Exam Questions DP-100

Designing and Implementing a Data Science Solution on Azure

https://www.2passeasy.com/dumps/DP-100/

NEW QUESTION 1
- (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_output = PipelineData("processed_data", datastore=datastore)
process_step = PythonScriptStep(script_name="process.py",
 arguments=["--data_for_train", data_output],
 outputs=[data_output], compute_target=aml_compute,
 source_directory=process_directory)
pipeline = Pipeline(workspace=ws, steps=[process_step])
```

Does the solution meet the goal?

A. Yes
B. No

Answer: B

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

NEW QUESTION 2
- (Exam Topic 3)
HOTSPOT
You register the following versions of a model.

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

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

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

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

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

A. Mastered
B. Not Mastered

Answer: A

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

NEW QUESTION 3
- (Exam Topic 3)
You create a Python script named train.py and save it in a folder named scripts. The script uses the scikit-learn framework to train a machine learning model.
You must run the script as an Azure Machine Learning experiment on your local workstation. You need to write Python code to initiate an experiment that runs the train.py script.
How should you complete the code segment? To answer, select the appropriate options in the answer area.
NOTE: Each correct selection is worth one point.

Answer Area

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

ws = Workspace.from_config()
py_sk = Environment('sklearn-training')
pkgs = CondaDependencies.create(pip_packages=['scikit-learn', 'azureml-defaults'])
py_sk.python.conda_dependencies = pkgs
script_config = ScriptRunConfig ( [ ▼ ] = 'scripts',
                                  [ ▼ ] = 'train.py',
                                  [ ▼ ] =py_sk)
experiment = Experiment(workspace=ws, name='training-experiment')
run = experiment.submit(config=script_config)
```

script
source_directory
resume_from
arguments

script
arguments
environment
compute_target

arguments
resume_from
environment
compute_target

A. Mastered
B. Not Mastered

Answer: A

Explanation:
Graphical user interface, text, application, table, Word Description automatically generated
Box 1: source_directory
source_directory: A local directory containing code files needed for a run. Box 2: script
Script: The file path relative to the source_directory of the script to be run. Box 3: environment
Reference:
https://docs.microsoft.com/en-us/python/api/azureml-core/azureml.core.scriptrunconfig

NEW QUESTION 4
- (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,
```

backfill
create_from_datasets
create_from_model

```
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
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 5
- (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)

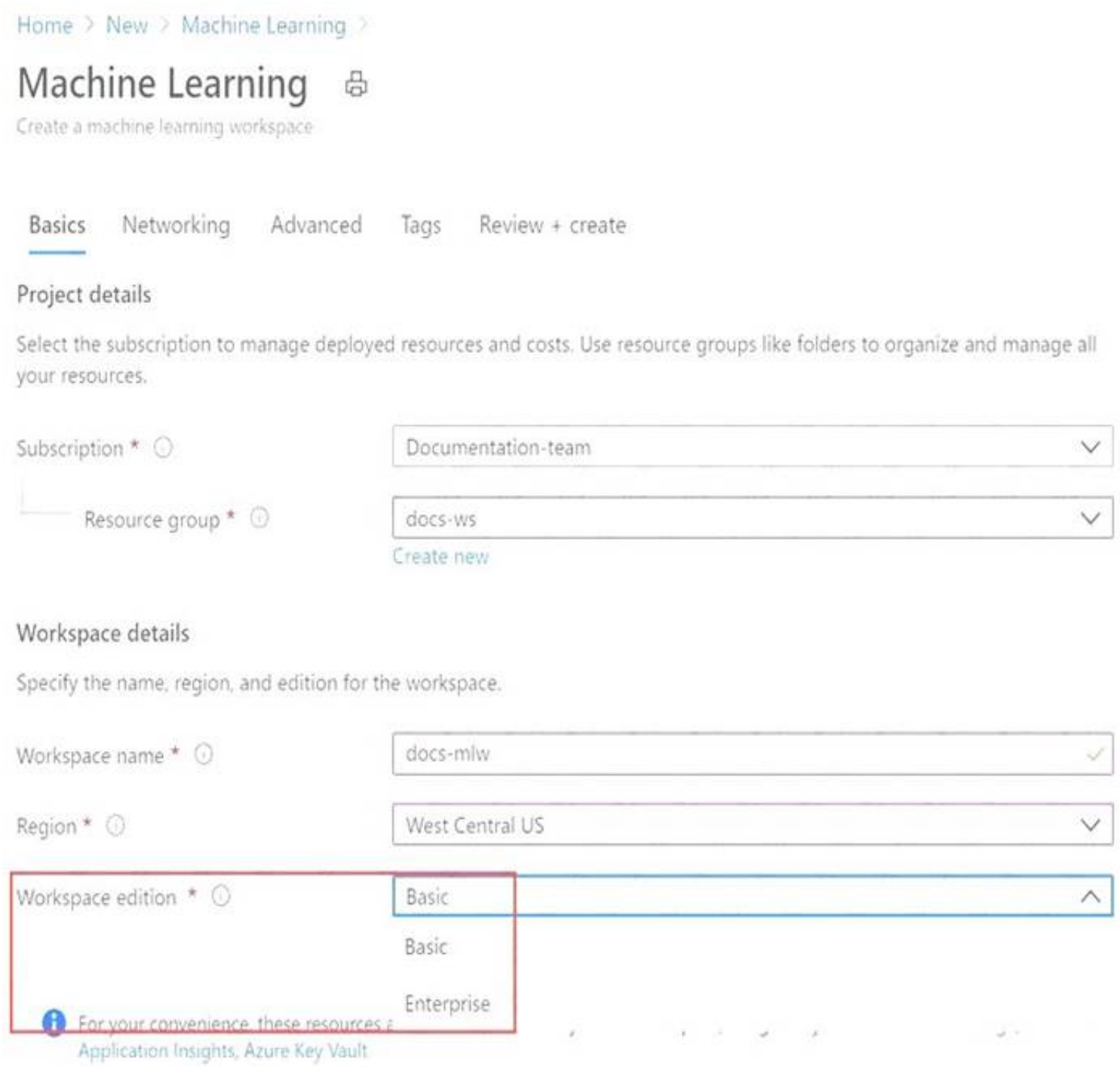
NEW QUESTION 6
- (Exam Topic 3)
You create a new Azure subscription. No resources are provisioned in the subscription. You need to create an Azure Machine Learning workspace.
What are three possible ways to achieve this goal? Each correct answer presents a complete solution. NOTE: Each correct selection 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 7
- (Exam Topic 3)
You plan to use a Deep Learning Virtual Machine (DLVM) to train deep learning models using Compute Unified Device Architecture (CUDA) computations.
You need to configure the DLVM to support CUDA. What should you implement?

A. Intel Software Guard Extensions (Intel SGX) technology
B. Solid State Drives (SSD)
C. Graphic Processing Unit (GPU)
D. Computer Processing Unit (CPU) speed increase by using overcloking
E. High Random Access Memory (RAM) configuration

Answer: C

Explanation:
A Deep Learning Virtual Machine is a pre-configured environment for deep learning using GPU instances.
References:
https://azuremarketplace.microsoft.com/en-au/marketplace/apps/microsoft-ads.dsvm-deep-learning

NEW QUESTION 8
- (Exam Topic 3)
You are hired as a data scientist at a winery. The previous data scientist used Azure Machine Learning. You need to review the models and explain how each model makes decisions.
Which explainer modules should you use? To answer, select the appropriate options in the answer area.
NOTE: Each correct selection is worth one point.

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

A. Mastered
B. Not Mastered

Answer: A

Explanation:
info based on
the given model and data sets. The meta explainers leverage all the libraries (SHAP, LIME, Mimic, etc.) that we have integrated or developed. The following are the meta explainers available in the SDK:
Tabular Explainer: Used with tabular datasets. Text Explainer: Used with text datasets. Image Explainer: Used with image datasets. Box 1: Tabular
Box 2: Text
Box 3: Image Reference:
https://medium.com/microsoftazure/automated-and-interpretable-machine-learning-d07975741298

NEW QUESTION 9
- (Exam Topic 3)
You are solving a classification task.
You must evaluate your model on a limited data sample by using k-fold cross validation. You start by configuring a k parameter as the number of splits.
You need to configure the k parameter for the cross-validation. Which value should you use?

A. k=0.5
B. k=0
C. k=5
D. k=1

Answer: C

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

NEW QUESTION 10
- (Exam Topic 3)
You create an Azure Machine Learning workspace.
You must create a custom role named DataScientist that meets the following requirements:
- Role members must not be able to delete the workspace.
- Role members must not be able to create, update, or delete compute resource in the workspace.
- Role members must not be able to add new users to the workspace.

You need to create a JSON file for the DataScientist role in the Azure Machine Learning workspace. The custom role must enforce the restrictions specified by the IT Operations team.
Which JSON code segment should you use?
A)

```
{
"Name": "DataScientist",
     "IsCustom": true,
     "Description": "Project Data Scientist role",
     "Actions": ["*"],
     "NotActions":[
     "Microsoft.MachineLearningServices/workspaces/*/delete",
     "Microsoft.MachineLearningServices/workspaces/computes/*/write",
     "Microsoft.MachineLearningServices/workspaces/computes/*/delete",
     "Microsoft.Authorization/*/write"
],
"AssignableScopes": [
  "/subscriptions/<id>/resourceGroups/ml-rg/providers/Microsoft/MachineLearningServices/workspaces/ml-ws"
  ]
}
```

B)

```
{
"Name": "DataScientist",
"IsCustom": true,
"Description": "Project Data Scientist role",
"Actions": ["*"],
"NotActions":[],
"AssignableScopes": [
    "/subscriptions/<id>/resourceGroups/ml-rg/providers/Microsoft.MachineLearningServices/workspaces/ml-ws"
}
```

C)

```
{
"Name": "DataScientist",
"IsCustom": true,
"Description": "Project Data Scientist role",
"Actions": ["Microsoft.MachineLearningServices/workspaces/*/delete",
     "Microsoft.MachineLearningServices/workspaces/computes/*/write",
     "Microsoft.MachineLearningServices/workspaces/computes/*/delete",
     "Microsoft.Authorization/*/write"
],
"NotActions":[],
"AssignaleScopes":[
     "/subscriptions/<id>/resourceGroups/ml-rg/providers/Microsoft.MachineLearningServices/workspaces/ml-ws"
]
}
```

D)

```
{
"Name": "DataScientist",
"IsCustom": true,
"Description": "Project Data Scientist role",
"Actions": [],
"NotActions":["*"],
"AssignableScopes": [
     "/subscriptions/<id>/resourceGroups/ml-rg/providers/Microsoft.MachineLearningServices/workspaces/ml-ws"
]
}
```

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

Answer: A

Explanation:
The following custom role can do everything in the workspace except for the following actions:

- It can't create or update a compute resource.
- It can't delete a compute resource.
- It can't add, delete, or alter role assignments.
- It can't delete the workspace.

To create a custom role, first construct a role definition JSON file that specifies the permission and scope for the role. The following example defines a custom role named "Data Scientist Custom" scoped at a specific workspace level:
data_scientist_custom_role.json :
{
"Name": "Data Scientist Custom", "IsCustom": true,
"Description": "Can run experiment but can't create or delete compute.", "Actions": ["*"],
"NotActions": [
"Microsoft.MachineLearningServices/workspaces/*/delete", "Microsoft.MachineLearningServices/workspaces/write",
"Microsoft.MachineLearningServices/workspaces/computes/*/write", "Microsoft.MachineLearningServices/workspaces/computes/*/delete",
"Microsoft.Authorization/*/write"

],
"AssignableScopes": ["/subscriptions/<subscription_id>/resourceGroups/<resource_group_name>/providers/Microsoft.MachineLearni
]
}
Reference:
https://docs.microsoft.com/en-us/azure/machine-learning/how-to-assign-roles

NEW QUESTION 10
- (Exam Topic 3) You are solving a classification task. The dataset is imbalanced.
You need to select an Azure Machine Learning Studio module to improve the classification accuracy. Which module should you use?

