Microsoft.DP-100.vSep-2024.by.Andien.178

Number: DP-100
Passing Score: 800
Time Limit: 120
File Version: 32.0

Exam Code: DP-100
Exam Name: Designing and Implementing a Data Science Solution on Azure

01 - Run experiments and train models

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
Section:
Explanation:
AUC_weighted is a Classification metric.
Note: AUC is the Area under the Receiver Operating Characteristic Curve. Weighted is the arithmetic mean of the score for each class, weighted by the number of true instances in each class.
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 a machine learning model by using the Azure Machine Learning designer. You publish the model as a real-time service on an Azure Kubernetes Service (AKS) inference compute cluster. You make no change to the deployed endpoint configuration.
You need to provide application developers with the information they need to consume the endpoint.
Which two values should you provide to application developers? Each correct answer presents part of the solution.
NOTE: Each correct selection is worth one point.

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

Correct Answer: C, E
Section:
Explanation:
Deploying an Azure Machine Learning model as a web service creates a REST API endpoint. You can send data to this endpoint and receive the prediction returned by the model.
You create a web service when you deploy a model to your local environment, Azure Container Instances, Azure Kubernetes Service, or field-programmable gate arrays (FPGA). You retrieve the URI used to access the web

service by using the Azure Machine Learning SDK. If authentication is enabled, you can also use the SDK to get the authentication keys or tokens.
Example:
# URL for the web service
scoring_uri = '<your web service URI>'
# If the service is authenticated, set the key or token key = '<your key or token>'
Reference: https://docs.microsoft.com/en-us/azure/machine-learning/how-to-consume-web-service

QUESTION 3
Note: This question is part of a series of questions that present the same scenario. Each question in the series contains a unique solution that might meet the stated goals. Some question sets might have more than one correct solution, while others might not have a correct solution.
After you answer a question in this section, you will NOT be able to return to it. As a result, these questions will not appear in the review screen.
You have a Python script named train.py in a local folder named scripts. The script trains a regression model by using scikit-learn. The script includes code to load a training data file which is also located in the scripts folder.
You must run the script as an Azure ML experiment on a compute cluster named aml-compute.
You need to configure the run to ensure that the environment includes the required packages for model training. You have instantiated a variable named aml-compute that references the target compute cluster.
Solution: Run the following code:

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

Does the solution meet the goal?

A. Yes
B. No

Correct Answer: B
Section:
Explanation:
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

QUESTION 4
You are performing clustering by using the K-means algorithm.

You need to define the possible termination conditions.
Which three conditions can you use? Each correct answer presents a complete solution.
NOTE: Each correct selection is worth one point.

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

Correct Answer: A, C, D
Section:
Explanation:
AD: The algorithm terminates when the centroids stabilize or when a specified number of iterations are completed.
C: A measure of how well the centroids represent the members of their clusters is the residual sum of squares or RSS, the squared distance of each vector from its centroid summed over all vectors. RSS is the objective function and our goal is to minimize it.
Reference:
https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/k-means-clustering https://nlp.stanford.edu/IR-book/html/htmledition/k-means-1.html

QUESTION 5
You are building a machine learning model for translating English language textual content into French language textual content.
You need to build and train the machine learning model to learn the sequence of the textual content.
Which type of neural network should you use?

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

Correct Answer: C
Section:
Explanation:
To translate a corpus of English text to French, we need to build a recurrent neural network (RNN).
Note: RNNs are designed to take sequences of text as inputs or return sequences of text as outputs, or both. They're called recurrent because the network's hidden layers have a loop in which the output and cell state from each time step become inputs at the next time step. This recurrence serves as a form of memory. It allows contextual information to flow through the network so that relevant outputs from previous time steps can be applied to network operations at the current time step.
Reference: https://towardsdatascience.com/language-translation-with-rnns-d84d43b40571

QUESTION 6
You create a binary classification model.
You need to evaluate the model performance.
Which two metrics can you use? Each correct answer presents a complete solution.
NOTE: Each correct selection is worth one point.

A. relative absolute error
B. precision
C. accuracy
D. mean absolute error

E. coefficient of determination

Correct Answer: B, C
Section:
Explanation:
The evaluation metrics available for binary classification models are: Accuracy, Precision, Recall, F1 Score, and AUC.
Note: A very natural question-is: 'Out of the individuals whom the model, how many were classified correctly (TP)?'
This question-can be answered by looking at the Precision of the model, which is the proportion of positives that are classified correctly.
Reference:
https://docs.microsoft.com/en-us/azure/machine-learning/studio/evaluate-model-performance

QUESTION 7
You create a script that trains a convolutional neural network model over multiple epochs and logs the validation loss after each epoch. The script includes arguments for batch size and learning rate.
You identify a set of batch size and learning rate values that you want to try.
You need to use Azure Machine Learning to find the combination of batch size and learning rate that results in the model with the lowest validation loss.
What should you do?

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

Correct Answer: E
Section:
Explanation:
Reference: https://docs.microsoft.com/en-us/azure/machine-learning/how-to-tune-hyperparameters

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 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:

```
data_store = Datastore.get(ws, “ml-data”)
data_input = DataReference(
    datastore = data_store,
    data_reference_name = “training_data”,
    path_on_datastore = “train/data.txt”)
data_output = PipelineData(“processed_data”, datastore=datastore)
process_step = PythonScriptStep(script_name= “process.py”,
    arguments=[ “- -data”, data_input], outputs=[data_output],
    compute_target=aml_compute, source_directory=process_directory)
train_step = PythonScriptStep(script_name= “train.py”,
    arguments=[“- -data”, data_output], inputs=[data_output],
    compute_target=aml_compute, source_directory=train_directory)
pipeline = Pipeline(workspace=ws, steps = [process_step, train_step])
```

Does the solution meet the goal?

A. Yes
B. No

Correct Answer: A
Section:
Explanation:
The two steps are present: process_step and train_step Data_input correctly references the data in the data store.
Note:
Data used in pipeline can be produced by one step and consumed in another step by providing a PipelineData object as an output of one step and an input of one or more subsequent steps.
PipelineData objects are also used when constructing Pipelines to describe step dependencies. To specify that a step requires the output of another step as input, use a PipelineData object in the constructor of both steps.
For example, the pipeline train step depends on the process_step_output output of the pipeline process step:
from azureml.pipeline.core import Pipeline, PipelineData from azureml.pipeline.steps import PythonScriptStep
datastore = ws.get_default_datastore()
process_step_output = PipelineData("processed_data", datastore=datastore) process_step = PythonScriptStep(script_name="process.py", arguments=["--data_for_train", process_step_output], outputs=[process_step_output], compute_target=aml_compute, source_directory=process_directory)
train_step = PythonScriptStep(script_name="train.py", arguments=["--data_for_train", process_step_output], inputs=[process_step_output], compute_target=aml_compute, source_directory=train_directory)
pipeline = Pipeline(workspace=ws, steps=[process_step, train_step])
Reference:
https://docs.microsoft.com/en-us/python/api/azureml-pipeline-core/azureml.pipeline.core.pipelinedata?view=azure-ml-py

QUESTION 9
You run an experiment that uses an AutoMLConfig class to define an automated machine learning task with a maximum of ten model training iterations. The task will attempt to find the best performing model based on a metric named accuracy.
You submit the experiment with the following code:

```
from azureml.core.experiment import Experiment
automl_experiment = Experiment(ws, 'automl_experiment')
automl_run = automl_experiment.submit(automl_config, show_output=True)
```

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

A. best_model = automl_run.get_details()
B. best_model = automl_run.get_metrics()
C. best_model = automl_run.get_file_names()[1]
D. best_model = automl_run.get_output()[1]

Correct Answer: D
Section:
Explanation:
The get_output method returns the best run and the fitted model.
Reference:
https://notebooks.azure.com/azureml/projects/azureml-getting-started/html/how-to-use-azureml/automated-machine-learning/classification/auto-ml-classification.ipynb

QUESTION 10
You plan to use the Hyperdrive feature of Azure Machine Learning to determine the optimal hyperparameter values when training a model.
You must use Hyperdrive to try combinations of the following hyperparameter values. You must not apply an early termination policy.
learning_rate: any value between 0.001 and 0.1
batch_size: 16, 32, or 64

You need to configure the sampling method for the Hyperdrive experiment.
Which two sampling methods can you use? Each correct answer is a complete solution.
NOTE: Each correct selection is worth one point.

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

Correct Answer: C, D
Section:
Explanation:
C: Bayesian sampling is based on the Bayesian optimization algorithm and makes intelligent choices on the hyperparameter values to sample next. It picks the sample based on how the previous samples performed, such that the new sample improves the reported primary metric.
Bayesian sampling does not support any early termination policy
Example:
from azureml.train.hyperdrive import BayesianParameterSampling
from azureml.train.hyperdrive import uniform, choice
param_sampling = BayesianParameterSampling({
"learning_rate": uniform(0.05, 0.1),
"batch_size": choice(16, 32, 64, 128)
}
)
D: In random sampling, hyperparameter values are randomly selected from the defined search space. Random sampling allows the search space to include both discrete and continuous hyperparameters.
Incorrect Answers:
B: Grid sampling can be used if your hyperparameter space can be defined as a choice among discrete values and if you have sufficient budget to exhaustively search over all values in the defined search space. Additionally, one can use automated early termination of poorly performing runs, which reduces wastage of resources.
Example, the following space has a total of six samples:
from azureml.train.hyperdrive import GridParameterSampling
from azureml.train.hyperdrive import choice
param_sampling = GridParameterSampling({
"num_hidden_layers": choice(1, 2, 3),
"batch_size": choice(16, 32)
}
)
Reference:
https://docs.microsoft.com/en-us/azure/machine-learning/how-to-tune-hyperparameters

QUESTION 11
You are training machine learning models in Azure Machine Learning. You use Hyperdrive to tune the hyperparameter.
In previous model training and tuning runs, many models showed similar performance.
You need to select an early termination policy that meets the following requirements:
accounts for the performance of all previous runs when evaluating the current run
avoids comparing the current run with only the best performing run to date
Which two early termination policies should you use? Each correct answer presents part of the solution.
NOTE: Each correct selection is worth one point.

A. Median stopping
B. Bandit

C. Default

D. Truncation selection

Correct Answer: A, D
Section:
Explanation:
The Median Stopping policy computes running averages across all runs and cancels runs whose best performance is worse than the median of the running averages.If no policy is specified, the hyperparameter tuning service will let all training runs execute to completion.
Reference:https://docs.microsoft.com/en-us/python/api/azureml-train- core/azureml.train.hyperdrive.medianstoppingpolicy
https://docs.microsoft.com/en-us/python/api/azureml-train- core/azureml.train.hyperdrive.truncationselectionpolicy
https://docs.microsoft.com/en-us/python/api/azureml-train- core/azureml.train.hyperdrive.banditpolicy

QUESTION 12
You use the Azure Machine Learning SDK in a notebook to run an experiment using a script file in an experiment folder.
The experiment fails.
You need to troubleshoot the failed experiment.
What are two possible ways to achieve this goal? Each correct answer presents a complete solution.

A. Use the get_metrics() method of the run object to retrieve the experiment run logs.

B. Use the get_details_with_logs() method of the run object to display the experiment run logs.

C. View the log files for the experiment run in the experiment folder.

D. View the logs for the experiment run in Azure Machine Learning studio.

E. Use the get_output() method of the run object to retrieve the experiment run logs.

Correct Answer: B, D
Section:
Explanation:
Use get_details_with_logs() to fetch the run details and logs created by the run.
You can monitor Azure Machine Learning runs and view their logs with the Azure Machine Learning studio.
Incorrect Answers:
A: You can view the metrics of a trained model using run.get_metrics(). E: get_output() gets the output of the step as PipelineData.
Reference:
https://docs.microsoft.com/en-us/python/api/azureml-pipeline-core/azureml.pipeline.core.steprun https://docs.microsoft.com/en-us/azure/machine-learning/how-to-monitor-view-training-logs

QUESTION 13
You are analyzing a dataset containing historical data from a local taxi company. You are developing a regression model.
You must predict the fare of a taxi trip.
You need to select performance metrics to correctly evaluate the regression model.
Which two metrics can you use? Each correct answer presents a complete solution?
NOTE: Each correct selection is worth one point.

A. a Root Mean Square Error value that is low

B. an R-Squared value close to 0

C. an F1 score that is low

D. an R-Squared value close to 1

E. an F1 score that is high

F. a Root Mean Square Error value that is high

Correct Answer: A, D
Section:
Explanation:
RMSE and R2 are both metrics for regression models.
A: 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.
D: 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.
Incorrect Answers:
C, E: F-score is used for classification models, not for regression models.
Reference: https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/evaluate-model

QUESTION 14
Note: This question is part of a series of questions that present the same scenario. Each question in the series contains a unique solution that might meet the stated goals. Some question sets might have more than one correct solution, while others might not have a correct solution.
After you answer a question in this section, you will NOT be able to return to it. As a result, these questions will not appear in the review screen.
You are using Azure Machine Learning to run an experiment that trains a classification model.
You want to use Hyperdrive to find parameters that optimize the AUC metric for the model. You configure a HyperDriveConfig for the experiment by running the following code:

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

You plan to use this configuration to run a script that trains a random forest model and then tests it with validation data. The label values for the validation data are stored in a variable named y_test variable, and the predicted probabilities from the model are stored in a variable named y_predicted.
You need to add logging to the script to allow Hyperdrive to optimize hyperparameters for the AUC metric.
Solution: Run the following code:

```
from sklearn.metrics import roc_auc_score
import logging
# code to train model omitted
auc = roc_auc_score(y_test, y_predicted)
logging.info("AUC: " + str(auc))
```

Does the solution meet the goal?

A. Yes
B. No

Correct Answer: A
Section:
Explanation:
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

QUESTION 15
Note: This question is part of a series of questions that present the same scenario. Each question in the series contains a unique solution that might meet the stated goals. Some question sets might have more than one

correct solution, while others might not have a correct solution.
After you answer a question in this section, you will NOT be able to return to it. As a result, these questions will not appear in the review screen.
You are using Azure Machine Learning to run an experiment that trains a classification model.
You want to use Hyperdrive to find parameters that optimize the AUC metric for the model. You configure a HyperDriveConfig for the experiment by running the following code:

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

You plan to use this configuration to run a script that trains a random forest model and then tests it with validation data. The label values for the validation data are stored in a variable named y_test variable, and the predicted probabilities from the model are stored in a variable named y_predicted.
You need to add logging to the script to allow Hyperdrive to optimize hyperparameters for the AUC metric.
Solution: Run the following code:

```
import 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

Correct Answer: B
Section:
Explanation:
Use a solution with logging.info(message) instead.
Note: Python printing/logging example: logging.info(message)
Destination: Driver logs, Azure Machine Learning designer
Reference: https://docs.microsoft.com/en-us/azure/machine-learning/how-to-debug-pipelines

QUESTION 16
Note: This question is part of a series of questions that present the same scenario. Each question in the series contains a unique solution that might meet the stated goals. Some question sets might have more than one correct solution, while others might not have a correct solution.
After you answer a question in this section, you will NOT be able to return to it. As a result, these questions will not appear in the review screen.
You are using Azure Machine Learning to run an experiment that trains a classification model.
You want to use Hyperdrive to find parameters that optimize the AUC metric for the model. You configure a HyperDriveConfig for the experiment by running the following code:

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

You plan to use this configuration to run a script that trains a random forest model and then tests it with validation data. The label values for the validation data are stored in a variable named y_test variable, and the predicted probabilities from the model are stored in a variable named y_predicted.
You need to add logging to the script to allow Hyperdrive to optimize hyperparameters for the AUC metric.
Solution: Run the following code:

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

Does the solution meet the goal?

A. Yes
B. No

Correct Answer: B
Section:
Explanation:
Use a solution with logging.info(message) instead.
Note: Python printing/logging example: logging.info(message)
Destination: Driver logs, Azure Machine Learning designer
Reference: https://docs.microsoft.com/en-us/azure/machine-learning/how-to-debug-pipelines

QUESTION 17
You use the following code to run a script as an experiment in Azure Machine Learning:

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

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

A. files = run.get_properties()
B. files= run.get_file_names()
C. files = run.get_details_with_logs()

D. files = run.get_metrics()

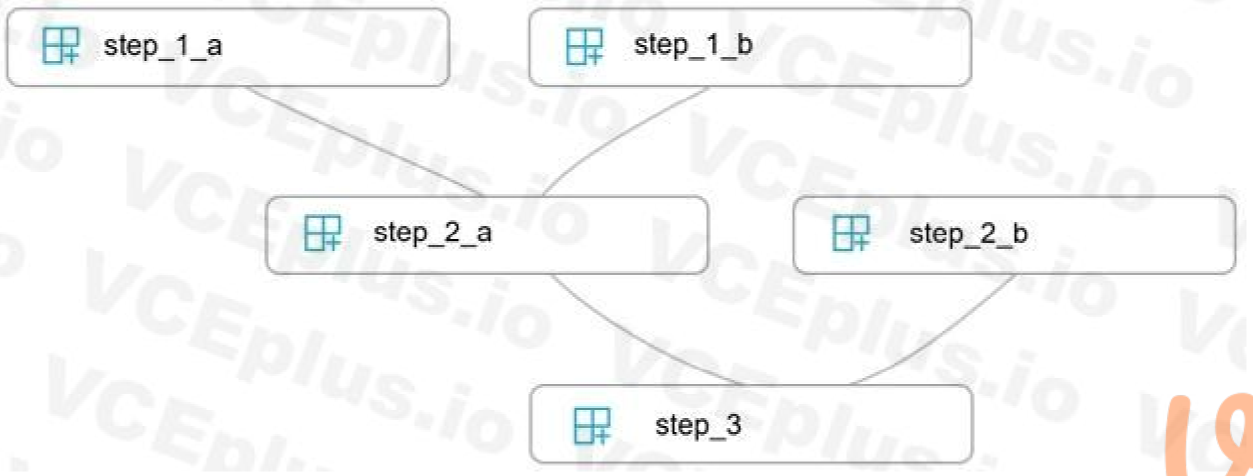
E. files = run.get_details()

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

QUESTION 18

You write five Python scripts that must be processed in the order specified in Exhibit A – which allows the same modules to run in parallel, but will wait for modules with dependencies.

You must create an Azure Machine Learning pipeline using the Python SDK, because you want to script to create the pipeline to be tracked in your version control system. You have created five PythonScriptSteps and have named the variables to match the module names.



You need to create the pipeline shown. Assume all relevant imports have been done.
Which Python code segment should you use?

A.
```
p = Pipeline(ws, steps=[[[[step_1_a, step_1_b], step_2_a],  step_2_b], step_3])
```

B.
```
pipeline_steps = {
    "Pipeline": {
        "run": step_3,
        "run_after": {[
            {"run": step_2_a,
                "run_after":
                    [{"run": step_1_a},
                     {"run": step_1_b}]
                },
            {"run": step_2_b}]
        }
    }
}
p = Pipeline(ws, steps=pipeline_steps)
```

C.

```
step_2_a.run_after(step_1_b)
step_2_a.run_after(step_1_a)
step_3.run_after(step_2_b)
step_3.run_after(step_2_a)
p = Pipeline(ws, steps=[step_3])
```

D.

```
p = Pipeline(ws, steps=[step_1_a, step_1_b, step_2_a, step_2_b, step_3])
```

Correct Answer: A
Section:
Explanation:
The steps parameter is an array of steps. To build pipelines that have multiple steps, place the steps in order in this array.
Reference: https://docs.microsoft.com/en-us/azure/machine-learning/how-to-use-parallel-run-step

QUESTION 19
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')
```

Correct Answer: B
Section:
Explanation:
Besides passing the dataset through the input parameters 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

QUESTION 20
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: Attach the mlvm virtual machine as a compute target in the Azure Machine Learning workspace. Install the Azure ML SDK on the Surface Book and run Python code to connect to the workspace. Run the training script as an experiment on the mlvm remote compute resource.
Does the solution meet the goal?

A. Yes
B. No

Correct Answer: A
Section:
Explanation:
Use the VM as a compute target.
Note: A compute target is a designated compute resource/environment where you run your training script or host your service deployment. This location may be your local machine or a cloud-based compute resource.
Reference:
https://docs.microsoft.com/en-us/azure/machine-learning/concept-compute-target

QUESTION 21
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

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

QUESTION 22
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. Run the training script as an experiment on the aks-cluster compute target.
Does the solution meet the goal?

A. Yes

B. No

Correct Answer: B

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

QUESTION 23
You create a batch inference pipeline by using the Azure ML SDK. You configure the pipeline parameters by executing the following code:

```
from azureml.contrib.pipeline.steps import ParallelRunConfig
parallel_run_config = ParallelRunConfig(
     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. the digit_identification.py script
B. the debug log
C. the Activity Log in the Azure portal for the Machine Learning workspace
D. the Inference Clusters tab in Machine Learning studio
E. a file named parallel_run_step.txt located in the output folder

Correct Answer: E
Section:
Explanation:
output_action (str): How the output is to be organized. Currently supported values are 'append_row' and 'summary_only'.
'append_row' - All values output by run() method invocations will be aggregated into one unique file named parallel_run_step.txt that is created in the output location. 'summary_only'
Reference:
https://docs.microsoft.com/en-us/python/api/azureml-contrib-pipeline-steps/azureml.contrib.pipeline.steps.parallelrunconfig

QUESTION 24
You plan to run a script as an experiment using a Script Run Configuration. The script uses modules from the scipy library as well as several Python packages that are not typically installed in a default conda environment.
You plan to run the experiment on your local workstation for small datasets and scale out the experiment by running it on more powerful remote compute clusters for larger datasets.
You need to ensure that the experiment runs successfully on local and remote compute with the least administrative effort.
What should you do?

A. Do not specify an environment in the run configuration for the experiment. Run the experiment by using the default environment.
B. Create a virtual machine (VM) with the required Python configuration and attach the VM as a compute target. Use this compute target for all experiment runs.
C. Create and register an Environment that includes the required packages. Use this Environment for all experiment runs.
D. Create a config.yaml file defining the conda packages that are required and save the file in the experiment folder.
E. Always run the experiment with an Estimator by using the default packages.

Correct Answer: C
Section:
Explanation:
If you have an existing Conda environment on your local computer, then you can use the service to create an environment object. By using this strategy, you can reuse your local interactive environment on remote runs.
Reference: https://docs.microsoft.com/en-us/azure/machine-learning/how-to-use-environments

QUESTION 25
You write a Python script that processes data in a comma-separated values (CSV) file.
You plan to run this script as an Azure Machine Learning experiment.
The script loads the data and determines the number of rows it contains using the following code:

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

run = Run.get_context()
data = pd.read_csv('./data.csv')
rows = (len(data))
# record row_count metric here
...
```

You need to record the row count as a metric named row_count that can be returned using the get_metrics method of the Run object after the experiment run completes.
Which code should you use?

A. run.upload_file(T3 row_count', './data.csv')
B. run.log('row_count', rows)
C. run.tag('row_count', rows)
D. run.log_table('row_count', rows)
E. run.log_row('row_count', rows)

Correct Answer: B
Section:
Explanation:
Log a numerical or string value to the run with the given name using log(name, value, description=''). 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)
Incorrect Answers:
E: Using log_row(name, description=None, **kwargs) creates a metric with multiple columns as described in kwargs. Each named parameter generates a column with the value specified. log_row can be called once to log an arbitrary tuple, or multiple times in a loop to generate a complete table.
Example: run.log_row("Y over X", x=1, y=0.4)
Reference: https://docs.microsoft.com/en-us/python/api/azureml-core/azureml.core.run

QUESTION 26
Note: This question-is part of a series of questions that present the same scenario. Each question-in the series contains a unique solution that might meet the stated goals. Some question-sets might have more than one correct solution, while others might not have a correct solution.
After you answer a question-in this section, you will NOT be able to return to it. As a result, these questions will not appear in the review screen.
You are creating a new experiment in Azure Machine Learning Studio.
One class has a much smaller number of observations than the other classes in the training set.
You need to select an appropriate data sampling strategy to compensate for the class imbalance.
Solution: You use the Synthetic Minority Oversampling Technique (SMOTE) sampling mode.
Does the solution meet the goal?

A. Yes
B. No

Correct Answer: A
Section:
Explanation:
SMOTE is used 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.
Reference:
https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/smote

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 are creating a new experiment in Azure Machine Learning Studio.
One class has a much smaller number of observations than the other classes in the training set.
You need to select an appropriate data sampling strategy to compensate for the class imbalance.
Solution: You use the Stratified split for the sampling mode.
Does the solution meet the goal?

A. Yes
B. No

Correct Answer: B
Section:
Explanation:
Instead use the Synthetic Minority Oversampling Technique (SMOTE) sampling mode.
Note: SMOTE is used 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.
Reference: https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/smote

QUESTION 28
You are creating a machine learning model.
You need to identify outliers in the data.
Which two visualizations can you use? Each correct answer presents a complete solution.
NOTE: Each correct selection is worth one point.

A. Venn diagram
B. Box plot
C. ROC curve
D. Random forest diagram
E. Scatter plot

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

QUESTION 29
You are evaluating a completed binary classification machine learning model.
You need to use the precision as the evaluation metric.
Which visualization should you use?

A. Violin plot
B. Gradient descent
C. Box plot
D. Binary classification confusion matrix

Correct Answer: D
Section:
Explanation:
Incorrect Answers:
A: A violin plot is a visual that traditionally combines a box plot and a kernel density plot.
B: Gradient descent is a first-order iterative optimization algorithm for finding the minimum of a function. To find a local minimum of a function using gradient descent, one takes steps proportional to the negative of the gradient (or approximate gradient) of the function at the current point.
C: A box plot lets you see basic distribution information about your data, such as median, mean, range and quartiles but doesn't show you how your data looks throughout its range.
Reference:
https://machinelearningknowledge.ai/confusion-matrix-and-performance-metrics-machine-learning/

QUESTION 30
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.

Correct Answer: A, D, F
Section:
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.

QUESTION 31
You are performing a filter-based feature selection for a dataset to build a multi-class classifier by using Azure Machine Learning Studio.
The dataset contains categorical features that are highly correlated to the output label column.
You need to select the appropriate feature scoring statistical method to identify the key predictors.

Which method should you use?

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

Correct Answer: D
Section:
Explanation:
Pearson's correlation statistic, or Pearson's correlation coefficient, is also known in statistical models as the r value. For any two variables, it returns a value that indicates the strength of the correlation Pearson's correlation coefficient is the test statistics that measures the statistical relationship, or association, between two continuous variables. It is known as the best method of measuring the association between variables of interest because it is based on the method of covariance. It gives information about the magnitude of the association, or correlation, as well as the direction of the relationship.
Incorrect Answers:
C: The two-way chi-squared test is a statistical method that measures how close expected values are to actual results.
Reference:
https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/filter-based-feature-selection https://www.statisticssolutions.com/pearsons-correlation-coefficient/

QUESTION 32
You plan to use automated machine learning to train a regression model. You have data that has features which have missing values, and categorical features with few distinct values.
You need to configure automated machine learning to automatically impute missing values and encode categorical features as part of the training task.
Which parameter and value pair should you use in the AutoMLConfig class?

A. featurization = 'auto'
B. enable_voting_ensemble = True
C. task = 'classification'
D. exclude_nan_labels = True
E. enable_tf = True

Correct Answer: A
Section:
Explanation:
Featurization str or FeaturizationConfig
Values: 'auto' / 'off' / FeaturizationConfig
Indicator for whether featurization step should be done automatically or not, or whether customized featurization should be used.
Column type is automatically detected. Based on the detected column type preprocessing/featurization is done as follows:
Categorical: Target encoding, one hot encoding, drop high cardinality categories, impute missing values.
Numeric: Impute missing values, cluster distance, weight of evidence.
DateTime: Several features such as day, seconds, minutes, hours etc.
Text: Bag of words, pre-trained Word embedding, text target encoding.
Reference:
https://docs.microsoft.com/en-us/python/api/azureml-train-automl-client/azureml.train.automl.automlconfig.automlconfig

QUESTION 33
You are building a regression model for estimating the number of calls during an event.
You need to determine whether the feature values achieve the conditions to build a Poisson regression model.
Which two conditions must the feature set contain? Each correct answer presents part of the solution.
NOTE: Each correct selection is worth one point.

A. The label data must be a negative value.
B. The label data must be whole numbers.
C. The label data must be non-discrete.
D. The label data must be a positive value.
E. The label data can be positive or negative.

Correct Answer: B, D
Section:
Explanation:
Poisson regression is intended for use in regression models that are used to predict numeric values, typically counts. Therefore, you should use this module to create your regression model only if the values you are trying to predict fit the following conditions:
The response variable has a Poisson distribution.
Counts cannot be negative. The method will fail outright if you attempt to use it with negative labels.
A Poisson distribution is a discrete distribution; therefore, it is not meaningful to use this method with non-whole numbers.
Reference:
https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/poisson-regression

QUESTION 34
Note: This question-is part of a series of questions that present the same scenario. Each question-in the series contains a unique solution that might meet the stated goals. Some question-sets might have more than one correct solution, while others might not have a correct solution.
After you answer a question-in this section, you will NOT be able to return to it. As a result, these questions will not appear in the review screen.
You are creating a new experiment in Azure Machine Learning Studio.
One class has a much smaller number of observations than the other classes in the training set.
You need to select an appropriate data sampling strategy to compensate for the class imbalance.
Solution: You use the Principal Components Analysis (PCA) sampling mode.
Does the solution meet the goal?

A. Yes
B. No

Correct Answer: B
Section:
Explanation:
Instead use the Synthetic Minority Oversampling Technique (SMOTE) sampling mode.
Note: SMOTE is used 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.
Incorrect Answers:
The Principal Component Analysis module in Azure Machine Learning Studio (classic) is used to reduce the dimensionality of your training data. The module analyzes your data and creates a reduced feature set that captures all the information contained in the dataset, but in a smaller number of features.
Reference: 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/principal-component-analysis

QUESTION 35
You are performing feature engineering on a dataset.
You must add a feature named CityName and populate the column value with the text London.
You need to add the new feature to the dataset.
Which Azure Machine Learning Studio module should you use?

A. Edit Metadata

B. Filter Based Feature Selection
C. Execute Python Script
D. Latent Dirichlet Allocation

Correct Answer: A
Section:
Explanation:
Typical metadata changes might include marking columns as features.
Reference:
https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/edit-metadata

QUESTION 36
You are evaluating a completed binary classification machine learning model.
You need to use the precision as the evaluation metric.
Which visualization should you use?

A. violin plot
B. Gradient descent
C. Scatter plot
D. Receiver Operating Characteristic (ROC) curve

Correct Answer: D
Section:
Explanation:
Receiver operating characteristic (or ROC) is a plot of the correctly classified labels vs. the incorrectly classified labels for a particular model.
Incorrect Answers:
A: A violin plot is a visual that traditionally combines a box plot and a kernel density plot.
B: Gradient descent is a first-order iterative optimization algorithm for finding the minimum of a function. To find a local minimum of a function using gradient descent, one takes steps proportional to the negative of the gradient (or approximate gradient) of the function at the current point.
C: A scatter plot graphs the actual values in your data against the values predicted by the model. The scatter plot displays the actual values along the X-axis, and displays the predicted values along the Y-axis. It also displays a line that illustrates the perfect prediction, where the predicted value exactly matches the actual value.
Reference:
https://docs.microsoft.com/en-us/azure/machine-learning/how-to-understand-automated-ml#confusion-matrix

QUESTION 37
You are solving a classification task.
You must evaluate your model on a limited data sample by using k-fold cross-validation. You start by configuring a k parameter as the number of splits.
You need to configure the k parameter for the cross-validation.
Which value should you use?

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

Correct Answer: B
Section:
Explanation:

Leave One Out (LOO) cross-validation
Setting K = n (the number of observations) yields n-fold and is called leave-one out cross-validation (LOO), a special case of the K-fold approach.
LOO CV is sometimes useful but typically doesn't shake up the data enough. The estimates from each fold are highly correlated and hence their average can have high variance. This is why the usual choice is K=5 or 10. It provides a good compromise for the bias-variance tradeoff.

QUESTION 38
You use the Azure Machine Learning service to create a tabular dataset named training_data. You plan to use this dataset in a training script.
You create a variable that references the dataset using the following code:
training_ds = workspace.datasets.get("training_data")
You define an estimator to run the script.
You need to set the correct property of the estimator to ensure that your script can access the training_data dataset.
Which property should you set?

A. environment_definition = {"training_data":training_ds}
B. inputs = [training_ds.as_named_input('training_ds')]
C. script_params = {"--training_ds":training_ds}
D. source_directory = training_ds

Correct Answer: B
Section:
Explanation:
Example:
Get the training dataset diabetes_ds = ws.datasets.get("Diabetes Dataset") # Create an estimator that uses the remote compute hyper_estimator = SKLearn(source_directory=experiment_folder, inputs=[diabetes_ds.as_named_input('diabetes')], # Pass the dataset as an input compute_target = cpu_cluster, conda_packages=['pandas','ipykernel','matplotlib'], pip_packages=['azureml-sdk','argparse','pyarrow'], entry_script='diabetes_training.py')
Reference: https://notebooks.azure.com/GraemeMalcolm/projects/azureml-primers/html/04%20-%20Optimizing%20Model%20Training.ipynb

QUESTION 39
You register a file dataset named csv_folder that references a folder. The folder includes multiple comma-separated values (CSV) files in an Azure storage blob container.
You plan to use the following code to run a script that loads data from the file dataset. You create and instantiate the following variables:

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

You have the following code:
```
from azureml.train.estimator import Estimator
file_dataset = ws.datasets.get('csv_folder')
estimator = Estimator(source_directory=script_folder,

compute_target = remote_cluster,
entry_script ='script.py')
run = experiment.submit(config=estimator)
run.wait_for_completion(show_output=True)
```
You need to pass the dataset to ensure that the script can read the files it references.
Which code segment should you insert to replace the code comment?

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

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

D. script_params={'--training_files': file_dataset},

Correct Answer: B
Section:
Explanation:
Example:
from azureml.train.estimator import Estimator
script_params = {
# to mount files referenced by mnist dataset
'--data-folder': mnist_file_dataset.as_named_input('mnist_opendataset').as_mount(),
'--regularization': 0.5
}
est = Estimator(source_directory=script_folder,
script_params=script_params,
compute_target=compute_target,
environment_definition=env,
entry_script='train.py')
Reference:
https://docs.microsoft.com/en-us/azure/machine-learning/tutorial-train-models-with-aml

QUESTION 40
You are creating a new Azure Machine Learning pipeline using the designer.
The pipeline must train a model using data in a comma-separated values (CSV) file that is published on a website. You have not created a dataset for this file.
You need to ingest the data from the CSV file into the designer pipeline using the minimal administrative effort.
Which module should you add to the pipeline in Designer?

A. Convert to CSV

B. Enter Data Manually

C. Import Data

D. Dataset

Correct Answer: D
Section:
Explanation:

QUESTION 41
You define a datastore named ml-data for an Azure Storage blob container. In the container, you have a folder named train that contains a file named data.csv. You plan to use the file to train a model by using the Azure Machine Learning
SDK.
You plan to train the model by using the Azure Machine Learning SDK to run an experiment on local compute.
You define a DataReference object by running the following code:

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

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

A.

```
import os
import argparse
import pandas as pd

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

B.

```
import os
import argparse
import pandas as pd

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

C.

```
import pandas as pd

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

D.

```
import os
import argparse
import pandas as pd

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

E.

```
import os
import argparse
import pandas as pd

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

Correct Answer: E
Section:
Explanation:
Example:
data_folder = args.data_folder # Load Train and Test data train_data = pd.read_csv(os.path.join(data_folder, 'data.csv'))
Reference:
https://www.element61.be/en/resource/azure-machine-learning-services-complete-toolbox-ai

QUESTION 42
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 an Azure Machine Learning service datastore in a workspace. The datastore contains the following files:
/data/2018/Q1.csv
/data/2018/Q2.csv
/data/2018/Q3.csv
/data/2018/Q4.csv
/data/2019/Q1.csv
All files store data in the following format:
id,f1,f2,I
1,1,2,0
2,1,1,1
3,2,1,0
4,2,2,1
You run the following code:

```
data_store = Datastore.register_azure_blob_container(workspace=ws,
 datastore_name= 'data_store',
 container_name= 'quarterly_data',
 account_name='companydata',
 account_key='NRPxk8duxbM3…'
 create_if_not_exists=False)
```

You need to create a dataset named training_data and load the data from all files into a single data frame by using the following code:

```
data_frame = training_data.to_pandas_dataframe()
```

Solution: Run the following code:

```
from azureml.core import Dataset
paths = (data_store, 'data/*/*.csv')
training_data = Dataset.Tabular.from_delimited_files(paths)
```

Does the solution meet the goal?

A. Yes

B. No

Correct Answer: B
Section:
Explanation:
Define paths with two file paths instead.
Use Dataset.Tabular_from_delimeted as the data isn't cleansed.
Reference:
https://docs.microsoft.com/en-us/azure/machine-learning/how-to-create-register-datasets

QUESTION 43
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 an Azure Machine Learning service datastore in a workspace. The datastore contains the following files:
/data/2018/Q1.csv
/data/2018/Q2.csv
/data/2018/Q3.csv
/data/2018/Q4.csv
/data/2019/Q1.csv
All files store data in the following format:
id,f1,f2,I
1,1,2,0
2,1,1,1
3,2,1,0
4,2,2,1
You run the following code:

```
data_store = Datastore.register_azure_blob_container(workspace=ws,
 datastore_name= 'data_store',
 container_name= 'quarterly_data',
 account_name='companydata',
 account_key='NRPxk8duxbM3…'
 create_if_not_exists=False)
```

You need to create a dataset named training_data and load the data from all files into a single data frame by using the following code:

```
data_frame = training_data.to_pandas_dataframe()
```

Solution: Run the following code:

```
from azureml.core import Dataset
paths = [(data_store, 'data/2018/*.csv'), (data_store, 'data/2019/*.csv')]
training_data = Dataset.File.from_files(paths)
```

Does the solution meet the goal?

A. Yes
B. No

Correct Answer: B
Section:
Explanation:
Use two file paths.
Use Dataset.Tabular_from_delimeted, instead of Dataset.File.from_files as the data isn't cleansed.

Note:
A FileDataset references single or multiple files in your datastores or public URLs. If your data is already cleansed, and ready to use in training experiments, you can download or mount the files to your compute as a FileDataset object.
A TabularDataset represents data in a tabular format by parsing the provided file or list of files. This provides you with the ability to materialize the data into a pandas or Spark DataFrame so you can work with familiar data preparation and training libraries without having to leave your notebook. You can create a TabularDataset object from .csv, .tsv, .parquet, .jsonl files, and from SQL query results.
Reference:
https://docs.microsoft.com/en-us/azure/machine-learning/how-to-create-register-datasets

QUESTION 44
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 an Azure Machine Learning service datastore in a workspace. The datastore contains the following files:
/data/2018/Q1.csv
/data/2018/Q2.csv
/data/2018/Q3.csv
/data/2018/Q4.csv
/data/2019/Q1.csv
All files store data in the following format:
id,f1,f2,I
1,1,2,0
2,1,1,1
3,2,1,0
4,2,2,1
You run the following code:

```
data_store = Datastore.register_azure_blob_container(workspace=ws,
 datastore_name= 'data_store',
 container_name= 'quarterly_data',
 account_name='companydata',
 account_key='NRPxk8duxbM3…'
 create_if_not_exists=False)
```

You need to create a dataset named training_data and load the data from all files into a single data frame by using the following code:

```
data_frame = training_data.to_pandas_dataframe()
```

Solution: Run the following code:

```
from azureml.core import Dataset
paths = [(data_store, 'data/2018/*.csv'),(data_store, 'data/2019/*.csv')]
training_data = Dataset.Tabular.from_delimited_files(paths)
```

Does the solution meet the goal?

A. Yes
B. No

Correct Answer: A
Section:
Explanation:
Use two file paths.
Use Dataset.Tabular_from_delimeted as the data isn't cleansed.
Note:
A TabularDataset represents data in a tabular format by parsing the provided file or list of files. This provides you with the ability to materialize the data into a pandas or Spark DataFrame so you can work with familiar data

preparation and training libraries without having to leave your notebook. You can create a TabularDataset object from .csv, .tsv, .parquet, .jsonl files, and from SQL query results.
Reference:
https://docs.microsoft.com/en-us/azure/machine-learning/how-to-create-register-datasets

QUESTION 45
You plan to use the Hyperdrive feature of Azure Machine Learning to determine the optimal hyperparameter values when training a model.
You must use Hyperdrive to try combinations of the following hyperparameter values:
learning_rate: any value between 0.001 and 0.1 batch_size: 16, 32, or 64
You need to configure the search space for the Hyperdrive experiment.
Which two parameter expressions should you use? Each correct answer presents part of the solution.
NOTE: Each correct selection is worth one point.

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

Correct Answer: B, D
Section:
Explanation:
B: Continuous hyperparameters are specified as a distribution over a continuous range of values. Supported distributions include: uniform(low, high) - Returns a value uniformly distributed between low and high
D: Discrete hyperparameters are specified as a choice among discrete values. choice can be:
one or more comma-separated values a range object any arbitrary list object
Reference:
https://docs.microsoft.com/en-us/azure/machine-learning/how-to-tune-hyperparameters

QUESTION 46
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_clasification	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')
best_iter = automl_ex.archived_time.find('11/11/2019 11:00:00 AM')
```

B.

```
from azureml.core import Workspace
from azureml.train.automl.run import AutoMLRun
automl_ex = ws.experiments.get('auto_ml_classification')
automl_run = AutoMLRun(automl_ex, 'AutoML_1234567890-123')
best_iter = automl_run.current_run
```

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
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]
```

E.

