Microsoft

Exam Questions DP-100

Designing and Implementing a Data Science Solution on Azure

About Exambible

Your Partner of IT Exam

Found in 1998

Exambible is a company specialized on providing high quality IT exam practice study materials, especially Cisco CCNA, CCDA, CCNP, CCIE, Checkpoint CCSE, CompTIA A+, Network+ certification practice exams and so on. We guarantee that the candidates will not only pass any IT exam at the first attempt but also get profound understanding about the certificates they have got. There are so many alike companies in this industry, however, Exambible has its unique advantages that other companies could not achieve.

Our Advances

* 99.9% Uptime

All examinations will be up to date.

* 24/7 Quality Support

We will provide service round the clock.

* 100% Pass Rate

Our guarantee that you will pass the exam.

* Unique Gurantee

If you do not pass the exam at the first time, we will not only arrange FULL REFUND for you, but also provide you another exam of your claim, ABSOLUTELY FREE!

NEW QUESTION 1
- (Exam Topic 3)
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.

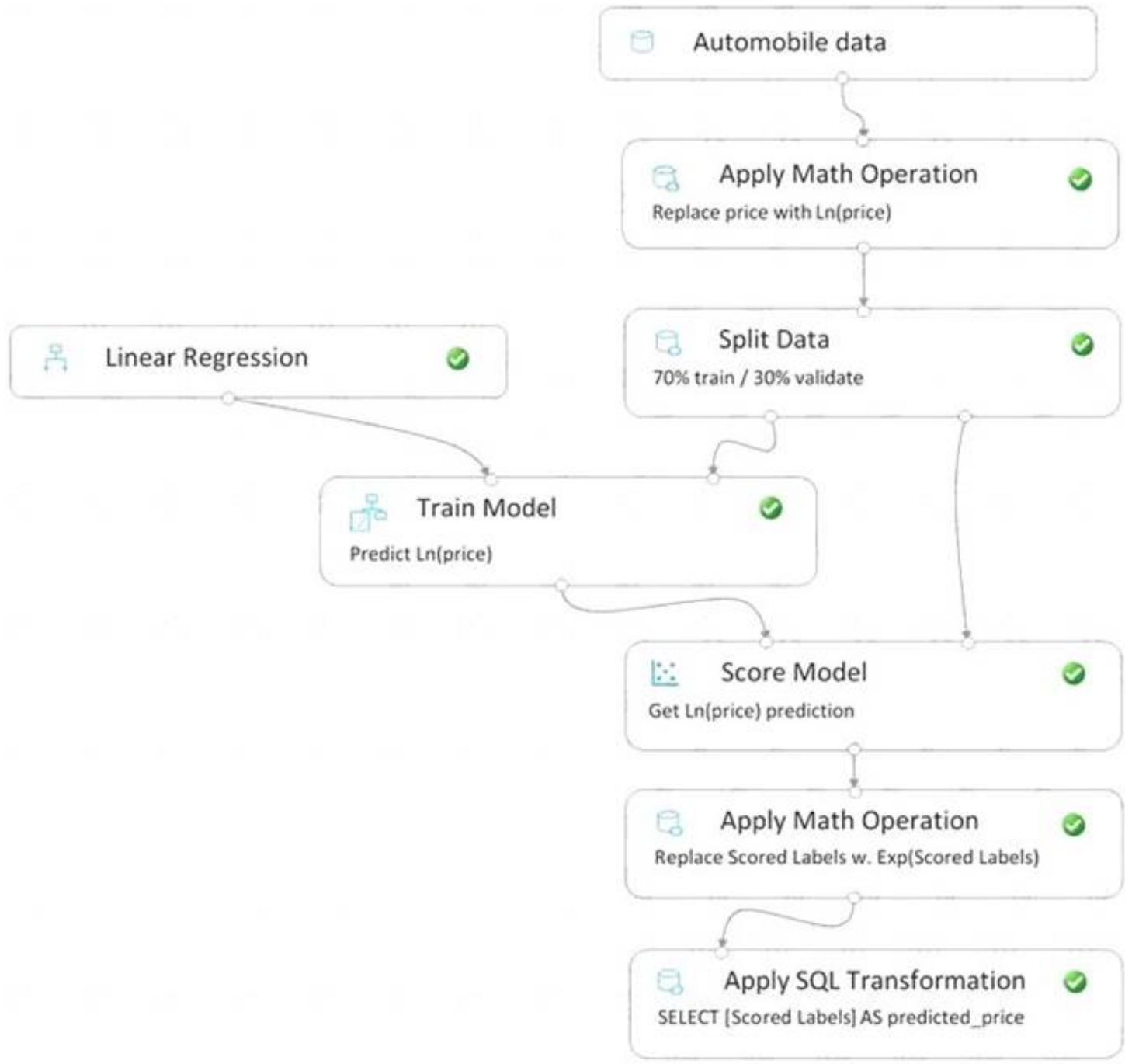
	Yes	No
The estimator will look for the files it needs to run an experiment in the training-experiment directory of the local compute environment.	○	○
The estimator will mount the local data-folder folder and make it available to the script through a parameter.	○	○
The train.py script file will be created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

A. Mastered
B. Not Mastered

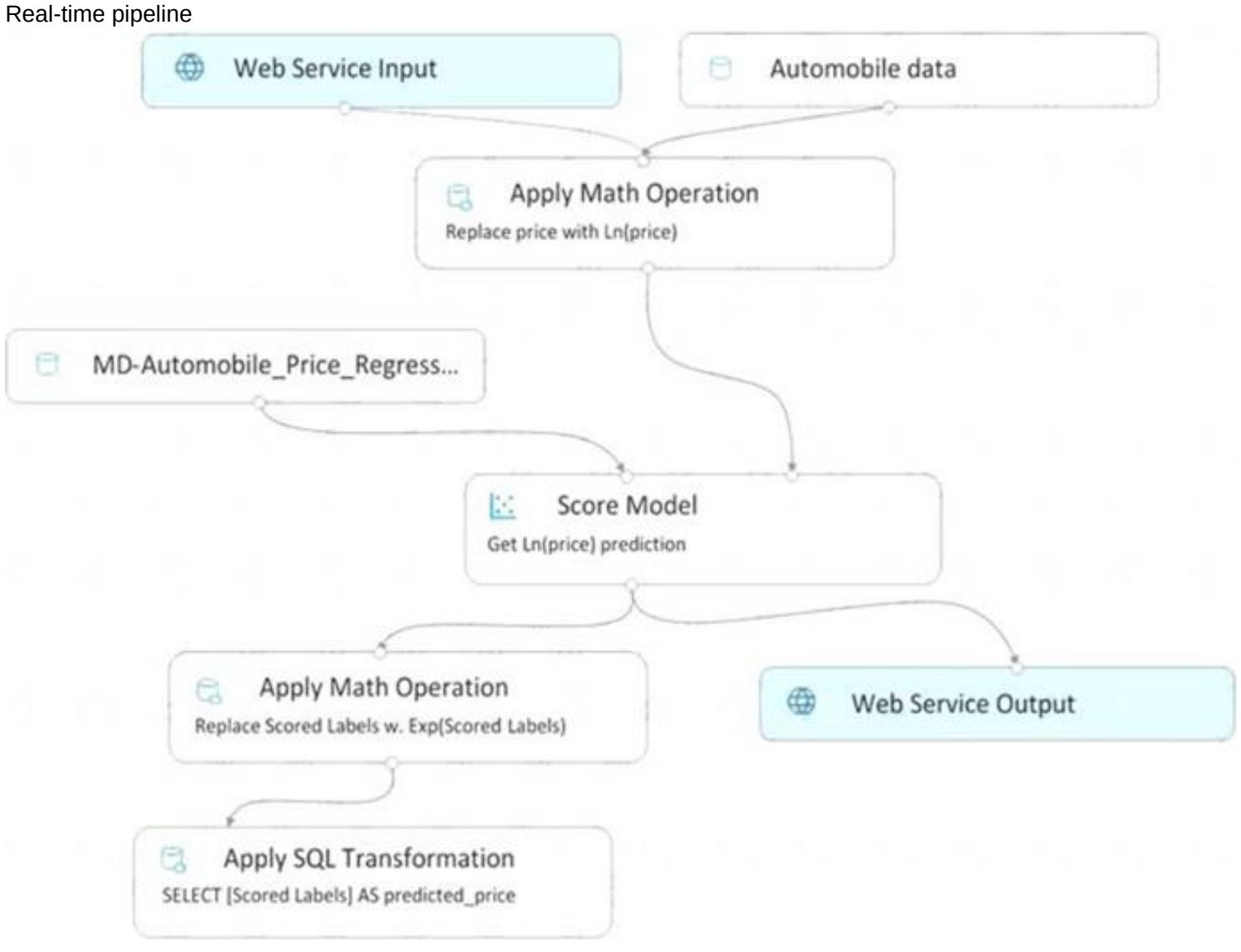
Answer: A

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.

NEW QUESTION 2
- (Exam Topic 3)
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.

Answer: ACE

NEW QUESTION 3
- (Exam Topic 3)
You create an Azure Machine Learning workspace.
You need to detect data drift between a baseline dataset and a subsequent target dataset by using the DataDriftDetector class.
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.

```
from azureml.core import Workspace, Dataset
from datetime import datetime

ws = Workspace.from_config()
dset = Dataset.get_by_name(ws, 'target')
baseline = target.time_before(datetime(2021, 2, 1))
features = ['windAngle', 'windSpeed', 'temperature', 'stationName']

monitor = DataDriftDetector.                    (ws, 'drift-monitor', baseline,
target, compute_target='cpu-cluster', frequency='Week', feature_list=None,
drift_threshold=.6, latency=24)

monitor = DataDriftDetector.get_by_name(ws, 'drift-monitor')
monitor = monitor.update(feature_list=features)
complete = monitor.                (datetime(2021, 1, 1), datetime.today())
```

backfill
create_from_datasets
create_from_model

backfill
list
update

A. Mastered
B. Not Mastered

Answer: A

Explanation:
Graphical user interface, text, application, Word Description automatically generated
Box 1: create_from_datasets
The create_from_datasets method creates a new DataDriftDetector object from a baseline tabular dataset and a target time series dataset.
Box 2: backfill
The backfill method runs a backfill job over a given specified start and end date.
Syntax: backfill(start_date, end_date, compute_target=None, create_compute_target=False) Reference:
https://docs.microsoft.com/en-us/python/api/azureml-datadrift/azureml.datadrift.datadriftdetector(class)

NEW QUESTION 4
- (Exam Topic 3)
You train a model and register it in your Azure Machine Learning workspace. You are ready to deploy the model as a real-time web service.
You deploy the model to an Azure Kubernetes Service (AKS) inference cluster, but the deployment fails because an error occurs when the service runs the entry script that is associated with the model deployment.
You need to debug the error by iteratively modifying the code and reloading the service, without requiring a re-deployment of the service for each code update.
What should you do?

A. Register a new version of the model and update the entry script to load the new version of the model from its registered path.
B. Modify the AKS service deployment configuration to enable application insights and re-deploy to AKS.
C. Create an Azure Container Instances (ACI) web service deployment configuration and deploy the model on ACI.
D. Add a breakpoint to the first line of the entry script and redeploy the service to AKS.
E. Create a local web service deployment configuration and deploy the model to a local Docker container.

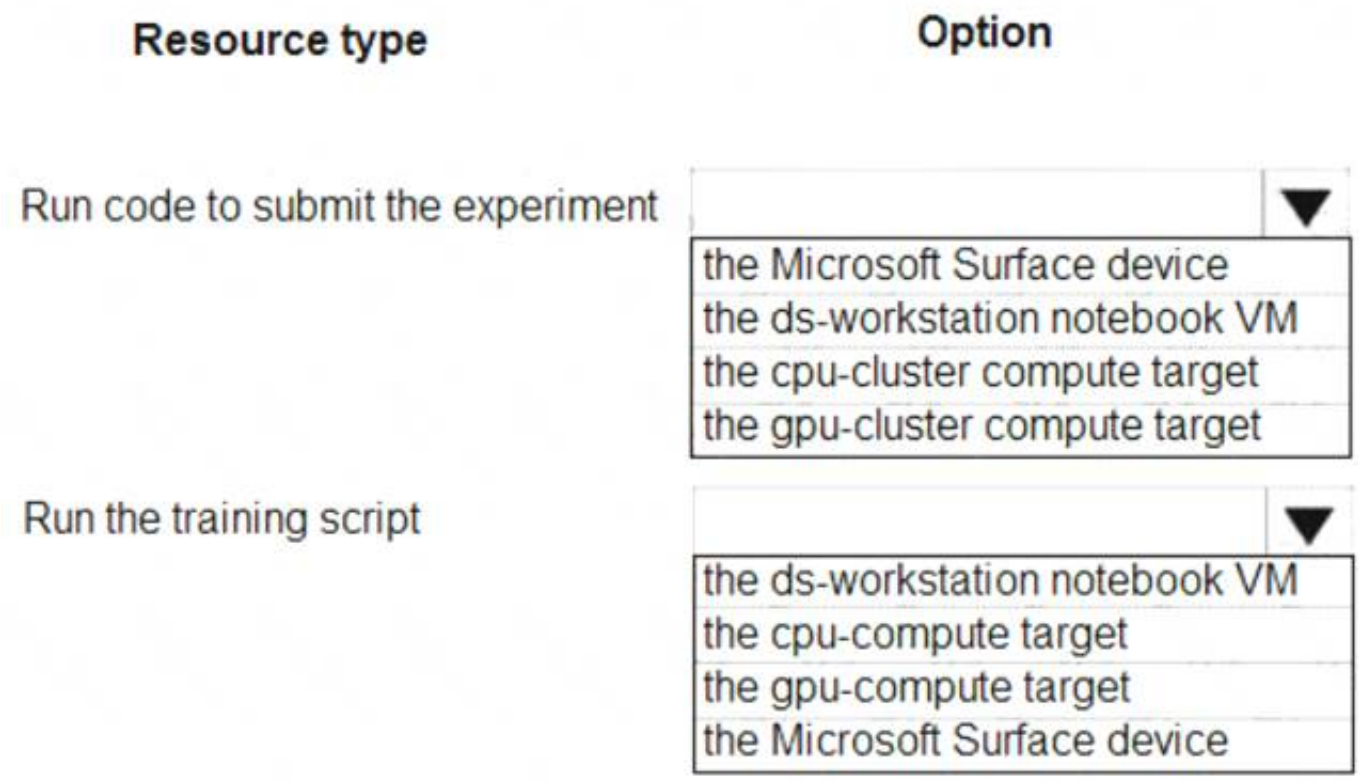
Answer: C

Explanation:
How to work around or solve common Docker deployment errors with Azure Container Instances (ACI) and Azure Kubernetes Service (AKS) using Azure Machine Learning.
The recommended and the most up to date approach for model deployment is via the Model.deploy() API using an Environment object as an input parameter. In this case our service will create a base docker image for you during deployment stage and mount the required models all in one call. The basic deployment tasks are:
* 1. Register the model in the workspace model registry.
* 2. Define Inference Configuration:
* a. Create an Environment object based on the dependencies you specify in the environment yaml file or use one of our procured environments.
* b. Create an inference configuration (InferenceConfig object) based on the environment and the scoring script.
* 3. Deploy the model to Azure Container Instance (ACI) service or to Azure Kubernetes Service (AKS).

NEW QUESTION 5
- (Exam Topic 3)
You are preparing to build a deep learning convolutional neural network model for image classification. You create a script to train the model using CUDA devices.
You must submit an experiment that runs this script in the Azure Machine Learning workspace. The following compute resources are available:

- a Microsoft Surface device on which Microsoft Office has been installed. Corporate IT policies prevent the installation of additional software
- a Compute Instance named ds-workstation in the workspace with 2 CPUs and 8 GB of memory
- an Azure Machine Learning compute target named cpu-cluster with eight CPU-based nodes
- an Azure Machine Learning compute target named gpu-cluster with four CPU and GPU-based nodes

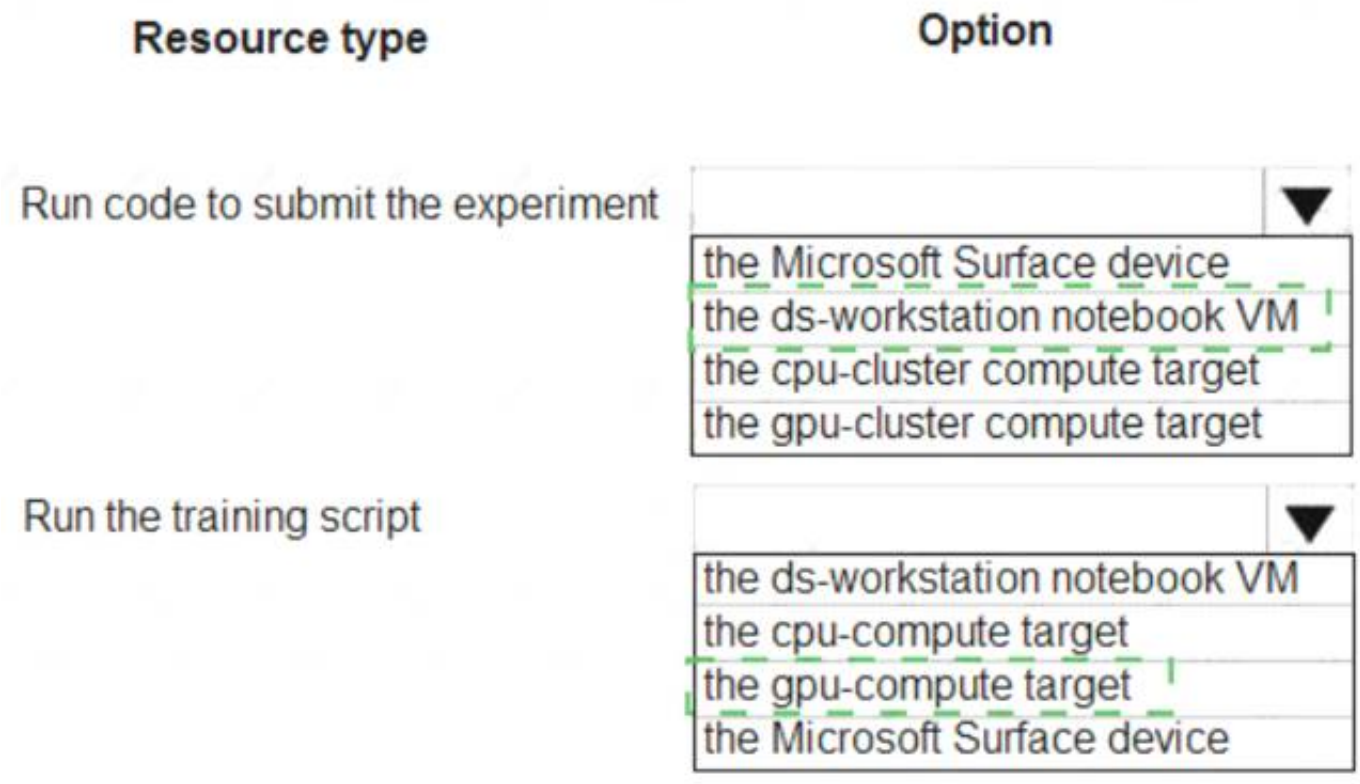
You need to specify the compute resources to be used for running the code to submit the experiment, and for running the script in order to minimize model training time.
Which resources should the data scientist use? To answer, select the appropriate options in the answer area.
NOTE: Each correct selection is worth one point.



A. Mastered
B. Not Mastered

Answer: A

Explanation:



NEW QUESTION 6
- (Exam Topic 3)
A set of CSV files contains sales records. All the CSV files have the same data schema.
Each CSV file contains the sales record for a particular month and has the filename sales.csv. Each file in stored in a folder that indicates the month and year when the data was recorded. The folders are in an Azure blob container for which a datastore has been defined in an Azure Machine Learning workspace. The folders are organized in a parent folder named sales to create the following hierarchical structure:

```
/sales
  /01-2019
    /sales.csv
  /02-2019
    /sales.csv
  /03-2019
    /sales.csv
  ...
```

At the end of each month, a new folder with that month's sales file is added to the sales folder.
You plan to use the sales data to train a machine learning model based on the following requirements:

- You must define a dataset that loads all of the sales data to date into a structure that can be easily converted to a dataframe.
- You must be able to create experiments that use only data that was created before a specific previous month, ignoring any data that was added after that month.
- You must register the minimum number of datasets possible.

You need to register the sales data as a dataset in Azure Machine Learning service workspace. What should you do?