A. Fisher Linear Discriminant Analysis.
B. Filter Based Feature Selection
C. Synthetic Minority Oversampling Technique (SMOTE)
D. Permutation Feature Importance

Answer: C

Explanation:
Use the SMOTE module in Azure Machine Learning Studio (classic) to increase the number of underepresented cases in a dataset used for machine learning. SMOTE is a better way of increasing the number of rare cases than simply duplicating existing cases.
You connect the SMOTE module to a dataset that is imbalanced. There are many reasons why a dataset might be imbalanced: the category you are targeting might be very rare in the population, or the data might simply be difficult to collect. Typically, you use SMOTE when the class you want to analyze is under-represented.
Reference:
https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/smote

NEW QUESTION 15
- (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 18
- (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

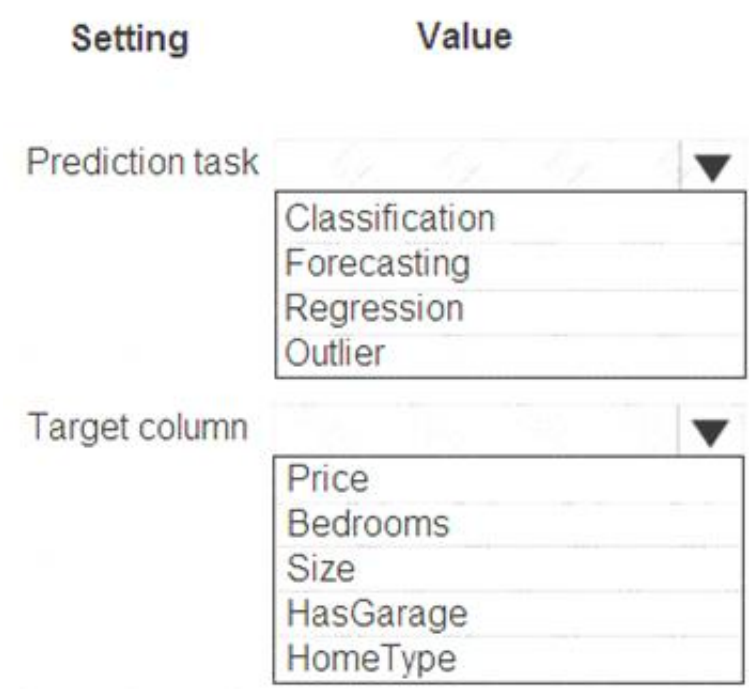
Answer: A

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 23
- (Exam Topic 3)
You have a dataset that includes home sales data for a city. The dataset includes the following columns.

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

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



A. Mastered
B. Not Mastered

Answer: A

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

NEW QUESTION 26
- (Exam Topic 3)
You have a multi-class image classification deep learning model that uses a set of labeled photographs. You create the following code to select hyperparameter values when training the model.

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

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

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

A. Mastered
B. Not Mastered

Answer: A

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

NEW QUESTION 29
- (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 30
- (Exam Topic 3)
You create a multi-class image classification deep learning model that uses a set of labeled images. You create a script file named train.py that uses the PyTorch 1.3 framework to train the model.
You must run the script by using an estimator. The code must not require any additional Python libraries to be installed in the environment for the estimator. The time required for model training must be minimized.
You need to define the estimator that will be used to run the script. Which estimator type should you use?

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

Answer: B

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

NEW QUESTION 34
- (Exam Topic 3)
You are building a binary classification model by using a supplied training set. The training set is imbalanced between two classes.
You need to resolve the data imbalance.
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. Penalize the classification
B. Resample the data set using under sampling or oversampling
C. Generate synthetic samples in the minority class.
D. Use accuracy as the evaluation metric of the model.
E. Normalize the training feature set.

Answer: ABD

Explanation:
References:
https://machinelearningmastery.com/tactics-to-combat-imbalanced-classes-in-your-machine-learning-dataset/

NEW QUESTION 38
- (Exam Topic 3)
You are a data scientist working for a hotel booking website company. You use the Azure Machine Learning service to train a model that identifies fraudulent transactions.
You must deploy the model as an Azure Machine Learning real-time web service using the Model.deploy method in the Azure Machine Learning SDK. The deployed web service must return real-time predictions of fraud based on transaction data input.
You need to create the script that is specified as the entry_script parameter for the InferenceConfig class used to deploy the model.
What should the entry script do?

A. Start a node on the inference cluster where the web service is deployed.
B. Register the model with appropriate tags and properties.
C. Create a Conda environment for the web service compute and install the necessary Python packages.
D. Load the model and use it to predict labels from input data.
E. Specify the number of cores and the amount of memory required for the inference compute.

Answer: D

Explanation:
The entry script receives data submitted to a deployed web service and passes it to the model. It then takes the response returned by the model and returns that to the client. The script is specific to your model. It must understand the data that the model expects and returns.
The two things you need to accomplish in your entry script are: Loading your model (using a function called init())
Running your model on input data (using a function called run()) Reference:
https://docs.microsoft.com/en-us/azure/machine-learning/how-to-deploy-and-where

NEW QUESTION 41
- (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

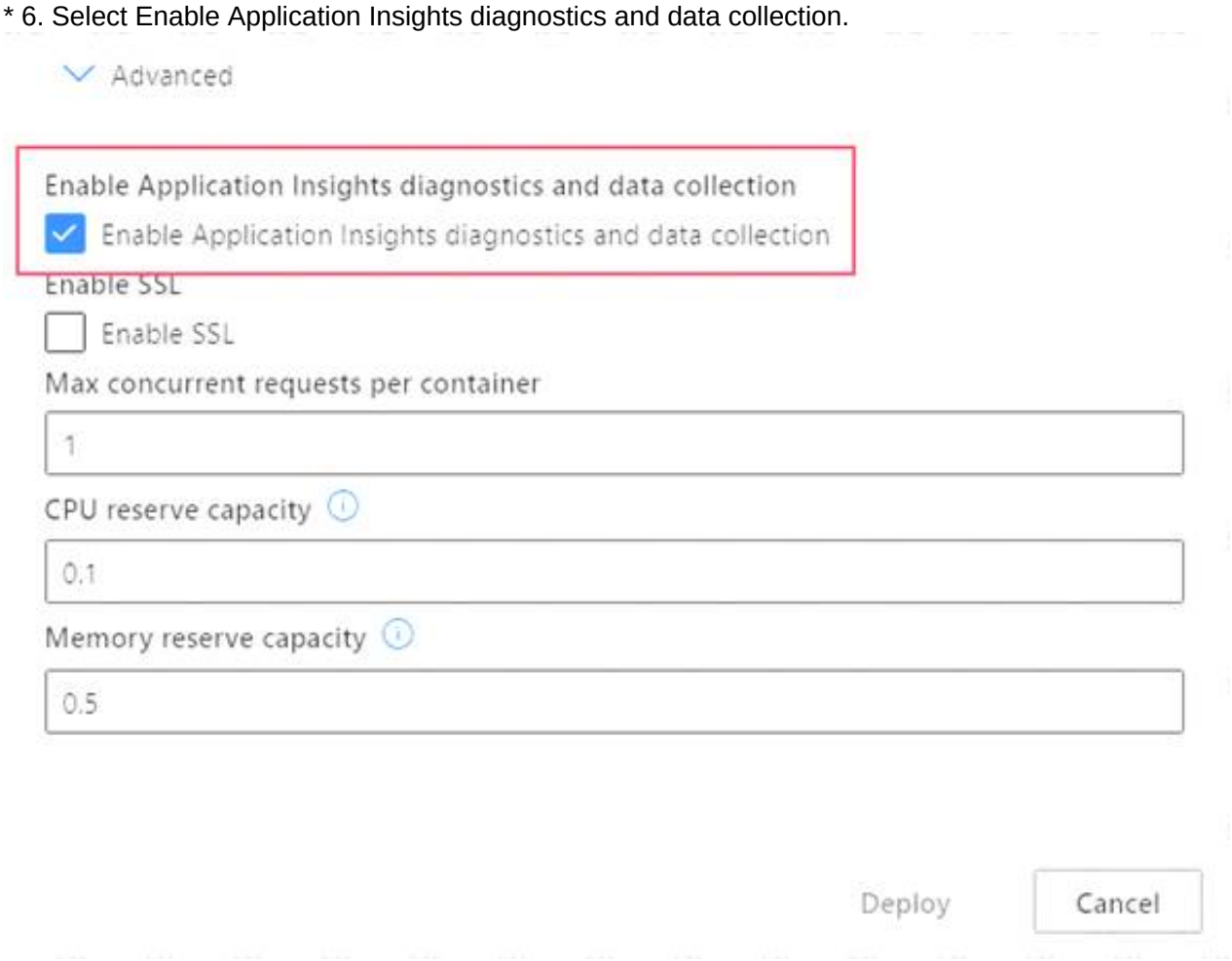
Explanation:
To translate a corpus of English text to French, we need to build a recurrent neural network (RNN).
Note: RNNs are designed to take sequences of text as inputs or return sequences of text as outputs, or both. They're called recurrent because the network's hidden layers have a loop in which 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 46
- (Exam Topic 3)
You deploy a real-time inference service for a trained model.
The deployed model supports a business-critical application, and it is important to be able to monitor the data submitted to the web service and the predictions the data generates.
You need to implement a monitoring solution for the deployed model using minimal administrative effort. What should you do?

A. View the explanation for the registered model in Azure ML studio.
B. Enable Azure Application Insights for the service endpoint and view logged data in the Azure portal.
C. Create an ML Flow tracking URI that references the endpoint, and view the data logged by ML Flow.
D. View the log files generated by the experiment used to train the model.

Answer: B

Explanation:
Configure logging with Azure Machine Learning studio
You can also enable Azure Application Insights from Azure Machine Learning studio. When you're ready to deploy your model as a web service, use the following steps to enable Application Insights:
* 1. Sign in to the studio at https://ml.azure.com.
* 2. Go to Models and select the model you want to deploy.
* 3. Select +Deploy.
* 4. Populate the Deploy model form.
* 5. Expand the Advanced menu.
* 6. Select Enable Application Insights diagnostics and data collection.



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

NEW QUESTION 47
- (Exam Topic 3)
You use the following code to run a script as an experiment in Azure Machine Learning:

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

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

A. files = run.get_properties()
B. files= run.get_file_names()
C. files = run.get_details_with_logs()
D. files = run.get_metrics()
E. files = run.get_details()