```
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 = automl_ex.get_runs('AutoML_1234567890-123')
```

Correct Answer: D
Section:
Explanation:
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

QUESTION 47
You have a comma-separated values (CSV) file containing data from which you want to train a classification model.
You are using the Automated Machine Learning interface in Azure Machine Learning studio to train the classification model. You set the task type to Classification.
You need to ensure that the Automated Machine Learning process evaluates only linear models.
What should you do?

A. Add all algorithms other than linear ones to the blocked algorithms list.
B. Set the Exit criterion option to a metric score threshold.

C. Clear the option to perform automatic featurization.

D. Clear the option to enable deep learning.

E. Set the task type to Regression.

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

QUESTION 48
Note: This question is part of a series of questions that present the same scenario. Each question in the series contains a unique solution that might meet the stated goals. Some question sets might have more than one correct solution, while others might not have a correct solution.
After you answer a question in this section, you will NOT be able to return to it. As a result, these questions will not appear in the review screen.
You plan to use a Python script to run an Azure Machine Learning experiment. The script creates a reference to the experiment run context, loads data from a file, identifies the set of unique values for the label column, and completes the experiment run:

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

The experiment must record the unique labels in the data as metrics for the run that can be reviewed later.
You must add code to the script to record the unique label values as run metrics at the point indicated by the comment.
Solution: Replace the comment with the following code:
run.upload_file('outputs/labels.csv', './data.csv')
Does the solution meet the goal?

A. Yes

B. No

Correct Answer: B
Section:
Explanation:
label_vals has the unique labels (from the statement label_vals = data['label'].unique()), and it has to be logged.
Note:
Instead use the run_log function to log the contents in label_vals:
for label_val in label_vals: run.log('Label Values', label_val)
Reference:
https://www.element61.be/en/resource/azure-machine-learning-services-complete-toolbox-ai

QUESTION 49
Note: This question is part of a series of questions that present the same scenario. Each question in the series contains a unique solution that might meet the stated goals. Some question sets might have more than one correct solution, while others might not have a correct solution.
After you answer a question in this section, you will NOT be able to return to it. As a result, these questions will not appear in the review screen.
You plan to use a Python script to run an Azure Machine Learning experiment. The script creates a reference to the experiment run context, loads data from a file, identifies the set of unique values for the label column, and completes the experiment run:

```
from azureml.core import Run
```

import pandas as pd
run = Run.get_context()
data = pd.read_csv('data.csv')
label_vals = data['label'].unique()
# Add code to record metrics here
run.complete()
The experiment must record the unique labels in the data as metrics for the run that can be reviewed later.
You must add code to the script to record the unique label values as run metrics at the point indicated by the comment.
Solution: Replace the comment with the following code:
run.log_table('Label Values', label_vals)
Does the solution meet the goal?

A. Yes
B. No

Correct Answer: B
Section:
Explanation:
Instead use the run_log function to log the contents in label_vals:
for label_val in label_vals: run.log('Label Values', label_val)
Reference:
https://www.element61.be/en/resource/azure-machine-learning-services-complete-toolbox-ai

QUESTION 50
Note: This question-is part of a series of questions that present the same scenario. Each question-in the series contains a unique solution that might meet the stated goals. Some question-sets might have more than one correct solution, while others might not have a correct solution.
After you answer a question-in this section, you will NOT be able to return to it. As a result, these questions will not appear in the review screen.
You plan to use a Python script to run an Azure Machine Learning experiment. The script creates a reference to the experiment run context, loads data from a file, identifies the set of unique values for the label column, and completes the experiment run:
from azureml.core import Run
import pandas as pd run = Run.get_context() data = pd.read_csv('data.csv') label_vals = data['label'].unique() # Add code to record metrics here run.complete()
The experiment must record the unique labels in the data as metrics for the run that can be reviewed later.
You must add code to the script to record the unique label values as run metrics at the point indicated by the comment.
Solution: Replace the comment with the following code:
for label_val in label_vals:
run.log('Label Values', label_val)
Does the solution meet the goal?

A. Yes
B. No

Correct Answer: A
Section:
Explanation:
The run_log function is used to log the contents in label_vals:
for label_val in label_vals: run.log('Label Values', label_val)
Reference: https://www.element61.be/en/resource/azure-machine-learning-services-complete-toolbox-ai

QUESTION 51

You are solving a classification task.
You must evaluate your model on a limited data sample by using k-fold cross-validation. You start by configuring a k parameter as the number of splits.
You need to configure the k parameter for the cross-validation.
Which value should you use?

A. k=0.5
B. k=0.01
C. k=5
D. k=1

Correct Answer: C
Section:
Explanation:
Leave One Out (LOO) cross-validation
Setting K = n (the number of observations) yields n-fold and is called leave-one out cross-validation (LOO), a special case of the K-fold approach.
LOO CV is sometimes useful but typically doesn't shake up the data enough. The estimates from each fold are highly correlated and hence their average can have high variance. This is why the usual choice is K=5 or 10. It provides a good compromise for the bias-variance tradeoff.

QUESTION 52
Note: This question is part of a series of questions that present the same scenario. Each question in the series contains a unique solution that might meet the stated goals. Some question sets might have more than one correct solution, while others might not have a correct solution.
After you answer a question in this section, you will NOT be able to return to it. As a result, these questions will not appear in the review screen.
You create a model to forecast weather conditions based on historical data.
You need to create a pipeline that runs a processing script to load data from a datastore and pass the processed data to a machine learning model training script.
Solution: Run the following code:

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

Does the solution meet the goal?

A. Yes
B. No

Correct Answer: B
Section:
Explanation:
The two steps are present: process_step and train_step
The training data input is not setup correctly.

Note:
Data used in pipeline can be produced by one step and consumed in another step by providing a PipelineData object as an output of one step and an input of one or more subsequent steps.
PipelineData objects are also used when constructing Pipelines to describe step dependencies. To specify that a step requires the output of another step as input, use a PipelineData object in the constructor of both steps.
For example, the pipeline train step depends on the process_step_output output of the pipeline process step:

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

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

QUESTION 53
HOTSPOT
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

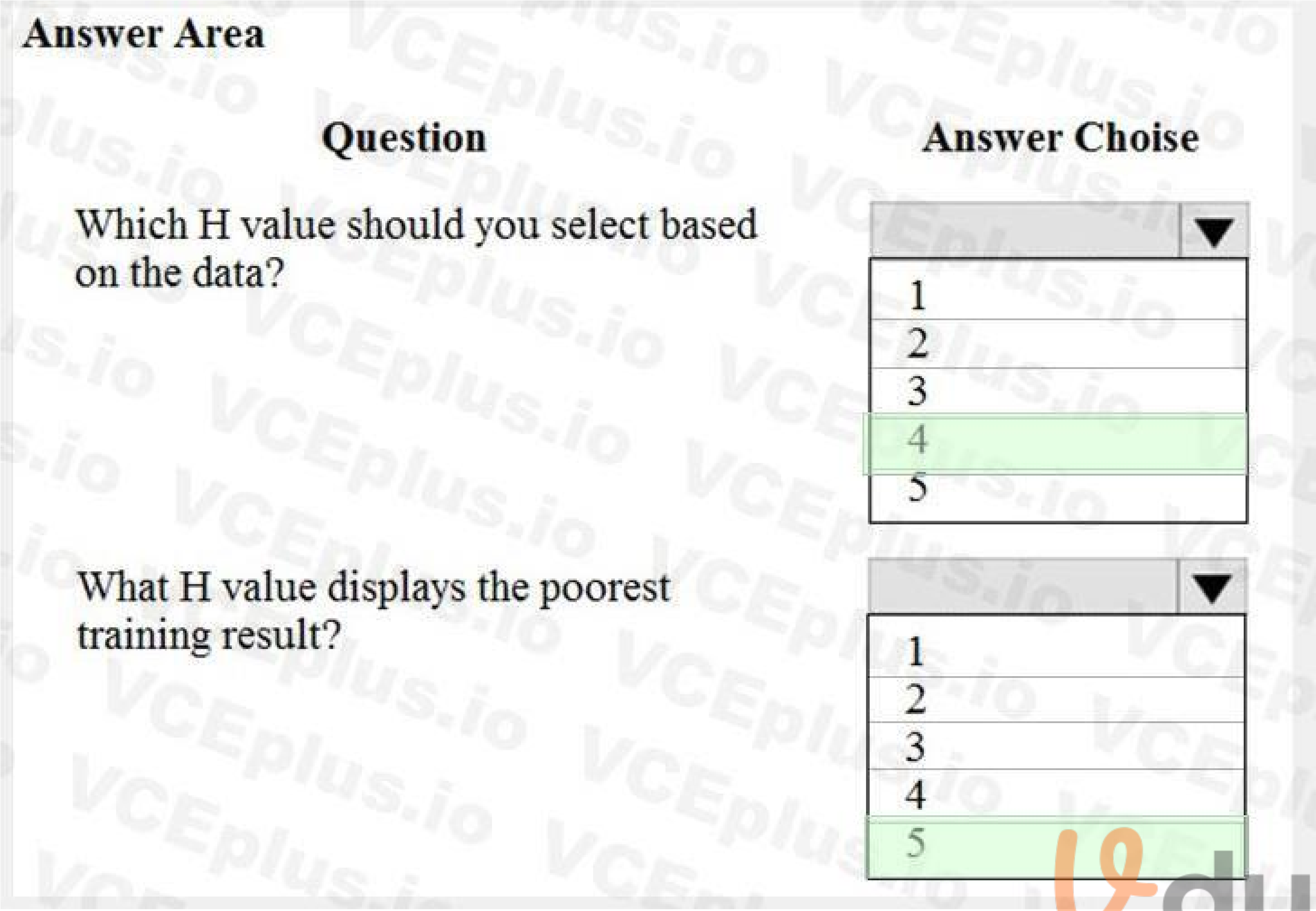
Use the drop-down menus to select the answer choice that answers each question based on the information presented in the graphic.

Hot Area:

Answer Area

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 Area:



Section:
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

QUESTION 54
DRAG DROP
You create machine learning models by using Azure Machine Learning.
You plan to train and score models by using a variety of compute contexts. You also plan to create a new compute resource in Azure Machine Learning studio.
You need to select the appropriate compute types.
Which compute types should you select? To answer, drag the appropriate compute types to the correct requirements. Each compute type may be used once, more than once, or not at all. You may need to drag the split bar between panes or scroll to view content.
NOTE: Each correct selection is worth one point.

Select and Place:

Compute types

- Attached compute
- Inference cluster
- Training cluster

Answer Area

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

Correct Answer:

Compute types

- Attached compute
- Inference cluster
- Training cluster

Answer Area

Requirement	Compute type
Train models by using the Azure Machine Learning designer.	Attached compute
Score new data through a trained model published as a real-time web service.	Inference cluster
Train models by using an Azure Databricks cluster.	Training cluster
Deploy models by using the Azure Machine Learning designer.	Attached compute

Section:

Explanation:

Box 1: Attached compute

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

Box 2: Inference cluster

Box 3: Training cluster

Box 4: Attached compute

QUESTION 55
DRAG DROP
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.

Select and Place:

Modules

- Export Data
- Tune Model Hyperparameters
- Cross Validate Model
- Evaluate Model
- Score Model
- Train Model

Answer Area

Correct Answer:

Modules

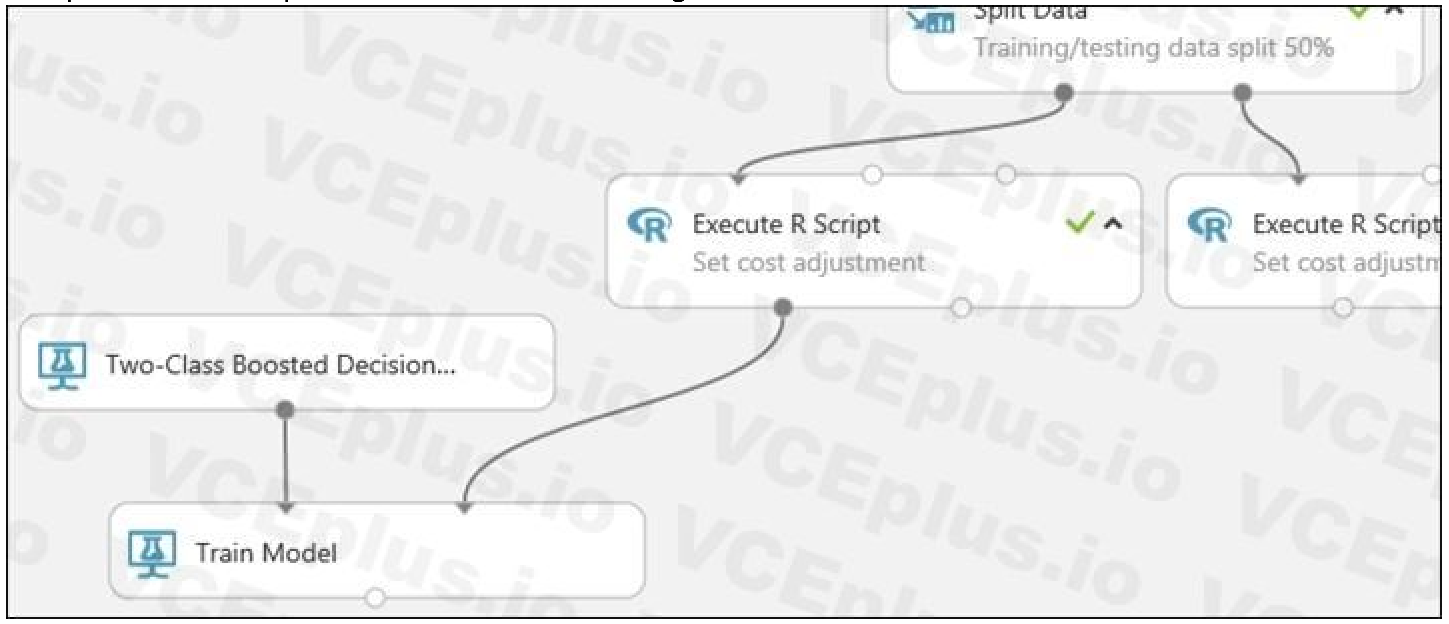
- Export Data
- Tune Model Hyperparameters
- Cross Validate Model

Answer Area

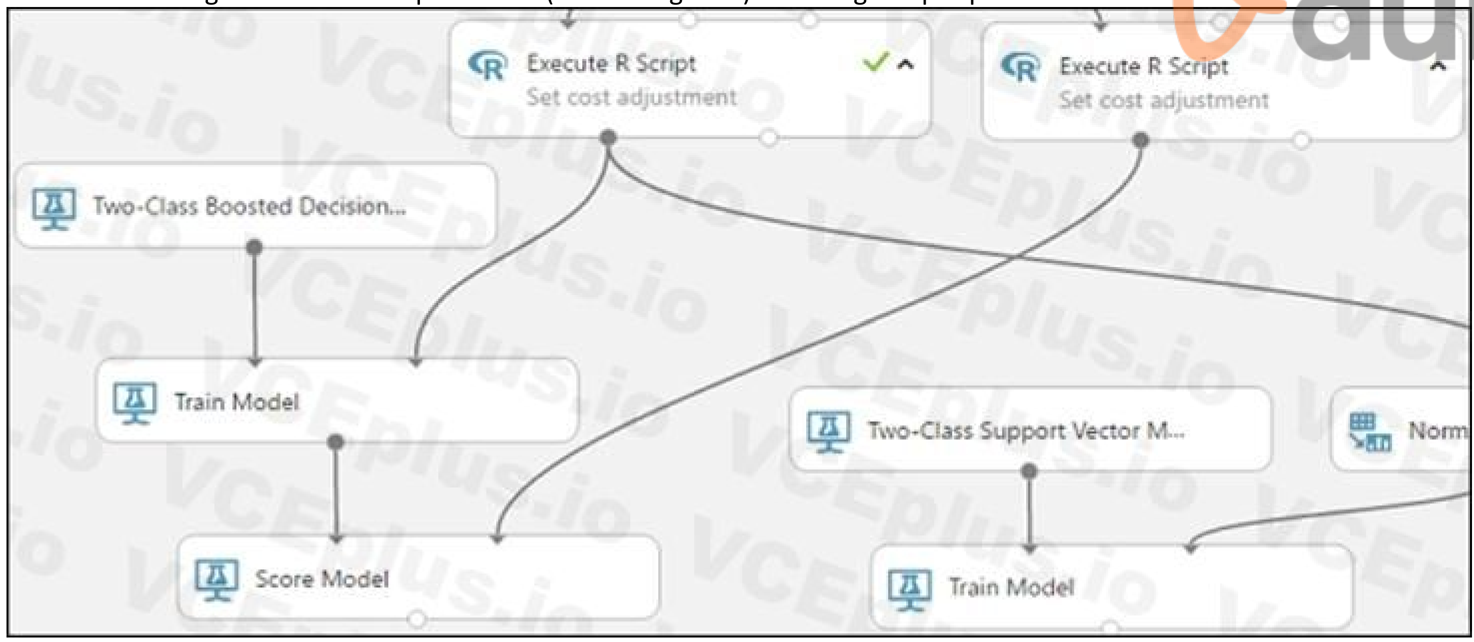
- Train Model
- Score Model
- Evaluate Model

Section:
Explanation:
Step 1: Train Model
Two-Class Boosted Decision Tree
First, set up the boosted decision tree model.
1. Find the Two-Class Boosted Decision Tree module in the module palette and drag it onto the canvas.

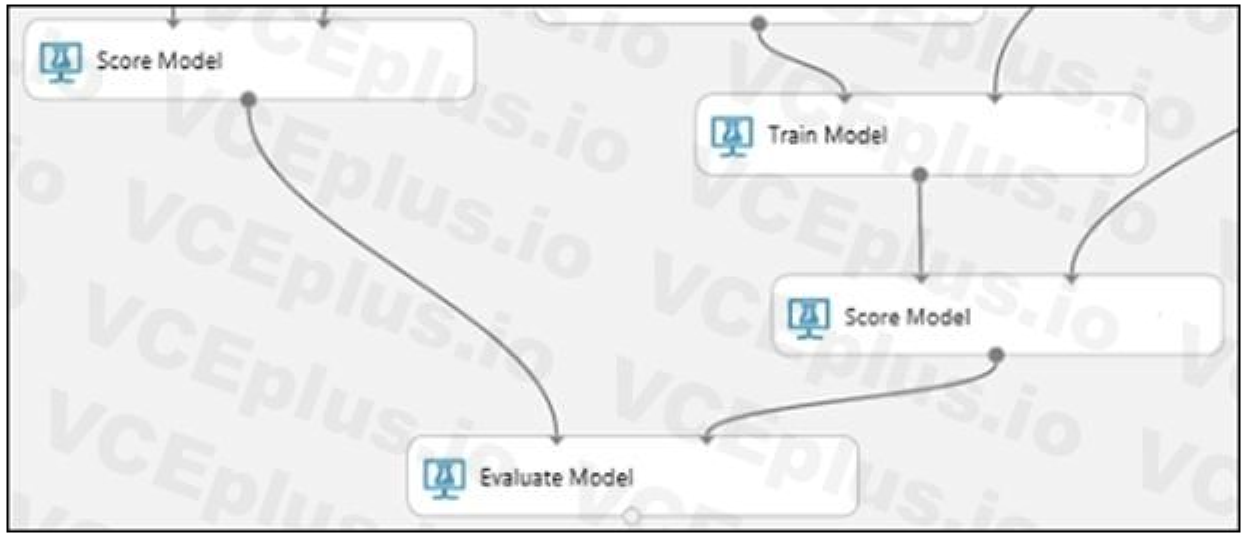
2. Find the Train Model module, drag it onto the canvas, and then connect the output of the Two-Class Boosted Decision Tree module to the left input port of the Train Model module.
The Two-Class Boosted Decision Tree module initializes the generic model, and Train Model uses training data to train the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



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



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



QUESTION 56

HOTSPOT

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.

Hot Area:

	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 Area:

	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.	○	○

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

QUESTION 57
HOTSPOT
You collect data from a nearby weather station. You have a pandas dataframe named weather_df that includes the following data:

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

The data is collected every 12 hours: noon and midnight.
You plan to use automated machine learning to create a time-series model that predicts temperature over the next seven days. For the initial round of training, you want to train a maximum of 50 different models.
You must use the Azure Machine Learning SDK to run an automated machine learning experiment to train these models.
You need to configure the automated machine learning run.
How should you complete the AutoMLConfig definition? To answer, select the appropriate options in the answer area.
NOTE: Each correct selection is worth one point.

Hot Area:

```
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=          ,
```

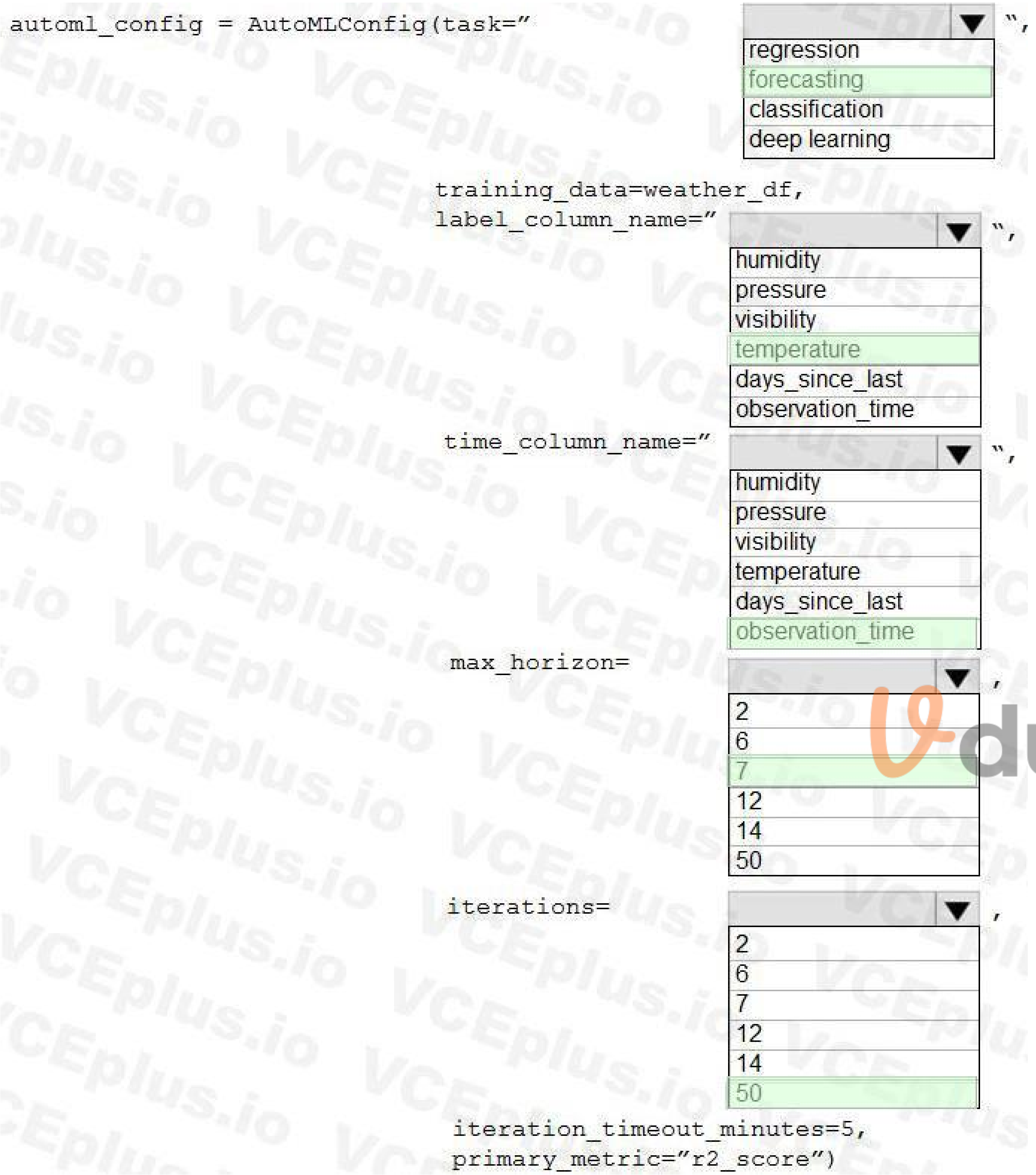
2
6
7
12
14
50

```
    iterations=          ,
```

2
6
7
12
14
50

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

Answer Area:



Section:

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

QUESTION 58
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

Correct Answer: B
Section:
Explanation:
train_step is missing.
Reference:
https://docs.microsoft.com/en-us/python/api/azureml-pipeline-core/azureml.pipeline.core.pipelinedata?view=azure-ml-py

QUESTION 59
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
Section:
Explanation:
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 60
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

Correct Answer: A
Section:
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

QUESTION 61
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

Correct Answer: B
Section:
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

QUESTION 62
Note: This question is part of a series of questions that present the same scenario. Each question in the series contains a unique solution that might meet the stated goals. Some question sets might have more than one correct solution, while others might not have a correct solution.
After you answer a question in this section, you will NOT be able to return to it. As a result, these questions will not appear in the review screen.
You have a Python script named train.py in a local folder named scripts. The script trains a regression model by using scikit-learn. The script includes code to load a training data file which is also located in the scripts folder.
You must run the script as an Azure ML experiment on a compute cluster named aml-compute.
You need to configure the run to ensure that the environment includes the required packages for model training. You have instantiated a variable named aml-compute that references the target compute cluster.
Solution: Run the following code:

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

Does the solution meet the goal?

A. Yes
B. No

Correct Answer: B
Section:
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

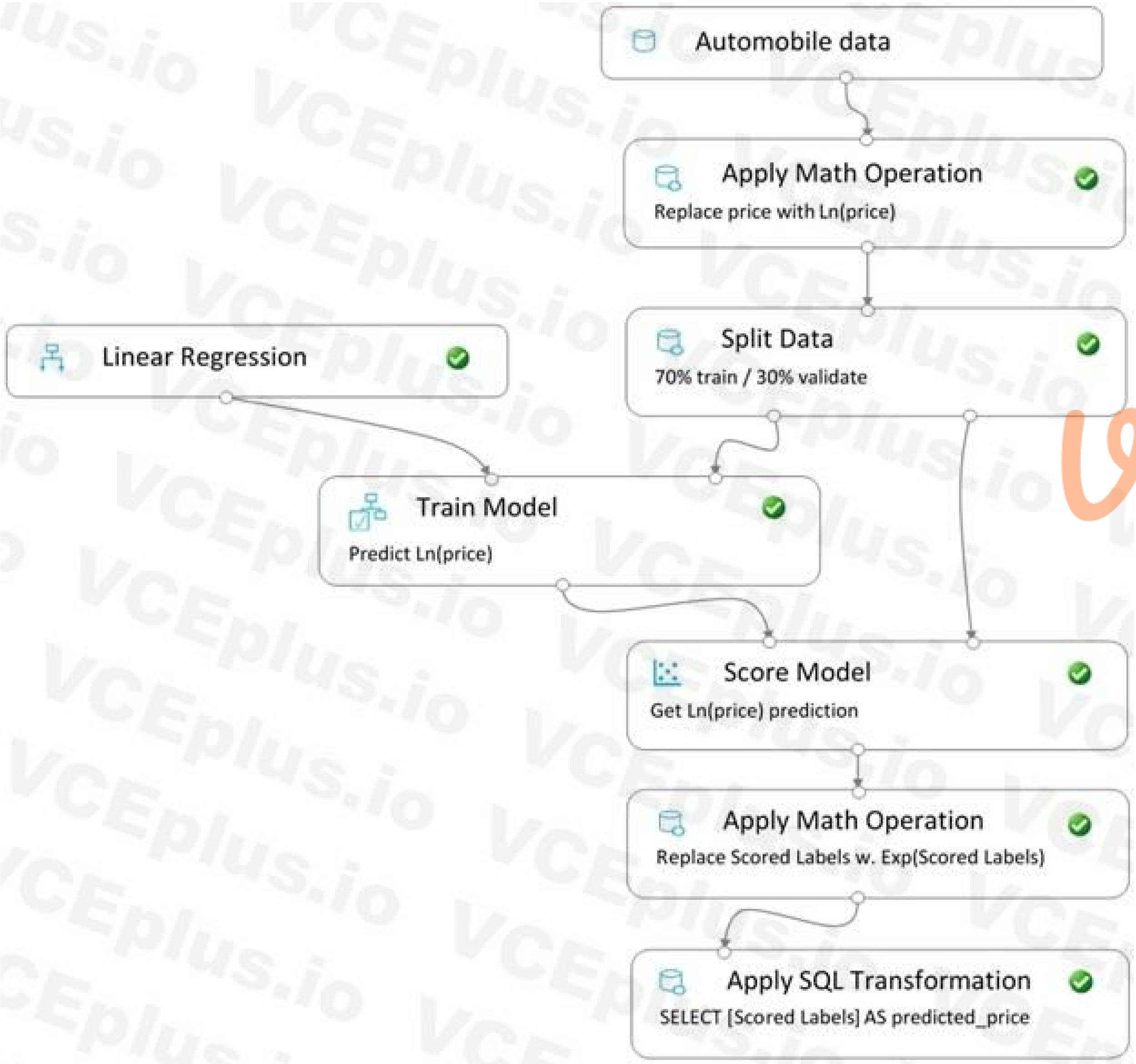
QUESTION 63
You create a multi-class image classification deep learning model that uses a set of labeled images. You create a script file named train.py that uses the PyTorch 1.3 framework to train the model.
You must run the script by using an estimator. The code must not require any additional Python libraries to be installed in the environment for the estimator. The time required for model training must be minimized.
You need to define the estimator that will be used to run the script.
Which estimator type should you use?

A. TensorFlow
B. PyTorch
C. SKLearn
D. Estimator

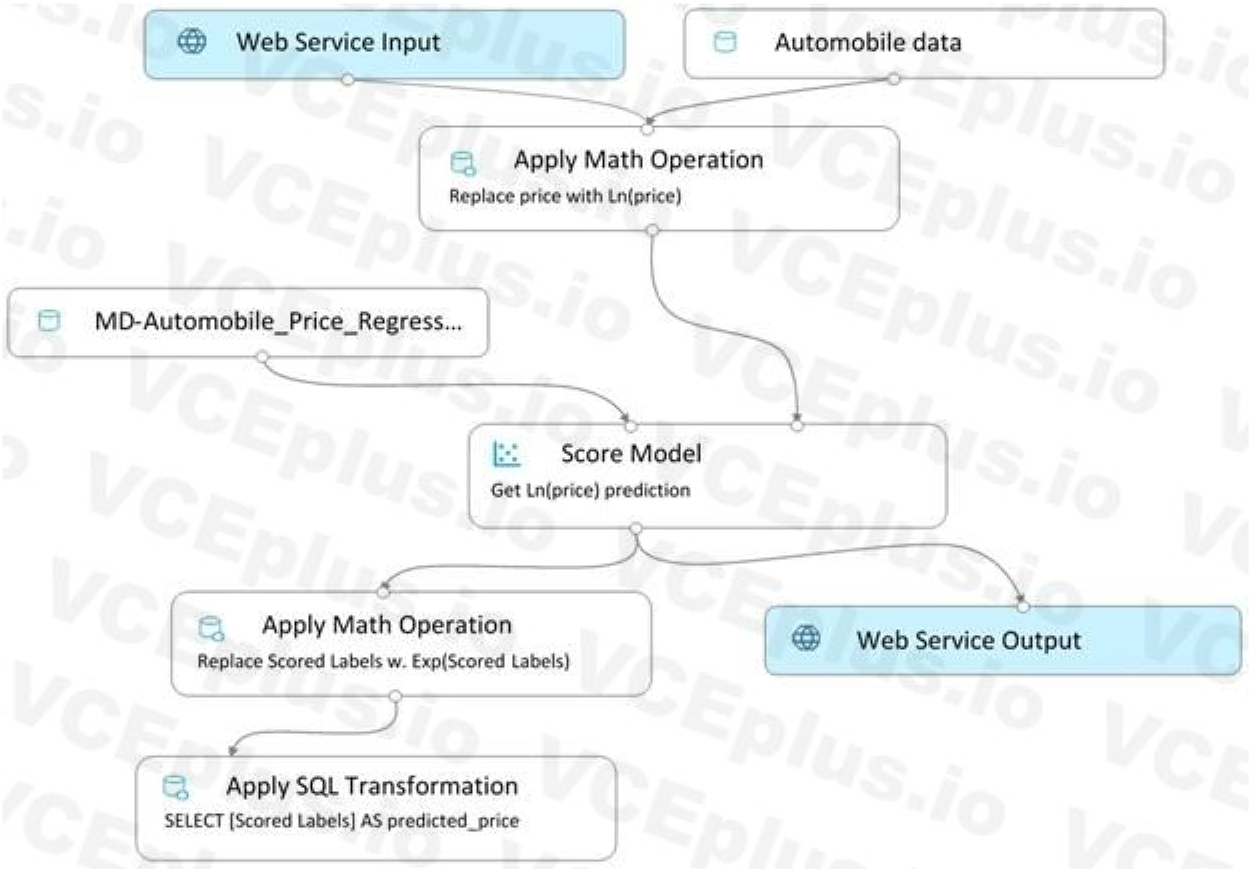
Correct Answer: B

Section:
Explanation:
For PyTorch, TensorFlow and Chainer tasks, Azure Machine Learning provides respective PyTorch, TensorFlow, and Chainer estimators to simplify using these frameworks.
Reference:
https://docs.microsoft.com/en-us/azure/machine-learning/how-to-train-ml-models

QUESTION 64
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: A, C, E
Section:

QUESTION 65
You use the Two-Class Neural Network module in Azure Machine Learning Studio to build a binary classification model. You use the Tune Model Hyperparameters module to tune accuracy for the model.
You need to configure the Tune Model Hyperparameters module.
Which two values should you use? Each correct answer presents part of the solution.
NOTE: Each correct selection is worth one point.

A. Number of hidden nodes
B. Learning Rate
C. The type of the normalizer
D. Number of learning iterations
E. Hidden layer specification

Correct Answer: D, E
Section:
Explanation:
D: For Number of learning iterations, specify the maximum number of times the algorithm should process the training cases.
E: For Hidden layer specification, select the type of network architecture to create.
Between the input and output layers you can insert multiple hidden layers. Most predictive tasks can be accomplished easily with only one or a few hidden layers.
Reference: https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/two-class-neural-network

QUESTION 66
You create a binary classification model by using Azure Machine Learning Studio.
You must tune hyperparameters by performing a parameter sweep of the model. The parameter sweep must meet the following requirements:
iterate all possible combinations of hyperparameters
minimize computing resources required to perform the sweep
You need to perform a parameter sweep of the model.
Which parameter sweep mode should you use?

A. Random sweep
B. Sweep clustering
C. Entire grid
D. Random grid

Correct Answer: D
Section:
Explanation:
Maximum number of runs on random grid: This option also controls the number of iterations over a random sampling of parameter values, but the values are not generated randomly from the specified range; instead, a matrix is created of all possible combinations of parameter values and a random sampling is taken over the matrix. This method is more efficient and less prone to regional oversampling or undersampling.
If you are training a model that supports an integrated parameter sweep, you can also set a range of seed values to use and iterate over the random seeds as well. This is optional, but can be useful for avoiding bias introduced by seed selection.
Incorrect Answers:
B: If you are building a clustering model, use Sweep Clustering to automatically determine the optimum number of clusters and other parameters.
C: Entire grid: When you select this option, the module loops over a grid predefined by the system, to try different combinations and identify the best learner. This option is useful for cases where you don't know what the best parameter settings might be and want to try all possible combination of values.
Reference:
https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/tune-model-hyperparameters

QUESTION 67
You are building a recurrent neural network to perform a binary classification.
You review the training loss, validation loss, training accuracy, and validation accuracy for each training epoch.
You need to analyze model performance.
You need to identify whether the classification model is overfitted.
Which of the following is correct?

A. The training loss stays constant and the validation loss stays on a constant value and close to the training loss value when training the model.
B. The training loss decreases while the validation loss increases when training the model.
C. The training loss stays constant and the validation loss decreases when training the model.
D. The training loss increases while the validation loss decreases when training the model.

Correct Answer: B
Section:

Explanation:
An overfit model is one where performance on the train set is good and continues to improve, whereas performance on the validation set improves to a point and then begins to degrade.
Reference:
https://machinelearningmastery.com/diagnose-overfitting-underfitting-lstm-models/

QUESTION 68
You use the Azure Machine Learning Python SDK to define a pipeline to train a model.
The data used to train the model is read from a folder in a datastore.
You need to ensure the pipeline runs automatically whenever the data in the folder changes.
What should you do?

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

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

QUESTION 69
You plan to run a Python script as an Azure Machine Learning experiment.
The script must read files from a hierarchy of folders. The files will be passed to the script as a dataset argument.
You must specify an appropriate mode for the dataset argument.
Which two modes can you use? Each correct answer presents a complete solution.
NOTE: Each correct selection is worth one point.

A. to_pandas_dataframe()
B. as_download()
C. as_upload()
D. as_mount()

Correct Answer: B
Section:
Explanation:
Reference: https://docs.microsoft.com/en-us/python/api/azureml-core/azureml.data.filedataset?view=azure-ml-py

QUESTION 70
DRAG DROP
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.

Select and Place:

Actions

- Configure a DataTransferStep() to fetch new image data from public web portal, running on the cpu-compute compute target.
- Configure an EstimatorStep() to run an estimator that runs the bird_classifier_train.py model training script on the gpu_compute compute target.
- Configure a PythonScriptStep() to run both image_fetcher.py and image_resize.py on the cpu-compute compute target.
- Configure an EstimatorStep() to run an estimator that runs the bird_classifier_train.py model training script on the cpu_compute compute target.
- Configure a PythonScriptStep() to run image_fetcher.py on the cpu-compute compute target.
- Configure a PythonScriptStep() to run image_resize.py on the cpu-compute compute target.
- Configure a PythonScriptStep() to run bird_classifier_train.py on the cpu-compute compute target.
- Configure a PythonScriptStep() to run bird_classifier_train.py on the gpu-compute compute target.

Answer Area

Correct Answer:

Actions

- Configure a PythonScriptStep() to run both image_fetcher.py and image_resize.py on the cpu-compute compute target.
- Configure an EstimatorStep() to run an estimator that runs the bird_classifier_train.py model training script on the cpu_compute compute target.
- Configure a PythonScriptStep() to run image_fetcher.py on the cpu-compute compute target.
- Configure a PythonScriptStep() to run bird_classifier_train.py on the cpu-compute compute target.
- Configure a PythonScriptStep() to run bird_classifier_train.py on the gpu-compute compute target.

Answer Area

- Configure a DataTransferStep() to fetch new image data from public web portal, running on the cpu-compute compute target.
- Configure a PythonScriptStep() to run image_resize.py on the cpu-compute compute target.
- Configure an EstimatorStep() to run an estimator that runs the bird_classifier_train.py model training script on the gpu_compute compute target.

Section:

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

QUESTION 71

HOTSPOT

You plan to use Hyperdrive to optimize the hyperparameters selected when training a model. You create the following code to define options for the hyperparameter experiment:

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

param_sampling = GridParameterSampling({
  "max_depth" : pe.choice(6, 7, 8, 9),
  "learning_rate" : pe.choice(0.05, 0.1, 0.15)
  })
hyperdrive_run_config = HyperDriveConfig(
  estimator = estimator,
  hyperparameter_sampling = param_sampling,
  policy = None,
  primary_metric_name = "auc",
  primary_metruc_goal = PrimaryMetricGoal.MAXIMIZE,
  max_total_runs = 50,
  max_concurrent_runs = 4)
```

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

Hot Area:

Answer Area

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

Answer Area:

Answer Area

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

Section:
Explanation:
Box 1: No max_total_runs (50 here)
The maximum total number of runs to create. This is the upper bound; there may be fewer runs when the sample space is smaller than this value.
Box 2: Yes
Policy EarlyTerminationPolicy
The early termination policy to use. If None - the default, no early termination policy will be used.
Box 3: No
Discrete hyperparameters are specified as a choice among discrete values. choice can be:
one or more comma-separated values
a range object
any arbitrary list object
Reference:
https://docs.microsoft.com/en-us/python/api/azureml-train-core/azureml.train.hyperdrive.hyperdriveconfig
https://docs.microsoft.com/en-us/azure/machine-learning/how-to-tune-hyperparameters

QUESTION 72
HOTSPOT
You are using Azure Machine Learning to train machine learning models. You need to compute target on which to remotely run the training script.
You run the following Python code:

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

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:

Answer Area

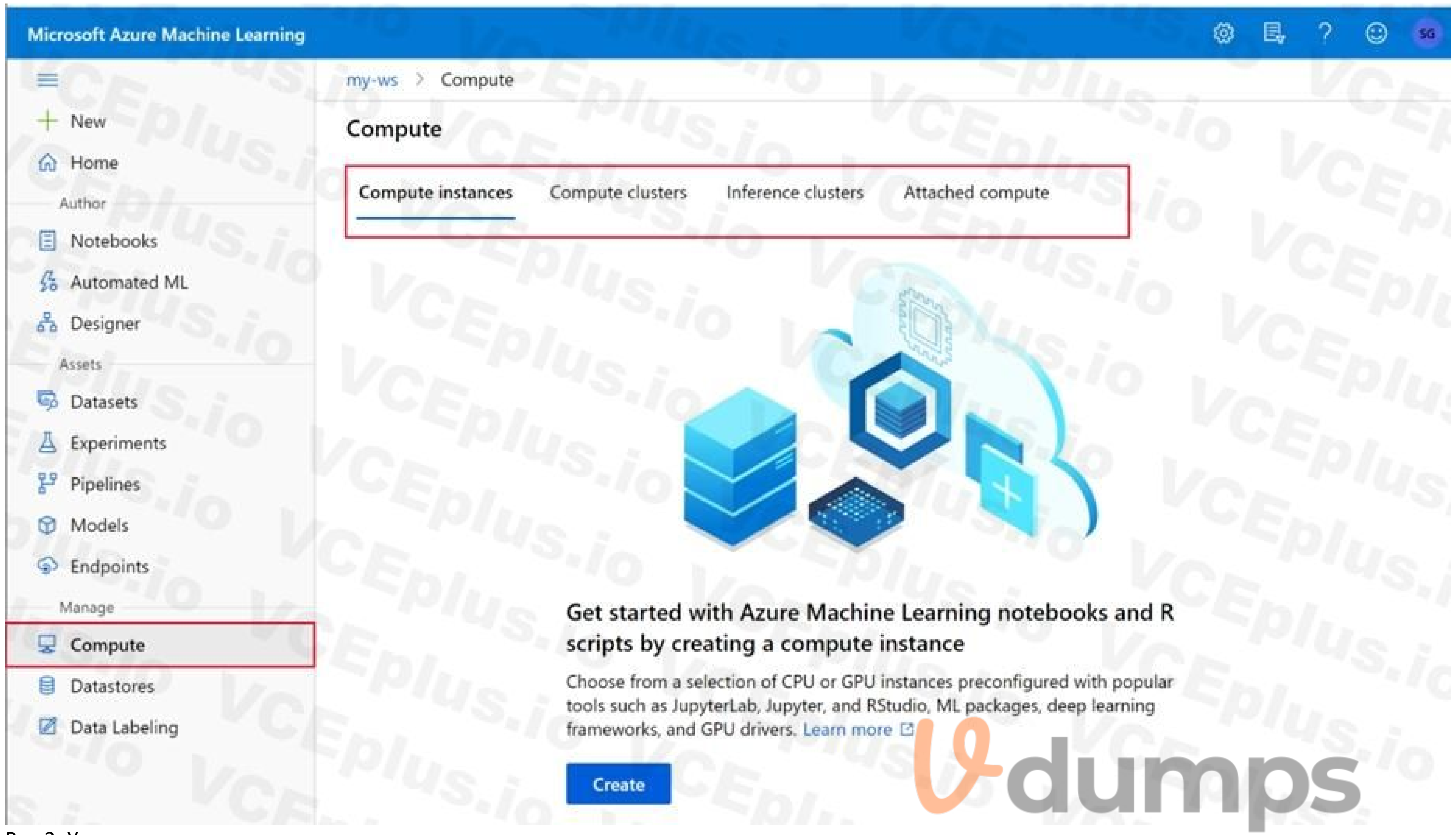
	Yes	No
The compute is created in the same region as the Machine Learning service workspace.	○	○
The compute resource created by the code is displayed as a compute cluster in Azure Machine Learning studio.	○	○
The minimum number of nodes will be zero.	○	○

Answer Area:

Answer Area

	Yes	No
The compute is created in the same region as the Machine Learning service workspace.	○	○
The compute resource created by the code is displayed as a compute cluster in Azure Machine Learning studio.	○	○
The minimum number of nodes will be zero.	○	○

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

QUESTION 73
HOTSPOT
You have an Azure blob container that contains a set of TSV files. The Azure blob container is registered as a datastore for an Azure Machine Learning service workspace. Each TSV file uses the same data schema.
You plan to aggregate data for all of the TSV files together and then register the aggregated data as a dataset in an Azure Machine Learning workspace by using the Azure Machine Learning SDK for Python.
You run the following code.

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

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

Hot Area:

Answer Area

	Yes	No
The myDataset_1 dataset can be converted into a pandas dataframe by using the following method: `using myDataset_1.to_pandas_dataframe()`		
The myDataset_1.to_path() method returns an array of file paths for all of the TSV files in the dataset.		
The myDataset_2 dataset can be converted into a pandas dataframe by using the following method: `myDataset_2.to_pandas_dataframe()`		

Answer Area:

Answer Area

	Yes	No
The myDataset_1 dataset can be converted into a pandas dataframe by using the following method: `using myDataset_1.to_pandas_dataframe()`		
The myDataset_1.to_path() method returns an array of file paths for all of the TSV files in the dataset.		
The myDataset_2 dataset can be converted into a pandas dataframe by using the following method: `myDataset_2.to_pandas_dataframe()`		

Section:

Explanation:

Box 1: No

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

Box 2: Yes

to_path() gets a list of file paths for each file stream defined by the dataset.

Box 3: Yes

TabularDataset.to_pandas_dataframe loads all records from the dataset into a pandas DataFrame.

TabularDataset represents data in a tabular format created by parsing the provided file or list of files.

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

Reference:

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

QUESTION 74

DRAG DROP

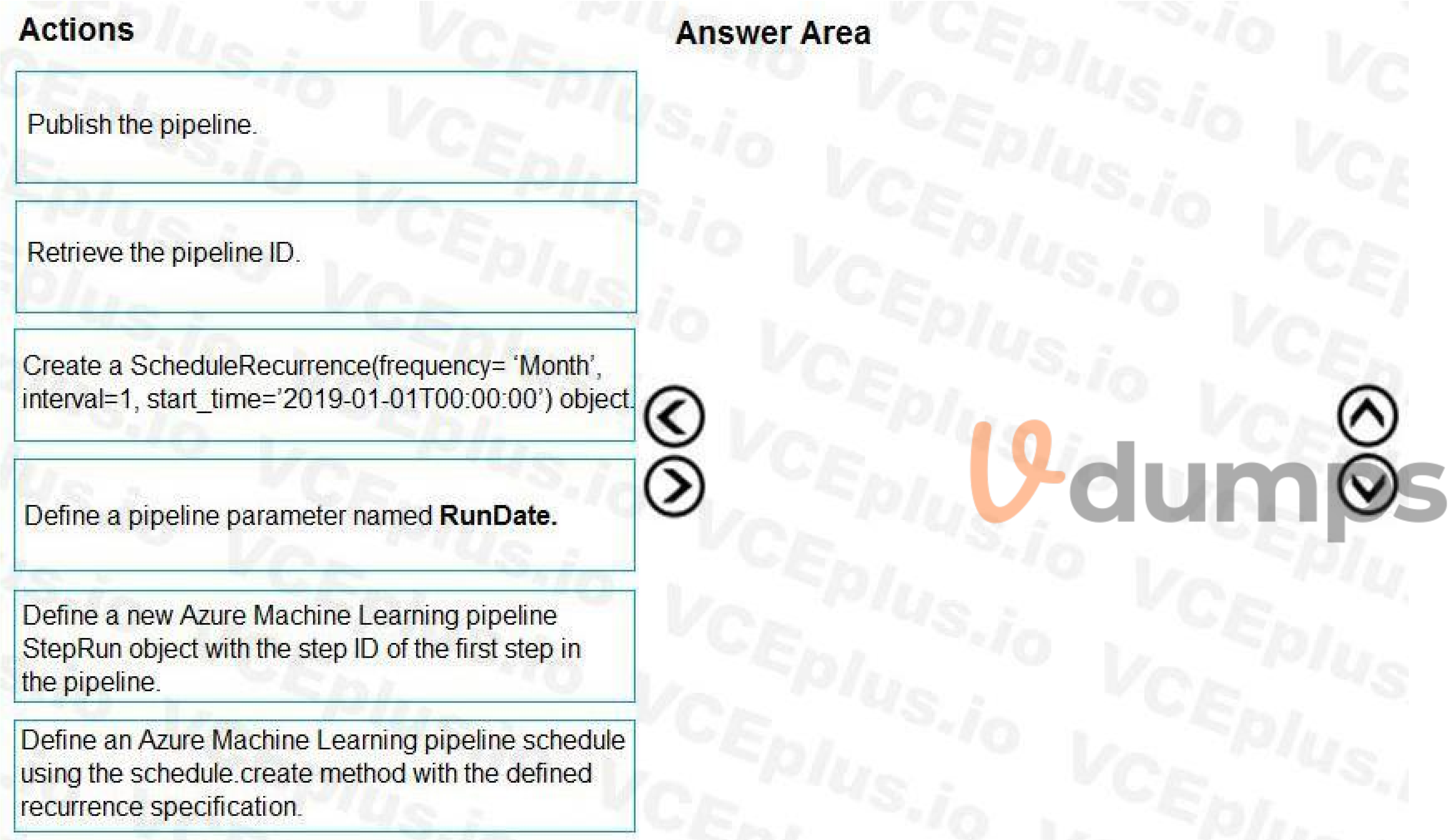
You create a multi-class image classification deep learning model.

The model must be retrained monthly with the new image data fetched from a public web portal. You create an Azure Machine Learning pipeline to fetch new data, standardize the size of images, and retrain the model.

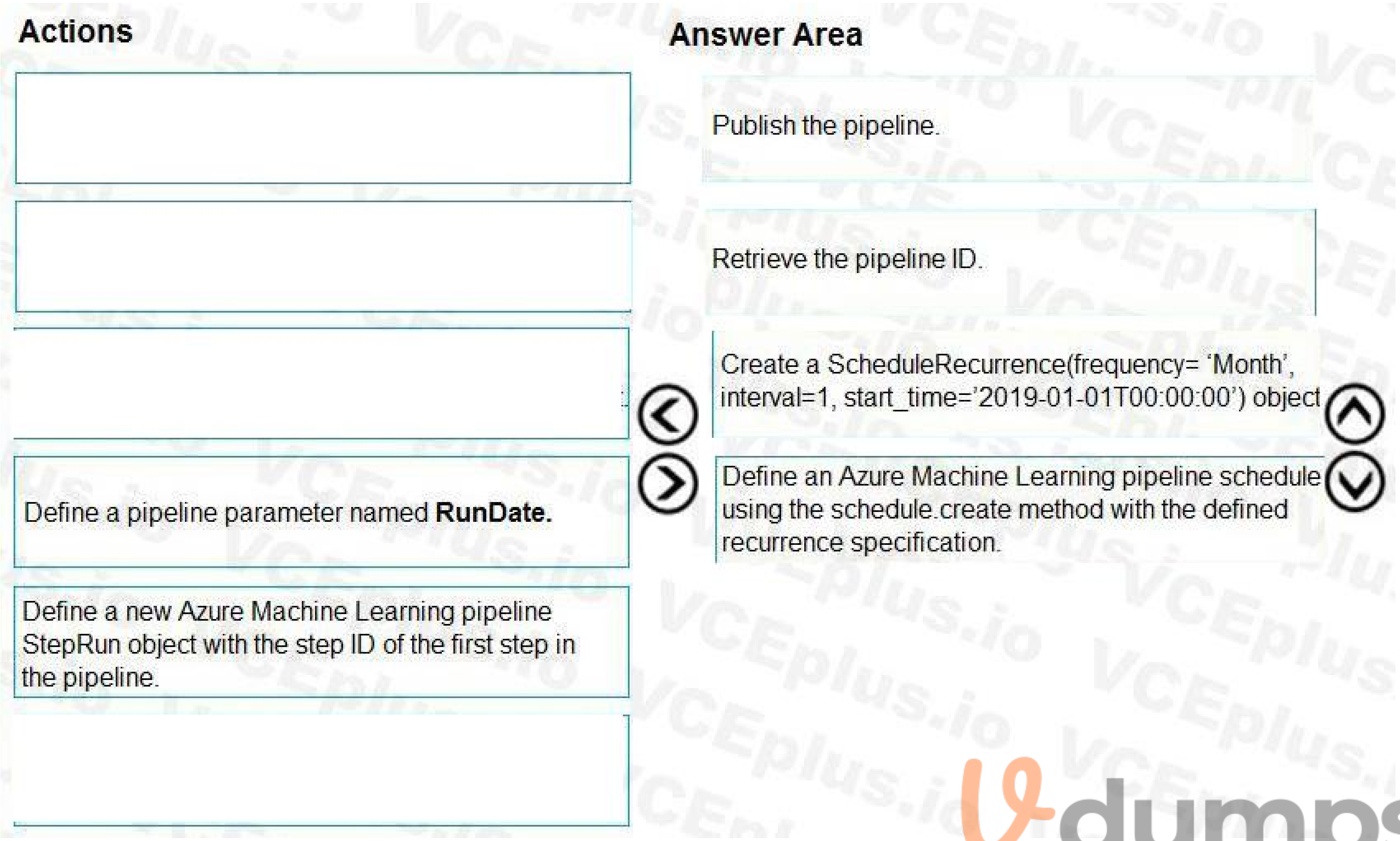
You need to use the Azure Machine Learning SDK to configure the schedule for the pipeline.