A. Create a tabular dataset that references the datastore and explicitly specifies each 'sales/mm-yyyy/ sales.csv' file every mont
B. Register the dataset with the name sales_dataset each month, replacing the existing dataset and specifying a tag named month indicating the month and year it was registere
C. Use this dataset for all experiments.
D. Create a tabular dataset that references the datastore and specifies the path 'sales/*/sales.csv', register the dataset with the name sales_dataset and a tag named month indicating the month and year it was registered, and use this dataset for all experiments.
E. Create a new tabular dataset that references the datastore and explicitly specifies each 'sales/mm-yyyy/ sales.csv' file every mont
F. Register the dataset with the name sales_dataset_MM-YYYY each month with appropriate MM and YYYY values for the month and yea
G. Use the appropriate month-specific dataset for experiments.
H. Create a tabular dataset that references the datastore and explicitly specifies each 'sales/mm-yyyy/ sales.csv' fil
I. Register the dataset with the name sales_dataset each month as a new version and with a tag named month indicating the month and year it was registere
J. Use this dataset for all experiments, identifying the version to be used based on the month tag as necessary.

Answer: B

Explanation:
Specify the path. Example:
The following code gets the workspace existing workspace and the desired datastore by name. And then passes the datastore and file locations to the path parameter to create a new TabularDataset, weather_ds.
from azureml.core import Workspace, Datastore, Dataset datastore_name = 'your datastore name'
# get existing workspace
workspace = Workspace.from_config()
# retrieve an existing datastore in the workspace by name datastore = Datastore.get(workspace, datastore_name)
# create a TabularDataset from 3 file paths in datastore datastore_paths = [(datastore, 'weather/2018/11.csv'), (datastore, 'weather/2018/12.csv'), (datastore, 'weather/2019/*.csv')]
weather_ds = Dataset.Tabular.from_delimited_files(path=datastore_paths)

NEW QUESTION 7
- (Exam Topic 3)
You are training machine learning models in Azure Machine Learning. You use Hyperdrive to tune the hyperparameters. 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. Bandit
B. Median stopping
C. Default
D. Truncation selection

Answer: BC

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.truncationselectionpoli https://docs.microsoft.com/en-us/python/api/azureml-train-core/azureml.train.hyperdrive.banditpolicy

NEW QUESTION 8
- (Exam Topic 3)
You create a new Azure subscription. No resources are provisioned in the subscription. You need to create an Azure Machine Learning workspace.
What are three possible ways to achieve this goal? Each correct answer presents a complete solution. NOTE: Each correct selection is worth one point.

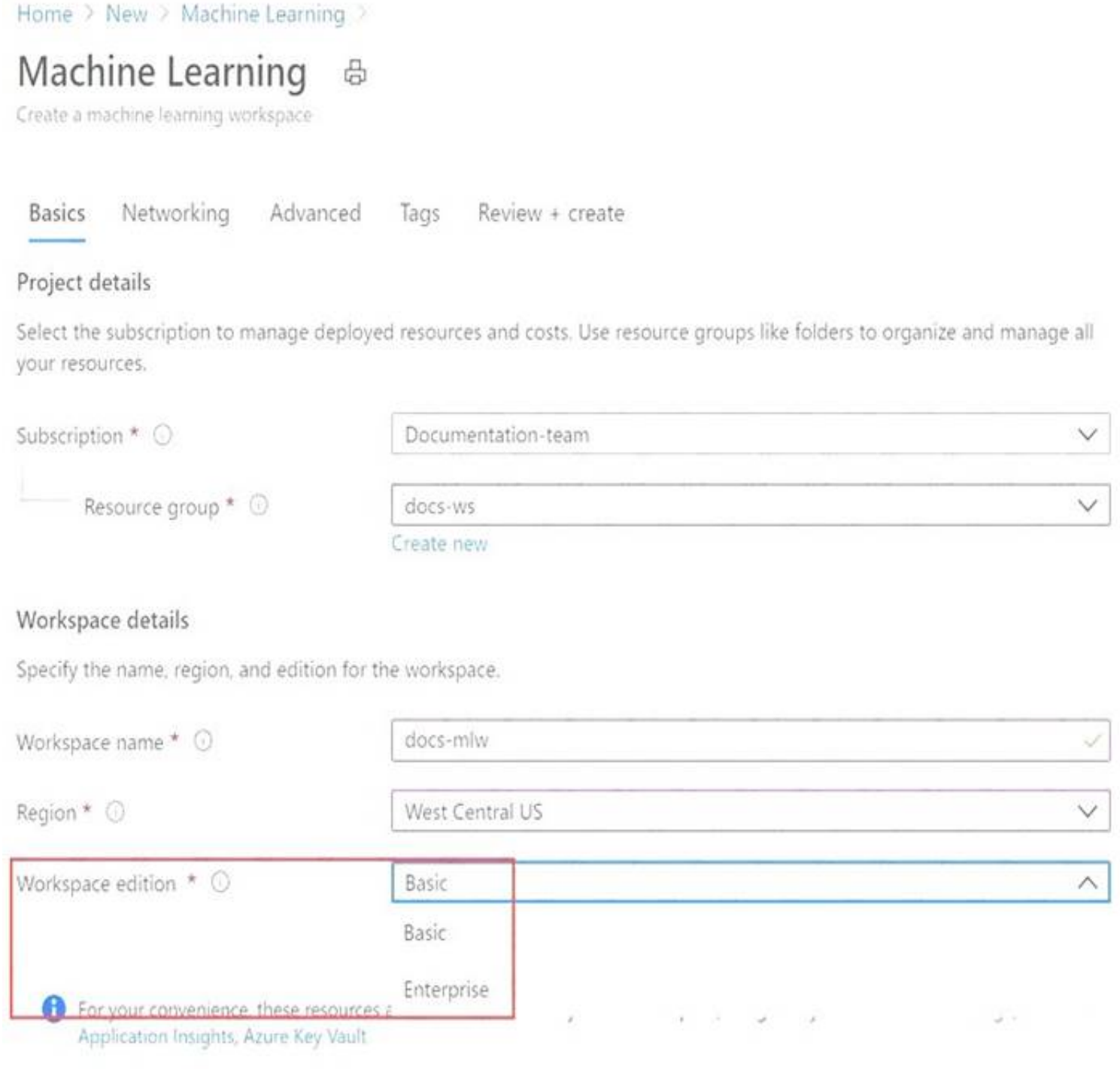
A. Run Python code that uses the Azure ML SDK library and calls the Workspace.create method with name, subscription_id, resource_group, and location parameters.
B. Use an Azure Resource Management template that includes a Microsoft.MachineLearningServices/ workspaces resource and its dependencies.
C. Use the Azure Command Line Interface (CLI) with the Azure Machine Learning extension to call the az group create function with --name and --location parameters, and then the az ml workspace create function, specifying –w and –g parameters for the workspace name and resource group.
D. Navigate to Azure Machine Learning studio and create a workspace.

E. Run Python code that uses the Azure ML SDK library and calls the Workspace.get method with name, subscription_id, and resource_group parameters.

Answer: BCD

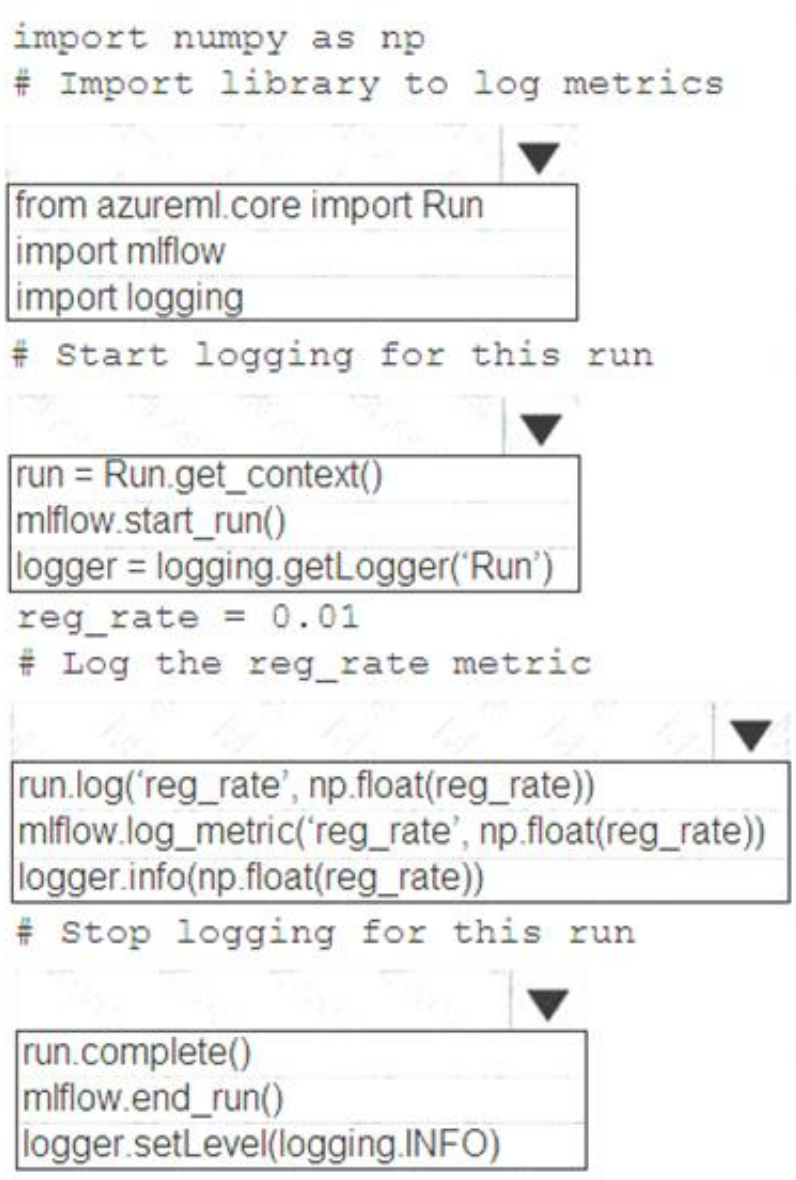
Explanation:
B: You can use an Azure Resource Manager template to create a workspace for Azure Machine Learning. Example:
{"type": "Microsoft.MachineLearningServices/workspaces",
…
C: You can create a workspace for Azure Machine Learning with Azure CLI Install the machine learning extension.
Create a resource group: az group create --name <resource-group-name> --location <location>
To create a new workspace where the services are automatically created, use the following command: az ml workspace create -w <workspace-name> -g <resource-group-name>
D: You can create and manage Azure Machine Learning workspaces in the Azure portal.

- Sign in to the Azure portal by using the credentials for your Azure subscription.
- In the upper-left corner of Azure portal, select + Create a resource.
- Use the search bar to find Machine Learning.
- Select Machine Learning.
- In the Machine Learning pane, select Create to begin.



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

NEW QUESTION 9
- (Exam Topic 3)
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.



A. Mastered
B. Not Mastered

Answer: A

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

NEW QUESTION 10
- (Exam Topic 3)
You are with a time series dataset in Azure Machine Learning Studio.
You need to split your dataset into training and testing subsets by using the Split Data module. Which splitting mode should you use?

A. Regular Expression Split
B. Split Rows with the Randomized split parameter set to true
C. Relative Expression Split
D. Recommender Split

Answer: B

Explanation:
Split Rows: Use this option if you just want to divide the data into two parts. You can specify the percentage of data to put in each split, but by default, the data is divided 50-50.
References:
https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/split-data

NEW QUESTION 10
- (Exam Topic 3)
You are creating a binary classification by using a two-class logistic regression model. You need to evaluate the model results for imbalance.
Which evaluation metric should you use?

A. Relative Absolute Error
B. AUC Curve
C. Mean Absolute Error
D. Relative Squared Error

Answer: B

Explanation:
One can inspect the true positive rate vs. the false positive rate in the Receiver Operating Characteristic (ROC) curve and the corresponding Area Under the Curve

(AUC) value. The closer this curve is to the upper left corner, the better the classifier's performance is (that is maximizing the true positive rate while minimizing the false positive rate). Curves that are close to the diagonal of the plot, result from classifiers that tend to make predictions that are close to random guessing.
References:
https://docs.microsoft.com/en-us/azure/machine-learning/studio/evaluate-model-performance#evaluating-a-bina

NEW QUESTION 12
- (Exam Topic 3)
You are moving a large dataset from Azure Machine Learning Studio to a Weka environment. You need to format the data for the Weka environment.
Which module should you use?

A. Convert to CSV
B. Convert to Dataset
C. Convert to ARFF
D. Convert to SVMLight

Answer: C

Explanation:
Use the Convert to ARFF module in Azure Machine Learning Studio, to convert datasets and results in Azure Machine Learning to the attribute-relation file format used by the Weka toolset. This format is known as ARFF.
The ARFF data specification for Weka supports multiple machine learning tasks, including data preprocessing, classification, and feature selection. In this format, data is organized by entites and their attributes, and is contained in a single text file.
References:
https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/convert-to-arff

NEW QUESTION 13
- (Exam Topic 3)
You deploy a model in Azure Container Instance.
You must use the Azure Machine Learning SDK to call the model API.
You need to invoke the deployed model using native SDK classes and methods.
How should you complete the command? To answer, select the appropriate options in the answer areas.
NOTE: Each correct selection is worth one point.

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

[▼]
- from azureml.core.webservice import requests
- from azureml.core.webservice import Webservice
- from azureml.core.webservice import LocalWebservice

```
import json
ws = Workspace.from_config()
service_name = "mlmodel1-service"
service = Webservice(name=service_name, workspace=ws)
x_new = [[2,101.5,1,24,21], [1,89.7,4,41,21]]
input_json = json.dumps({"data": x_new})
```

[▼]
- predictions = service.run(input_json)
- predictions = requests.post(service.scoring_uri, input_json)
- predictions = service.deserialize(ws, input_json)

A. Mastered
B. Not Mastered

Answer: A

Explanation:
Box 1: from azureml.core.webservice import Webservice
The following code shows how to use the SDK to update the model, environment, and entry script for a web service to Azure Container Instances:
from azureml.core import Environment
from azureml.core.webservice import Webservice
from azureml.core.model import Model, InferenceConfig Box 2: predictions = service.run(input_json)
Example: The following code demonstrates sending data to the service: import json
test_sample = json.dumps({'data': [[1, 2, 3, 4, 5, 6, 7, 8, 9, 10],
[10, 9, 8, 7, 6, 5, 4, 3, 2, 1]
]})
test_sample = bytes(test_sample, encoding='utf8') prediction = service.run(input_data=test_sample)
print(prediction) Reference:
https://docs.microsoft.com/bs-latn-ba/azure/machine-learning/how-to-deploy-azure-container-instance https://docs.microsoft.com/en-us/azure/machine-learning/how-to-troubleshoot-deployment

NEW QUESTION 15
- (Exam Topic 3)
You are conducting feature engineering to prepuce data for further analysis. The data includes seasonal patterns on inventory requirements.
You need to select the appropriate method to conduct feature engineering on the data. Which method should you use?

A. Exponential Smoothing (ETS) function.
B. One Class Support Vector Machine module
C. Time Series Anomaly Detection module
D. Finite Impulse Response (FIR) Filter module.

Answer: D

NEW QUESTION 18
- (Exam Topic 3)
You are preparing to use the Azure ML SDK to run an experiment and need to create compute. You run the following code:

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

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

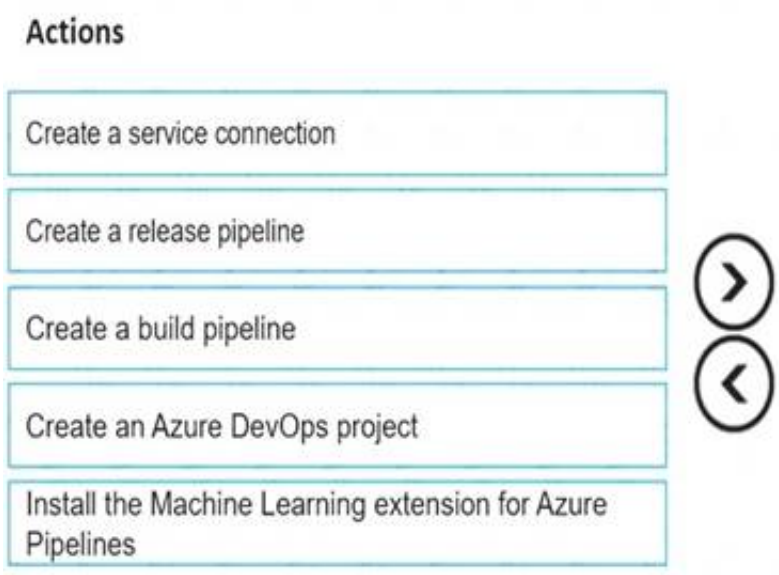
	Yes	No
If a training cluster named aml-cluster already exists in the workspace, it will be deleted and replaced.	○	○
The `wait_for_completion()` method will not return until the aml-cluster compute has four active nodes.	○	○
If the code creates a new aml-cluster compute target, it may be preempted due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

A. Mastered
B. Not Mastered

Answer: A

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

NEW QUESTION 21
- (Exam Topic 3)
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.

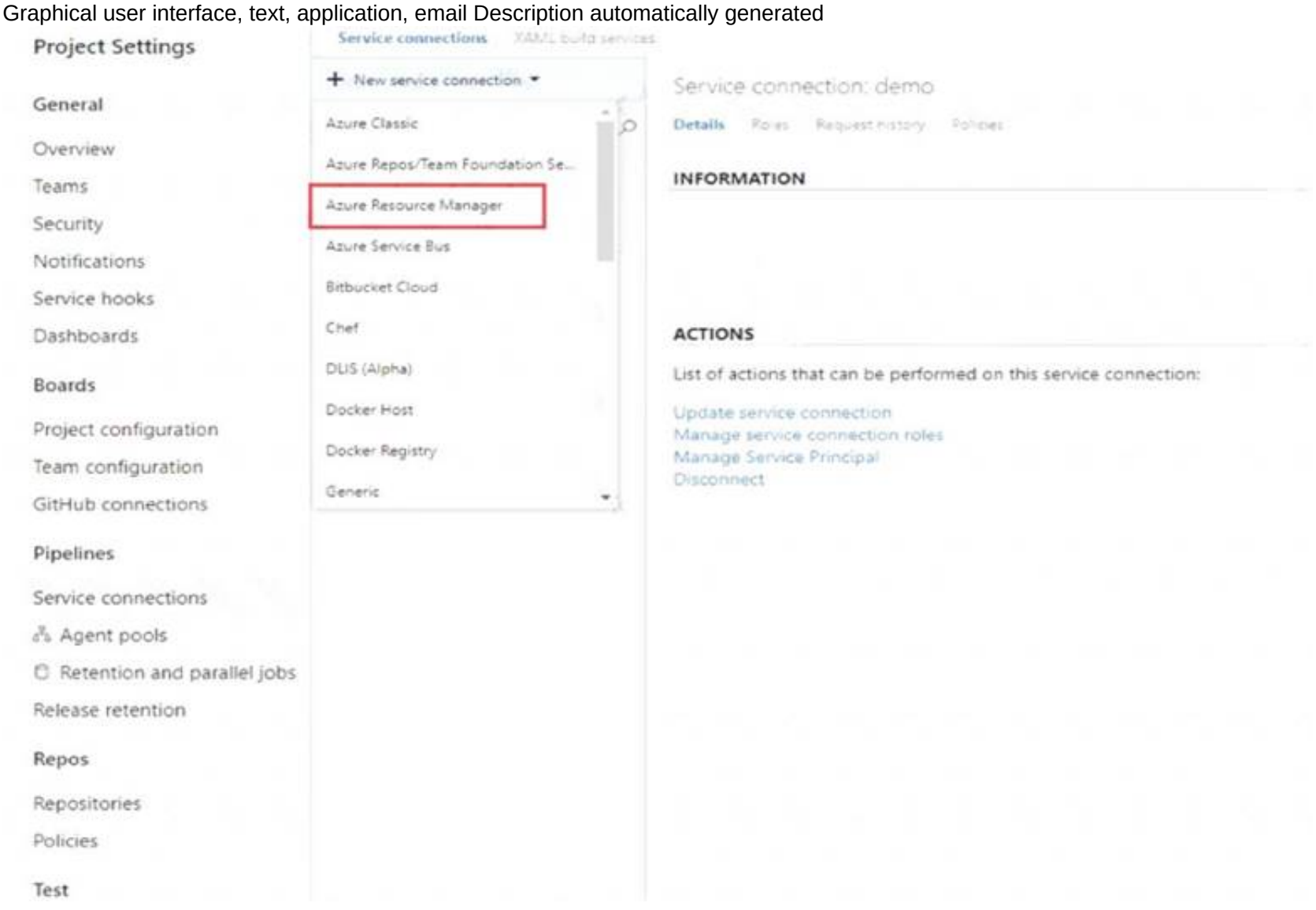


A. Mastered
B. Not Mastered

Answer: A

Explanation:
Graphical user interface, text, application, email Description automatically generated
Step 1: Create an Azure DevOps project
Step 2: Create a release pipeline
- Sign in to your Azure DevOps organization and navigate to your project.
- 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
Graphical user interface, text, application, email Description automatically generated