Answer: B

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

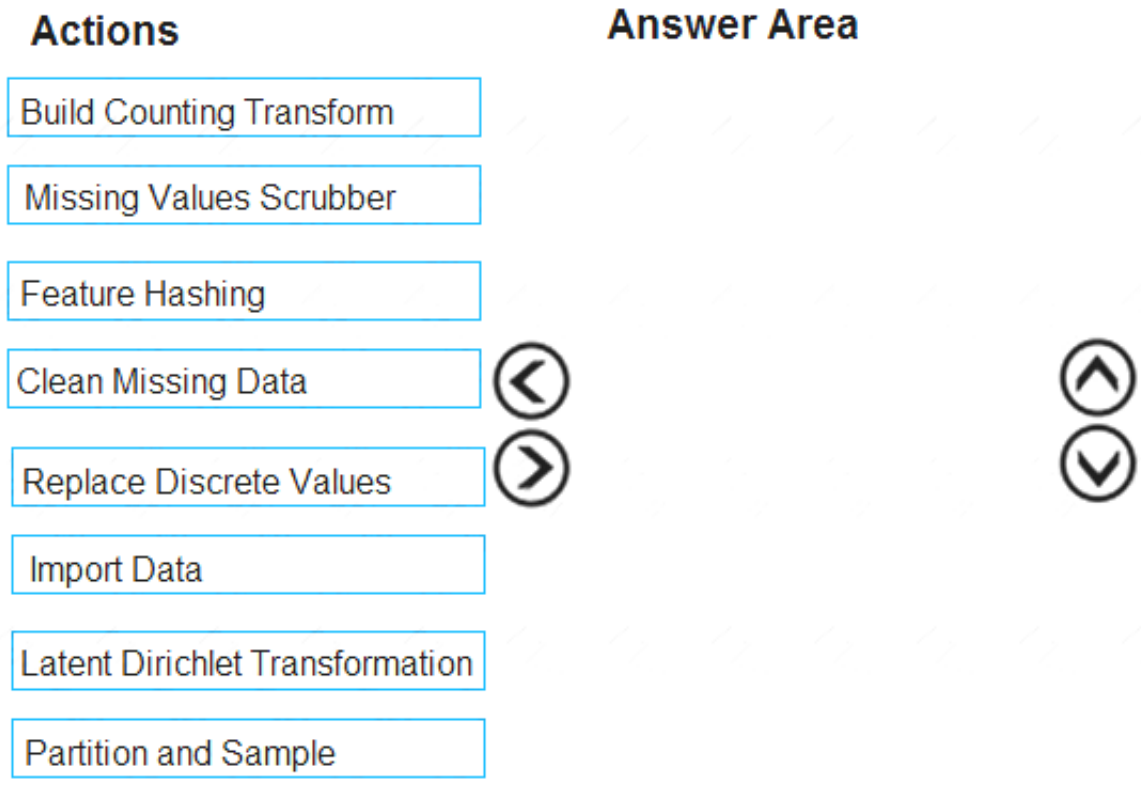
NEW QUESTION 49
- (Exam Topic 3)
You create an Azure Machine Learning workspace named ML-workspace. You also create an Azure Databricks workspace named DB-workspace. DB-workspace contains a cluster named DB-cluster.
You must use DB-cluster to run experiments from notebooks that you import into DB-workspace.
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 51
- (Exam Topic 3)
You are creating an experiment by using Azure Machine Learning Studio.
You must divide the data into four subsets for evaluation. There is a high degree of missing values in the data. You must prepare the data for analysis.
You need to select appropriate methods for producing the experiment.
Which three modules should you run in sequence? To answer, move the appropriate actions from the list of actions to the answer area and arrange them in the correct order.
NOTE: More than one order of answer choices is correct. You will receive credit for any of the correct orders you select.



A. Mastered
B. Not Mastered

Answer: A

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

NEW QUESTION 52
- (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 creating a new experiment in Azure Learning learning Studio.
One class has a much smaller number of observations than the other classes in the training
You need to select an appropriate data sampling strategy to compensate for the class imbalance. Solution: You use the Synthetic Minority Oversampling Technique (SMOTE) sampling mode. Does the solution meet the goal?

A. Yes
B. No

Answer: A

Explanation:
SMOTE is used to increase the number of underepresented cases in a dataset used for machine learning. SMOTE is a better way of increasing the number of rare cases than simply duplicating existing cases.
References:
https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/smote

NEW QUESTION 57
- (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 59
- (Exam Topic 3)
You create a multi-class image classification deep learning model that uses the PyTorch deep learning framework.
You must configure Azure Machine Learning Hyperdrive to optimize the hyperparameters for the classification model.
You need to define a primary metric to determine the hyperparameter values that result in the model with the best accuracy score.
Which three actions must you perform? Each correct answer presents part of the solution.
NOTE: Each correct selection is worth one point.

A. Set the primary_metric_goal of the estimator used to run the bird_classifier_train.py script to maximize.
B. Add code to the bird_classifier_train.py script to calculate the validation loss of the model and log it as a float value with the key loss.
C. Set the primary_metric_goal of the estimator used to run the bird_classifier_train.py script to minimize.
D. Set the primary_metric_name of the estimator used to run the bird_classifier_train.py script to accuracy.
E. Set the primary_metric_name of the estimator used to run the bird_classifier_train.py script to loss.
F. Add code to the bird_classifier_train.py script to calculate the validation accuracy of the model and log it as a float value with the key accuracy.

Answer: ADF

Explanation:
AD:
primary_metric_name="accuracy", primary_metric_goal=PrimaryMetricGoal.MAXIMIZE
Optimize the runs to maximize "accuracy". Make sure to log this value in your training script. Note:
primary_metric_name: The name of the primary metric to optimize. The name of the primary metric needs to exactly match the name of the metric logged by the training script.
primary_metric_goal: It can be either PrimaryMetricGoal.MAXIMIZE or PrimaryMetricGoal.MINIMIZE and determines whether the primary metric will be maximized or minimized when evaluating the runs.
F: The training script calculates the val_accuracy and logs it as "accuracy", which is used as the primary metric.