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

Select and Place:



Correct Answer:



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

QUESTION 75
HOTSPOT
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.

Hot Area:

Answer Area

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

Answer Area:

Answer Area

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

Section:
Explanation:
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.

QUESTION 76
HOTSPOT
You have a Python data frame named salesData in the following format:

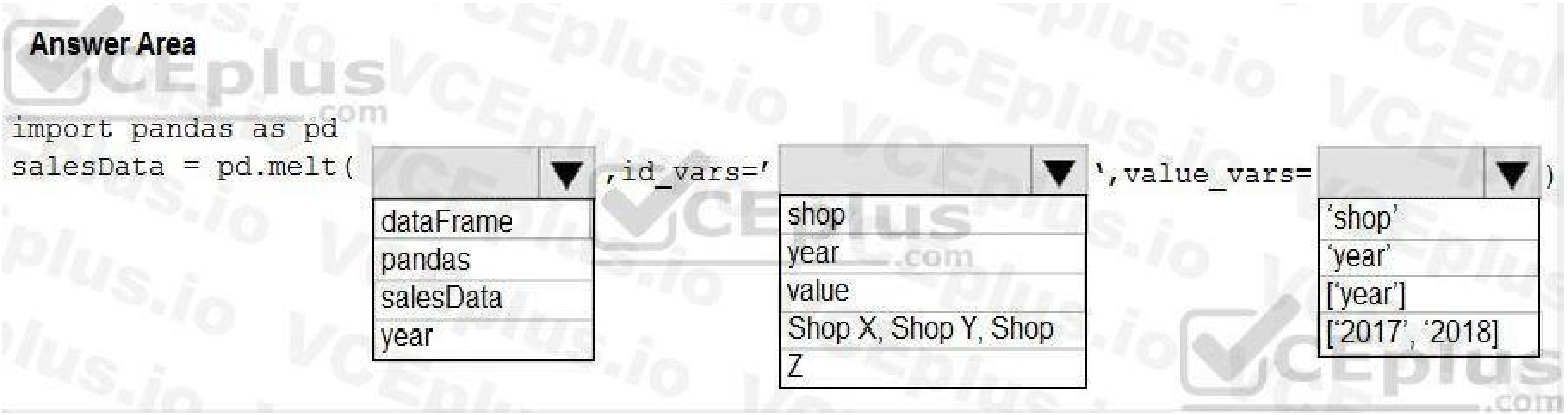
	shop	2017	2018
0	Shop X	34	25
1	Shop Y	65	76
2	Shop Z	48	55

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.

Hot Area:



Answer Area:

Section:
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

QUESTION 77

HOTSPOT

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.

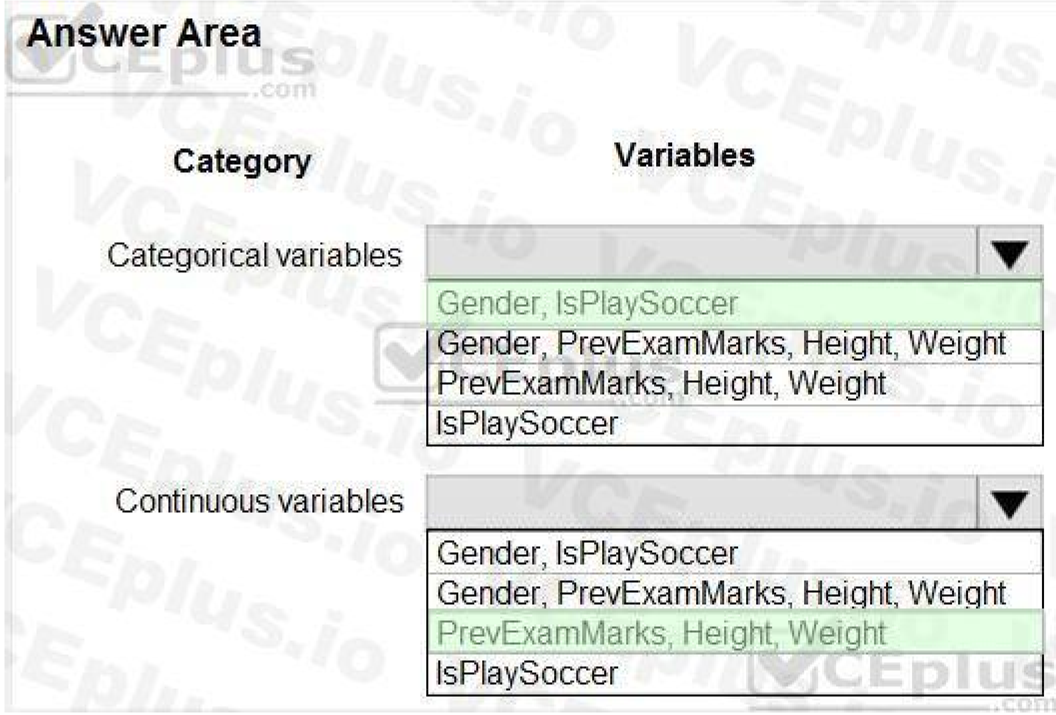
NOTE: Each correct selection is worth one point.

Hot Area:

Answer Area

Category	Variables
Categorical variables	▼
	Gender, IsPlaySoccer
	Gender, PrevExamMarks, Height, Weight
	PrevExamMarks, Height, Weight
	IsPlaySoccer
Continuous variables	▼
	Gender, IsPlaySoccer
	Gender, PrevExamMarks, Height, Weight
	PrevExamMarks, Height, Weight
	IsPlaySoccer

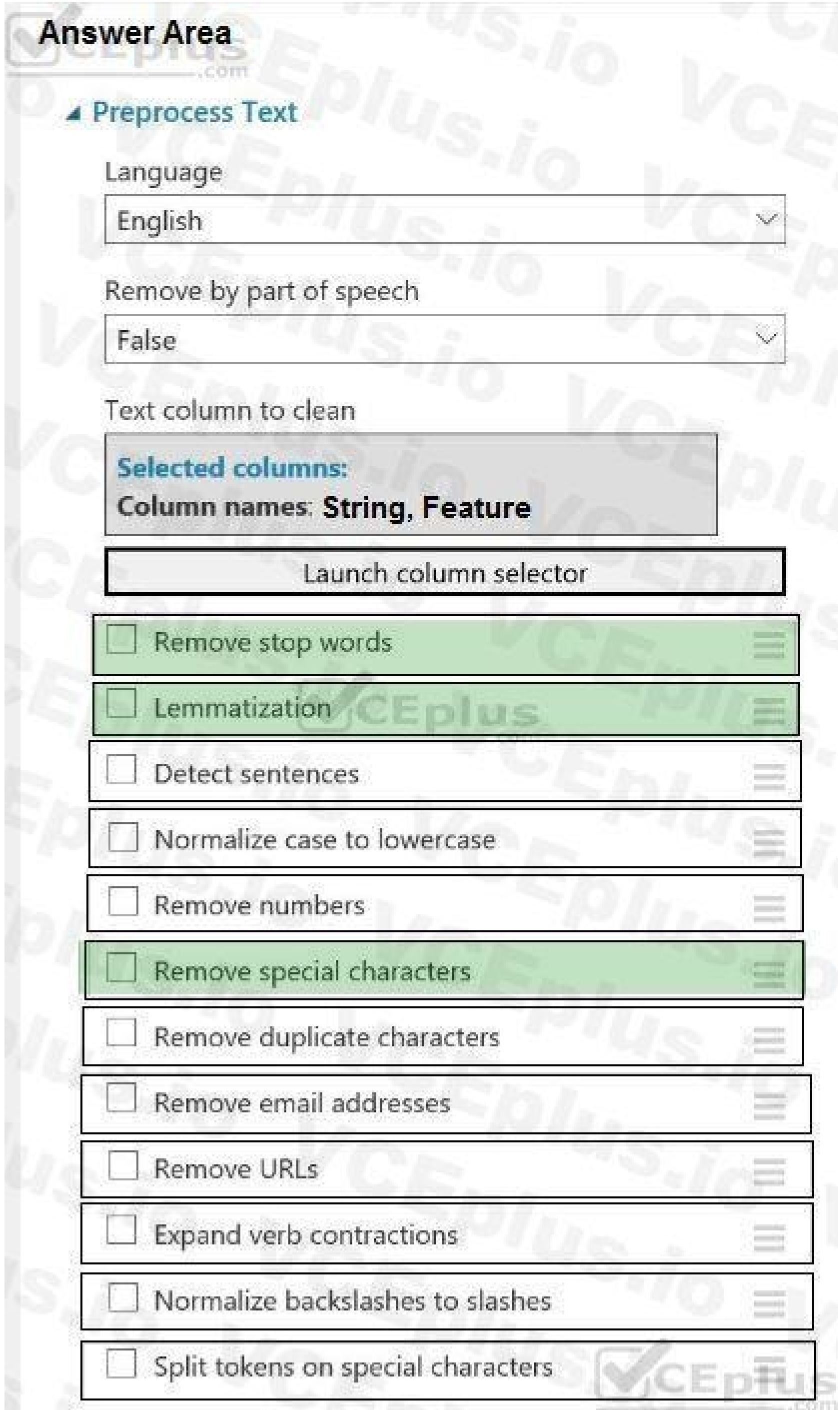
Answer Area:



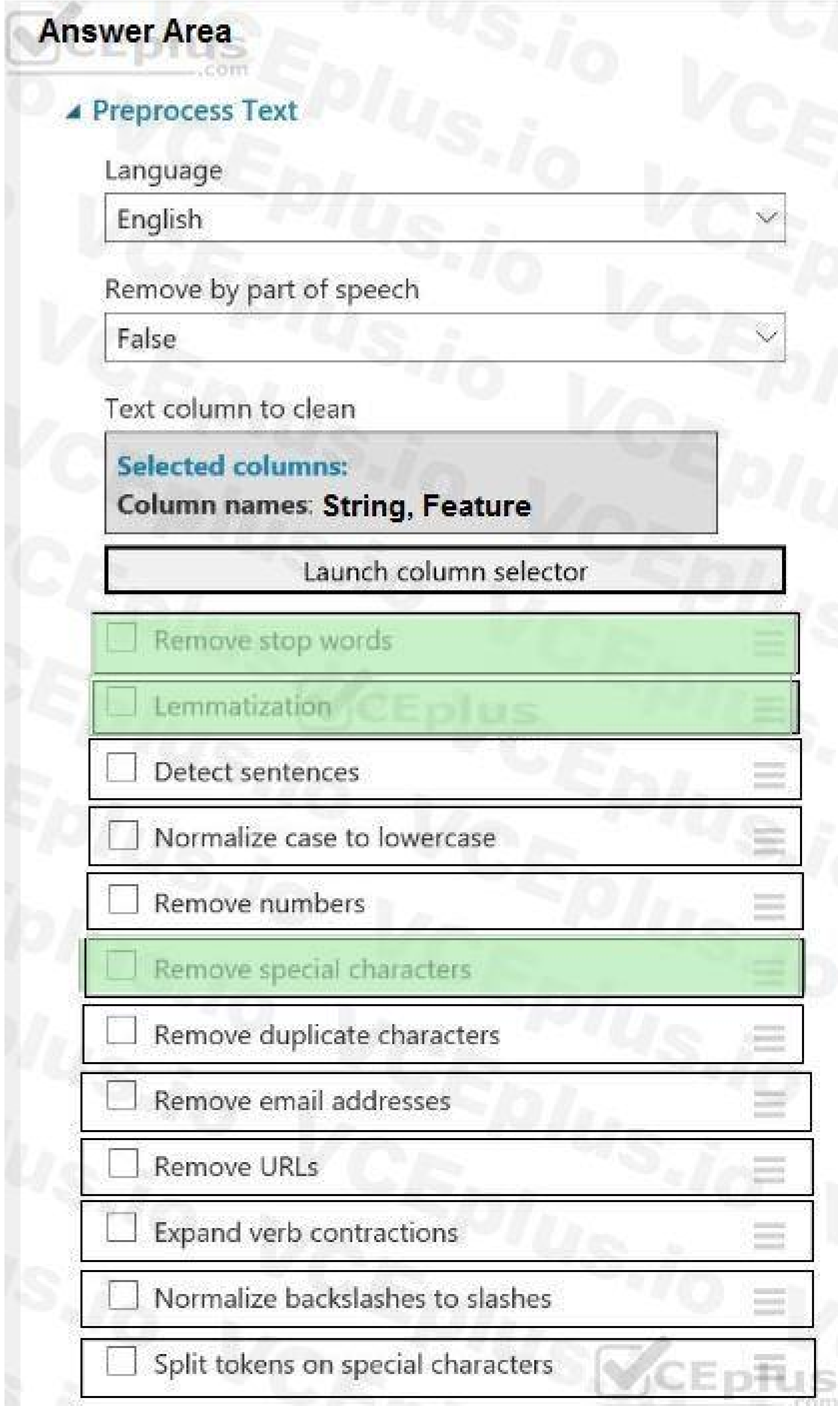
Section:
Explanation:
References:
https://www.edureka.co/blog/classification-algorithms/

QUESTION 78
HOTSPOT
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.

Hot Area:


Answer Area
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 Area:



Section:
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

QUESTION 79

DRAG DROP

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

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

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

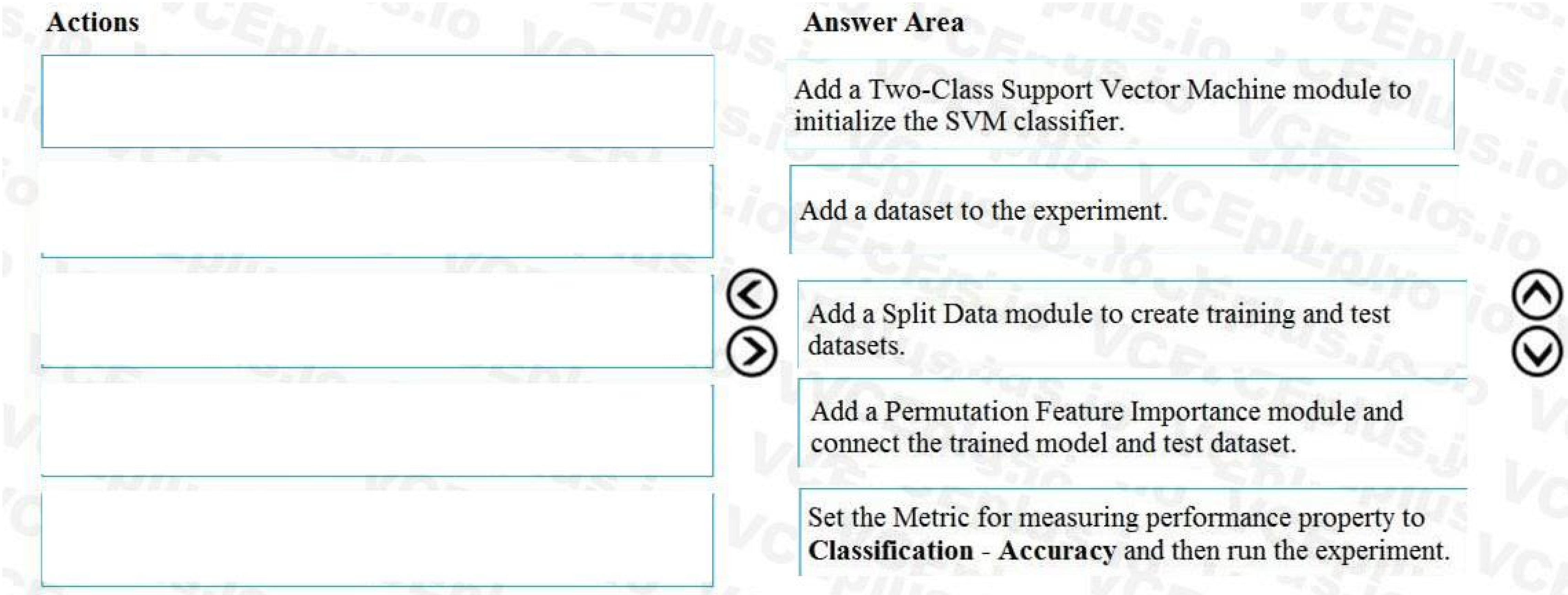
Select and Place:

Actions

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

Answer Area

Correct Answer:



Section:
Explanation:
Step 1: Add a Two-Class Support Vector Machine module to initialize the SVM classifier.
Step 2: Add a dataset to the experiment
Step 3: Add a Split Data module to create training and test dataset.
To generate a set of feature scores requires that you have an already trained model, as well as a test dataset.
Step 4: Add a Permutation Feature Importance module and connect to the trained model and test dataset.
Step 5: Set the Metric for measuring performance property to Classification - Accuracy and then run the experiment.
Reference:
https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/two-class-support-vector-machine
https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/permutation-feature-importance

QUESTION 80
HOTSPOT
You are using the Hyperdrive feature in Azure Machine Learning to train a model.
You configure the Hyperdrive experiment by running the following code:

```
from azureml.train.hyperdrive import RandomParameterSampling
param_sampling = RandomParameterSampling( {
      "learning_rate": normal(10, 3),
      "keep_probability": uniform(0.05, 0.1),
      "batch_size": choice(16, 32, 64, 128)
      "number_of_hidden_layers": choice(range(3,5))
   }
)
```

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:

	Yes	No
By defining sampling in this manner, every possible combination of the parameters will be tested.	○	○
Random values of the learning_rate parameter will be selected from a normal distribution with a mean of 10 and a standard deviation of 3.	○	○
The keep_probability parameter value will always be either **0.05** or **0.1**.	○	○
Random values for the number_of_hidden_layers parameter will be selected from a normal distribution with a mean of 3 and a standard deviation of 5.	○	○

Answer Area:

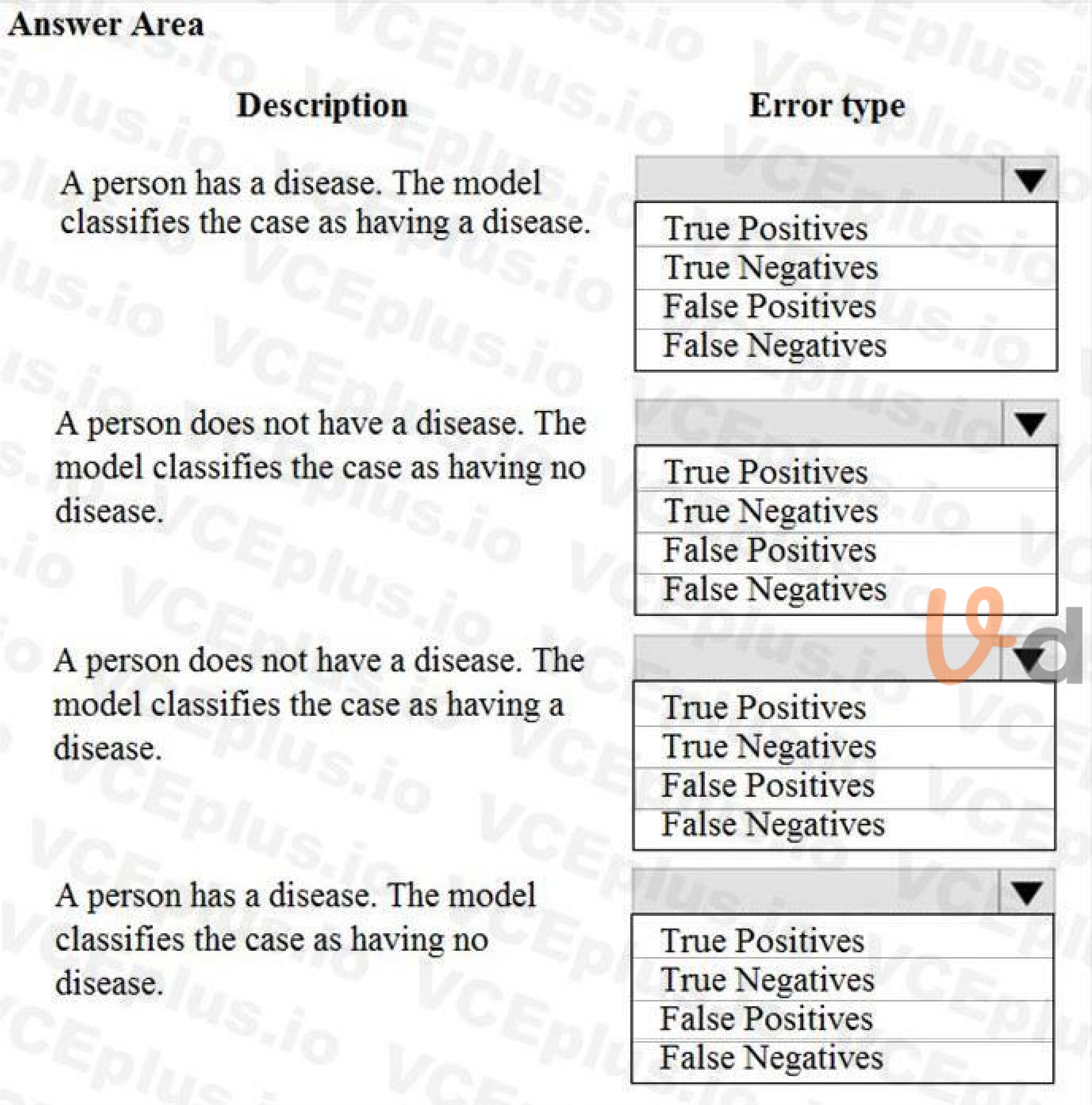
	Yes	No
By defining sampling in this manner, every possible combination of the parameters will be tested.	○	○
Random values of the learning_rate parameter will be selected from a normal distribution with a mean of 10 and a standard deviation of 3.	○	○
The keep_probability parameter value will always be either **0.05** or **0.1**.	○	○
Random values for the number_of_hidden_layers parameter will be selected from a normal distribution with a mean of 3 and a standard deviation of 5.	○	○

Section:
Explanation:
Box 1: Yes
In random sampling, hyperparameter values are randomly selected from the defined search space. Random sampling allows the search space to include both discrete and continuous hyperparameters.
Box 2: Yes
learning_rate has a normal distribution with mean value 10 and a standard deviation of 3.
Box 3: No
keep_probability has a uniform distribution with a minimum value of 0.05 and a maximum value of 0.1.
Box 4: No
number_of_hidden_layers takes on one of the values [3, 4, 5].
Reference:
https://docs.microsoft.com/en-us/azure/machine-learning/how-to-tune-hyperparameters

QUESTION 81
HOTSPOT
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.

Hot Area:



Answer Area:

Answer Area

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

Section:
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:

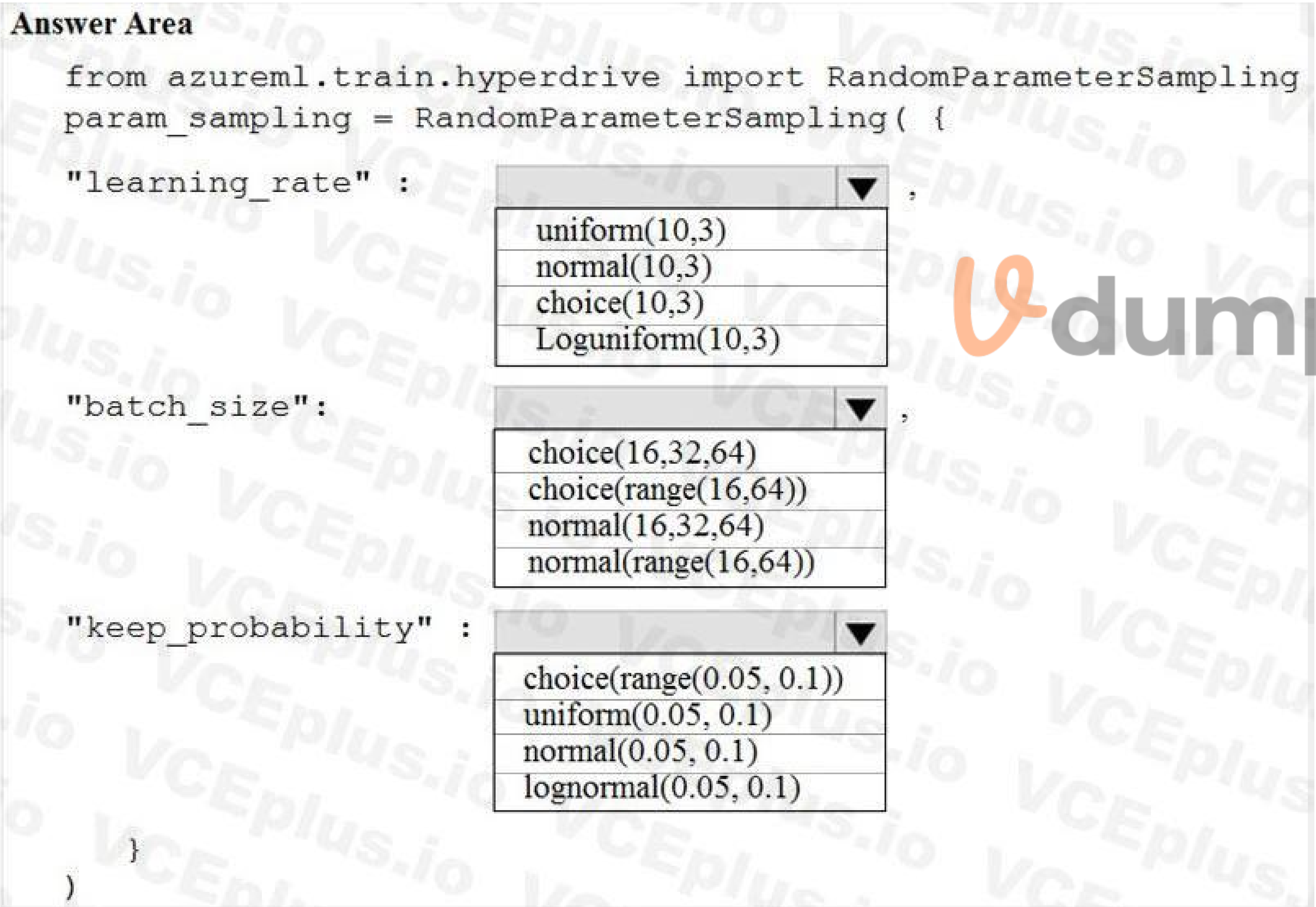
Reference:
https://developers.google.com/machine-learning/crash-course/classification/true-false-positive-negative

QUESTION 82

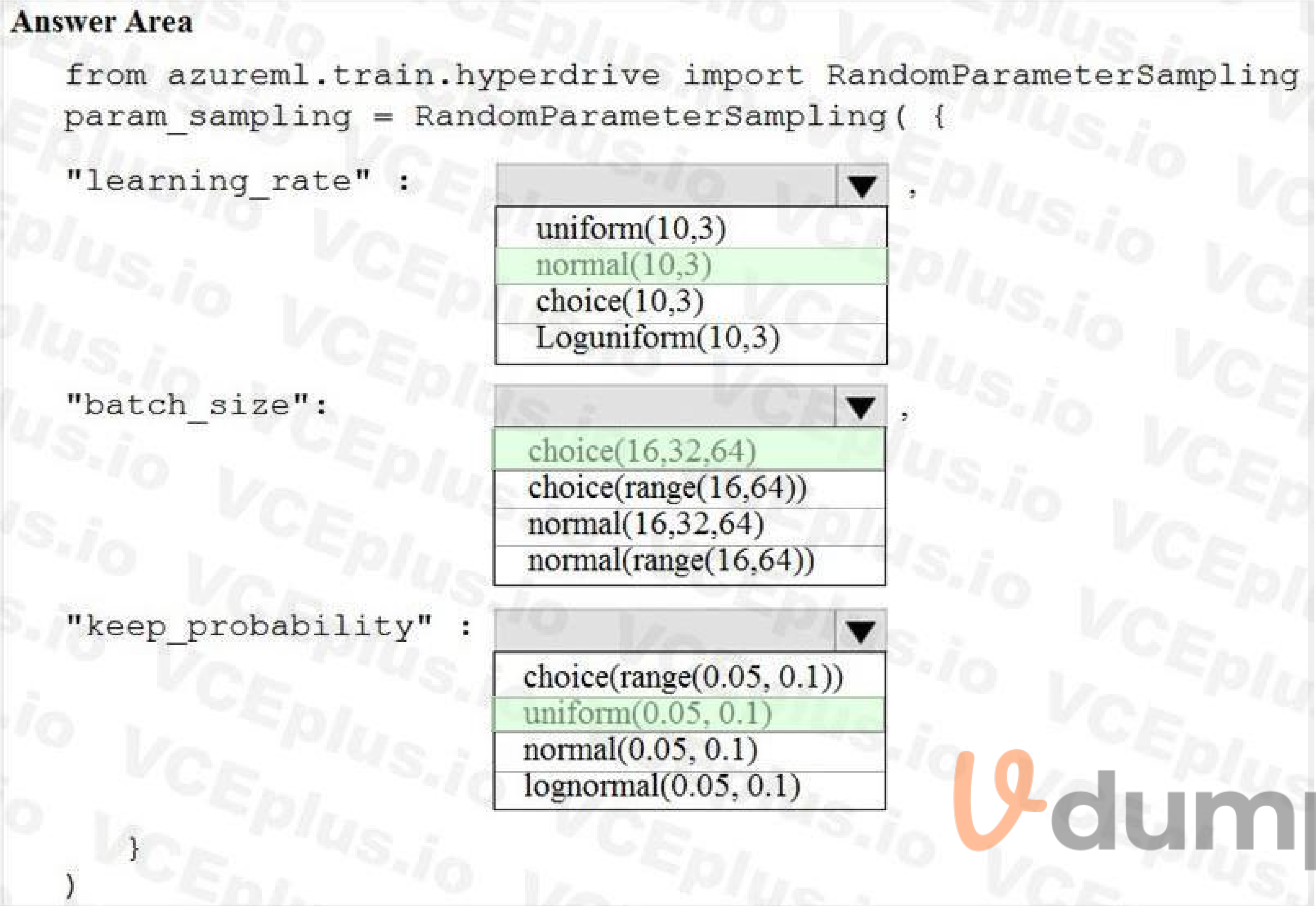
HOTSPOT

You are using the Azure Machine Learning Service to automate hyperparameter exploration of your neural network classification model.
You must define the hyperparameter space to automatically tune hyperparameters using random sampling according to following requirements:
The learning rate must be selected from a normal distribution with a mean value of 10 and a standard deviation of 3.
Batch size must be 16, 32 and 64.
Keep probability must be a value selected from a uniform distribution between the range of 0.05 and 0.1.
You need to use the param_sampling method of the Python API for the Azure Machine Learning Service.
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:



Section:

Explanation:

In random sampling, hyperparameter values are randomly selected from the defined search space. Random sampling allows the search space to include both discrete and continuous hyperparameters.

Example:

```
from azureml.train.hyperdrive import RandomParameterSampling
param_sampling = RandomParameterSampling( {
"learning_rate": normal(10, 3),
"keep_probability": uniform(0.05, 0.1),
"batch_size": choice(16, 32, 64)
}
```

Reference:

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

QUESTION 83

DRAG DROP

You create a training pipeline using the Azure Machine Learning designer. You upload a CSV file that contains the data from which you want to train your model.

You need to use the designer to create a pipeline that includes steps to perform the following tasks:

Select the training features using the pandas filter method.

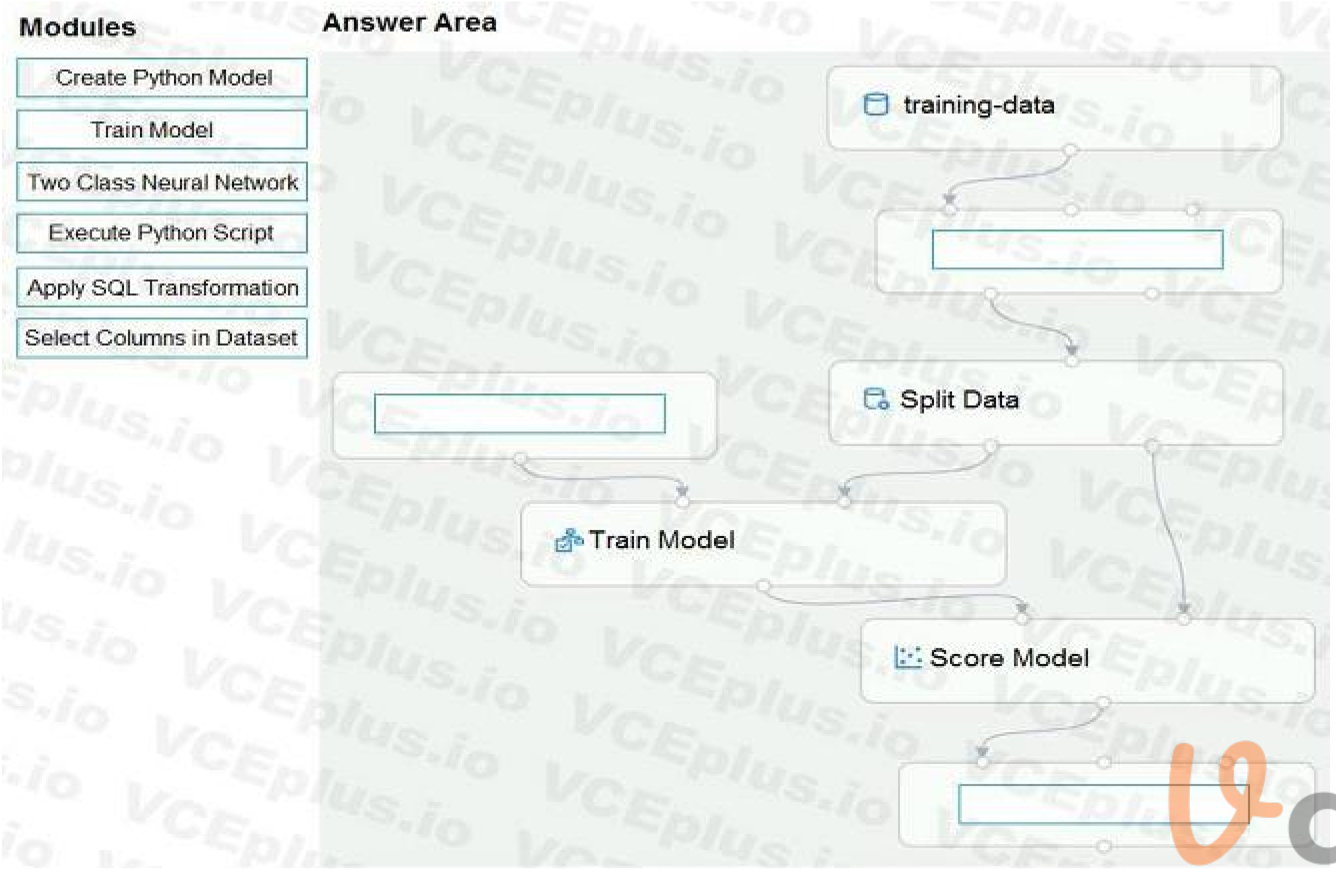
Train a model based on the naive_bayes.GaussianNB algorithm.

Return only the Scored Labels column by using the query SELECT [Scored Labels] FROM t1;

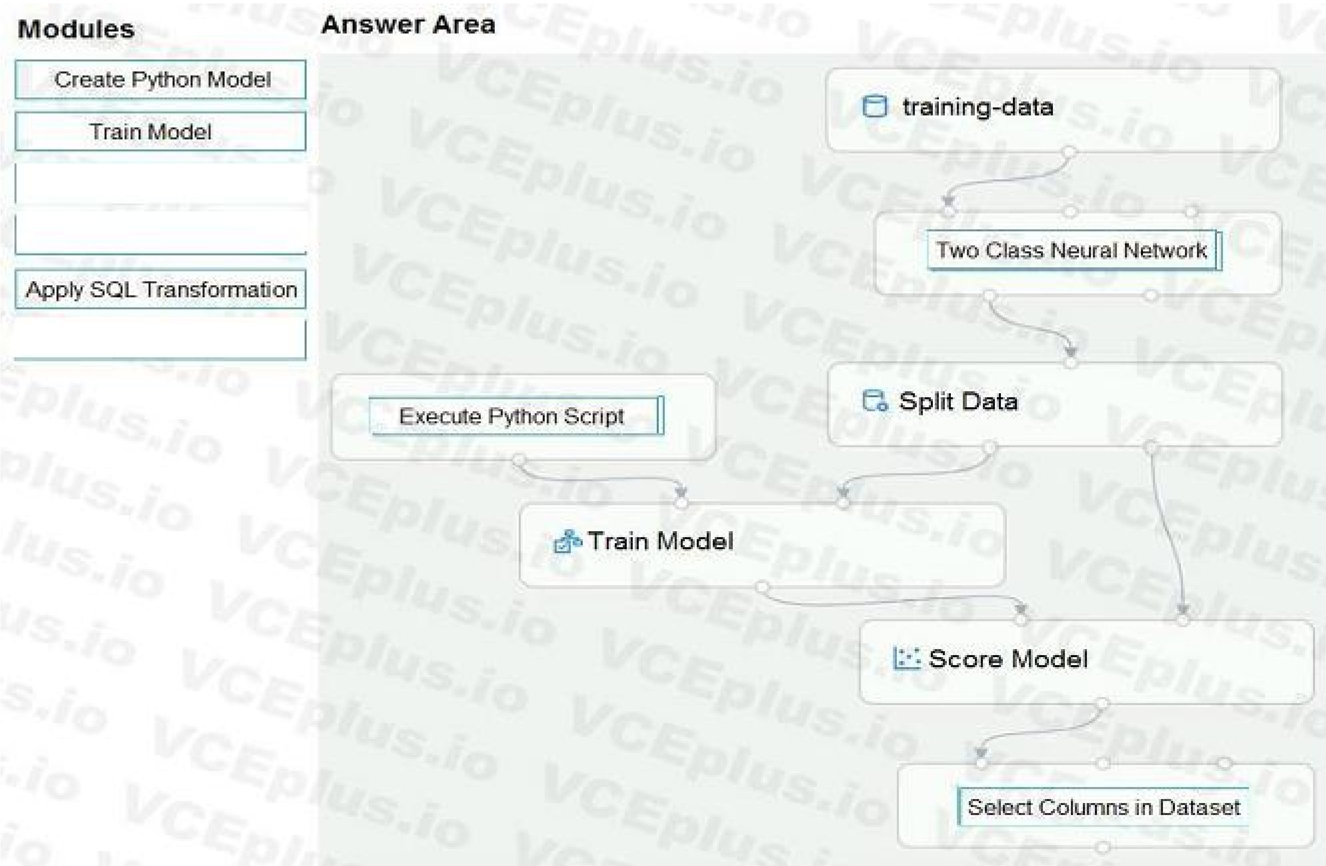
Which modules should you use? To answer, drag the appropriate modules to the appropriate locations. Each module name may be used once, more than once, or not at all. You may need to drag the split bar between panes

or scroll to view content.
NOTE: Each correct selection is worth one point.

Select and Place:



Correct Answer:



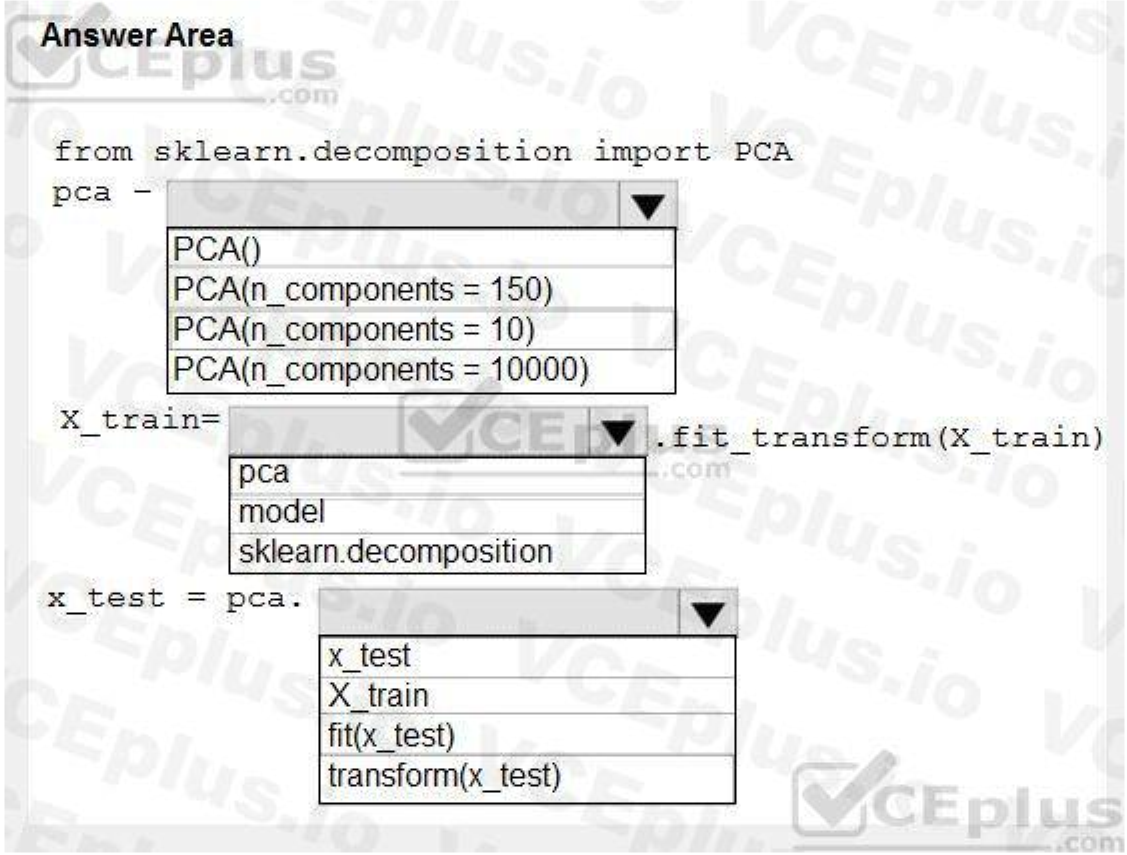
Section:
Explanation:
Box 1: Two-Class Neural Network
The Two-Class Neural Network creates a binary classifier using a neural network algorithm.
Train a model based on the naive_bayes.GaussianNB algorithm.
Box 2: Execute python script
Select the training features using the pandas filter method
Box 3: Select Columns in DataSet
Return only the Scored Labels column by using the query SELECT [Scored Labels] FROM t1;
Reference:
https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/two-class-neural-network

QUESTION 84
HOTSPOT
You have a dataset created for multiclass classification tasks that contains a normalized numerical feature set with 10,000 data points and 150 features.
You use 75 percent of the data points for training and 25 percent for testing. You are using the scikit-learn machine learning library in Python. You use X to denote the feature set and Y to denote class labels.
You create the following Python data frames:

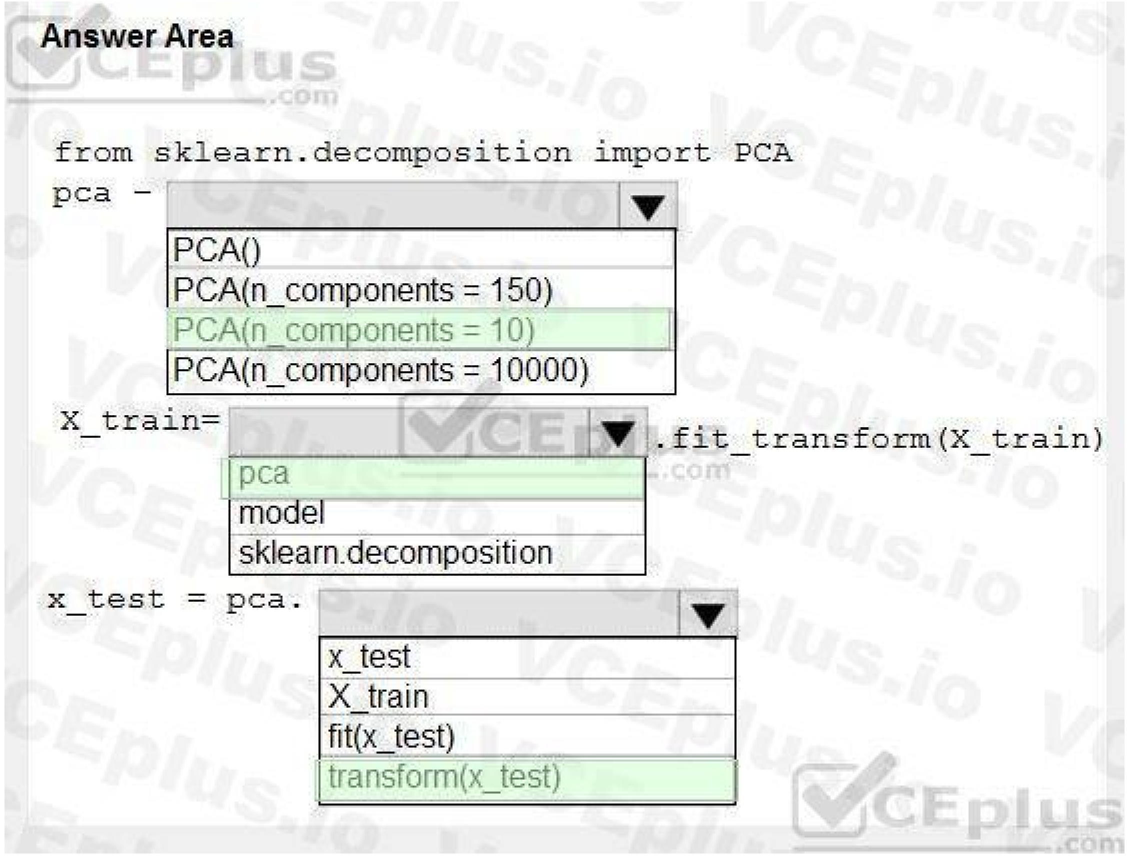
Name	Description
X_train	training feature set
Y_train	training class labels
x_train	testing feature set
y_train	testing class labels

You need to apply the Principal Component Analysis (PCA) method to reduce the dimensionality of the feature set to 10 features in both training and testing sets.
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:



Section:

Explanation:

Box 1: PCA(n_components = 10)

Need to reduce the dimensionality of the feature set to 10 features in both training and testing sets.

