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.certbus.com/dp-100.html

100% Passing Guarantee
100% Money Back Assurance

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

- **Instant Download** After Purchase
- **100% Money Back** Guarantee
- **365 Days** Free Update
- **800,000+** Satisfied Customers

QUESTION 1

You are creating a classification model for a banking company to identify possible instances of credit card fraud. You plan to create the model in Azure Machine Learning by using automated machine learning. The training dataset that you are using is highly unbalanced.

You need to evaluate the classification model. Which primary metric should you use?

A. normalized_mean_absolute_error

B. AUC_weighted

C. accuracy

D. normalized_root_mean_squared_error

E. spearman_correlation

Correct Answer: B

AUC_weighted is a Classification metric.

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

Incorrect Answers:

A: normalized_mean_absolute_error is a regression metric, not a classification metric.

C: When comparing approaches to imbalanced classification problems, consider using metrics beyond accuracy such as recall, precision, and AUROC. It may be that switching the metric you optimize for during parameter selection or model selection is enough to provide desirable performance detecting the minority class.

D: normalized_root_mean_squared_error is a regression metric, not a classification metric.

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

QUESTION 2

You create an Azure Machine Learning compute resource to train models. The compute resource is configured as follows:

1.

Minimum nodes: 2

2.

Maximum nodes: 4

You must decrease the minimum number of nodes and increase the maximum number of nodes to the following values:

1. Minimum nodes: 0

2. Maximum nodes: 8

You need to reconfigure the compute resource.

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

NOTE: Each correct selection is worth one point.

A. Use the Azure Machine Learning studio.

B. Run the update method of the AmlCompute class in the Python SDK.

C. Use the Azure portal.

D. Use the Azure Machine Learning designer.

E. Run the refresh_state() method of the BatchCompute class in the Python SDK.

Correct Answer: ABC

A: You can manage assets and resources in the Azure Machine Learning studio.

B: The update(min_nodes=None, max_nodes=None, idle_seconds_before_scaledown=None) of the AmlCompute class updates the ScaleSettings for this AmlCompute target.

C: To change the nodes in the cluster, use the UI for your cluster in the Azure portal.

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

QUESTION 3

HOTSPOT

You create a Python script named train.py and save it in a folder named scripts. The script uses the scikit-learn framework to train a machine learning model.

You must run the script as an Azure Machine Learning experiment on your local workstation.

You need to write Python code to initiate an experiment that runs the train.py script.

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.

Hot Area:

Answer Area

```
from azureml.core import Experiment, ScriptRunConfig, Environment
from azureml.core.conda_dependencies import CondaDependencies
from azureml.core import Workspace

ws = Workspace.from_config()
py_sk = Environment('sklearn-training')
pkgs = CondaDependencies.create(pip_packages=['scikit-learn', 'azureml-defaults'])
py_sk.python.conda_dependencies = pkgs
script_config = ScriptRunConfig (
```

[▼] = 'scripts',
- script
- source_directory
- resume_from
- arguments

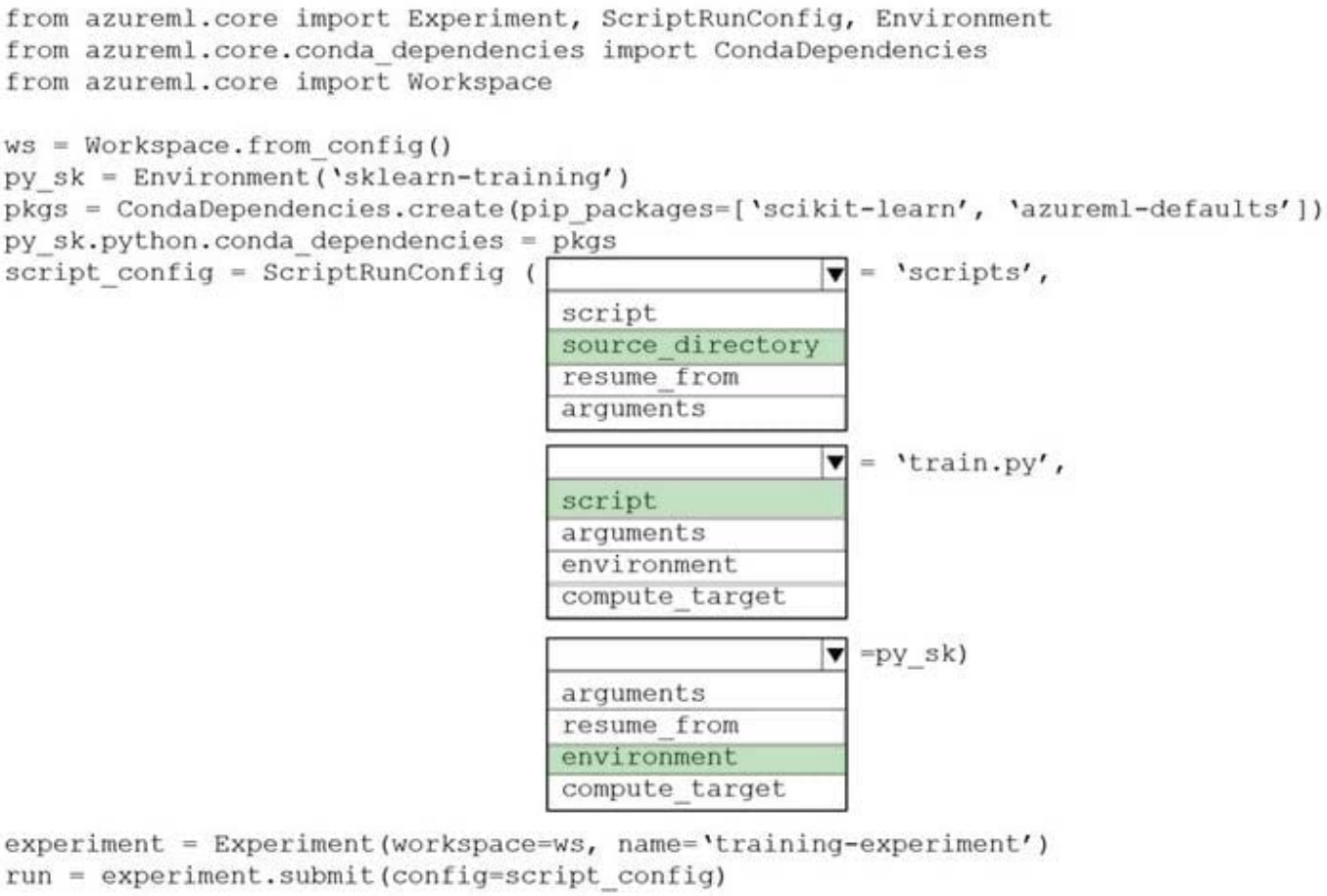
[▼] = 'train.py',
- script
- arguments
- environment
- compute_target

[▼] =py_sk)
- arguments
- resume_from
- environment
- compute_target

```
experiment = Experiment(workspace=ws, name='training-experiment')
run = experiment.submit(config=script_config)
```

Correct Answer:

Answer Area



Box 1: source_directory

source_directory: A local directory containing code files needed for a run.

Box 2: script

Script: The file path relative to the source_directory of the script to be run.

Box 3: environment

Reference:

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

QUESTION 4

HOTSPOT

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

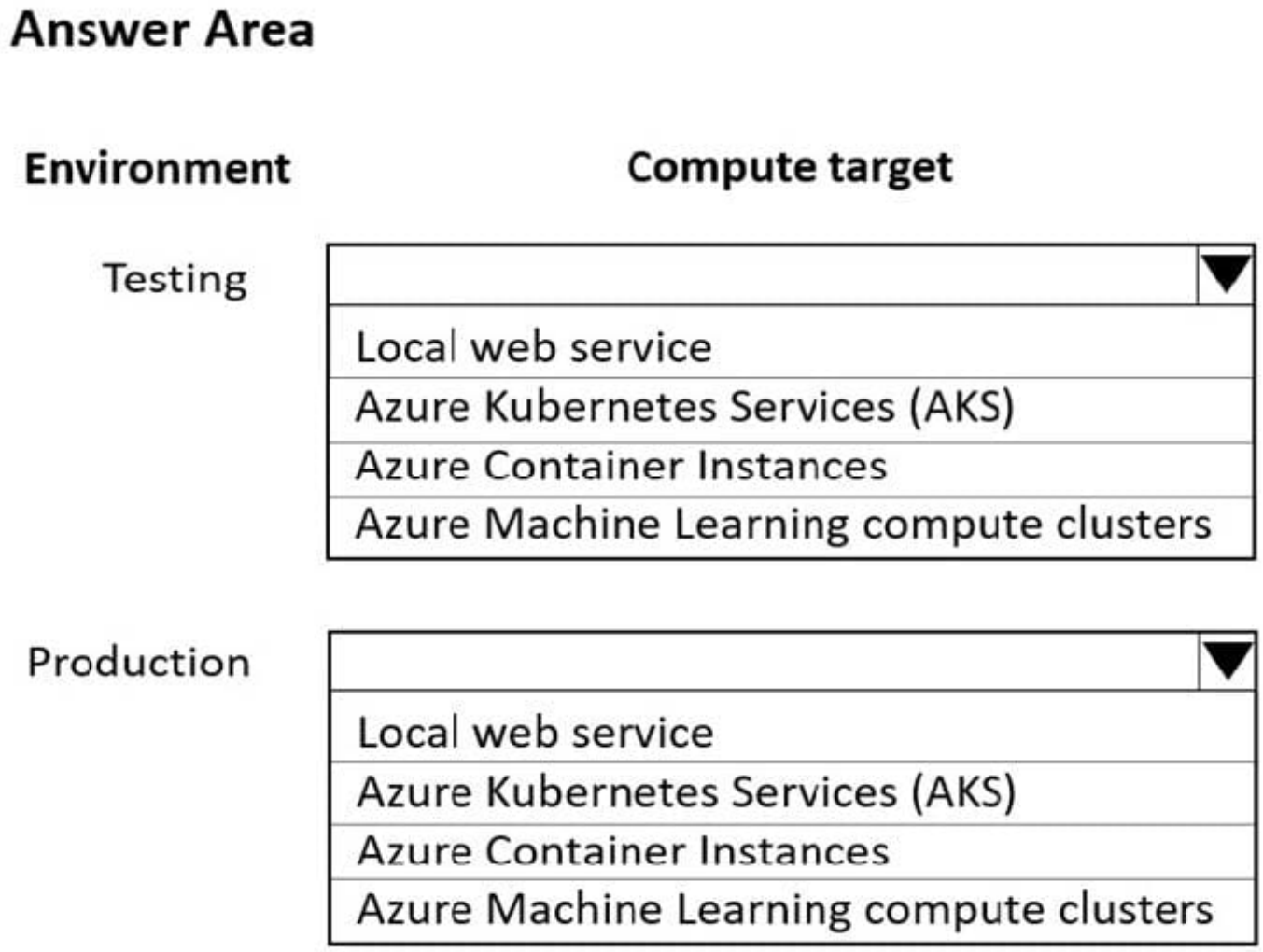
The compute target for testing must minimize cost and deployment efforts. The compute target for production must provide fast response time, autoscaling of the deployed service, and support real-time inferencing.

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.

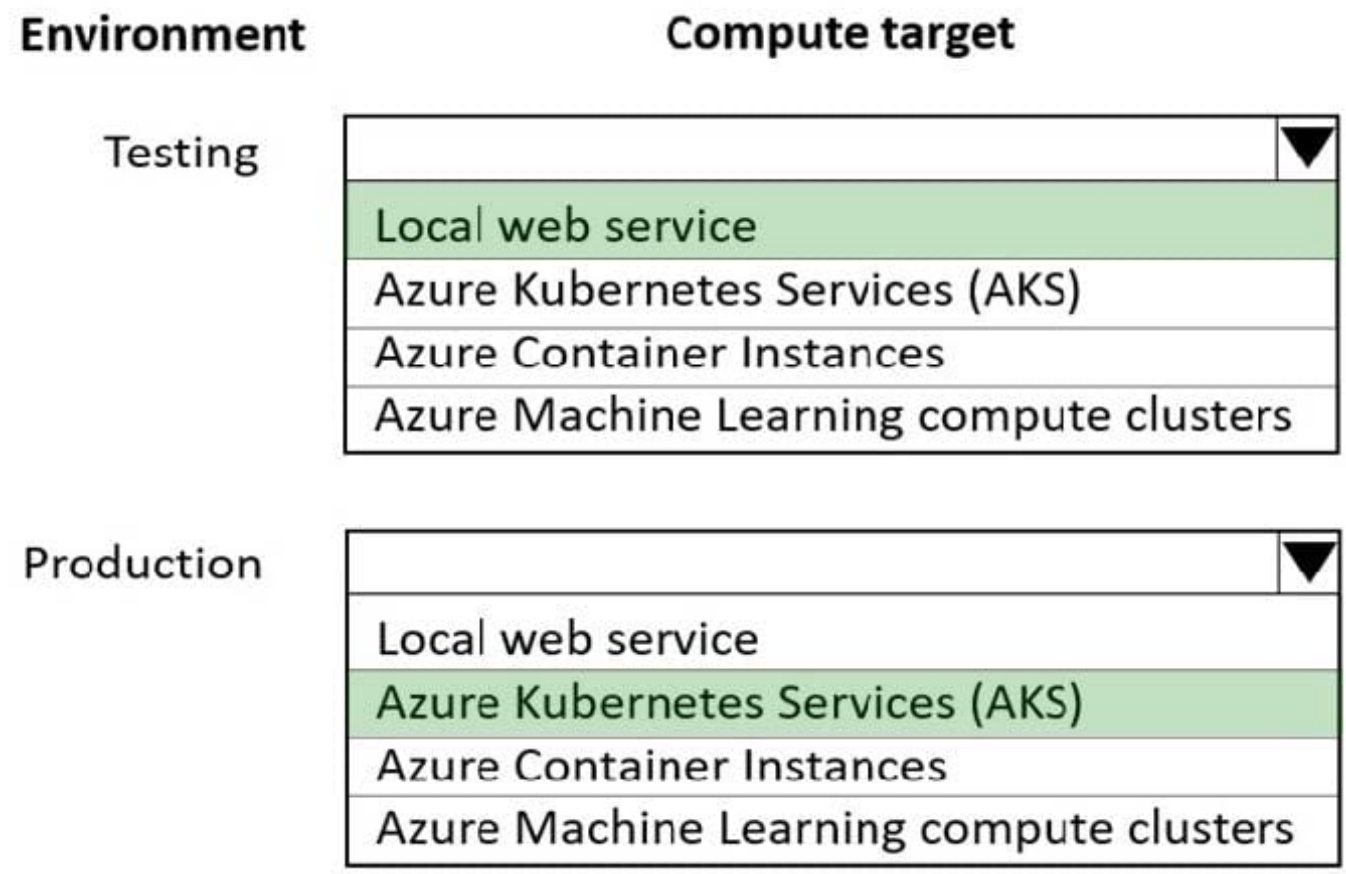
NOTE: Each correct selection is worth one point.

Hot Area:



Correct Answer:

Answer Area



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

QUESTION 5

You are planning to host practical training to acquaint learners with data visualization creation using Python. Learner devices are able to connect to the internet.

Learner devices are currently NOT configured for Python development. Also, learners are unable to install software on their devices as they lack administrator permissions. Furthermore, they are unable to access Azure subscriptions.

It is imperative that learners are able to execute Python-based data visualization code.

Which of the following actions should you take?

A. You should consider configuring the use of Azure Container Instance.

B. You should consider configuring the use of Azure BatchAI.

C. You should consider configuring the use of Azure Notebooks.

D. You should consider configuring the use of Azure Kubernetes Service.

Correct Answer: C

Reference: https://notebooks.azure.com/

QUESTION 6

HOTSPOT

You have machine learning models produce unfair predictions across sensitive features.

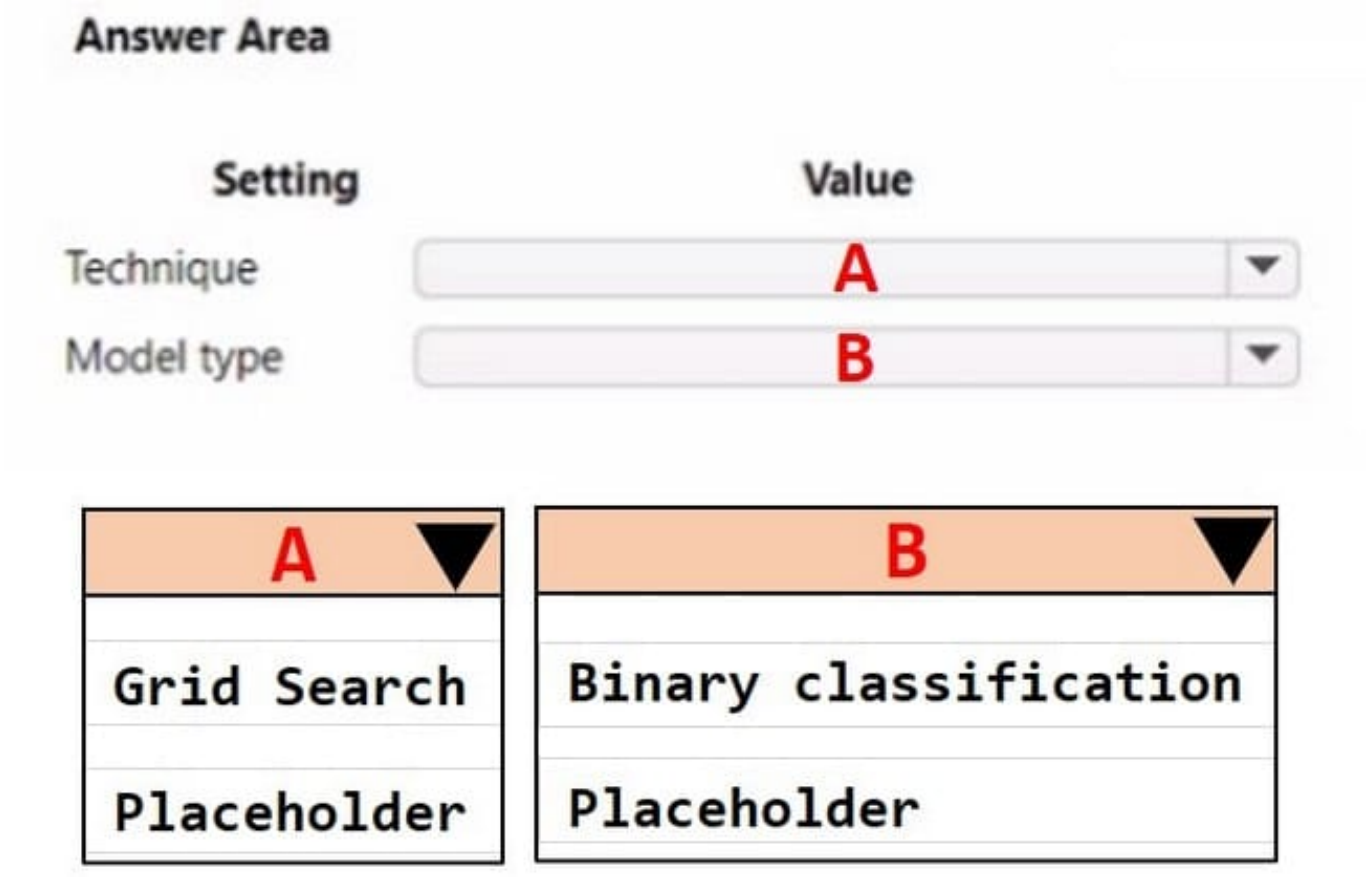
You must use a post-processing technique to apply a constraint to the models to mitigate their unfairness.

You need to select a post-processing technique and model type.

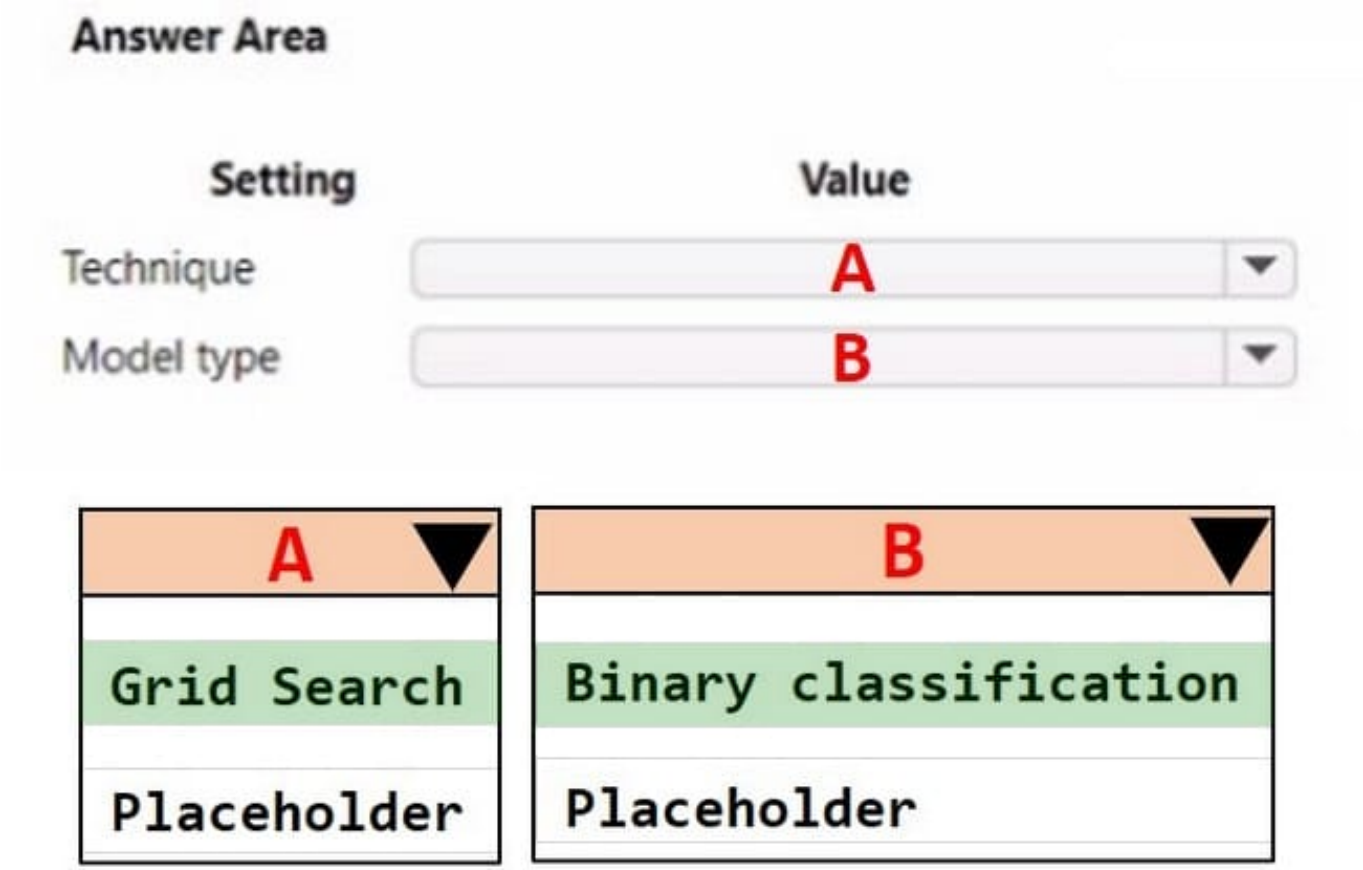
What should you use? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.

Hot Area:



Correct Answer:



QUESTION 7

You use the Azure Machine Learning Python SDK to define a pipeline that consists of multiple steps.

When you run the pipeline, you observe that some steps do not run. The cached output from a previous run is used instead.

You need to ensure that every step in the pipeline is run, even if the parameters and contents of the source directory have not changed since the previous run.

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. Use a PipelineData object that references a datastore other than the default datastore.

B. Set the regenerate_outputs property of the pipeline to True.

C. Set the allow_reuse property of each step in the pipeline to False.

D. Restart the compute cluster where the pipeline experiment is configured to run.

E. Set the outputs property of each step in the pipeline to True.

Correct Answer: BC

B: If regenerate_outputs is set to True, a new submit will always force generation of all step outputs, and disallow data reuse for any step of this run. Once this run is complete, however, subsequent runs may reuse the results of this run.

C: Keep the following in mind when working with pipeline steps, input/output data, and step reuse.

If data used in a step is in a datastore and allow_reuse is True, then changes to the data change won\\'t be detected. If the data is uploaded as part of the snapshot (under the step\\'s source_directory), though this is not recommended, then the

hash will change and will trigger a rerun.

Reference:

https://docs.microsoft.com/en-us/python/api/azureml-pipeline-core/azureml.pipeline.core.pipelinestep

https://github.com/Azure/MachineLearningNotebooks/blob/master/how-to-use-azureml/machine-learning-pipelines/intro-to-pipelines/aml-pipelines-getting-started.ipynb

QUESTION 8

Note: This question is part of a series of questions that present the same scenario. Each question in the series contains a unique solution that might meet the stated goals. Some question sets might have more than one correct solution, while

others might not have a correct solution.

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

You are creating a model to predict the price of a student\\'s artwork depending on the following variables:

the student\\'s length of education, degree type, and art form.

You start by creating a linear regression model.

You need to evaluate the linear regression model.

Solution: Use the following metrics: Mean Absolute Error, Root Mean Absolute Error, Relative Absolute Error, Relative Squared Error, and the Coefficient of Determination.

Does the solution meet the goal?

A. Yes

B. No

Correct Answer: A

The following metrics are reported for evaluating regression models. When you compare models, they are ranked by the metric you select for evaluation.

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

Root mean squared error (RMSE) creates a single value that summarizes the error in the model. By squaring the difference, the metric disregards the difference between over-prediction and under-prediction.

Relative absolute error (RAE) is the relative absolute difference between expected and actual values; relative because the mean difference is divided by the arithmetic mean.

Relative squared error (RSE) similarly normalizes the total squared error of the predicted values by dividing by the total squared error of the actual values.

Mean Zero One Error (MZOE) indicates whether the prediction was correct or not. In other words:

ZeroOneLoss(x,y) = 1 when x!=y; otherwise 0.

Coefficient of determination, often referred to as R2, represents the predictive power of the model as a value between 0 and 1. Zero means the model is random (explains nothing); 1 means there is a perfect fit. However, caution should be

used in interpreting R2 values, as low values can be entirely normal and high values can be suspect.

AUC.

Reference:

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

QUESTION 9

You are evaluating a completed binary classification machine. You need to use the precision as the evaluation metric. Which visualization should you use?

A. scatter plot

B. coefficient of determination

C. Receiver Operating Characteristic CROC) curve

D. Gradient descent

Correct Answer: C

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

References:

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

QUESTION 10

Note: This question is part of a series of questions that present the same scenario. Each question in the series contains a unique solution that might meet the stated goals. Some question sets might have more than one correct solution, while

others might not have a correct solution.

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

You train a classification model by using a logistic regression algorithm.

You must be able to explain the model\\'s predictions by calculating the importance of each feature, both as an overall global relative importance value and as a measure of local importance for a specific set of predictions.

You need to create an explainer that you can use to retrieve the required global and local feature importance values.

Solution: Create a PFIExplainer.

Does the solution meet the goal?

A. Yes

B. No

Correct Answer: A

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

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

QUESTION 11

HOTSPOT

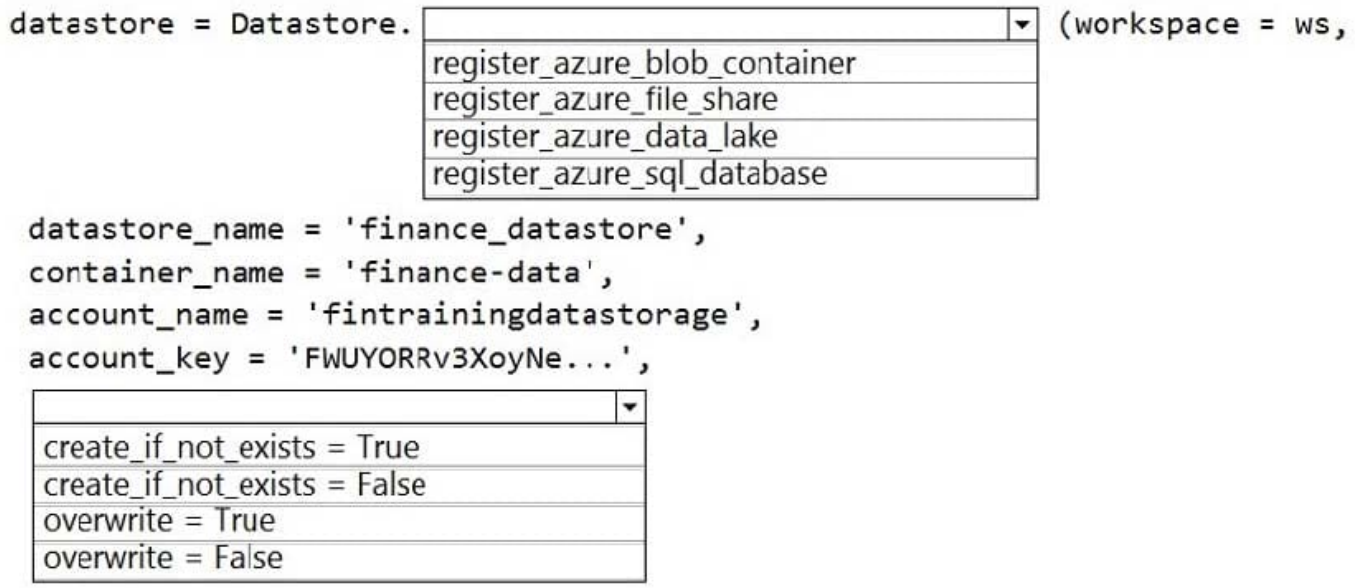
The finance team asks you to train a model using data in an Azure Storage blob container named finance-data.

You need to register the container as a datastore in an Azure Machine Learning workspace and ensure that an error will be raised if the container does not exist.

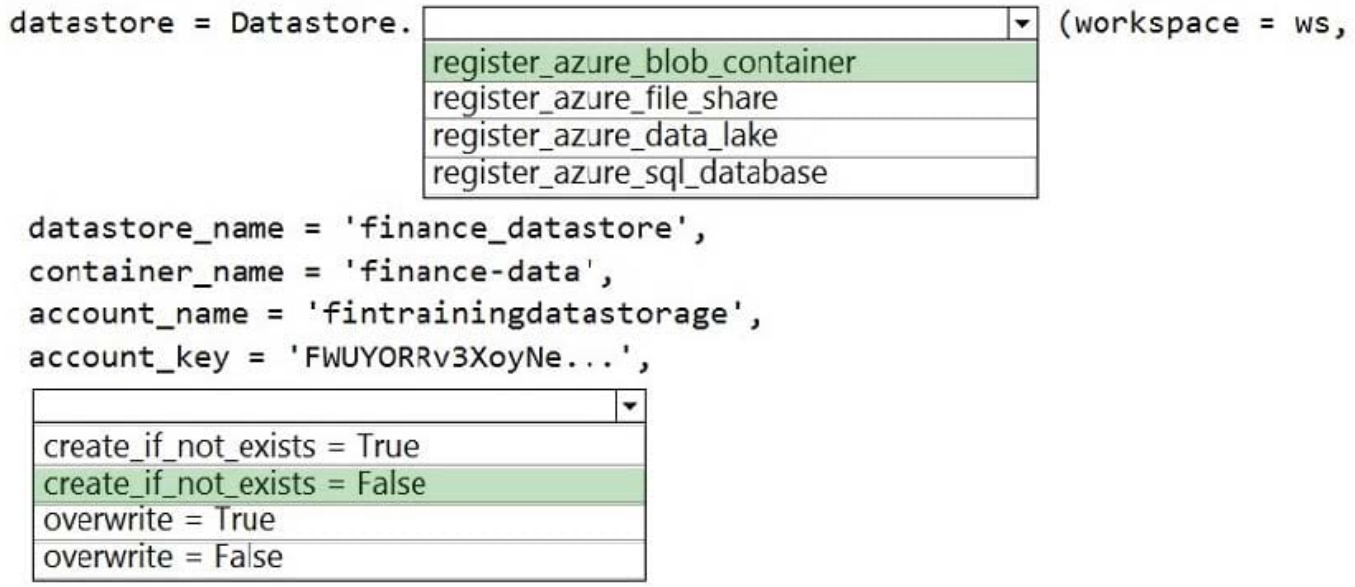
How should you complete the code? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.

Hot Area:



Correct Answer:



Box 1: register_azure_blob_container

Register an Azure Blob Container to the datastore.

Box 2: create_if_not_exists = False

Create the file share if it does not exists, defaults to False.

Reference:

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

QUESTION 12

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_input = PipelineData("raw_data", datastore=rawdatastore)
data_output = PipelineData("processed_data", datastore=datastore)
process_step = PythonScriptStep(script_name="process.py",
 arguments=["--data_for_train", data_input],
 outputs=[data_output], compute_target=aml_compute,
 source_directory=process_directory)
train_step = PythonScriptStep(script_name="train.py",
 arguments=["--data_for_train", data_input], inputs=[data_output],
 compute_target=aml_compute, source_directory=train_directory)
pipeline = Pipeline(workspace=ws, steps=[process_step, train_step])
```

Does the solution meet the goal?

A. Yes

B. No

Correct Answer: B

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

Compare with this example, the pipeline train step depends on the process_step_output output of the pipeline process step:

from azureml.pipeline.core import Pipeline, PipelineData from azureml.pipeline.steps import PythonScriptStep

datastore = ws.get_default_datastore()

process_step_output = PipelineData("processed_data", datastore=datastore) process_step = PythonScriptStep(script_name="process.py", arguments=["--data_for_train", process_step_output],

outputs=[process_step_output],

compute_target=aml_compute,

source_directory=process_directory)

train_step = PythonScriptStep(script_name="train.py",

arguments=["--data_for_train", process_step_output],

inputs=[process_step_output],

compute_target=aml_compute,

source_directory=train_directory)

pipeline = Pipeline(workspace=ws, steps=[process_step, train_step])

Reference:

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

---

QUESTION 13

You use Azure Machine Learning to train a model based on a dataset named dataset1.

You define a dataset monitor and create a dataset named dataset2 that contains new data.

You need to compare dataset1 and dataset2 by using the Azure Machine Learning SDK for Python.

Which method of the DataDriftDetector class should you use?

A. run

B. get

C. backfill

D. update

Correct Answer: C

A backfill run is used to see how data changes over time.

Reference: https://docs.microsoft.com/en-us/python/api/azureml-datadrift/azureml.datadrift.datadriftdetector.datadriftdetector

---

QUESTION 14

You train and publish a machine learning model.

You need to run a pipeline that retrains the model based on a trigger from an external system.

What should you configure?

A. Azure Data Catalog

B. Azure Batch

C. Azure Logic App

Correct Answer: C

An Azure Logic App allows for more complex triggering logic or behavior.

To use an Azure Logic App to trigger a Machine Learning pipeline, you\\'ll need the REST endpoint for a published Machine Learning pipeline. Create and publish your pipeline. Then find the REST endpoint of your PublishedPipeline by using

the pipeline ID.

You can find the pipeline ID in Azure Machine Learning studio

published_pipeline = PublishedPipeline.get(ws, id="")

published_pipeline.endpoint

Note: Azure Logic Apps is a cloud platform where you can create and run automated workflows with little to no code. By using the visual designer and selecting from prebuilt operations, you can quickly build a workflow that integrates and

manages your apps, data, services, and systems.

Azure Logic Apps simplifies the way that you connect legacy, modern, and cutting-edge systems across cloud, on premises, and hybrid environments and provides low-code-no-code tools for you to develop highly scalable integration

solutions for your enterprise and business-to-business (B2B) scenarios.

Reference: https://learn.microsoft.com/en-us/azure/machine-learning/v1/how-to-trigger-published-pipeline

https://learn.microsoft.com/en-us/azure/logic-apps/logic-apps-overview

QUESTION 15

DRAG DROP

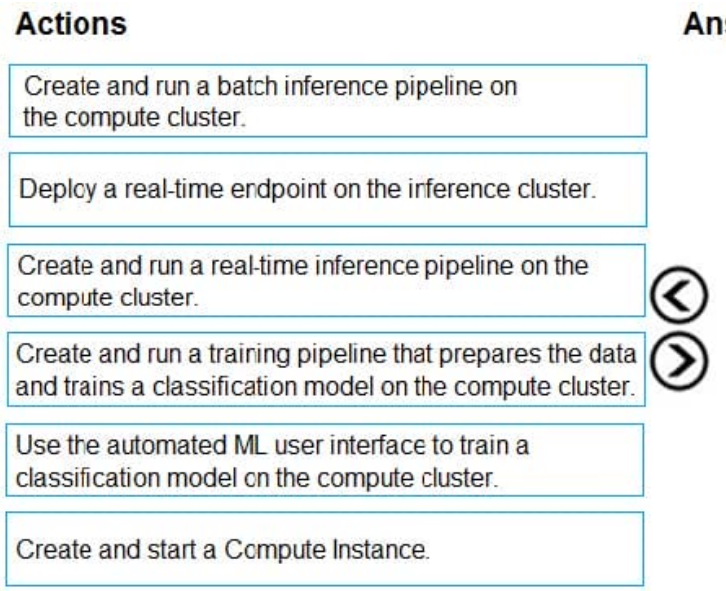
You have an Azure Machine Learning workspace that contains a CPU-based compute cluster and an Azure Kubernetes Services (AKS) inference cluster. You create a tabular dataset containing data that you plan to use to create a

classification model.

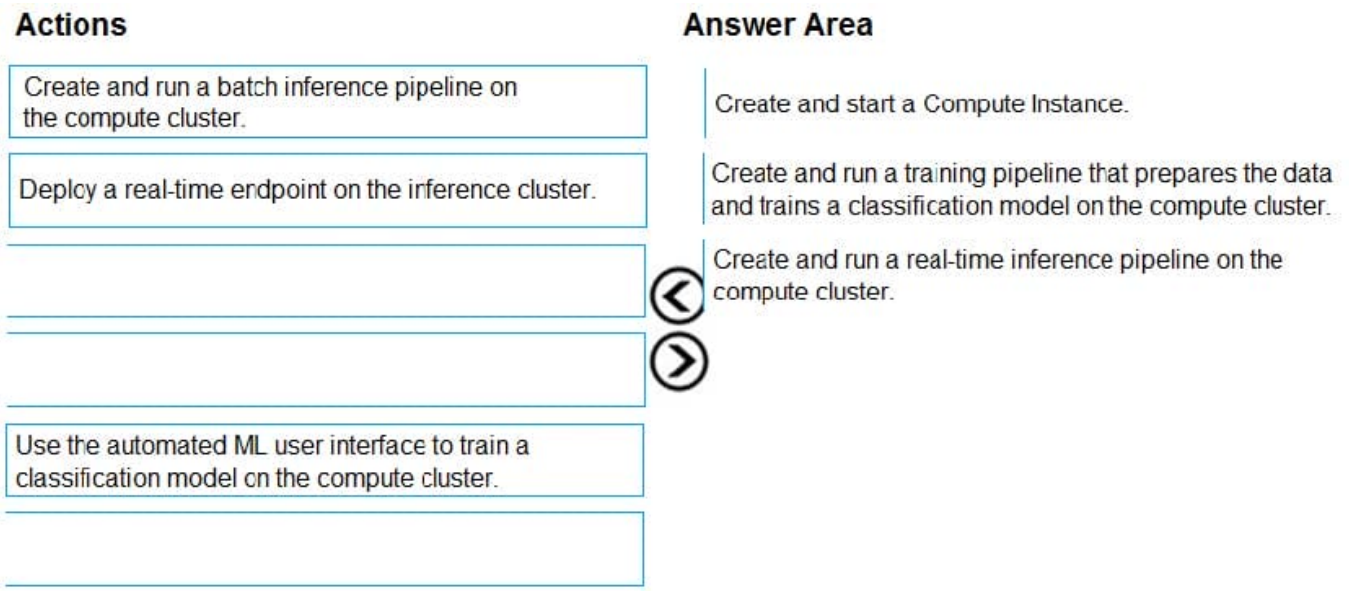
You need to use the Azure Machine Learning designer to create a web service through which client applications can consume the classification model by submitting new data and getting an immediate prediction as a response.

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/