NEW QUESTION 61
- (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.sklearn import SKLearn
sk_est = SKLearn(source_directory='./scripts',
 compute_target=aml-compute,
 entry_script='train.py')
```

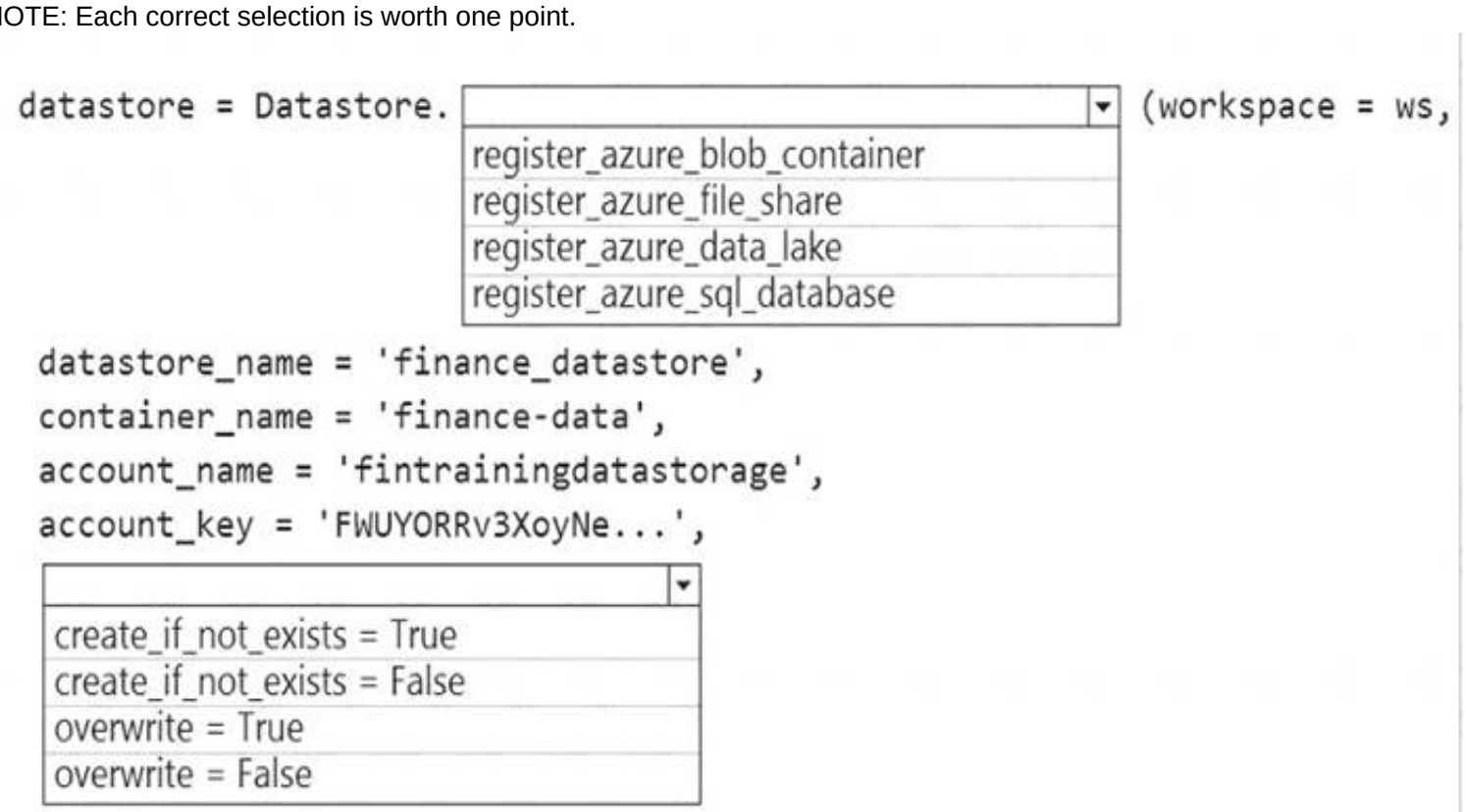
Does the solution meet the goal?

A. Yes
B. No

Answer: A

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 64
- (Exam Topic 3)
The finance team asks you to train a model using data in an Azure Storage blob container named finance-data. You need to register the container as a datastore in an Azure Machine Learning workspace and ensure that an
error will be raised if the container does not exist.
How should you complete the code? To answer, select the appropriate options in the answer area.
NOTE: Each correct selection is worth one point.



A. Mastered
B. Not Mastered

Answer: A

Explanation:
Box 1: register_azure_blob_container
Register an Azure Blob Container to the datastore.
Box 2: create_if_not_exists = False
Create the file share if it does not exists, defaults to False. Reference:
https://docs.microsoft.com/en-us/python/api/azureml-core/azureml.core.datastore.datastore

NEW QUESTION 69
- (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 72

- (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 creating a model to predict the price of a student's artwork depending on the following variables: the student's length of education, degree type, and art form.
You start by creating a linear regression model. You need to evaluate the linear regression model.
Solution: Use the following metrics: Mean Absolute Error, Root Mean Absolute Error, Relative Absolute Error, Relative Squared Error, and the Coefficient of Determination.
Does the solution meet the goal?

A. Yes
B. No

Answer: A

Explanation:
The following metrics are reported for evaluating regression models. When you compare models, they are ranked by the metric you select for evaluation.
Mean absolute error (MAE) measures how close the predictions are to the actual outcomes; thus, a lower score is better.
Root mean squared error (RMSE) creates a single value that summarizes the error in the model. By squaring the difference, the metric disregards the difference between over-prediction and under-prediction.
Relative absolute error (RAE) is the relative absolute difference between expected and actual values; relative because the mean difference is divided by the arithmetic mean.
Relative squared error (RSE) similarly normalizes the total squared error of the predicted values by dividing by the total squared error of the actual values.
Mean Zero One Error (MZOE) indicates whether the prediction was correct or not. In other words: ZeroOneLoss(x,y) = 1 when x!=y; otherwise 0.
Coefficient of determination, often referred to as R2, represents the predictive power of the model as a value between 0 and 1. Zero means the model is random (explains nothing); 1 means there is a perfect fit. However, caution should be used in interpreting R2 values, as low values can be entirely normal and high values can be suspect.
AUC.
References:
https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/evaluate-model

NEW QUESTION 76
- (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?

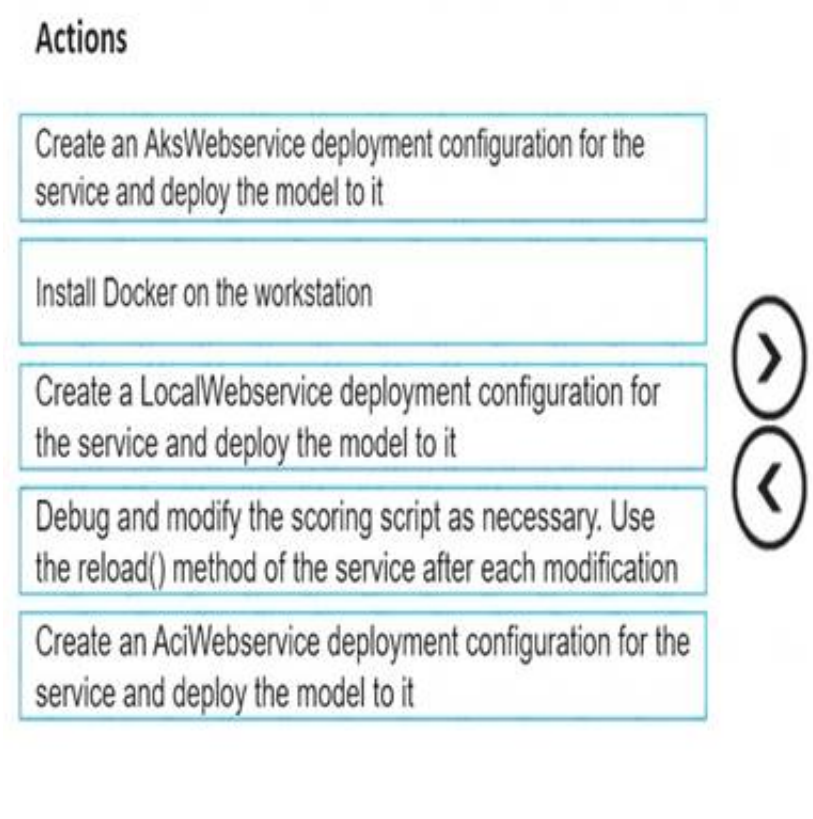
A. Yes
B. No

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 81
- (Exam Topic 3)
You train and register a model by using the Azure Machine Learning SDK on a local workstation. Python 3.6 and Visual Studio Code are installed on the

workstation.
When you try to deploy the model into production as an Azure Kubernetes Service (AKS)-based web service, you experience an error in the scoring script that causes deployment to fail.
You need to debug the service on the local workstation before deploying the service to production.
Which four actions should you perform in sequence? To answer, move the appropriate actions from the list of actions to the answer area and arrange them in the correct order.



A. Mastered
B. Not Mastered

Answer: A

Explanation:
Graphical user interface, text, application, email Description automatically generated
Step 1: Install Docker on the workstation
Prerequisites include having a working Docker installation on your local system. Build or download the dockerfile to the compute node.
Step 2: Create an AksWebservice deployment configuration and deploy the model to it
To deploy a model to Azure Kubernetes Service, create a deployment configuration that describes the compute resources needed.
# If deploying to a cluster configured for dev/test, ensure that it was created with enough
# cores and memory to handle this deployment configuration. Note that memory is also used by
# things such as dependencies and AML components.
deployment_config = AksWebservice.deploy_configuration(cpu_cores = 1, memory_gb = 1)
service = Model.deploy(ws, "myservice", [model], inference_config, deployment_config, aks_target) service.wait_for_deployment(show_output = True)
print(service.state) print(service.get_logs())
Step 3: Create a LocalWebservice deployment configuration for the service and deploy the model to it
To deploy locally, modify your code to use LocalWebservice.deploy_configuration() to create a deployment configuration. Then use Model.deploy() to deploy the service.
Step 4: Debug and modify the scoring script as necessary. Use the reload() method of the service after each modification.
During local testing, you may need to update the score.py file to add logging or attempt to resolve any problems that you've discovered. To reload changes to the score.py file, use reload(). For example, the following code reloads the script for the service, and then sends data to it.
Reference:
https://docs.microsoft.com/en-us/azure/machine-learning/how-to-deploy-azure-kubernetes-service https://docs.microsoft.com/en-us/azure/machine-learning/how-to-troubleshoot-deployment-local

NEW QUESTION 85
- (Exam Topic 3)
You have a Python data frame named salesData in the following format:

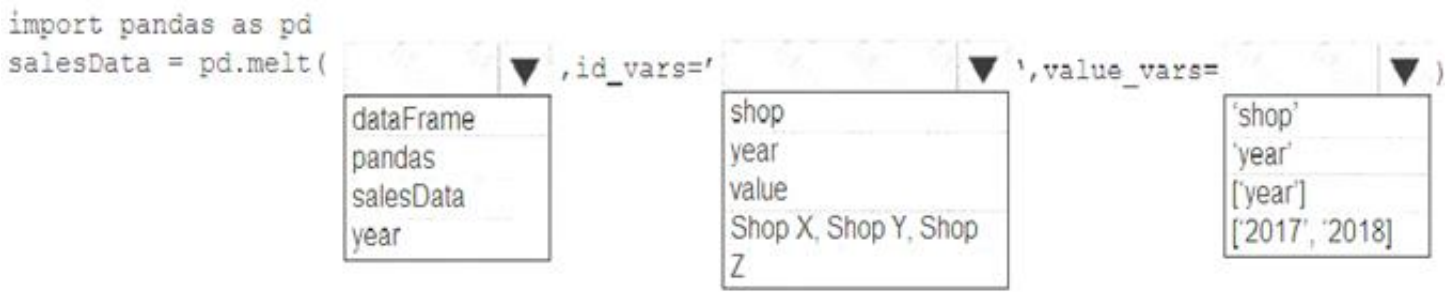
	shop	2017	2018
0	Shop X	34	25
1	Shop Y	65	76
2	Shop Z	48	55

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



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 88
- (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?

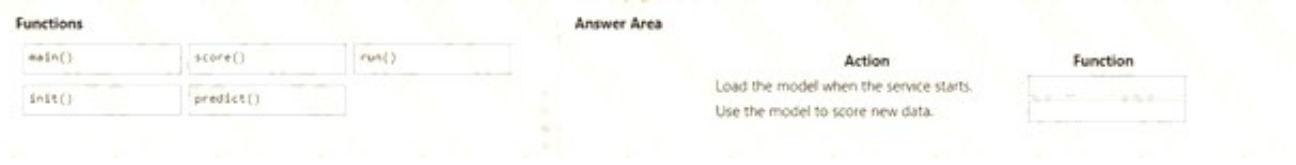
A. Yes
B. No

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 93
- (Exam Topic 3)
You use Azure Machine Learning to deploy a model as a real-time web service.

You need to create an entry script for the service that ensures that the model is loaded when the service starts and is used to score new data as it is received.
Which functions should you include in the script? To answer, drag the appropriate functions to the correct actions. Each function may be used once, more than once, or not at all. You may need to drag the split bar between panes or scroll to view content
NOTE: Each correct selection is worth one point.



A. Mastered
B. Not Mastered

Answer: A

Explanation:
Box 1: init()
The entry script has only two required functions, init() and run(data). These functions are used to initialize the service at startup and run the model using request data passed in by a client. The rest of the script handles loading and running the model(s).
Box 2: run() Reference:
https://docs.microsoft.com/en-us/azure/machine-learning/how-to-deploy-existing-model

NEW QUESTION 98
- (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

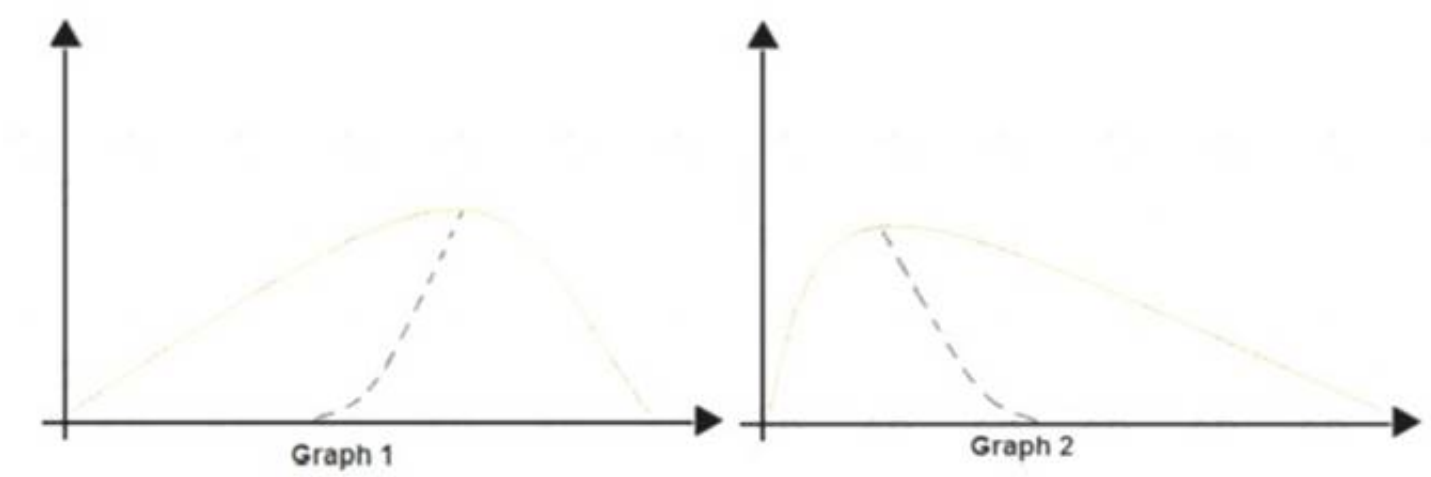
NEW QUESTION 99
- (Exam Topic 3)
Note: This question is part of a series of questions that present the same scenario. Each question in the series contains a unique solution that might meet the stated goals. Some question sets might have more than one correct solution, while others might not have a correct 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 creating a new experiment in Azure Machine Learning Studio.