Example:

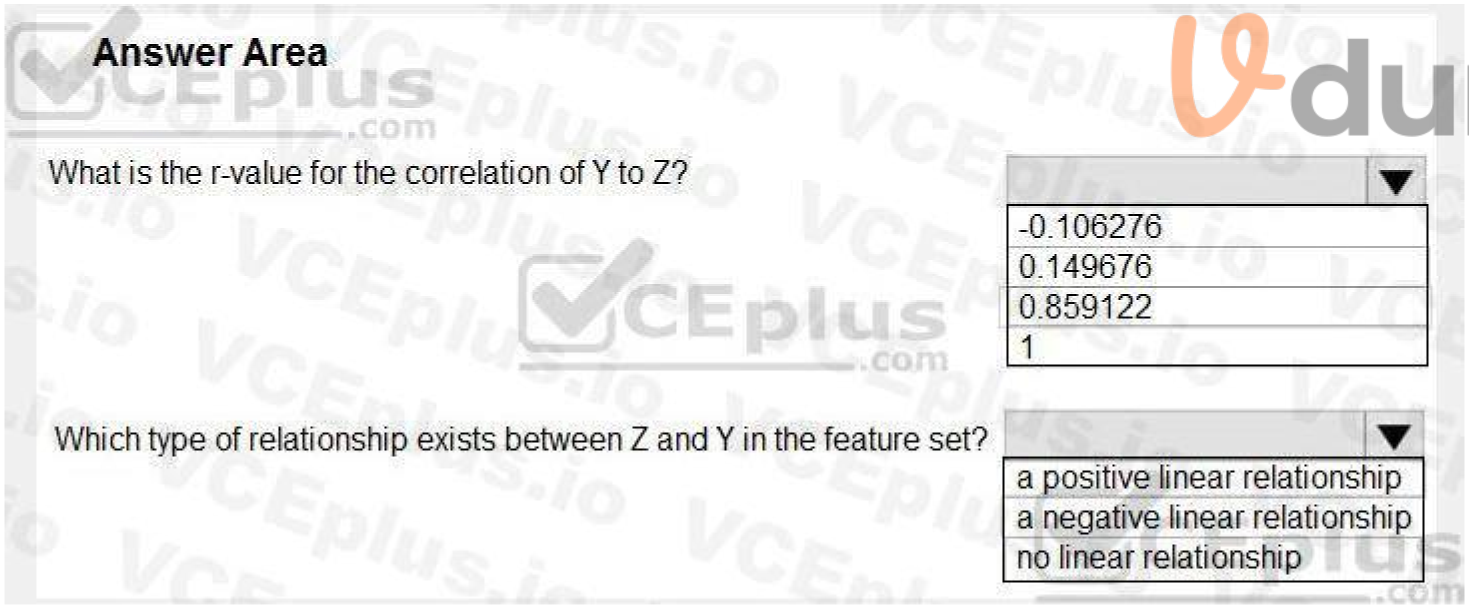
from sklearn.decomposition import PCA
pca = PCA(n_components=2) ;2 dimensions
principalComponents = pca.fit_transform(x)
Box 2: pca
fit_transform(X[, y])fits the model with X and apply the dimensionality reduction on X.
Box 3: transform(x_test)
transform(X) applies dimensionality reduction to X.
References:
https://scikit-learn.org/stable/modules/generated/sklearn.decomposition.PCA.html

QUESTION 85
HOTSPOT
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:

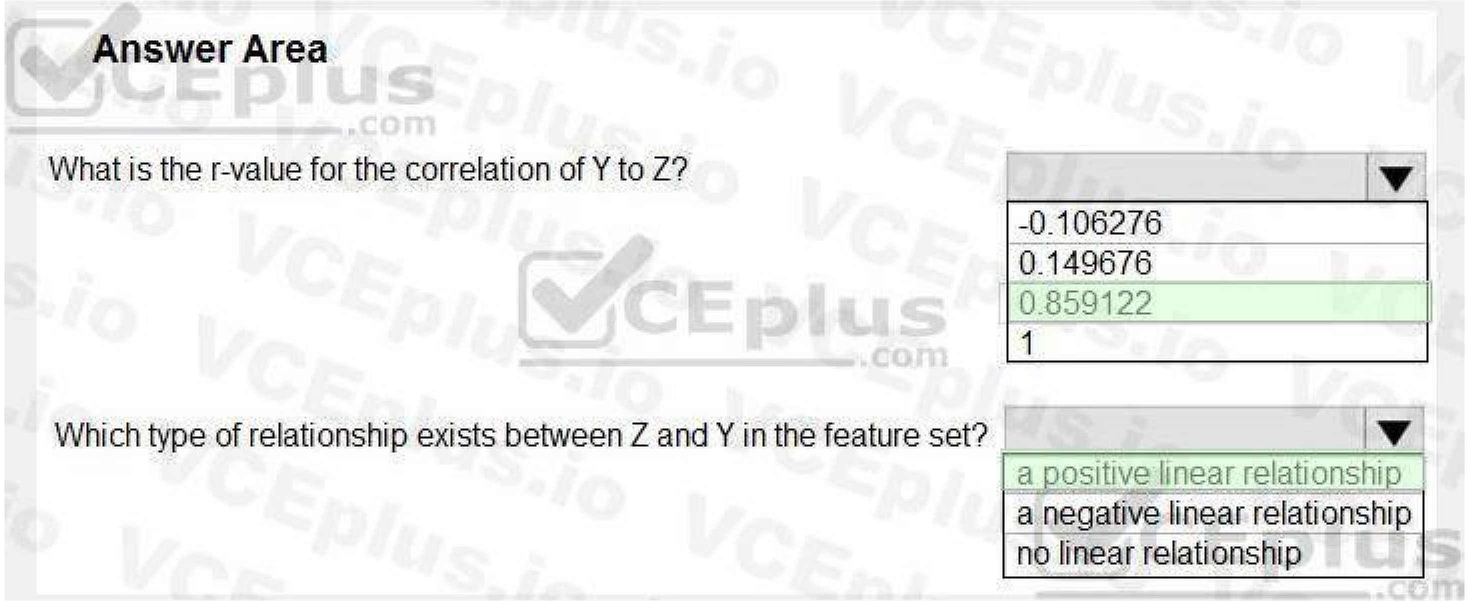
	X	Y	Z
X	1	0.149676	-0.106276
Y	0.149676	1	0.859122
Z	-0.106276	0.859122	1

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.

Hot Area:



Answer Area:



Section:
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

QUESTION 86
DRAG DROP
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.

Select and Place:

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()
```

Correct Answer:

Code segments

- log_list
- log_row

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()
```

Section:

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.

QUESTION 87
HOTSPOT
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.

Hot Area:

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',
                 [dropdown 1] =split1_ds,
                 [dropdown 2] =split2_ds
                 [dropdown 3] ='price')
```

Dropdown 1:
- X
- Y
- X_valid
- Y_valid
- training_data

Dropdown 2:
- X
- Y
- X_valid
- Y_valid
- validation_data
- training_data

Dropdown 3:
- y
- y_valid
- y_max
- label_column_name
- exclude_nan_labels

Answer Area:

Answer Area

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

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

X
Y
X_valid
Y_valid
training_data

X
Y
X_valid
Y_valid
validation_data
training_data

y
y_valid
y_max
label_column_name
exclude_nan_labels

Section:
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

QUESTION 88
HOTSPOT
You have a multi-class image classification deep learning model that uses a set of labeled photographs. You create the following code to select hyperparameter values when training the model.

```
from azureml.train.hyperdrive import BayesianParameterSampling
param_sampling = BayesianParametersSampling ({
 "learning_rate": uniform(0.01, 0.1),
 "batch_size": choice(16, 32, 64, 128)}
)
```

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:

	Yes	No
Hyperparameter combinations for the runs are selected based on how previous samples performed in the previous experiment run.	○	○
The learning rate value 0.09 might be used during model training.	○	○
You can define an early termination policy for this hyperparameter tuning run.	○	○

Answer Area:

	Yes	No
Hyperparameter combinations for the runs are selected based on how previous samples performed in the previous experiment run.	○	○
The learning rate value 0.09 might be used during model training.	○	○
You can define an early termination policy for this hyperparameter tuning run.	○	○

Section:
Explanation:
Box 1: Yes
Hyperparameters are adjustable parameters you choose to train a model that govern the training process itself. Azure Machine Learning allows you to automate hyperparameter exploration in an efficient manner, saving you significant time and resources. You specify the range of hyperparameter values and a maximum number of training runs. The system then automatically launches multiple simultaneous runs with different parameter configurations and finds the configuration that results in the best performance, measured by the metric you choose. Poorly performing training runs are automatically early terminated, reducing wastage of compute resources. These resources are instead used to explore other hyperparameter configurations.
Box 2: Yes
uniform(low, high) - Returns a value uniformly distributed between low and high
Box 3: No
Bayesian sampling does not currently support any early termination policy.
Reference:
https://docs.microsoft.com/en-us/azure/machine-learning/how-to-tune-hyperparameters

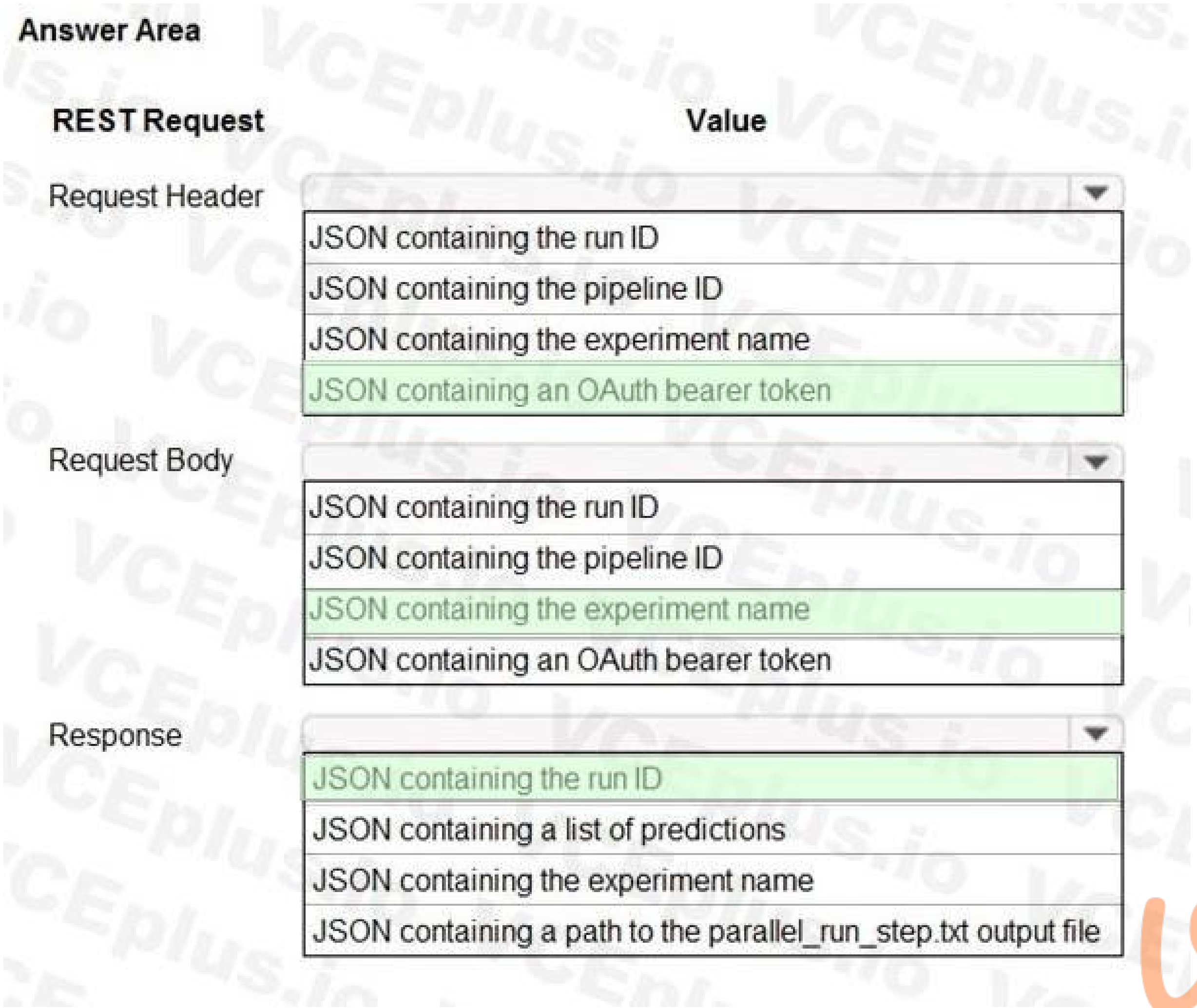
QUESTION 89
HOTSPOT
You publish a batch inferencing pipeline that will be used by a business application.
The application developers need to know which information should be submitted to and returned by the REST interface for the published pipeline.
You need to identify the information required in the 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.

Hot Area:

Answer Area

REST Request	Value
Request Header	JSON containing the run ID JSON containing the pipeline ID JSON containing the experiment name JSON containing an OAuth bearer token
Request Body	JSON containing the run ID JSON containing the pipeline ID JSON containing the experiment name JSON containing an OAuth bearer token
Response	JSON containing the run ID JSON containing a list of predictions JSON containing the experiment name JSON containing a path to the parallel_run_step.txt output file

Answer Area:



Section:
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:

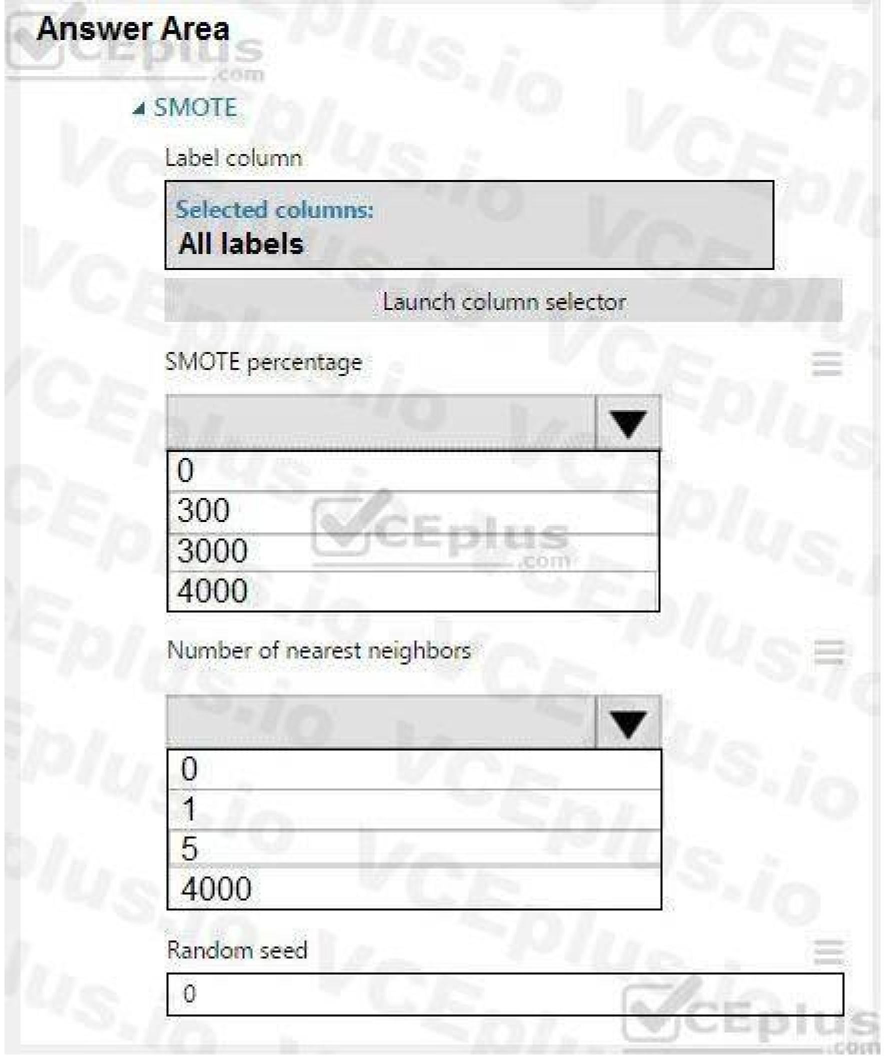
```
rest_endpoint = published_pipeline.endpoint
response = requests.post(rest_endpoint,
headers=auth_header,
json={"ExperimentName": "batch_scoring",
"ParameterAssignments": {"process_count_per_node": 6}})
run_id = response.json()["Id"]
```

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

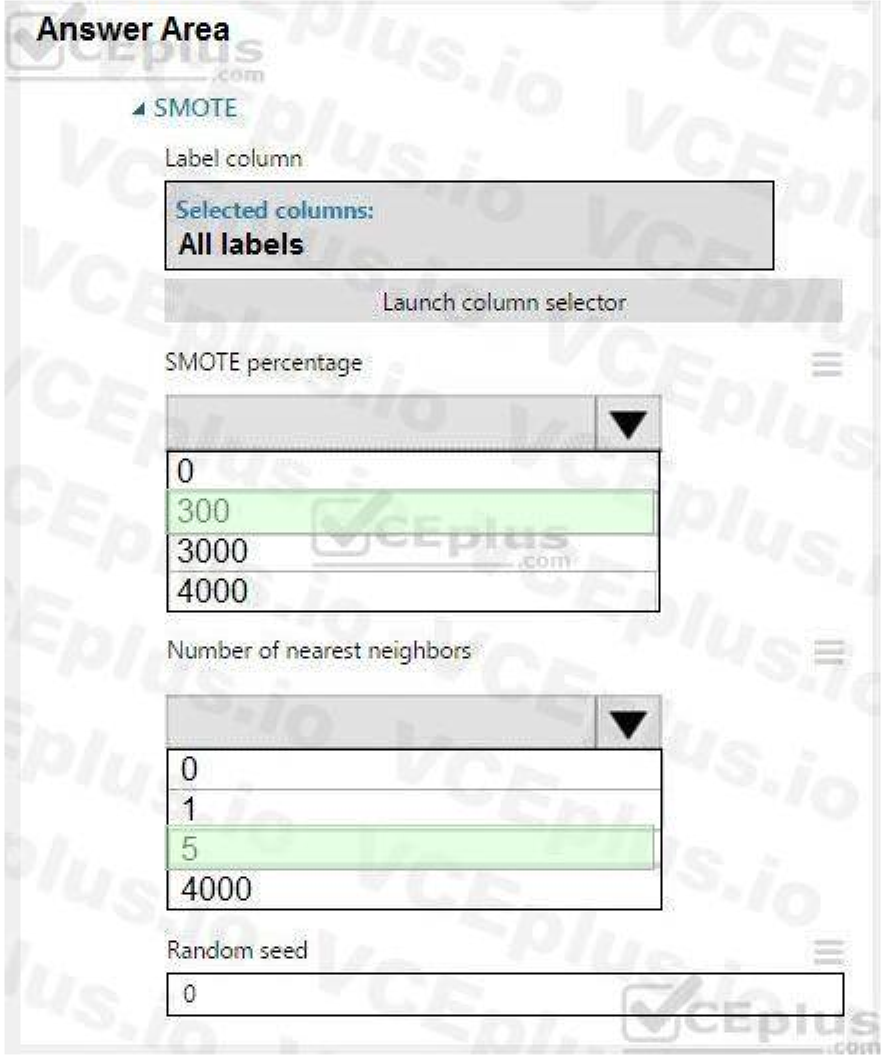
QUESTION 90
HOTSPOT
You create an experiment in Azure Machine Learning Studio. You add a training dataset that contains 10,000 rows. The first 9,000 rows represent class 0 (90 percent).
The remaining 1,000 rows represent class 1 (10 percent).

The training set is imbalances between two classes. You must increase the number of training examples for class 1 to 4,000 by using 5 data rows. You add the Synthetic Minority Oversampling Technique (SMOTE) module to the experiment.
You need to configure the module.
Which values should you use? To answer, select the appropriate options in the dialog box in the answer area.
NOTE: Each correct selection is worth one point.

Hot Area:



Answer Area:



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

QUESTION 91
HOTSPOT
You are running Python code interactively in a Conda environment. The environment includes all required Azure Machine Learning SDK and MLflow packages.
You must use MLflow to log metrics in an Azure Machine Learning experiment named mlflow-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.

Hot Area:

Answer Area

```
import mlflow
from azureml.core import Workspace
ws = Workspace.from_config()
# Set the MLflow logging target
[ ▼ ]
  mlflow.tracking.client = ws
  mlflow.set_tracking_uri(ws.get_mlflow_tracking_uri())
  mlflow.log_param('workspace', ws)
# Configure experiment
[ ▼ ]
  mlflow-experiment = Run.get_context()
  mlflow.get_run('mlflow-experiment')
  mlflow.set_experiment('mlflow-experiment')
# Begin the experiment run
with [ ▼ ]
       mlflow.active_run
       mlflow.start_run()
       Run.get_context()
    # Log my_metric with value 1.00
    [ ▼ ] ('my_metric', 1.00)
      run.log()
      mlflow.log_metric
      print
print("Finished!")
```

Answer Area:

Answer Area

```
import mlflow
from azureml.core import Workspace
ws = Workspace.from_config()
# Set the MLflow logging target
```

mlflow.tracking.client = ws
mlflow.set_tracking_uri(ws.get_mlflow_tracking_uri())
mlflow.log_param('workspace', ws)

```
# Configure experiment
```

mlflow-experiment = Run.get_context()
mlflow.get_run('mlflow-experiment')
mlflow.set_experiment('mlflow-experiment')

```
# Begin the experiment run
with
```

mlflow.active_run
mlflow.start_run()
Run.get_context()

```
    # Log my_metric with value 1.00
```

run.log()
mlflow.log_metric
print

```
('my_metric', 1.00)
print("Finished!")
```

Section:

Explanation:

Box 1: mlflow.set_tracking_uri(ws.get_mlflow_tracking_uri())

In the following code, the get_mlflow_tracking_uri() method assigns a unique tracking URI address to the workspace, ws, and set_tracking_uri() points the MLflow tracking URI to that address.

mlflow.set_tracking_uri(ws.get_mlflow_tracking_uri())

Box 2: mlflow.set_experiment(experiment_name)

Set the MLflow experiment name with set_experiment() and start your training run with start_run().

Box 3: mlflow.start_run()

Box 4: mlflow.log_metric

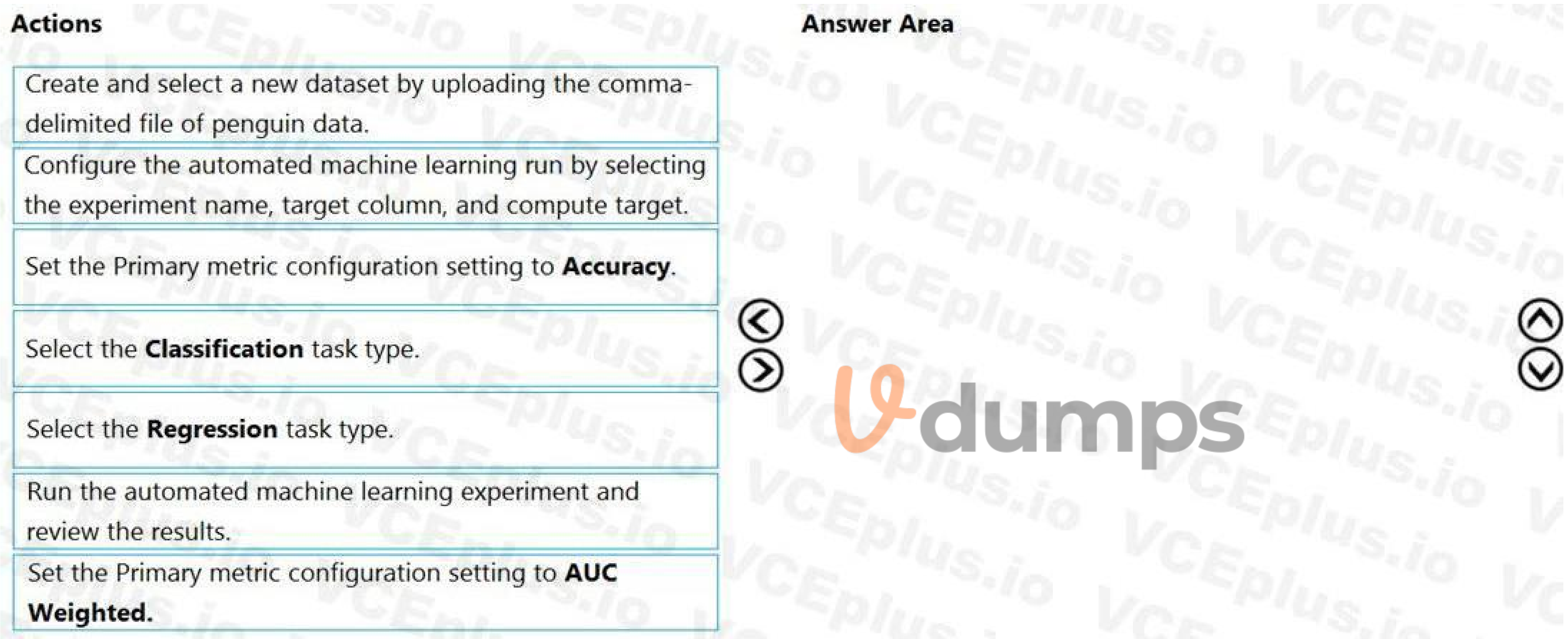
Then use log_metric() to activate the MLflow logging API and begin logging your training run metrics.

Reference:

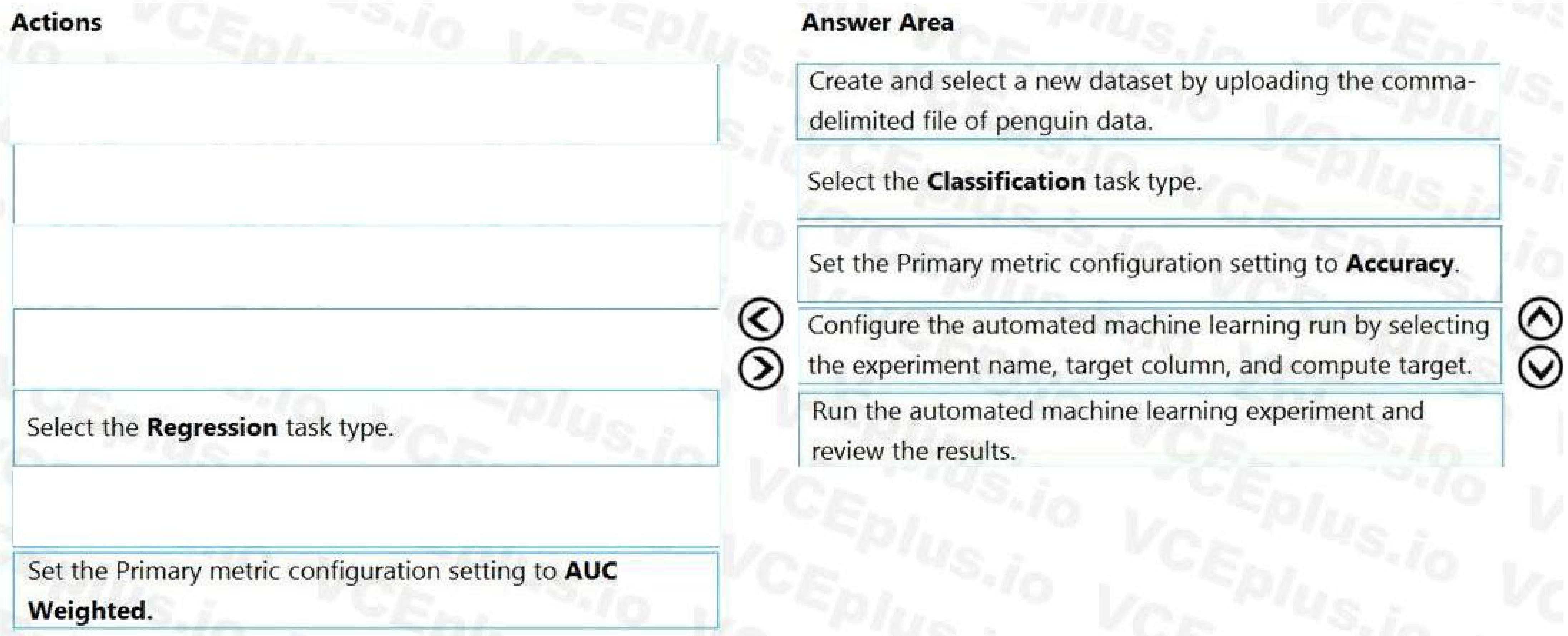
https://docs.microsoft.com/en-us/azure/machine-learning/how-to-use-mlflow

QUESTION 92
DRAG DROP
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 three 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 correct order.

Select and Place:



Correct Answer:



Section:
Explanation:
Step 1:Create and select a new dataset by uploading he command-delimited file of penguin data.
Step 2: Select the Classification task type
Step 3: Set the Primary metric configuration setting to Accuracy.
The available metrics you can select is determined by the task type you choose.
Primary metrics for classification scenarios:
Post thresholded metrics, like accuracy, average_precision_score_weighted, norm_macro_recall, and precision_score_weighted may not optimize as well for datasets which are very small, have very large class skew (class imbalance), or when the expected metric value is very close to 0.0 or 1.0. In those cases, AUC_weighted can be a better choice for the primary metric.
Step 4: Configure the automated machine learning run by selecting the experiment name, target column, and compute target
Step 5: Run the automated machine learning experiment and review the results.
Reference:
https://docs.microsoft.com/en-us/azure/machine-learning/how-to-configure-auto-train

QUESTION 93
HOTSPOT
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.

Hot Area:

Answer Area

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 Area:

Answer Area

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

Section:

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:

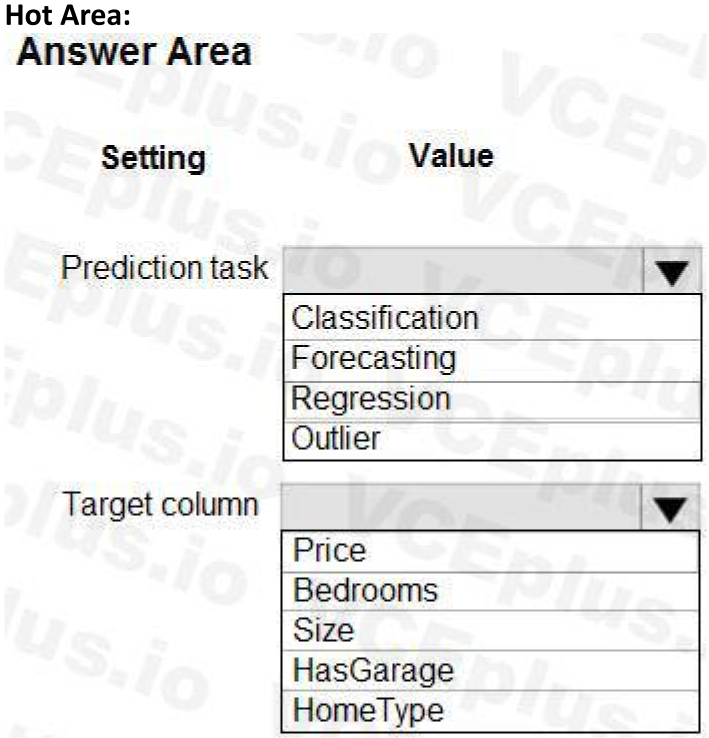
Tabular Explainer: Used with tabular datasets.
Text Explainer: Used with text datasets.
Image Explainer: Used with image datasets.
Box 1: Tabular
Box 2: Text
Box 3: Image
Incorrect Answers:
Hierarchical Attention Network (HAN)
HAN was proposed by Yang et al. in 2016. Key features of HAN that differentiates itself from existing approaches to document classification are (1) it exploits the hierarchical nature of text data and (2) attention mechanism is adapted for document classification.
Reference:
https://medium.com/microsoftazure/automated-and-interpretable-machine-learning-d07975741298

QUESTION 94
HOTSPOT
You have a dataset that includes home sales data for a city. The dataset includes the following columns.

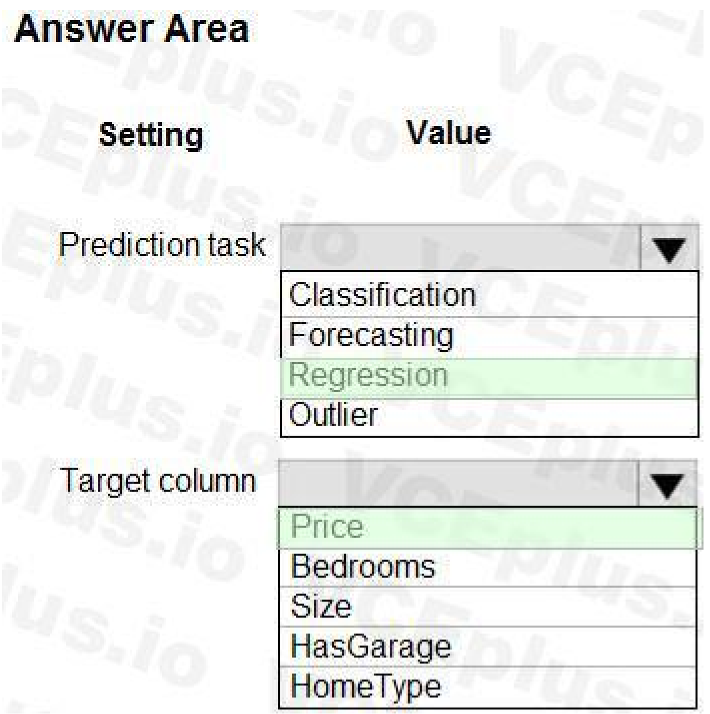
Name	Description
Price	The sales price for the house.
Bedrooms	The number of bedrooms in the house.
Size	The size of the house in square feet.
HasGarage	A binary value indicating whether or not the house has a garage.
HomeType	The category of home, for example, apartment, townhouse, single-family home.

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.

Hot Area:



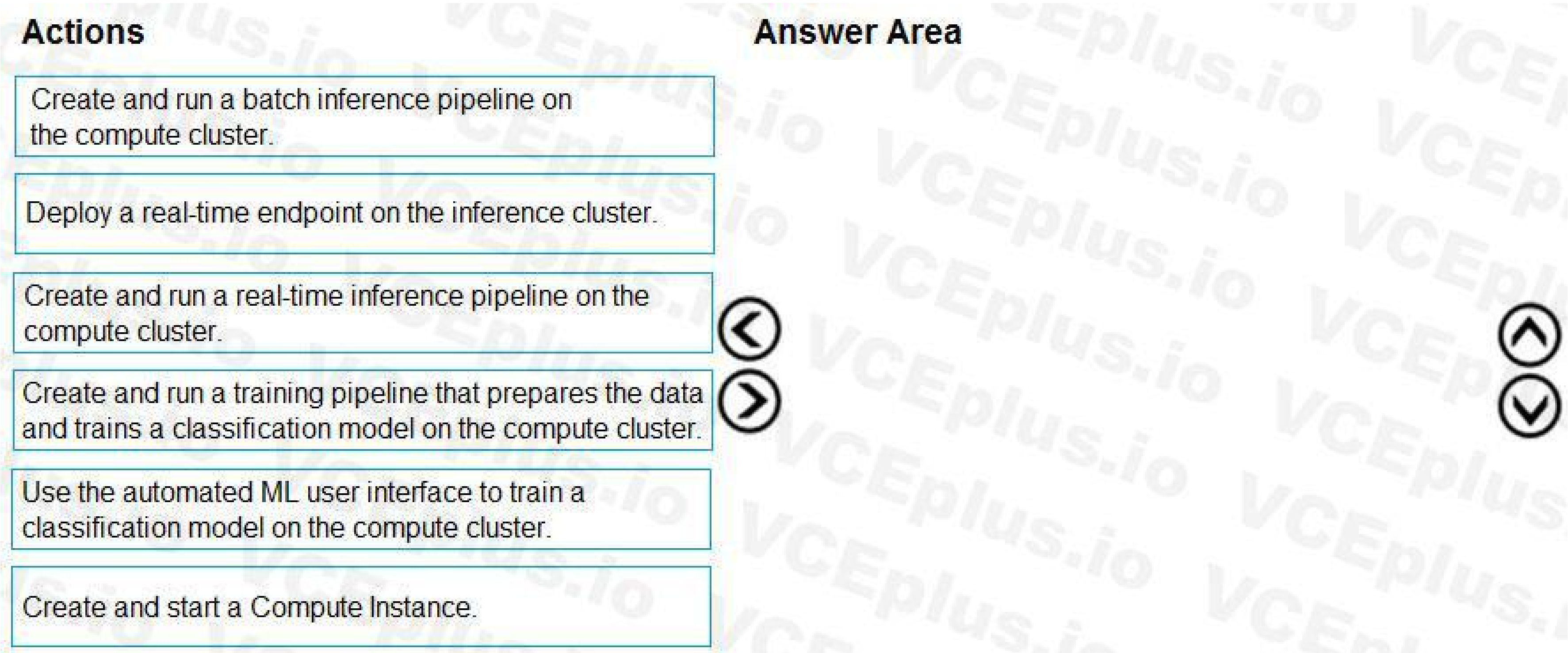
Answer Area:



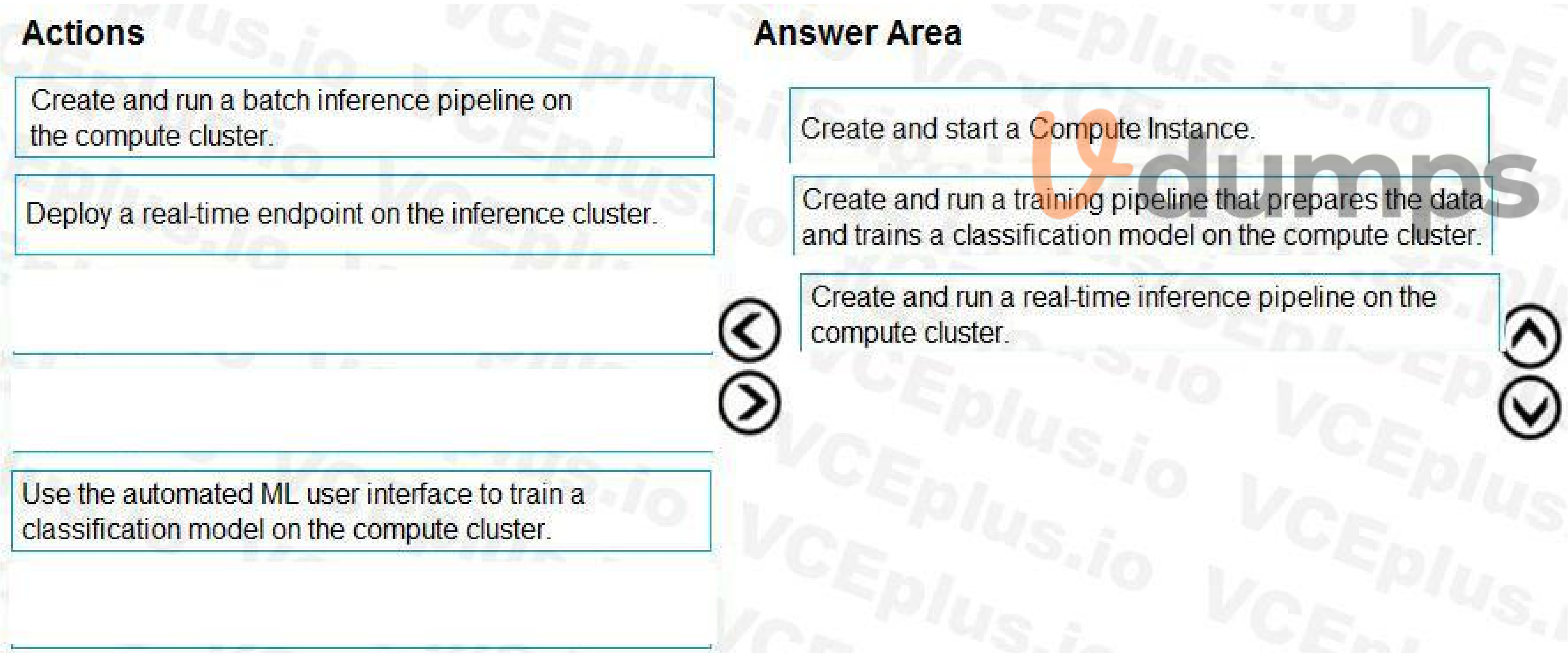
Section:
Explanation:
Box 1: Regression
Regression is a supervised machine learning technique used to predict numeric values.
Box 2: Price
Reference:
https://docs.microsoft.com/en-us/learn/modules/create-regression-model-azure-machine-learning-designer

QUESTION 95
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:



Section:
Explanation:
Step 1: Create and start a Compute Instance To train and deploy models using Azure Machine Learning designer, you need compute on which to run the training process, test the model, and host the model in a deployed service.
There are four kinds of compute resource you can create:
Compute Instances: Development workstations that data scientists can use to work with data and models.
Compute Clusters: Scalable clusters of virtual machines for on-demand processing of experiment code.
Inference Clusters: Deployment targets for predictive services that use your trained models.
Attached Compute: Links to existing Azure compute resources, such as Virtual Machines or Azure Databricks clusters.
Step 2: Create and run a training pipeline..
After you've used data transformations to prepare the data, you can use it to train a machine learning model. Create and run a training pipeline

Step 3: Create and run a real-time inference pipeline
After creating and running a pipeline to train the model, you need a second pipeline that performs the same data transformations for new data, and then uses the trained model to inference (in other words, predict) label values based on its features. This pipeline will form the basis for a predictive service that you can publish for applications to use.
Reference:
https://docs.microsoft.com/en-us/learn/modules/create-classification-model-azure-machine-learning-designer/

QUESTION 96
HOTSPOT
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.

Hot Area:

Answer Area

```
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')

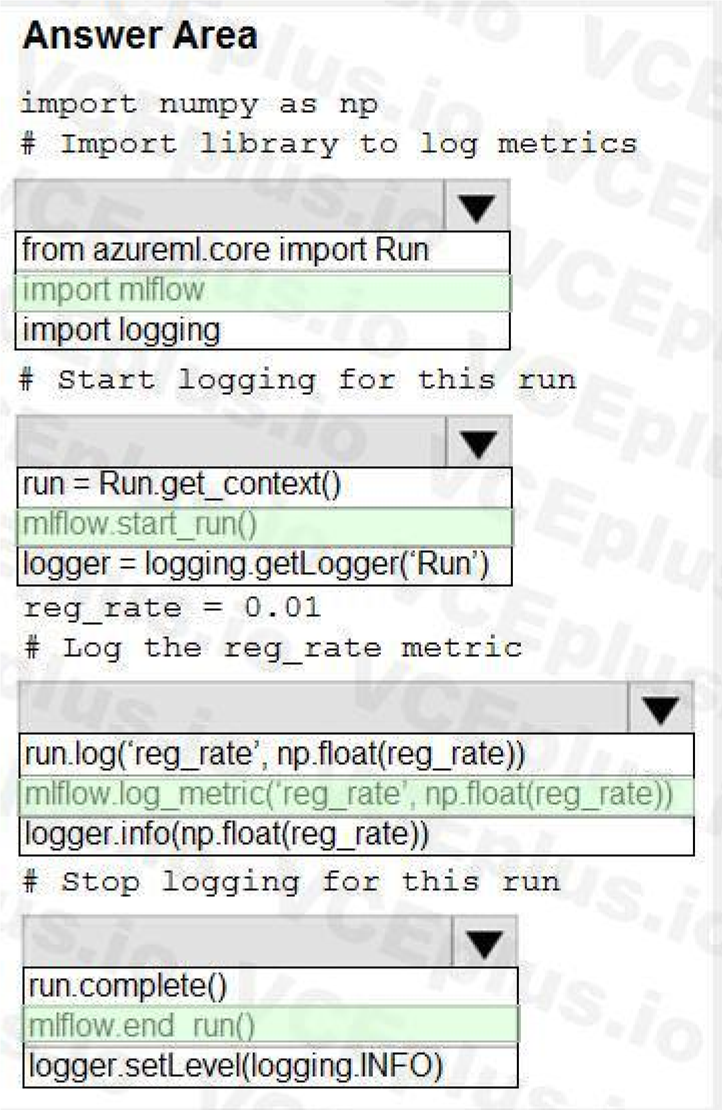
```
reg_rate = 0.01
# Log the reg_rate metric
```

- 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 Area:



Section:
Explanation:
Box 1: import mlflow
Import the mlflow and Workspace classes to access MLflow's tracking URI and configure your workspace.
Box 2: mlflow.start_run()
Set the MLflow experiment name with set_experiment() and start your training run with start_run().
Box 3: mlflow.log_metric(' ..')
Use log_metric() to activate the MLflow logging API and begin logging your training run metrics.
Box 4: mlflow.end_run()
Close the run:
run.endRun()
Reference:
https://docs.microsoft.com/en-us/azure/machine-learning/how-to-use-mlflow

02 - Run experiments and train models
Case study
Overview
You are a data scientist in a company that provides data science for professional sporting events. Models will use global and local market data to meet the following business goals:
Understand sentiment of mobile device users at sporting events based on audio from crowd reactions.
Assess a user's tendency to respond to an advertisement.

Customize styles of ads served on mobile devices.
Use video to detect penalty events
Current environment
Media used for penalty event detection will be provided by consumer devices. Media may include images and videos captured during the sporting event and shared using social media. The images and videos will have varying sizes and formats.
The data available for model building comprises of seven years of sporting event media. The sporting event media includes; recorded video transcripts or radio commentary, and logs from related social media feeds captured during the sporting events.
Crowd sentiment will include audio recordings submitted by event attendees in both mono and stereo formats.
Penalty detection and sentiment
Data scientists must build an intelligent solution by using multiple machine learning models for penalty event detection.
Data scientists must build notebooks in a local environment using automatic feature engineering and model building in machine learning pipelines.
Notebooks must be deployed to retrain by using Spark instances with dynamic worker allocation.
Notebooks must execute with the same code on new Spark instances to recode only the source of the data.
Global penalty detection models must be trained by using dynamic runtime graph computation during training.
Local penalty detection models must be written by using BrainScript.
Experiments for local crowd sentiment models must combine local penalty detection data.
Crowd sentiment models must identify known sounds such as cheers and known catch phrases. Individual crowd sentiment models will detect similar sounds.
All shared features for local models are continuous variables.
Shared features must use double precision. Subsequent layers must have aggregate running mean and standard deviation metrics available.
Advertisements
During the initial weeks in production, the following was observed:
Ad response rated declined.
Drops were not consistent across ad styles.
The distribution of features across training and production data are not consistent
Analysis shows that, of the 100 numeric features on user location and behavior, the 47 features that come from location sources are being used as raw features. A suggested experiment to remedy the bias and variance issue is to engineer 10 linearly uncorrelated features.
Initial data discovery shows a wide range of densities of target states in training data used for crowd sentiment models.
All penalty detection models show inference phases using a Stochastic Gradient Descent (SGD) are running too slow.
Audio samples show that the length of a catch phrase varies between 25%-47% depending on region The performance of the global penalty detection models shows lower variance but higher bias when comparing training and validation sets. Before implementing any feature changes, you must confirm the bias and variance using all training and validation cases.
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.
Sampling must guarantee mutual and collective exclusively between local and global segmentation models that share the same features.
Local market segmentation models will be applied before determining a user's propensity to respond to an advertisement.
Ad response models must support non-linear boundaries of features.
The ad propensity model uses a cut threshold is 0.45 and retrains occur if weighted Kappa deviated from 0.1 +/- 5%.
The ad propensity model uses cost factors shown in the following diagram:

		Actual	
		1	**0**
Predicted	**0**	1	2
	1	2	1

The ad propensity model uses proposed cost factors shown in the following diagram:

		Actual	
		1	0
Predicted	0	1	5
	1	5	1

Performance curves of current and proposed cost factor scenarios are shown in the following diagram:

QUESTION 1
You need to implement a scaling strategy for the local penalty detection data.
Which normalization type should you use?