Select AzureMLWorkspace for the scope level, then fill in the following subsequent parameters. Graphical user interface, text, application Description automatically generated
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

NEW QUESTION 25
- (Exam Topic 3)
You use Azure Machine Learning to train and register a model.
You must deploy the model into production as a real-time web service to an inference cluster named service-compute that the IT department has created in the

Azure Machine Learning workspace.
Client applications consuming the deployed web service must be authenticated based on their Azure Active Directory service principal.
You need to write a script that uses the Azure Machine Learning SDK to deploy the model. The necessary modules have been imported.
How should you complete the code? To answer, select the appropriate options in the answer area.
NOTE: Each correct selection is worth one point.

```
# Assume the necessary modules have been imported
deploy_target = [▼] (ws, "service-compute")
```
AksCompute
AmlCompute
RemoteCompute
BatchCompute

```
deployment_config = [▼] .deploy_configuration(cpu_cores=1, memory_gb=1,
```
AksWebservice
AciWebservice
LocalWebService

```
[▼] )
```
token_auth_enabled=True
token_auth_enabled=False
auth_enabled=True
auth_enabled=False

```
service = Model.deploy(ws, "ml-service",
 [model], inference_config, deployment_config, deploy_target)
service.wait_for_deployment(show_output = True)
```

A. Mastered
B. Not Mastered

Answer: A

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

NEW QUESTION 26
- (Exam Topic 3)
A coworker registers a datastore in a Machine Learning services workspace by using the following code:

```
Datastore.register_azure_blob_container(workspace=ws,
 datastore_name='demo_datastore',
 container_name='demo_datacontainer',
 account_name='demo_account',
 account_key='0A0A0A-0A0A00A-0A00A0A0A0A0A',
 create_if_not_exists=True)
```

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

```
import azureml.core
from azureml.core import Workspace, Datastore
ws = Workspace.from_config()
datastore = [▼] .get( [▼] , ' [▼]
```
Workspace
Datastore
Experiment
Run

ws
run
experiment
log

demo_datastore
demo_datacontainer
demo_account
Datastore

A. Mastered
B. Not Mastered

Answer: A

Explanation:
Box 1: DataStore
To get a specific datastore registered in the current workspace, use the get() static method on the Datastore class:
# Get a named datastore from the current workspace
datastore = Datastore.get(ws, datastore_name='your datastore name') Box 2: ws
Box 3: demo_datastore Reference:
https://docs.microsoft.com/en-us/azure/machine-learning/how-to-access-data

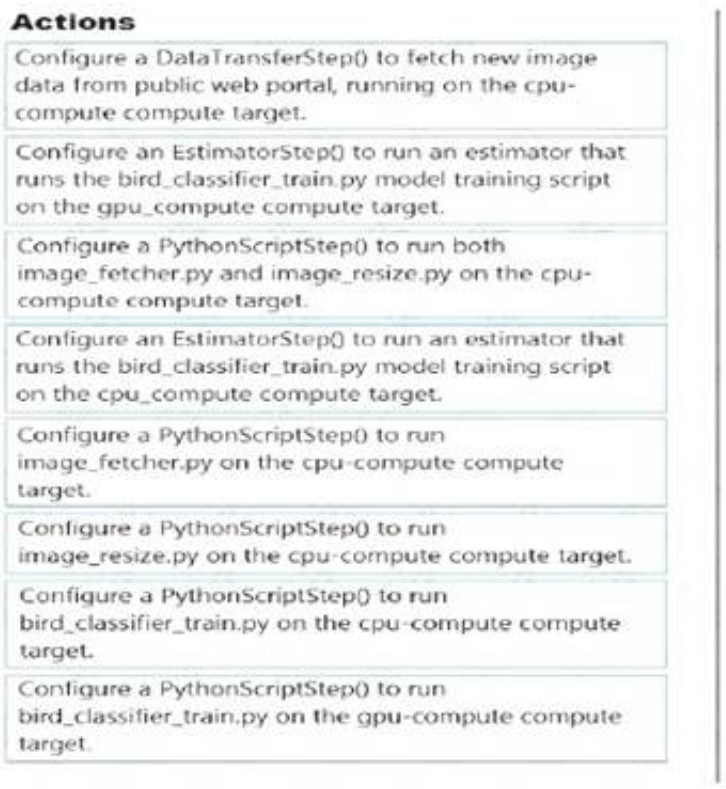
NEW QUESTION 28
- (Exam Topic 3)
You are building a machine learning model for translating English language textual content into French language textual content.
You need to build and train the machine learning model to learn the sequence of the textual content. Which type of neural network should you use?

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

Answer: C

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.
References:
https://towardsdatascience.com/language-translation-with-rnns-d84d43b40571

NEW QUESTION 32
- (Exam Topic 3)
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.



A. Mastered
B. Not Mastered

Answer: A

Explanation:
Step 1: Configure a DataTransferStep() to fetch new image data…
Step 2: Configure a PythonScriptStep() to run image_resize.y on the cpu-compute compute target. Step 3: Configure the EstimatorStep() to run training script on the gpu_compute computer target.
The PyTorch estimator provides a simple way of launching a PyTorch training job on a compute target. Reference:
https://docs.microsoft.com/en-us/azure/machine-learning/how-to-train-pytorch

NEW QUESTION 37
- (Exam Topic 3)
You create an Azure Machine Learning workspace named ML-workspace. You also create an Azure Databricks workspace named DB-workspace. DB-workspace contains a cluster named DB-cluster.
You must use DB-cluster to run experiments from notebooks that you import into DB-workspace.
You need to use ML-workspace to track MLflow metrics and artifacts generated by experiments running on DB-cluster. The solution must minimize the need for

custom code.
What should you do?

A. From DB-cluster, configure the Advanced Logging option.
B. From DB-workspac
C. configure the Link Azure ML workspace option.
D. From ML-workspac
E. create an attached compute.
F. From ML-workspac
G. create a compute cluster.

Answer: B

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

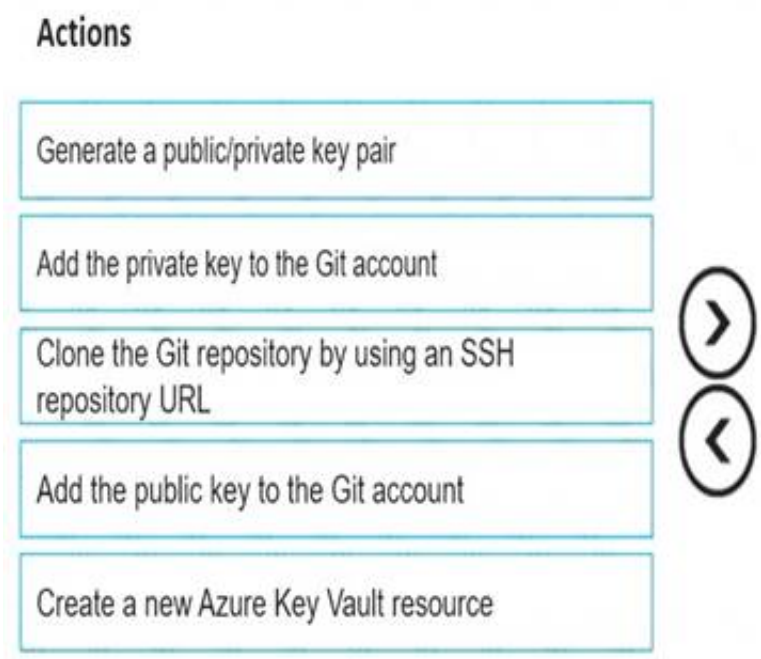
NEW QUESTION 42
- (Exam Topic 3)
You create and register a model in an Azure Machine Learning workspace.
You must use the Azure Machine Learning SDK to implement a batch inference pipeline that uses a ParallelRunStep to score input data using the model. You must specify a value for the ParallelRunConfig compute_target setting of the pipeline step.
You need to create the compute target. Which class should you use?

A. BatchCompute
B. AdlaCompute
C. AmlCompute
D. Aks Compute

Answer: C

Explanation:
Compute target to use for ParallelRunStep. This parameter may be specified as a compute target object or the string name of a compute target in the workspace.
The compute_target target is of AmlCompute or string.
Note: An Azure Machine Learning Compute (AmlCompute) is a managed-compute infrastructure that allows you to easily create a single or multi-node compute.
The compute is created within your workspace region as a resource that can be shared with other users
Reference:
https://docs.microsoft.com/en-us/python/api/azureml-contrib-pipeline-steps/azureml.contrib.pipeline.steps.parall https://docs.microsoft.com/en-us/python/api/azureml-core/azureml.core.compute.amlcompute(class)

NEW QUESTION 43
- (Exam Topic 3)
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.



A. Mastered
B. Not Mastered

Answer: A

Explanation:
Graphical user interface, text, application, chat or text message Description automatically generated
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

NEW QUESTION 45
- (Exam Topic 3)
You plan to create a speech recognition deep learning model. The model must support the latest version of Python.
You need to recommend a deep learning framework for speech recognition to include in the Data Science Virtual Machine (DSVM).
What should you recommend?

A. Apache Drill
B. Tensorflow
C. Rattle
D. Weka

Answer: B

Explanation:
TensorFlow is an open source library for numerical computation and large-scale machine learning. It uses Python to provide a convenient front-end API for building applications with the framework
TensorFlow can train and run deep neural networks for handwritten digit classification, image recognition, word embeddings, recurrent neural networks, sequence-to-sequence models for machine translation, natural language processing, and PDE (partial differential equation) based simulations.
References:
https://www.infoworld.com/article/3278008/what-is-tensorflow-the-machine-learning-library-explained.html

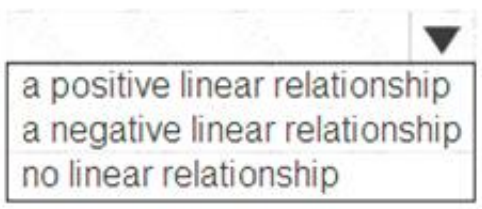
NEW QUESTION 50
- (Exam Topic 3)
You have a feature set containing the following numerical features: X, Y, and Z.
The Poisson correlation coefficient (r-value) of X, Y, and Z features is shown in the following image:

	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.

What is the r-value for the correlation of Y to Z?

Which type of relationship exists between Z and Y in the feature set?



A. Mastered
B. Not Mastered

Answer: A

Explanation:
Box 1: 0.859122
Box 2: a positively linear relationship
+1 indicates a strong positive linear relationship
-1 indicates a strong negative linear correlation
0 denotes no linear relationship between the two variables. References:
https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/compute-linear-correlation

NEW QUESTION 53
- (Exam Topic 3)
Note: This question is part of a series of questions that present the same scenario. Each question in the series contains a unique solution that might meet the stated goals. Some question sets might have more than one correct solution, while others might not have a correct solution.
After you answer a question in this section, you will NOT be able to return to it. As a result, these questions will not appear in the review screen.
You have a Python script named train.py in a local folder named scripts. The script trains a regression model by using scikit-learn. The script includes code to load a training data file which is also located in the scripts folder.
You must run the script as an Azure ML experiment on a compute cluster named aml-compute.
You need to configure the run to ensure that the environment includes the required packages for model training. You have instantiated a variable named aml-

compute that references the target compute cluster.
Solution: Run the following code:

```
from azureml.train.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

Answer: B

Explanation:
The scikit-learn estimator provides a simple way of launching a scikit-learn training job on a compute target. It is implemented through the SKLearn class, which can be used to support single-node CPU training.
Example:
from azureml.train.sklearn import SKLearn
}
estimator = SKLearn(source_directory=project_folder, compute_target=compute_target, entry_script='train_iris.py'
)
Reference:
https://docs.microsoft.com/en-us/azure/machine-learning/how-to-train-scikit-learn

NEW QUESTION 58
- (Exam Topic 3)
You need to select a pre built development environment for a series of data science experiments. You must use the R language for the experiments.
Which three environments can you use? Each correct answer presents a complete solution. NOTE: Each correct selection is worth one point.

A. MI.NET Library on a local environment
B. Azure Machine Learning Studio
C. Data Science Virtual Machine (OSVM)
D. Azure Data bricks
E. Azure Cognitive Services

Answer: ABD

NEW QUESTION 62
- (Exam Topic 3)
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.

```
from sklearn.decomposition import PCA
pca =
```
PCA()
PCA(n_components = 150)
PCA(n_components = 10)
PCA(n_components = 10000)

```
X_train=
```
pca
model
sklearn.decomposition

```
.fit_transform(X_train)
x_test = pca.
```
x_test
X_train
fit(x_test)
transform(x_test)

A. Mastered
B. Not Mastered

Answer: A

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

NEW QUESTION 65
- (Exam Topic 3)
You use the Azure Machine Learning Python SDK to define a pipeline to train a model.
The data used to train the model is read from a folder in a datastore.
You need to ensure the pipeline runs automatically whenever the data in the folder changes. What should you do?

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

Answer: D

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

NEW QUESTION 66
- (Exam Topic 3)
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.

Answer: BD

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. 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

NEW QUESTION 70
- (Exam Topic 3)
Note: This question is part of a series of questions that present the same scenario. Each question in the series contains a unique solution that might meet the stated goals. Some question sets might have more than one correct solution, while others might not have a correct solution.
After you answer a question in this section, you will NOT be able to return to it. As a result, these questions will not appear in the review screen.
You create a model to forecast weather conditions based on historical data.
You need to create a pipeline that runs a processing script to load data from a datastore and pass the processed data to a machine learning model training script.
Solution: Run the following code:

```
datastore = ws.get_default_datastore()
data_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?

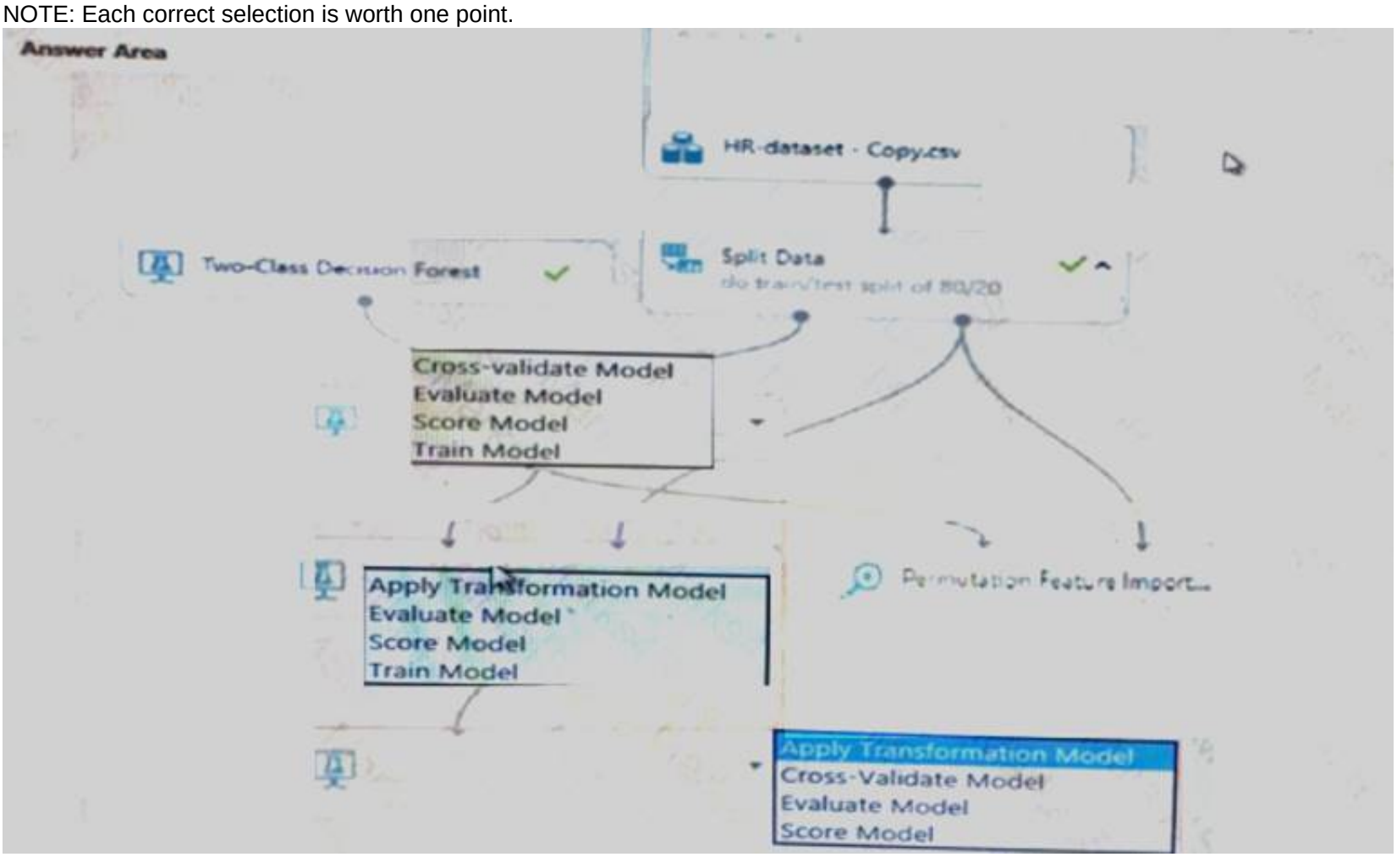
A. Yes

B. No

Answer: B

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=azu

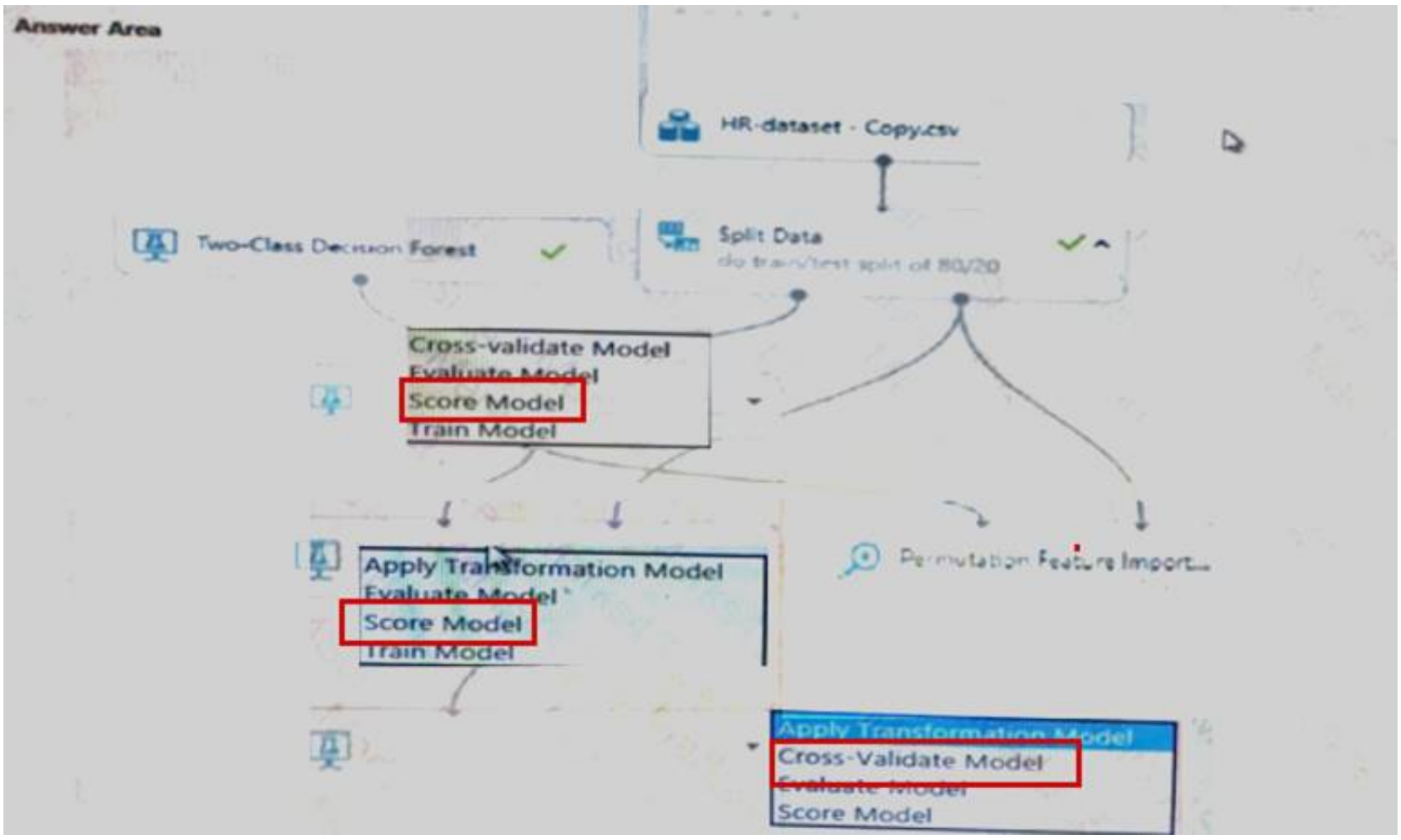
NEW QUESTION 75
- (Exam Topic 3)
You create a binary classification model using Azure Machine Learning Studio.
You must use a Receiver Operating Characteristic (RO C) curve and an F1 score to evaluate the model. You need to create the required business metrics.
How should you complete the experiment? To answer, select the appropriate options in the dialog box in the answer area.
NOTE: Each correct selection is worth one point.