One class has a much smaller number of observations than the other classes in the training set. You need to select an appropriate data sampling strategy to compensate for the class imbalance. Solution: You use the Scale and Reduce sampling mode.
Does the solution meet the goal?

A. Yes
B. No

Answer: B

Explanation:
Instead use the Synthetic Minority Oversampling Technique (SMOTE) sampling mode.
Note: SMOTE is used to increase the number of underepresented cases in a dataset used for machine learning. SMOTE is a better way of increasing the number of rare cases than simply duplicating existing cases.
References:
https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/smote

NEW QUESTION 104
- (Exam Topic 3)
You are analyzing the asymmetry in a statistical distribution.
The following image contains two density curves that show the probability distribution of two datasets.



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

Question	Answer choice
Which type of distribution is shown for the dataset density curve of Graph 1?	Negative skew Positive skew Normal distribution Bimodal distribution
Which type of distribution is shown for the dataset density curve of Graph 2?	Negative skew Positive skew Normal distribution Bimodal distribution

A. Mastered
B. Not Mastered

Answer: A

Explanation:
Box 1: Positive skew
Positive skew values means the distribution is skewed to the right. Box 2: Negative skew
Negative skewness values mean the distribution is skewed to the left. References:
https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/compute-elementary-statistic

NEW QUESTION 108
- (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 113
- (Exam Topic 3)
Your team is building a data engineering and data science development environment. The environment must support the following requirements:
- support Python and Scala
- compose data storage, movement, and processing services into automated data pipelines
- the same tool should be used for the orchestration of both data engineering and data science
- support workload isolation and interactive workloads
- enable scaling across a cluster of machines You need to create the environment.

What should you do?

A. Build the environment in Apache Hive for HDInsight and use Azure Data Factory for orchestration.
B. Build the environment in Azure Databricks and use Azure Data Factory for orchestration.
C. Build the environment in Apache Spark for HDInsight and use Azure Container Instances for orchestration.
D. Build the environment in Azure Databricks and use Azure Container Instances for orchestration.

Answer: B

Explanation:
In Azure Databricks, we can create two different types of clusters.
- Standard, these are the default clusters and can be used with Python, R, Scala and SQL
- High-concurrency

Azure Databricks is fully integrated with Azure Data Factory.

NEW QUESTION 116
- (Exam Topic 3)
You plan to use the Hyperdrive feature of Azure Machine Learning to determine the optimal hyperparameter values when training a model.
You must use Hyperdrive to try combinations of the following hyperparameter values:
• learning_rate: any value between 0.001 and 0.1
• batch_size: 16, 32, or 64
You need to configure the search space for the Hyperdrive experiment.
Which two parameter expressions should you use? Each correct answer presents part of the solution. NOTE: Each correct selection is worth one point.

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

Answer: BD

Explanation:
B: Continuous hyperparameters are specified as a distribution over a continuous range of values. Supported distributions include:
- uniform(low, high) - Returns a value uniformly distributed between low and high

D: Discrete hyperparameters are specified as a choice among discrete values. choice can be:
- one or more comma-separated values
- a range object
- any arbitrary list object Reference:

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

NEW QUESTION 117
- (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: for label_val in label_vals:
run.log('Label Values', label_val) Does the solution meet the goal?

A. Yes
B. No

Answer: A

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

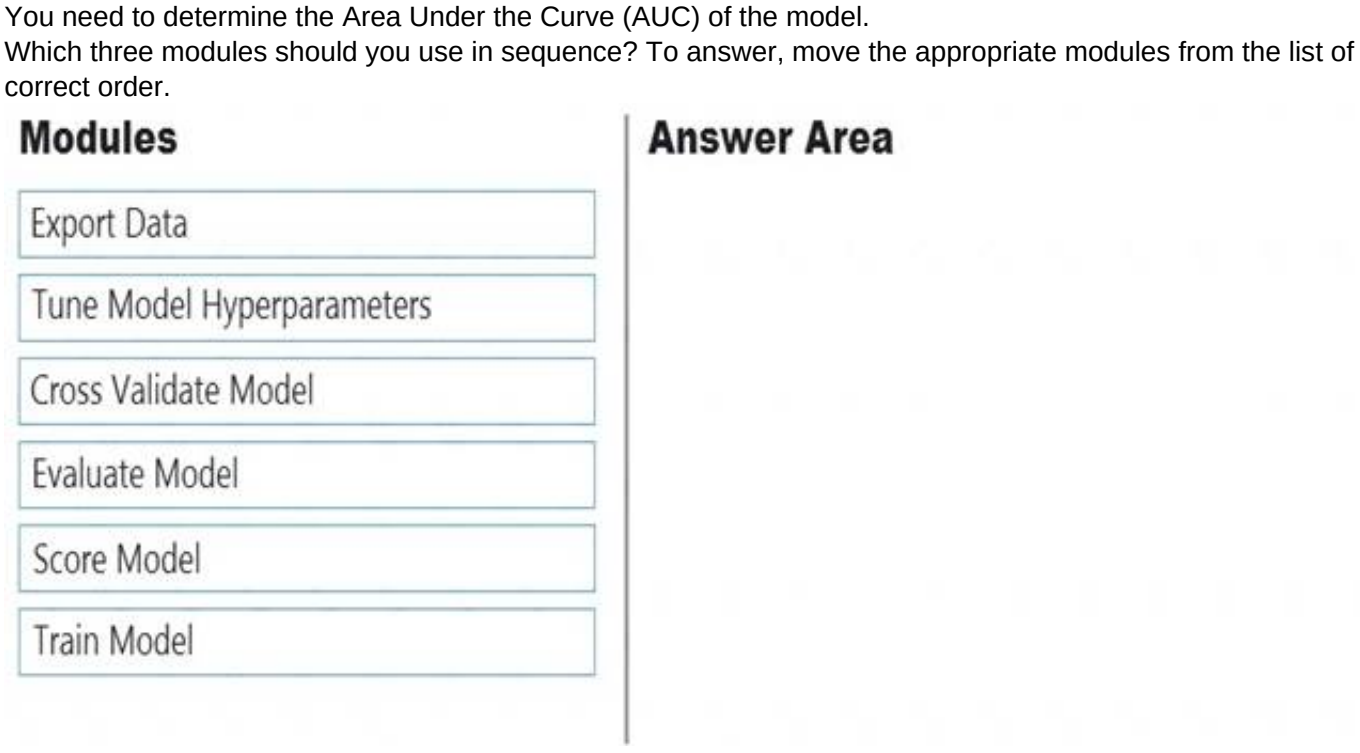
NEW QUESTION 121
- (Exam Topic 3)
You plan to use the Hyperdrive feature of Azure Machine Learning to determine the optimal hyperparameter values when training a model.
You must use Hyperdrive to try combinations of the following hyperparameter values. You must not apply an early termination policy.
learning_rate: any value between 0.001 and 0.1
• batch_size: 16, 32, or 64
You need to configure the sampling method for the Hyperdrive experiment
Which two sampling methods can you use? Each correct answer is a complete solution. NOTE: Each correct selection is worth one point.

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

Answer: CD

Explanation:
C: Bayesian sampling is based on the Bayesian optimization algorithm and makes intelligent choices on the hyperparameter values to sample next. It picks the sample based on how the previous samples performed, such that the new sample improves the reported primary metric.
Bayesian sampling does not support any early termination policy Example:
from azureml.train.hyperdrive import BayesianParameterSampling from azureml.train.hyperdrive import uniform, choice param_sampling = BayesianParameterSampling({ "learning_rate": uniform(0.05, 0.1),
"batch_size": choice(16, 32, 64, 128)
}
)
D: In random sampling, hyperparameter values are randomly selected from the defined search space. Random sampling allows the search space to include both discrete and continuous hyperparameters.
Reference:
https://docs.microsoft.com/en-us/azure/machine-learning/how-to-tune-hyperparameters

NEW QUESTION 125
- (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.

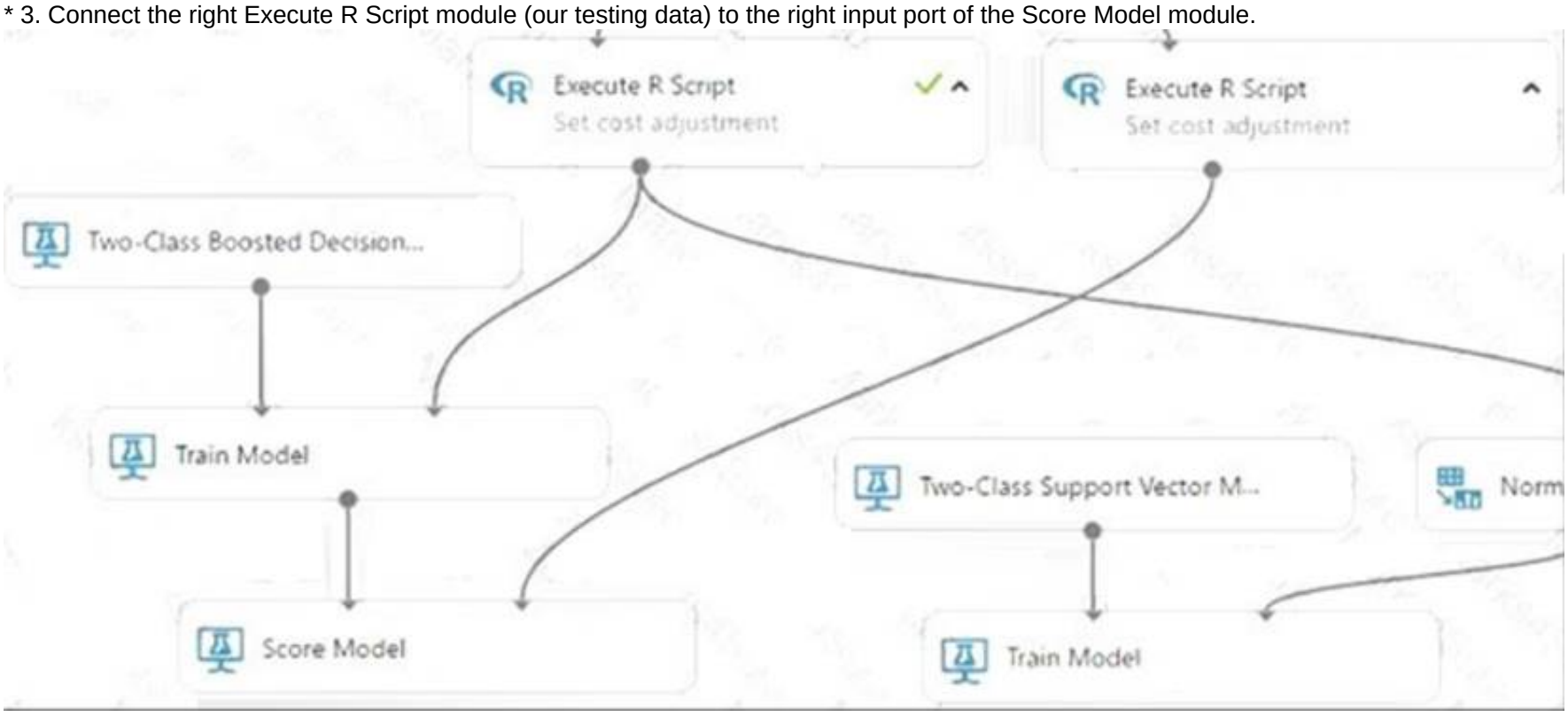


A. Mastered
B. Not Mastered

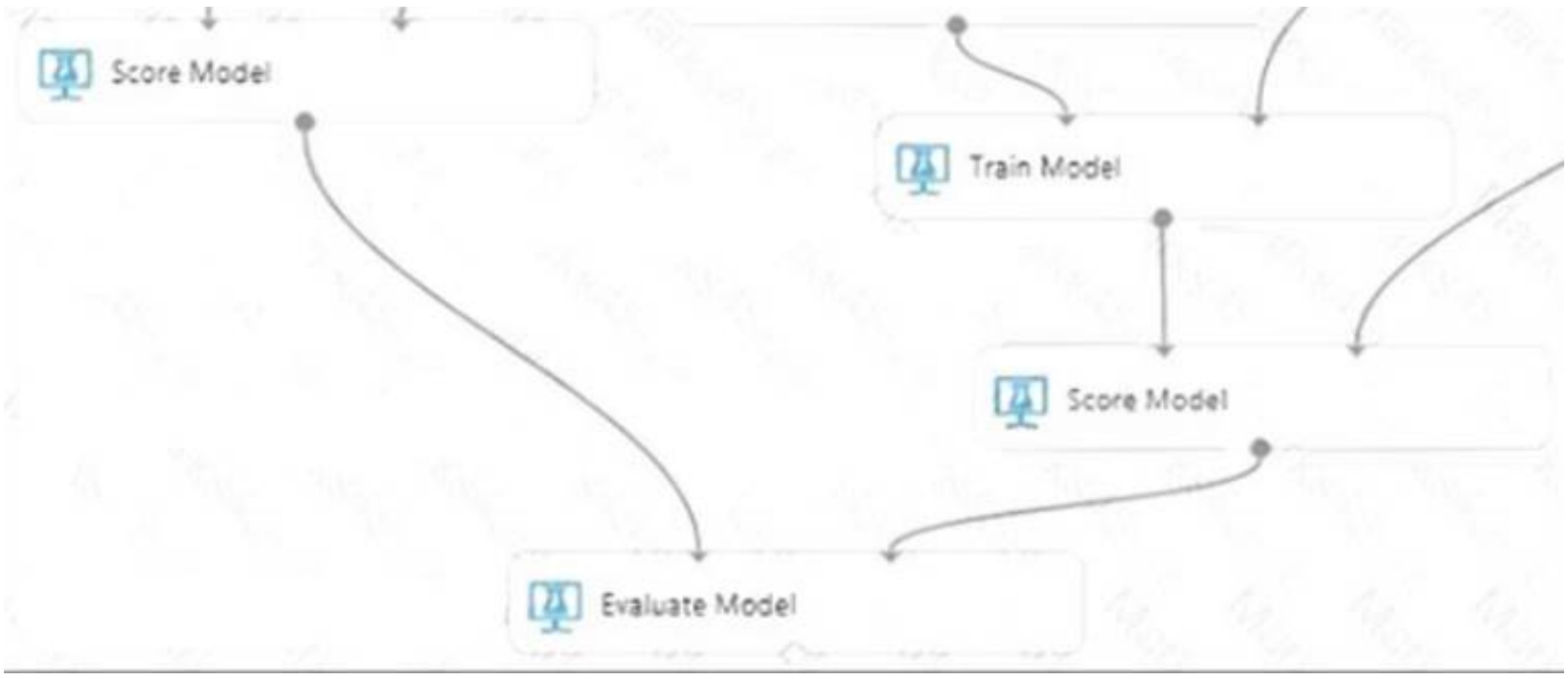
Answer: A

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 128
- (Exam Topic 3)
You publish a batch inferencing pipeline that will be used by a business application.
The application developers need to know which information should be submitted to and returned by the REST interface for the published pipeline.
You need to identify the information required in the REST request and returned as a response from the published pipeline.
Which values should you use in the REST request and to expect in the response? To answer, select the appropriate options in the answer area.
NOTE: Each correct selection is worth one point.

Answer Area

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

A. Mastered
B. Not Mastered

Answer: A

Explanation:
Box 1: JSON containing an OAuth bearer token Specify your authentication header in the request.
To run the pipeline from the REST endpoint, you need an OAuth2 Bearer-type authentication header. Box 2: JSON containing the experiment name
Add a JSON payload object that has the experiment name. Example:
rest_endpoint = published_pipeline.endpoint response = requests.post(rest_endpoint, headers=auth_header, json={"ExperimentName": "batch_scoring", "ParameterAssignments": {"process_count_per_node": 6}}) run_id = response.json()["Id"]
Box 3: JSON containing the run ID
Make the request to trigger the run. Include code to access the Id key from the response dictionary to get the value of the run ID.
Reference:
https://docs.microsoft.com/en-us/azure/machine-learning/tutorial-pipeline-batch-scoring-classification

NEW QUESTION 130
- (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.