A. Streaming
B. Weight
C. Batch
D. Cosine

Correct Answer: C
Section:
Explanation:
Post batch normalization statistics (PBN) is the Microsoft Cognitive Toolkit (CNTK) version of how to evaluate the population mean and variance of Batch Normalization which could be used in inference Original Paper.
In CNTK, custom networks are defined using the BrainScriptNetworkBuilder and described in the CNTK network description language "BrainScript." Scenario:
Local penalty detection models must be written by using BrainScript.
Reference:
https://docs.microsoft.com/en-us/cognitive-toolkit/post-batch-normalization-statistics

QUESTION 2
You need to implement a feature engineering strategy for the crowd sentiment local models.
What should you do?

A. Apply an analysis of variance (ANOVA).
B. Apply a Pearson correlation coefficient.

C. Apply a Spearman correlation coefficient.
D. Apply a linear discriminant analysis.

Correct Answer: D
Section:
Explanation:
The linear discriminant analysis method works only on continuous variables, not categorical or ordinal variables.
Linear discriminant analysis is similar to analysis of variance (ANOVA) in that it works by comparing the means of the variables.
Scenario:
Data scientists must build notebooks in a local environment using automatic feature engineering and model building in machine learning pipelines. Experiments for local crowd sentiment models must combine local penalty detection data. All shared features for local models are continuous variables.
Incorrect Answers:
B: The Pearson correlation coefficient, sometimes called Pearson's R test, is a statistical value that measures the linear relationship between two variables. By examining the coefficient values, you can infer something about the strength of the relationship between the two variables, and whether they are positively correlated or negatively correlated.
C: Spearman's correlation coefficient is designed for use with non-parametric and non-normally distributed data. Spearman's coefficient is a nonparametric measure of statistical dependence between two variables, and is sometimes denoted by the Greek letter rho. The Spearman's coefficient expresses the degree to which two variables are monotonically related. It is also called Spearman rank correlation, because it can be used with ordinal variables.
Reference:
https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/fisher-linear-discriminant-analysis https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/compute-linear-correlation

QUESTION 3
You need to implement a model development strategy to determine a user's tendency to respond to an ad.
Which technique should you use?

A. Use a Relative Expression Split module to partition the data based on centroid distance.
B. Use a Relative Expression Split module to partition the data based on distance travelled to the event.
C. Use a Split Rows module to partition the data based on distance travelled to the event.
D. Use a Split Rows module to partition the data based on centroid distance.

Correct Answer: A
Section:
Explanation:
Split Data partitions the rows of a dataset into two distinct sets.
The Relative Expression Split option in the Split Data module of Azure Machine Learning Studio is helpful when you need to divide a dataset into training and testing datasets using a numerical expression.
Relative Expression Split: Use this option whenever you want to apply a condition to a number column. The number could be a date/time field, a column containing age or dollar amounts, or even a percentage. For example, you might want to divide your data set depending on the cost of the items, group people by age ranges, or separate data by a calendar date.
Scenario:
Local market segmentation models will be applied before determining a user's propensity to respond to an advertisement. The distribution of features across training and production data are not consistent
Reference: https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/split-data

QUESTION 4
You need to implement a new cost factor scenario for the ad response models as illustrated in the performance curve exhibit.
Which technique should you use?

A. Set the threshold to 0.5 and retrain if weighted Kappa deviates +/- 5% from 0.45.
B. Set the threshold to 0.05 and retrain if weighted Kappa deviates +/- 5% from 0.5.
C. Set the threshold to 0.2 and retrain if weighted Kappa deviates +/- 5% from 0.6.
D. Set the threshold to 0.75 and retrain if weighted Kappa deviates +/- 5% from 0.15.

Correct Answer: A
Section:
Explanation:
Scenario:
Performance curves of current and proposed cost factor scenarios are shown in the following diagram:

The ad propensity model uses a cut threshold is 0.45 and retrains occur if weighted Kappa deviated from 0.1 +/- 5%.

QUESTION 5
HOTSPOT
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.

Hot Area:

Answer Area

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

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

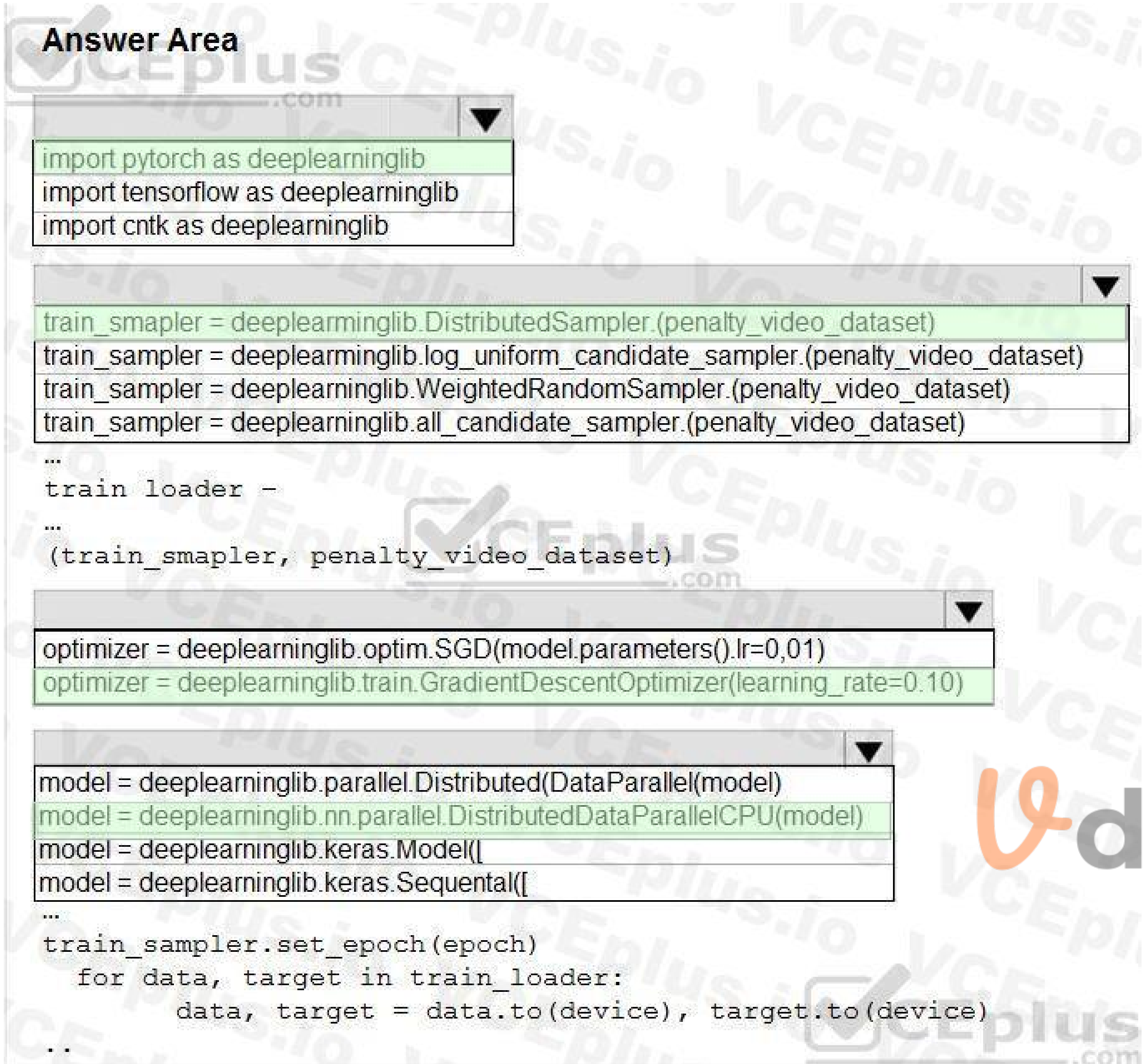
```
...
train loader –
...
(train_smapler, penalty_video_dataset)
```

- 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 Area:


Answer Area

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

- train_smapler = deeplearninglib.DistributedSampler.(penalty_video_dataset)
- train_sampler = deeplearninglib.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)

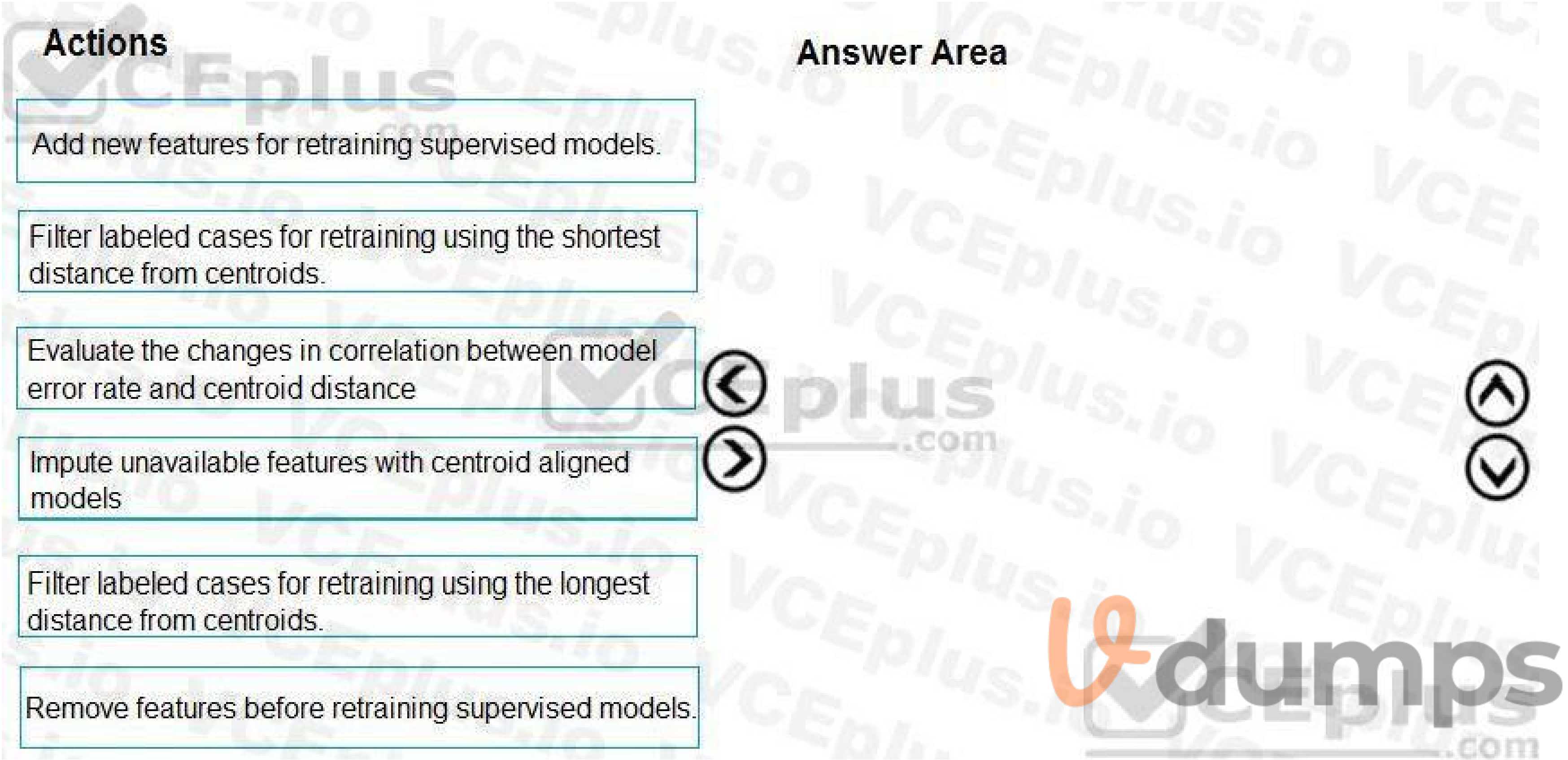
- model = deeplearninglib.parallel.Distributed(DataParallel(model)
- model = deeplearninglib.nn.parallel.DistributedDataParallelCPU(model)
- model = deeplearninglib.keras.Model([
- model = deeplearninglib.keras.Sequental([

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

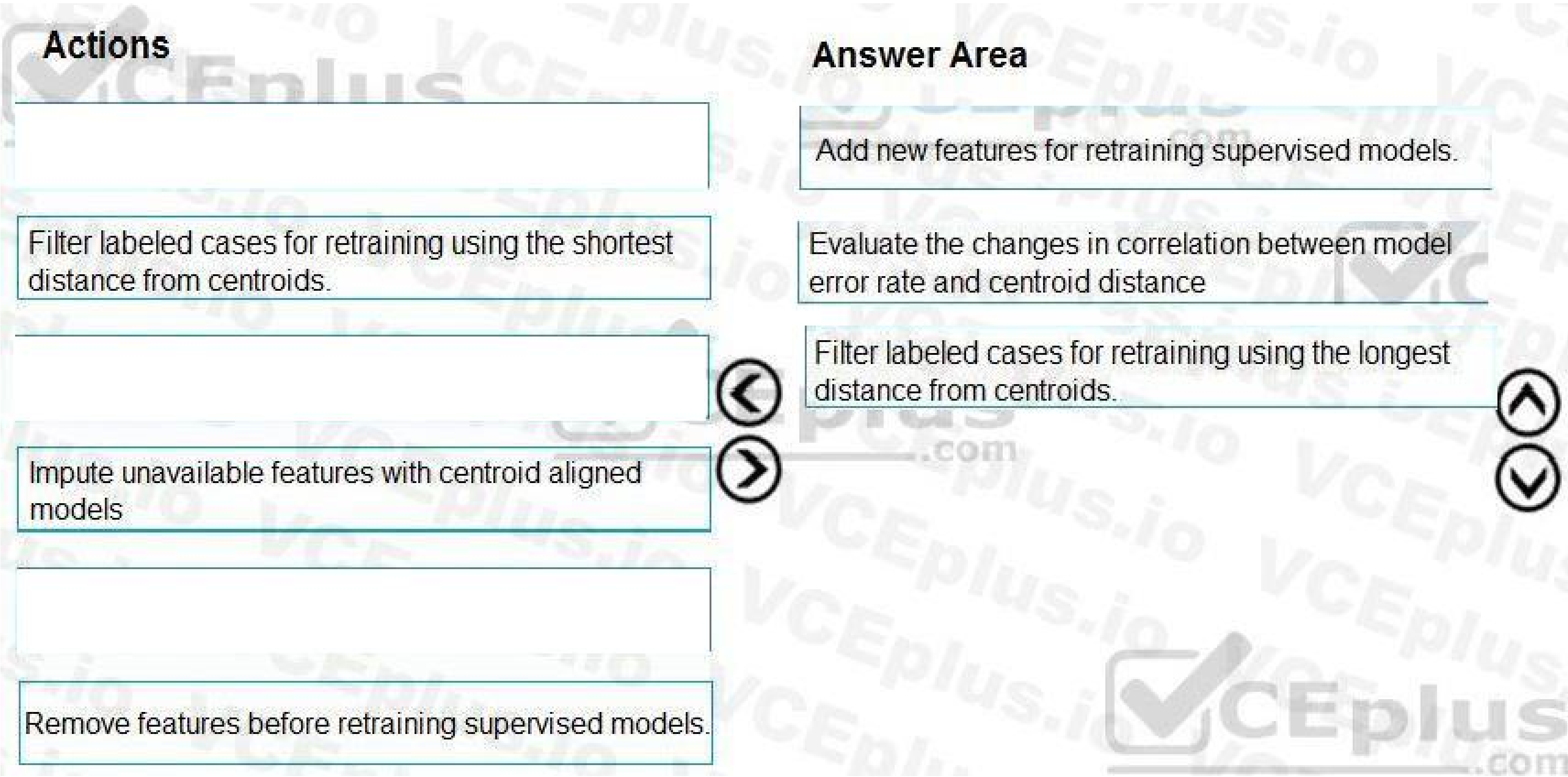
Section:
Explanation:
Box 1: import pytorch as deeplearninglib
Box 2: ..DistributedSampler(Sampler)..
DistributedSampler(Sampler):
Sampler that restricts data loading to a subset of the dataset.
It is especially useful in conjunction with class:`torch.nn.parallel.DistributedDataParallel`. In such case, 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

QUESTION 6
DRAG DROP
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.

Select and Place:



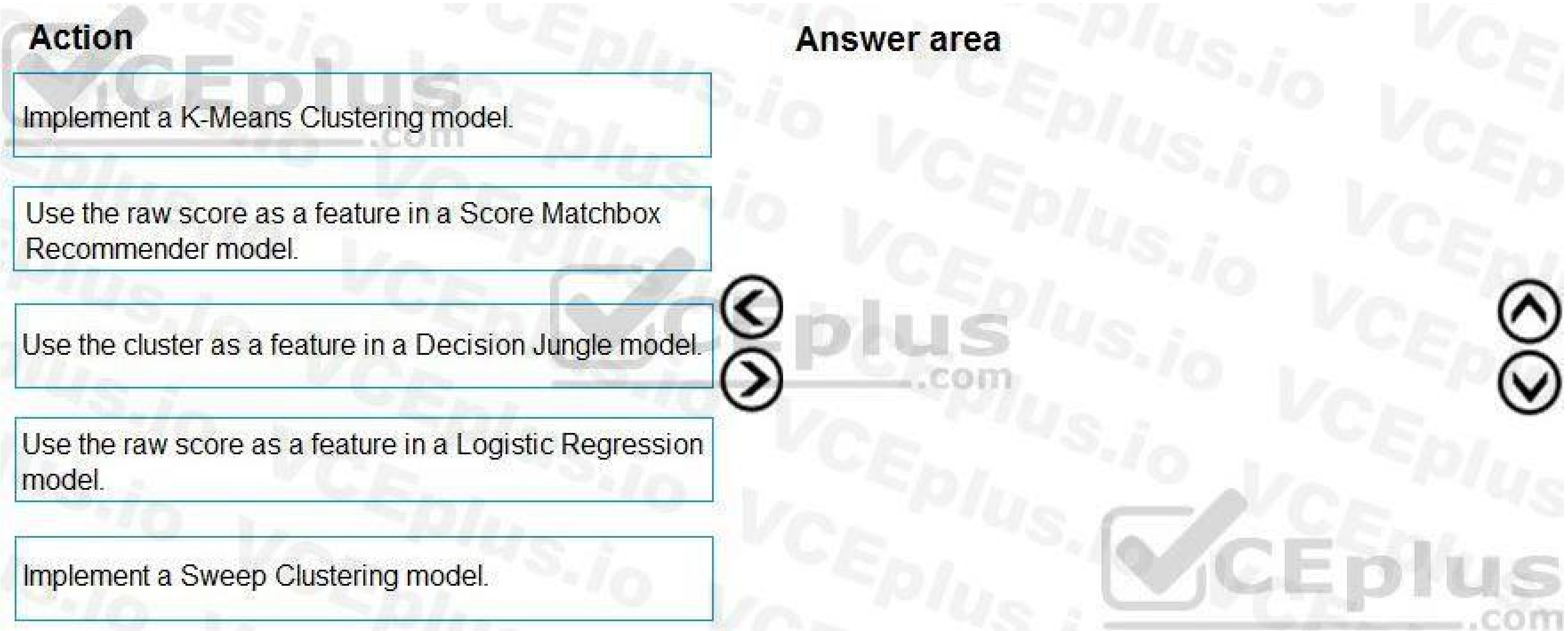
Correct Answer:



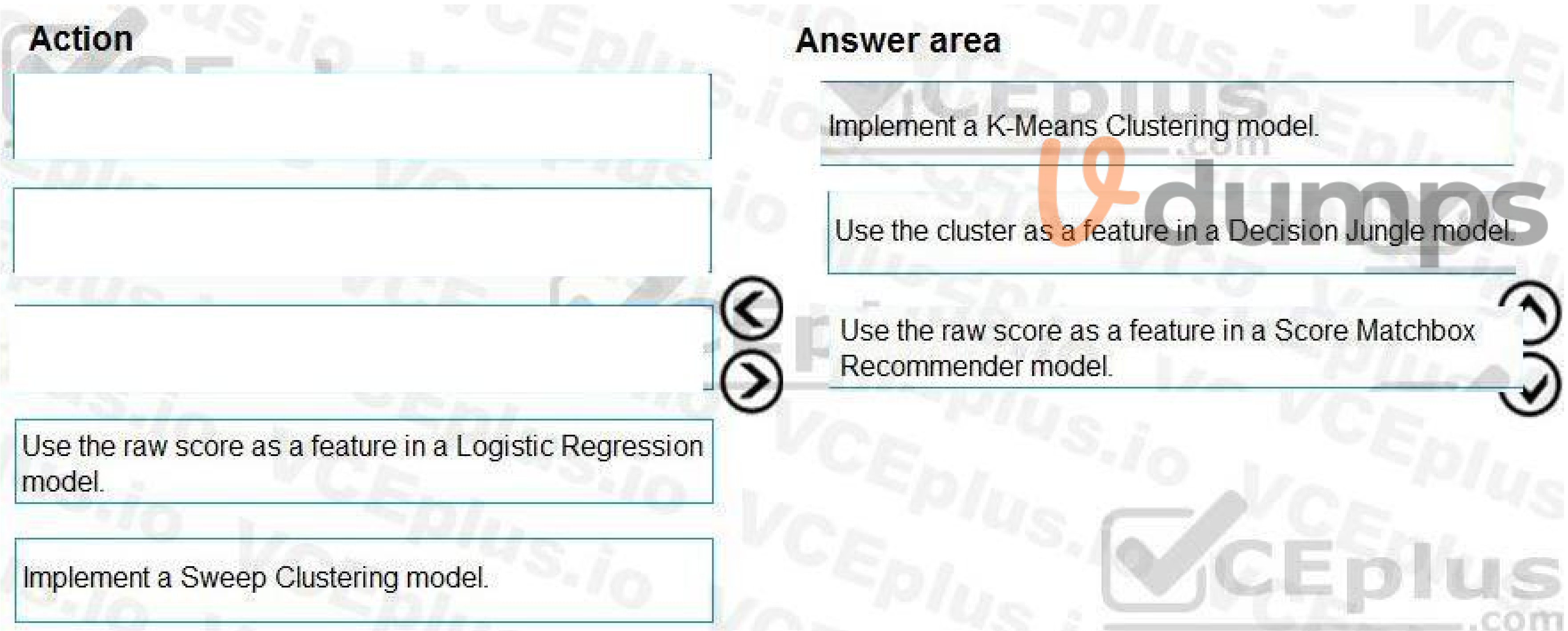
Section:
Explanation:
Scenario:
Experiments for local crowd sentiment models must combine local penalty detection data.
Crowd sentiment models must identify known sounds such as cheers and known catch phrases. Individual crowd sentiment models will detect similar sounds.
Note: Evaluate the changed in correlation between model error rate and centroid distance
In machine learning, a nearest centroid classifier or nearest prototype classifier is a classification model that assigns to observations the label of the class of training samples whose mean (centroid) is closest to the observation.
References:
https://en.wikipedia.org/wiki/Nearest_centroid_classifier
https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/sweep-clustering

QUESTION 7
DRAG DROP
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.

Select and Place:



Correct Answer:

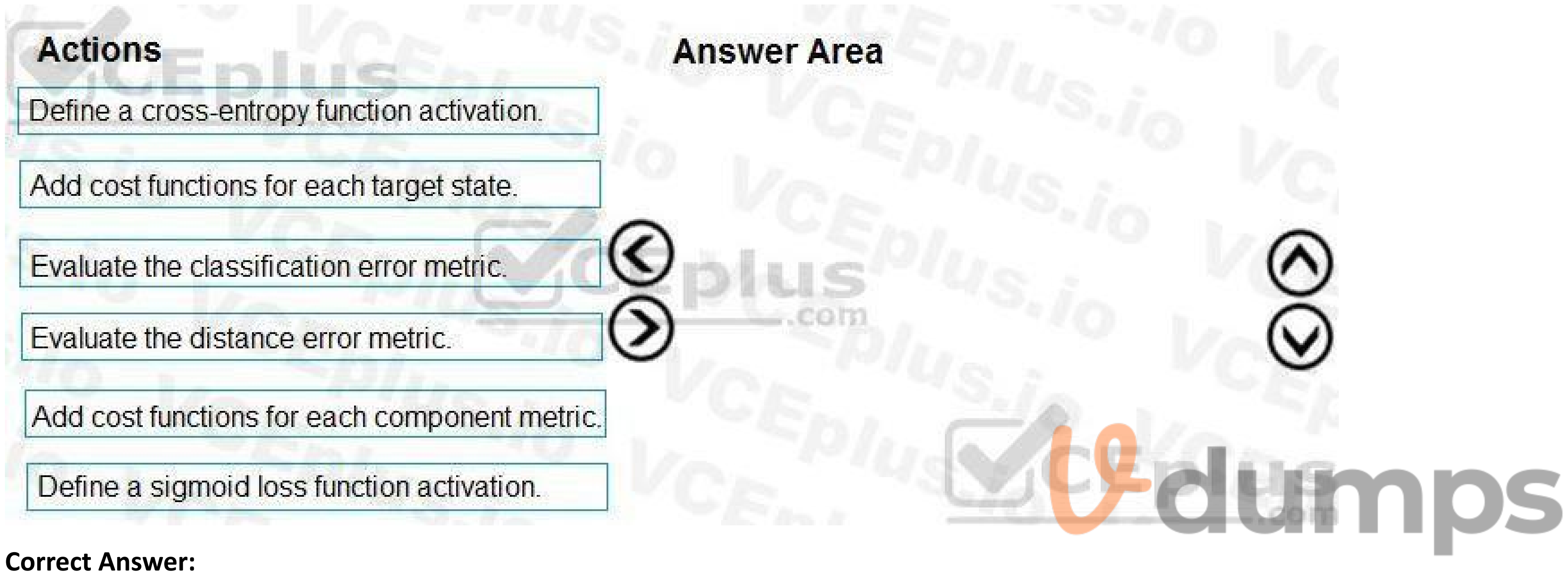


Section:
Explanation:
Step 1: Implement a K-Means Clustering model
Step 2: Use the cluster as a feature in a Decision jungle model.
Decision jungles are non-parametric models, which can represent non-linear decision boundaries.
Step 3: Use the raw score as a feature in a Score Matchbox Recommender model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

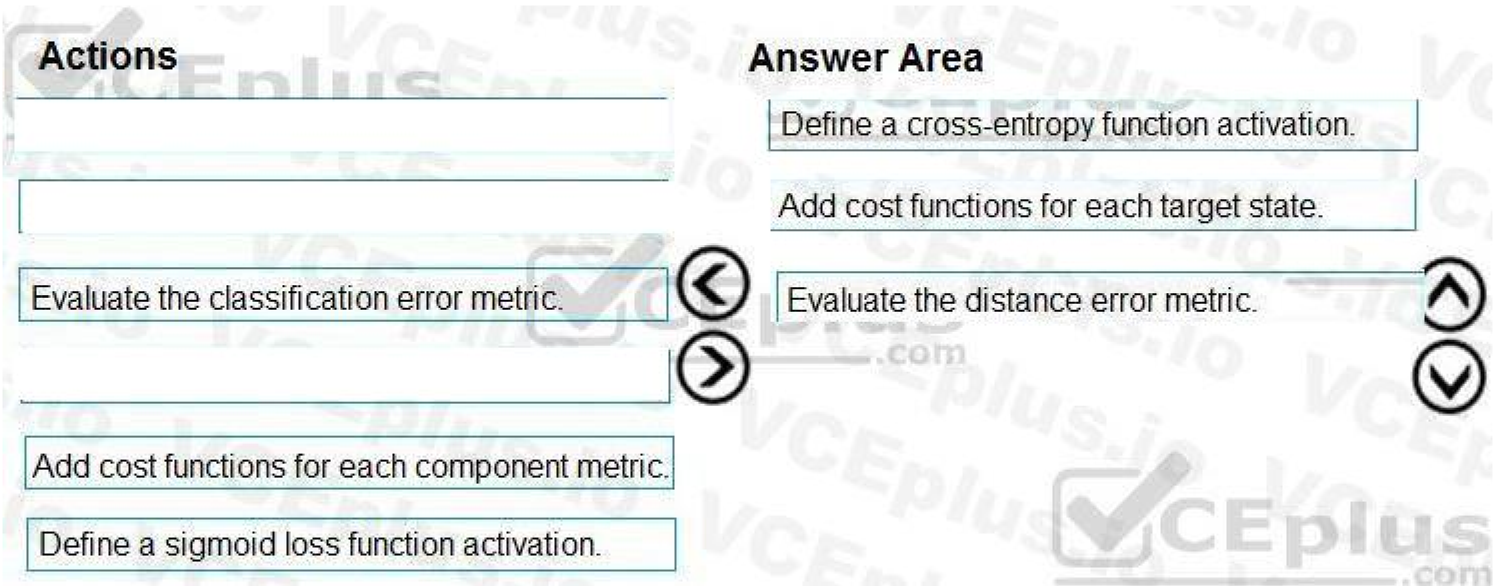
Ad response models must support non-linear boundaries of features.
References:
https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/multiclass-decision-jungle
https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/score-matchbox-recommender

QUESTION 8
DRAG DROP
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.

Select and Place:



Correct Answer:



Section:
Explanation:
Step 1: Define a cross-entropy function activation
When using a neural network to perform classification and prediction, it is usually better to use cross-entropy error than classification error, and somewhat better to use cross-entropy error than mean squared error to evaluate the quality of the neural network.
Step 2: Add cost functions for each target state.
Step 3: Evaluated the distance error metric.

References:
https://www.analyticsvidhya.com/blog/2018/04/fundamentals-deep-learning-regularization-techniques/

03 - Run experiments and train models

Case study

This is a case study. Case studies are not timed separately. You can use as much exam time as you would like to complete each case. However, there may be additional case studies and sections on this exam. You must manage your time to ensure that you are able to complete all questions included on this exam in the time provided.

To answer the questions included in a case study, you will need to reference information that is provided in the case study. Case studies might contain exhibits and other resources that provide more information about the scenario that is described in the case study. Each question is independent of the other questions in this case study.

At the end of this case study, a review screen will appear. This screen allows you to review your answers and to make changes before you move to the next section of the exam. After you begin a new section, you cannot return to this section.

To start the case study To display the first question in this case study, click the Next button. Use the buttons in the left pane to explore the content of the case study before you answer the questions. Clicking these buttons displays information such as business requirements, existing environment, and problem statements. If the case study has an All Information tab, note that the information displayed is identical to the information displayed on the subsequent tabs. When you are ready to answer a question, click the Question button to return to the question.

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. You add both files to Azure Machine Learning Studio as separate datasets to the starting point for an experiment. Both datasets contain 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 a 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

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.

Data issues

Missing values

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 datasets also contain 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. 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-parametric statistics to measure relationships.

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

Model training
Permutation Feature Importance
Given a trained model and a test dataset, you must compute the Permutation Feature Importance scores of feature variables. You must be determined the absolute fit for the model.
Hyperparameters
You must configure hyperparameters in the model learning process to speed the learning phase. 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, must 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.
Cross-validation
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. You must complete this task before the data goes through the sampling process.
Linear regression module
When you train a Linear Regression module, you must 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. The distribution of features across multiple training models must be 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.

QUESTION 1
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? Each correct answer presents part of the solution.
NOTE: Each correct selection is worth one point.

A. Create Scatterplot
B. Summarize Data
C. Clip Values
D. Replace Discrete Values
E. Build Counting Transform

Correct Answer: A, B, C
Section:
Explanation:
B: To have a global view, the summarize data module can be used. Add the module and connect it to the data set that needs to be visualized. A: One way to quickly identify Outliers visually is to create scatter plots.
C: The easiest way to treat the outliers in Azure ML is to use the Clip Values module. It can identify and optionally replace data values that are above or below a specified threshold.
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.
Reference:
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

QUESTION 2
You need to select a feature extraction method.
Which method should you use?

A. Mutual information
B. Pearson's correlation
C. Spearman correlation

D. Fisher Linear Discriminant Analysis

Correct Answer: C
Section:
Explanation:
Spearman's rank correlation coefficient assesses how well the relationship between two variables can be described using a monotonic function.
Note: Both Spearman's and Kendall's can be formulated as special cases of a more general correlation coefficient, and they are both appropriate in this scenario.
Scenario: The MedianValue and AvgRoomsInHouse columns both hold data in numeric format. You need to select a feature selection algorithm to analyze the relationship between the two columns in more detail.
Incorrect Answers:
B: The Spearman correlation between two variables is equal to the Pearson correlation between the rank values of those two variables; while Pearson's correlation assesses linear relationships, Spearman's correlation assesses monotonic relationships (whether linear or not).
Reference: https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/feature-selection-modules

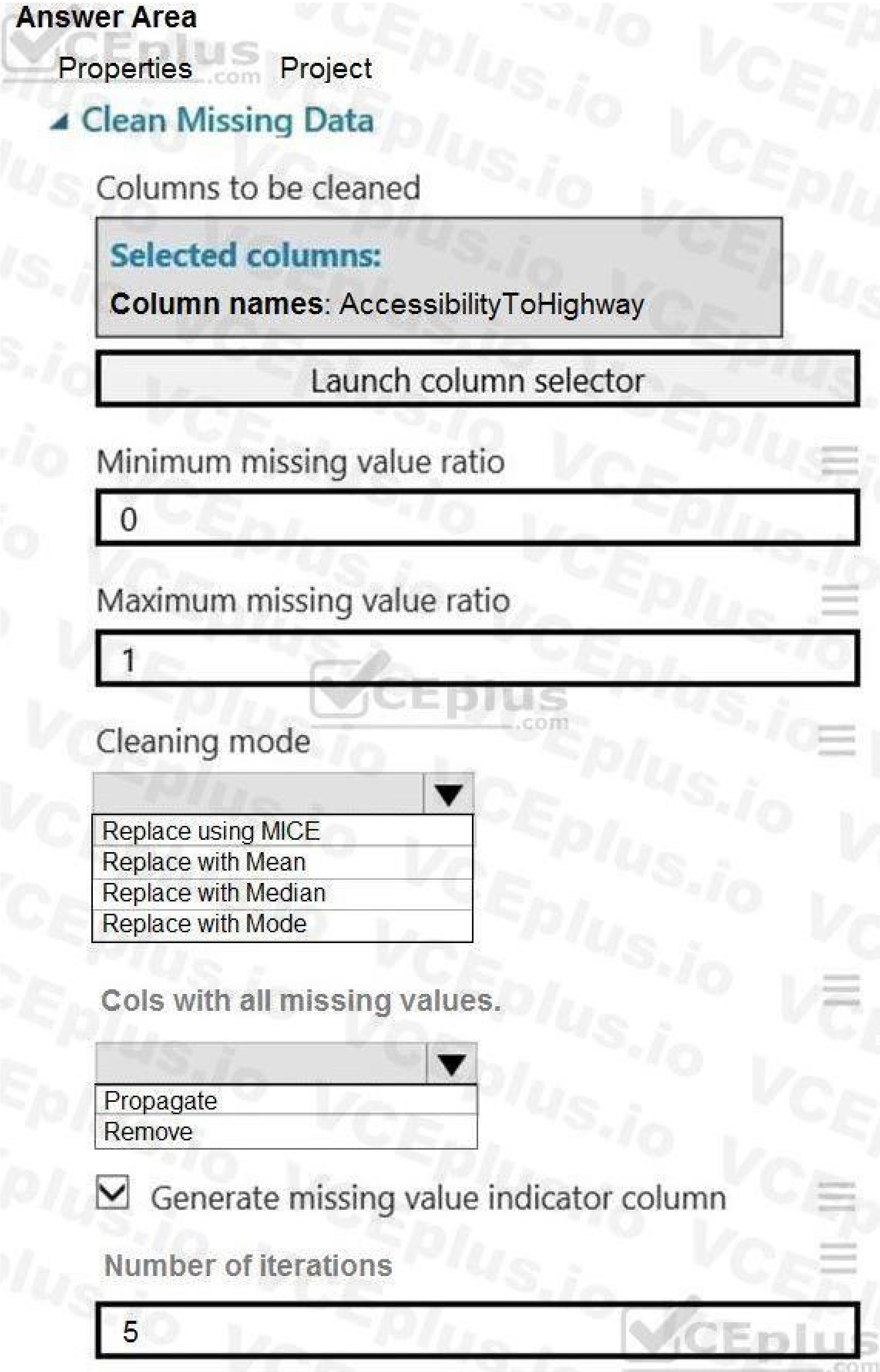
QUESTION 3
You need to select a feature extraction method.
Which method should you use?

A. Mutual information
B. Mood's median test
C. Kendall correlation
D. Permutation Feature Importance

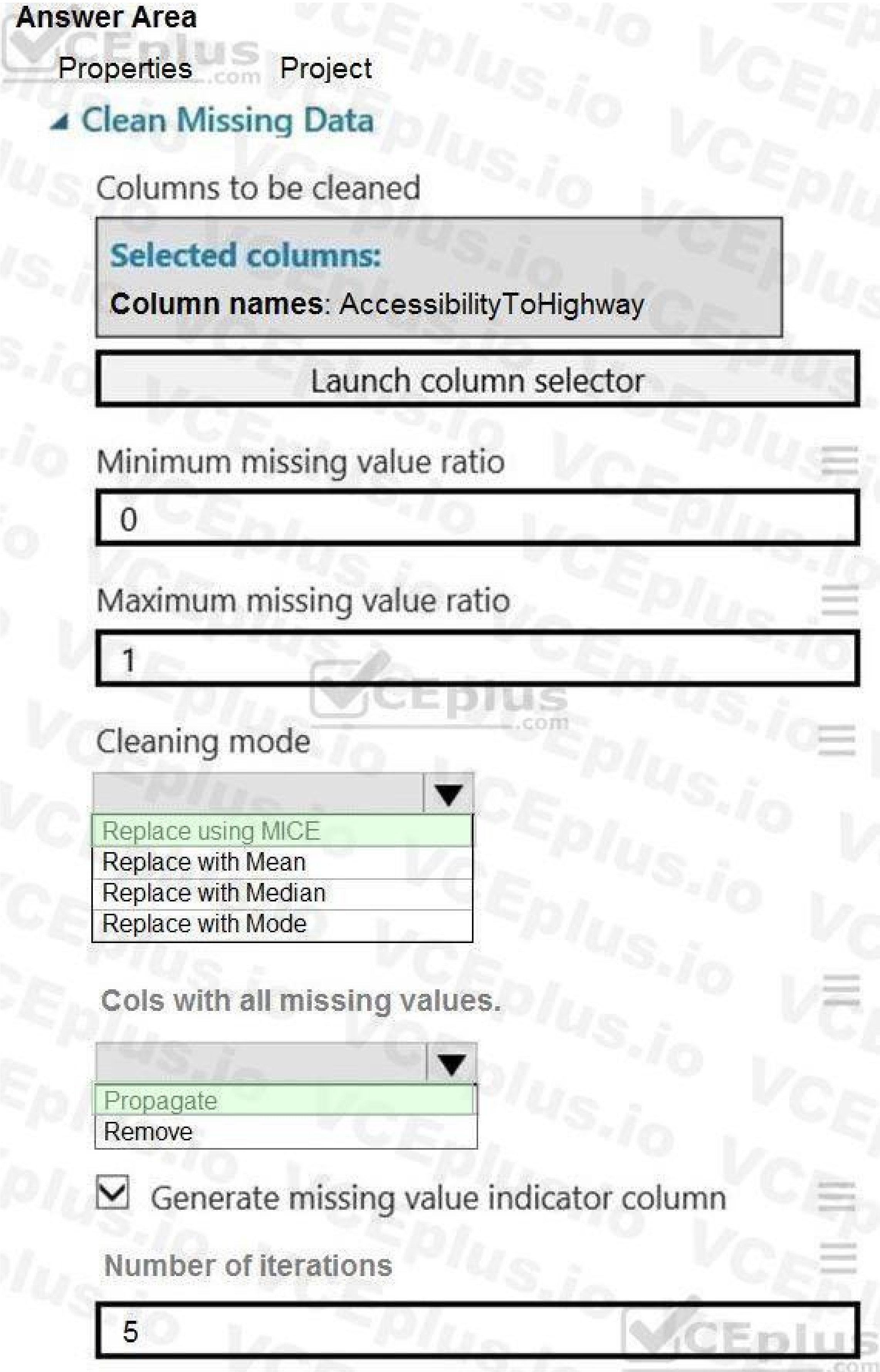
Correct Answer: C
Section:
Explanation:
In statistics, the Kendall rank correlation coefficient, commonly referred to as Kendall's tau coefficient (after the Greek letter Ď„), is a statistic used to measure the ordinal association between two measured quantities. It is a supported method of the Azure Machine Learning Feature selection.
Note: Both Spearman's and Kendall's can be formulated as special cases of a more general correlation coefficient, and they are both appropriate in this scenario.
Scenario: The MedianValue and AvgRoomsInHouse columns both hold data in numeric format. You need to select a feature selection algorithm to analyze the relationship between the two columns in more detail.
Reference:
https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/feature-selection-modules

QUESTION 4
HOTSPOT
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.

Hot Area:


Answer Area
Properties
Project
Clean Missing Data
Columns to be cleaned
Selected columns:
Column names: AccessibilityToHighway
Launch column selector
Minimum missing value ratio
0
Maximum missing value ratio
1
Cleaning mode
Replace using MICE
Replace with Mean
Replace with Median
Replace with Mode
Cols with all missing values.
Propagate
Remove
Generate missing value indicator column
Number of iterations
5

Answer Area:



Section:

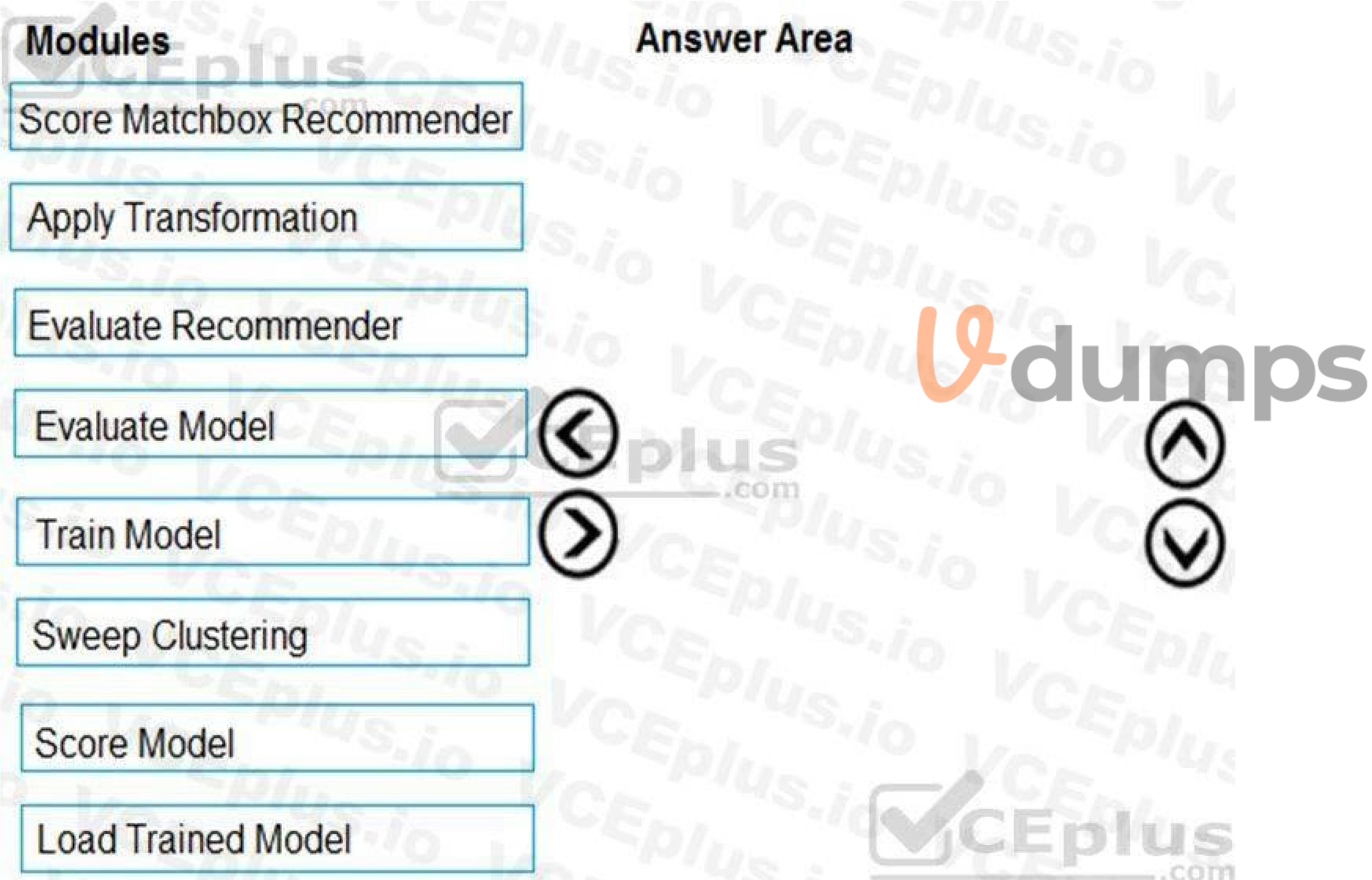
Explanation:

Box 1: Replace using MICE Replace using MICE: For each missing value, this option assigns a new value, which is calculated by using a method described in the statistical literature as "Multivariate Imputation using Chained Equations" or "Multiple Imputation by Chained Equations". With a multiple imputation method, each variable with missing data is modeled conditionally using the other variables in the data before filling in the missing values.

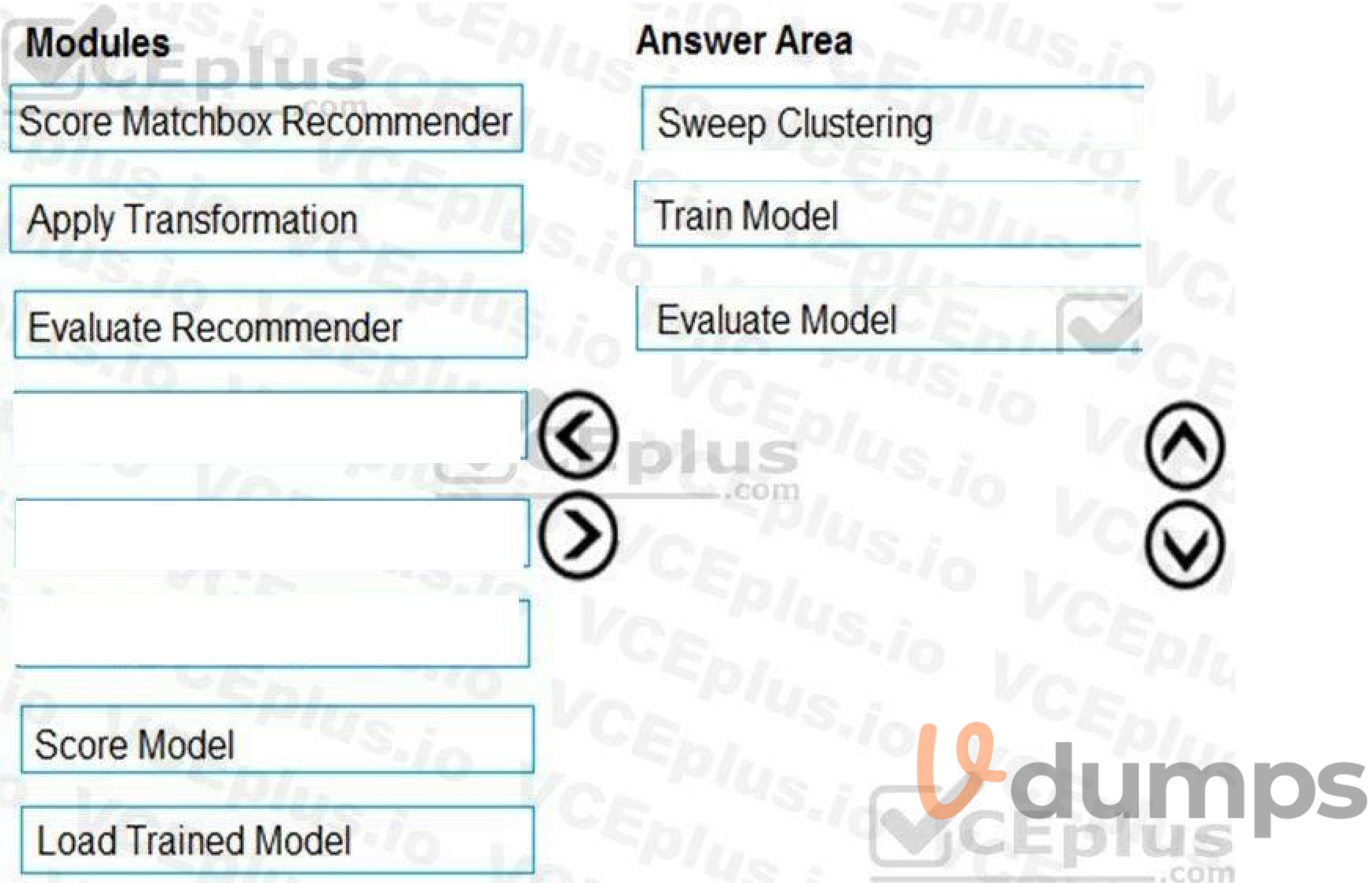
Scenario: The AccessibilityToHighway column in both datasets contains missing values. The missing data must be replaced with new data so that it is modeled conditionally using the other variables in the data before filling in the missing values.
Box 2: Propagate
Cols with all missing values indicate if columns of all missing values should be preserved in the output.
References:
https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/clean-missing-data

QUESTION 5
DRAG DROP
You need to produce a visualization for the diagnostic test evaluation according to the data visualization requirements.
Which three modules should you recommend be used in sequence? To answer, move the appropriate modules from the list of modules to the answer area and arrange them in the correct order.