A. Mastered
B. Not Mastered

Answer: A

Explanation:



NEW QUESTION 80
- (Exam Topic 3)
You plan to explore demographic data for home ownership in various cities. The data is in a CSV file with the following format:
age,city,income,home_owner 21,Chicago,50000,0 35,Seattle,120000,1 23,Seattle,65000,0 45,Seattle,130000,1 18,Chicago,48000,0
You need to run an experiment in your Azure Machine Learning workspace to explore the data and log the results. The experiment must log the following information:

- the number of observations in the dataset
- a box plot of income by home_owner
- a dictionary containing the city names and the average income for each city

You need to use the appropriate logging methods of the experiment's run object to log the required information.
How should you complete the code? To answer, drag the appropriate code segments to the correct locations. Each code segment may be used once, more than once, or not at all. You may need to drag the split bar between panes or scroll to view content.
NOTE: Each correct selection is worth one point.

Code segments

log
log_list
log_row
log_table
log_image

Answer Area

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

A. Mastered
B. Not Mastered

Answer: A

Explanation:
Box 1: log
The number of observations in the dataset. run.log(name, value, description='')
Scalar values: Log a numerical or string value to the run with the given name. Logging a metric to a run causes that metric to be stored in the run record in the experiment. You can log the same metric multiple times within a run, the result being considered a vector of that metric.
Example: run.log("accuracy", 0.95)
Box 2: log_image
A box plot of income by home_owner.
log_image Log an image to the run record. Use log_image to log a .PNG image file or a matplotlib plot to the run. These images will be visible and comparable in the run record.
Example: run.log_image("ROC", plot=plt) Box 3: log_table
A dictionary containing the city names and the average income for each city. log_table: Log a dictionary object to the run with the given name.

NEW QUESTION 84
- (Exam Topic 3)
You have a Python data frame named salesData in the following format:

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

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

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

You need to use the pandas.melt() function in Python to perform the transformation.
How should you complete the code segment? To answer, select the appropriate options in the answer area. NOTE: Each correct selection is worth one point.
Answer Area

```
import pandas as pd
salesData = pd.melt(  [dataFrame / pandas / salesData / year]  ,id_vars='  [shop / year / value / Shop X, Shop Y, Shop Z]  ',value_vars=  ['shop' / 'year' / ['year'] / ['2017', '2018']]  )
```

A. Mastered
B. Not Mastered

Answer: A

Explanation:
Box 1: dataFrame
Syntax: pandas.melt(frame, id_vars=None, value_vars=None, var_name=None, value_name='value', col_level=None)[source]
Where frame is a DataFrame Box 2: shop
Paramter id_vars id_vars : tuple, list, or ndarray, optional Column(s) to use as identifier variables.
Box 3: ['2017','2018']
value_vars : tuple, list, or ndarray, optional
Column(s) to unpivot. If not specified, uses all columns that are not set as id_vars. Example:
df = pd.DataFrame({'A': {0: 'a', 1: 'b', 2: 'c'},
'B': {0: 1, 1: 3, 2: 5},
'C': {0: 2, 1: 4, 2: 6}})
pd.melt(df, id_vars=['A'], value_vars=['B', 'C']) A variable value
0 a B 1
1 b B 3
2 c B 5
3 a C 2
4 b C 4
5 c C 6
References:
https://pandas.pydata.org/pandas-docs/stable/reference/api/pandas.melt.html

NEW QUESTION 87
- (Exam Topic 3)
Note: This question is part of a series of questions that present the same scenario. Each question in the series contains a unique solution that might meet the stated goals. Some question sets might have more than one correct solution, while others might not have a correct solution.
After you answer a question in this section, you will NOT be able to return to it. As a result, these questions will not appear in the review screen.
You are using Azure Machine Learning to run an experiment that trains a classification model.

You want to use Hyperdrive to find parameters that optimize the AUC metric for the model. You configure a HyperDriveConfig for the experiment by running the following code:

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

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

Answer: A

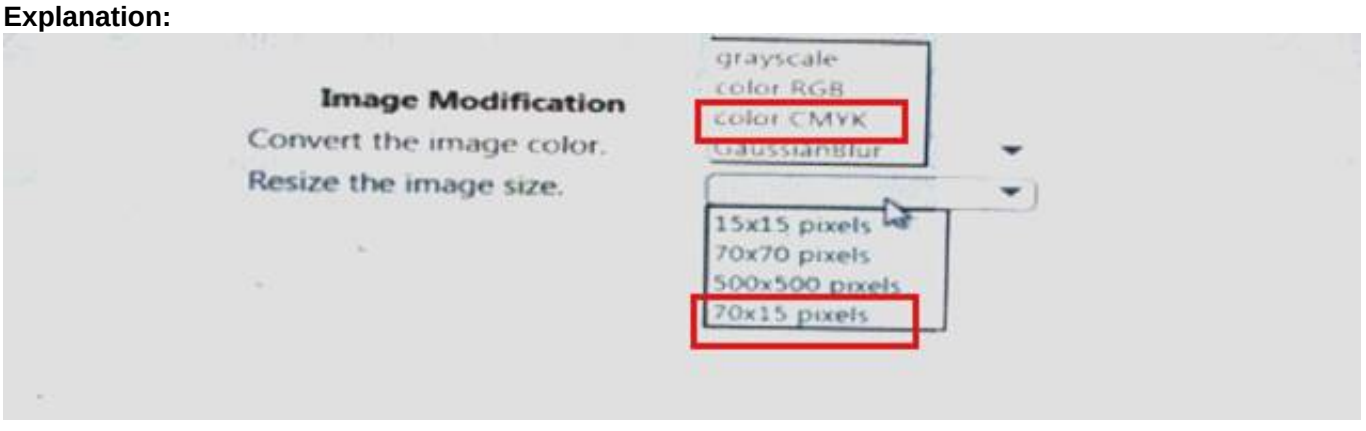
Explanation:
Python printing/logging example: logging.info(message)
Destination: Driver logs, Azure Machine Learning designer
Reference:
https://docs.microsoft.com/en-us/azure/machine-learning/how-to-debug-pipelines

NEW QUESTION 90
- (Exam Topic 3)
You arc I mating a deep learning model to identify cats and dogs. You have 25,000 color images. You must meet the following requirements:
• Reduce the number of training epochs.
• Reduce the size of the neural network.
• Reduce over-fitting of the neural network.
You need to select the image modification values.
Which value should you use? To answer, select the appropriate Options in the answer area. NOTE: Each correct selection is worth one point.

A. Mastered
B. Not Mastered

Answer: A

Explanation:



NEW QUESTION 94
- (Exam Topic 3)
You define a datastore named ml-data for an Azure Storage blob container. In the container, you have a folder named train that contains a file named data.csv.
You plan to use the file to train a model by using the Azure Machine Learning SDK.
You plan to train the model by using the Azure Machine Learning SDK to run an experiment on local compute.
You define a DataReference object by running the following code:

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

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

A.
```
import os
import argparse
import pandas as pd

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

B.
```
import os
import argparse
import pandas as pd

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

C.
```
import pandas as pd

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

D.
```
import os
import argparse
import pandas as pd

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

E.
```
import os
import argparse
import pandas as pd

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

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

Answer: E

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

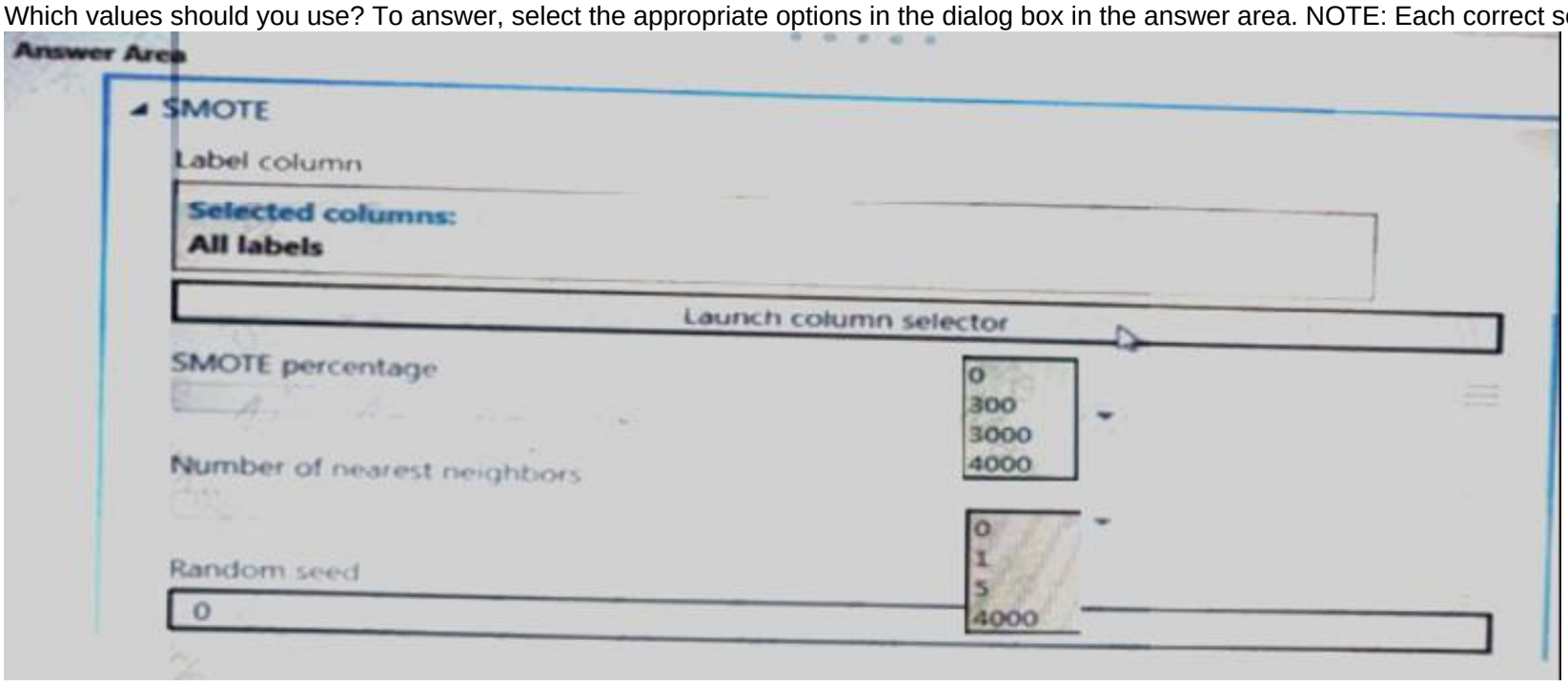
NEW QUESTION 97
- (Exam Topic 3)
You use the following code to define the steps for a pipeline: from azureml.core import Workspace, Experiment, Run from azureml.pipeline.core import Pipeline from azureml.pipeline.steps import PythonScriptStep ws = Workspace.from_config()
. . .
step1 = PythonScriptStep(name="step1", ...) step2 = PythonScriptsStep(name="step2", ...) pipeline_steps = [step1, step2]
You need to add code to run the steps.
Which two code segments can you use to achieve this goal? Each correct answer presents a complete solution.
NOTE: Each correct selection is worth one point.

A. experiment = Experiment(workspace=ws, name='pipeline-experiment')run = experiment.submit(config=pipeline_steps)
B. run = Run(pipeline_steps)
C. pipeline = Pipeline(workspace=ws, steps=pipeline_steps) experiment = Experiment(workspace=ws, name='pipeline-experiment')run = experiment.submit(pipeline)
D. pipeline = Pipeline(workspace=ws, steps=pipeline_steps)run = pipeline.submit(experiment_name='pipeline-experiment')

Answer: CD

Explanation:
After you define your steps, you build the pipeline by using some or all of those steps.
# Build the pipeline. Example:
pipeline1 = Pipeline(workspace=ws, steps=[compare_models])
# Submit the pipeline to be run
pipeline_run1 = Experiment(ws, 'Compare_Models_Exp').submit(pipeline1) Reference:
https://docs.microsoft.com/en-us/azure/machine-learning/how-to-create-machine-learning-pipelines

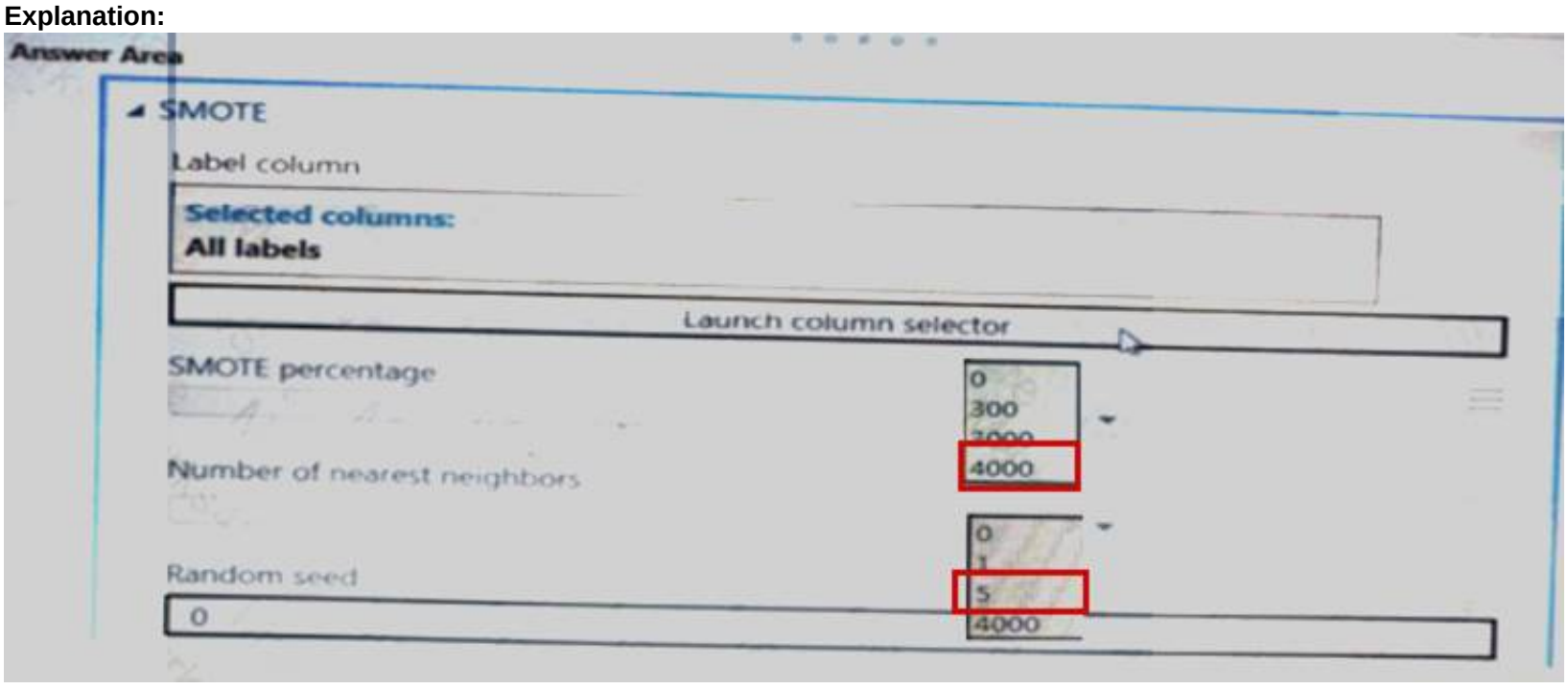
NEW QUESTION 99
- (Exam Topic 3)
You create an experiment in Azure Machine Learning Studio- You add a training dataset that contains 10.000 rows. The first 9.000 rows represent class 0 (90 percent). The first 1.000 rows represent class 1 (10 percent).
The training set is unbalanced between two Classes. You must increase the number of training examples for class 1 to 4,000 by using data rows. You add the Synthetic Minority Oversampling Technique (SMOTE) module to the experiment.
You need to configure the module.
Which values should you use? To answer, select the appropriate options in the dialog box in the answer area. NOTE: Each correct selection is worth one point.



A. Mastered
B. Not Mastered

Answer: A

Explanation:



NEW QUESTION 102
- (Exam Topic 3)
You arc creating a new experiment in Azure Machine Learning Studio. You have a small dataset that has missing values in many columns. The data does not require the application of predictors for each column. You plan to use the Clean Missing Data module to handle the missing data.
You need to select a data cleaning method. Which method should you use?

A. Synthetic Minority
B. Replace using Probabilistic PAC
C. Replace using MICE
D. Normalization

Answer: B

NEW QUESTION 104
- (Exam Topic 3)
You develop and train a machine learning model to predict fraudulent transactions for a hotel booking website. Traffic to the site varies considerably. The site experiences heavy traffic on Monday and Friday and much lower traffic on other days. Holidays are also high web traffic days. You need to deploy the model as an Azure Machine Learning real-time web service endpoint on compute that can dynamically scale up and down to support demand. Which deployment compute option should you use?

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

Answer: D

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

NEW QUESTION 106
- (Exam Topic 3)
You are performing sentiment analysis using a CSV file that includes 12,000 customer reviews written in a short sentence format. You add the CSV file to Azure Machine Learning Studio and configure it as the starting point dataset of an experiment. You add the Extract N-Gram Features from Text module to the experiment to extract key phrases from the customer review column in the dataset.
You must create a new n-gram dictionary from the customer review text and set the maximum n-gram size to trigrams.
What should you select? To answer, select the appropriate options in the answer area.
NOTE: Each correct selection is worth one point.

A. Mastered
B. Not Mastered

Answer: A

Explanation:
Vocabulary mode: Create
For Vocabulary mode, select Create to indicate that you are creating a new list of n-gram features. N-Grams size: 3
For N-Grams size, type a number that indicates the maximum size of the n-grams to extract and store. For example, if you type 3, unigrams, bigrams, and trigrams will be created.
Weighting function: Leave blank
The option, Weighting function, is required only if you merge or update vocabularies. It specifies how terms in the two vocabularies and their scores should be weighted against each other.
References:
https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/extract-n-gram-features-from

NEW QUESTION 110
- (Exam Topic 3)
You are building recurrent neural network to perform a binary classification.
The training loss, validation loss, training accuracy, and validation accuracy of each training epoch has been provided. You need to identify whether the classification model is over fitted.
Which of the following is correct?

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

Answer: B

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.
References:
https://machinelearningmastery.com/diagnose-overfitting-underfitting-lstm-models/

NEW QUESTION 111
- (Exam Topic 3)
You are using Azure Machine Learning to train machine learning models. You need a compute target on which to remotely run the training script. You run the following Python code:

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

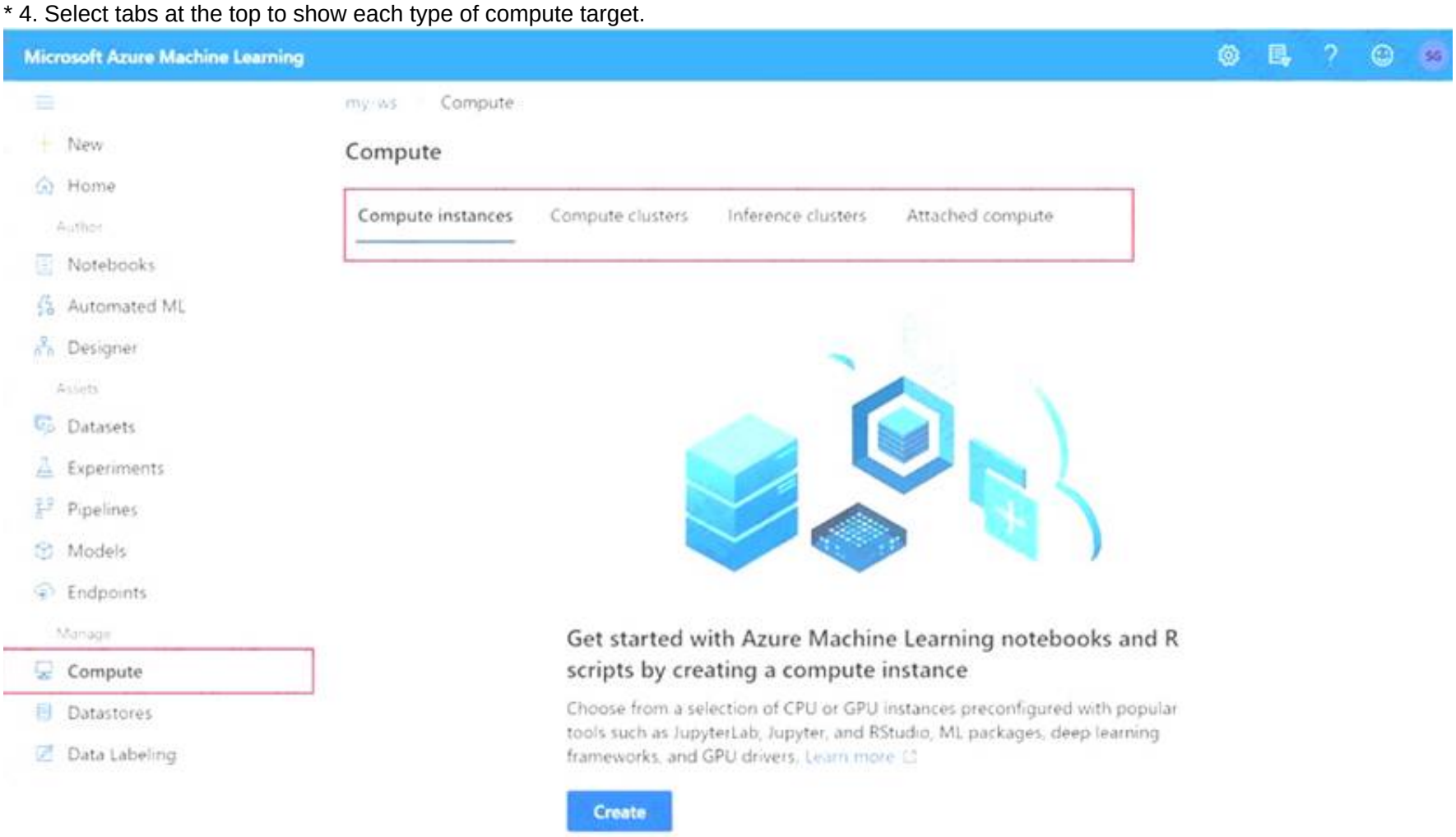
Answer Area

	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.	○	○

A. Mastered
B. Not Mastered

Answer: A

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



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

NEW QUESTION 113
- (Exam Topic 3)
An organization uses Azure Machine Learning service and wants to expand their use of machine learning. You have the following compute environments. The organization does not want to create another compute
environment.

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

You need to determine which compute environment to use for the following scenarios.
Which compute types should you use? To answer, drag the appropriate compute environments to the correct scenarios. Each compute environment may be used once, more than once, or not at all. You may need to drag the split bar between panes or scroll to view content.
NOTE: Each correct selection is worth one point.

Environments

nb_server
aks_cluster
mlc_cluster

Answer Area

Scenario	Environment
Run an Azure Machine Learning Designer training pipeline.	Environment
Deploying a web service from the Azure Machine Learning designer.	Environment

A. Mastered
B. Not Mastered

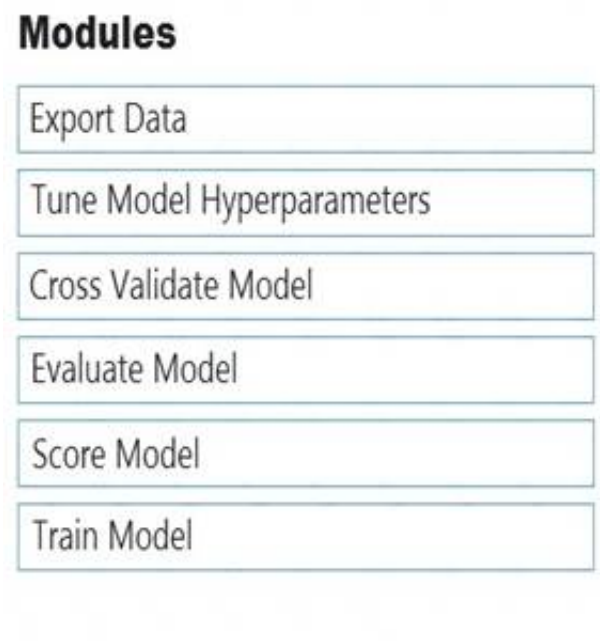
Answer: A

Explanation:
Box 1: nb_server

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

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

NEW QUESTION 116
- (Exam Topic 3)
You are building an experiment using the Azure Machine Learning designer.
You split a dataset into training and testing sets. You select the Two-Class Boosted Decision Tree as the algorithm.
You need to determine the Area Under the Curve (AUC) of the model.
Which three modules should you use in sequence? To answer, move the appropriate modules from the list of modules to the answer area and arrange them in the correct order.

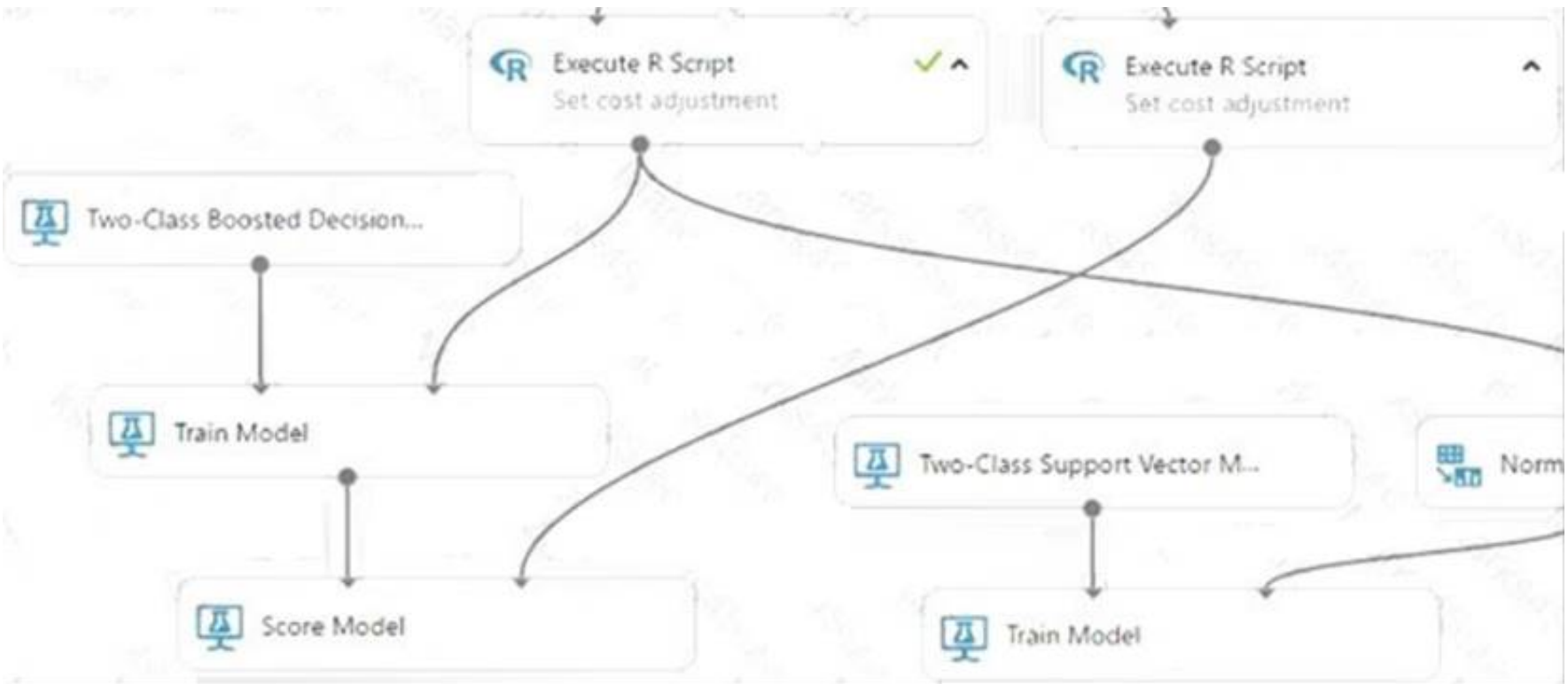


A. Mastered
B. Not Mastered

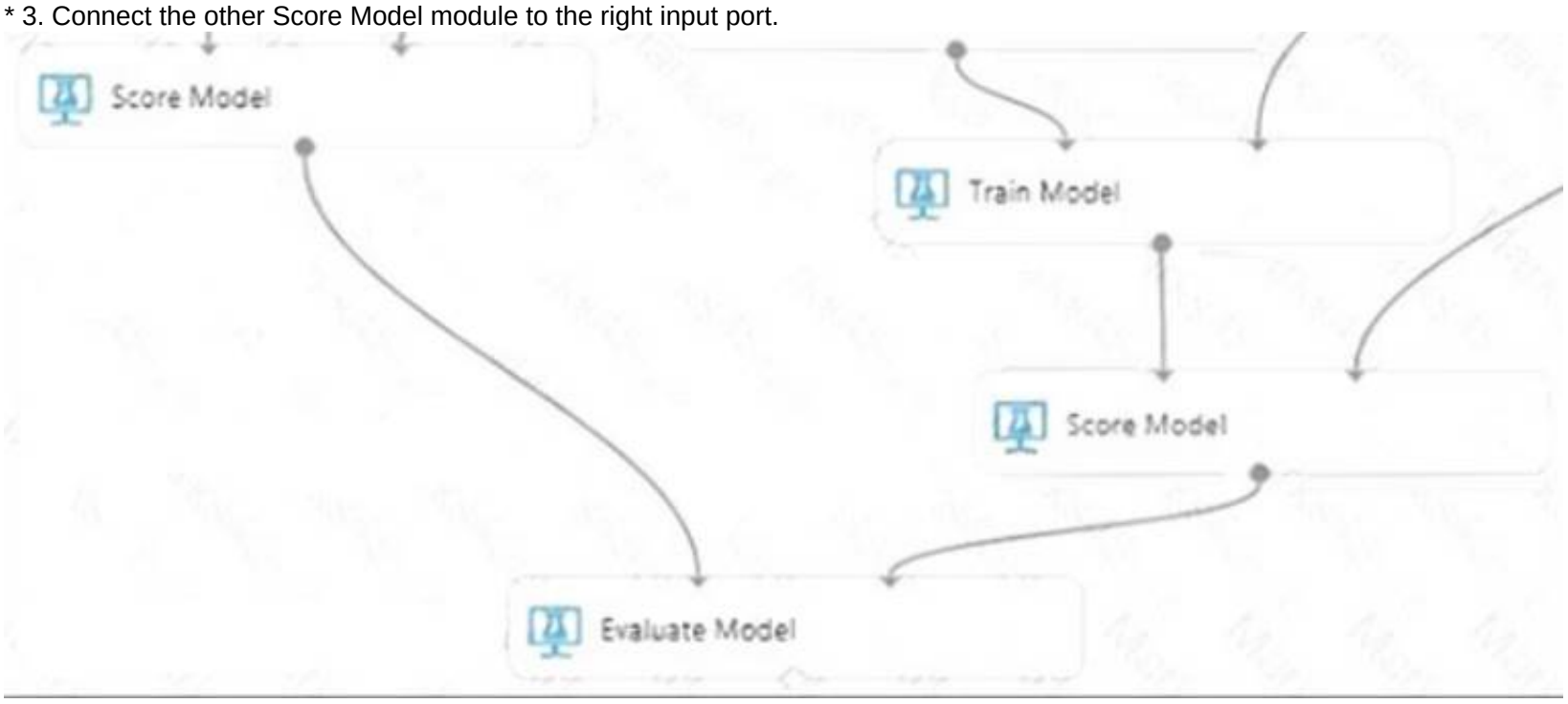
Answer: A

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

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



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



NEW QUESTION 120
- (Exam Topic 3)
You configure a Deep Learning Virtual Machine for Windows.
You need to recommend tools and frameworks to perform the following:
- Build deep neural network (DNN) models
- Perform interactive data exploration and visualization

Which tools and frameworks should you recommend? To answer, drag the appropriate tools to the correct tasks. Each tool 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.

Tools

Vowpal Wabbit

PowerBI Desktop

Azure Data Factory

Microsoft Cognitive Toolkit

Answer Area

Task	Tool
Build DNN models	Tool
Enable interactive data exploration and visualization	Tool

A. Mastered
B. Not Mastered

Answer: A

Explanation:
Box 1: Vowpal Wabbit

Use the Train Vowpal Wabbit Version 8 module in Azure Machine Learning Studio (classic), to create a machine learning model by using Vowpal Wabbit.
Box 2: PowerBI Desktop
Power BI Desktop is a powerful visual data exploration and interactive reporting tool
BI is a name given to a modern approach to business decision making in which users are empowered to find, explore, and share insights from data across the enterprise.
References:
https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/train-vowpal-wabbit-version- https://docs.microsoft.com/en-us/azure/architecture/data-guide/scenarios/interactive-data-exploration

NEW QUESTION 123
- (Exam Topic 3)
You are creating a machine learning model in Python. The provided dataset contains several numerical columns and one text column. The text column represents a product's category. The product category will always be one of the following:

- Bikes
- Cars
- Vans
- Boats

You are building a regression model using the scikit-learn Python package.
You need to transform the text data to be compatible with the scikit-learn Python package.
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.

A. Mastered
B. Not Mastered

Answer: A

Explanation:
Box 1: pandas as df
Pandas takes data (like a CSV or TSV file, or a SQL database) and creates a Python object with rows and columns called data frame that looks very similar to table in a statistical software (think Excel or SPSS for example.
Box 2: transpose[ProductCategoryMapping] Reshape the data from the pandas Series to columns. Reference:
https://datascienceplus.com/linear-regression-in-python/

NEW QUESTION 126
- (Exam Topic 3)

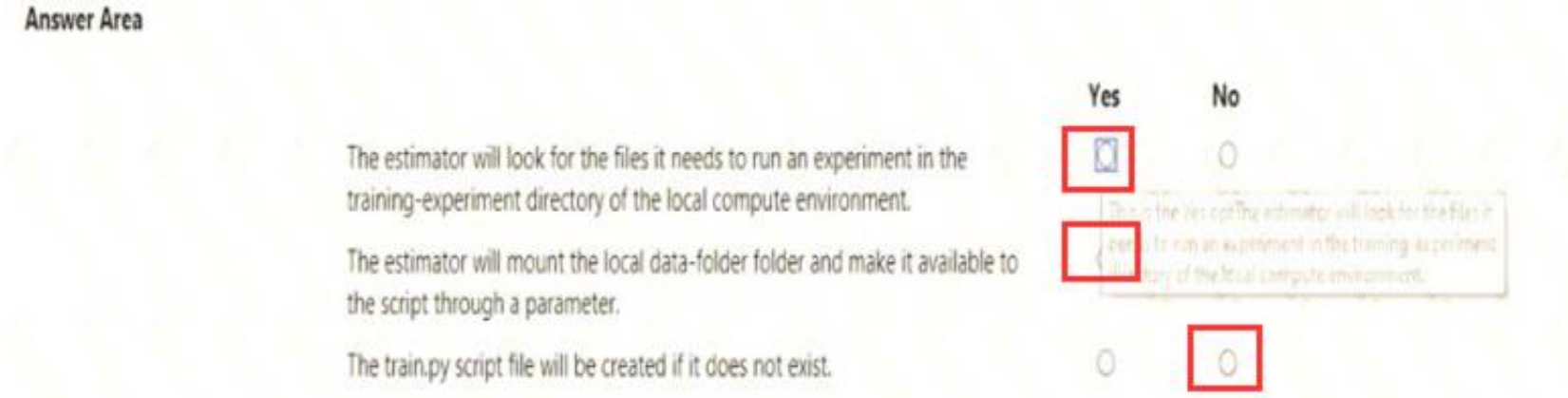
```
train_cluster = ComputeTarget(workspace=work_space, name='train-cluster')
estimator = Estimator(source_directory =
 'training-experiment',
 script_params = {'--data-folder': data_source.as_mount(), '--regularization': 0.8},
 compute_target = train_cluster,
 entry_script = 'train.py',
 conda_packages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

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

A. Mastered
B. Not Mastered

Answer: A

Explanation:



NEW QUESTION 129
- (Exam Topic 3)
Note: This question is part of a series of questions that present the same scenario. Each question in the series contains a unique solution that might meet the stated goals. Some question sets might have more than one correct solution, while others might not have a correct solution.
After you answer a question in this section, you will NOT be able to return to it. As a result, these questions will not appear in the review screen.
You are a data scientist using Azure Machine Learning Studio.
You need to normalize values to produce an output column into bins to predict a target column. Solution: Apply a Quantiles normalization with a QuantileIndex normalization.
Does the solution meet the GOAL?

A. Yes
B. No

Answer: B

Explanation:
Use the Entropy MDL binning mode which has a target column. References:
https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/group-data-into-bins

NEW QUESTION 130
- (Exam Topic 3)
You use the Azure Machine Learning designer to create and run a training pipeline. You then create a real-time inference pipeline.
You must deploy the real-time inference pipeline as a web service.
What must you do before you deploy the real-time inference pipeline?

A. Run the real-time inference pipeline.
B. Create a batch inference pipeline.
C. Clone the training pipeline.
D. Create an Azure Machine Learning compute cluster.

Answer: D

Explanation:
You need to create an inferencing cluster. Deploy the real-time endpoint
After your AKS service has finished provisioning, return to the real-time inferencing pipeline to complete deployment.
- Select Deploy above the canvas.
- Select Deploy new real-time endpoint.
- Select the AKS cluster you created.
- Select Deploy. Reference:

https://docs.microsoft.com/en-us/azure/machine-learning/tutorial-designer-automobile-price-deploy

NEW QUESTION 131
- (Exam Topic 3)
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?

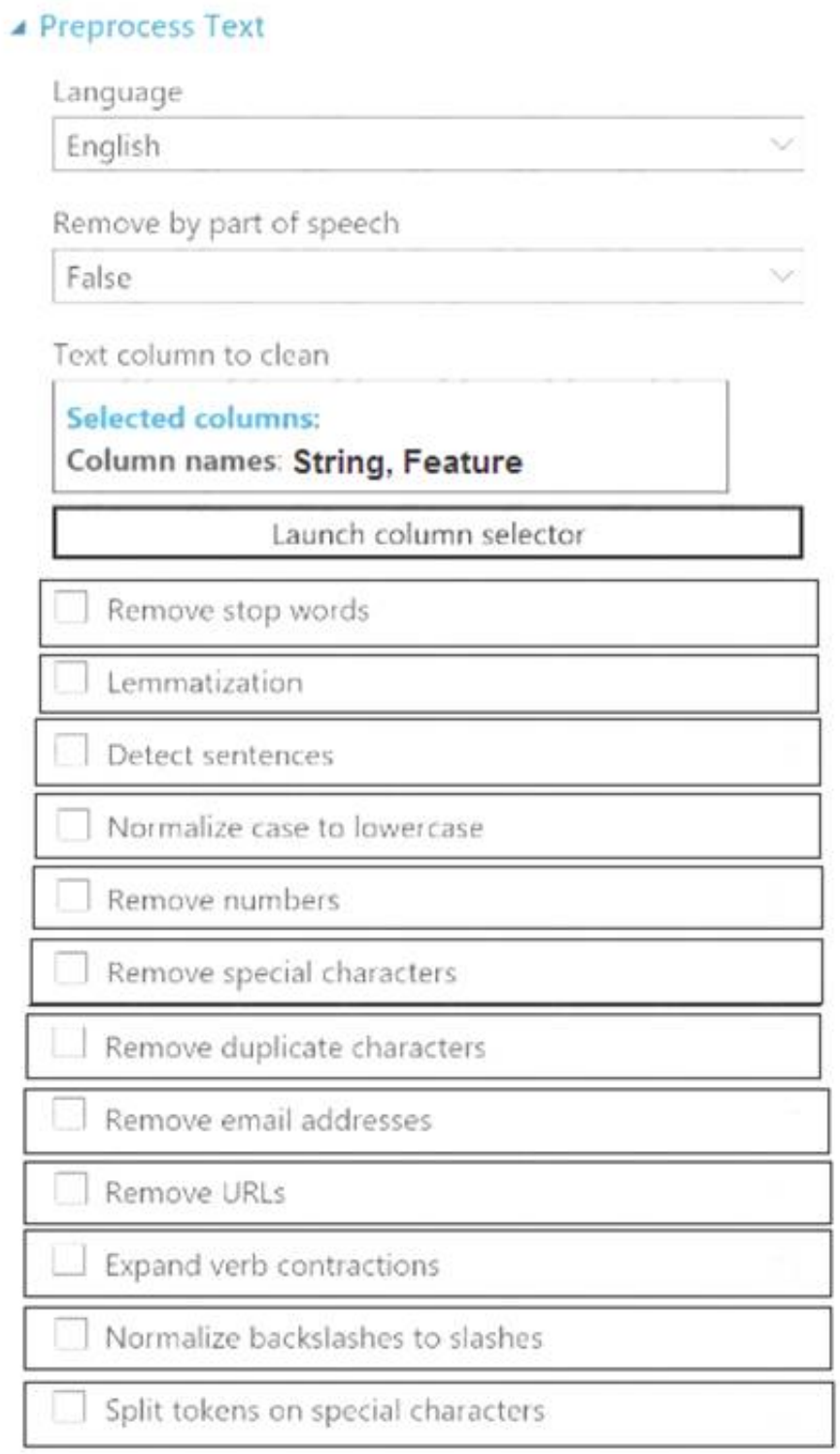
A. Edit Metadata
B. Preprocess Text
C. Execute Python Script
D. Latent Dirichlet Allocation

Answer: A

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

NEW QUESTION 136
- (Exam Topic 3)
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.



A. Mastered
B. Not Mastered

Answer: A

Explanation:
Box 1: Remove stop words
Remove words to optimize information retrieval.

Remove stop words: Select this option if you want to apply a predefined stopword list to the text column. Stop word removal is performed before any other processes.
Box 2: Lemmatization
Ensure that multiple related words from a single canonical form. Lemmatization converts multiple related words to a single canonical form Box 3: Remove special characters
Remove special characters: Use this option to replace any non-alphanumeric special characters with the pipe | character.
References:
https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/preprocess-text

NEW QUESTION 138
- (Exam Topic 3)
Note: This question is part of a series of questions that present the same scenario. Each question in the series contains a unique solution that might meet the stated goals. Some question sets might have more than one correct solution, while others might not have a correct solution.
After you answer a question in this section, you will NOT be able to return to it. As a result, these questions will not appear in the review screen.
You are using Azure Machine Learning Studio to perform feature engineering on a dataset. You need to normalize values to produce a feature column grouped into bins.
Solution: Apply an Entropy Minimum Description Length (MDL) binning mode. Does the solution meet the goal?

A. Yes
B. No

Answer: A

Explanation:
Entropy MDL binning mode: This method requires that you select the column you want to predict and the column or columns that you want to group into bins. It then makes a pass over the data and attempts to determine the number of bins that minimizes the entropy. In other words, it chooses a number of bins that allows the data column to best predict the target column. It then returns the bin number associated with each row of your data in a column named <colname>quantized.
References:
https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/group-data-into-bins

NEW QUESTION 139
- (Exam Topic 3)
You use Azure Machine Learning designer to create a real-time service endpoint. You have a single Azure Machine Learning service compute resource. You train the model and prepare the real-time pipeline for deployment You need to publish the inference pipeline as a web service. Which compute type should you use?

A. HDInsight
B. Azure Databricks
C. Azure Kubernetes Services
D. the existing Machine Learning Compute resource
E. a new Machine Learning Compute resource

Answer: C

Explanation:
Azure Kubernetes Service (AKS) can be used real-time inference. Reference:
https://docs.microsoft.com/en-us/azure/machine-learning/concept-compute-target

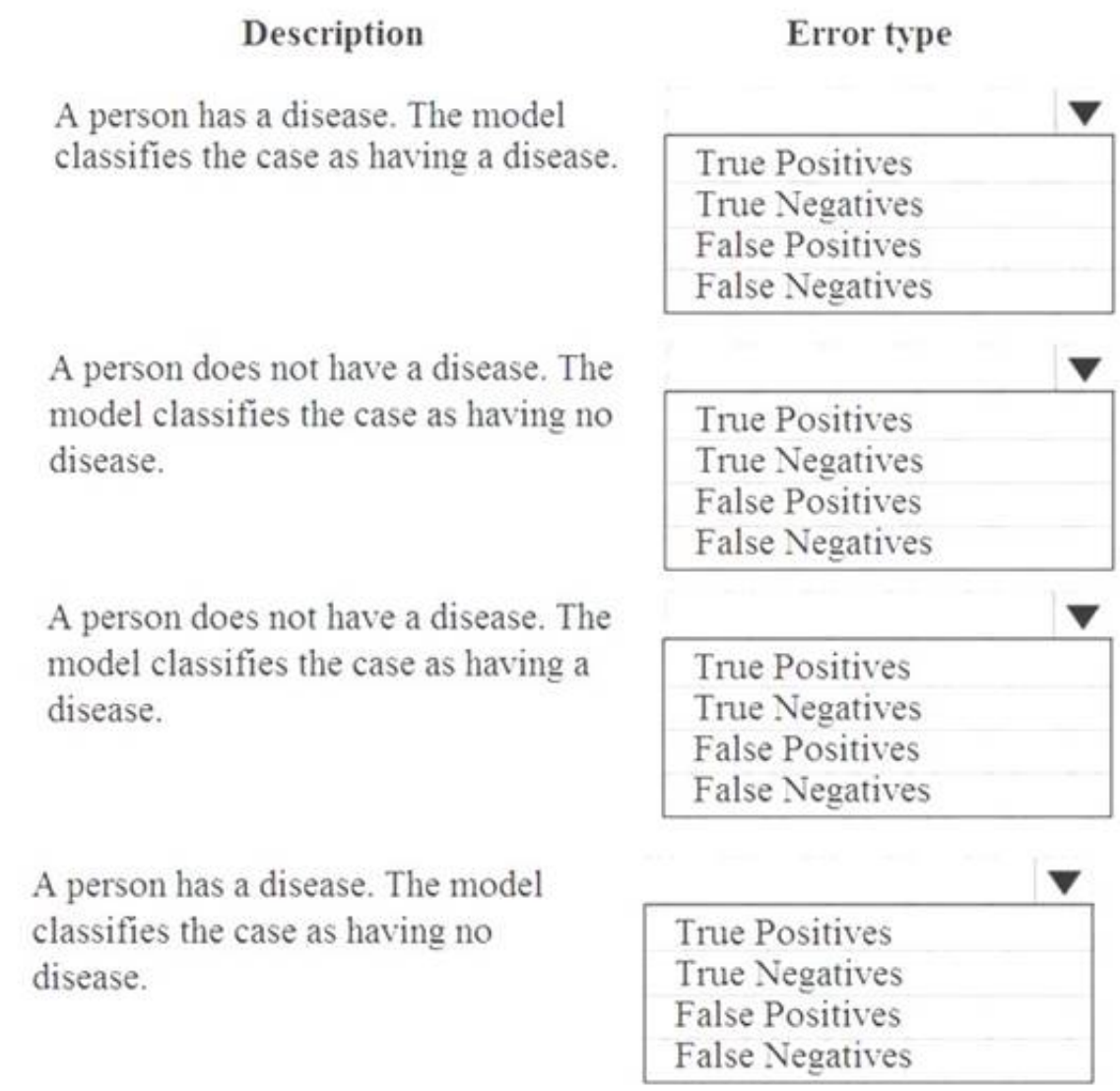
NEW QUESTION 144
- (Exam Topic 3)
You are creating a machine learning model. You have a dataset that contains null rows.
You need to use the Clean Missing Data module in Azure Machine Learning Studio to identify and resolve the null and missing data in the dataset.
Which parameter should you use?

A. Replace with mean
B. Remove entire column
C. Remove entire row
D. Hot Deck

Answer: B

Explanation:
Remove entire row: Completely removes any row in the dataset that has one or more missing values. This is useful if the missing value can be considered randomly missing.
References:
https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/clean-missing-data

NEW QUESTION 146
- (Exam Topic 3)
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.



A. Mastered
B. Not Mastered

Answer: A

Explanation:
Box 1: True Positive
A true positive is an outcome where the model correctly predicts the positive class Box 2: True Negative
A true negative is an outcome where the model correctly predicts the negative class. Box 3: False Positive
A false positive is an outcome where the model incorrectly predicts the positive class. Box 4: False Negative
A false negative is an outcome where the model incorrectly predicts the negative class. Note: Let's make the following definitions:
"Wolf" is a positive class. "No wolf" is a negative class.
We can summarize our "wolf-prediction" model using a 2x2 confusion matrix that depicts all four possible outcomes:
Reference:
https://developers.google.com/machine-learning/crash-course/classification/true-false-positive-negative

NEW QUESTION 149
- (Exam Topic 3)
Note: This question is part of a series of questions that present the same scenario. Each question in the series contains a unique solution that might meet the stated goals. Some question sets might have more than one correct solution, while others might not have a correct solution.
After you answer a question in this section, you will NOT be able to return to it. As a result, these questions will not appear in the review screen.
You train a classification model by using a logistic regression algorithm.
You must be able to explain the model's predictions by calculating the importance of each feature, both as an overall global relative importance value and as a measure of local importance for a specific set of predictions.
You need to create an explainer that you can use to retrieve the required global and local feature importance values.
Solution: Create a PFIExplainer. Does the solution meet the goal?

A. Yes
B. No

Answer: A

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

NEW QUESTION 151
- (Exam Topic 3)
You are developing a linear regression model in Azure Machine Learning Studio. You run an experiment to compare different algorithms.
The following image displays the results dataset output:

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

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

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

A. Mastered
B. Not Mastered

Answer: A

Explanation:
Box 1: Boosted Decision Tree Regression
Mean absolute error (MAE) measures how close the predictions are to the actual outcomes; thus, a lower score is better.
Box 2:
Online Gradient Descent: If you want the algorithm to find the best parameters for you, set Create trainer mode option to Parameter Range. You can then specify multiple values for the algorithm to try.
References:
https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/evaluate-model https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/linear-regression

NEW QUESTION 152
- (Exam Topic 3)
Note: This question is part of a series of questions that present the same scenario. Each question in the series contains a unique solution that might meet the stated goals. Some question sets might have more than one correct solution, while others might not have a correct solution.
After you answer a question in this section, you will NOT be able to return to it. As a result, these questions will not appear in the review screen.
You plan to use a Python script to run an Azure Machine Learning experiment. The script creates a reference to the experiment run context, loads data from a file, identifies the set of unique values for the label column, and completes the experiment run:
from azureml.core import Run import pandas as pd
run = Run.get_context()
data = pd.read_csv('data.csv') label_vals = data['label'].unique()
# Add code to record metrics here run.complete()
The experiment must record the unique labels in the data as metrics for the run that can be reviewed later.
You must add code to the script to record the unique label values as run metrics at the point indicated by the comment.
Solution: Replace the comment with the following code: run.log_table('Label Values', label_vals)
Does the solution meet the goal?

A. Yes
B. No

Answer: B

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

NEW QUESTION 153
- (Exam Topic 3)
Note: This question is part of a series of questions that present the same scenario. Each question in the series contains a unique solution that might meet the stated goals. Some question sets might have more than one correct solution, while others might not have a correct solution.
After you answer a question in this section, you will NOT be able to return to it. As a result, these questions will not appear in the review screen.
You are analyzing a numerical dataset which contain missing values in several columns.
You must clean the missing values using an appropriate operation without affecting the dimensionality of the feature set.
You need to analyze a full dataset to include all values.
Solution: Use the last Observation Carried Forward (IOCF) method to impute the missing data points. Does the solution meet the goal?

A. Yes
B. No

Answer: B

Explanation:
Instead use the Multiple Imputation by Chained Equations (MICE) method.
Replace using MICE: For each missing value, this option assigns a new value, which is calculated by using a method described in the statistical literature as "Multivariate Imputation using Chained Equations" or "Multiple Imputation by Chained Equations". With a multiple imputation method, each variable with missing data is modeled conditionally using the other variables in the data before filling in the missing values.
Note: Last observation carried forward (LOCF) is a method of imputing missing data in longitudinal studies. If a person drops out of a study before it ends, then his or her last observed score on the dependent variable is used for all subsequent (i.e., missing) observation points. LOCF is used to maintain the sample size and to reduce the bias caused by the attrition of participants in a study.
References:
https://methods.sagepub.com/reference/encyc-of-research-design/n211.xml https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3074241/

NEW QUESTION 157
- (Exam Topic 3)
You create an Azure Machine Learning compute resource to train models. The compute resource is configured as follows:
- Minimum nodes: 2
- Maximum nodes: 4

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

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

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

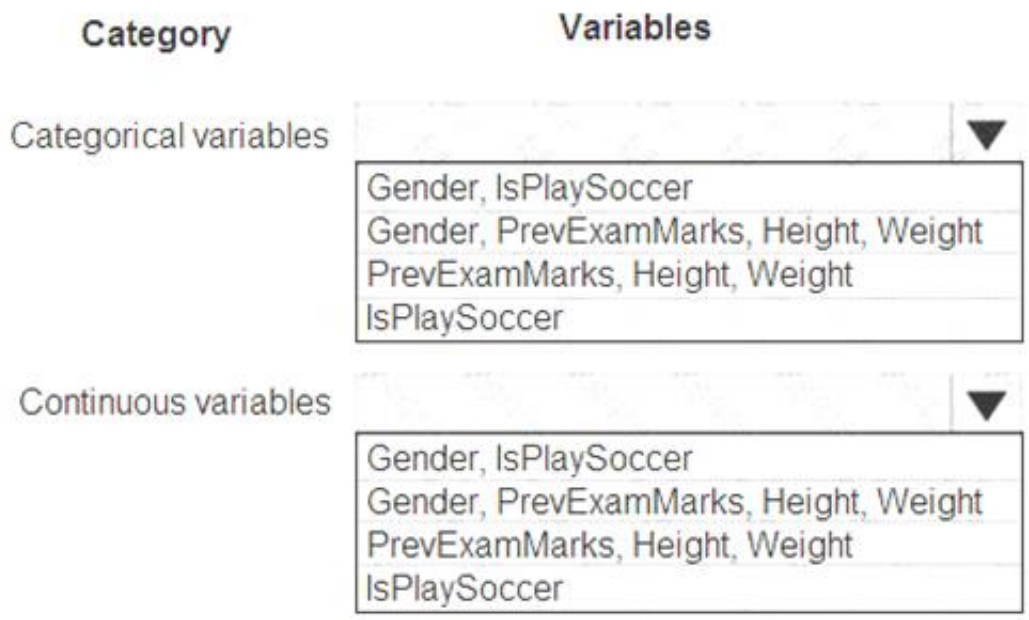
Answer: ABC

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

NEW QUESTION 158
- (Exam Topic 3)
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.



A. Mastered
B. Not Mastered

Answer: A

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

NEW QUESTION 159
- (Exam Topic 3)
Note: This question is part of a series of questions that present the same scenario. Each question in the series contains a unique solution that might meet the stated goals. Some question sets might have more than one correct solution, while others might not have a correct solution.
After you answer a question in this section, you will NOT be able to return to it. As a result, these questions will not appear in the review screen.
You are a data scientist using Azure Machine Learning Studio.
You need to normalize values to produce an output column into bins to predict a target column. Solution: Apply a Quantiles binning mode with a PQuantile normalization.
Does the solution meet the goal?

A. Yes
B. No

Answer: B

Explanation:
Use the Entropy MDL binning mode which has a target column. References:
https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/group-data-into-bins

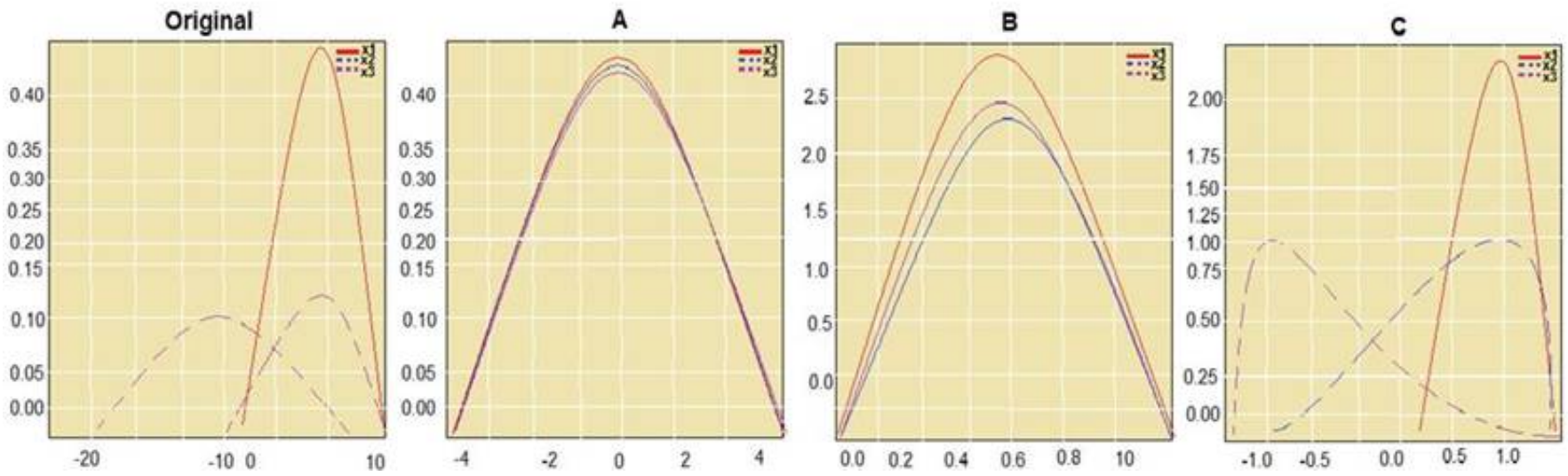
NEW QUESTION 164
- (Exam Topic 3)
You create a script that trains a convolutional neural network model over multiple epochs and logs the validation loss after each epoch. The script includes arguments for batch size and learning rate.
You identify a set of batch size and learning rate values that you 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

Answer: E

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

NEW QUESTION 169
- (Exam Topic 3)
You are performing feature scaling by using the scikit-learn Python library for x.1 x2, and x3 features. Original and scaled data is shown in the following image.



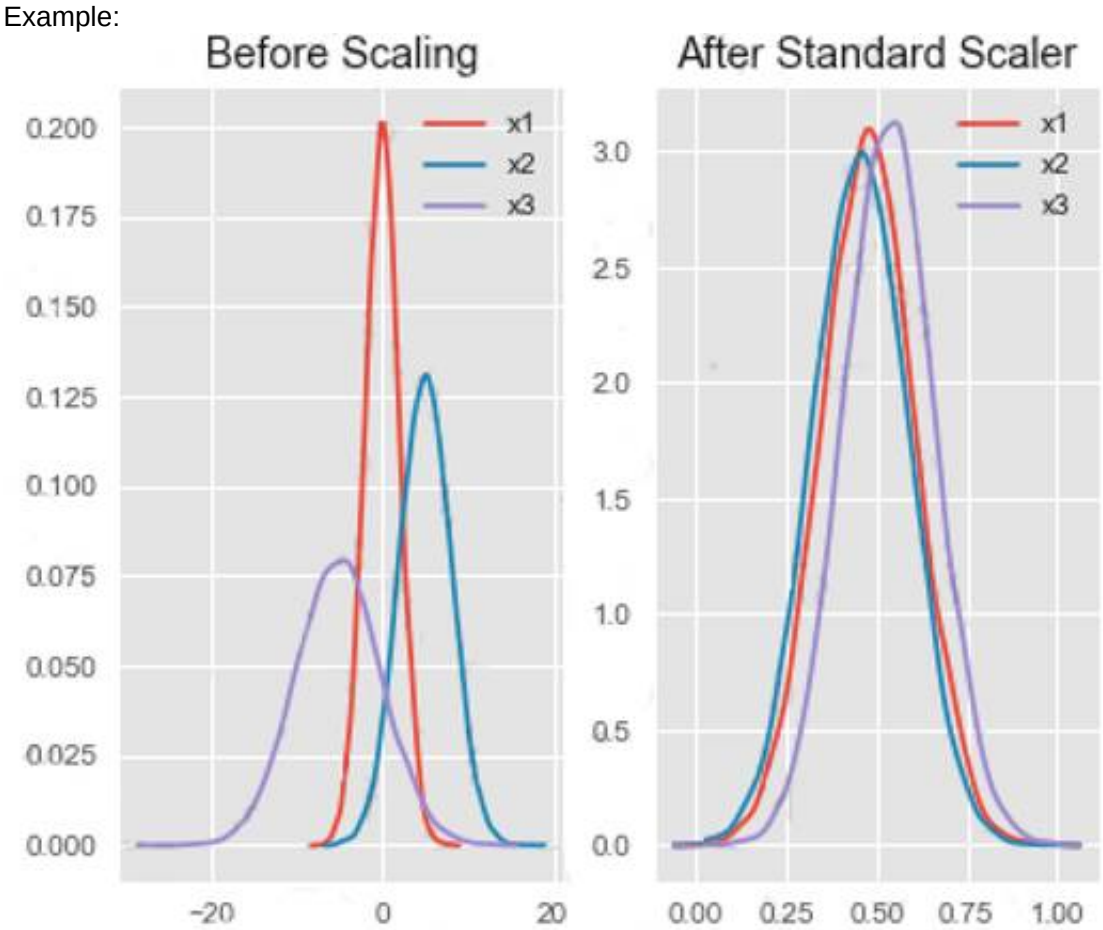
Use the drop-down menus to select the answer choice that answers each question based on the information presented in the graphic.
NOTE: Each correct selection is worth one point.

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

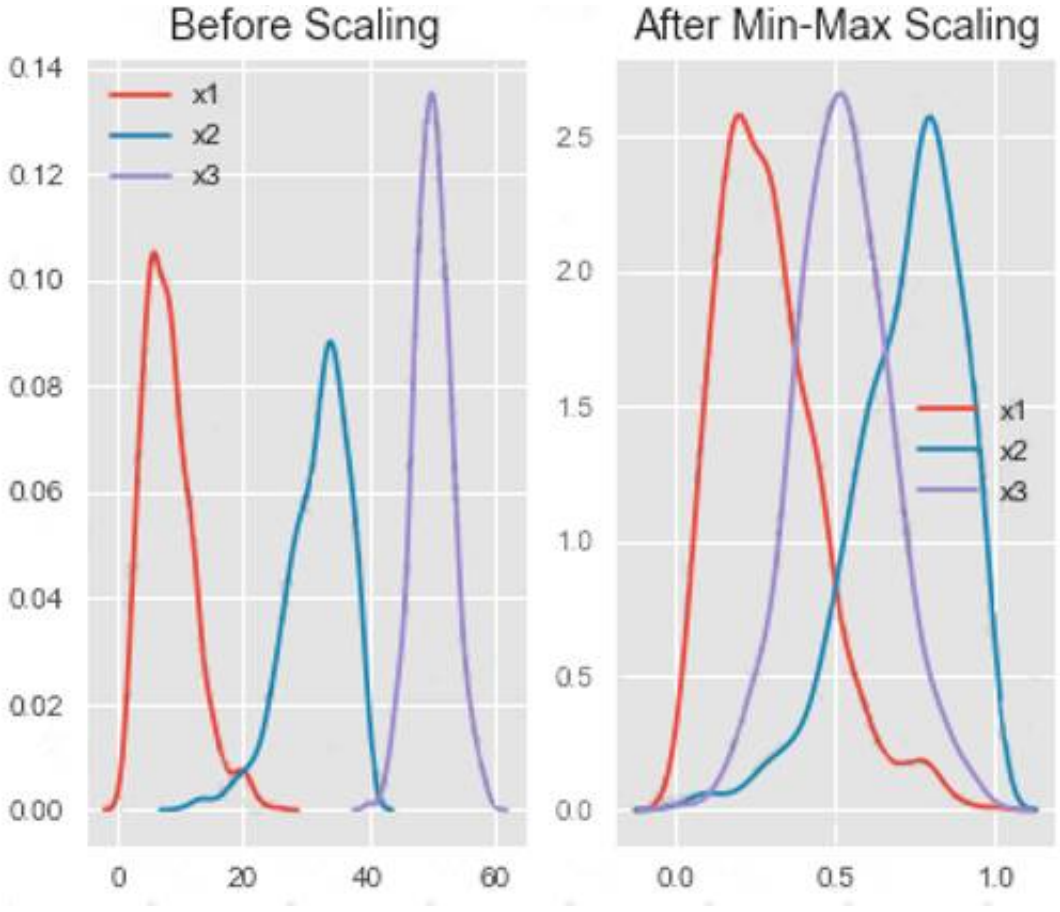
A. Mastered
B. Not Mastered

Answer: A

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



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



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

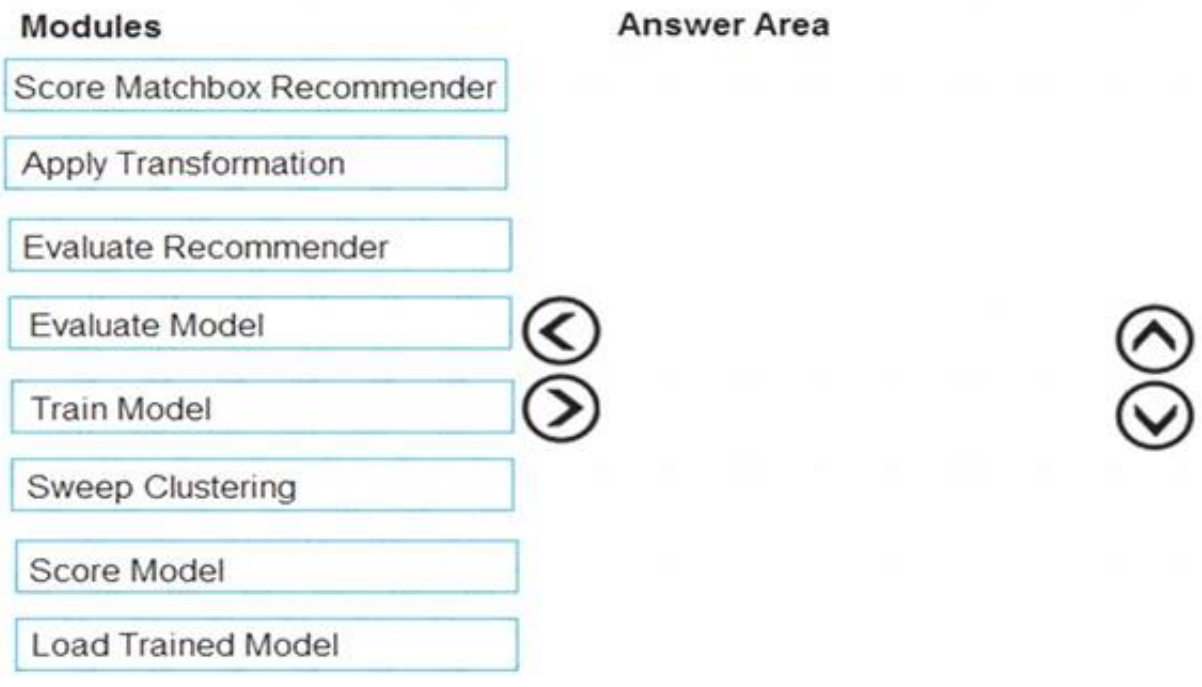
NEW QUESTION 170
- (Exam Topic 3)
An organization creates and deploys a multi-class image classification deep learning model that uses a set of labeled photographs.
The software engineering team reports there is a heavy inferencing load for the prediction web services during the summer. The production web service for the model fails to meet demand despite having a fully-utilized compute cluster where the web service is deployed.
You need to improve performance of the image classification web service with minimal downtime and minimal administrative effort.
What should you advise the IT Operations team to do?

A. Increase the minimum node count of the compute cluster where the web service is deployed.
B. Create a new compute cluster by using larger VM sizes for the nodes, redeploy the web service to that cluster, and update the DNS registration for the service endpoint to point to the new cluster.
C. Increase the VM size of nodes in the compute cluster where the web service is deployed.
D. Increase the node count of the compute cluster where the web service is deployed.

Answer: D

Explanation:
The Azure Machine Learning SDK does not provide support scaling an AKS cluster. To scale the nodes in the cluster, use the UI for your AKS cluster in the Azure Machine Learning studio. You can only change the node count, not the VM size of the cluster.
Reference:
https://docs.microsoft.com/en-us/azure/machine-learning/how-to-create-attach-kubernetes

NEW QUESTION 172
- (Exam Topic 2)
You need to produce a visualization for the diagnostic test evaluation according to the data visualization requirements.
Which three modules should you recommend be used in sequence? To answer, move the appropriate modules from the list of modules to the answer area and arrange them in the correct order.



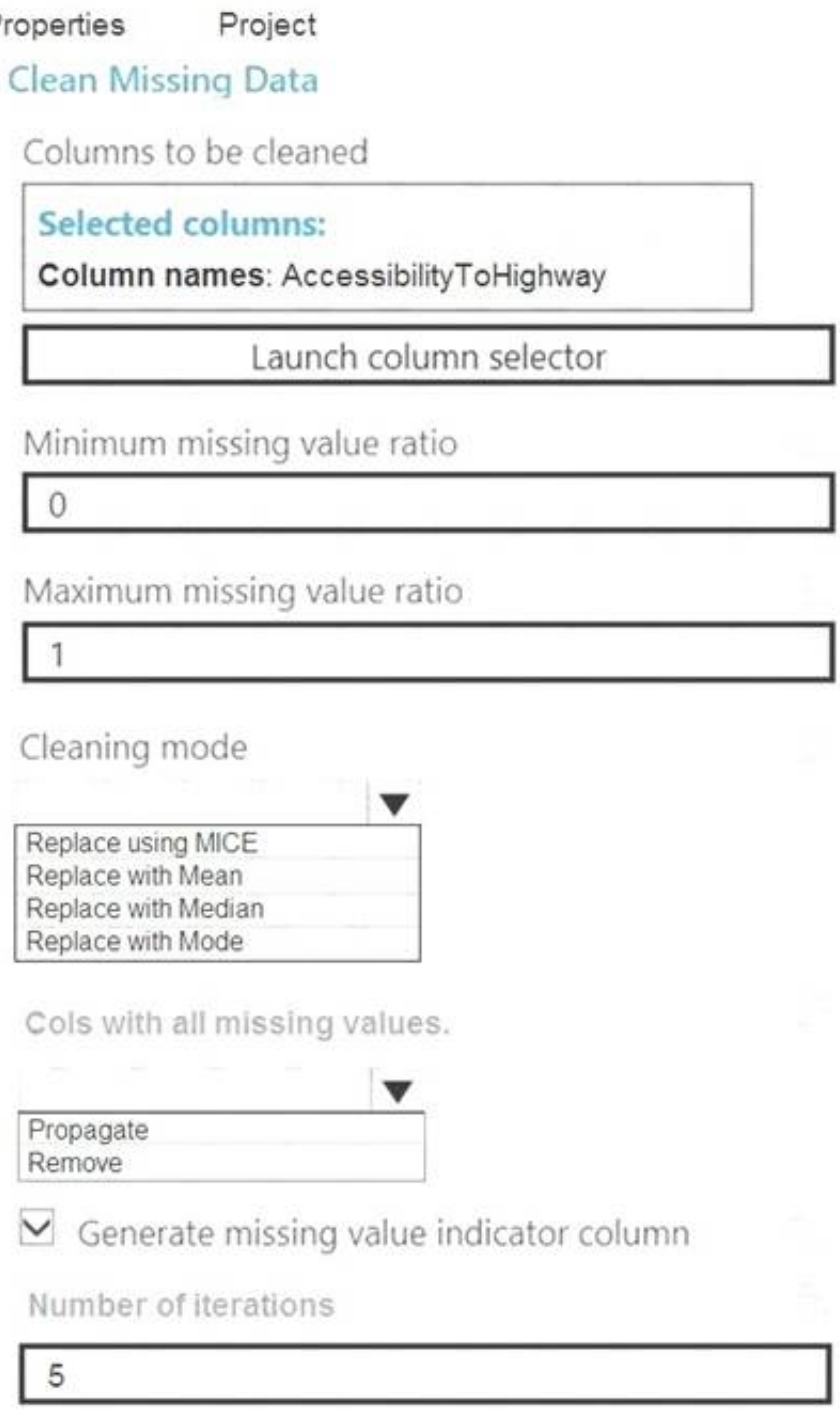
A. Mastered

B. Not Mastered

Answer: A

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

NEW QUESTION 173
- (Exam Topic 2)
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.



A. Mastered
B. Not Mastered

Answer: A

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

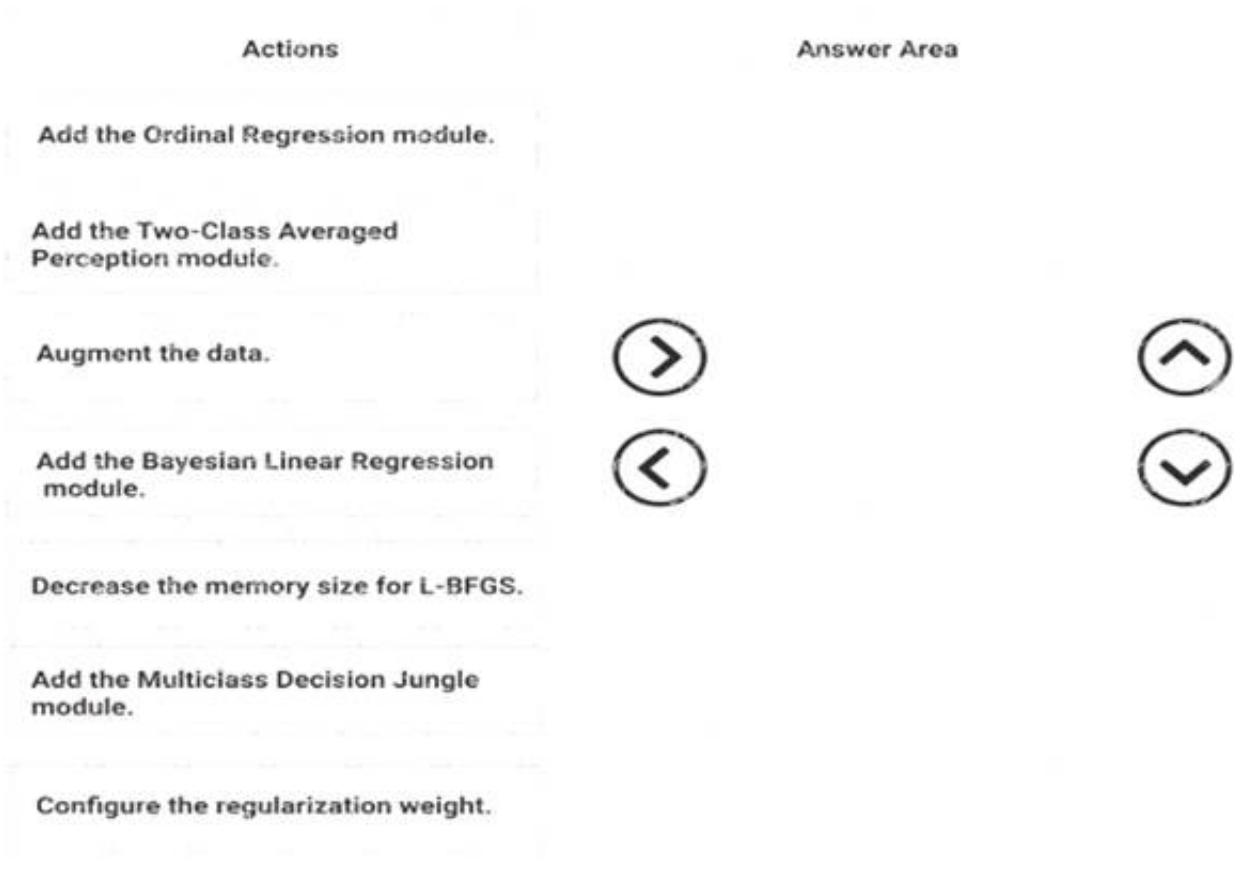
NEW QUESTION 176
- (Exam Topic 2)
You need to select a feature extraction method. Which method should you use?

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

Answer: C

Explanation:
In statistics, the Kendall rank correlation coefficient, commonly referred to as Kendall's tau coefficient (after the Greek letter), 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.
Scenario: When you train a Linear Regression module using a property dataset that shows data for property prices for a large city, you need to determine the best features to use in a model. You can choose standard metrics provided to measure performance before and after the feature importance process completes. You must ensure that the distribution of the features across multiple training models is consistent.
References:
https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/feature-selection-modules

NEW QUESTION 177
- (Exam Topic 2)
You need to correct the model fit issue.
Which three actions should you perform in sequence? To answer, move the appropriate actions from the list of actions to the answer area and arrange them in the correct order.

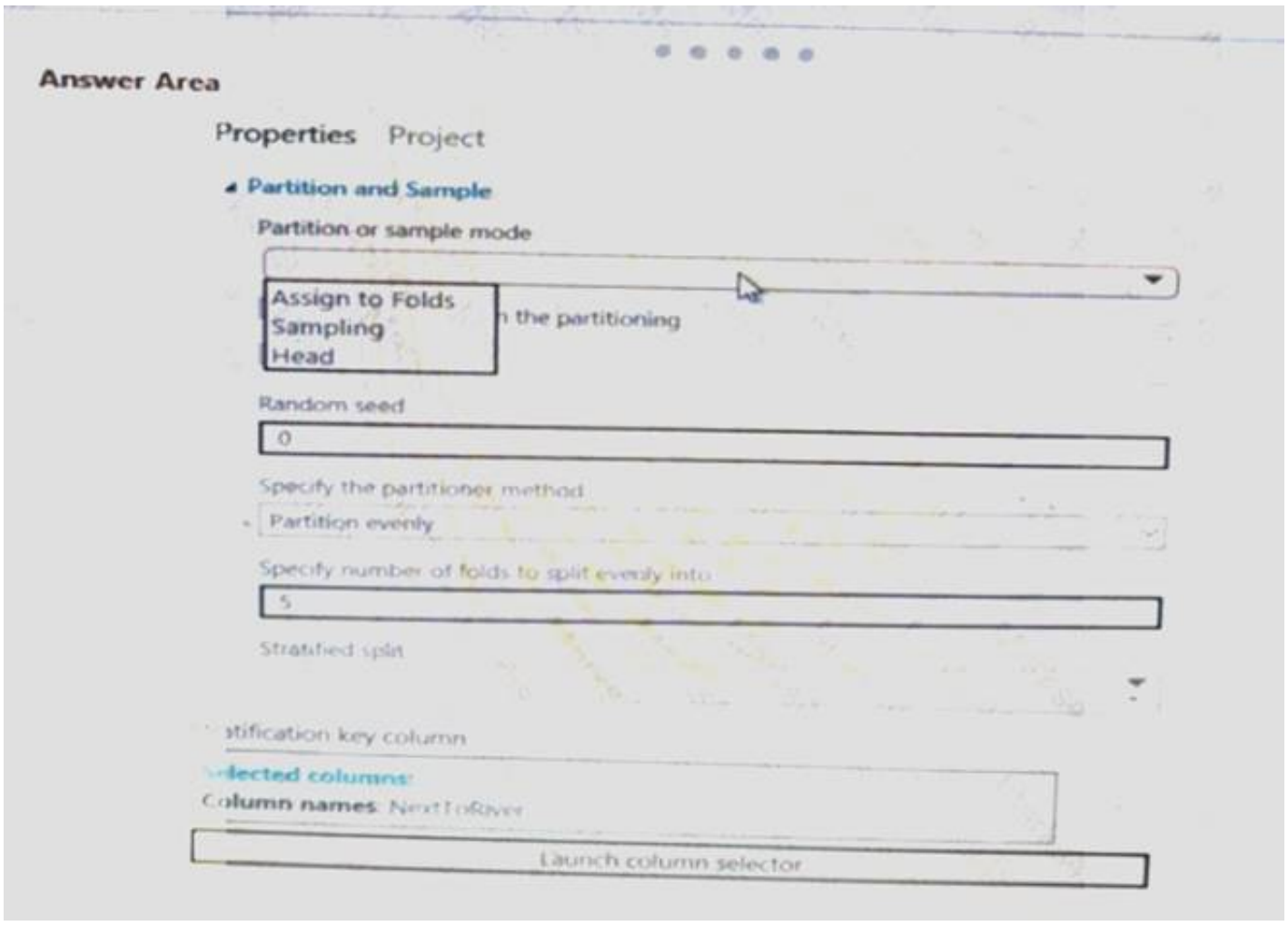


A. Mastered
B. Not Mastered

Answer: A

Explanation:
Step 1: Augment the data
Scenario: Columns in each dataset contain missing and null values. The datasets also contain many outliers.
Step 2: Add the Bayesian Linear Regression module.
Scenario: You produce a regression model to predict property prices by using the Linear Regression and Bayesian Linear Regression modules.
Step 3: Configure the regularization weight.
Regularization typically is used to avoid overfitting. For example, in L2 regularization weight, type the value to use as the weight for L2 regularization. We recommend that you use a non-zero value to avoid overfitting.
Scenario:
Model fit: The model shows signs of overfitting. You need to produce a more refined regression model that reduces the overfitting.

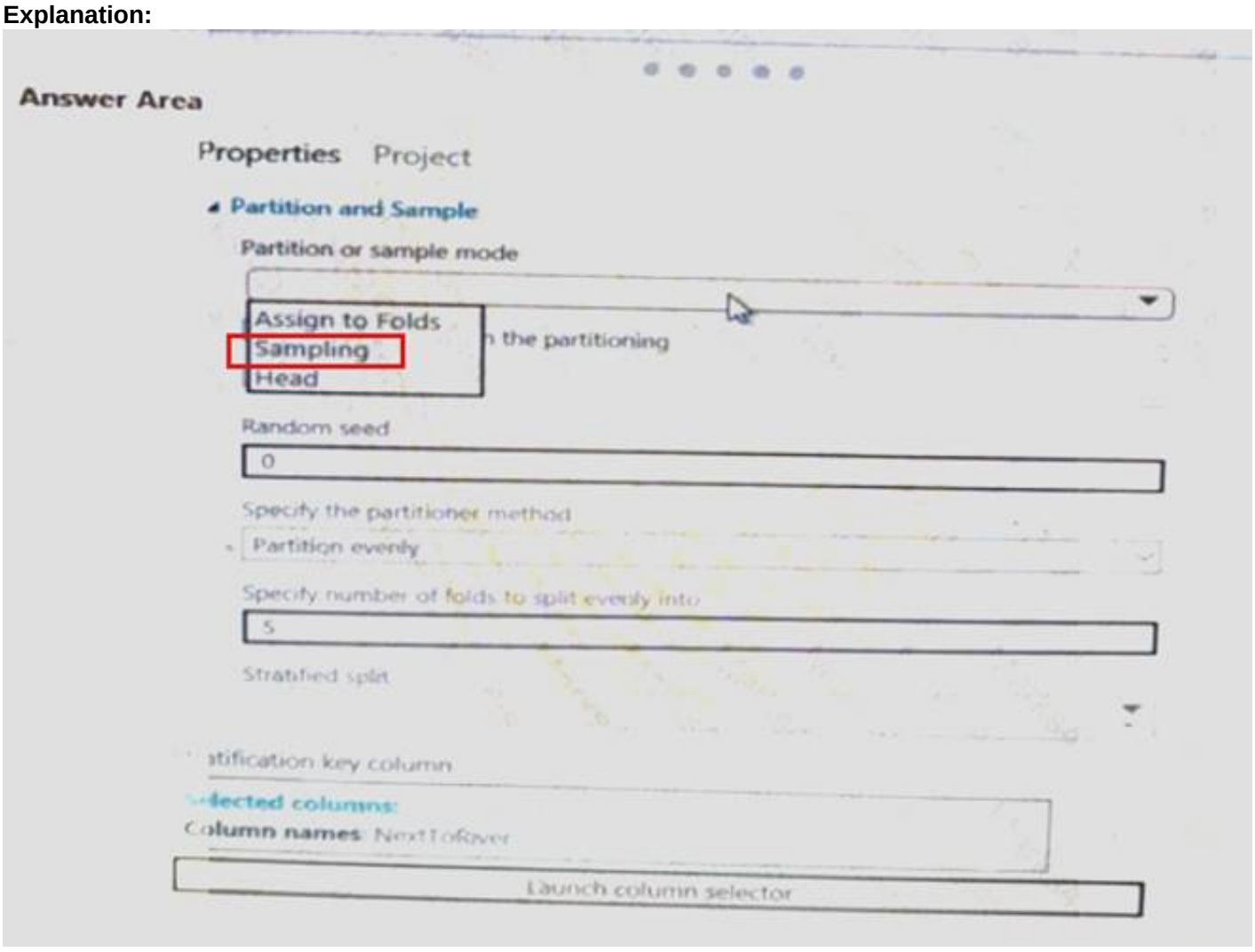
NEW QUESTION 182
- (Exam Topic 2)
You need to identify the methods for dividing the data according, to the testing requirements.
Which properties should you select? To answer, select the appropriate option-, m the answer area. NOTE: Each correct selection is worth one point.



A. Mastered
B. Not Mastered

Answer: A

Explanation:



NEW QUESTION 187
- (Exam Topic 1)
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.

Answer: D

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.

NEW QUESTION 192
- (Exam Topic 1)
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.

Answer: A

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

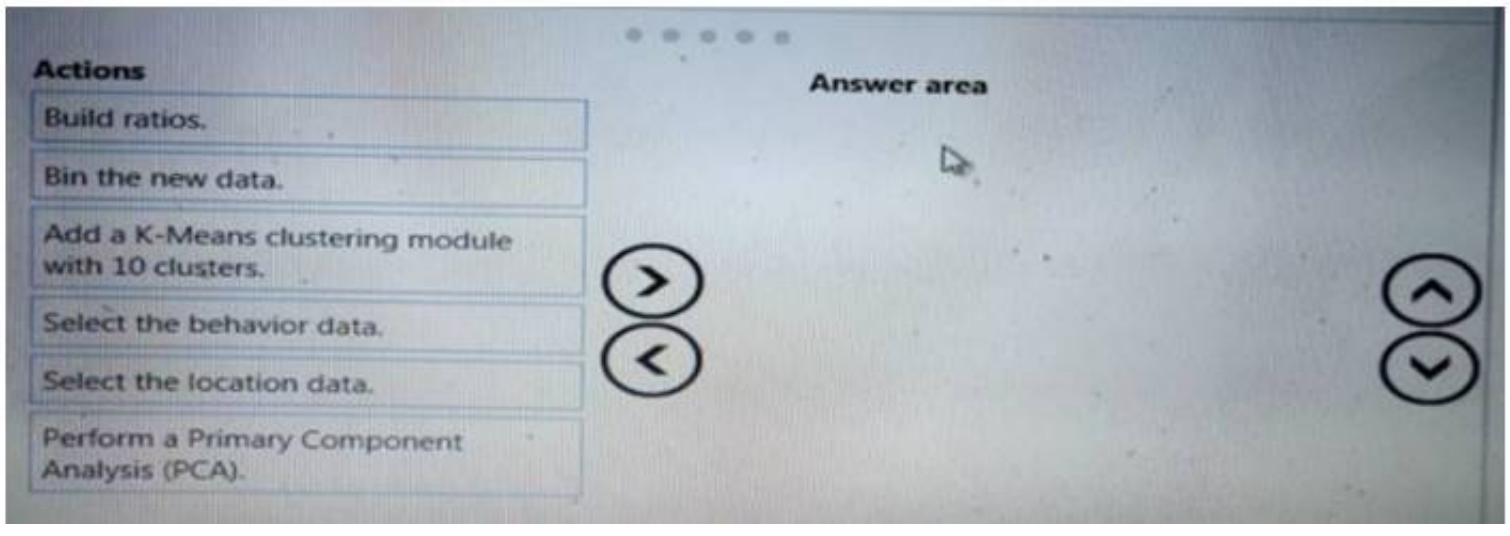
NEW QUESTION 193
- (Exam Topic 1)
You need to implement a scaling strategy for the local penalty detection data.
Which normalization type should you use?

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

Answer: C

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. References:
https://docs.microsoft.com/en-us/cognitive-toolkit/post-batch-normalization-statistics

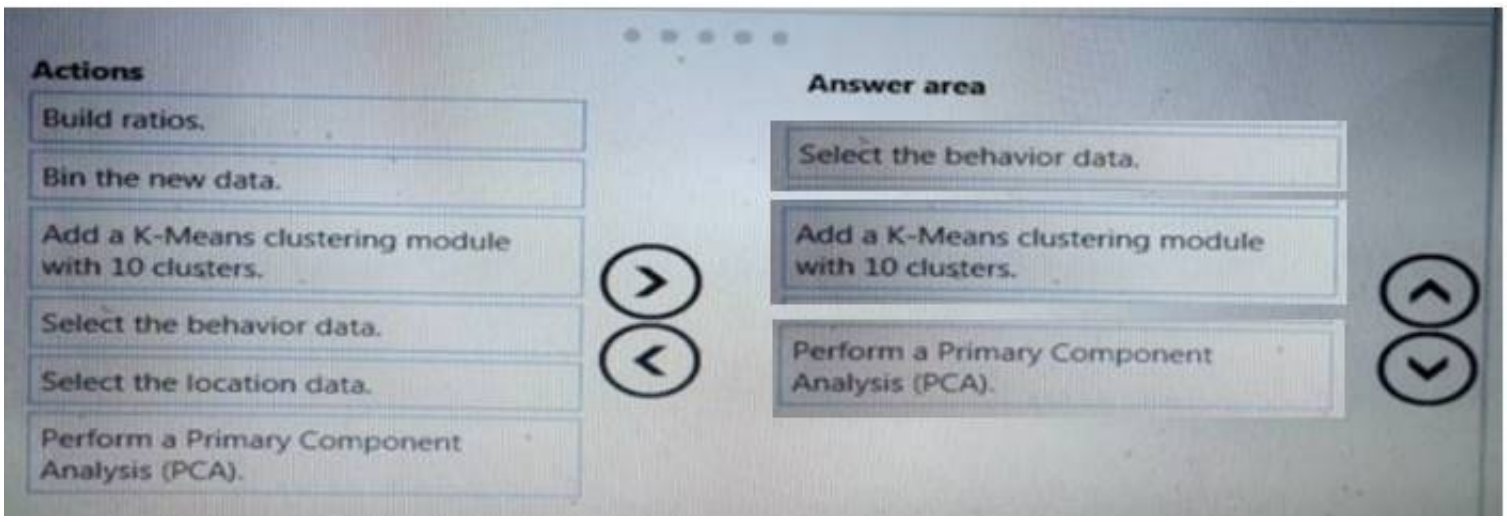
NEW QUESTION 195
- (Exam Topic 1)
You need to modify the inputs for the global penalty event model to address the bias and variance issue. Which three actions should you perform in sequence? To answer, move the appropriate actions from the list of actions to the answer area and arrange them in the correct order.



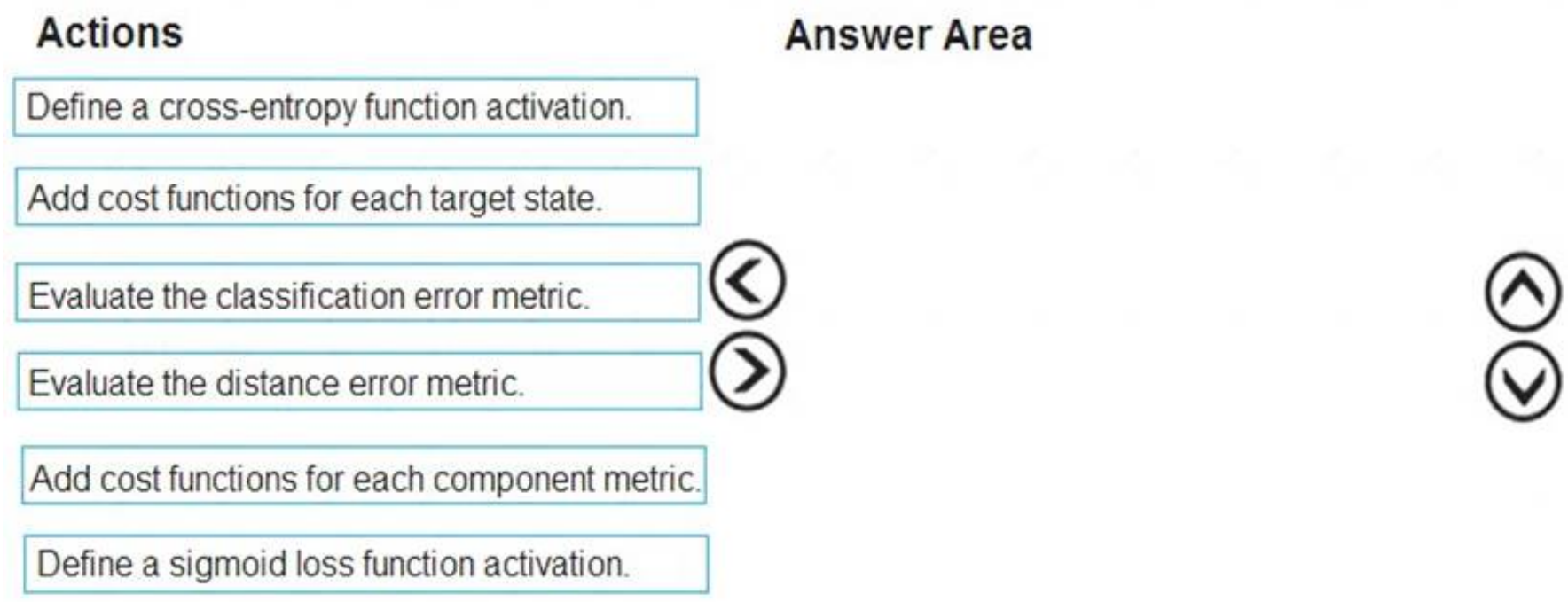
A. Mastered
B. Not Mastered

Answer: A

Explanation:



NEW QUESTION 199
- (Exam Topic 1)
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.

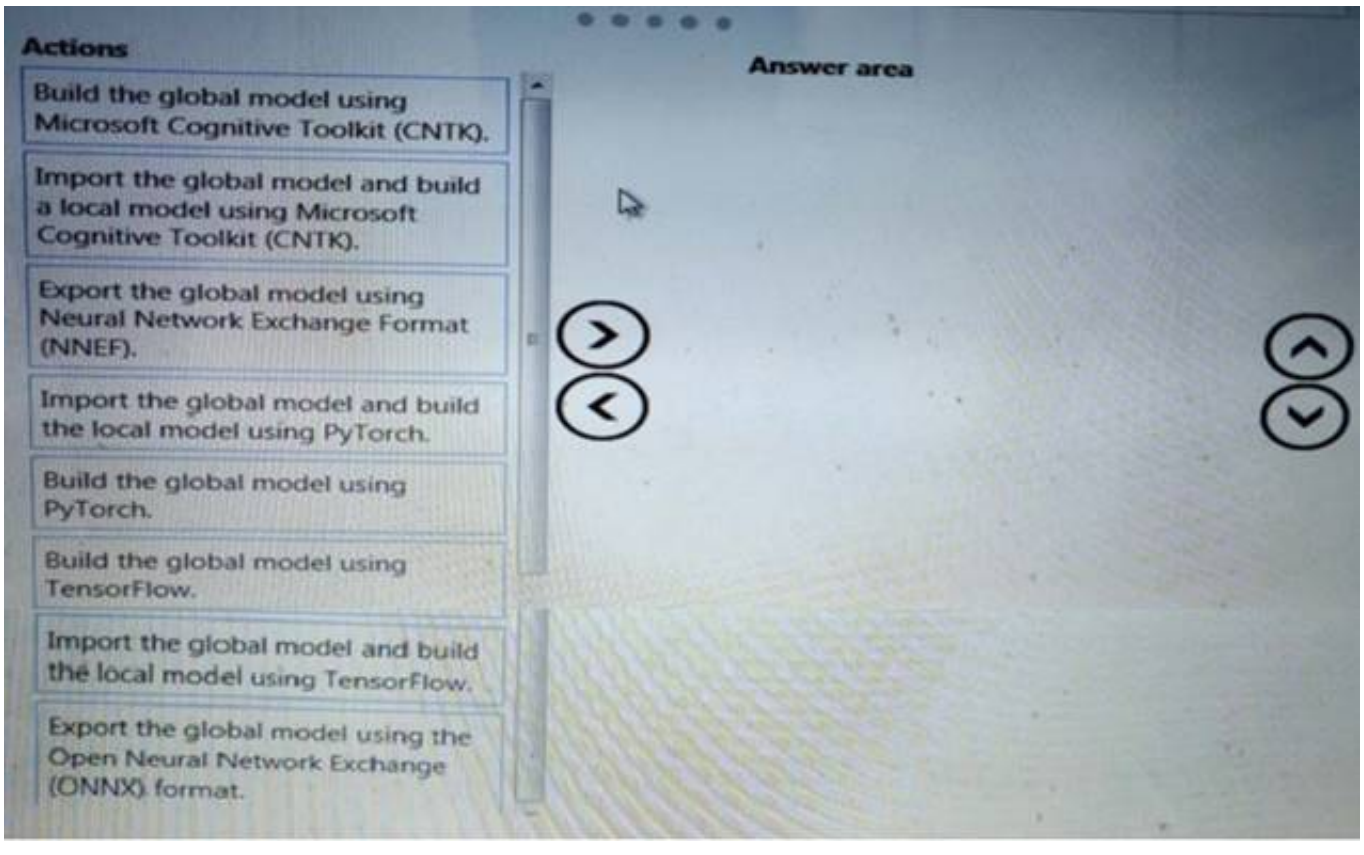


A. Mastered
B. Not Mastered

Answer: A

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/

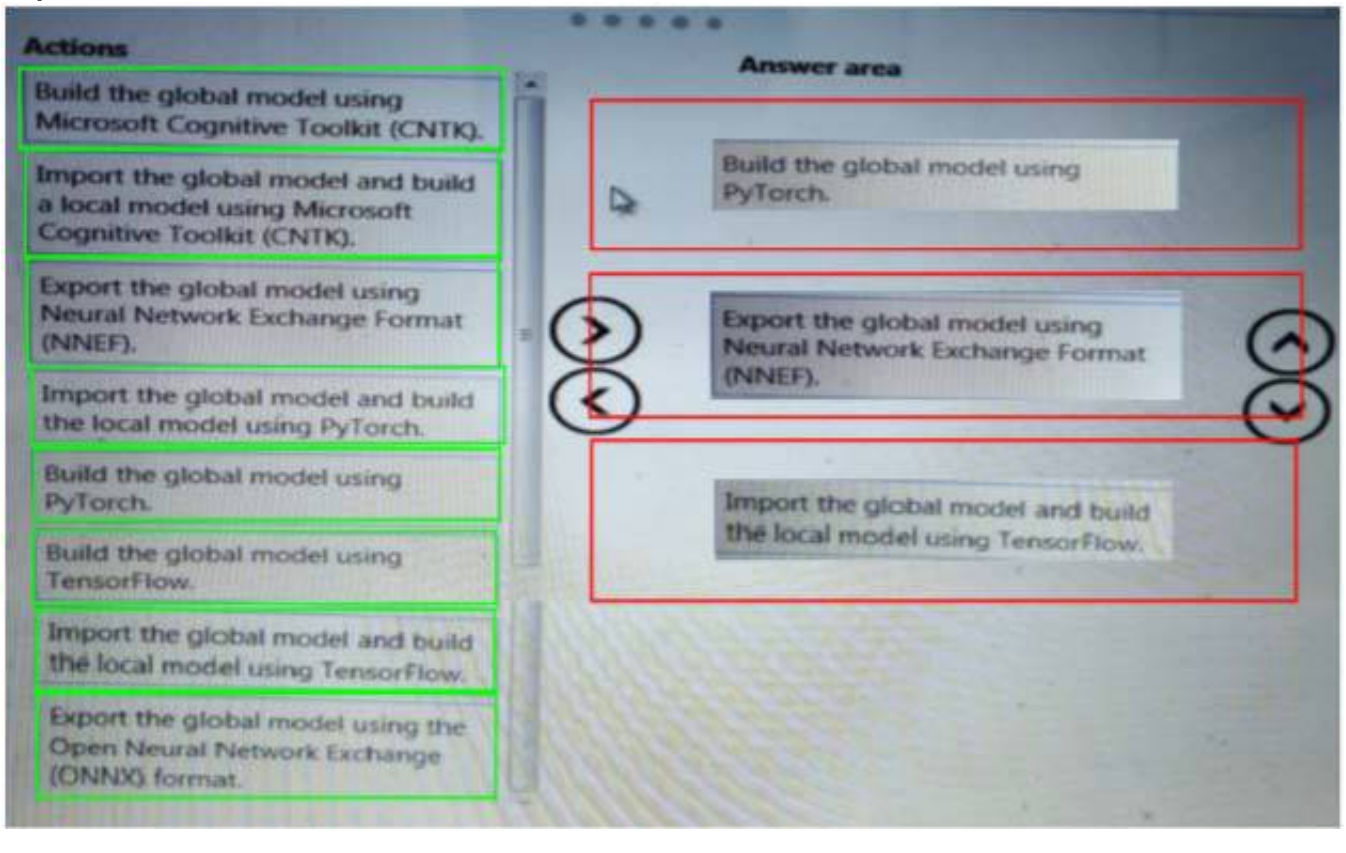
NEW QUESTION 201
- (Exam Topic 1)
You need to define a process for penalty event detection.
Which three actions should you perform in sequence? To answer, move the appropriate actions from the list of actions to the answer area and arrange them in the correct order.



A. Mastered
B. Not Mastered

Answer: A

Explanation:



NEW QUESTION 205
.....

Relate Links

100% Pass Your DP-100 Exam with Exambible Prep Materials

https://www.exambible.com/DP-100-exam/

Contact us

We are proud of our high-quality customer service, which serves you around the clock 24/7.

Viste - https://www.exambible.com/