```
from sklearn import linear_model
import
```

- pandas as df
- numpy as df
- scipy as df

```
dataset = df.read_csv("data\\ProductSales.csv")
ProductCategoryMapping = {"Bikes":1, "Cars":2, "Boats": 3,
"Vans": 4}
dataset['ProductCategoryMapping'] =
dataset['ProductCategory'].
```

- map[ProductCategoryMapping]
- reduce[ProductCategoryMapping]
- transpose[ProductCategoryMapping]

```
regr = linear_model.LinearRegression()
X_train = dataset[['ProductCategoryMapping', 'ProductSize',
'ProductCost']]
y_train = dataset[['Sales']]
regr.fit(X_train, y_train)
```

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 133
- (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.

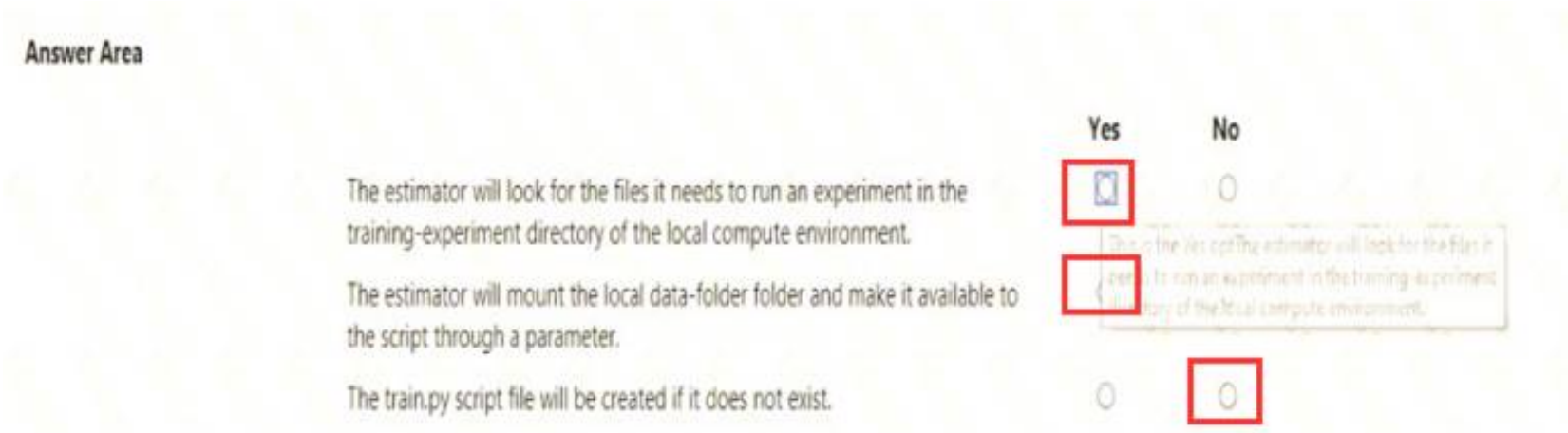
Answer Area

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

A. Mastered
B. Not Mastered

Answer: A

Explanation:



NEW QUESTION 137
- (Exam Topic 3)
You have the following code. The code prepares an experiment to run a script:

```
from azureml.core import Workspace, Experiment, Run, ScriptRunConfig

ws = Workspace.from_config()
script_config = ScriptRunConfig(source_directory='experiment_files',
                                script='experiment.py')