Select and Place:



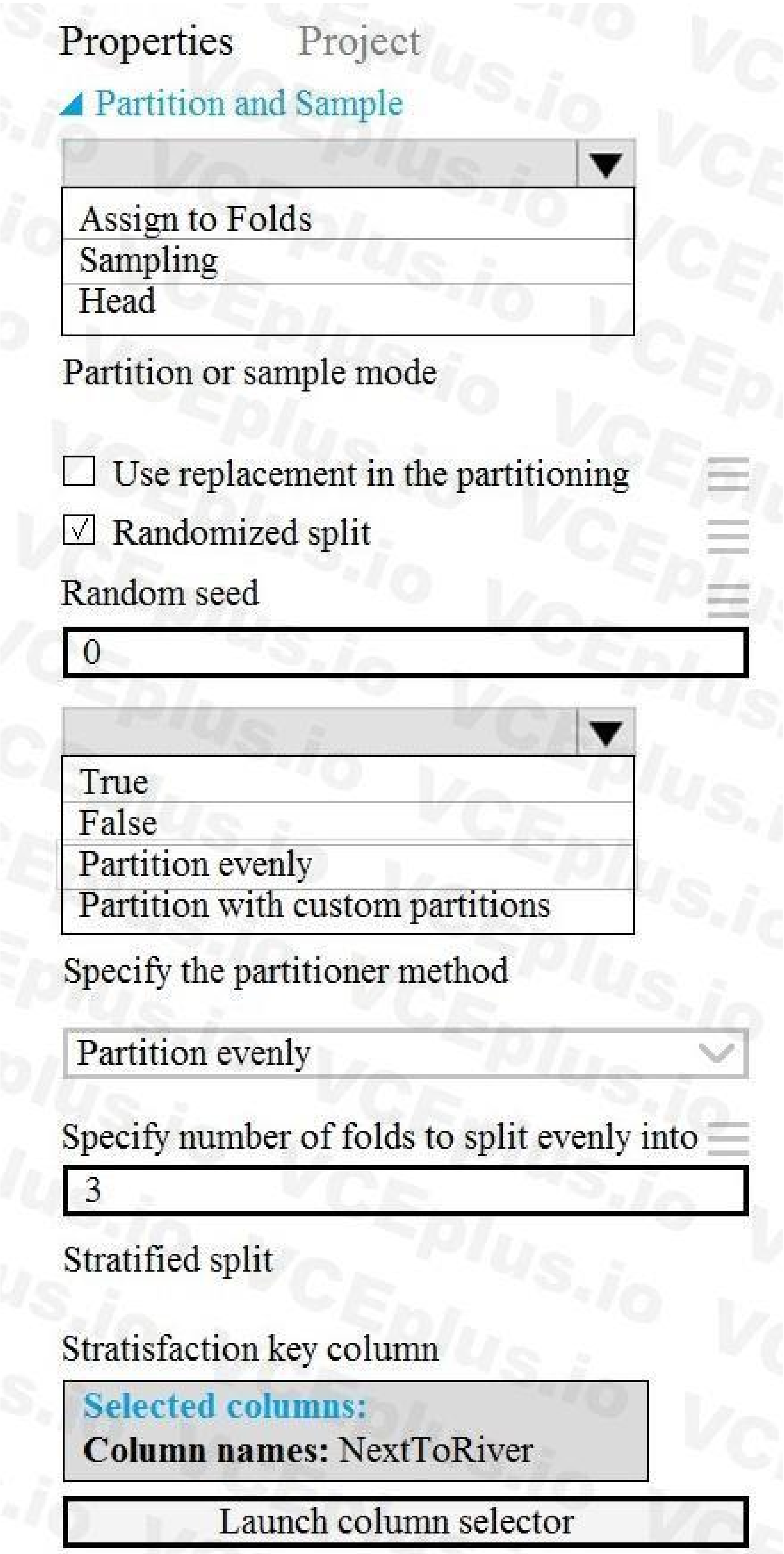
Correct Answer:



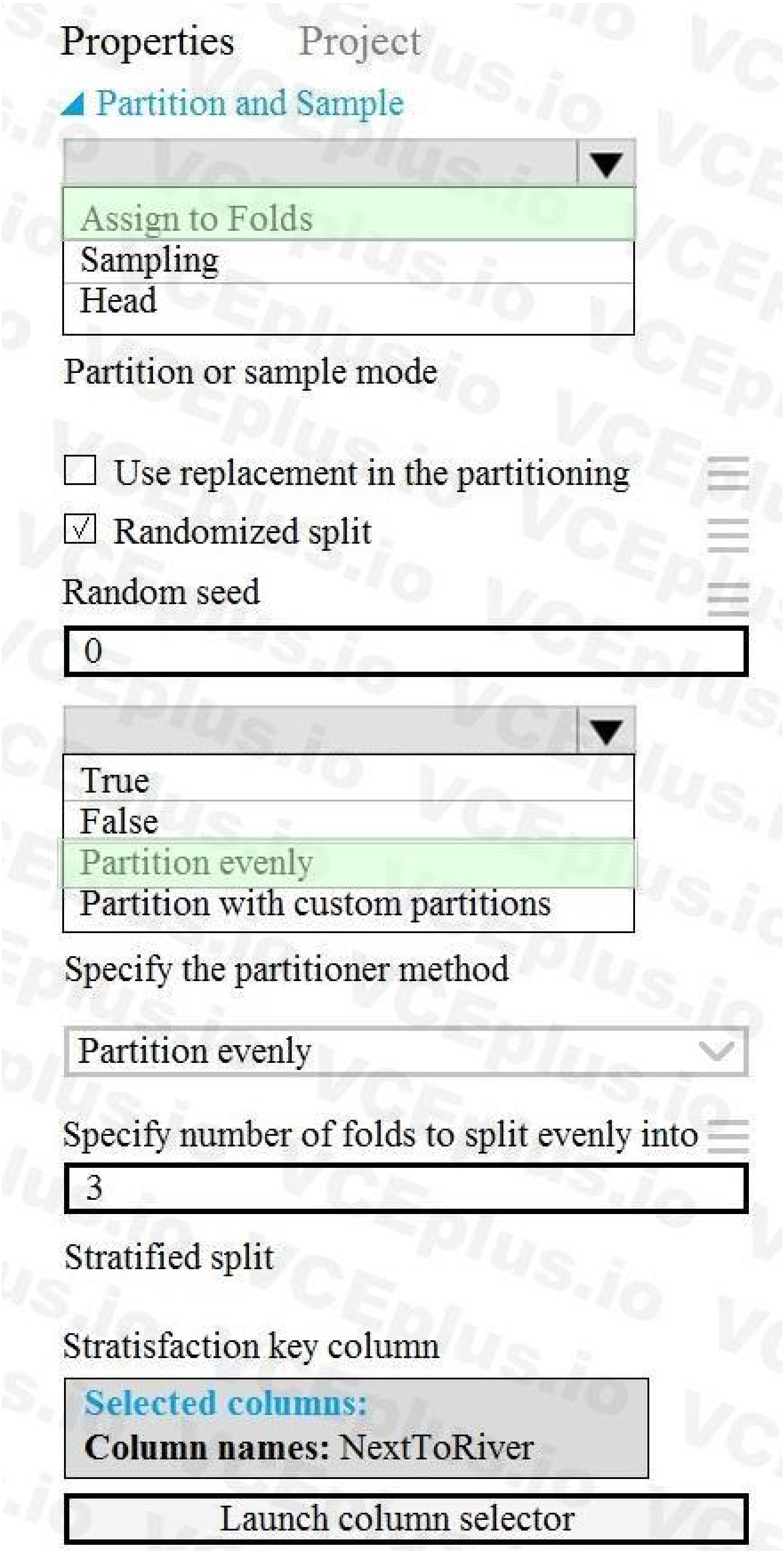
Section:
Explanation:
Step 1: Sweep Clustering
Start by using the "Tune Model Hyperparameters" module to select the best sets of parameters for each of the models we're considering.
One of the interesting things about the "Tune Model Hyperparameters" module is that it not only outputs the results from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

QUESTION 6
HOTSPOT
You need to identify the methods for dividing the data according to the testing requirements.
Which properties should you select? To answer, select the appropriate options in the answer area.
NOTE: Each correct selection is worth one point.

Hot Area:


Properties
Project
Partition and Sample
Assign to Folds
Sampling
Head
Partition or sample mode
Use replacement in the partitioning
Randomized split
Random seed
0
True
False
Partition evenly
Partition with custom partitions
Specify the partitioner method
Partition evenly
Specify number of folds to split evenly into
3
Stratified split
Stratisfaction key column
Selected columns:
Column names: NextToRiver
Launch column selector

Answer Area:



Section:

Explanation:

Scenario: Testing

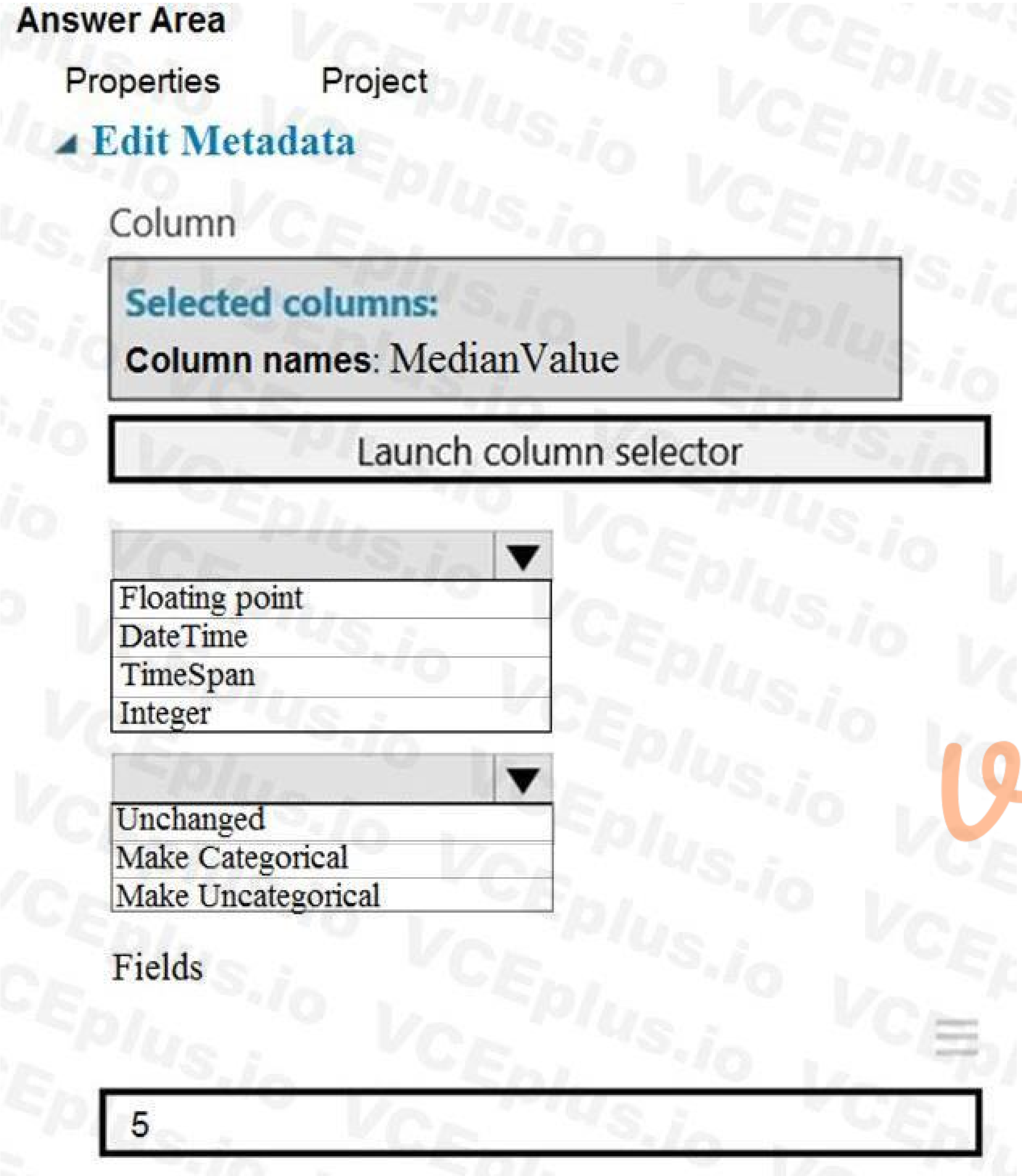
You must produce multiple partitions of a dataset based on sampling using the Partition and Sample module in Azure Machine Learning Studio.

Box 1: Assign to folds

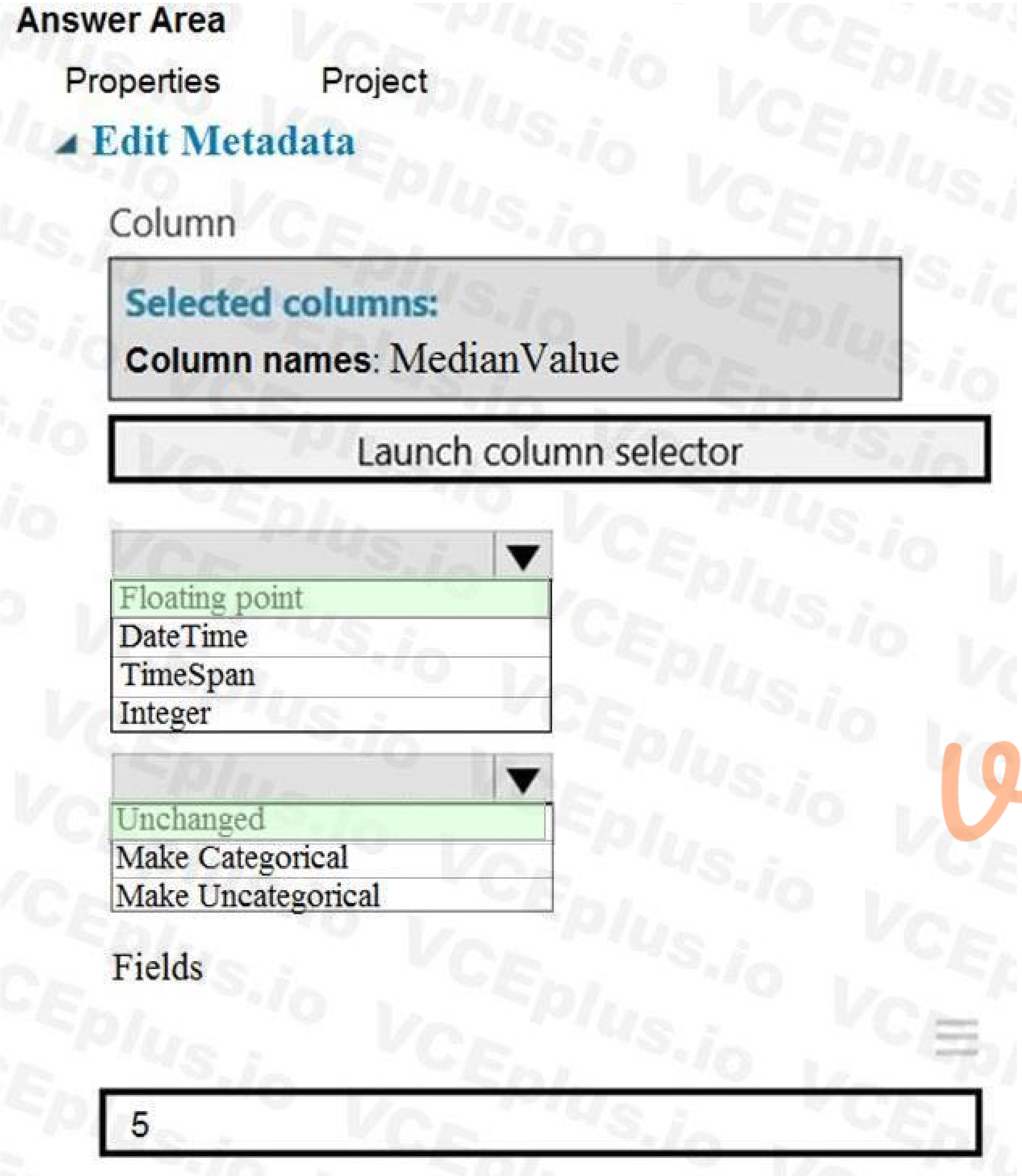
Use Assign to folds option when you want to divide the dataset into subsets of the data. This option is also useful when you want to create a custom number of folds for cross-validation, or to split rows into several groups.
Not Head: Use Head mode to get only the first n rows. This option is useful if you want to test a pipeline on a small number of rows, and don't need the data to be balanced or sampled in any way.
Not Sampling: The Sampling option supports simple random sampling or stratified random sampling. This is useful if you want to create a smaller representative sample dataset for testing.
Box 2: Partition evenly
Specify the partitioner method: Indicate how you want data to be apportioned to each partition, using these options:
Partition evenly: Use this option to place an equal number of rows in each partition. To specify the number of output partitions, type a whole number in the Specify number of folds to split evenly into text box.
Reference:
https://docs.microsoft.com/en-us/azure/machine-learning/algorithm-module-reference/partition-and-sample

QUESTION 7
HOTSPOT
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.

Hot Area:


Answer Area
Properties
Project
Edit Metadata
Column
Selected columns:
Column names: MedianValue
Launch column selector
Floating point
DateTime
TimeSpan
Integer
Unchanged
Make Categorical
Make Uncategorical
Fields
5

Answer Area:

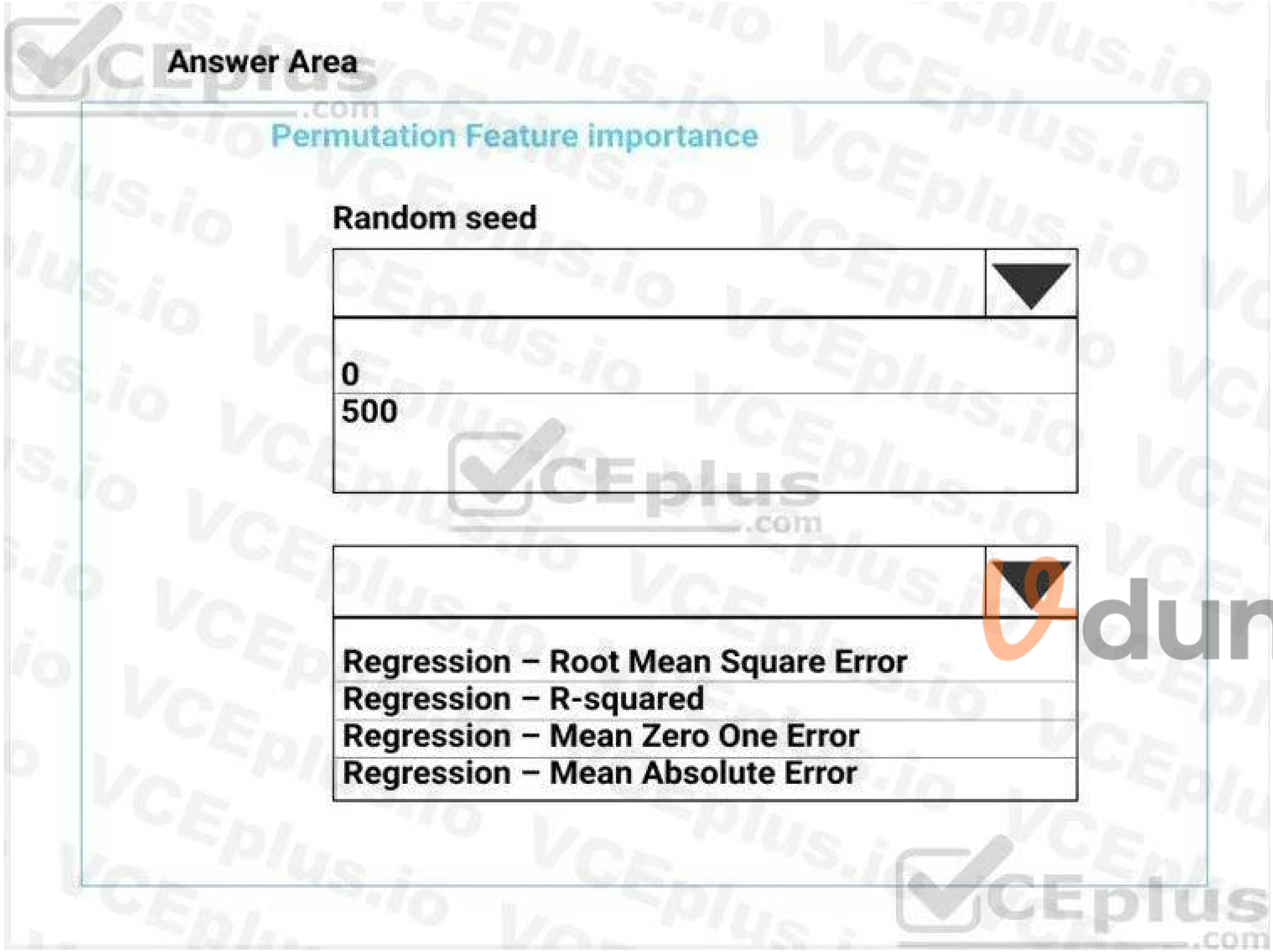


Section:
Explanation:
Box 1: Floating point
Need floating point for Median values.
Scenario: An initial investigation shows that the datasets are identical in structure apart from the MedianValue column. The smaller Paris dataset contains the MedianValue in text format, whereas the larger London dataset 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.

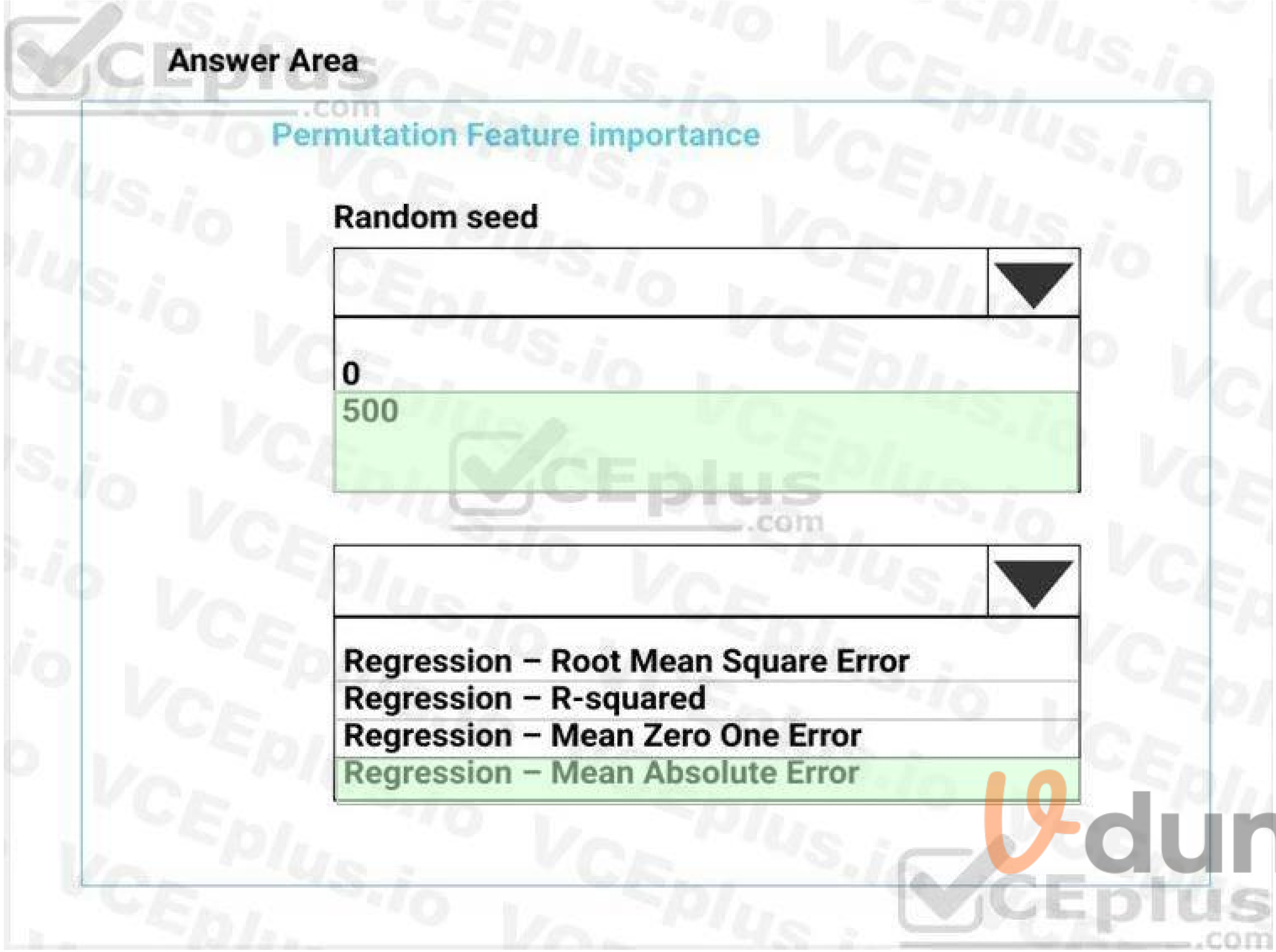
QUESTION 8

HOTSPOT
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.

Hot Area:



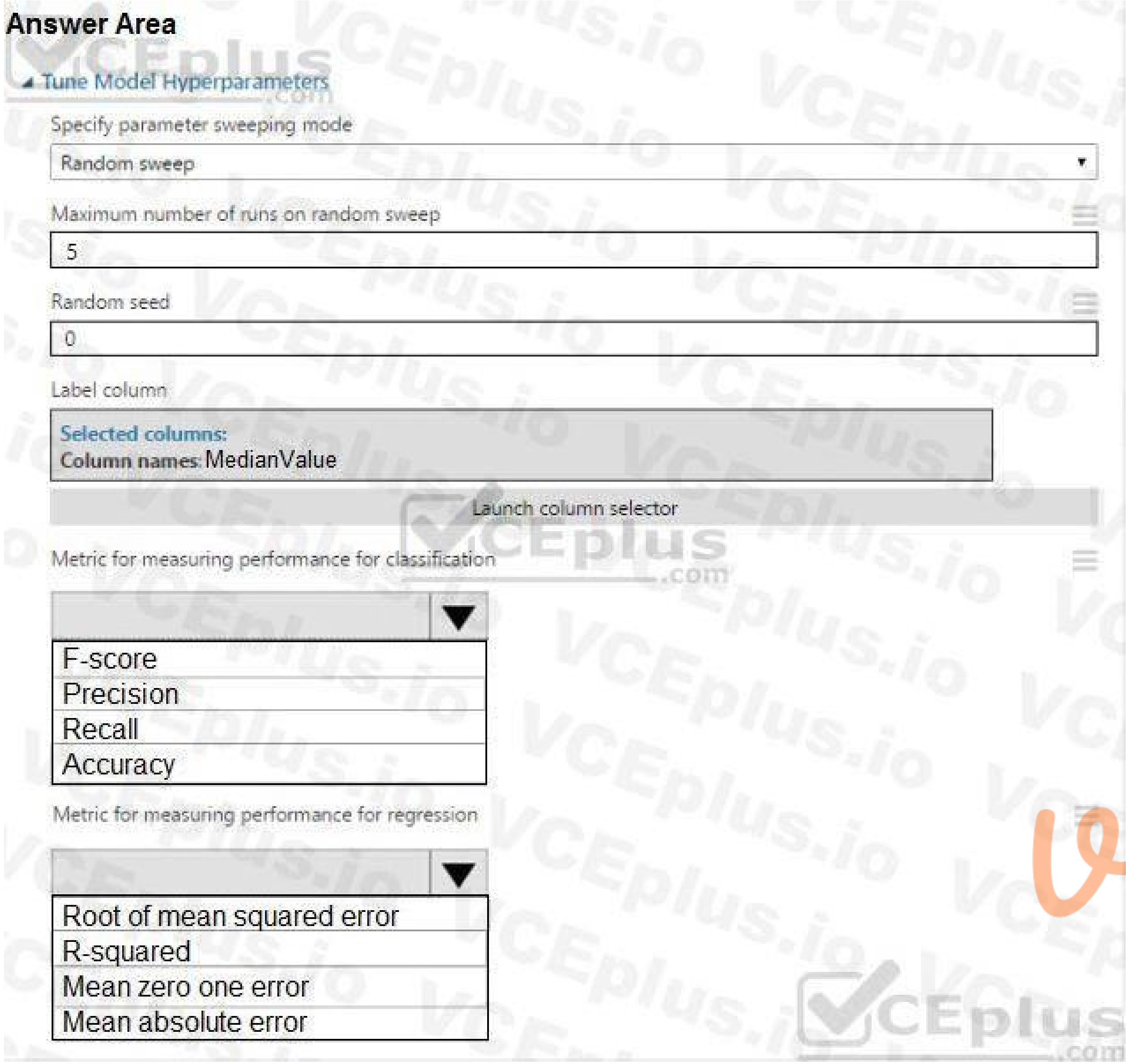
Answer Area:



Section:
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

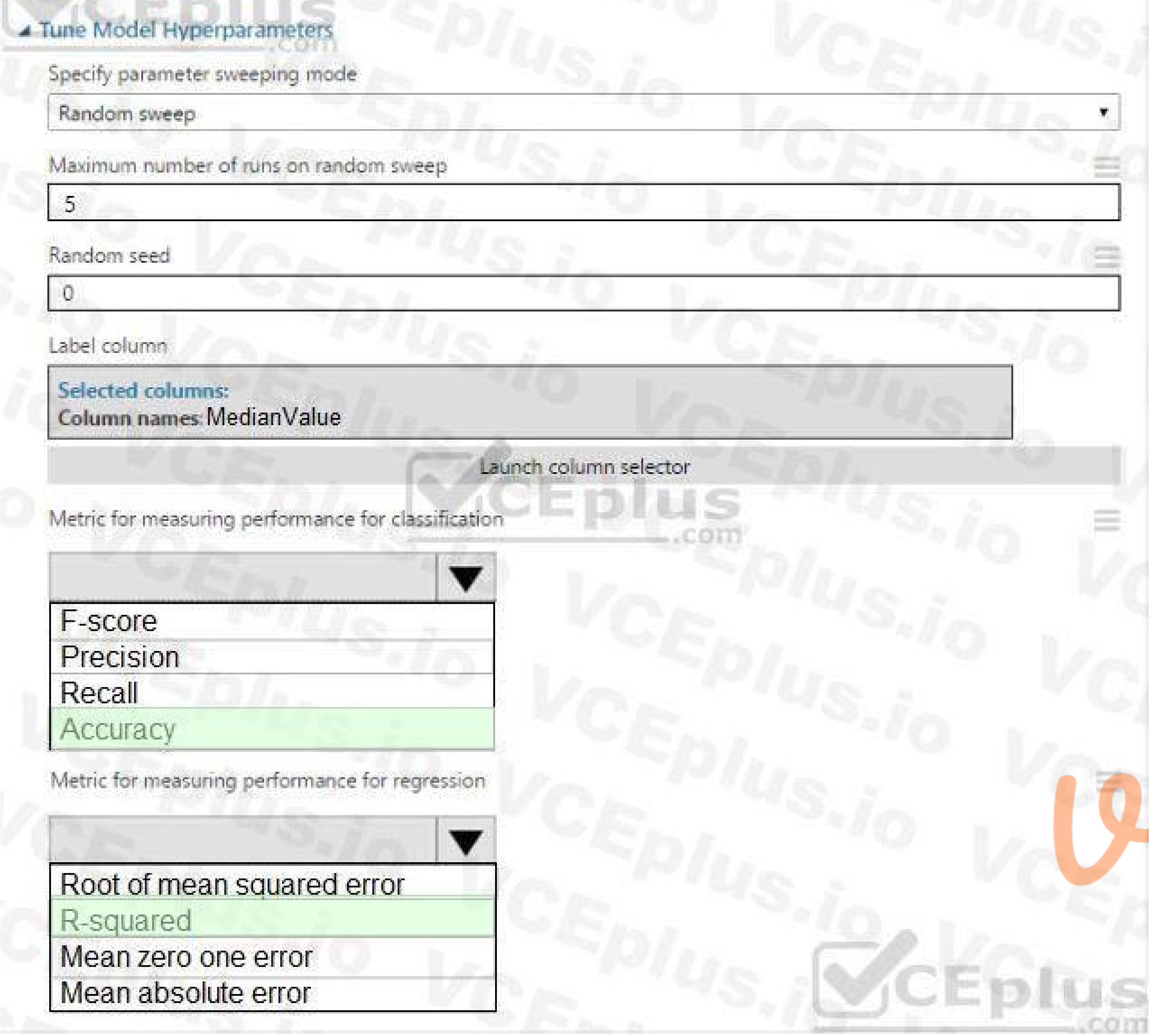
QUESTION 9
HOTSPOT
You need to set up the Permutation Feature Importance module according to the model training requirements.
Which properties should you select? To answer, select the appropriate options in the answer area.
NOTE: Each correct selection is worth one point.

Hot Area:


Answer Area
Tune Model Hyperparameters
Specify parameter sweeping mode
Random sweep
Maximum number of runs on random sweep
5
Random seed
0
Label column
Selected columns:
Column names: MedianValue
Launch column selector
Metric for measuring performance for classification
F-score
Precision
Recall
Accuracy
Metric for measuring performance for regression
Root of mean squared error
R-squared
Mean zero one error
Mean absolute error

Answer Area:

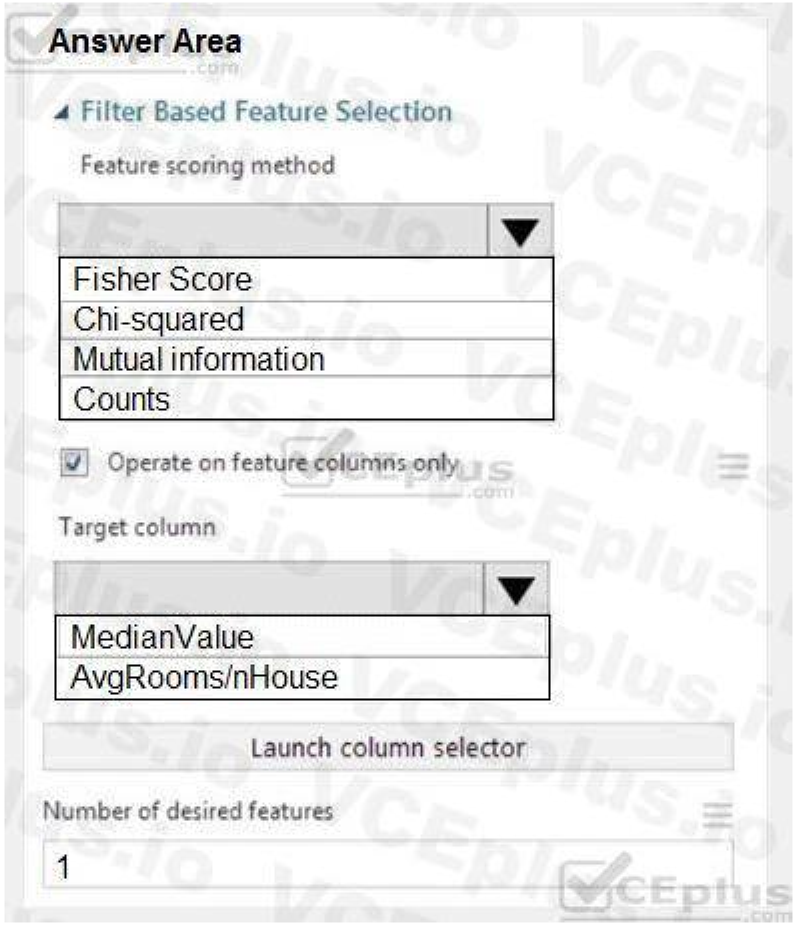
Answer Area



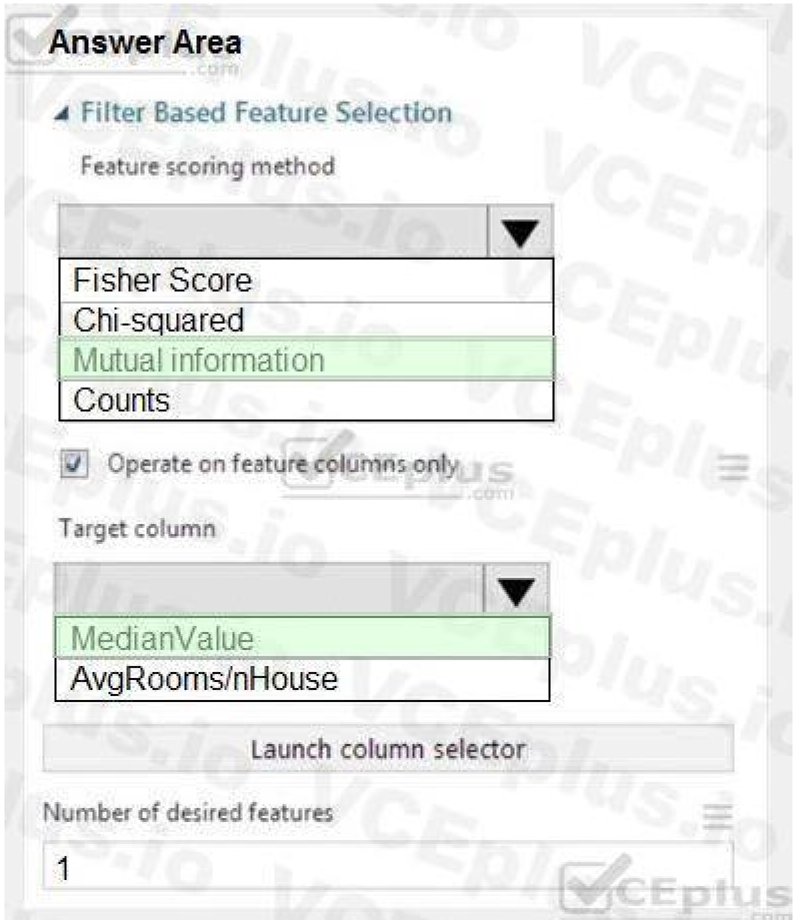
Section:
Explanation:
Box 1: Accuracy
Scenario: You want to configure hyperparameters in the model learning process to speed the learning phase by using hyperparameters. In addition, this configuration should cancel the lowest performing runs at each evaluation interval, thereby directing effort and resources towards models that are more likely to be successful.
Box 2: R-Squared

QUESTION 10
HOTSPOT
You need to configure the Feature Based Feature Selection module based on the experiment requirements and datasets.
How should you configure the module properties? To answer, select the appropriate options in the dialog box in the answer area.
NOTE: Each correct selection is worth one point.

Hot Area:



Answer Area:

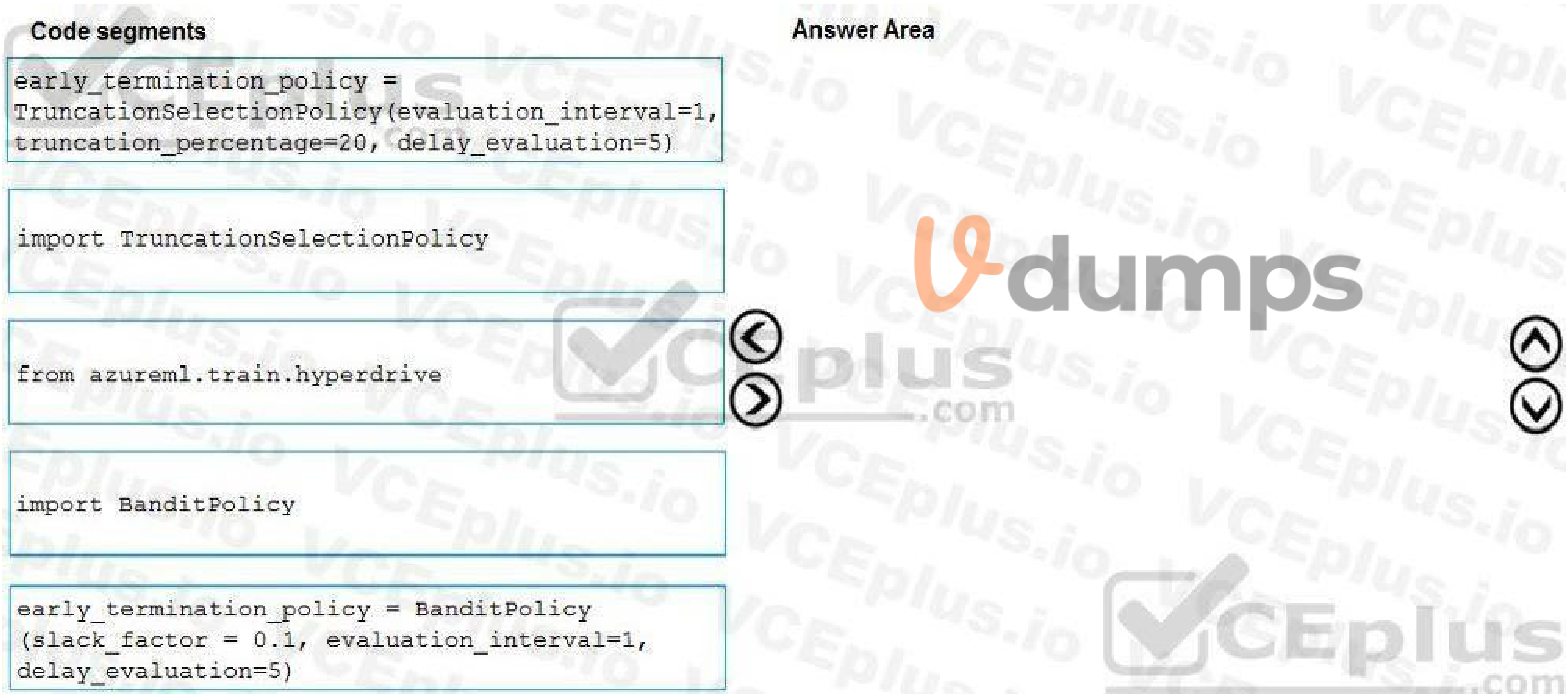


Section:

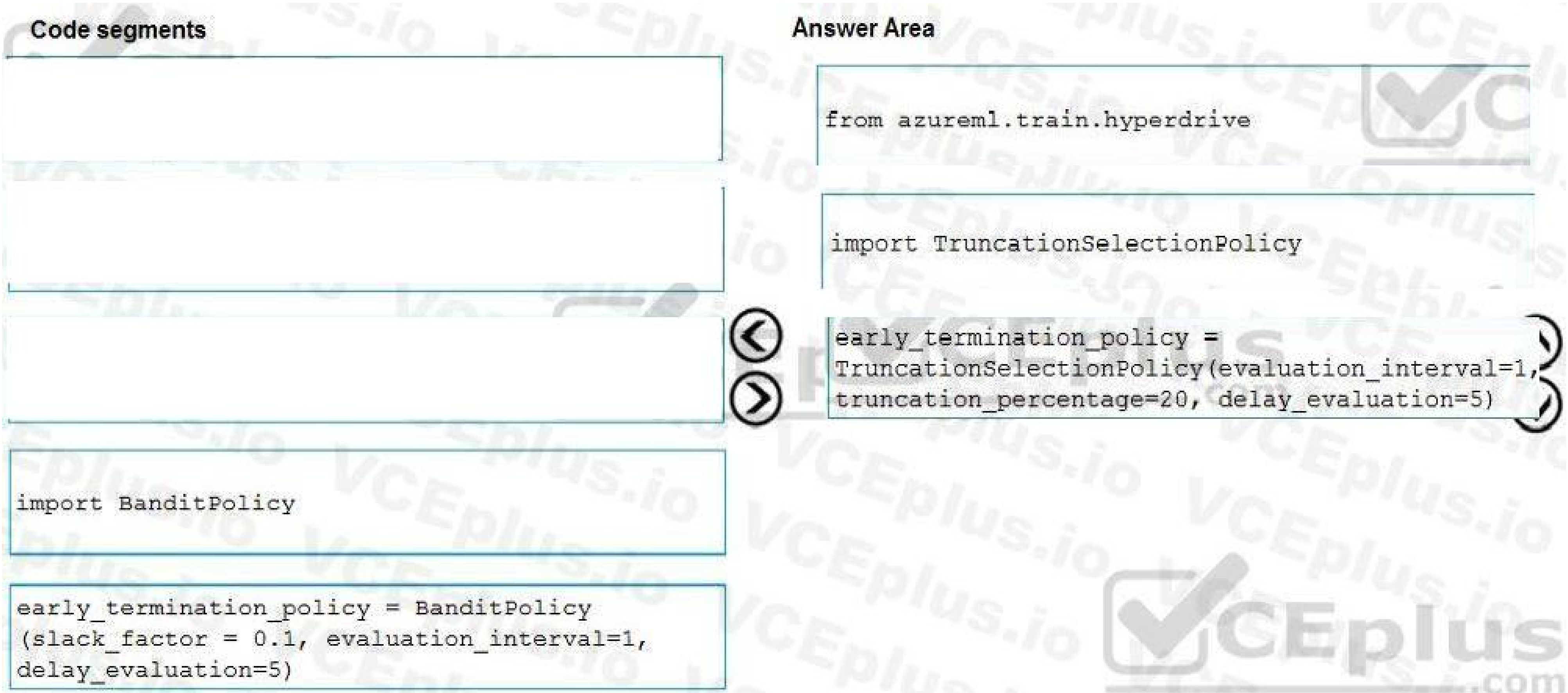
Explanation:
Box 1: Mutual Information.
The mutual information score is particularly useful in feature selection because it maximizes the mutual information between the joint distribution and target variables in datasets with many dimensions.
Box 2: MedianValue
MedianValue is the feature column, , it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

QUESTION 11
DRAG DROP
You need to implement an early stopping criteria policy for model training.
Which three code segments should you use to develop the solution? To answer, move the appropriate code segments from the list of code segments to the answer area and arrange them in the correct order.
NOTE: More than one order of answer choices is correct. You will receive credit for any of the correct orders you select.

