn't have column names, column indices can be used instead, expressed as integers.
This parameter is applicable to training_data and validation_data parameters.
Incorrect Answers:
X: The training features to use when fitting pipelines during an experiment. This setting is being deprecated. Please use training_data and label_column_name instead.
Y: The training labels to use when fitting pipelines during an experiment. This is the value your model will predict. This setting is being deprecated. Please use training_data and label_column_name instead.
X_valid: Validation features to use when fitting pipelines during an experiment.
If specified, then y_valid or sample_weight_valid must also be specified.
Y_valid: Validation labels to use when fitting pipelines during an experiment.
Both X_valid and y_valid must be specified together.
exclude_nan_labels: Whether to exclude rows with NaN values in the label. The default is True.
y_max: y_max (float)
Maximum value of y for a regression experiment. The combination of y_min and y_max are used to normalize test set metrics based on the input data range. If not specified, the maximum value is inferred from the data.
Reference:
https://docs.microsoft.com/en-us/python/api/azureml-train-automl-client/azureml.train.automl.automlconfig.automlconfig?view=azure-ml-py

NEW QUESTION: 117
You are creating a machine learning model that can predict the species of a penguin from its measurements.
You have a file that contains measurements for free species of penguin in comma delimited format.
The model must be optimized for area under the received operating characteristic curve performance metric averaged for each class.
You need to use the Automated Machine Learning user interface in Azure Machine Learning studio to run an experiment and find the best performing model.
Which five 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 collect order.

Answer:



Explanation:



NEW QUESTION: 118

Hotspot Question

You must use an Azure Data Science Virtual Machine (DSVM) as a compute target.

You need to attach an existing DSVM to the workspace by using the Azure Machine Learning SDK for Python.

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

NOTE: Each correct selection is worth one point.

Answer Area

```
from azureml.core.compute import RemoteCompute, ComputeTarget
compute_target_name = "dsvm"
config = RemoteCompute.                    (resource_id='<resource_id>',
 ssh_port=22, username='<username>', private_key_file='./.ssh/id_rsa')
compute = ComputeTarget.                   (ws, compute_target_name, config)
compute.wait_for_completion(show_output=True)
```

- attach_configuration
- get_credentials
- detach

- detach
- create
- attach

Answer:

Answer Area

```
from azureml.core.compute import RemoteCompute, ComputeTarget
compute_target_name = "dsvm"
config = RemoteCompute.                    (resource_id='<resource_id>',
 ssh_port=22, username='<username>', private_key_file='./.ssh/id_rsa')
compute = ComputeTarget.                   (ws, compute_target_name, config)
compute.wait_for_completion(show_output=True)
```

- attach_configuration
- get_credentials
- detach

- detach
- create
- attach

exam-tests.com
Microsoft

Explanation:

https://learn.microsoft.com/en-us/python/api/azureml-core/azureml.core.compute.remote.remotecompute?view=azure-ml-py

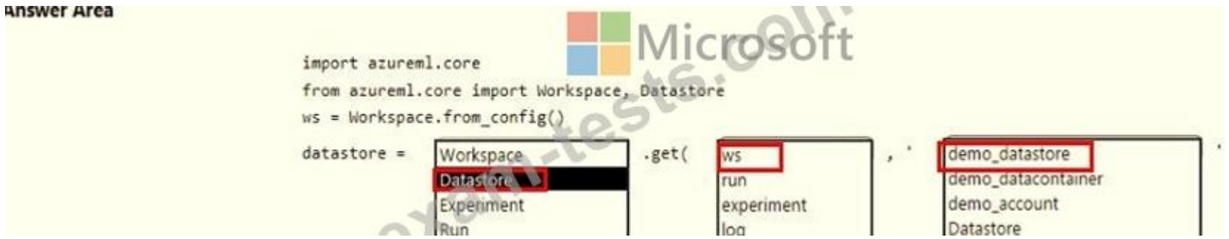
NEW QUESTION: 119

A coworker registers a datastore in a Machine Learning services workspace by using the following code:

You need to write code to access the datastore from a notebook.

Answer Area

Answer:



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

NEW QUESTION: 120

You plan to implement an Azure Machine Learning solution. You have the following requirements:

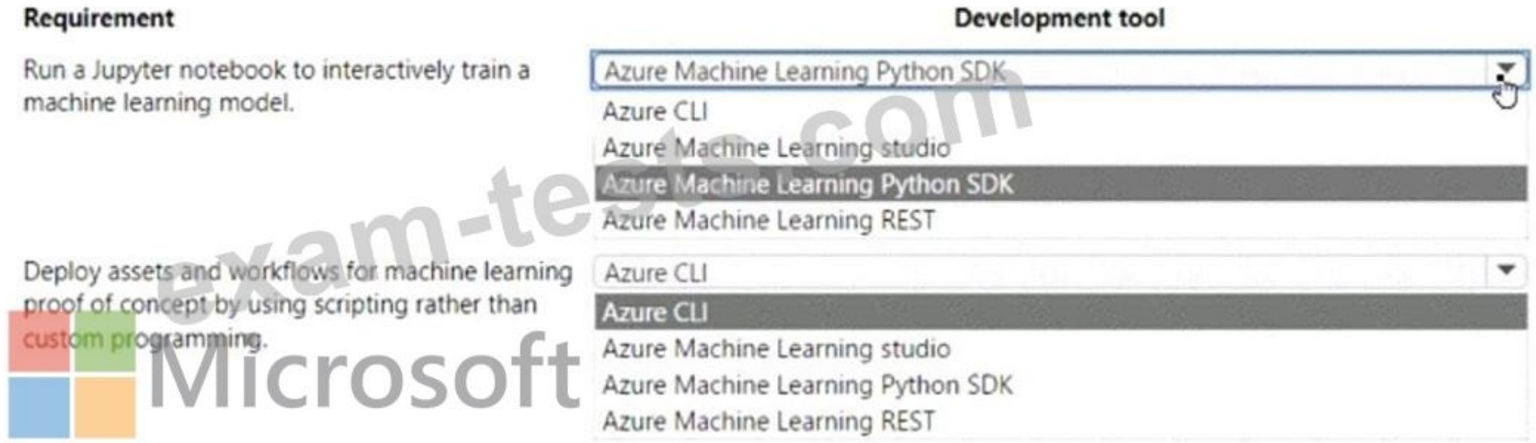
* Run a Jupyter notebook to interactively tram a machine learning model.
* Deploy assets and workflows for machine learning proof of concept by using scripting rather than custom programming.

You need to select a development technique for each requirement

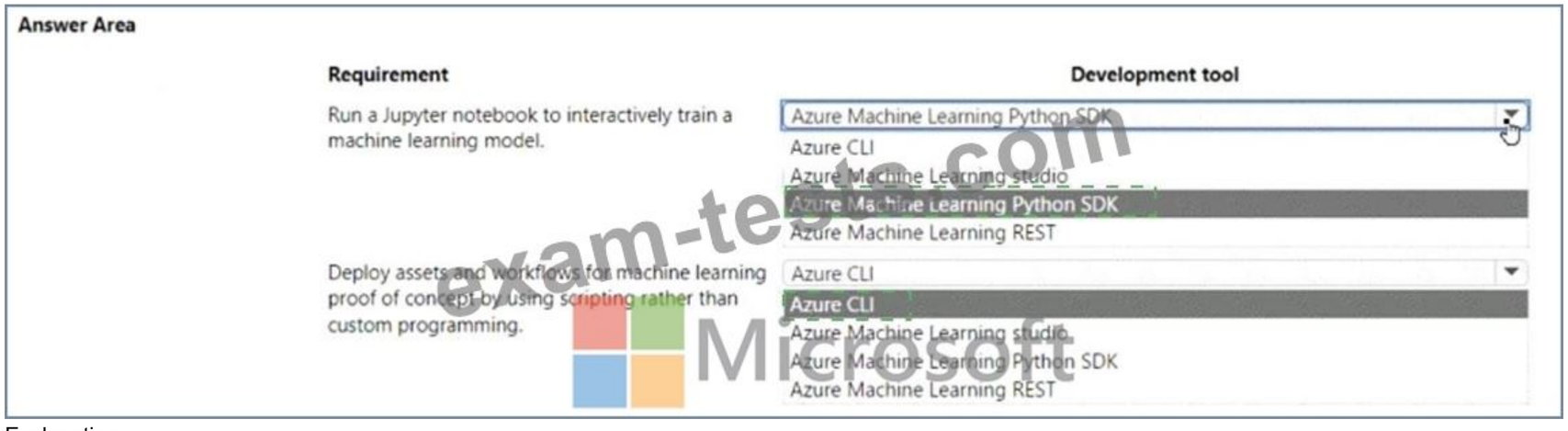
Which development technique should you use? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.

Answer Area

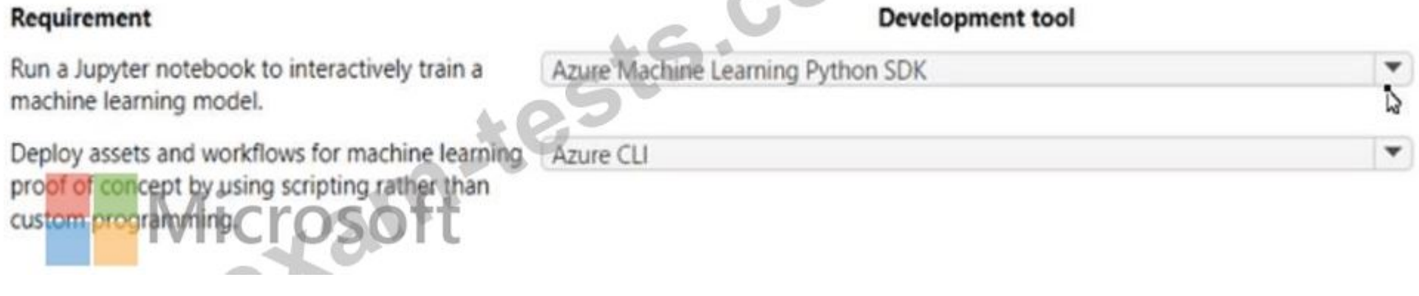


Answer:



Explanation:

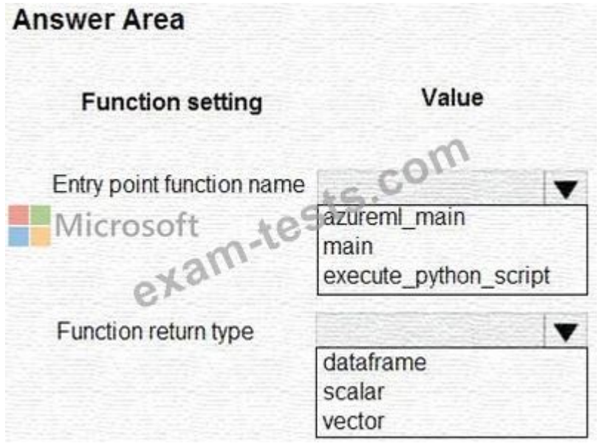
Answer Area



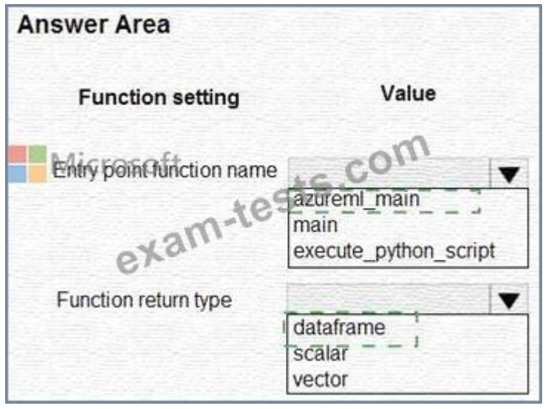
NEW QUESTION: 121

You use Azure Machine Learning Studio to build a machine learning experiment.
You need to divide data into two distinct datasets.
Which module should you use?
A. Assign Data to Clusters
B. Load Trained Model
C. Partition and Sample
D. Tune Model-Hyperparameters
Answer: C (LEAVE A REPLY)
Explanation/Reference:
Explanation:
Partition and Sample with the Stratified split option outputs multiple datasets, partitioned using the rules you specified.
References:
https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/partition-and-sample

NEW QUESTION: 122
You create an Azure Machine Learning dataset containing automobile price data The dataset includes 10,000 rows and 10 columns You use Azure Machine Learning Designer to transform the dataset by using an Execute Python Script component and custom code.
The code must combine three columns to create a new column.
You need to configure the code function.
Which configurations should you use? lo answer, select the appropriate options in the answer area NOTE: Each correct selection is worth one point.



Answer:



Explanation:

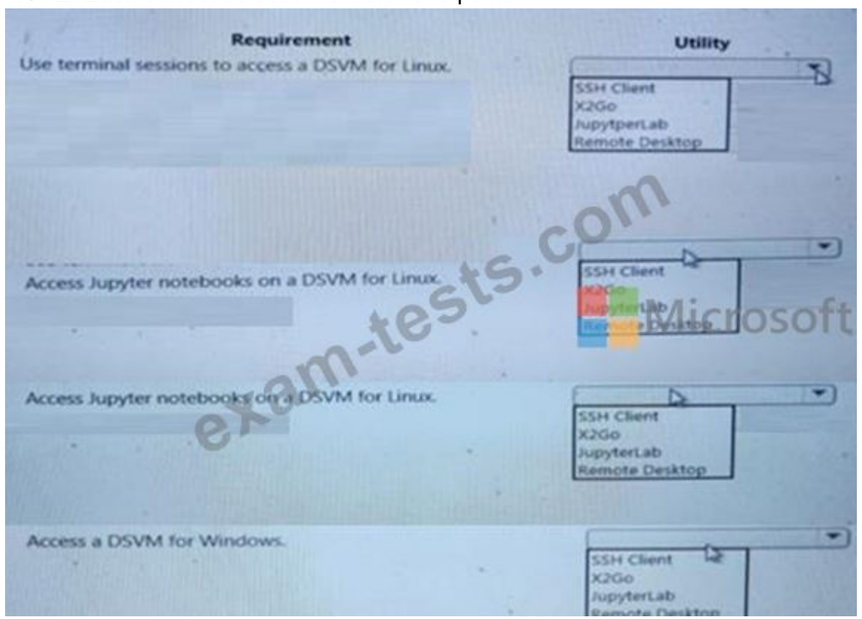
NEW QUESTION: 123

You use Data Science Virtual Machines (DSVMs) for Windows and Linux in Azure.

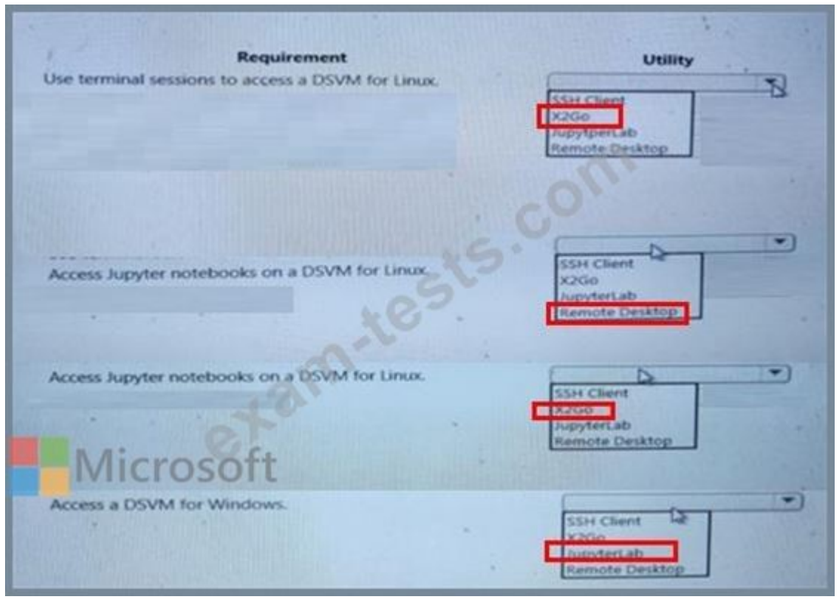
You need to access the DSVMs.

Which utilities should you use? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.



Answer:



NEW QUESTION: 124

Hotspot Question

You create an Azure Machine Learning workspace. You use the Azure Machine Learning Python SDK v2 to create a compute cluster.

The compute cluster must run a training script. Costs associated with running the training script must be minimized.

You need to complete the Python script to create the compute cluster.

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

NOTE: Each correct selection is worth one point.

Answer Area

```
from azure.ai.ml.entities import AmlCompute
try:
    cpu_cluster = ml_client.compute.get("cpu-cluster")
except Exception:
    cpu_cluster = [AmlCompute | ComputeInstance | KubernetesCompute](
        name="cpu-cluster",
        size="STANDARD_DS3_V2",
        max_instances=4,
        [tier="LowPriority" | min_instances=0 | min_instances=1],
    )
    cpu_cluster =
ml_client.begin_create_or_update(cpu_cluster)
)
```


Answer:

Answer Area

```
from azure.ai.ml.entities import AmlCompute
try:
  cpu_cluster = ml_client.compute.get("cpu-cluster")
except Exception:
  cpu_cluster = [AmlCompute | ComputeInstance | KubernetesCompute] (
    name="cpu-cluster",
    size="STANDARD_DS3_V2",
    max_instances=4,
    [tier="LowPriority" | min_instances=0 | min_instances=1],
    )
  cpu_cluster =
ml_client.begin_create_or_update(cpu_cluster)
```


NEW QUESTION: 125

You use the Azure Machine Learning SDK for Python to create a pipeline that includes the following step:

```
step = PythonScriptStep(name= "step1",
                        script_name= "script1.py",
                        compute_target=aml_compute,
                        source_directory=source_directory)
```

The output of the step run must be cached and reused on subsequent runs when the source_directory value has not changed.

You need to define the step.

What should you include in the step definition?

A. allow_reuse

B. version

C. data.as_input(name=?

D. hash_paths

Answer: A (LEAVE A REPLY)

https://learn.microsoft.com/en-us/python/api/azureml-pipeline-steps/azureml.pipeline.steps.python_script_step.pythonscriptstep?view=azure-ml-py

NEW QUESTION: 126

You create an Azure Databricks workspace and a linked Azure Machine Learning workspace.
You have the following Python code segment in the Azure Machine Learning workspace:
import mlflow
import mlflow.azureml
import azureml.mlflow
import azureml.core
from azureml.core import Workspace
subscription_id = 'subscription_id'
resourse_group = 'resource_group_name'
workspace_name = 'workspace_name'
ws = Workspace.get(name=workspace_name,
subscription_id=subscription_id,
resource_group=resource_group)
experimentName = "/Users/{user_name}/{experiment_folder}/{experiment_name}" mlflow.set_experiment(experimentName) uri = ws.get_mlflow_tracking_uri() mlflow.set_tracking_uri(uri) Instructions: For each of the following statements, select Yes if the statement is true. Otherwise, select No.
NOTE: Each correct selection is worth one point.

	Yes	No
A resource group and Azure Machine Learning workspace will be created.	○	○
An Azure Databricks experiment will be tracked only in the Azure Machine Learning workspace.	○	○
The epoch loss metric is set to be tracked.	○	○

Answer:

	Yes	No
A resource group and Azure Machine Learning workspace will be created.	○	●
An Azure Databricks experiment will be tracked only in the Azure Machine Learning workspace.	●	○
The epoch loss metric is set to be tracked.	●	○

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

NEW QUESTION: 127
Note: This question is part of a series of questions that present the same scenario. Each question in the series contains a unique solution that might meet the stated goals. Some question sets might have more than one correct solution, while others might not have a correct solution.

After you answer a question in this section, you will NOT be able to return to it. As a result, these questions will not appear in the review screen.
You have a Python script named train.py in a local folder named scripts. The script trains a regression model by using scikit-learn. The script includes code to load a training data file which is also located in the scripts folder.
You must run the script as an Azure ML experiment on a compute cluster named aml-compute.
You need to configure the run to ensure that the environment includes the required packages for model training. You have instantiated a variable named aml-compute that references the target compute cluster.
Solution: Run the following code:

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

Does the solution meet the goal?
A. Yes
B. No
Answer: B (LEAVE A REPLY)
There is a missing line: conda_packages=['scikit-learn'], which is needed.
Correct example:
sk_est = Estimator(source_directory='./my-sklearn-proj',
script_params=script_params,
compute_target=compute_target,
entry_script='train.py',
conda_packages=['scikit-learn'])
Note:
The Estimator class represents a generic estimator to train data using any supplied framework.
This class is designed for use with machine learning frameworks that do not already have an Azure Machine Learning pre-configured estimator. Pre-configured estimators exist for Chainer, PyTorch, TensorFlow, and SKLearn.
Example:
from azureml.train.estimator import Estimator
script_params = {
# to mount files referenced by mnist dataset
'--data-folder': ds.as_named_input('mnist').as_mount(),
'--regularization': 0.8
}
Reference:
https://docs.microsoft.com/en-us/python/api/azureml-train-core/azureml.train.estimator.estimator

NEW QUESTION: 128
You are creating a binary classification by using a two-class logistic regression model.
You need to evaluate the model results for imbalance.
Which evaluation metric should you use?
A. Relative Absolute Error
B. AUC Curve
C. Mean Absolute Error
D. Relative Squared Error

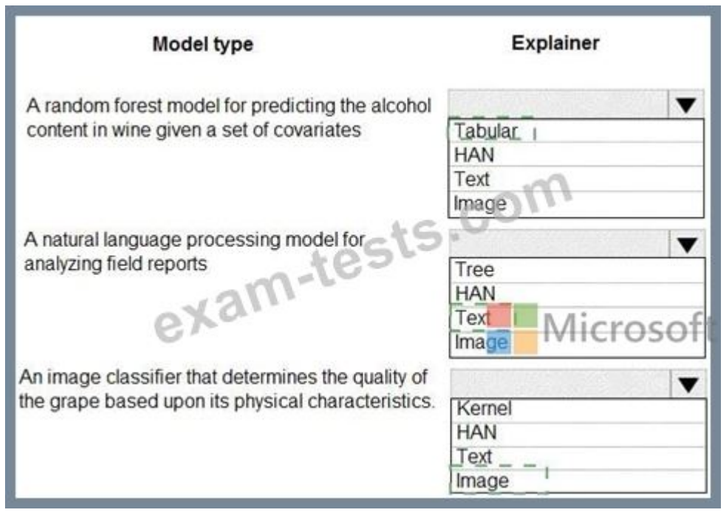
Answer: B (LEAVE A REPLY)
One can inspect the true positive rate vs. the false positive rate in the Receiver Operating Characteristic (ROC) curve and the corresponding Area Under the Curve (AUC) value. The closer this curve is to the upper left corner, the better the classifier's performance is (that is maximizing the true positive rate while minimizing the false positive rate). Curves that are close to the diagonal of the plot, result from classifiers that tend to make predictions that are close to random guessing.
References:
https://docs.microsoft.com/en-us/azure/machine-learning/studio/evaluate-model-performance#evaluating-a-binar

NEW QUESTION: 129
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 an Azure Machine Learning model.
You plan to deploy the model to an online endpoint.
You need to ensure that applications will be able to use the authentication method with a non- expiring artifact to access the model.
Solution: Create a Kubernetes online endpoint and set the value of its auth_mode parameter to aml_token. Deploy the model to the online endpoint.
Does the solution meet the goal?
A. No
B. Yes
Answer: A (LEAVE A REPLY)

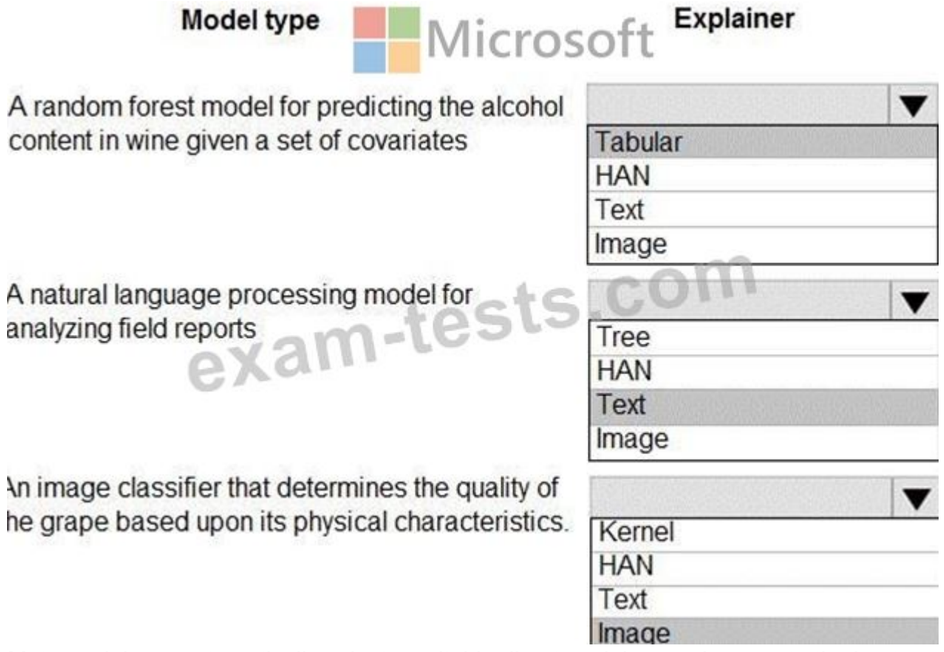
NEW QUESTION: 130
You are hired as a data scientist at a winery. The previous data scientist used Azure Machine Learning.
You need to review the models and explain how each model makes decisions.
Which explainer modules should you use? To answer, select the appropriate options in the answer area.
NOTE: Each correct selection is worth one point.

Model type	Explainer
A random forest model for predicting the alcohol content in wine given a set of covariates	Tabular / HAN / Text / Image
A natural language processing model for analyzing field reports	Tree / HAN / Text / Image
An image classifier that determines the quality of the grape based upon its physical characteristics.	Kernel / HAN / Text / Image

Answer:



Explanation:



Meta explainers automatically select a suitable direct explainer and generate the best explanation info based on the given model and data sets. The meta explainers leverage all the libraries (SHAP, LIME, Mimic, etc.) that we have integrated or developed. The following are the meta explainers available in the SDK:

Tabular Explainer: Used with tabular datasets.

Text Explainer: Used with text datasets.

Image Explainer: Used with image datasets.

Box 1: Tabular

Box 2: Text

Box 3: Image

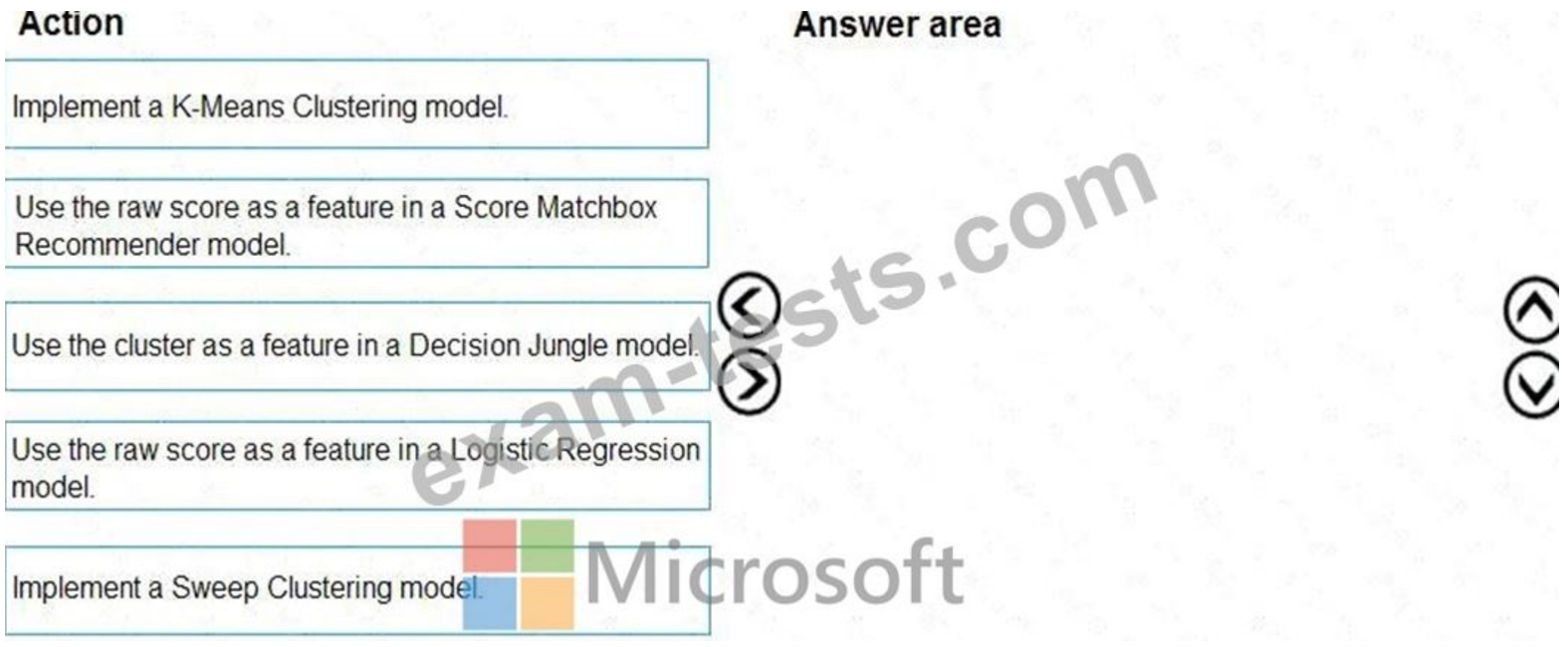
Reference:

https://medium.com/microsoftazure/automated-and-interpretable-machine-learning-d07975741298

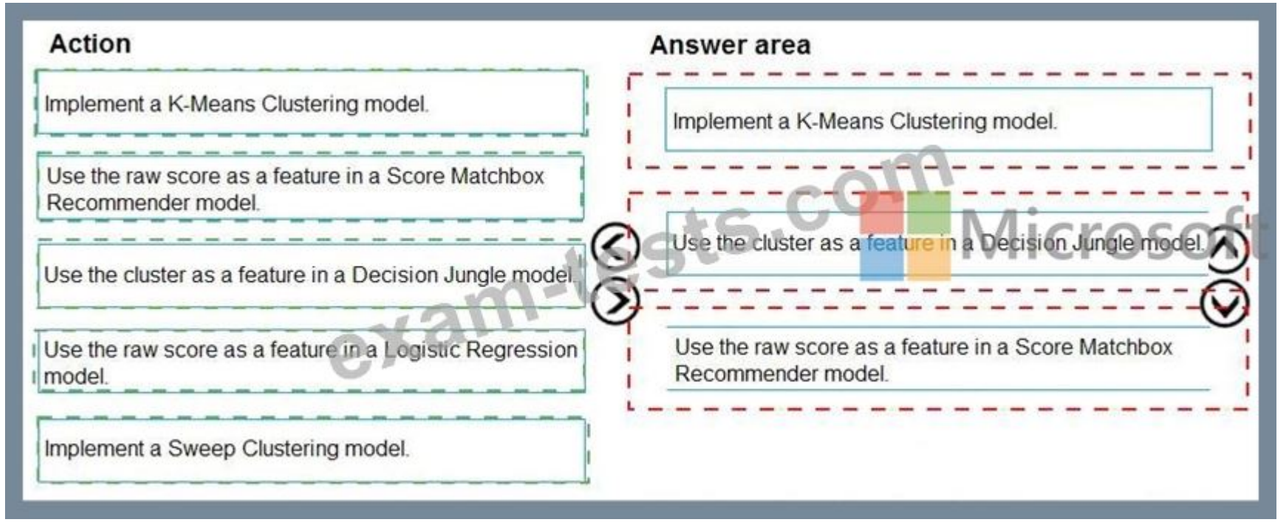
NEW QUESTION: 131

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.

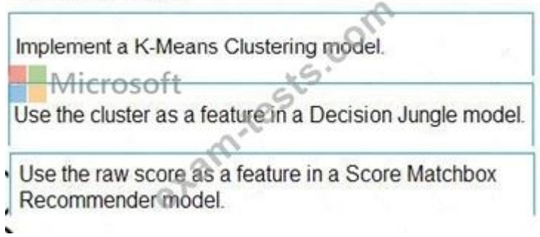


Answer:



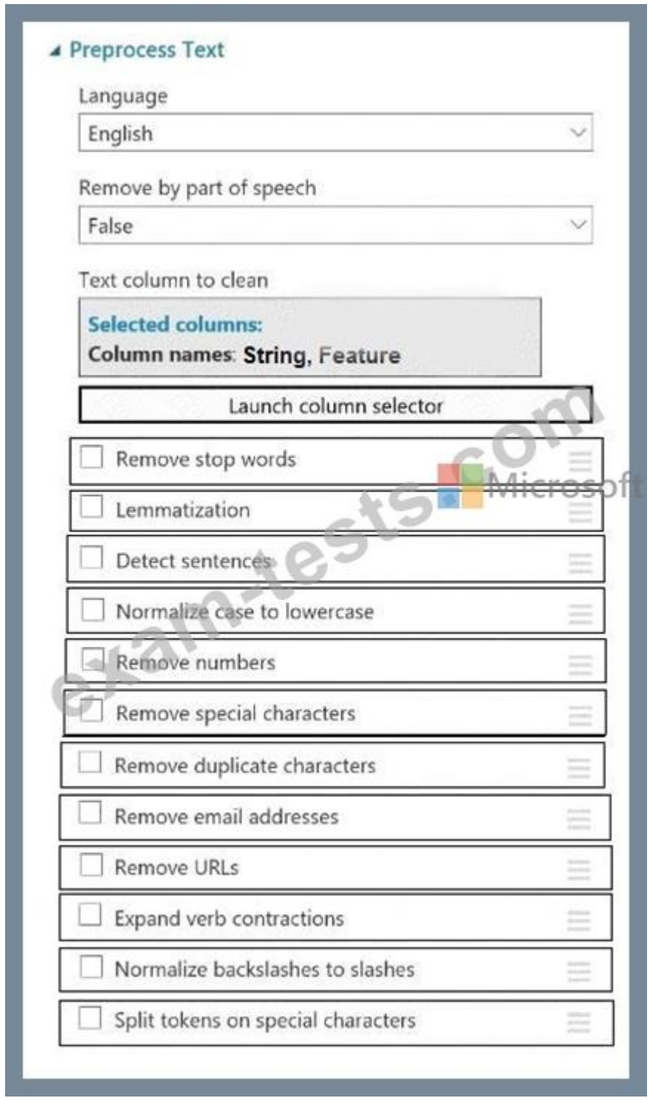
Explanation:

Answer area

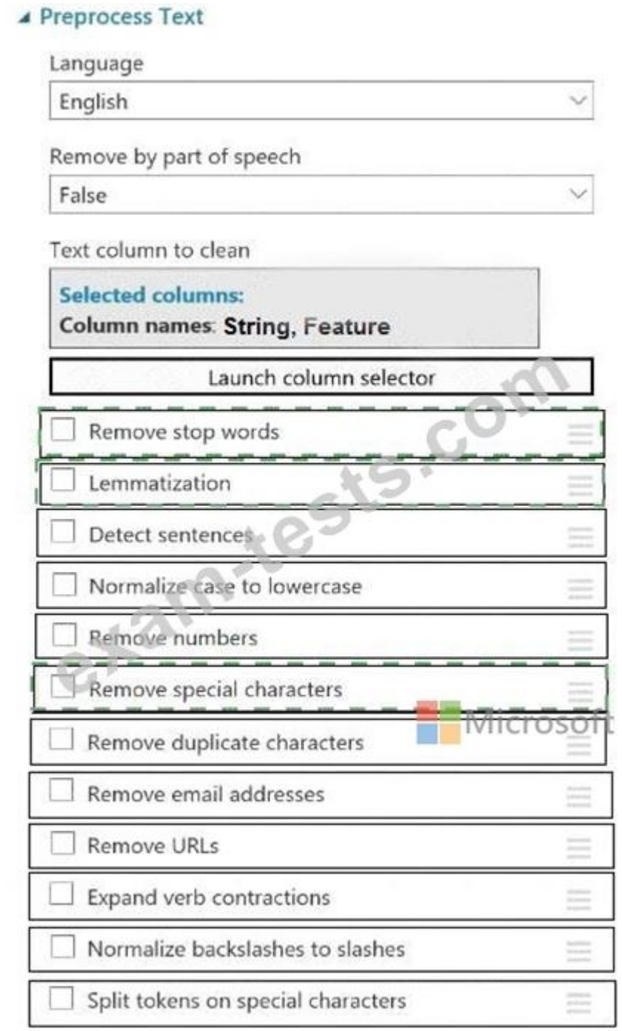


Step 1: Implement a K-Means Clustering model
Step 2: Use the cluster as a feature in a Decision jungle model.
Decision jungles are non-parametric models, which can represent non-linear decision boundaries.
Step 3: Use the raw score as a feature in a Score Matchbox Recommender model The goal of creating 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- recommender

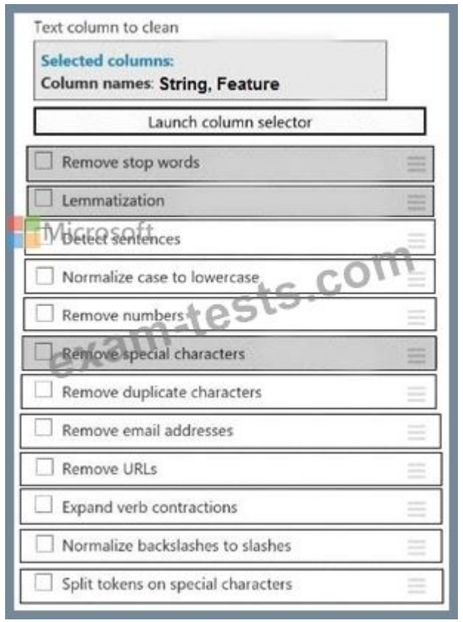
NEW QUESTION: 132
You plan to preprocess text from CSV files. You load the Azure Machine Learning Studio default stop words list.
You need to configure the Preprocess Text module to meet the following requirements:
* Ensure that multiple related words from a single canonical form.
* Remove pipe characters from text.
* Remove words to optimize information retrieval.
Which three options should you select? To answer, select the appropriate options in the answer area.
NOTE: Each correct selection is worth one point.


Preprocess Text
Language
English
Remove by part of speech
False
Text column to clean
Selected columns:
Column names: String, Feature
Launch column selector
Remove stop words
Lemmatization
Detect sentences
Normalize case to lowercase
Remove numbers
Remove special characters
Remove duplicate characters
Remove email addresses
Remove URLs
Expand verb contractions
Normalize backslashes to slashes
Split tokens on special characters

Answer:



Explanation



Box 1: Remove stop words

Remove words to optimize information retrieval.

Remove stop words: Select this option if you want to apply a predefined stopword list to the text column. Stop word removal is performed before any other processes.

Box 2: Lemmatization

Ensure that multiple related words from a single canonical form.

Lemmatization converts multiple related words to a single canonical form Box 3: Remove special characters Remove special characters: Use this option to replace any non-alphanumeric special characters with the pipe | character.

References:

https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/preprocess-text

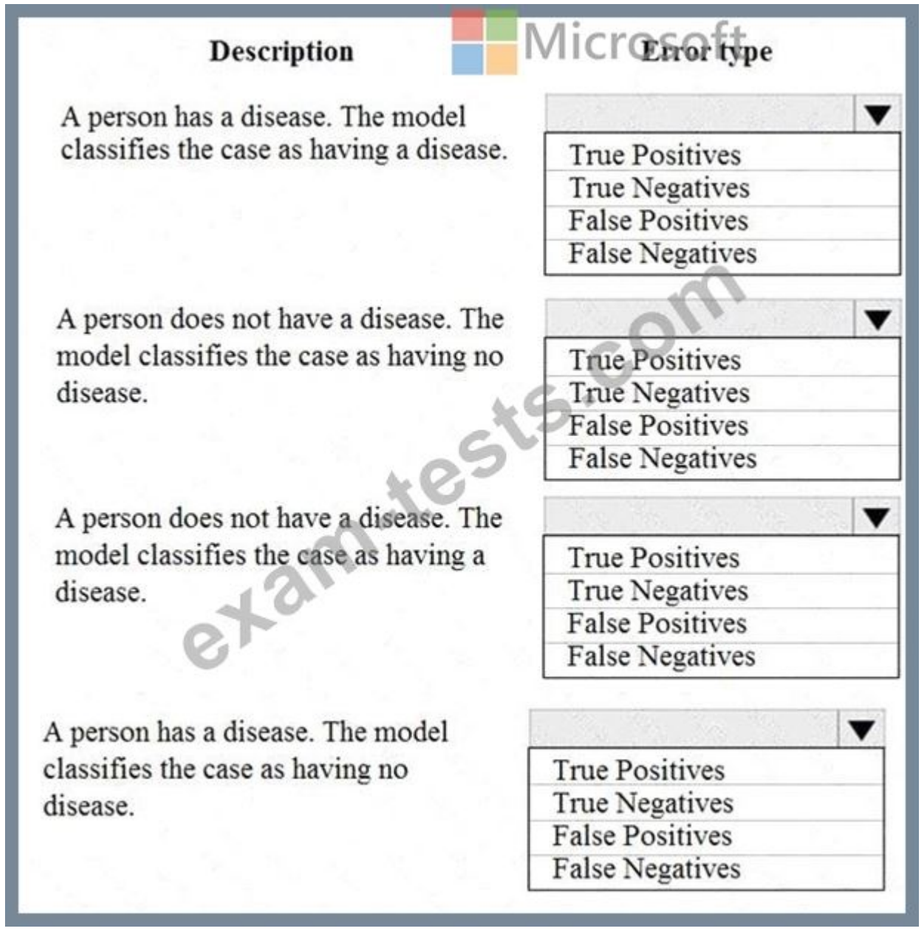
NEW QUESTION: 133

You create a binary classification model to predict whether a person has a disease.

You need to detect possible classification errors.

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

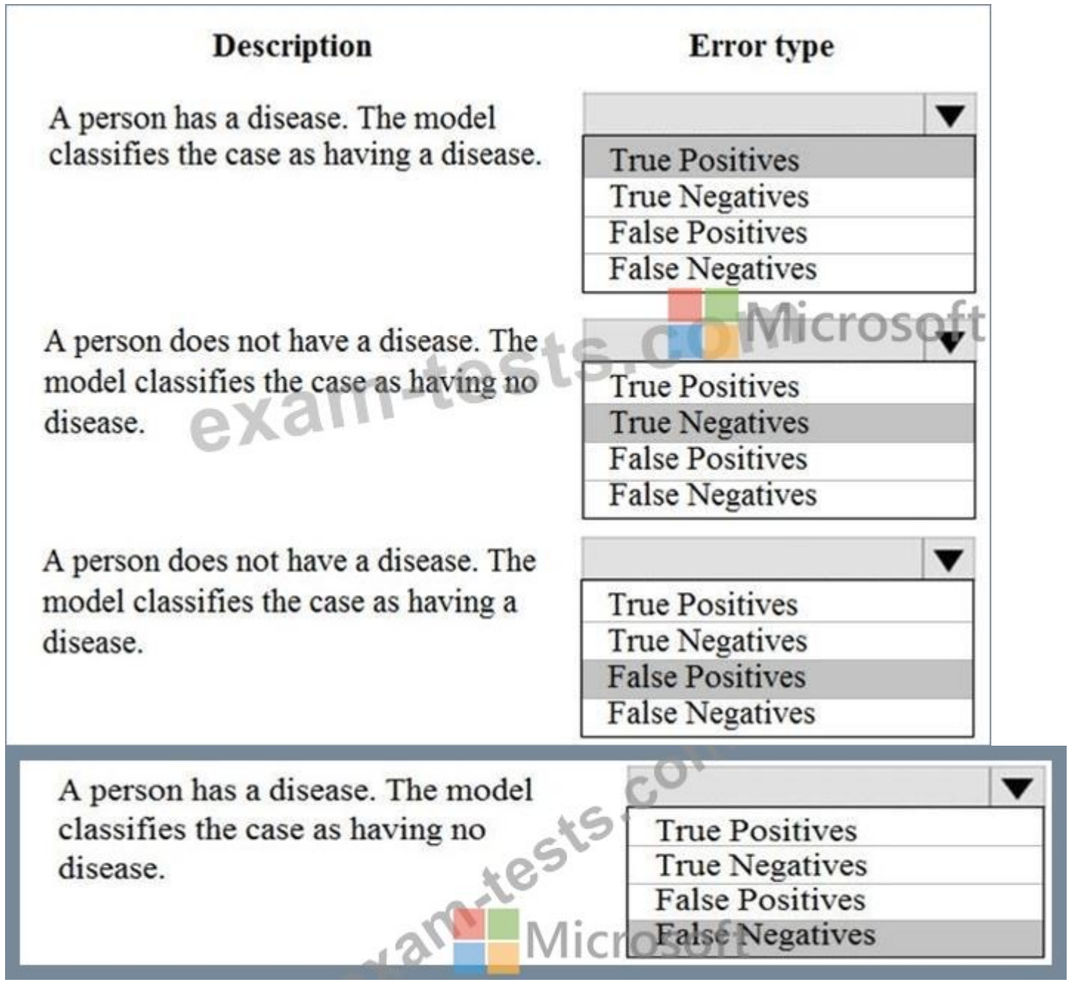
NOTE: Each correct selection is worth one point.


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

Answer:

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

Explanation:



Box 1: True Positive

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

Box 3: False Positive

A false positive is an outcome where the model incorrectly predicts the positive class.

Box 4: False Negative

A false negative is an outcome where the model incorrectly predicts the negative class.

Note: Let's make the following definitions:

"Wolf" is a positive class.

"No wolf" is a negative class.

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

Reference:

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

NEW QUESTION: 134

You train a classification model by using a decision tree algorithm.

You create an estimator by running the following Python code. The variable feature_names is a list of all feature names, and class_names is a list of all class names.

```
from interpret.ext.blackbox import TabularExplainer
explainer = TabularExplainer(model,
                             x_train,
                             features=feature_names,
                             classes=class_names)
```

You need to explain the predictions made by the model for all classes by determining the importance of all features.

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

NOTE: Each correct selection is worth one point.

	Yes	No
The SHAP TreeExplainer will be used to interpret the model.	○	○
If you omit the features and classes parameters in the TabularExplainer instantiation, the explainer still works as expected.	○	○
You could interpret the model by using a MimicExplainer instead of a TabularExplainer.	○	○

Answer:

	Yes	No
The SHAP TreeExplainer will be used to interpret the model.	○	○
If you omit the features and classes parameters in the TabularExplainer instantiation, the explainer still works as expected.	○	○
You could interpret the model by using a MimicExplainer instead of a TabularExplainer.	○	○

Explanation

	Yes	No
The SHAP TreeExplainer will be used to interpret the model.	○	○
If you omit the features and classes parameters in the TabularExplainer instantiation, the explainer still works as expected.	○	○
You could interpret the model by using a MimicExplainer instead of a TabularExplainer.	○	○

Box 1: Yes

TabularExplainer calls one of the three SHAP explainers underneath (TreeExplainer, DeepExplainer, or KernelExplainer).

Box 2: Yes

To make your explanations and visualizations more informative, you can choose to pass in feature names and output class names if doing classification.

Box 3: No
TabularExplainer automatically selects the most appropriate one for your use case, but you can call each of its three underlying explainers underneath (TreeExplainer, DeepExplainer, or KernelExplainer) directly.
Reference:
https://docs.microsoft.com/en-us/azure/machine-learning/how-to-machine-learning-interpretability-aml