DP-100 Q&As

Designing and Implementing a Data Science Solution on Azure

Pass Microsoft DP-100 Exam with 100% Guarantee

Free Download Real Questions & Answers **PDF** and **VCE** file from:

https://www.pass2lead.com/dp-100.html

100% Passing Guarantee
100% Money Back Assurance

Following Questions and Answers are all new published by Microsoft Official Exam Center

QUESTION 1

HOTSPOT

You have an Azure Machine Learning workspace.

You run the following code in a Python environment in which the configuration file for your workspace has been downloaded.

```
from azureml.core import Workspace
from azureml.core import Experiment
import pandas as pd
import datetime as dt
ws = Workspace.from_config()
experiment = Experiment(workspace=ws, name='my_experiment')
run = experiment.start_logging()
print('run_time', dt.datetime.now())

row_count = (len(data))
run.log('observations', row_count)
run.complete()
```

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.

Hot Area:

Statements	Yes	No
An error will occur if an experiment named my_experiment does not already exist in the workspace.	○	○
If the experiment does not exist, it will be created. If the experiment does exist, the code will create a new run of the existing experiment.	○	○
After the code completes, a metric named run_time is recorded in the experiment run. The metric will contain the date and time for the run.	○	○
After the code completes, the data.csv file will be available in the run's output.	○	○

Correct Answer:

Statements	Yes	No
An error will occur if an experiment named my_experiment does not already exist in the workspace.	○	●
If the experiment does not exist, it will be created. If the experiment does exist, the code will create a new run of the existing experiment.	●	○
After the code completes, a metric named run_time is recorded in the experiment run. The metric will contain the date and time for the run.	○	●
After the code completes, the data.csv file will be available in the run's output.	○	●

QUESTION 2

You are evaluating a completed binary classification machine learning model.

You need to use the precision as the valuation metric.

Which visualization should you use?

A. Binary classification confusion matrix

B. box plot

C. Gradient descent

D. coefficient of determination

Correct Answer: A

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

QUESTION 3

You need to resolve the local machine learning pipeline performance issue. What should you do?

A. Increase Graphic Processing Units (GPUs).

B. Increase the learning rate.

C. Increase the training iterations.

D. Increase Central Processing Units (CPUs).

Correct Answer: A

QUESTION 4

You create an Azure Machine Learning pipeline named pipeline 1 with two steps that contain Python scripts. Data processed by the first step is passed to the second step.

You must update the content of the downstream data source of pipeline 1 and run the pipeline again.

You need to ensure the new run of pipeline 1 fully processes the updated content.

Solution: Change the value of the compute.target parameter of the PythonScriptStep object in the two steps.

Does the solution meet the goal?

A. Yes

B. No

Correct Answer: B

QUESTION 5

You are determining if two sets of data are significantly different from one another by using Azure Machine Learning Studio.

Estimated values in one set of data may be more than or less than reference values in the other set of data. You must produce a distribution that has a constant Type I error as a function of the correlation.

You need to produce the distribution.

Which type of distribution should you produce?