script_experiment = Experiment(workspace=ws, name='script-experiment')
```

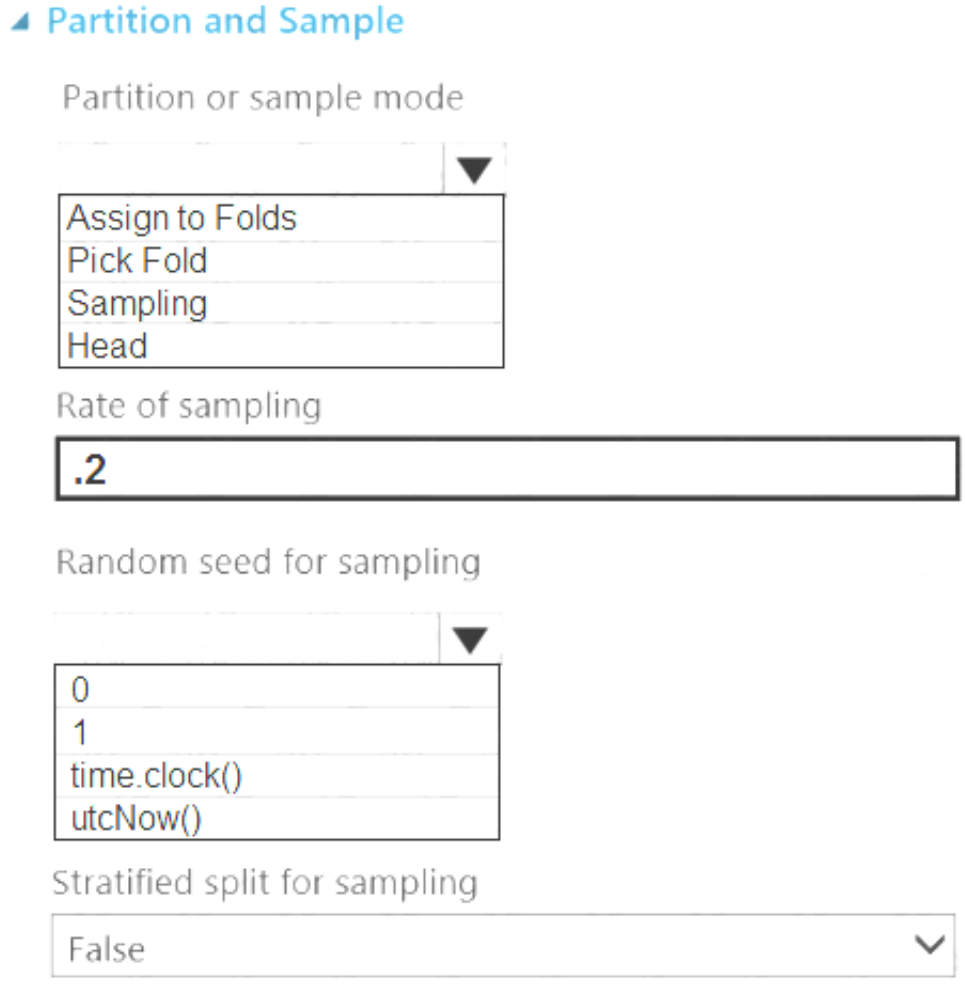
The experiment must be run on local computer using the default environment. You need to add code to start the experiment and run the script.
Which code segment should you use?

A. run = script_experiment.start_logging()
B. run = Run(experiment=script_experiment)
C. ws.get_run(run_id=experiment.id)
D. run = script_experiment.submit(config=script_config)

Answer: D

Explanation:
The experiment class submit method submits an experiment and return the active created run.
Syntax: submit(config, tags=None, **kwargs) Reference:
https://docs.microsoft.com/en-us/python/api/azureml-core/azureml.core.experiment.experiment

NEW QUESTION 140
- (Exam Topic 3)
You are retrieving data from a large datastore by using Azure Machine Learning Studio.
You must create a subset of the data for testing purposes using a random sampling seed based on the system clock.
You add the Partition and Sample module to your experiment. You need to select the properties for the module.
Which values should you select? To answer, select the appropriate options in the answer area.
NOTE: Each correct selection is worth one point.

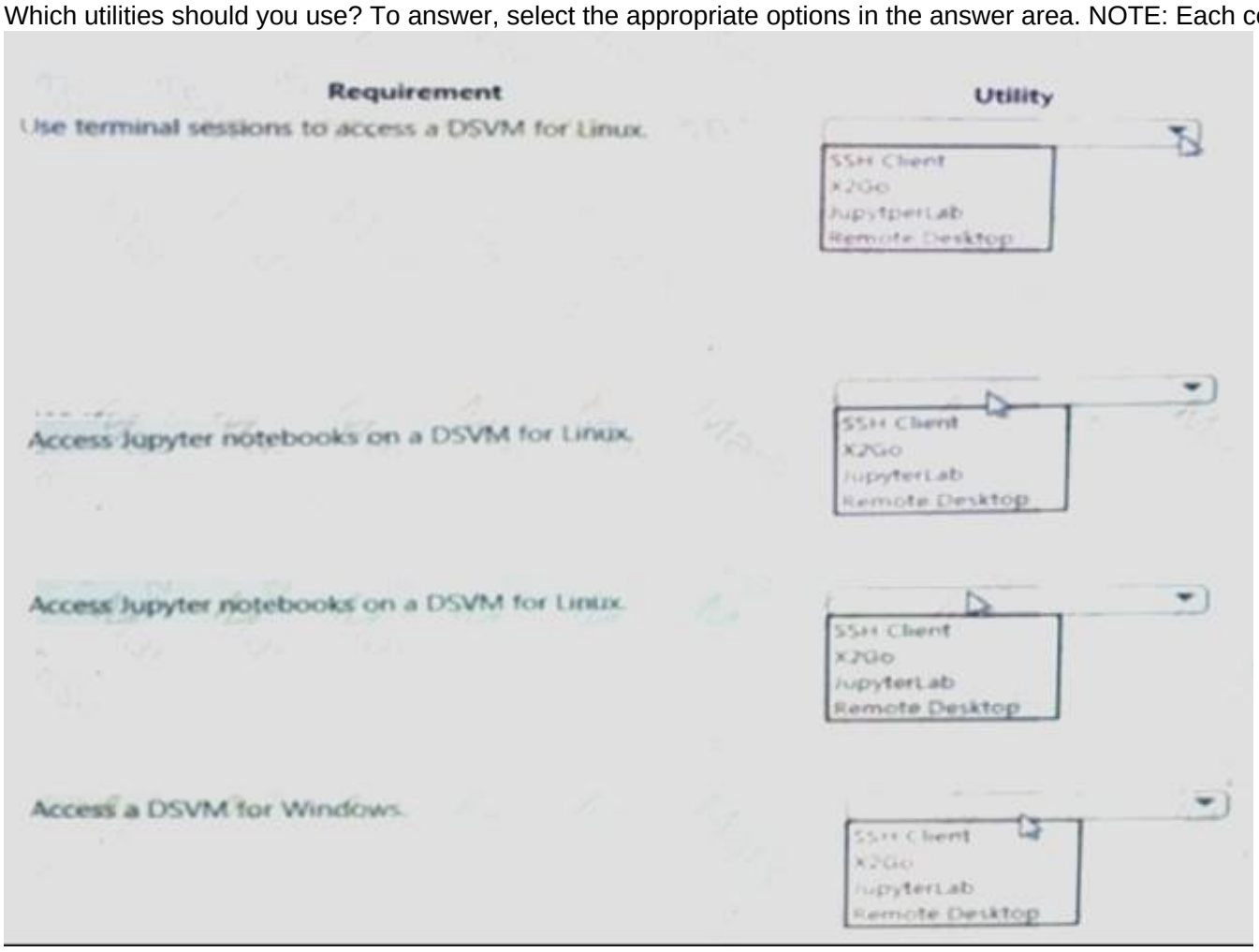


A. Mastered
B. Not Mastered

Answer: A

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

NEW QUESTION 143
- (Exam Topic 3)
You use Data Science Virtual Machines (DSVMs) for Windows and Linux in Azure. You need to access the DSVMs.
Which utilities should you use? To answer, select the appropriate options in the answer area. NOTE: Each correct selection is worth one point.



A. Mastered
B. Not Mastered

Answer: A

Explanation:

NEW QUESTION 144
- (Exam Topic 3)
You are creating a new Azure Machine Learning pipeline using the designer.
The pipeline must train a model using data in a comma-separated values (CSV) file that is published on a website. You have not created a dataset for this file.
You need to ingest the data from the CSV file into the designer pipeline using the minimal administrative effort.
Which module should you add to the pipeline in Designer?

A. Convert to CSV
B. Enter Data Manually D
C. Import Data
D. Dataset

Answer: D

Explanation:
The preferred way to provide data to a pipeline is a Dataset object. The Dataset object points to data that lives in or is accessible from a datastore or at a Web URL. The Dataset class is abstract, so you will create an instance of either a FileDataset (referring to one or more files) or a TabularDataset that's created by from one or more files with delimited columns of data.
Example:
from azureml.core import Dataset
iris_tabular_dataset = Dataset.Tabular.from_delimited_files([(def_blob_store, 'train-dataset/iris.csv')]) Reference:
https://docs.microsoft.com/en-us/azure/machine-learning/how-to-create-your-first-pipeline

NEW QUESTION 146
- (Exam Topic 3)
You run an automated machine learning experiment in an Azure Machine Learning workspace. Information about the run is listed in the table below:

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

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

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

B)

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

C)

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

D)

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

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

Answer: A

Explanation:
The get_output method on automl_classifier returns the best run and the fitted model for the last invocation. Overloads on get_output allow you to retrieve the best run and fitted model for any logged metric or for a particular iteration.
In [ ]:
best_run, fitted_model = local_run.get_output() Reference:
https://notebooks.azure.com/azureml/projects/azureml-getting-started/html/how-to-use-azureml/automated-mach

NEW QUESTION 149
- (Exam Topic 3)
You create a deep learning model for image recognition on Azure Machine Learning service using GPU-based training.
You must deploy the model to a context that allows for real-time GPU-based inferencing. You need to configure compute resources for model inferencing.
Which compute type should you use?

A. Azure Container Instance
B. Azure Kubernetes Service
C. Field Programmable Gate Array
D. Machine Learning Compute

Answer: B

Explanation:
You can use Azure Machine Learning to deploy a GPU-enabled model as a web service. Deploying a model on Azure Kubernetes Service (AKS) is one option. The AKS cluster provides a GPU resource that is used by the model for inference.
Inference, or model scoring, is the phase where the deployed model is used to make predictions. Using GPUs instead of CPUs offers performance advantages on highly parallelizable computation.
Reference:
https://docs.microsoft.com/en-us/azure/machine-learning/how-to-deploy-inferencing-gpus

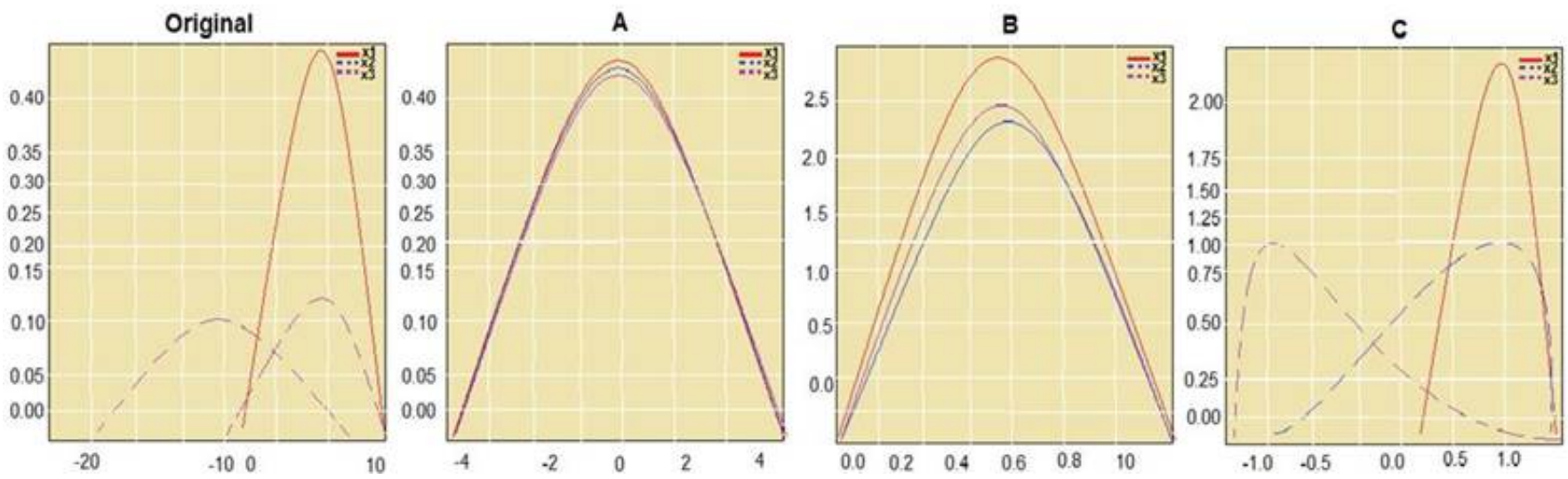
NEW QUESTION 151
- (Exam Topic 3)
You create a script that trains a convolutional neural network model over multiple epochs and logs the validation loss after each epoch. The script includes arguments for batch size and learning rate.
You identify a set of batch size and learning rate values that you 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 156
- (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.



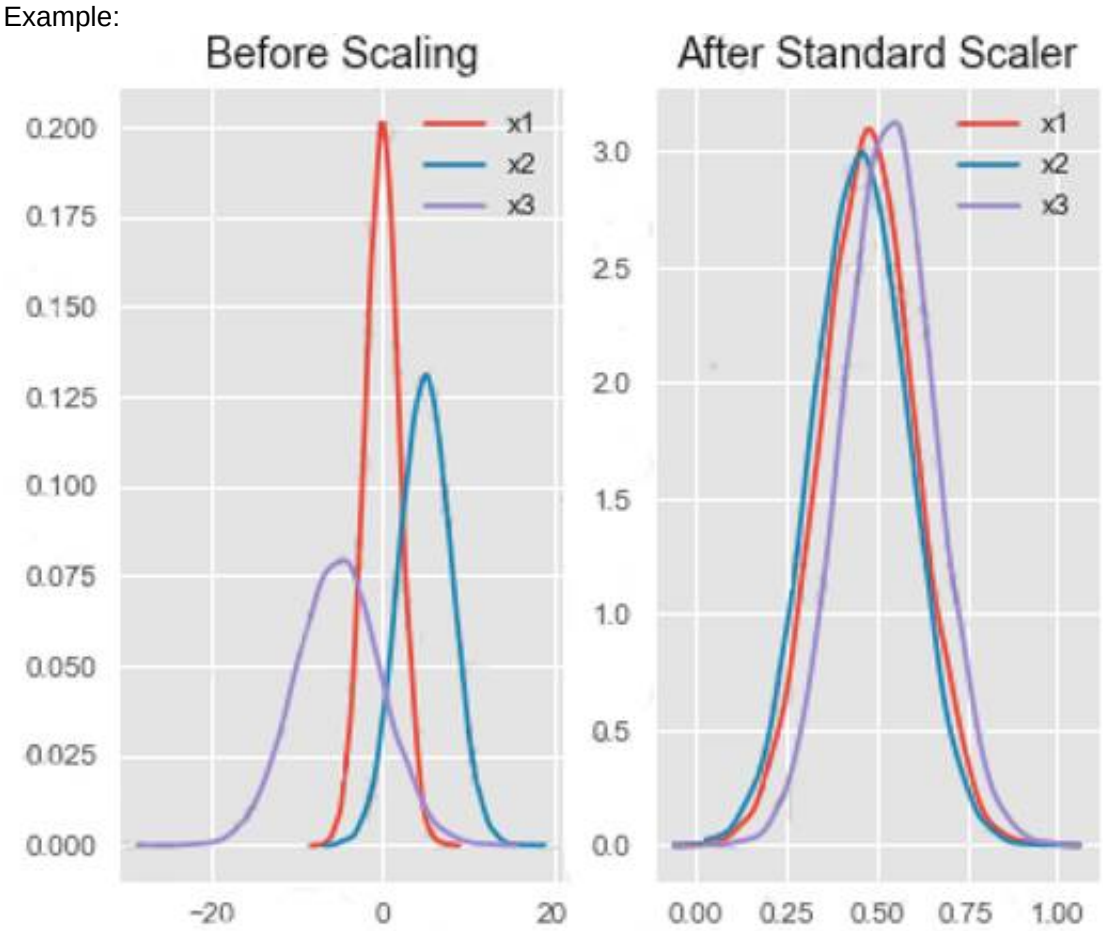
Use the drop-down menus to select the answer choice that answers each question based on the information presented in the graphic.
NOTE: Each correct selection is worth one point.

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

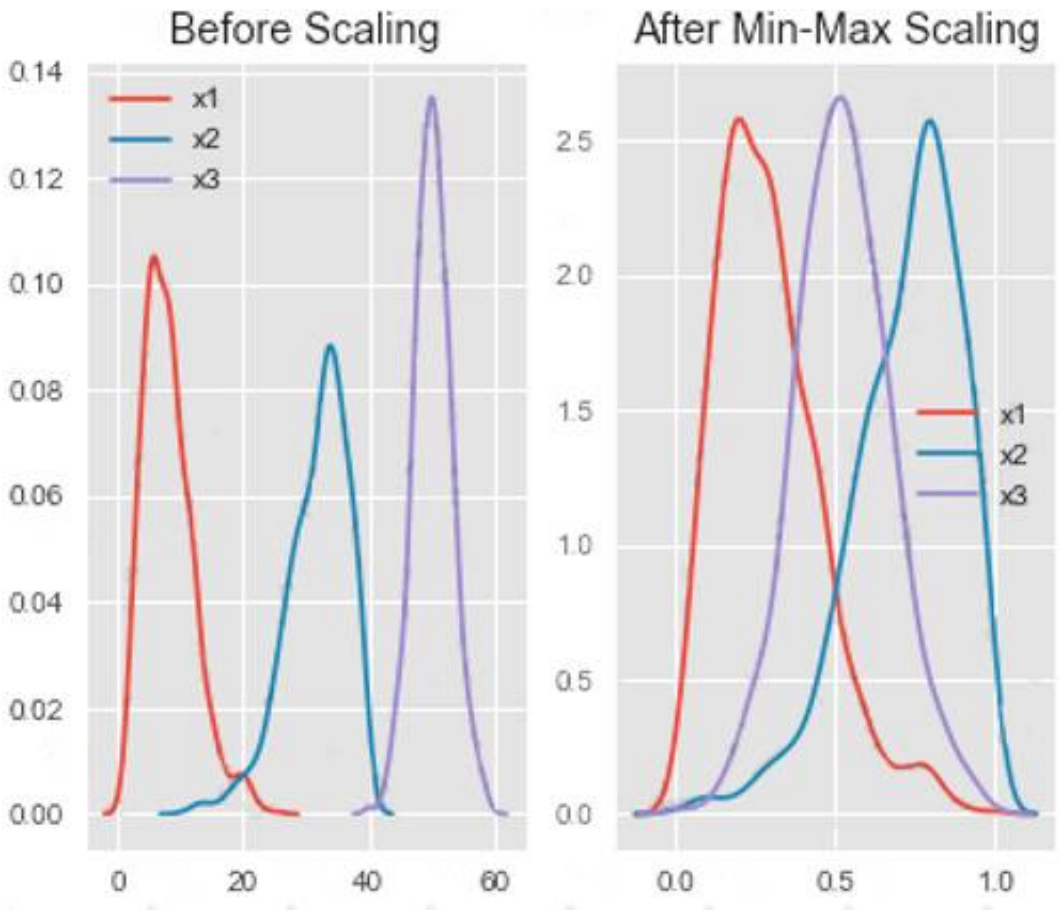
A. Mastered
B. Not Mastered

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 158
- (Exam Topic 3)
You are a lead data scientist for a project that tracks the health and migration of birds. You create a multi-class image classification deep learning model that uses a set of labeled bird photographs collected by experts.
You have 100,000 photographs of birds. All photographs use the JPG format and are stored in an Azure blob container in an Azure subscription.
You need to access the bird photograph files in the Azure blob container from the Azure Machine Learning service workspace that will be used for deep learning model training. You must minimize data movement.
What should you do?

A. Create an Azure Data Lake store and move the bird photographs to the store.
B. Create an Azure Cosmos DB database and attach the Azure Blob containing bird photographs storage to the database.
C. Create and register a dataset by using TabularDataset class that references the Azure blob storage containing bird photographs.
D. Register the Azure blob storage containing the bird photographs as a datastore in Azure Machine Learning service.
E. Copy the bird photographs to the blob datastore that was created with your Azure Machine Learning service workspace.

Answer: D

Explanation:
We recommend creating a datastore for an Azure Blob container. When you create a workspace, an Azure blob container and an Azure file share are automatically registered to the workspace.
Reference:
https://docs.microsoft.com/en-us/azure/machine-learning/how-to-access-data

NEW QUESTION 159
- (Exam Topic 3)
You use Azure Machine Learning to train a model based on a dataset named dataset1. You define a dataset monitor and create a dataset named dataset2 that contains new data.
You need to compare dataset1 and dataset2 by using the Azure Machine Learning SDK for Python. Which method of the DataDriftDetector class should you use?

A. run
B. get
C. backfill
D. update

Answer: C

Explanation:
A backfill run is used to see how data changes over time. Reference:
https://docs.microsoft.com/en-us/python/api/azureml-datadrift/azureml.datadrift.datadriftdetector.datadriftdetect

NEW QUESTION 163
- (Exam Topic 3)
You create a binary classification model. The model is registered in an Azure Machine Learning workspace. You use the Azure Machine Learning Fairness SDK to assess the model fairness.
You develop a training script for the model on a local machine.
You need to load the model fairness metrics into Azure Machine Learning studio. What should you do?

A. Implement the download_dashboard_by_upload_id function
B. Implement the creace_group_metric_sec function

C. Implement the upload_dashboard_dictionary function
D. Upload the training script

Answer: C

Explanation:
import azureml.contrib.fairness package to perform the upload:
from azureml.contrib.fairness import upload_dashboard_dictionary, download_dashboard_by_upload_id Reference:
https://docs.microsoft.com/en-us/azure/machine-learning/how-to-machine-learning-fairness-aml

NEW QUESTION 165
- (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)
```

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

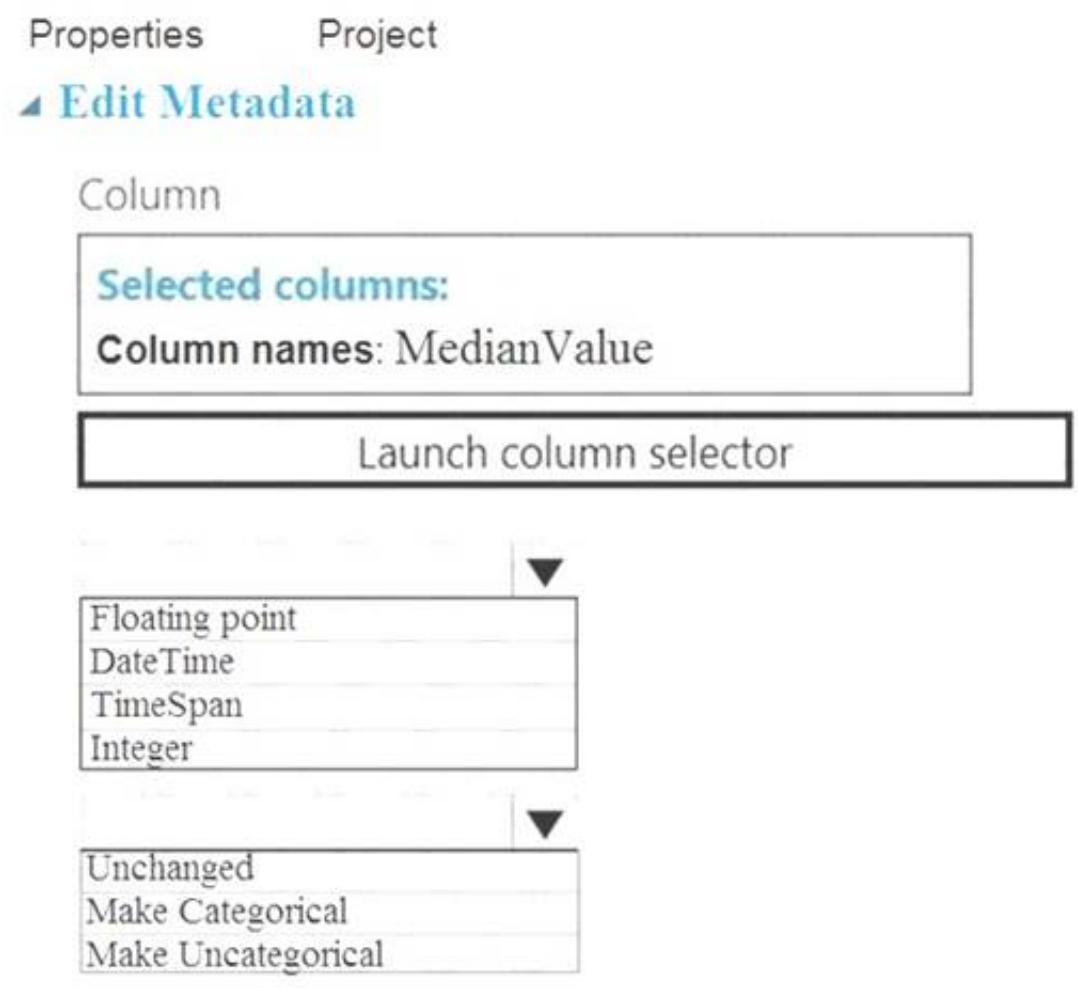
```
import numpy as np
from sklearn.metrics import roc_auc_score
from azureml.core.run import Run
run = Run.get_context()
# code to train model omitted
auc = roc_auc_score(y_test, y_predicted)
run.log("AUC", np.float(auc))
```

Does the solution meet the goal?

A. Yes
B. No

Answer: A

NEW QUESTION 170
- (Exam Topic 2)
You need to configure the Edit Metadata module so that the structure of the datasets match.
Which configuration options should you select? To answer, select the appropriate options in the answer area.
NOTE: Each correct selection is worth one point.



A. Mastered
B. Not Mastered

Answer: A

Explanation:
Box 1: Floating point
Need floating point for Median values.
Scenario: An initial investigation shows that the datasets are identical in structure apart from the MedianValue column. The smaller Paris dataset contains the MedianValue in text format, whereas the larger London dataset contains the MedianValue in numerical format.
Box 2: Unchanged
Note: Select the Categorical option to specify that the values in the selected columns should be treated as categories.
For example, you might have a column that contains the numbers 0,1 and 2, but know that the numbers actually mean "Smoker", "Non smoker" and "Unknown". In that case, by flagging the column as categorical you can ensure that the values are not used in numeric calculations, only to group data.

NEW QUESTION 172
- (Exam Topic 2)
You need to select a feature extraction method. Which method should you use?

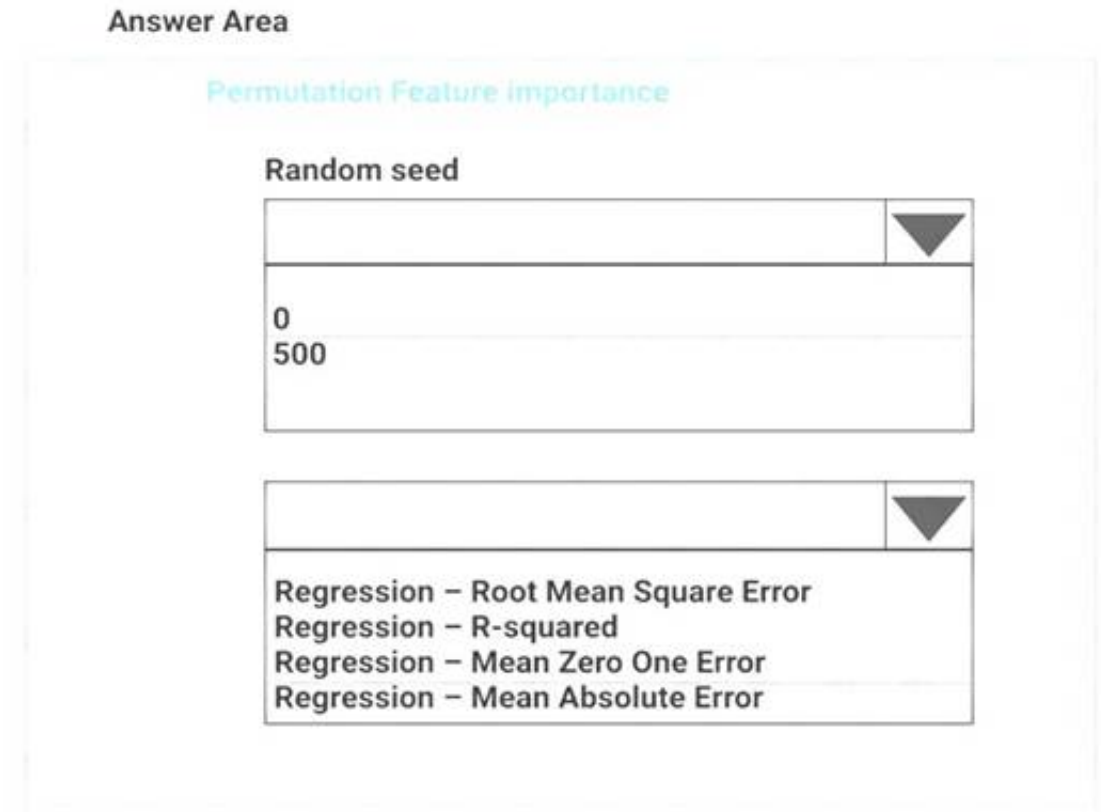
A. Mutual information
B. Mood's median test
C. Kendall correlation
D. Permutation Feature Importance

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 configure the Permutation Feature Importance module for the model training requirements. What should you do? To answer, select the appropriate options in the dialog box in the answer area. NOTE: Each correct selection is worth one point.

Answer Area



A. Mastered
B. Not Mastered

Answer: A

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

https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/permutation-feature-importan

NEW QUESTION 179
- (Exam Topic 1)
You need to select an environment that will meet the business and data requirements. Which environment should you use?

A. Azure HDInsight with Spark MLlib
B. Azure Cognitive Services
C. Azure Machine Learning Studio
D. Microsoft Machine Learning Server

Answer: D

NEW QUESTION 183
.....

THANKS FOR TRYING THE DEMO OF OUR PRODUCT

Visit Our Site to Purchase the Full Set of Actual DP-100 Exam Questions With Answers.

We Also Provide Practice Exam Software That Simulates Real Exam Environment And Has Many Self-Assessment Features. Order the DP-100 Product From:

https://www.2passeasy.com/dumps/DP-100/

Money Back Guarantee

DP-100 Practice Exam Features:

* DP-100 Questions and Answers Updated Frequently

* DP-100 Practice Questions Verified by Expert Senior Certified Staff

* DP-100 Most Realistic Questions that Guarantee you a Pass on Your FirstTry

* DP-100 Practice Test Questions in Multiple Choice Formats and Updatesfor 1 Year