Select and Place:



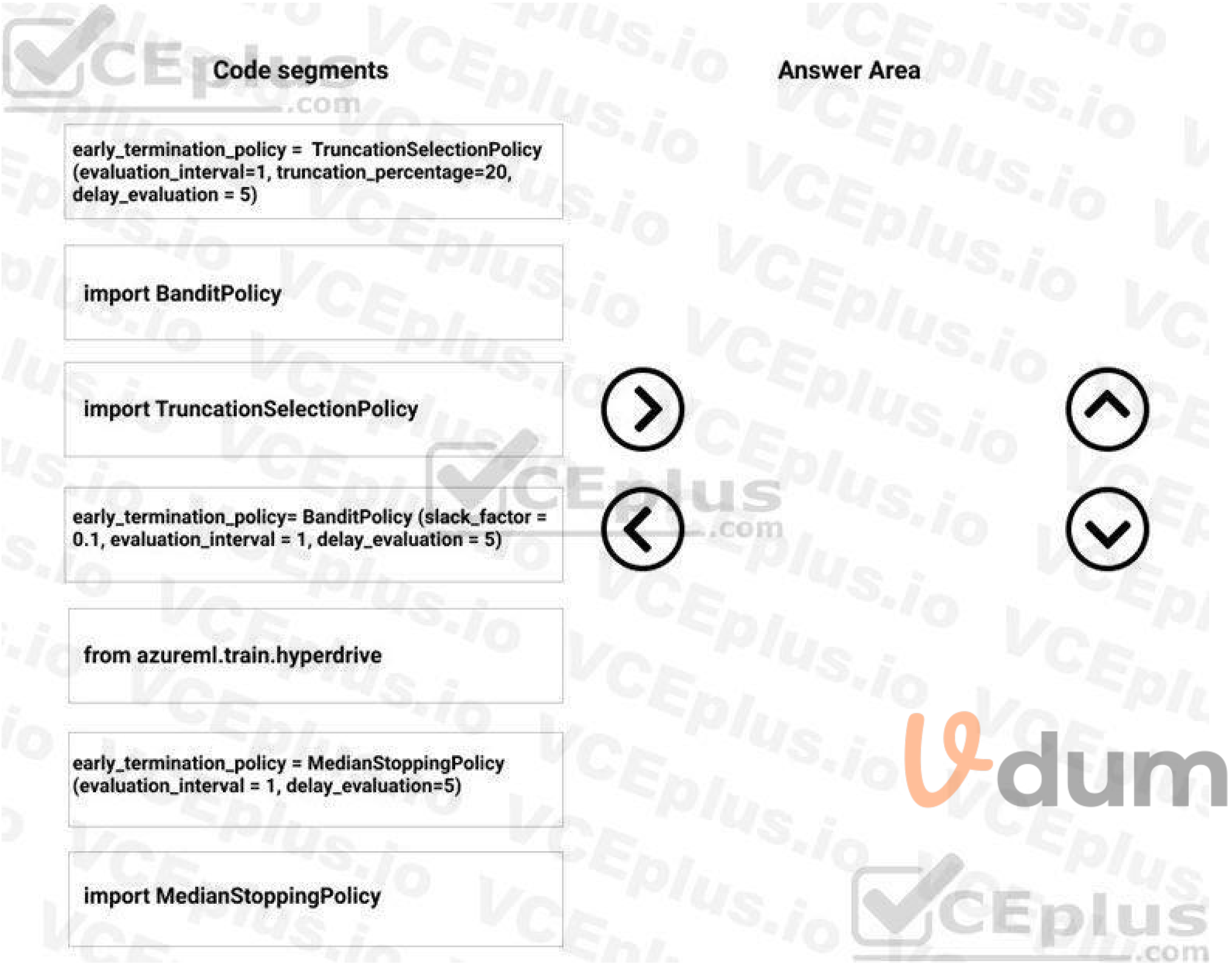
Correct Answer:



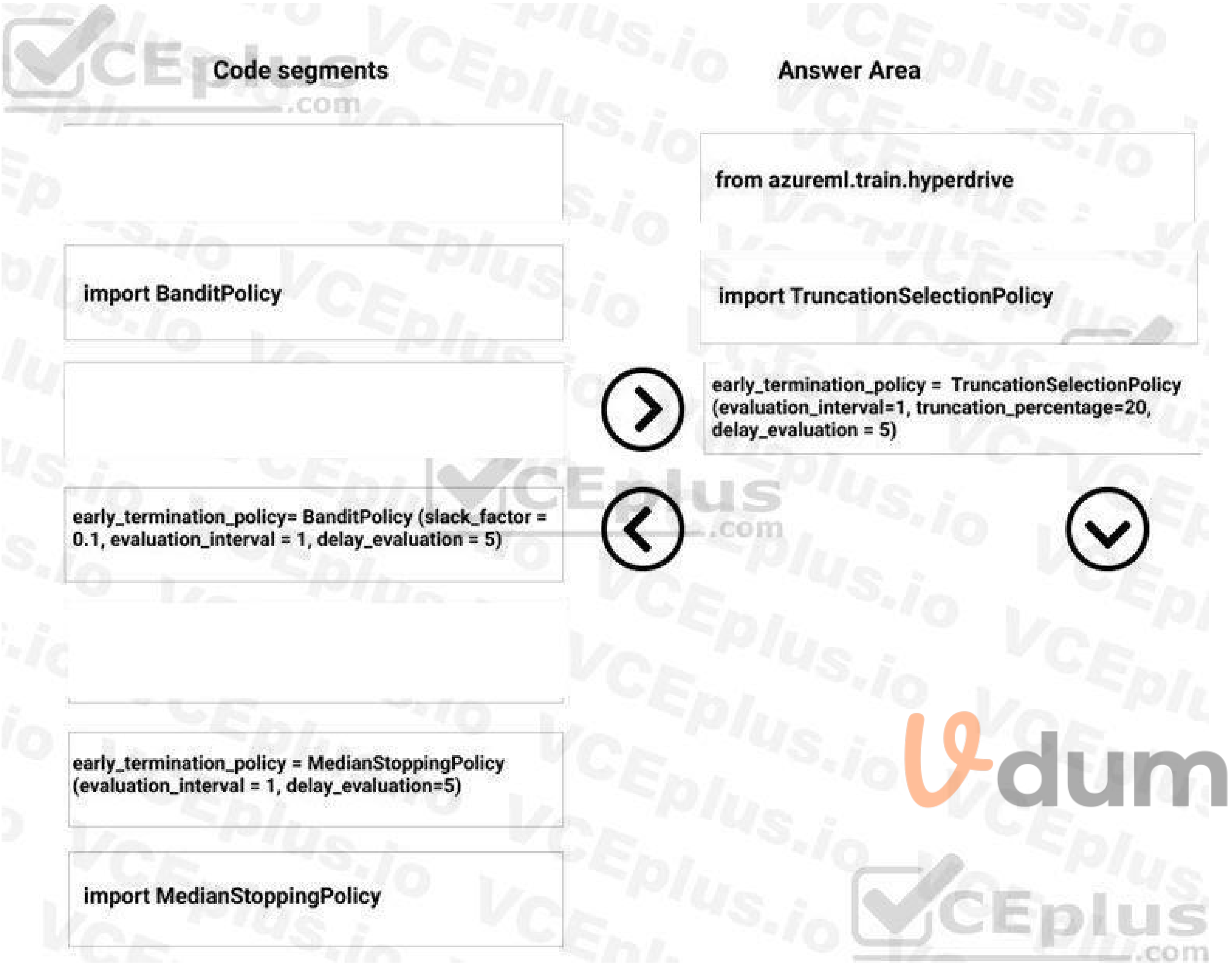
Section:
Explanation:
You need to implement an early stopping criterion on models that provides savings without terminating promising jobs.
Truncation selection cancels a given percentage of lowest performing runs at each evaluation interval. Runs are compared based on their performance on the primary metric and the lowest X% are terminated.
Example:
from azureml.train.hyperdrive import TruncationSelectionPolicy early_termination_policy = TruncationSelectionPolicy(evaluation_interval=1, truncation_percentage=20, delay_evaluation=5)
Incorrect Answers:
Bandit is a termination policy based on slack factor/slack amount and evaluation interval. The policy early terminates any runs where the primary metric is not within the specified slack factor / slack amount with respect to the best performing training run.
Example:
from azureml.train.hyperdrive import BanditPolicy
early_termination_policy = BanditPolicy(slack_factor = 0.1, evaluation_interval=1, delay_evaluation=5
References:
https://docs.microsoft.com/en-us/azure/machine-learning/service/how-to-tune-hyperparameters

QUESTION 12
DRAG DROP
You need to implement early stopping criteria as stated in the model training requirements.
Which three code segments should you use to develop the solution? To answer, move the appropriate code segments from the list of code segments to the answer area and arrange them in the correct order.
NOTE: More than one order of answer choices is correct. You will receive the credit for any of the correct orders you select.

Select and Place:



Correct Answer:



Section:
Explanation:
Step 1: from azureml.train.hyperdrive
Step 2: Import TruncationCelectionPolicy
Truncation selection cancels a given percentage of lowest performing runs at each evaluation interval. Runs are compared based on their performance on the primary metric and the lowest X% are terminated.
Scenario: You must configure hyperparameters in the model learning process to speed the learning phase. 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.
Step 3: early_terminiation_policy = TruncationSelectionPolicy..
Example:
from azureml.train.hyperdrive import TruncationSelectionPolicy
early_termination_policy = TruncationSelectionPolicy(evaluation_interval=1, truncation_percentage=20, delay_evaluation=5)
In this example, the early termination policy is applied at every interval starting at evaluation interval 5. A run will be terminated at interval 5 if its performance at interval 5 is in the lowest 20% of performance of all runs at interval 5.
Incorrect Answers:
Median:
Median stopping is an early termination policy based on running averages of primary metrics reported by the runs. This policy computes running averages across all training runs and terminates runs whose performance is worse than the median of the running averages.
Slack:

Bandit is a termination policy based on slack factor/slack amount and evaluation interval. The policy early terminates any runs where the primary metric is not within the specified slack factor / slack amount with respect to the best performing training run.
References:
https://docs.microsoft.com/en-us/azure/machine-learning/service/how-to-tune-hyperparameters

Exam D

QUESTION 1
You have a dataset that is stored m an Azure Machine Learning workspace.
You must perform a data analysis for differentiate privacy by using the SmartNoise SDK.
You need to measure the distribution of reports for repeated queries to ensure that they are balanced Which type of test should you perform?

A. Bias
B. Accuracy
C. Privacy
D. Utility

Correct Answer: B
Section:

QUESTION 2
HOTSPOT
You have a binary classifier that predicts positive cases of diabetes within two separate age groups.
The classifier exhibits a high degree of disparity between the age groups.
You need to modify the output of the classifier to maximize its degree of fairness across the age groups and meet the following requirements:
• Eliminate the need to retrain the model on which the classifier is based.
• Minimize the disparity between true positive rates and false positive rates across age groups.
Which algorithm and panty constraint should you use? To answer, select the appropriate options in the answer are
A. NOTE: Each correct selection is worth one point.

Answer:

Hot Area:

Answer Area:

Section:
Explanation:

QUESTION 3
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. Compute target
B. Kubernetes online endpoint
C. Model files
D. Online endpoint
E. Environment

Correct Answer: A, C
Section:

QUESTION 4
You create an Azure Machine Learning pipeline named pipeline1 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 pipeline1 and run the pipeline again You need to ensure the new run of pipeline1 fully processes the updated content.
Solution: Set the allow_reuse parameter of the PythonScriptStep object of both steps to False Does the solution meet the goal?

A. Yes
B. No

Correct Answer: B
Section:

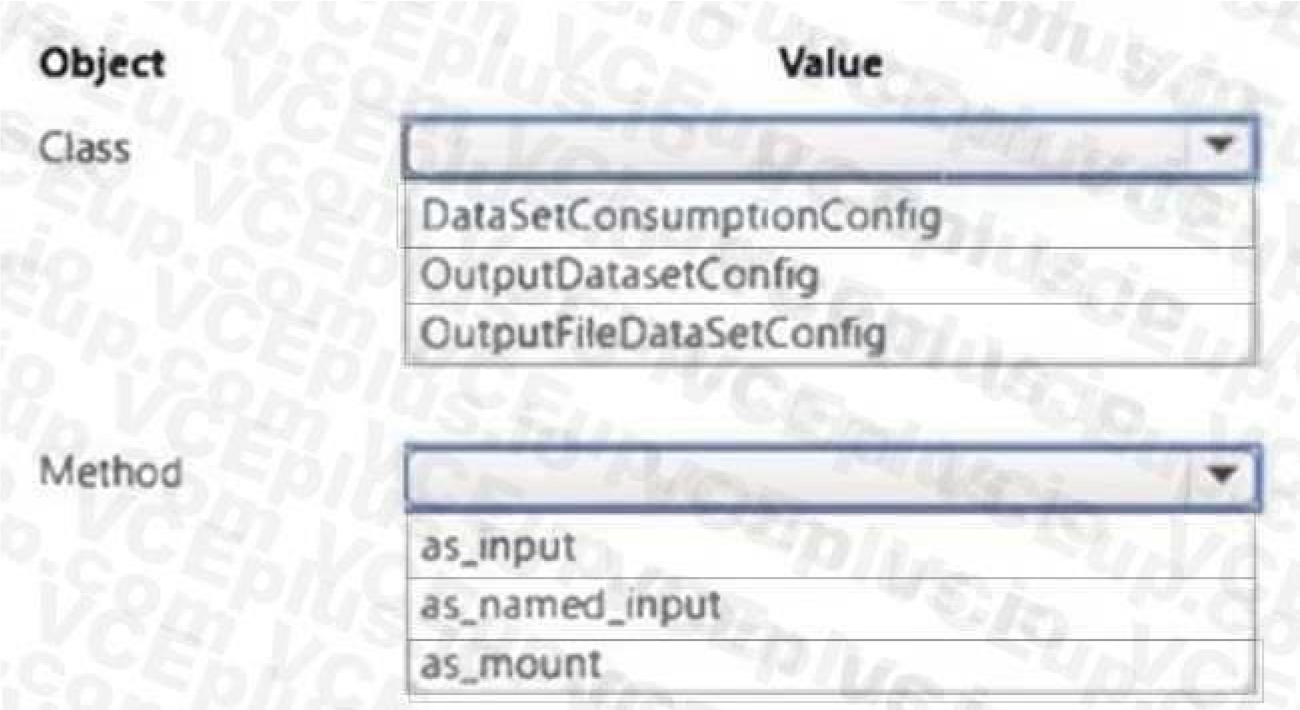
QUESTION 5
You create an Azure Machine Learning pipeline named pipeline 1 with two steps that contain Python scnpts. Data processed by the first step is passed to the second step.
You must update the content of the downstream data source of pipeline 1 and run the pipeline again.
You need to ensure the 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

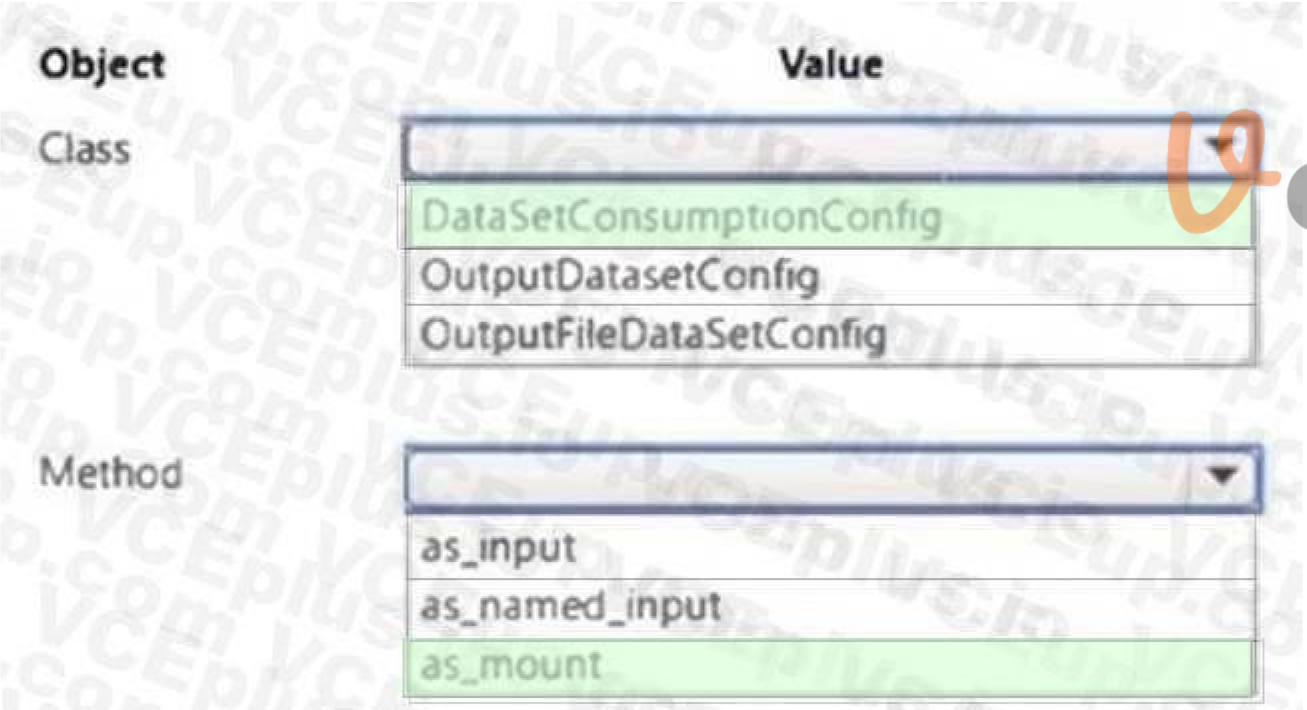
Correct Answer: B
Section:

QUESTION 6
HOTSPOT
You plan to implement a two-step pipeline by using the Azure Machine Learning SDK for Python.
The pipeline will pass temporary data from the first step to the second step.
You need to identify the class and the corresponding method that should be used in the second step to access temporary data generated by the first step in the pipeline.
Which class and method should you identify? To answer, select the appropriate options in the answer area.
NOTE: Each correct selection is worth one point

Hot Area:



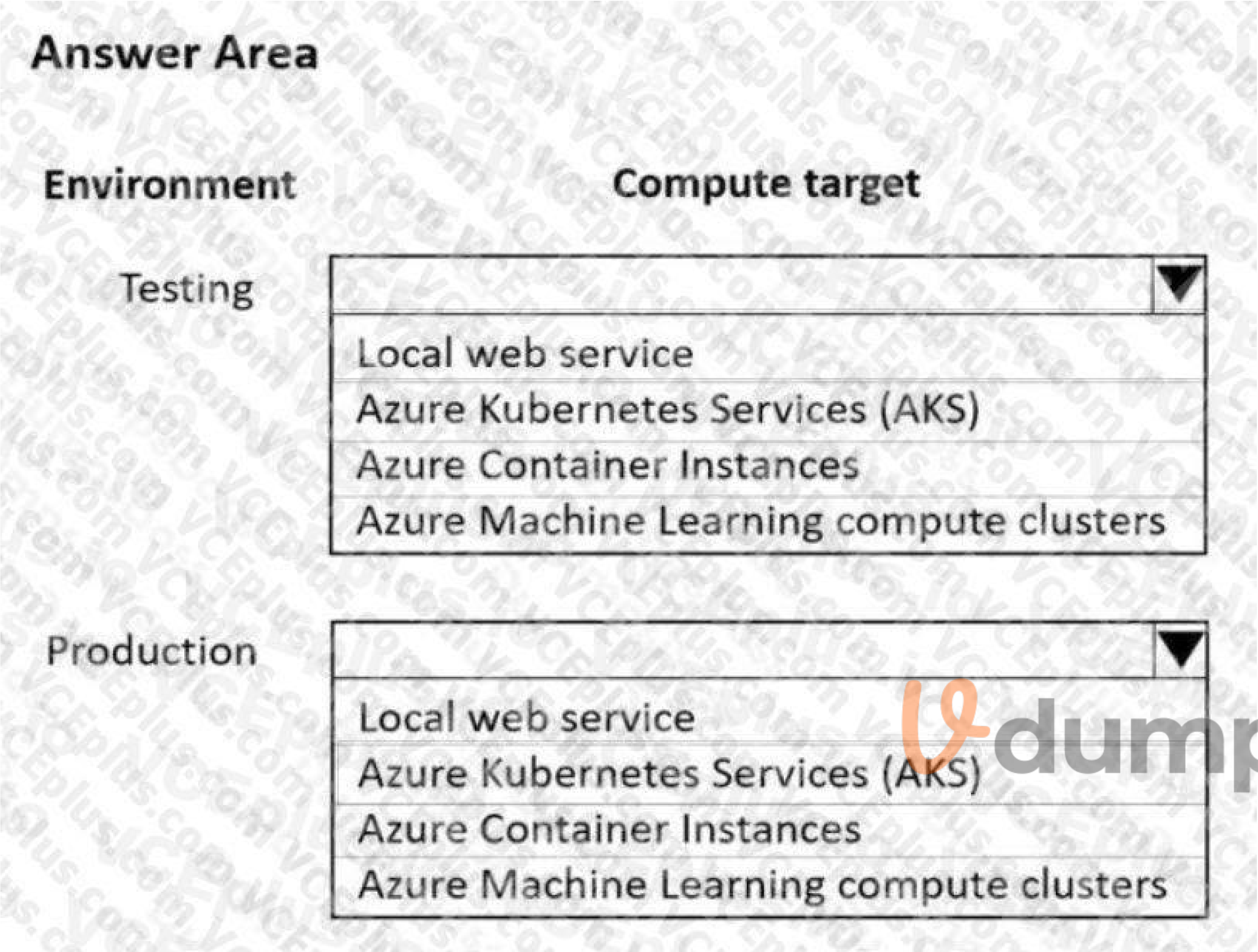
Answer Area:



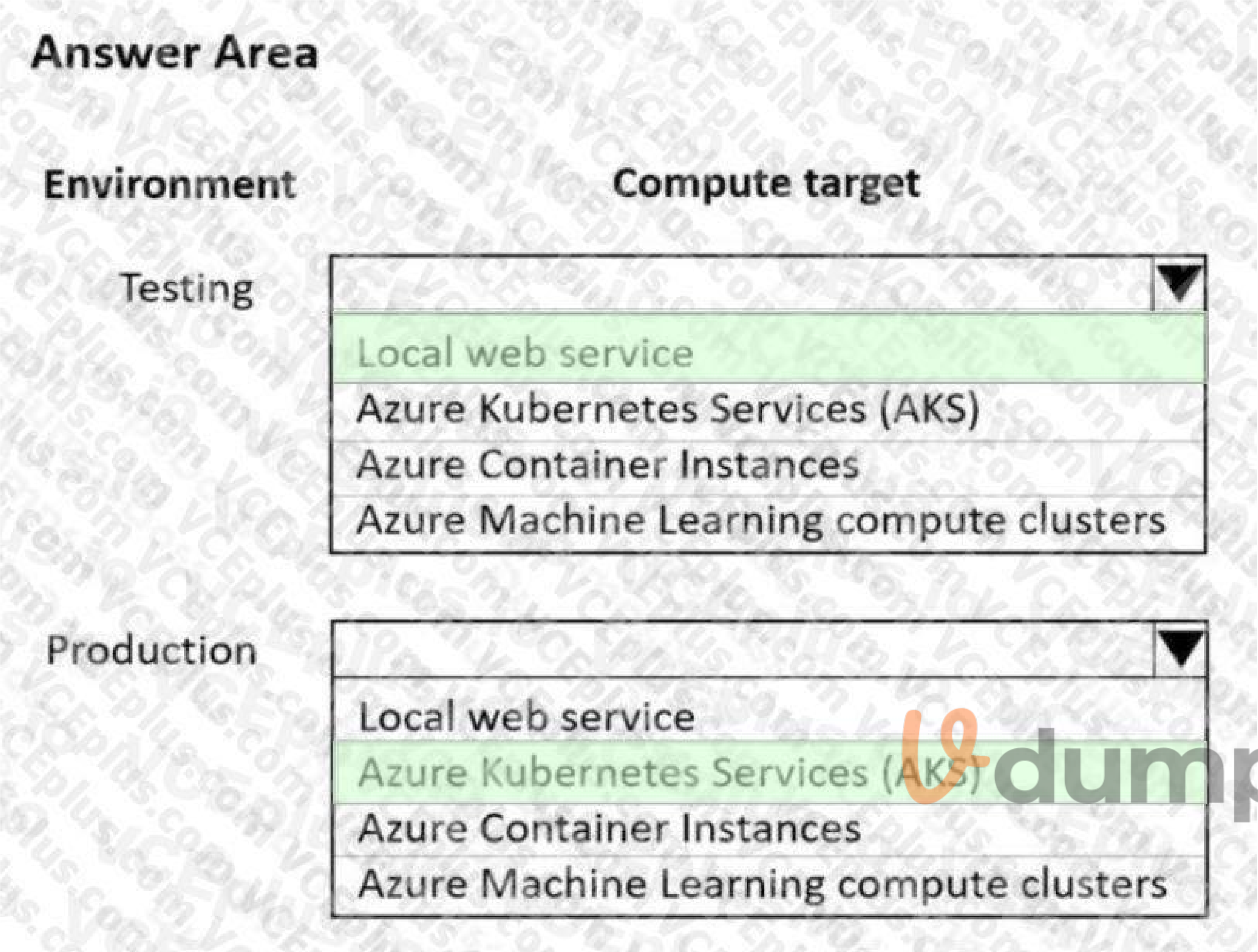
Section:
Explanation:

QUESTION 7
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:



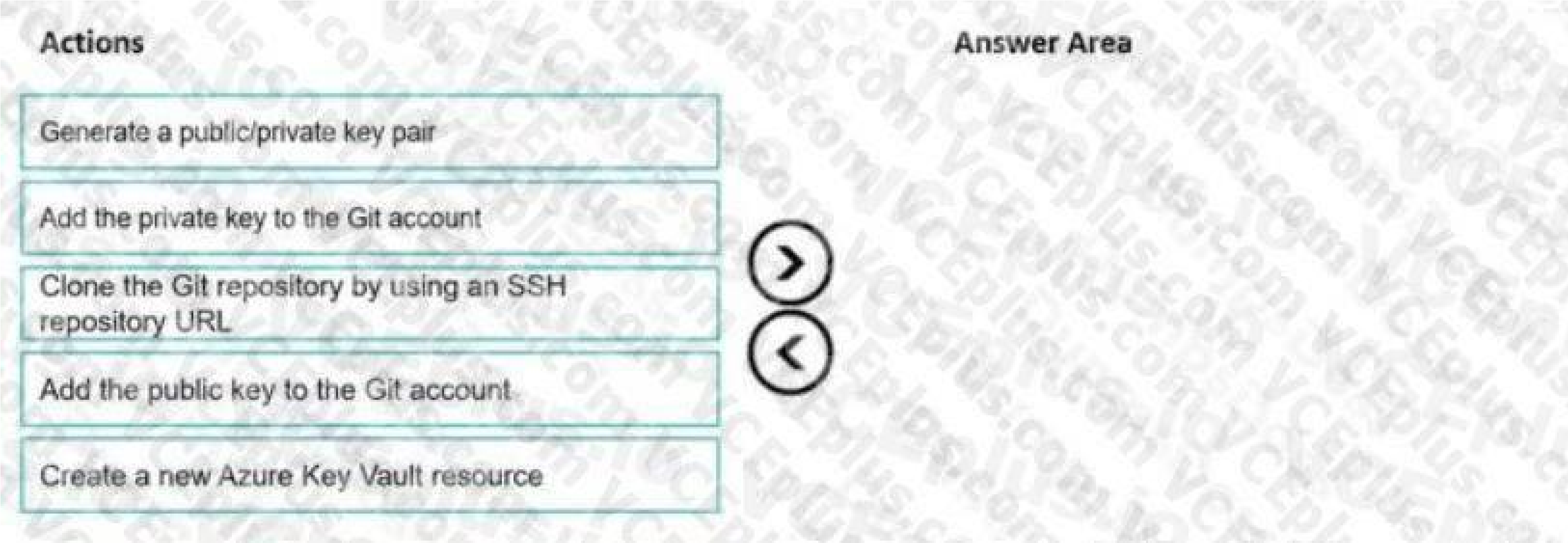
Answer Area:



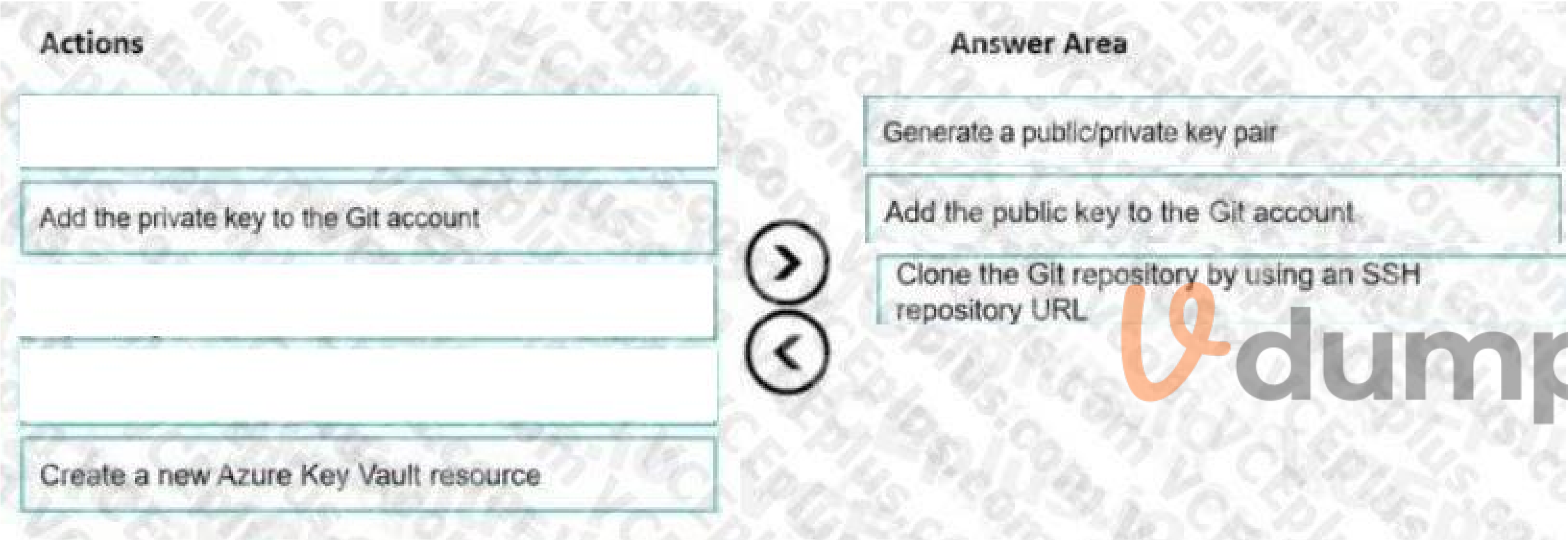
Section:
Explanation:
Box 1: Local web service
The Local web service compute target is used for testing/debugging. Use it for limited testing and troubleshooting. Hardware acceleration depends on use of libraries in the local system.
Box 2: Azure Kubernetes Service (AKS)
Azure Kubernetes Service (AKS) 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 8
DRAG DROP
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.

Select and Place:



Correct Answer:



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

QUESTION 9
HOTSPOT
You are the owner of an Azure Machine Learning workspace.

You must prevent the creation or deletion of compute resources by using a custom role. You must allow all other operations inside the workspace.
You need to configure the custom role.
How should you complete the configuration? To answer, select the appropriate options in the answer area.
NOTE: Each correct selection is worth one point.

Hot Area:

Answer Area:

Section:
Explanation:
Box 1: Microsoft.MachineLearningServices/workspaces/*/read
Reader role: Read-only actions in the workspace. Readers can list and view assets, including datastore credentials, in a workspace. Readers can't create or update these assets.
Box 2: Microsoft.MachineLearningServices/workspaces/*/write
If the roles include Actions that have a wildcard (*), the effective permissions are computed by subtracting the NotActions from the allowed Actions.
Box 3: Box 2: Microsoft.MachineLearningServices/workspaces/computes/*/delete
Box 4: Microsoft.MachineLearningServices/workspaces/computes/*/write
Reference: https://docs.microsoft.com/en-us/azure/role-based-access-control/overview#how-azure-rbac-determines-if-a-user-has-access-to-a-resource

QUESTION 10
HOTSPOT
You create an Azure Machine Learning workspace named workspace1. You assign a custom role to a user of workspace1.
The custom role has the following JSON definition:

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

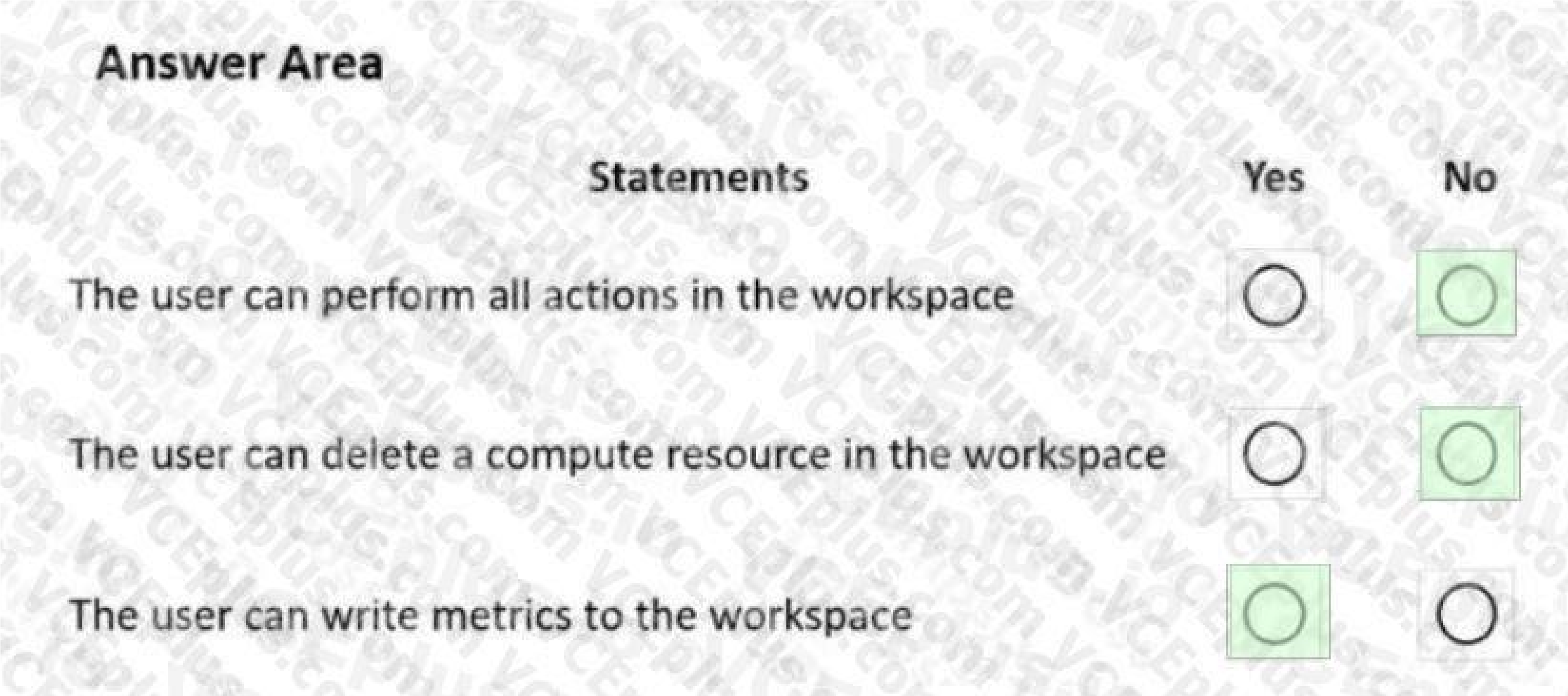
Instructions: For each of the following statements, select Yes if the statement is true. Otherwise, select No.
NOTE: Each correct selection is worth one point.

Hot Area:

Answer Area

Statements	Yes	No
The user can perform all actions in the workspace	○	○
The user can delete a compute resource in the workspace	○	○
The user can write metrics to the workspace	○	○

Answer Area:

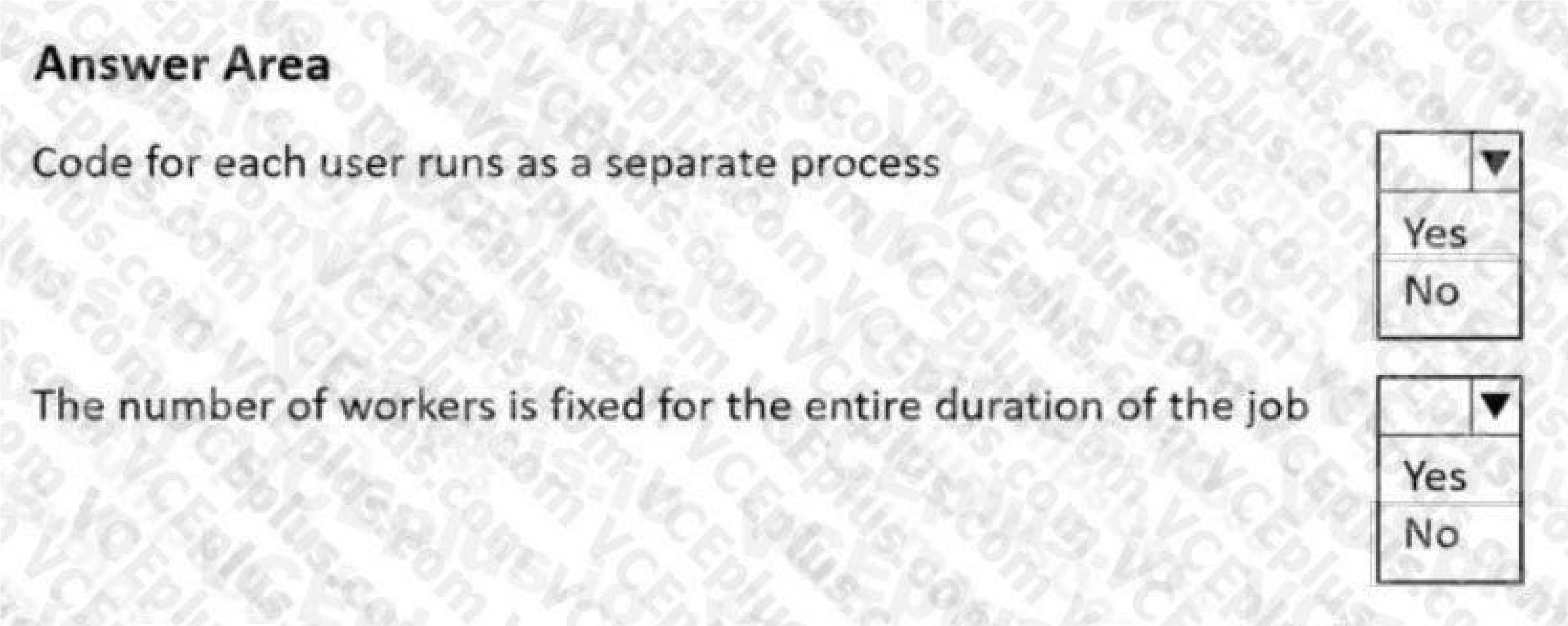


Section:
Explanation:
Box 1: No
The actions listed in NotActions are prohibited.
If the roles include Actions that have a wildcard (*), the effective permissions are computed by subtracting the NotActions from the allowed Actions.
Box 2: No
Deleting compute resources in the workspace is in the NotActions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

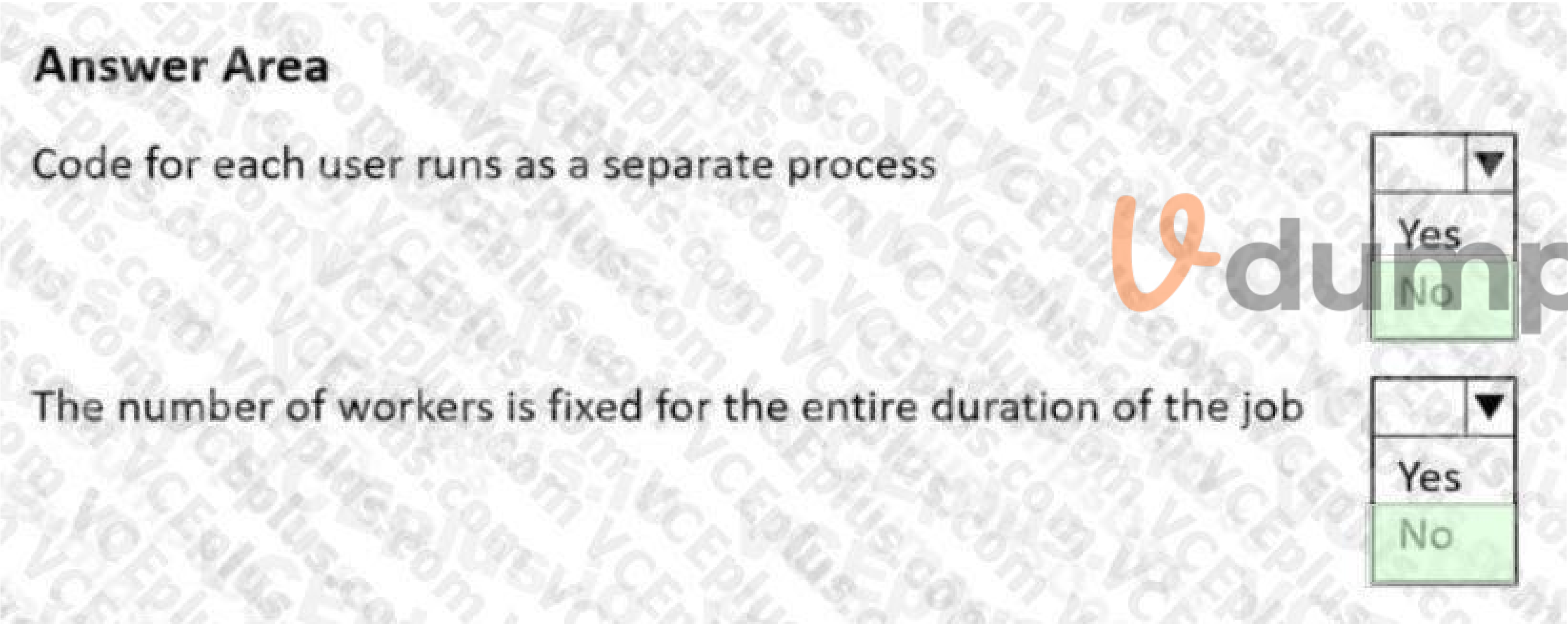
QUESTION 11
HOTSPOT
You create a new Azure Databricks workspace.
You configure a new cluster for long-running tasks with mixed loads on the compute cluster as shown in the image below.

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.

Hot Area:



Answer Area:



Section:
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

QUESTION 12
HOTSPOT
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.

Hot Area:

Answer Area

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 Area:

Answer Area

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	●	○

Section:
Explanation:
Box 1: No
Environment is a required parameter. The environment to use for the run. If no environment is specified, azureml.core.runconfig.DEFAULT_CPU_IMAGE will be used as the Docker image for the run.
The following example shows how to instantiate a new environment. from azureml.core import Environment myenv =
Environment(name="myenv")
Box 2: Yes
Parameter compute_target: The compute target where training will happen. This can either be a ComputeTarget object, the name of an existing ComputeTarget, or the string "local". If no compute target is specified, your local machine will be used.

Box 3: Yes
Parameter source_directory. A local directory containing code files needed for a run.
Parameter script. The file path relative to the source_directory of the script to be run.
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

QUESTION 13
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)
```

Answer 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)
```

Section:
Explanation:
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 14
You have an Azure Machine Learning workspace.
You plan to run a job to tram a model as an MLflow model output.
You need to specify the output mode of the MLflow model.
Which three modes can you specify? Each correct answer presents a complete solution.
NOTE: Each correct selection is worth one point.

A. rw_mount
B. ro mount
C. upload

D. download
E. direct

Correct Answer: B, C, E
Section:

QUESTION 15
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. You connect to a terminal session from the Notebooks page in Azure Machine Learning studio.
You plan to add a new Jupyter kernel that will be accessible from the same terminal session.
You need to perform the task that must be completed before you can add the new kernel.
Solution: Create an environment.
Does the solution meet the goal?

A. Yes
B. No

Correct Answer: A
Section:

QUESTION 16
You have an Azure Machine Learning workspace.
You plan to use the workspace to set up automated machine learning training for an image classification model.
You need to choose the primary metric to optimize the model training.
Which primary metric should you choose?

A. r2_score
B. mean_absolute_error
C. accuracy
D. root_mean_squared_log_error

Correct Answer: C
Section:

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 an Azure Machine Learning workspace. You connect to a terminal session from the Notebooks page in Azure Machine Learning studio.
You plan to add a new Jupyter kernel that will be accessible from the same terminal session.
You need to perform the task that must be completed before you can add the new kernel.
Solution: Delete the Python 3.6 - AzureML kernel.
Does the solution meet the goal?

A. Yes
B. No

Correct Answer: B
Section:

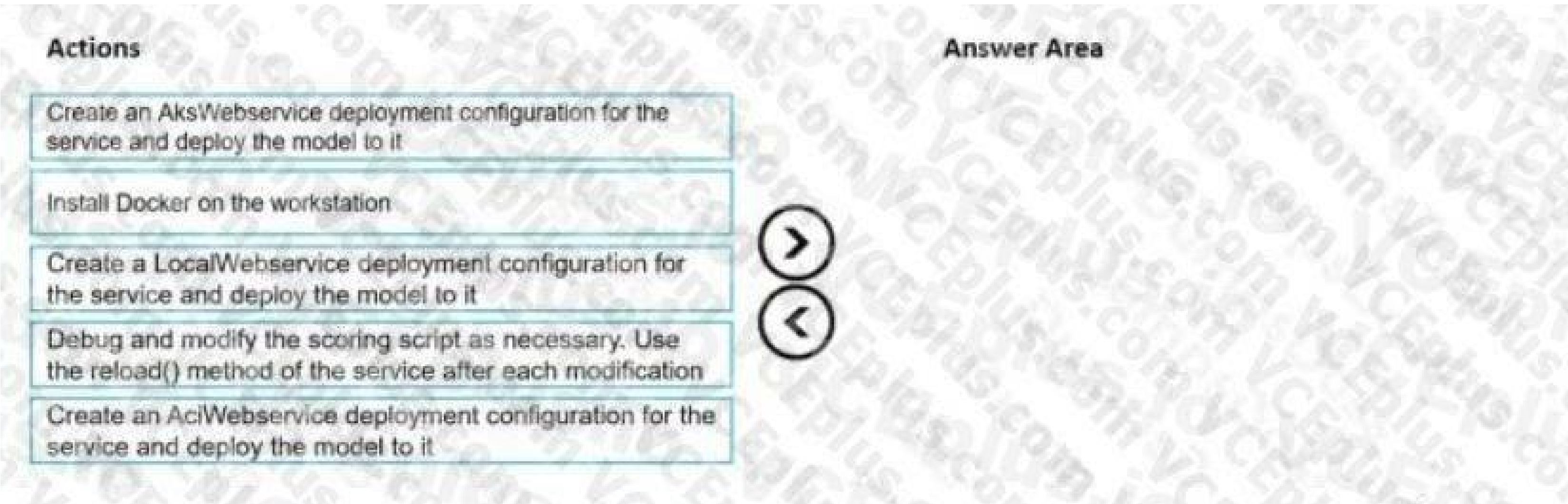
QUESTION 18
Note: This question is part of a series of questions that present the same scenario. Each question in the series contains a unique solution that might meet the stated goals. Some question sets might have more than one correct solution, while others might not have a correct solution.
After 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. You connect to a terminal session from the Notebooks page in Azure Machine Learning studio.
You plan to add a new Jupyter kernel that will be accessible from the same terminal session.
You need to perform the task that must be completed before you can add the new kernel.
Solution: Delete the Python 3.8 - AzureML kernel.
Does the solution meet the goal?