A. Unpaired t-test with a two-tail option

B. Unpaired t-test with a one-tail option

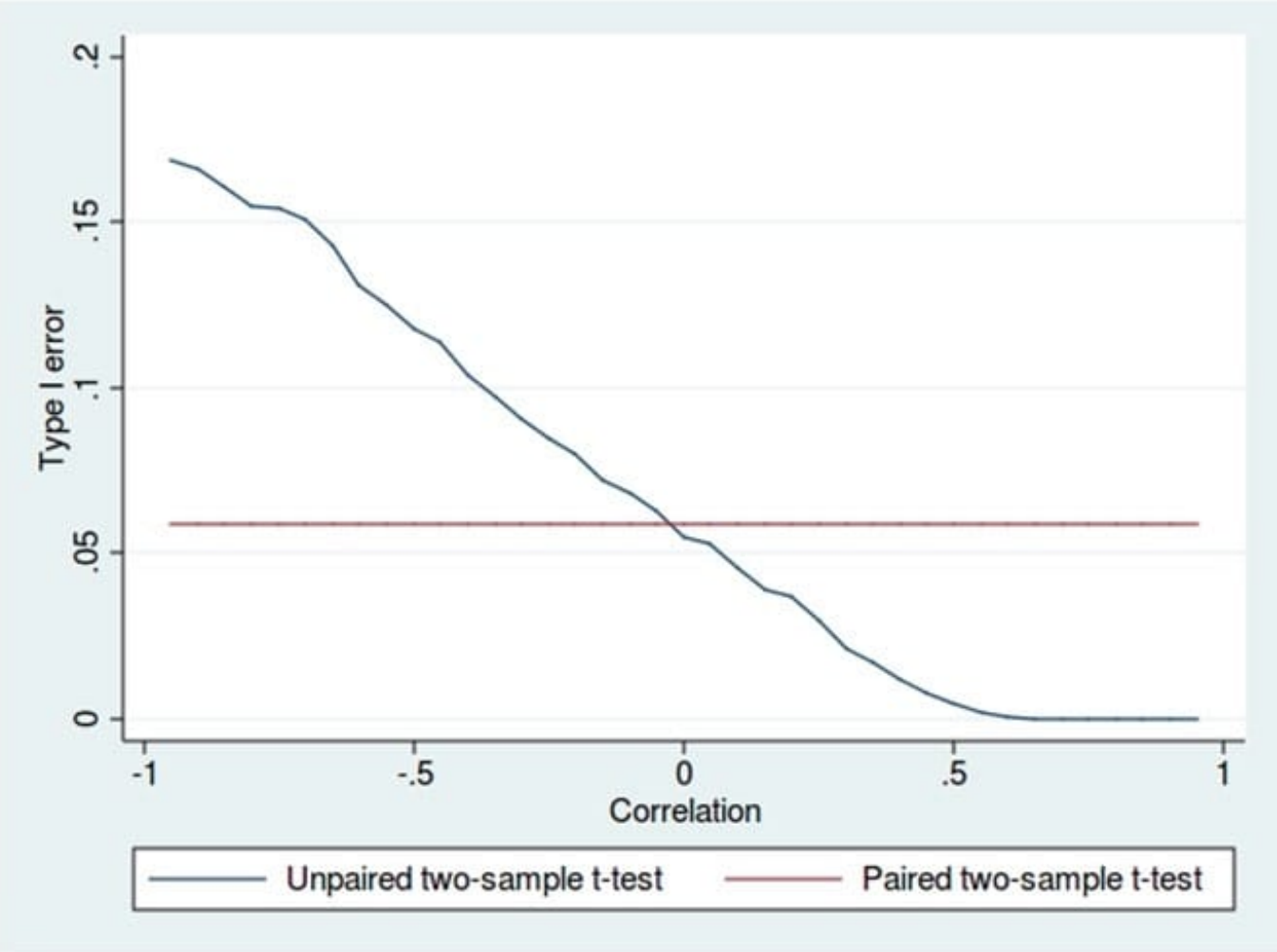
C. Paired t-test with a one-tail option

D. Paired t-test with a two-tail option

Correct Answer: D

Choose a one-tail or two-tail test. The default is a two-tailed test. This is the most common type of test, in which the expected distribution is symmetric around zero.

Example: Type I error of unpaired and paired two-sample t-tests as a function of the correlation. The simulated random numbers originate from a bivariate normal distribution with a variance of 1.



Reference: https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/test-hypothesis-using-t-test

https://en.wikipedia.org/wiki/Student%27s_t-test

QUESTION 6

You manage an Azure Machine Learning workspace. You develop a machine learning model.

You must deploy the model to use a low-priority VM with a pricing discount.

You need to deploy the model.

Which compute target should you use?

A. Azure Kubernetes Service (AKS)

B. Azure Machine Learning compute clusters

C. Azure Container Instances (ACI)

D. Local deployment

Correct Answer: B

How batch deployment works with low priority VMs

Azure Machine Learning Batch Deployments provides several capabilities that make it easy to consume and benefit from low priority VMs:

*

Batch deployment jobs consume low priority VMs by running on Azure Machine Learning compute clusters created with low priority VMs. Once a deployment is associated with a low priority VMs\\' cluster, all the jobs produced by such deployment will use low priority VMs. Per-job configuration is not possible.

*

Batch deployment jobs automatically seek the target number of VMs in the available compute cluster based on the number of tasks to submit. If VMs are preempted or unavailable, batch deployment jobs attempt to replace the lost capacity by queuing the failed tasks to the cluster.

*

Etc.

Reference: https://learn.microsoft.com/en-us/azure/machine-learning/how-to-use-low-priority-batch

QUESTION 7

You create a workspace by using Azure Machine Learning Studio.

You must run a Python SDK v2 notebook in the workspace by using Azure Machine Learning Studio. You must preserve the current values of variables set in the notebook for the current instance.

You need to maintain the state of the notebook.

What should you do?

A. Change the compute.

B. Change the current kernel

C. Stop the compute.

D. Stop the current kernel.

Correct Answer: B

QUESTION 8

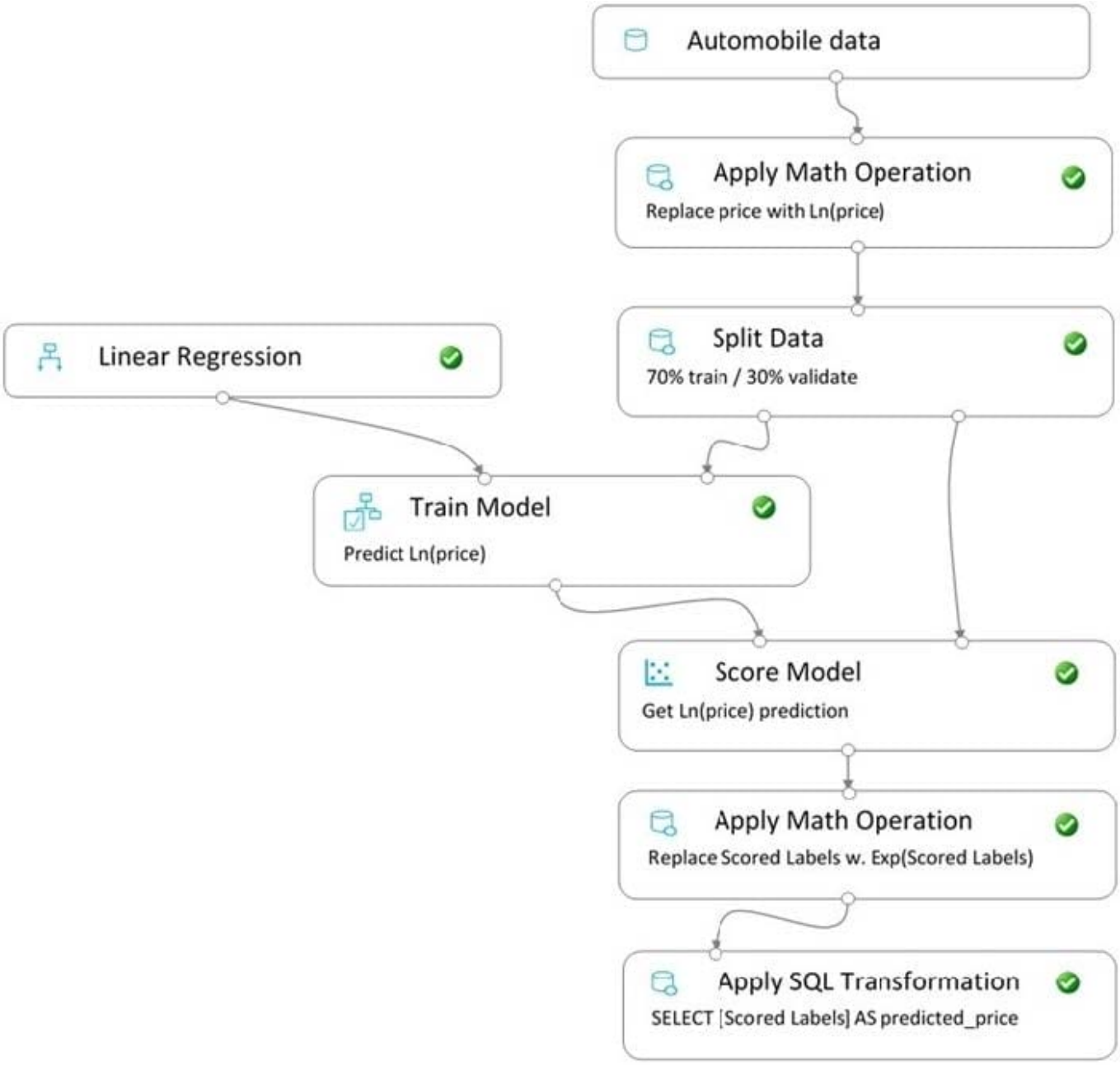
You create a pipeline in designer to train a model that predicts automobile prices.

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

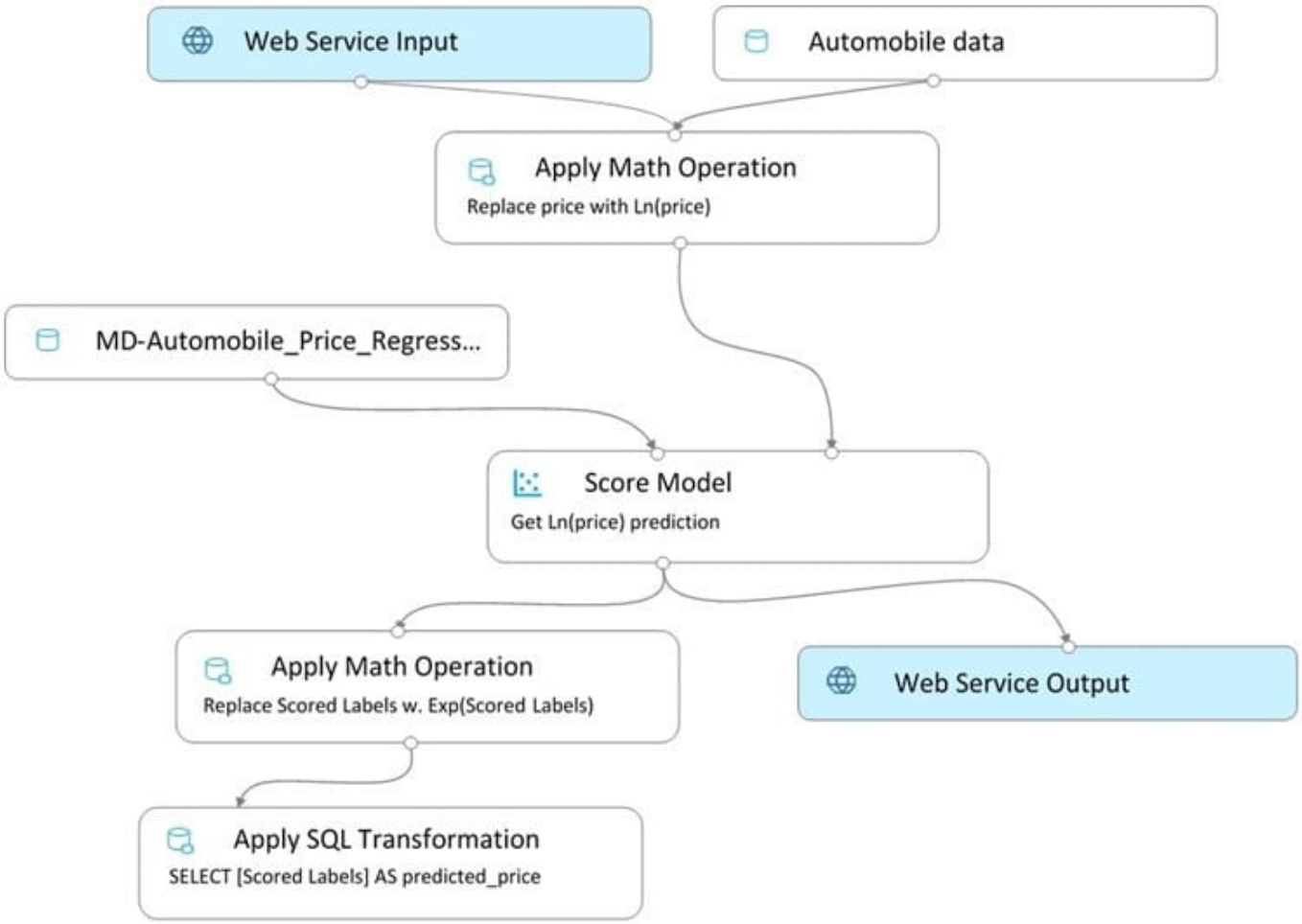
the predicted price.

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

Training pipeline



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



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

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

NOTE: Each correct selection is worth one point.

A. Connect the output of the Apply SQL Transformation to the Web Service Output module.

B. Replace the Web Service Input module with a data input that does not include the price column.

C. Add a Select Columns module before the Score Model module to select all columns other than price.

D. Replace the training dataset module with a data input that does not include the price column.

E. Remove the Apply Math Operation module that replaces price with its natural log from the data flow.

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

Correct Answer: ACE

QUESTION 9

You use the following Python code in a notebook to deploy a model as a web service:

from azureml.core.webservice import AciWebservice

from azureml.core.model import InferenceConfig

inference_config = InferenceConfig(runtime=\\'python\\', source_directory=\\'model_files\\', entry_script=\\'score.py\\', conda_file=\\'env.yml\\')

deployment_config = AciWebservice.deploy_configuration(cpu_cores=1, memory_gb=1)

service = Model.deploy(ws, \\'my-service\\', [model], inference_config, deployment_config)

service.wait_for_deployment(True)

The deployment fails.

You need to use the Python SDK in the notebook to determine the events that occurred during service deployment an initialization.

Which code segment should you use?

A. service.state

B. service.get_logs()

C. service.serialize()

D. service.environment

Correct Answer: B

The first step in debugging errors is to get your deployment logs. In Python: service.get_logs()

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

QUESTION 10

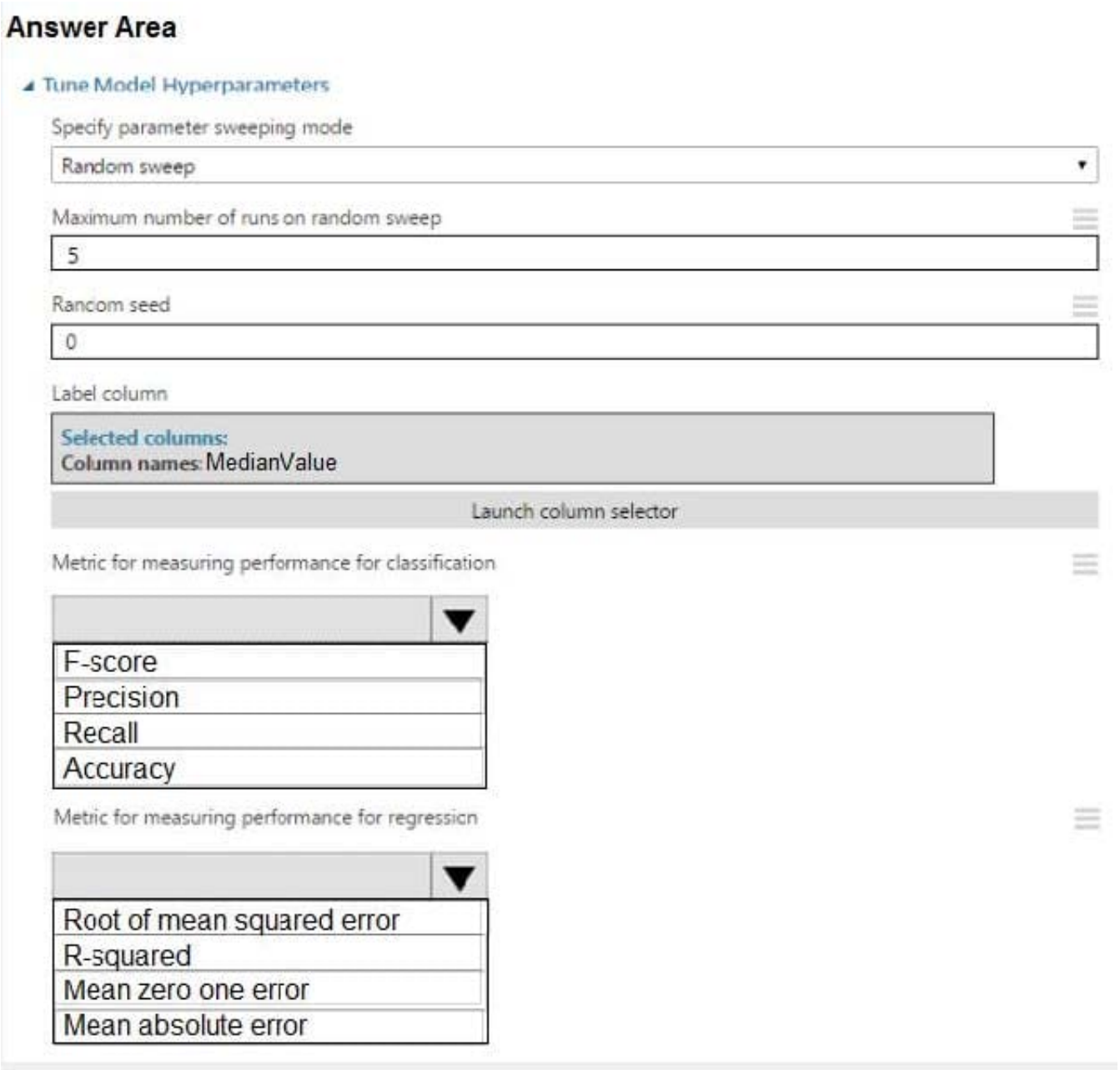
You need to set up the Permutation Feature Importance module according to the model training requirements.

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

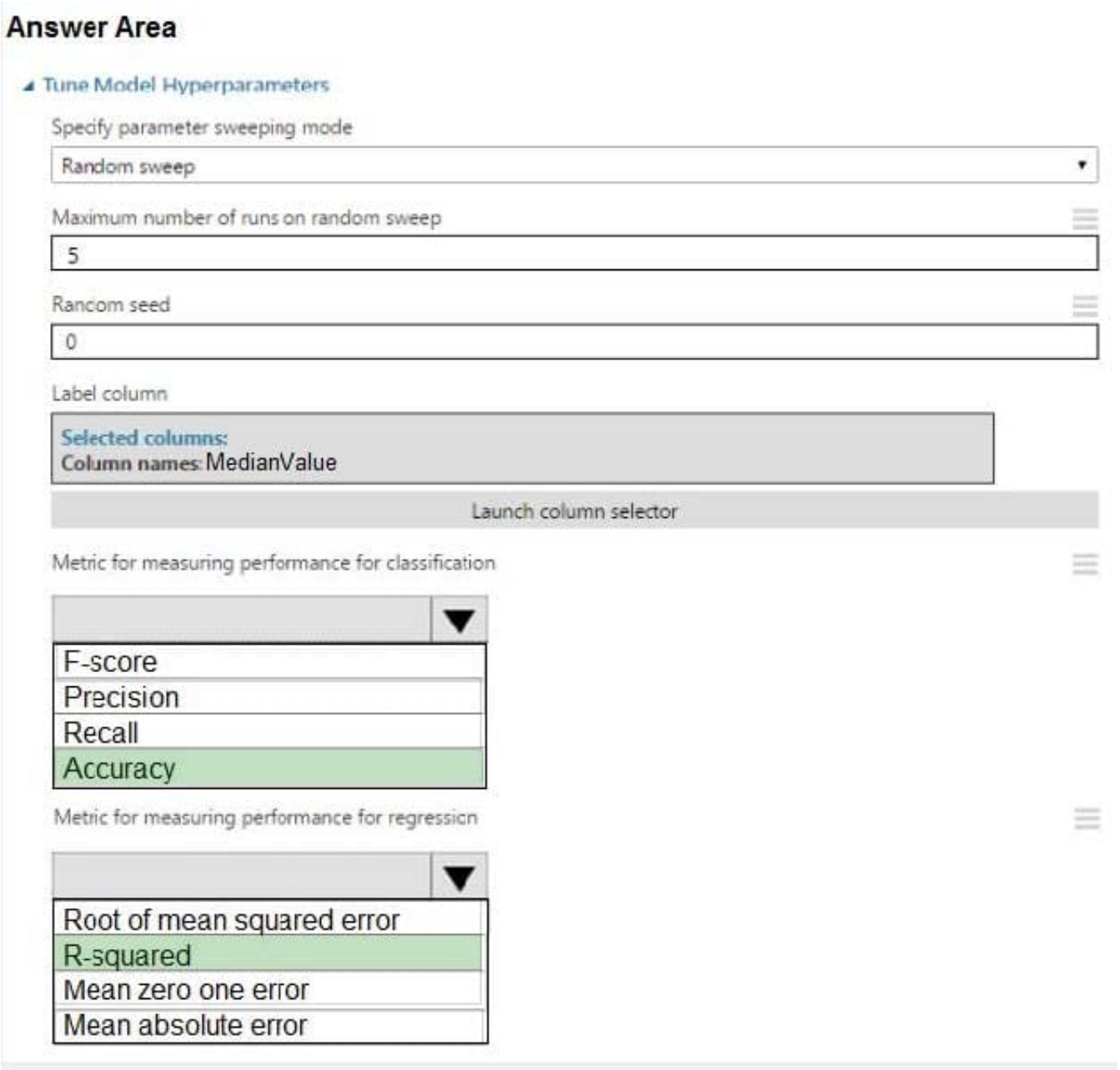
NOTE: Each correct selection is worth one point.

Hot Area:

Answer Area



Correct Answer:



Box 1: Accuracy

Scenario: 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.

Box 2: R-Squared

QUESTION 11

You are implementing hyperparameter tuning by using Bayesian sampling for an Azure ML Python SDK v2-based model training from a notebook. The notebook is in an Azure Machine Learning workspace. The notebook uses a training

script that runs on a compute cluster with 20 nodes.

The code implements Bandit termination policy with slack_factor set to 02 and a sweep job with max_concurrent_trials set to 10.

You must increase effectiveness of the tuning process by improving sampling convergence.

You need to select which sampling convergence to use.

What should you select?

A. Set the value of slack.factor of earty.termination policy to 0.1.

B. Set the value of max_concurrent_trials to 4.

C. Set the value of slack_factor of eartyjermination policy to 0.9.

D. Set the value of max.concurrentjrials to 20.

Correct Answer: C

QUESTION 12

DRAG DROP

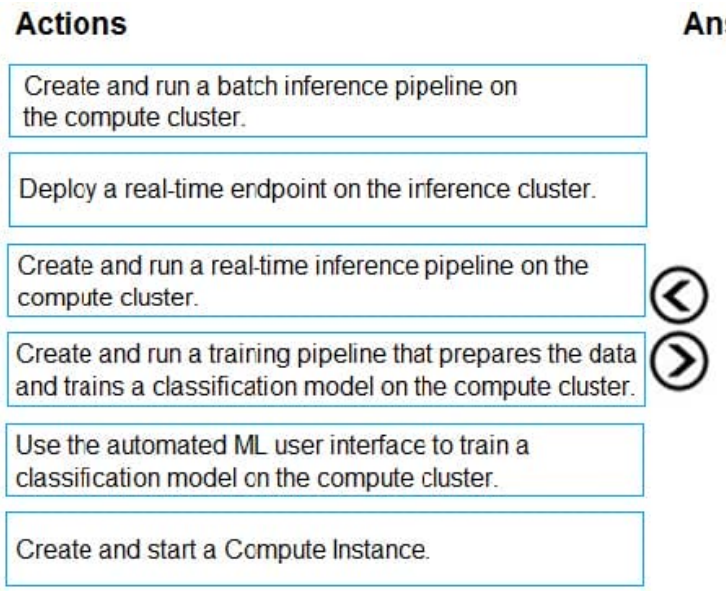
You have an Azure Machine Learning workspace that contains a CPU-based compute cluster and an Azure Kubernetes Services (AKS) inference cluster. You create a tabular dataset containing data that you plan to use to create a

classification model.

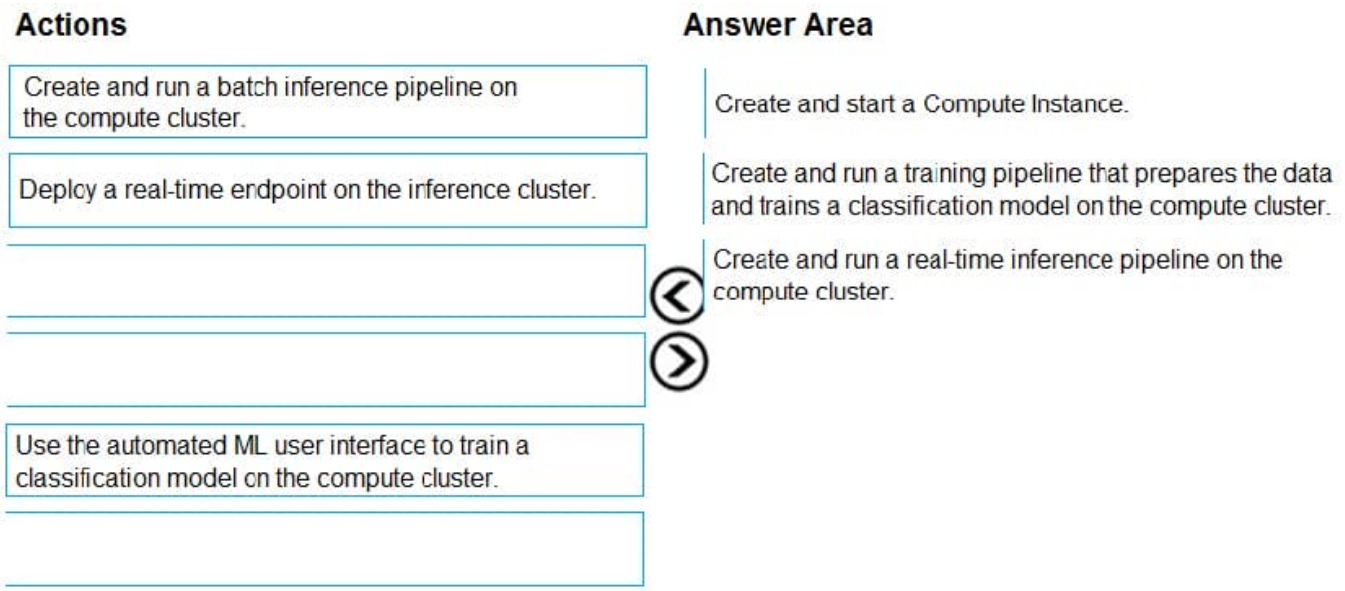
You need to use the Azure Machine Learning designer to create a web service through which client applications can consume the classification model by submitting new data and getting an immediate prediction as a response.

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

Select and Place:



Correct Answer:



Step 1: Create and start a Compute Instance

To train and deploy models using Azure Machine Learning designer, you need compute on which to run the training process, test the model, and host the model in a deployed service.

There are four kinds of compute resource you can create:

Compute Instances: Development workstations that data scientists can use to work with data and models.

Compute Clusters: Scalable clusters of virtual machines for on-demand processing of experiment code.

Inference Clusters: Deployment targets for predictive services that use your trained models.

Attached Compute: Links to existing Azure compute resources, such as Virtual Machines or Azure Databricks clusters.

Step 2: Create and run a training pipeline..

After you\\'ve used data transformations to prepare the data, you can use it to train a machine learning model. Create and run a training pipeline

Step 3: Create and run a real-time inference pipeline

After creating and running a pipeline to train the model, you need a second pipeline that performs the same data transformations for new data, and then uses the trained model to inference (in other words, predict) label values based on its

features. This pipeline will form the basis for a predictive service that you can publish for applications to use.

Reference:

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

QUESTION 13

You are implementing hyperparameter tuning for a model training from a notebook. The notebook is in an Azure Machine Learning workspace.

You must configure a grid sampling method over the search space for the num_hidden_layers and batch_size hyperparameters.

You need to identify the hyperparameters for the grid sampling.

Which hyperparameter sampling approach should you use?

A. uniform

B. qlognormal

C. choice

D. normal

Correct Answer: B

Discrete hyperparameters (as this is Grid Sample, see note below)

Discrete hyperparameters are specified as a Choice among discrete values. Choice can be:

one or more comma-separated values

a range object

any arbitrary list object

Python

from azure.ai.ml.sweep import Choice

command_job_for_sweep = command_job(

batch_size=Choice(values=[16, 32, 64, 128]),

number_of_hidden_layers=Choice(values=range(1,5)),

)

In this case, batch_size one of the values [16, 32, 64, 128] and number_of_hidden_layers takes one of the values [1, 2, 3, 4].

The following advanced discrete hyperparameters can also be specified using a distribution:

QUniform(min_value, max_value, q) - Returns a value like round(Uniform(min_value, max_value) / q) * q

QLogUniform(min_value, max_value, q) - Returns a value like round(exp(Uniform(min_value, max_value)) / q) * q

QNormal(mu, sigma, q) - Returns a value like round(Normal(mu, sigma) / q) * q

*-> QLogNormal(mu, sigma, q) - Returns a value like round(exp(Normal(mu, sigma)) / q) * q

Note: Grid sampling

Grid sampling can only be employed when all hyperparameters are discrete, and is used to try every possible combination of parameters in the search space.

For example, in the following code example, grid sampling is used to try every possible combination of discrete batch_size and learning_rate value:

from azureml.train.hyperdrive import GridParameterSampling, choice

param_space = {

\\'--batch_size\\': choice(16, 32, 64),

\\'--learning_rate\\': choice(0.01, 0.1, 1.0)

}

param_sampling = GridParameterSampling(param_space)

Reference:

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

---

QUESTION 14

DRAG DROP

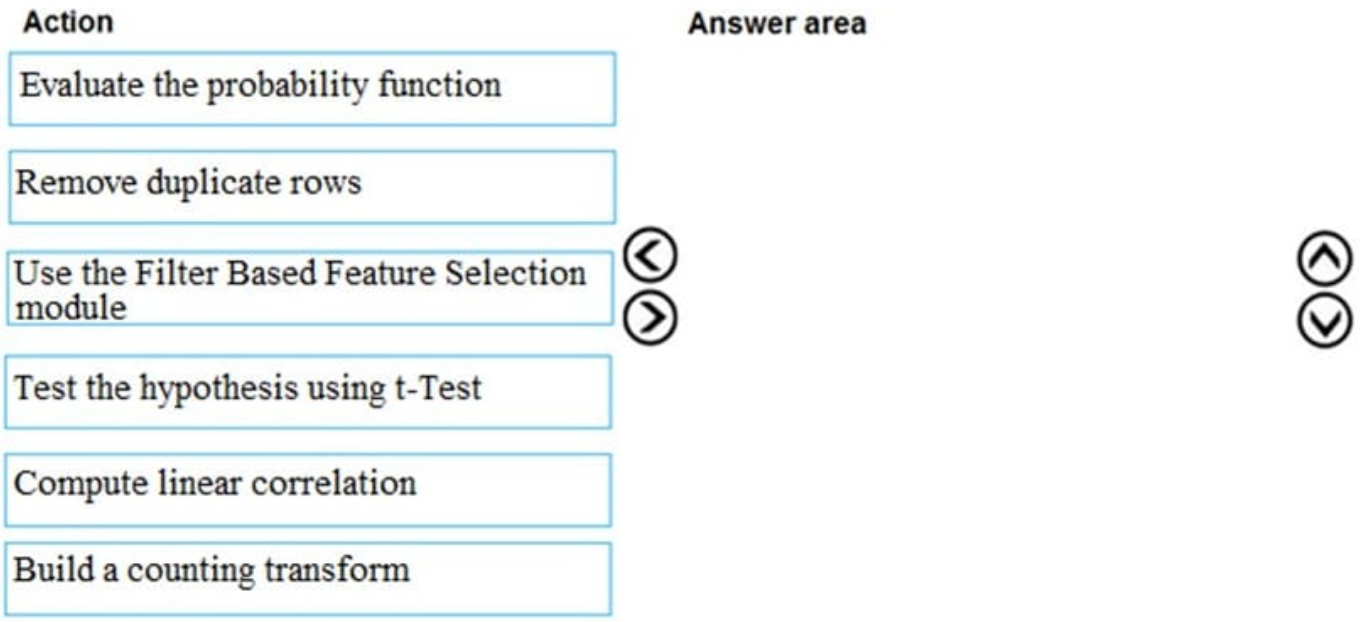
You are producing a multiple linear regression model in Azure Machine Learning Studio.

Several independent variables are highly correlated.

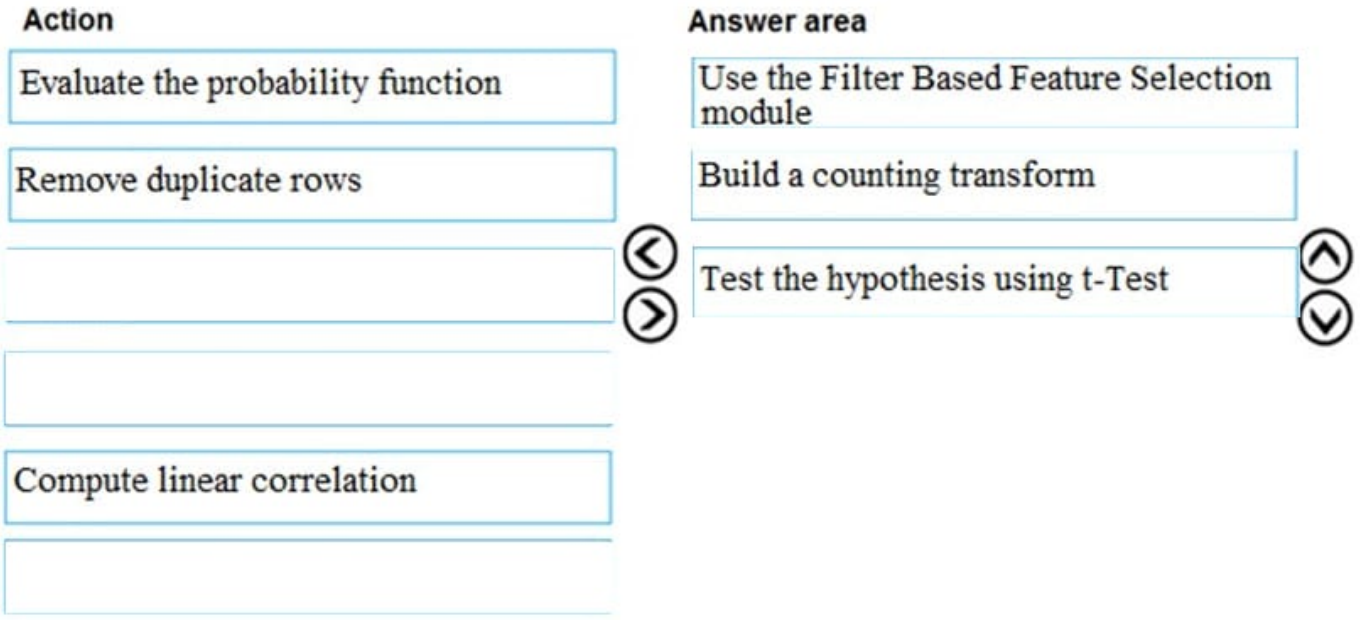
You need to select appropriate methods for conducting effective feature engineering on all the data.

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.

Select and Place:



Correct Answer:



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-selection

https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/build-counting-transform

QUESTION 15

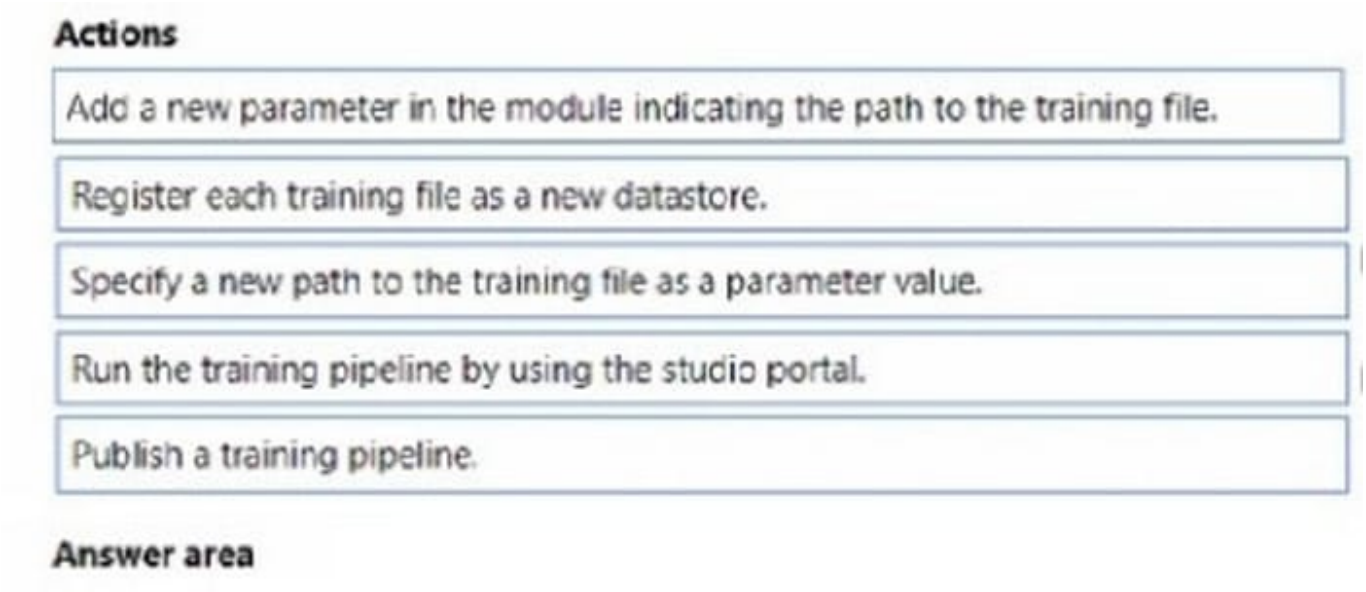
DRAG DROP

You use a training pipeline in the Azure Machine Learning designer. You register a datastore named ds1. The datastore contains multiple training data files. You use the Import Data module with the configured datastore.

You need to retrain a model on a different set of data files.

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.

Select and Place:



Correct Answer:

Actions

- Add a new parameter in the module indicating the path to the training file.
- Specify a new path to the training file as a parameter value.
- Run the training pipeline by using the studio portal.
- Publish a training pipeline.

Answer area

Register each training file as a new datastore.