A. Yes
B. No

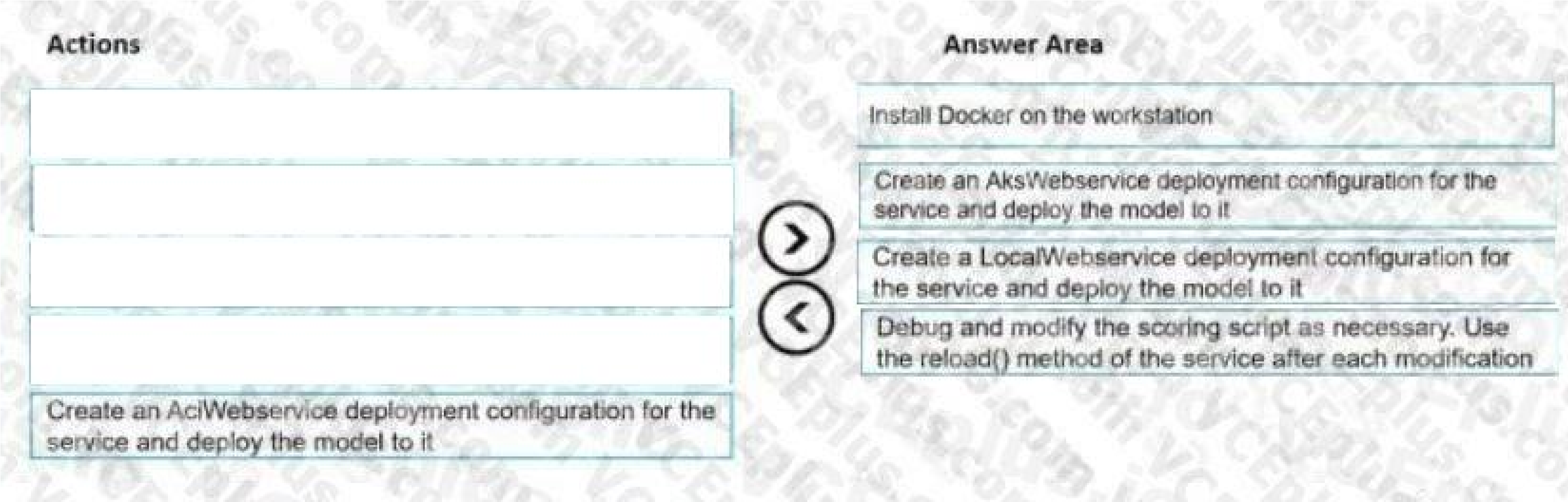
Correct Answer: B
Section:

QUESTION 19
DRAG DROP
You train and register a model by using the Azure Machine Learning SDK on a local workstation. Python 3.6 and Visual Studio Code are installed on the workstation.
When you try to deploy the model into production as an Azure Kubernetes Service (AKS)-based web service, you experience an error in the scoring script that causes deployment to fail.
You need to debug the service on the local workstation before deploying the service to production.
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:



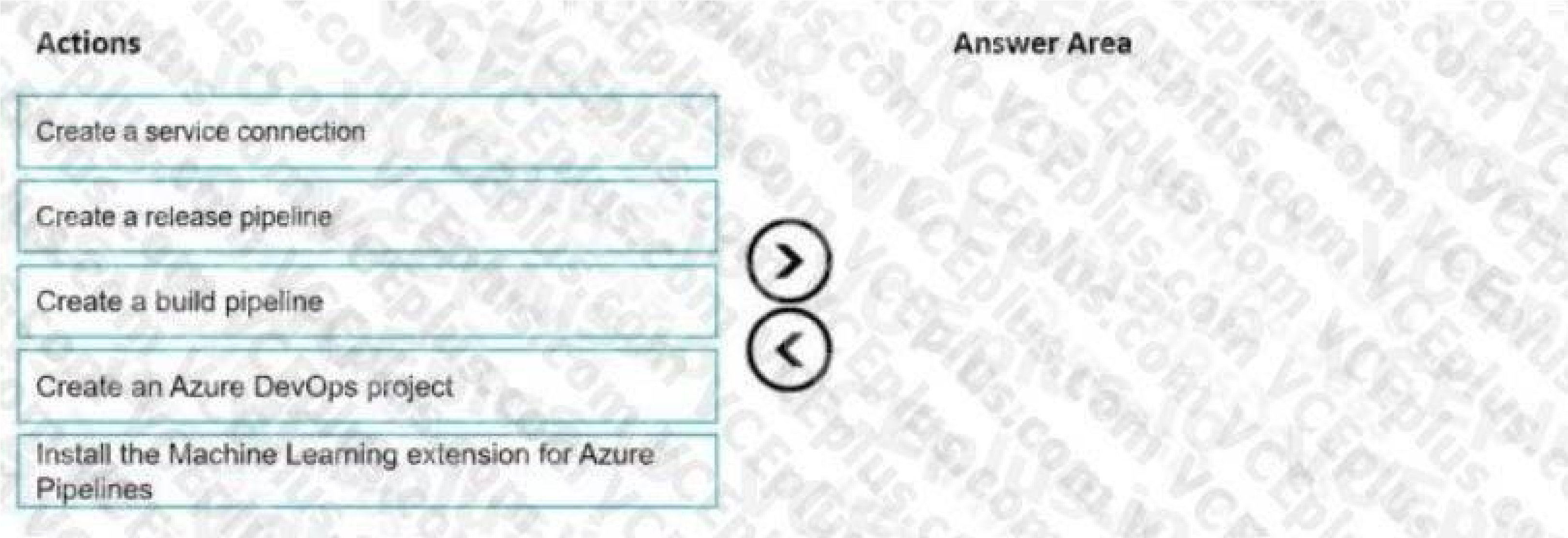
Correct Answer:



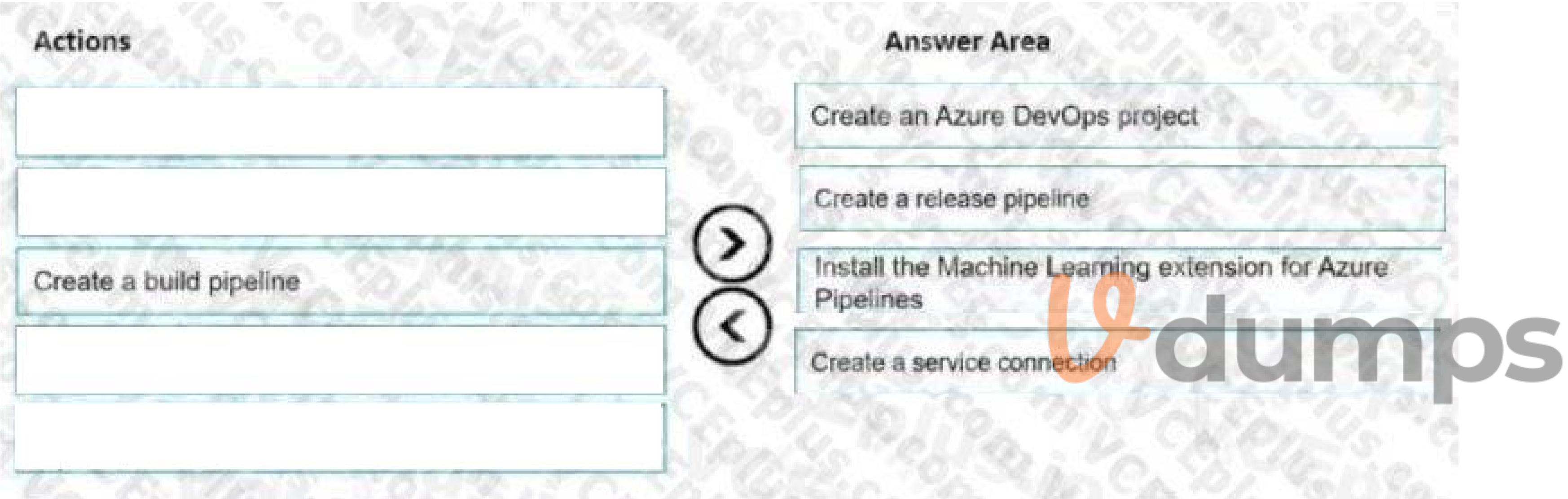
Section:
Explanation:
Step 1: Install Docker on the workstation
Prerequisites include having a working Docker installation on your local system. Build or download the dockerfile to the compute node.
Step 2: Create an AksWebservice deployment configuration and deploy the model to it To deploy a model to Azure Kubernetes Service, create a deployment configuration that describes the compute resources needed.
# 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)
service.wait_for_deployment(show_output = True) print(service.state) print(service.get_logs())
Step 3: Create a LocalWebservice deployment configuration for the service and deploy the model to it
To deploy locally, modify your code to use LocalWebservice.deploy_configuration() to create a deployment configuration.
Then use Model.deploy() to deploy the service.
Step 4: Debug and modify the scoring script as necessary. Use the reload() method of the service after each modification.
During local testing, you may need to update the score.py file to add logging or attempt to resolve any problems that you've discovered. To reload changes to the score.py file, use reload(). For example, the following code reloads the script for the service, and then sends data to it.
Incorrect Answers:
AciWebservice: The types of web services that can be deployed are LocalWebservice, which will deploy a model locally, and AciWebservice and AksWebservice, which will deploy a model to Azure Container Instances (ACI) and Azure
Kubernetes Service (AKS), respectively.
Reference:
https://docs.microsoft.com/en-us/azure/machine-learning/how-to-deploy-azure-kubernetes-service
https://docs.microsoft.com/en-us/azure/machine-learning/how-to-troubleshoot-deployment-local

QUESTION 20
DRAG DROP
You create an Azure Machine Learning workspace and a new Azure DevOps organization. You register a model in the workspace and deploy the model to the target environment.
All new versions of the model registered in the workspace must automatically be deployed to the target environment.
You need to configure Azure Pipelines to deploy the model.
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.

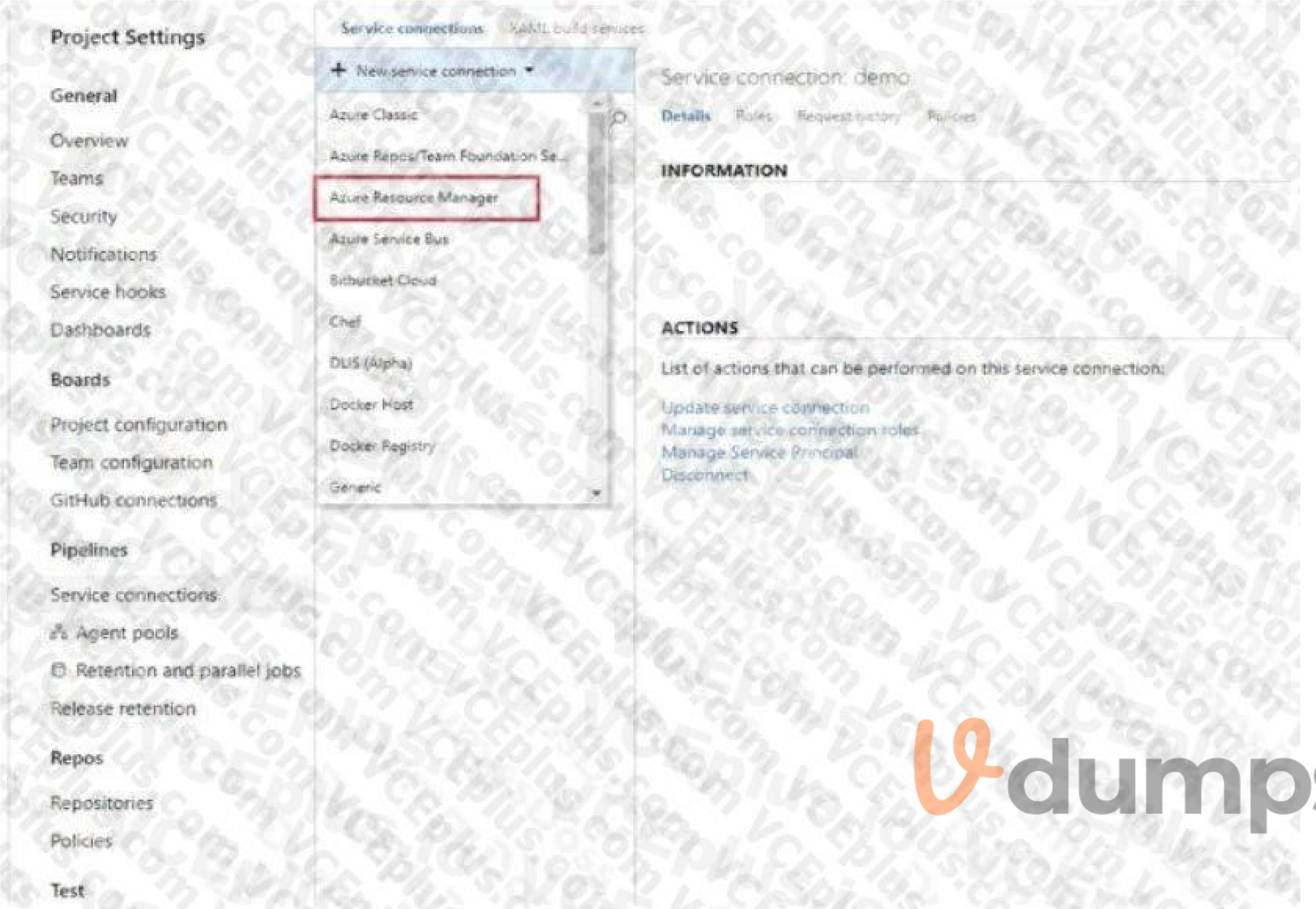
Select and Place:



Correct Answer:



Section:
Explanation:
Step 1: Create an Azure DevOps project
Step 2: Create a release pipeline
1. Sign in to your Azure DevOps organization and navigate to your project.
2. Go to Pipelines, and then select New pipeline.
Step 3: Install the Machine Learning extension for Azure Pipelines You must install and configure the Azure CLI and ML extension.
Step 4: Create a service connection
How to set up your service connection



Select AzureMLWorkspace for the scope level, then fill in the following subsequent parameters.

Note: How to enable model triggering in a release pipeline
Go to your release pipeline and add a new artifact. Click on AzureML Model artifact then select the appropriate AzureML service connection and select from the available models in your workspace. Enable the deployment trigger on your model artifact as shown here. Every time a new version of that model is registered, a release pipeline will be triggered.
Reference:
https://marketplace.visualstudio.com/items?itemName=ms-air-aiagility.vss-services-azureml https://docs.microsoft.com/en- us/azure/devops/pipelines/targets/azure-machine-learning

QUESTION 21
You use the Azure Machine Learning Python SDK to create a batch inference pipeline.
You must publish the batch inference pipeline so that business groups in your organization can use the pipeline. Each business group must be able to specify a different location for the data that the pipeline submits to the model for scoring.
You need to publish the pipeline.
What should you do?

A. Create multiple endpoints for the published pipeline service and have each business group submit jobs to its own endpoint.
B. Define a PipelineParameter object for the pipeline and use it to specify the business group-specific input dataset for each pipeline run.
C. Define a OutputFileDatasetConfig object for the pipeline and use the object to specify the business group-specific input dataset for each pipeline run.
D. Have each business group run the pipeline on local compute and use a local file for the input data.

Correct Answer: C
Section:

QUESTION 22
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.

Answer Area

Setting	Value
Technique	Grid Search
Model type	Binary classification

NOTE: Each correct selection is worth one point.

A. See below image

Correct Answer: A
Section:
Explanation:

Answer Area

Setting	Value
Technique	Grid Search
Model type	Binary classification

QUESTION 23

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.from_config()
```

Does the solution meet the goal?

A. Yes
B. No

Correct Answer: B
Section:

QUESTION 24

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

Correct Answer: B
Section:

QUESTION 25
You use Azure Machine Learning studio to analyze a dataset containing a decimal column named column1. You need to verity that the column1 values are normally distributed.
Which static should you use?

A. Profile
B. Type
C. Max
D. Mean

Correct Answer: A
Section:

QUESTION 26
You use the Azure Machine learning SDK foe Python to create a pipeline that includes the following step:
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. hash_path
C. data-as_input(name-..)
D. version

Correct Answer: A
Section:

QUESTION 27
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:

Answer 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.		

Answer Area:

Answer 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.		

Section:

Explanation:

QUESTION 28

You have an Azure Machine Learning workspace. You build a deep learning model.
You need to publish a GPU-enabled model as a web service.
Which two compute targets can you use? Each correct answer presents a complete solution.
NOTE: Each correct selection is worth one point.

A. Azure Kubernetes Service (AKS)
B. Azure Container Instances (ACI)
C. Local web service
D. Azure Machine Learning compute clusters

Correct Answer: A, B
Section:

QUESTION 29
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 nonexpiring artifact to access the model.
Solution:
Create a managed 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. Yes
B. No

Correct Answer: B
Section:

QUESTION 30
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 nonexpiring artifact to access the model.
Solution:
Create a managed online endpoint with the default authentication settings. Deploy the model to the online endpoint.
Does the solution meet the goal?

A. Yes
B. No

Correct Answer: B
Section:
Explanation:

QUESTION 31
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.

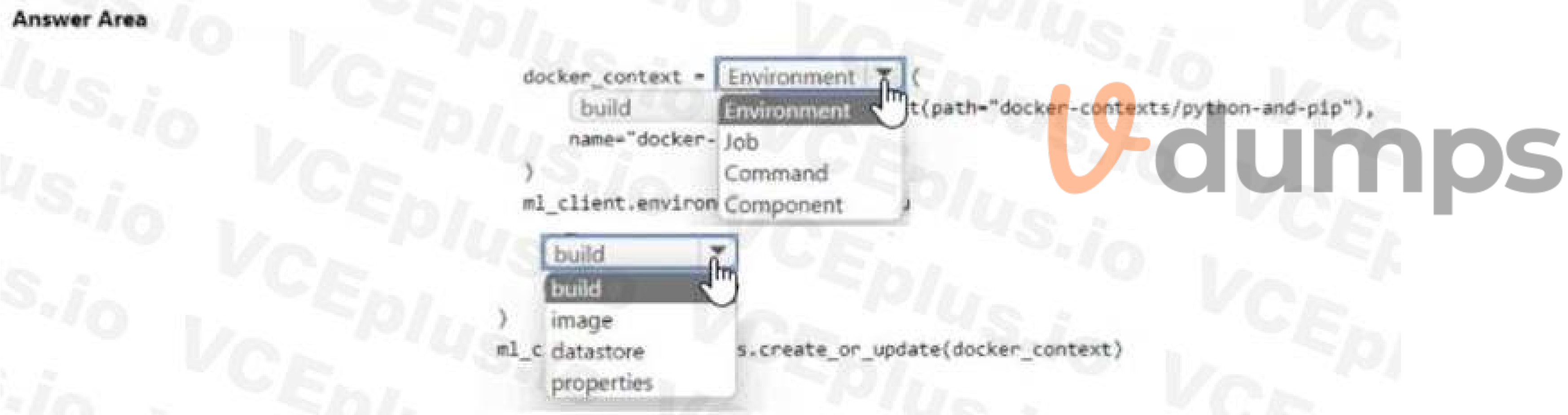
NOTE: Each correct selection is worth one point.

A. PythonScriptStep
B. AutoMLStep
C. CommandStep
D. StepRun

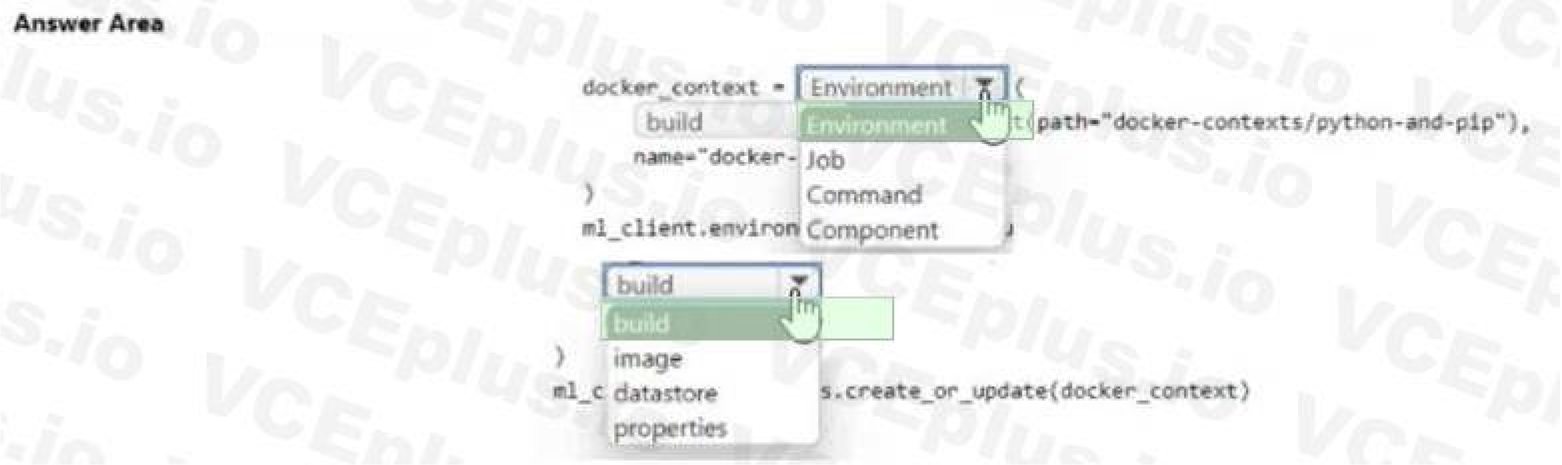
Correct Answer: A, C
Section:

QUESTION 32
HOTSPOT
You manage an Azure Machine Learning workspace.
You must define the execution environments for your jobs and encapsulate the dependencies for your code.
You need to configure the environment from a Docker build context.
How should you complete the rode segment? To answer, select the appropriate option in the answer area.
NOTE: Each correct selection is worth one point.
Answer:

Hot Area:



Answer Area:



Section:
Explanation:

Answer Area

```
docker_context = Environment (
    build =BuildContext(path="docker-contexts/python-and-pip"),
    name="docker-context "
)
ml_client.environments.create_or_update(docker_context)
```

QUESTION 33
You have a dataset that contains records of patients tested for diabetes. The datasei includes the patient s age.
You plan to create an analysis that will report the mean age value from the differentially private data derived from the dataset-
You need to identify the epsilon value to use in the analysis that minimizes the risk of exposing the actual data.
Which epsilon value should you use?

A. -1.5
B. -0.5
C. 0.5
D. 1.5

Correct Answer: C
Section:

QUESTION 34
You create a binary classification model. You use the Fairlearn package to assess model fairness. You must eliminate the need to retrain the model. You need to implement the Fair learn package. Which algorithm should you use?

A. fairlearn.reductions.ExponentiatedGradient
B. fatrlearn.reductions.GridSearch
C. fair Icarn.postprocessing.ThresholdOplimizer
D. fairlearn.preprocessing.CorrelationRemover

Correct Answer: D
Section:

QUESTION 35
HOTSPOT
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.

Hot Area:

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 Area:

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.	○	○

Section:

Explanation:

QUESTION 36

DRAG DROP

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.

Select and Place:

Actions

- Add the explanation component to the Responsible AI Insights dashboard.
- Add the error analysis component to the Responsible AI Insights dashboard.
- Add the causal component to the Responsible AI Insights dashboard.
- Load and configure the Responsible AI Insights dashboard constructor component.
- Add the counterfactuals component to the Responsible AI Insights dashboard.
- Use the Gather Responsible AI Insights dashboard component to present the dashboard.

Answer Area

Correct Answer:

Actions

- Add the explanation component to the Responsible AI Insights dashboard.
- Add the error analysis component to the Responsible AI Insights dashboard.
- Add the causal component to the Responsible AI Insights dashboard.

Answer Area

- Load and configure the Responsible AI Insights dashboard constructor component.
- Add the counterfactuals component to the Responsible AI Insights dashboard.
- Use the Gather Responsible AI Insights dashboard component to present the dashboard.

Section:

Explanation:

QUESTION 37
You have an Azure Machine Learning (ML) model deployed to an online endpoint.
You need to review container logs from the endpoint by using Azure Ml Python SDK v2. The logs must include the console log from the inference server with print/log statements from the models scoring script.
What should you do first?

A. Create an instance of the the MLClient class.
B. Create an instance of the OnlineDeploymentOperations class.
C. Connect by using SSH to the inference server.
D. Connect by using Docker tools to the inference server.

Correct Answer: A
Section:

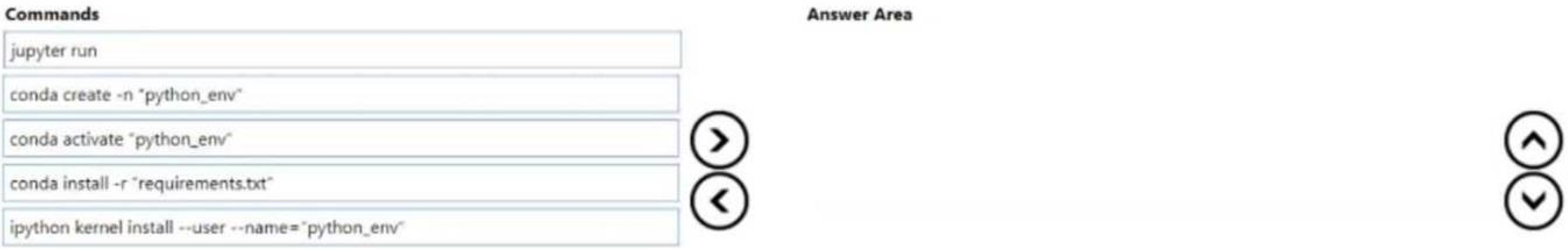
QUESTION 38
You train and publish a machine teaming model.
You need to run a pipeline that retrains the model based on a trigger from an external system.
What should you configure?

A. Azure Data Catalog
B. Azure Batch
C. Azure logic App

Correct Answer: C
Section:

QUESTION 39
DRAG DROP
You manage an Azure Machine Learning workspace named workspace1 with a compute instance named compute1. You connect to compute! by using a terminal window from wofkspace1. You create a file named "requirements.txt" containing Python dependencies to include Jupyler.
You need to add a new Jupyter kernel to compute1.
Which four commands should you use? 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:

Commands

jupyter run

Answer Area

conda create -n "python_env"

conda activate "python_env"

conda install -r "requirements.txt"

ipython kernel install --user --name="python_env"

Section:
Explanation:

QUESTION 40
You create a workspace to include a compute instance by using Azure Machine Learning Studio. You are developing a Python SDK v2 notebook in the workspace. You need to use Intellisense in the notebook. What should you do?

A. Start the compute instance.
B. Run a %pip magic function on the compute instance.
C. Run a !pip magic function on the compute instance.
D. Stop the compute instance.

Correct Answer: B
Section:

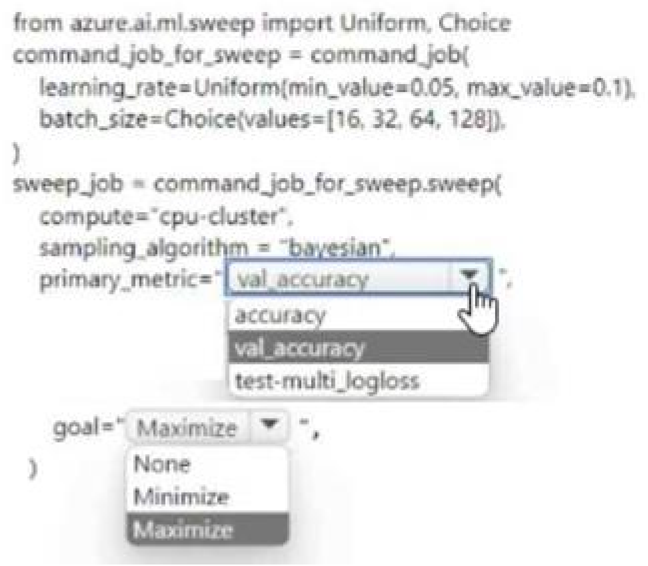
QUESTION 41
HOTSPOT
You use Azure Machine Learning to train a machine learning model.
You use the following training script in Python to perform logging:

```
import mlflow
mlflow.log_metric("accuracy", float(val_accuracy))
```

You must use a Python script to define a sweep job.
You need to provide the primary metric and goal you want hyperparameter tuning to optimize.
NOTE: Each correct selection is worth one point.

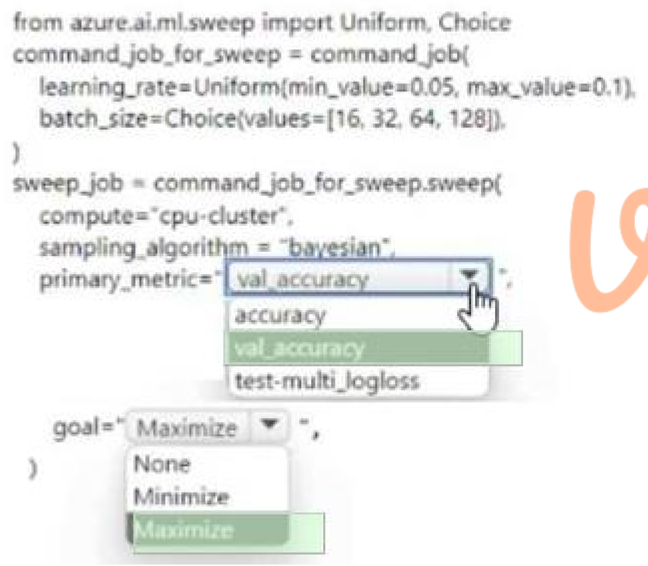
Hot Area:

Answer Area



Answer Area:

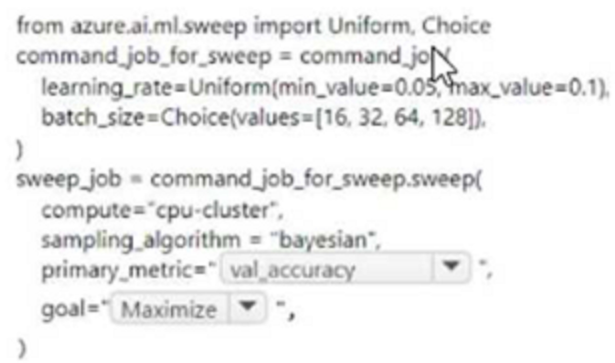
Answer Area



Section:

Explanation:

Answer Area



QUESTION 42

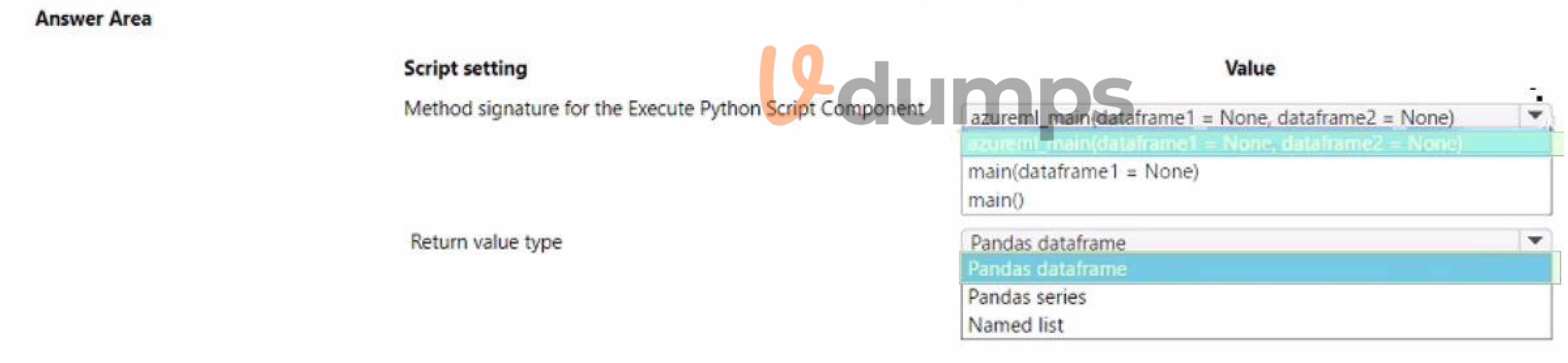
HOTSPOT

You are using the Azure Machine Learning designer to transform a dataset by using an Execute Python Script component and custom code.

You need to define the method signature for the Execute Python Script component and return value type.
What should you define? To answer, select the appropriate options in the answer area.
NOTE: Each correct selection is worth one point.

Hot Area:

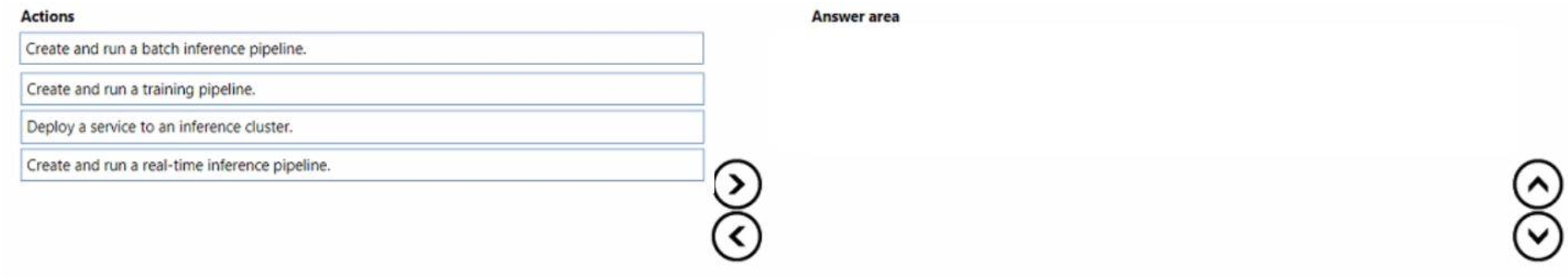
Answer Area:



Section:
Explanation:

QUESTION 43
DRAG DROP
You are developing a machine learning solution by using the Azure Machine Learning designer.
You need to create a web service that applications can use to submit data feature values and retrieve a predicted label.
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:

Section:
Explanation:
Create and run a training pipeline.
Deploy a service to an inference cluster.
Create and run a real-time inference pipeline.

QUESTION 44
You manage an Azure Machine Learning workspace.
You must provide explanations for the behavior of the models with feature importance measures.
You need to configure a Responsible AI dashboard in Azure Machine Learning.
Which dashboard component should you configure?

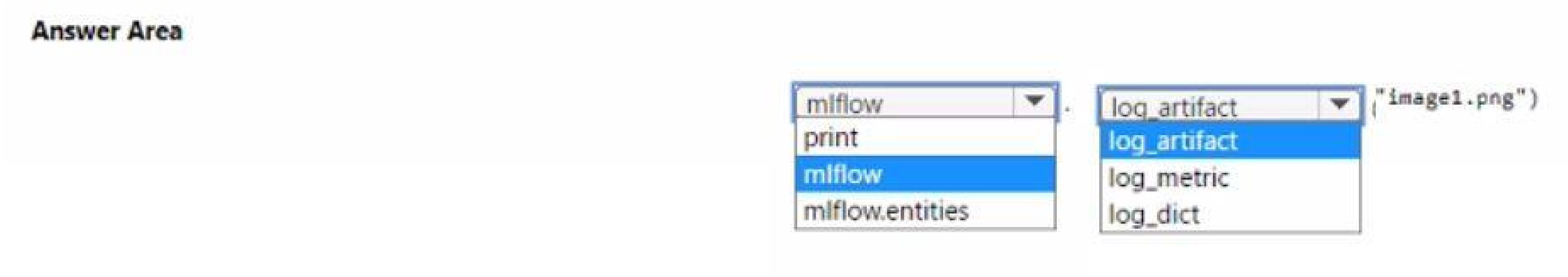
A. Fairness assessment
B. Counterfactual what-if
C. Interpretability
D. Casual inference

Correct Answer: C
Section:

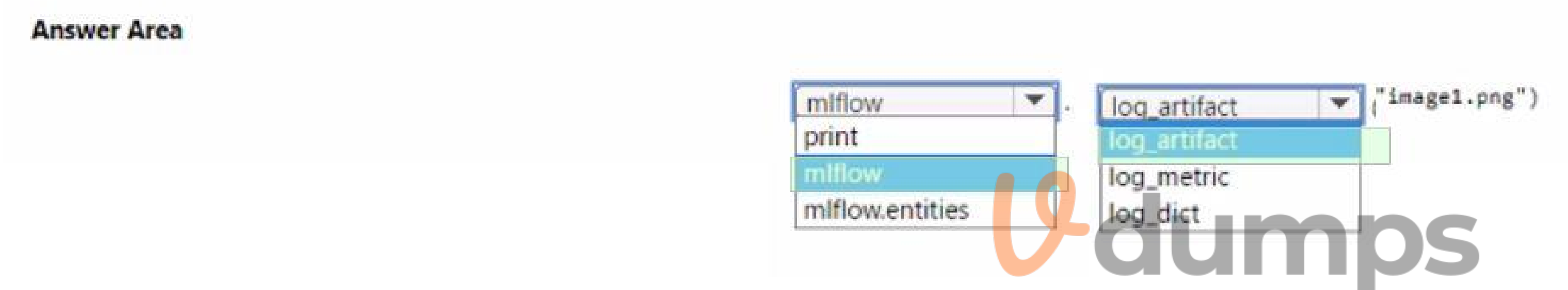
QUESTION 45

HOTSPOT
You create an Azure Machine Learning workspace.
You plan to write an Azure Machine Learning SDK for Python v2 script that logs an image for an experiment. The logged image must be available from the images tab in Azure Machine Learning Studio.
You need to complete the script.
Which code segments should you use? To answer, select the appropriate options in the answer area.
NOTE: Each correct selection is worth one point.

Hot Area:



Answer Area:



Section:
Explanation:

QUESTION 46
You manage an Azure Machine Learning workspace.
You must log multiple metrics by using MLflow.
You need to maximize logging performance.
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. MLflowClient.log_batch
B. mlflowlog_metrics
C. mlflow.log_param
D. mlflow.log. metric

Correct Answer: A, B
Section:

QUESTION 47
You manage an Azure Machine Learning workspace.
You need to define an environment from a Docker image by using the Azure Machine Learning Python SDK v2.
Which parameter should you use?

A. conda_file
B. image
C. build
D. properties

Correct Answer: B
Section:

QUESTION 48
You use Azure Machine Learning studio to analyze an mltable data asset containing a decimal column named column1. You need to verify that the column1 values are normally distributed.
Which statistic should you use?

A. Max
B. Type
C. Profile
D. Mean

Correct Answer: C
Section:

QUESTION 49
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 use Azure Machine Learning designer to load the following datasets into an experiment:

Dataset1

Age	Length	Width
3	22	13
7	11	96
18	32	85

Dataset2

Age	Length	Width
11	101	65
6	98	23
33	22	54
17	52	12

You need to create a dataset that has the same columns and header row as the input datasets and contains all rows from both input datasets.
Solution: Use the Add Rows module.
Does the solution meet the goal?

A. Yes
B. No

Correct Answer: B

Section:

QUESTION 50

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 use Azure Machine Learning designer to load the following datasets into an experiment:

Dataset1

Age	Length	Width
3	22	13
7	11	96
18	32	85

Dataset2

Age	Length	Width
11	101	65
6	98	23
33	22	54
17	52	12

You need to create a dataset that has the same columns and header row as the input datasets and contains all rows from both input datasets.

Solution: Use the Join Data module.

Does the solution meet the goal?

A. Yes

B. No

Correct Answer: B

Section:

QUESTION 51

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 use Azure Machine Learning designer to load the following datasets into an experiment:

Dataset1

Age	Length	Width
3	22	13
7	11	96
18	32	85

Dataset2

Age	Length	Width
11	101	65
6	98	23
33	22	54
17	52	12

You need to create a dataset that has the same columns and header row as the input datasets and contains all rows from both input datasets.
Solution: Use the Execute Python Script module.
Does the solution meet the goal?

A. Yes
B. No

Correct Answer: B
Section:

QUESTION 52
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 use Azure Machine Learning designer to load the following datasets into an experiment:

Dataset1

Age	Length	Width
3	22	13
7	11	96
18	32	85

Dataset2

Age	Length	Width
11	101	65
6	98	23
33	22	54
17	52	12

You need to create a dataset that has the same columns and header row as the input datasets and contains all rows from both input datasets.
Solution: Use the Apply Transformation module.

Does the solution meet the goal?

A. Yes
B. No

Correct Answer: B
Section:

QUESTION 53
You have an Azure Machine Learning workspace.
You plan to tune a model hyperparameter when you train the model.
You need to define a search space that returns a normally distributed value.
Which parameter should you use?

A. QUniform
B. LogUniform
C. Uniform
D. LogNormal

Correct Answer: A
Section:

QUESTION 54
You have an Azure Machine Learning workspace named WS1.
You plan to use the Responsible AI dashboard to assess MLflow models that you will register in WS1.
You need to identify the library you should use to register the MLflow models.
Which library should you use?

A. PyTorch
B. mlpy
C. TensorFlow
D. scikit-learn

Correct Answer: A
Section:

QUESTION 55
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

Correct Answer: A
Section:

QUESTION 56

You have an Azure Machine Learning workspace named Workspace 1 Workspace! has a registered Mlflow model named model 1 with PyFunc flavor
You plan to deploy model1 to an online endpoint named endpoint1 without egress connectivity by using Azure Machine learning Python SDK vl
You have the following code:

```
blue_deployment = ManagedOnlineDeployment(
    name="blue",
    endpoint_name=endpoint1,
    model=model1,
    instance_type="Standard_F4s_v2",
    instance_count=1
)
```

You need to add a parameter to the ManagedOnlineDeployment object to ensure the model deploys successfully
Solution: Add the environment parameter.
Does the solution meet the goal?

A. Yes
B. No

Correct Answer: B
Section:

QUESTION 57
You have an Azure Machine Learning workspace named Workspace 1 Workspace! has a registered Mlflow model named model 1 with PyFunc flavor
You plan to deploy model1 to an online endpoint named endpointl without egress connectivity by using Azure Machine learning Python SDK vl
You have the following code:

```
blue_deployment = ManagedOnlineDeployment(
    name="blue",
    endpoint_name=endpoint1,
    model=model1,
    instance_type="Standard_F4s_v2",
    instance_count=1
)
```


You need to add a parameter to the ManagedOnllneDeployment object to ensure the model deploys successfully
Solution: Add the with_package parameter.
Does the solution meet the goal?

A. Yes
B. No

Correct Answer: B
Section:

QUESTION 58
You have an Azure Machine Learning workspace named Workspace 1 Workspace! has a registered Mlflow model named model 1 with PyFunc flavor
You plan to deploy model1 to an online endpoint named endpoint1 without egress connectivity by using Azure Machine learning Python SDK vl
You have the following code:

```
blue_deployment = ManagedOnlineDeployment(
    name="blue",
    endpoint_name=endpoint1,
    model=model1,
    instance_type="Standard_F4s_v2",
    instance_count=1
)
```

You need to add a parameter to the ManagedOnllneDeployment object to ensure the model deploys successfully
Solution: Add the scoring_script parameter.

Does the solution meet the goal?

A. Yes
B. No

Correct Answer: A
Section:

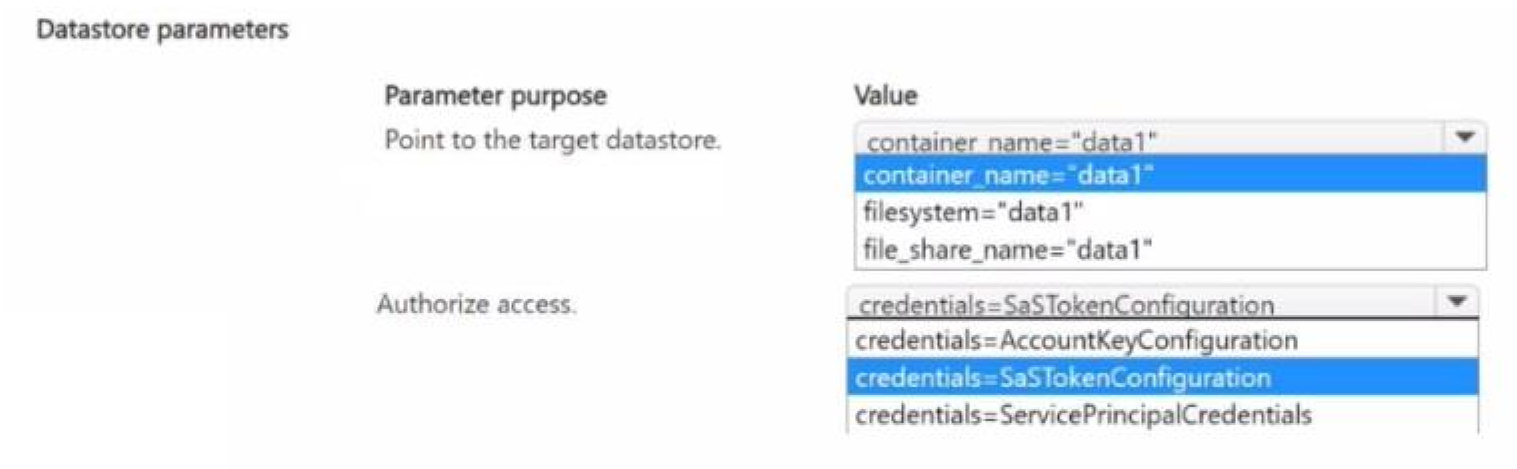
QUESTION 59
You manage an Azure Machine Learning workspace. You plan to import data from Azure Data Lake Storage Gen2. You need to build a URI that represents the storage location. Which protocol should you use?

A. abfss
B. https
C. adl
D. wasbs

Correct Answer: A
Section:

QUESTION 60
HOTSPOT
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.

Hot Area:



Answer Area:

Section:
Explanation:

QUESTION 61
You manage an Azure Machine Learning workspace. The development environment for 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 to use a user-assigned identity.
Solution: Pass the UserAssignedIdentity class object to the SynapseSparkCompute class.
Does the solution meet the goat?

A. Yes
B. No

Correct Answer: B
Section:

QUESTION 62
You manage an Azure Machine Learning workspace. The development environment for 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 to use a user-assigned identity.
Solution: Initialize the DefaultAzureCredential class.
Does the solution meet the goal?

A. Yes
B. No

Correct Answer: B
Section: