Microsoft

Exam Questions DP-100

Designing and Implementing a Data Science Solution on Azure

About Exambible

Your Partner of IT Exam

Found in 1998

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

Our Advances

* 99.9% Uptime

All examinations will be up to date.

* 24/7 Quality Support

We will provide service round the clock.

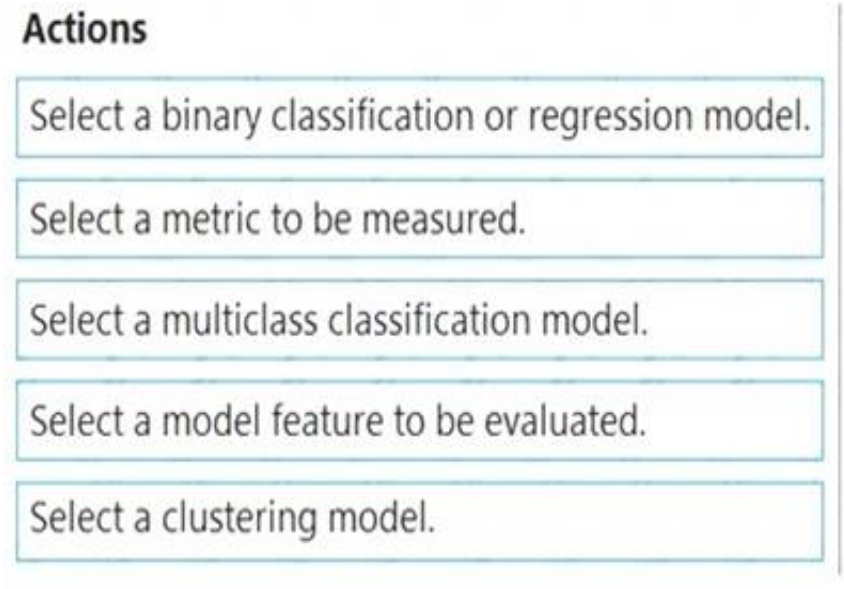
* 100% Pass Rate

Our guarantee that you will pass the exam.

* Unique Gurantee

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

NEW QUESTION 1
- (Exam Topic 3)
You have several machine learning models registered in an Azure Machine Learning workspace. You must use the Fairlearn dashboard to assess fairness in a selected model.
Which three actions should you perform in sequence? To answer, move the appropriate actions from the list of actions to the answer area and arrange them in the correct order.



A. Mastered
B. Not Mastered

Answer: A

Explanation:
Graphical user interface, text, application Description automatically generated
Step 1: Select a model feature to be evaluated.
Step 2: Select a binary classification or regression model.
Register your models within Azure Machine Learning. For convenience, store the results in a dictionary,
which maps the id of the registered model (a string in name:version format) to the predictor itself. Example:
model_dict = {}
lr_reg_id = register_model("fairness_logistic_regression", lr_predictor) model_dict[lr_reg_id] = lr_predictor
svm_reg_id = register_model("fairness_svm", svm_predictor) model_dict[svm_reg_id] = svm_predictor
Step 3: Select a metric to be measured Precompute fairness metrics.
Create a dashboard dictionary using Fairlearn's metrics package. Reference:
https://docs.microsoft.com/en-us/azure/machine-learning/how-to-machine-learning-fairness-aml

NEW QUESTION 2
- (Exam Topic 3)
HOTSPOT
You create a script for training a machine learning model in Azure Machine Learning service. You create an estimator by running the following code:

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

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

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

A. Mastered
B. Not Mastered

Answer: A

Explanation:
Box 1: Yes
Parameter source_directory is a local directory containing experiment configuration and code files needed for a training job.
Box 2: Yes
script_params is a dictionary of command-line arguments to pass to the training script specified in entry_script.
Box 3: No
Box 4: Yes
The conda_packages parameter is a list of strings representing conda packages to be added to the Python environment for the experiment.

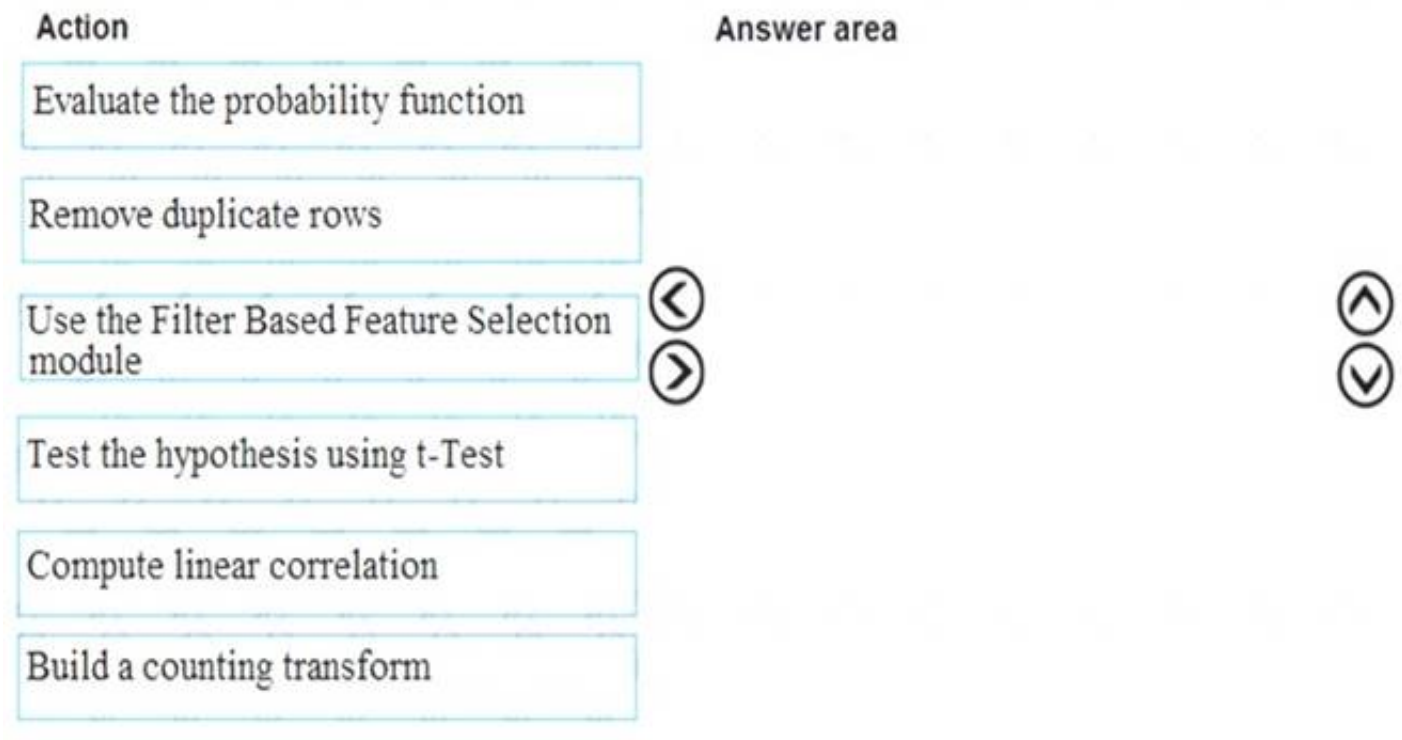
NEW QUESTION 3
- (Exam Topic 3)
You train a machine learning model.
You must deploy the model as a real-time inference service for testing. The service requires low CPU utilization and less than 48 MB of RAM. The compute target for the deployed service must initialize automatically while minimizing cost and administrative overhead.
Which compute target should you use?

A. Azure Kubernetes Service (AKS) inference cluster
B. Azure Machine Learning compute cluster
C. Azure Container Instance (ACI)
D. attached Azure Databricks cluster

Answer: C

Explanation:
Azure Container Instances (ACI) are suitable only for small models less than 1 GB in size. Use it for low-scale CPU-based workloads that require less than 48 GB of RAM.
Note: Microsoft recommends using single-node Azure Kubernetes Service (AKS) clusters for dev-test of larger models.
Reference:
https://docs.microsoft.com/id-id/azure/machine-learning/how-to-deploy-and-where

NEW QUESTION 4
- (Exam Topic 3)
You are producing a multiple linear regression model in Azure Machine Learning Studio. Several independent variables are highly correlated.
You need to select appropriate methods for conducting effective feature engineering on all the data.
Which three actions should you perform in sequence? To answer, move the appropriate actions from the list of actions to the answer area and arrange them in the correct order.



A. Mastered
B. Not Mastered

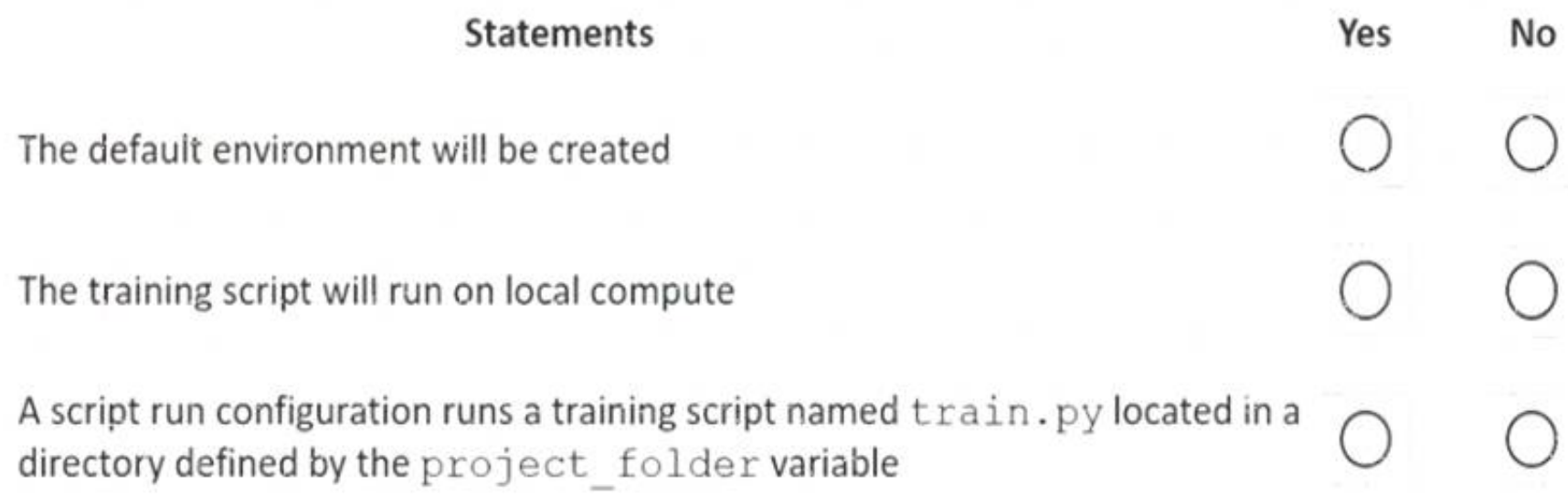
Answer: A

Explanation:
Step 1: Use the Filter Based Feature Selection module
Filter Based Feature Selection identifies the features in a dataset with the greatest predictive power.
The module outputs a dataset that contains the best feature columns, as ranked by predictive power. It also outputs the names of the features and their scores from the selected metric.
Step 2: Build a counting transform
A counting transform creates a transformation that turns count tables into features, so that you can apply the transformation to multiple datasets.
Step 3: Test the hypothesis using t-Test References:
https://docs.microsoft.com/bs-latn-ba/azure/machine-learning/studio-module-reference/filter-based-feature-selec
https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/build-counting-transform

NEW QUESTION 5
- (Exam Topic 3)
You use an Azure Machine Learning workspace. You create the following Python code:

```
from azureml.core import ScriptRunConfig
src = ScriptRunConfig(source_directory=project_folder,
                      script='train.py'
                      environment=myenv)
```

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



A. Mastered
B. Not Mastered

Answer: A

Explanation:
Graphical user interface, text, application Description automatically generated
Box 1: No
Environment is a required parameter. The environment to use for the run. If no environment is specified, azureml.core.runconfig.DEFAULT_CPU_IMAGE will be used as the Docker image for the run.
The following example shows how to instantiate a new environment. from azureml.core import Environment
myenv = Environment(name="myenv") Box 2: Yes

Parameter compute_target: The compute target where training will happen. This can either be a ComputeTarget object, the name of an existing ComputeTarget, or the string "local". If no compute target is specified, your local machine will be used.
Box 3: Yes
Parameter source_directory. A local directory containing code files needed for a run. Parameter script. The file path relative to the source_directory of the script to be run. Reference:
https://docs.microsoft.com/en-us/python/api/azureml-core/azureml.core.scriptrunconfig https://docs.microsoft.com/en-us/python/api/azureml-core/azureml.core.environment.environment

NEW QUESTION 6
- (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 Stratified split for the 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 7
- (Exam Topic 3)
You create a training pipeline using the Azure Machine Learning designer. You upload a CSV file that contains the data from which you want to train your model.
You need to use the designer to create a pipeline that includes steps to perform the following tasks:

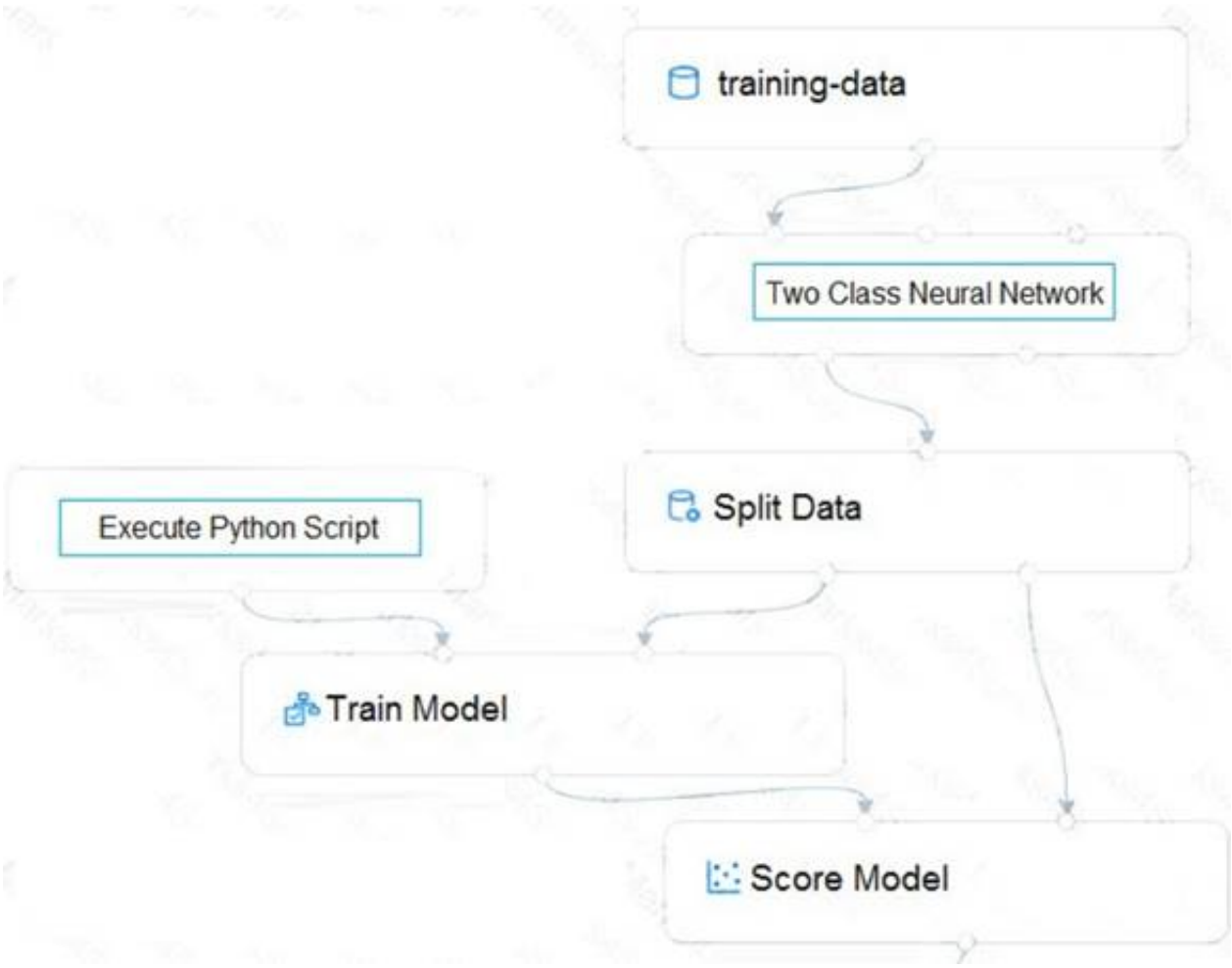
- Select the training features using the pandas filter method.
- Train a model based on the naive_bayes.GaussianNB algorithm.
- Return only the Scored Labels column by using the query SELECT [Scored Labels] FROM t1; Which modules should you use? To answer, drag the appropriate modules to the appropriate locations. Each module name may be used once, more than once, or not at all. You may need to drag the split bar between panes or scroll to view content.

NOTE: Each correct selection is worth one point.

A. Mastered
B. Not Mastered

Answer: A

Explanation:



NEW QUESTION 8
- (Exam Topic 3)
Your Azure Machine Learning workspace has a dataset named real_estate_data. A sample of the data in the dataset follows.

postal_code	num_bedrooms	sq_feet	garage	price
12345	3	1300	0	23,9000
54321	1	950	0	11,0000
12346	2	1200	1	15,0000

You want to use automated machine learning to find the best regression model for predicting the price column. You need to configure an automated machine learning experiment using the Azure Machine Learning SDK. How should you complete the code? To answer, select the appropriate options in the answer area.
NOTE: Each correct selection is worth one point.

Answer Area

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

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

▼ =split1_ds,
- X
- Y
- X_valid
- Y_valid
- training_data

▼ =split2_ds
- X
- Y
- X_valid
- Y_valid
- validation_data
- training_data

▼ ='price')
- y
- y_valid
- y_max
- label_column_name
- exclude_nan_labels

A. Mastered
B. Not Mastered

Answer: A

Explanation:
Box 1: training_data
The training data to be used within the experiment. It should contain both training features and a label column (optionally a sample weights column). If training_data is specified, then the label_column_name parameter must also be specified.
Box 2: validation_data
Provide validation data: In this case, you can either start with a single data file and split it into training and validation sets or you can provide a separate data file for the validation set. Either way, the validation_data parameter in your AutoMLConfig object assigns which data to use as your validation set.
Example, the following code example explicitly defines which portion of the provided data in dataset to use for training and validation.
dataset = Dataset.Tabular.from_delimited_files(data)
training_data, validation_data = dataset.random_split(percentage=0.8, seed=1) automl_config = AutoMLConfig(compute_target = aml_remote_compute, task = 'classification',
primary_metric = 'AUC_weighted', training_data = training_data,
validation_data = validation_data, label_column_name = 'Class'
)
Box 3: label_column_name label_column_name:
The name of the label column. If the input data is from a pandas.DataFrame which doesn't have column names, column indices can be used instead, expressed as integers.
This parameter is applicable to training_data and validation_data parameters. Reference:
https://docs.microsoft.com/en-us/python/api/azureml-train-automl-client/azureml.train.automl.automlconfig.auto

NEW QUESTION 9
- (Exam Topic 3)
You train a model and register it in your Azure Machine Learning workspace. You are ready to deploy the model as a real-time web service.
You deploy the model to an Azure Kubernetes Service (AKS) inference cluster, but the deployment fails because an error occurs when the service runs the entry script that is associated with the model deployment.
You need to debug the error by iteratively modifying the code and reloading the service, without requiring a re-deployment of the service for each code update.
What should you do?

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

Answer: C

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

NEW QUESTION 10
- (Exam Topic 3)
You are preparing to build a deep learning convolutional neural network model for image classification. You create a script to train the model using CUDA devices.
You must submit an experiment that runs this script in the Azure Machine Learning workspace. The following compute resources are available:

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

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

Resource type	Option
Run code to submit the experiment	the Microsoft Surface device the ds-workstation notebook VM the cpu-cluster compute target the gpu-cluster compute target
Run the training script	the ds-workstation notebook VM the cpu-compute target the gpu-compute target the Microsoft Surface device

A. Mastered
B. Not Mastered

Answer: A

Explanation:

Resource type	Option
Run code to submit the experiment	the Microsoft Surface device the ds-workstation notebook VM the cpu-cluster compute target the gpu-cluster compute target
Run the training script	the ds-workstation notebook VM the cpu-compute target the gpu-compute target the Microsoft Surface device

NEW QUESTION 10
- (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.dnn import TensorFlow
sk_est = TensorFlow(source_directory='./scripts',
 compute_target=aml-compute,
 entry_script='train.py')
```

Does the solution meet the goal?

A. Yes
B. No

Answer: B

Explanation:

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

NEW QUESTION 15
- (Exam Topic 3)
Note: This question is part of a series of questions that present the same scenario. Each question in the series contains a unique solution that might meet the stated goals. Some question sets might have more than one correct solution, while others might not have a correct solution.
After you answer a question in this section, you will NOT be able to return to it. As a result, these questions will not appear in the review screen.
You plan to use a Python script to run an Azure Machine Learning experiment. The script creates a reference to the experiment run context, loads data from a file, identifies the set of unique values for the label column, and completes the experiment run:

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

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

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

A. Yes
B. No

Answer: A

Explanation:
run.log_list log a list of values to the run with the given name using log_list. Example: run.log_list("accuracies", [0.6, 0.7, 0.87])
Note:
Data= pd.read_csv('data.csv')
Data is read into a pandas.DataFrame, which is a two-dimensional, size-mutable, potentially heterogeneous tabular data.
label_vals =data['label'].unique
label_vals contains a list of unique label values. Reference:
https://www.element61.be/en/resource/azure-machine-learning-services-complete-toolbox-ai https://docs.microsoft.com/en-us/python/api/azureml-core/azureml.core.run(class) https://pandas.pydata.org/docs/reference/api/pandas.DataFrame.html

NEW QUESTION 17
- (Exam Topic 3)
You use the Two-Class Neural Network module in Azure Machine Learning Studio to build a binary classification model. You use the Tune Model Hyperparameters module to tune accuracy for the model.
You need to select the hyperparameters that should be tuned using the Tune Model Hyperparameters module. Which two hyperparameters should you use? Each correct answer presents part of the solution.
NOTE: Each correct selection is worth one point.

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

Answer: DE

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

NEW QUESTION 22
- (Exam Topic 3)
You are running a training experiment on remote compute in Azure Machine Learning.
The experiment is configured to use a conda environment that includes the mlflow and azureml-contrib-run packages.
You must use MLflow as the logging package for tracking metrics generated in the experiment. You need to complete the script for the experiment.

How should you complete the code? To answer, select the appropriate options in the answer area.
NOTE: Each correct selection is worth one point.

```
import numpy as np
# Import library to log metrics
[ from azureml.core import Run | import mlflow | import logging ]
# Start logging for this run
[ run = Run.get_context() | mlflow.start_run() | logger = logging.getLogger('Run') ]
reg_rate = 0.01
# Log the reg_rate metric
[ run.log('reg_rate', np.float(reg_rate)) | mlflow.log_metric('reg_rate', np.float(reg_rate)) | logger.info(np.float(reg_rate)) ]
# Stop logging for this run
[ run.complete() | mlflow.end_run() | logger.setLevel(logging.INFO) ]
```

A. Mastered
B. Not Mastered

Answer: A

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

NEW QUESTION 26
- (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 27
- (Exam Topic 3)
You are creating a binary classification by using a two-class logistic regression model. You need to evaluate the model results for imbalance.
Which evaluation metric should you use?

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

Answer: B

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

NEW QUESTION 28
- (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 32
- (Exam Topic 3)
You use the Azure Machine Learning service to create a tabular dataset named training.data. You plan to use this dataset in a training script.
You create a variable that references the dataset using the following code: training_ds = workspace.datasets.get("training_data")
You define an estimator to run the script.
You need to set the correct property of the estimator to ensure that your script can access the training.data dataset
Which property should you set?
A)

```
inputs = [training_ds.as_named_input('training_ds')]
```

B)

```
script_params = {"--training_ds":training_ds}
```

C)

```
environment_definition = {"training_data":training_ds}
```

D)

```
source_directory = training_ds
```

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

Answer: A

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

NEW QUESTION 36
- (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 40
- (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 41
- (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 44
- (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 45
- (Exam Topic 3)

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

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

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

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

Answer: B

Explanation:
Log a numerical or string value to the run with the given name using log(name, value, description=''). Logging a metric to a run causes that metric to be stored in the run record in the experiment. You can log the same metric multiple times within a run, the result being considered a vector of that metric.
Example: run.log("accuracy", 0.95) Reference:
https://docs.microsoft.com/en-us/python/api/azureml-core/azureml.core.run

NEW QUESTION 48
- (Exam Topic 3)
A coworker registers a datastore in a Machine Learning services workspace by using the following code:

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

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

```
import azureml.core
from azureml.core import Workspace, Datastore
ws = Workspace.from_config()
datastore = [Workspace / Datastore / Experiment / Run].get( [ws / run / experiment / log] , ' [demo_datastore / demo_datacontainer / demo_account / Datastore]
```

A. Mastered
B. Not Mastered

Answer: A

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

NEW QUESTION 50
- (Exam Topic 3)
You are the owner of an Azure Machine Learning workspace.
You must prevent the creation or deletion of compute resources by using a custom role. You must allow all other operations inside the workspace.
You need to configure the custom role.
How should you complete the configuration? To answer, select the appropriate options in the answer area.
NOTE: Each correct selection is worth one point.

Answer Area

```
{
 "Name": "Data Scientist Custom",
 "IsCustom": true
 "Description": "Description"
 "Actions": [
```

[Microsoft.MachineLearningServices/workspaces/*/read | Microsoft.MachineLearningServices/workspaces/computes/*/write | Microsoft.MachineLearningServices/workspaces/delete]

[Microsoft.MachineLearningServices/workspaces/*/write | Microsoft.MachineLearningServices/workspaces/computes/*/write | Microsoft.MachineLearningServices/workspaces/delete]

```
 ],
 "NotActions": [
```

[Microsoft.MachineLearningServices/workspaces/*/read | Microsoft.MachineLearningServices/workspaces/*/write | Microsoft.MachineLearningServices/workspaces/computes/*/delete]

[Microsoft.MachineLearningServices/workspaces/*/read | Microsoft.MachineLearningServices/workspaces/*/write | Microsoft.MachineLearningServices/workspaces/computes/*/write]

```
 ],
 "AssignableScopes": [
   "/subscriptions/<subscription_id>"
 ]
}
```

A. Mastered
B. Not Mastered

Answer: A

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

NEW QUESTION 52
- (Exam Topic 3)
You create and register a model in an Azure Machine Learning workspace.
You must use the Azure Machine Learning SDK to implement a batch inference pipeline that uses a ParallelRunStep to score input data using the model. You must specify a value for the ParallelRunConfig compute_target setting of the pipeline step.
You need to create the compute target. Which class should you use?

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

Answer: C

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

NEW QUESTION 54

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

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

Does the solution meet the goal?

A. Yes
B. No

Answer: B

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

NEW QUESTION 56
- (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)
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 61
- (Exam Topic 3)
You have a dataset created for multiclass classification tasks that contains a normalized numerical feature set with 10,000 data points and 150 features.
You use 75 percent of the data points for training and 25 percent for testing. You are using the scikit-learn machine learning library in Python. You use X to denote the feature set and Y to denote class labels.
You create the following Python data frames:

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

You need to apply the Principal Component Analysis (PCA) method to reduce the dimensionality of the feature set to 10 features in both training and testing sets.
How should you complete the code segment? To answer, select the appropriate options in the answer area. NOTE: Each correct selection is worth one point.

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

```
X_train=
```
.fit_transform(X_train)

pca
model
sklearn.decomposition

```
x_test = pca.
```

x_test
X_train
fit(x_test)
transform(x_test)

A. Mastered
B. Not Mastered

Answer: A

Explanation:
Box 1: PCA(n_components = 10)
Need to reduce the dimensionality of the feature set to 10 features in both training and testing sets. Example:
from sklearn.decomposition import PCA pca = PCA(n_components=2) ;2 dimensions principalComponents = pca.fit_transform(x)
Box 2: pca
fit_transform(X[, y])fits the model with X and apply the dimensionality reduction on X. Box 3: transform(x_test)
transform(X) applies dimensionality reduction to X. References:
https://scikit-learn.org/stable/modules/generated/sklearn.decomposition.PCA.html

NEW QUESTION 63
- (Exam Topic 3)
You are implementing a machine learning model to predict stock prices. The model uses a PostgreSQL database and requires GPU processing.
You need to create a virtual machine that is pre-configured with the required tools. What should you do?

A. Create a Data Science Virtual Machine (DSVM) Windows edition.
B. Create a Geo Al Data Science Virtual Machine (Geo-DSVM) Windows edition.
C. Create a Deep Learning Virtual Machine (DLVM) Linux edition.
D. Create a Deep Learning Virtual Machine (DLVM) Windows edition.
E. Create a Data Science Virtual Machine (DSVM) Linux edition.

Answer: E

NEW QUESTION 67
- (Exam Topic 3)
You create a machine learning model by using the Azure Machine Learning designer. You publish the model as a real-time service on an Azure Kubernetes Service (AKS) inference compute cluster. You make no changes to the deployed endpoint configuration.
You need to provide application developers with the information they need to consume the endpoint.
Which two values should you provide to application developers? Each correct answer presents part of the solution.
NOTE: Each correct selection is worth one point.

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

Answer: CE

Explanation:
Deploying an Azure Machine Learning model as a web service creates a REST API endpoint. You can send data to this endpoint and receive the prediction returned by the model.
You create a web service when you deploy a model to your local environment, Azure Container Instances, Azure Kubernetes Service, or field-programmable gate arrays (FPGA). You retrieve the URI used to access the web service by using the Azure Machine Learning SDK. If authentication is enabled, you can also use the SDK to get the authentication keys or tokens.
Example:
# URL for the web service
scoring_uri = '<your web service URI>'
# If the service is authenticated, set the key or token key = '<your key or token>'
Reference:
https://docs.microsoft.com/en-us/azure/machine-learning/how-to-consume-web-service

NEW QUESTION 68
- (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?

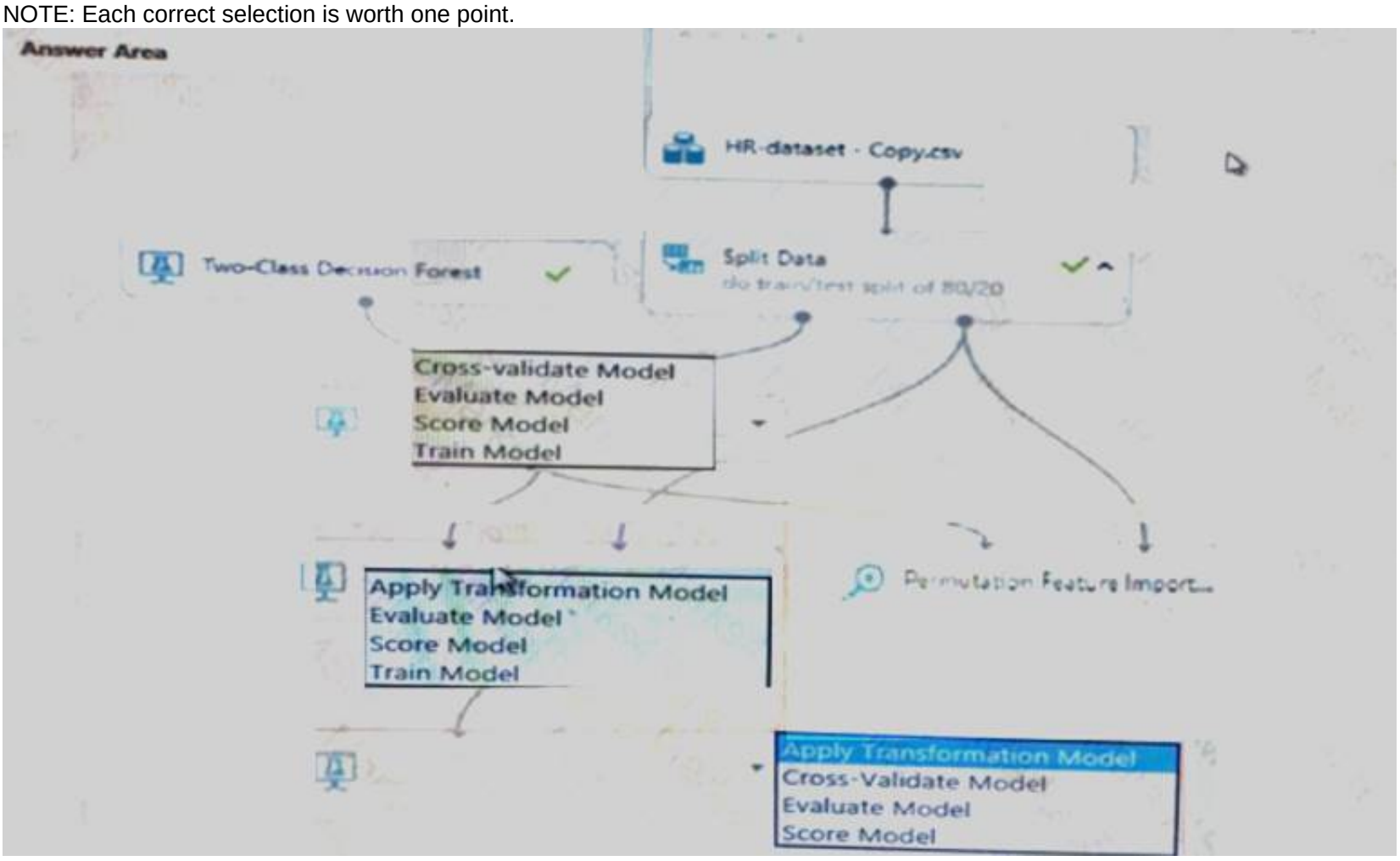
A. Yes
B. No

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

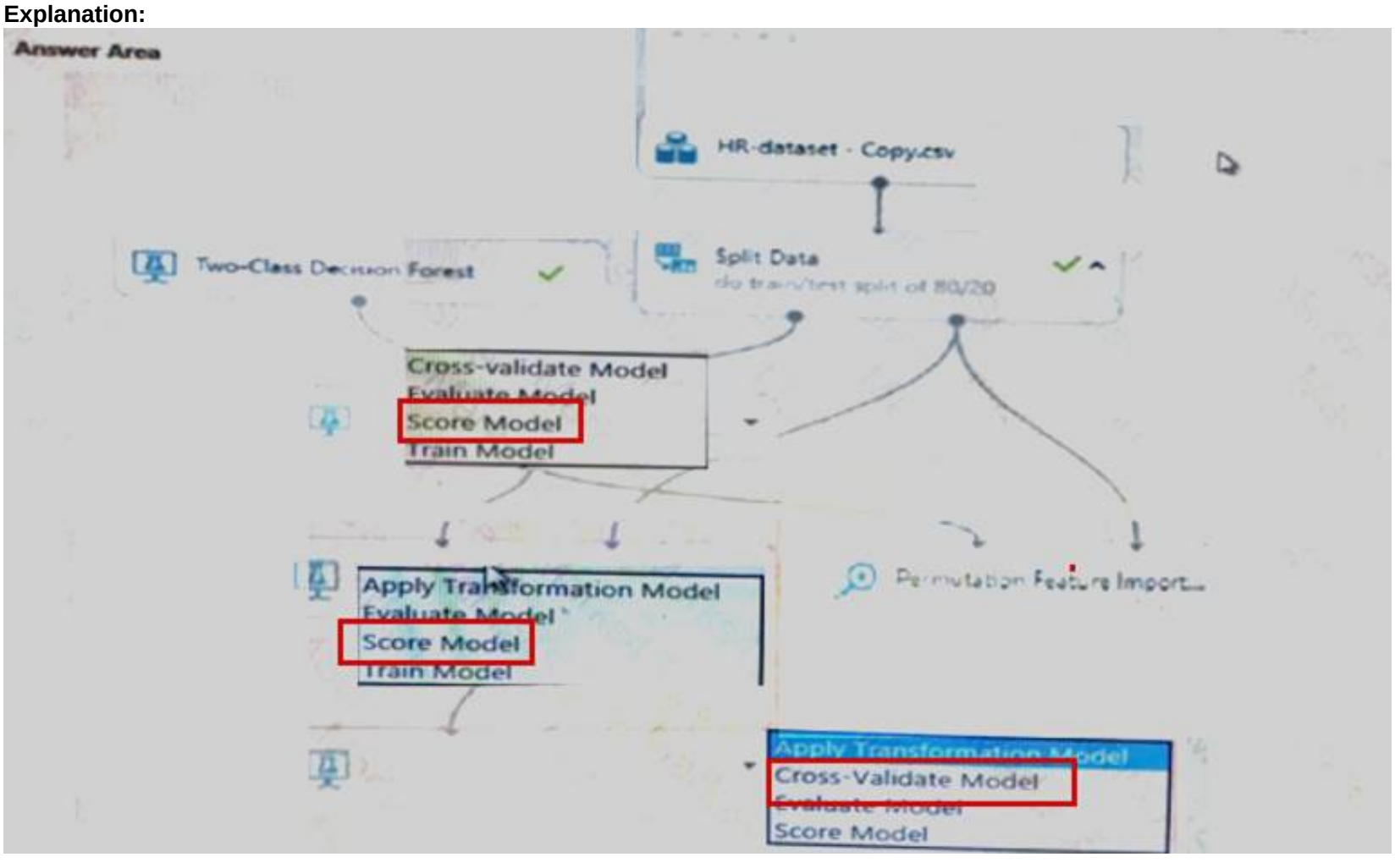
NEW QUESTION 71
- (Exam Topic 3)
You create a binary classification model using Azure Machine Learning Studio.
You must use a Receiver Operating Characteristic (RO C) curve and an F1 score to evaluate the model. You need to create the required business metrics.
How should you complete the experiment? To answer, select the appropriate options in the dialog box in the answer area.
NOTE: Each correct selection is worth one point.



A. Mastered
B. Not Mastered

Answer: A

Explanation:



NEW QUESTION 73
- (Exam Topic 3)
You deploy a model as an Azure Machine Learning real-time web service using the following code.

```
# ws, model, interference_config, and deployment_config defined previously
service = Model.deploy(ws, 'classification-service', [model], inference_config, deployment_config)
service.wait_for_deployment(True)
```

The deployment fails.
You need to troubleshoot the deployment failure by determining the actions that were performed during deployment and identifying the specific action that failed.
Which code segment should you run?

A. service.get_logs()
B. service.state
C. service.serialize()
D. service.update_deployment_state()

Answer: A

Explanation:
You can print out detailed Docker engine log messages from the service object. You can view the log for ACI, AKS, and Local deployments. The following example demonstrates how to print the logs.
if you already have the service object handy print(service.get_logs())
if you only know the name of the service (note there might be multiple services with the same name but different version number)
print(ws.webservices['mysvc'].get_logs()) Reference:
https://docs.microsoft.com/en-us/azure/machine-learning/how-to-troubleshoot-deployment

NEW QUESTION 77
- (Exam Topic 3)
You are planning to register a trained model in an Azure Machine Learning workspace.
You must store additional metadata about the model in a key-value format. You must be able to add new metadata and modify or delete metadata after creation.
You need to register the model. Which parameter should you use?

A. description
B. model_framework
C. cags
D. properties

Answer: D

Explanation:
azureml.core.Model.properties:
Dictionary of key value properties for the Model. These properties cannot be changed after registration, however new key value pairs can be added.
Reference:
https://docs.microsoft.com/en-us/python/api/azureml-core/azureml.core.model.model

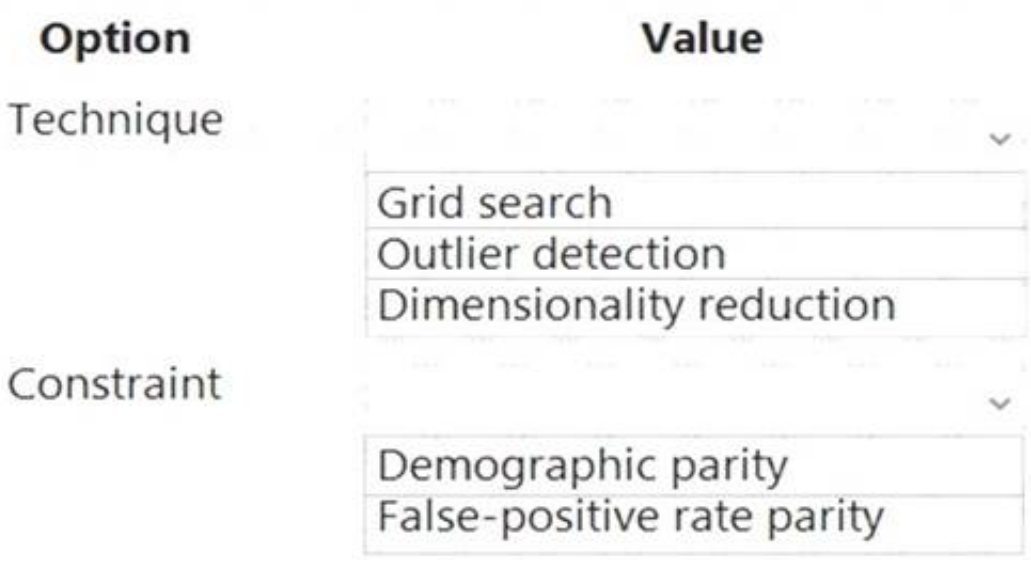
NEW QUESTION 79
- (Exam Topic 3)
You are analyzing a dataset containing historical data from a local taxi company. You arc developing a regression a regression model.
You must predict the fare of a taxi trip.
You need to select performance metrics to correctly evaluate the- regression model. Which two metrics can you use? Each correct answer presents a complete solution. NOTE: Each correct selection is worth one point.

A. an F1 score that is high
B. an R Squared value dose to 1
C. an R-Squared value close to 0
D. a Root Mean Square Error value that is high
E. a Root Mean Square Error value that is low
F. an F 1 score that is low.

Answer: BE

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

NEW QUESTION 80
- (Exam Topic 3)
A biomedical research company plans to enroll people in an experimental medical treatment trial.
You create and train a binary classification model to support selection and admission of patients to the trial. The model includes the following features: Age, Gender, and Ethnicity.
The model returns different performance metrics for people from different ethnic groups.
You need to use Fairlearn to mitigate and minimize disparities for each category in the Ethnicity feature. Which technique and constraint 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:
Graphical user interface, text, application, chat or text message Description automatically generated
Box 1: Grid Search
Fairlearn open-source package provides postprocessing and reduction unfairness mitigation algorithms: ExponentiatedGradient, GridSearch, and ThresholdOptimizer.
Note: The Fairlearn open-source package provides postprocessing and reduction unfairness mitigation algorithms types:

- Reduction: These algorithms take a standard black-box machine learning estimator (e.g., a LightGBM model) and generate a set of retrained models using a sequence of re-weighted training datasets.
- Post-processing: These algorithms take an existing classifier and the sensitive feature as input.

Box 2: Demographic parity
The Fairlearn open-source package supports the following types of parity constraints: Demographic parity, Equalized odds, Equal opportunity, and Bounded group loss.
Reference:
https://docs.microsoft.com/en-us/azure/machine-learning/concept-fairness-ml

NEW QUESTION 84
- (Exam Topic 3)
You are performing sentiment analysis using a CSV file that includes 12,000 customer reviews written in a short sentence format. You add the CSV file to Azure Machine Learning Studio and configure it as the starting point dataset of an experiment. You add the Extract N-Gram Features from Text module to the experiment to extract key phrases from the customer review column in the dataset.
You must create a new n-gram dictionary from the customer review text and set the maximum n-gram size to trigrams.
What should you select? To answer, select the appropriate options in the answer area.
NOTE: Each correct selection is worth one point.

A. Mastered
B. Not Mastered

Answer: A

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

NEW QUESTION 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 create an Azure Machine Learning service datastore in a workspace. The datastore contains the following files:
• /data/2018/Q1.csv
• /data/2018/Q2.csv
• /data/2018/Q3.csv
• /data/2018/Q4.csv
• /data/2019/Q1.csv

All files store data in the following format: id,f1,f2i
1,1.2,0
2,1,1,
1 3,2.1,0
You run the following code:

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

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

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

Solution: Run the following code:

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

Does the solution meet the goal?

A. Yes
B. No

Answer: B

Explanation:
Use two file paths.
Use Dataset.Tabular_from_delimeted, instead of Dataset.File.from_files as the data isn't cleansed. Reference:
https://docs.microsoft.com/en-us/azure/machine-learning/how-to-create-register-datasets

NEW QUESTION 91
- (Exam Topic 3)
You are analyzing a raw dataset that requires cleaning.
You must perform transformations and manipulations by using Azure Machine Learning Studio. You need to identify the correct modules to perform the transformations.
Which modules should you choose? To answer, drag the appropriate modules to the correct scenarios. Each module 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.

Answer Area

Methods	Scenario	Module
Clean Missing Data	Replace missing values by removing rows and columns.	
SMOTE	Increase the number of low-incidence examples in the dataset.	
Convert to Indicator Values	Convert a categorical feature into a binary indicator.	
Remove Duplicate Rows	Remove potential duplicates from a dataset.	
Threshold Filter		

A. Mastered
B. Not Mastered

Answer: A

Explanation:
Box 1: Clean Missing Data Box 2: SMOTE
Use the SMOTE module in Azure Machine Learning Studio 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.
Box 3: Convert to Indicator Values
Use the Convert to Indicator Values module in Azure Machine Learning Studio. The purpose of this module is to convert columns that contain categorical values into a series of binary indicator columns that can more easily be used as features in a machine learning model.
Box 4: Remove Duplicate Rows References:
https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/smote https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/convert-to-indicator-values

NEW QUESTION 95
- (Exam Topic 3)
You are 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 Oversampling Technique (SMOTE)
B. Replace using MICE
C. Replace using; Probabilistic PCA
D. Normalization

Answer: C

Explanation:
Replace using Probabilistic PCA: Compared to other options, such as Multiple Imputation using Chained Equations (MICE), this option has the advantage of not requiring the application of predictors for each column. Instead, it approximates the covariance for the full dataset. Therefore, it might offer better performance for datasets that have missing values in many columns.
References:
https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/clean-missing-data

NEW QUESTION 99
- (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 104
- (Exam Topic 3)
Note: This question is part of a series of questions that present the same scenario. Each question in the series contains a unique solution that might meet the stated goals. Some question sets might have more than one correct solution, while others might not have a correct solution.
After you answer a question in this section, you will NOT be able to return to it. As a result, these questions will not appear in the review screen.
You train and register a machine learning model.
You plan to deploy the model as a real-time web service. Applications must use key-based authentication to use the model.
You need to deploy the web service.
Solution:
Create an AksWebservice instance.
Set the value of the auth_enabled property to False.
Set the value of the token_auth_enabled property to True.
Deploy the model to the service. Does the solution meet the goal?

A. Yes
B. No

Answer: B

Explanation:
Instead use only auth_enabled = TRUE Note: Key-based authentication.
Web services deployed on AKS have key-based auth enabled by default. ACI-deployed services have
key-based auth disabled by default, but you can enable it by setting auth_enabled = TRUE when creating the ACI web service. The following is an example of creating an ACI deployment configuration with key-based auth enabled.
deployment_config <- aci_webservice_deployment_config(cpu_cores = 1, memory_gb = 1,
auth_enabled = TRUE) Reference:
https://azure.github.io/azureml-sdk-for-r/articles/deploying-models.html

NEW QUESTION 109
- (Exam Topic 3)
You plan to use a Data Science Virtual Machine (DSVM) with the open source deep learning frameworks Caffe2 and Theano. You need to select a pre configured DSVM to support the framework.
What should you create?

A. Data Science Virtual Machine for Linux (CentOS)
B. Data Science Virtual Machine for Windows 2012
C. Data Science Virtual Machine for Windows 2016
D. Geo AI Data Science Virtual Machine with ArcGIS
E. Data Science Virtual Machine for Linux (Ubuntu)

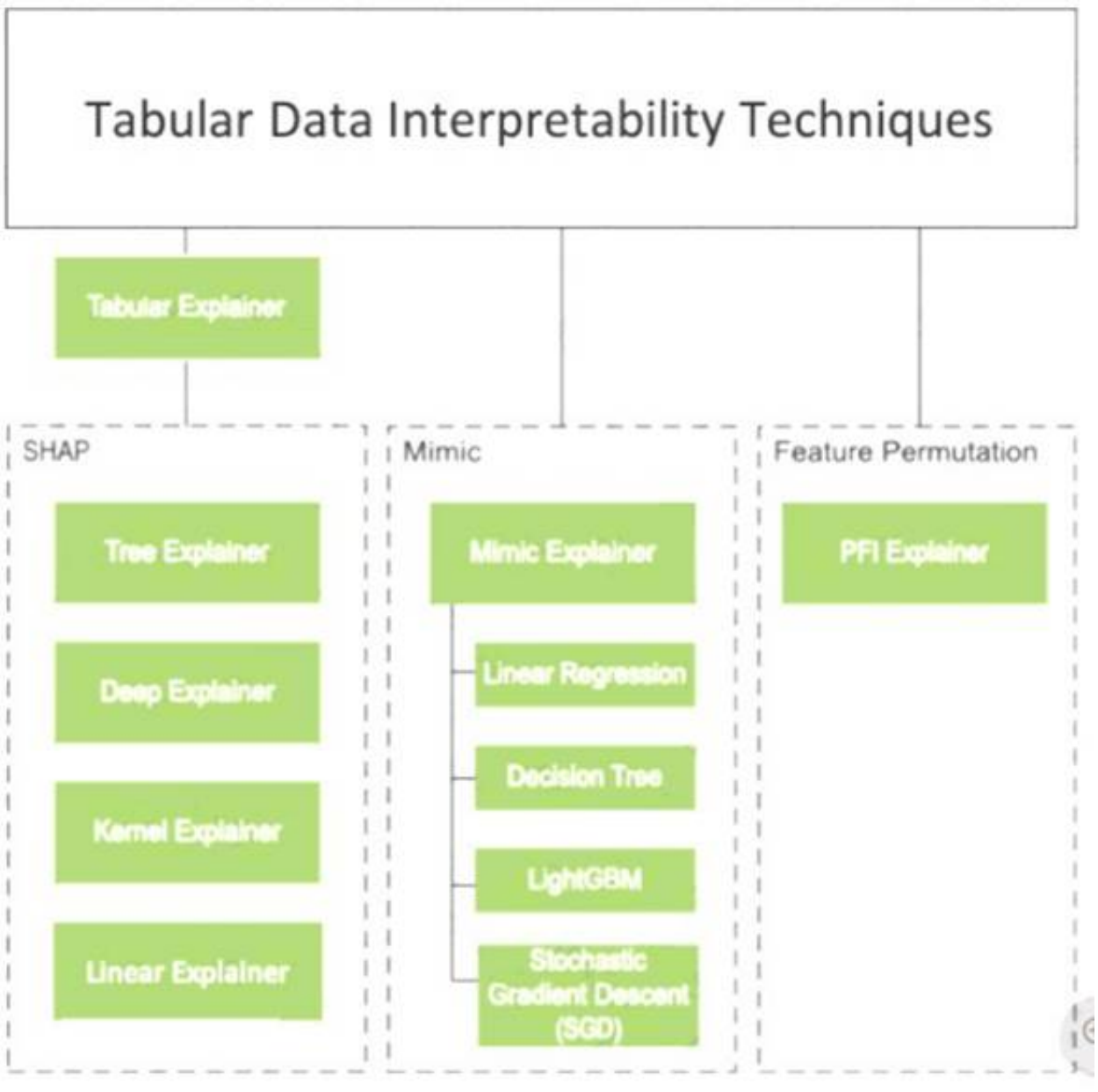
Answer: E

NEW QUESTION 111
- (Exam Topic 3)
Note: This question is part of a series of questions that present the same scenario. Each question in the series contains a unique solution that might meet the stated goals. Some question sets might have more
than one correct solution, while others might not have a correct solution.
After you answer a question in this section, you will NOT be able to return to it. As a result, these questions will not appear in the review screen.
You train a classification model by using a logistic regression algorithm.
You must be able to explain the model's predictions by calculating the importance of each feature, both as an overall global relative importance value and as a measure of local importance for a specific set of predictions.
You need to create an explainer that you can use to retrieve the required global and local feature importance values.
Solution: Create a TabularExplainer. Does the solution meet the goal?

A. Yes
B. No

Answer: B

Explanation:
Instead use Permutation Feature Importance Explainer (PFI). Note 1:



Note 2: Permutation Feature Importance Explainer (PFI): Permutation Feature Importance is a technique used to explain classification and regression models. At a high level, the way it works is by randomly shuffling data one feature at a time for the entire dataset and calculating how much the performance metric of interest changes. The larger the change, the more important that feature is. PFI can explain the overall behavior of any underlying model but does not explain individual

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

NEW QUESTION 113
- (Exam Topic 3)
You train a classification model by using a decision tree algorithm.
You create an estimator by running the following Python code. The variable feature_names is a list of all feature names, and class_names is a list of all class names.
from interpret.ext.blackbox import TabularExplainer

```
explainer = TabularExplainer(model,
                             x_train,
                             features=feature_names,
                             classes=classe_names)
```

You need to explain the predictions made by the model for all classes by determining the importance of all features.
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 SHAP TreeExplainer will be used to interpret the model.	○	○
If you omit the features and classes parameters in the TabularExplainer instantiation, the explainer still works as expected.	○	○
You could interpret the model by using a MimicExplainer instead of a TabularExplainer.	○	○

A. Mastered
B. Not Mastered

Answer: A

Explanation:
s and visualizations more informative, you can choose to pass in feature names and output class names if doing classification.
Box 3: No
TabularExplainer automatically selects the most appropriate one for your use case, but you can call each of its three underlying explainers underneath (TreeExplainer, DeepExplainer, or KernelExplainer) directly.
Reference:
https://docs.microsoft.com/en-us/azure/machine-learning/how-to-machine-learning-interpretability-aml

NEW QUESTION 115
- (Exam Topic 3)
You are performing feature engineering on a dataset.
You must add a feature named CityName and populate the column value with the text London.
You need to add the new feature to the dataset.
Which Azure Machine Learning Studio module should you use?

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

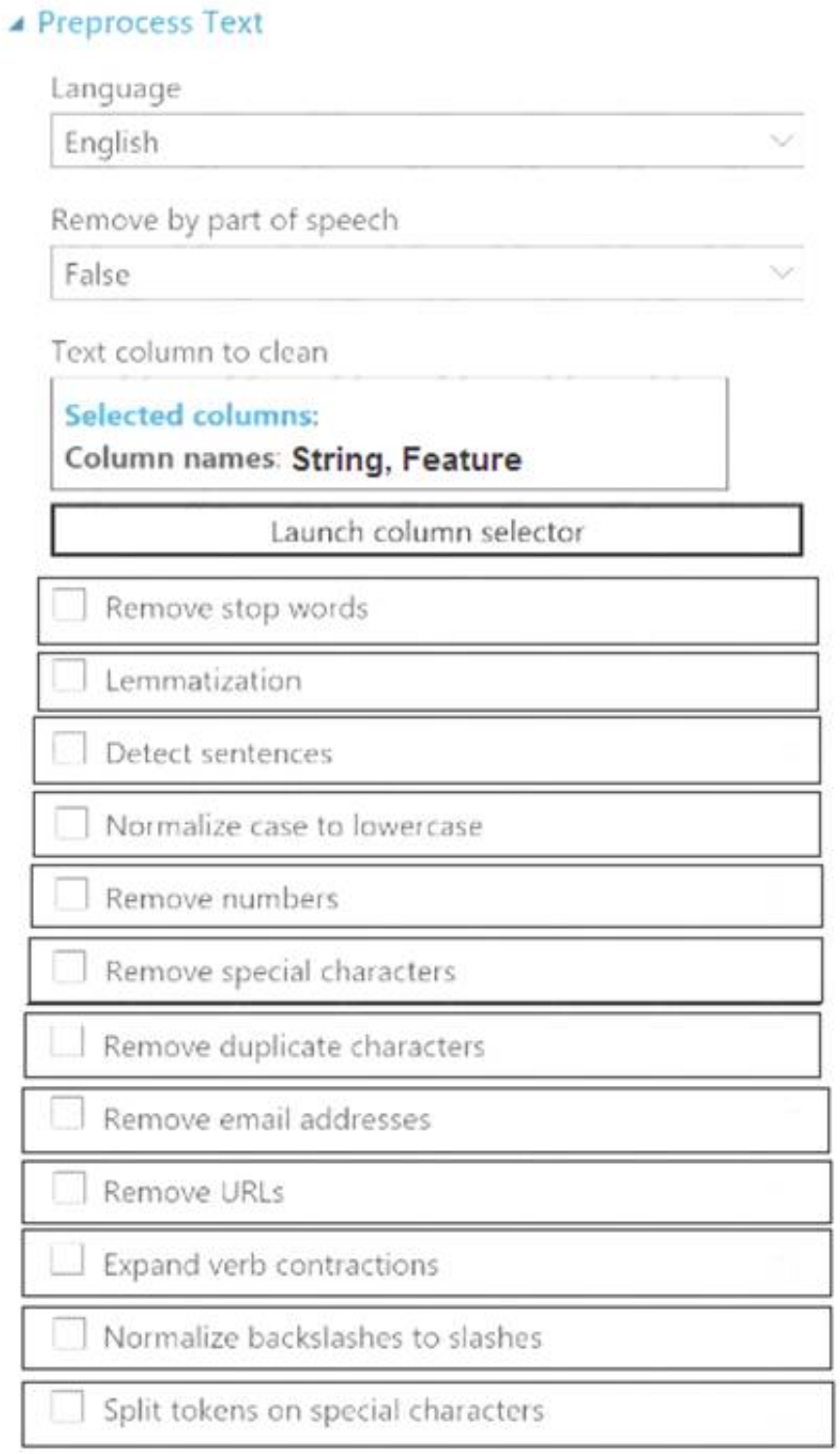
Answer: A

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

NEW QUESTION 120
- (Exam Topic 3)
You plan to preprocess text from CSV files. You load the Azure Machine Learning Studio default stop words list.
You need to configure the Preprocess Text module to meet the following requirements:

- Ensure that multiple related words from a single canonical form.
- Remove pipe characters from text.
- Remove words to optimize information retrieval.

Which three options should you select? To answer, select the appropriate options in the answer area. NOTE: Each correct selection is worth one point.



A. Mastered
B. Not Mastered

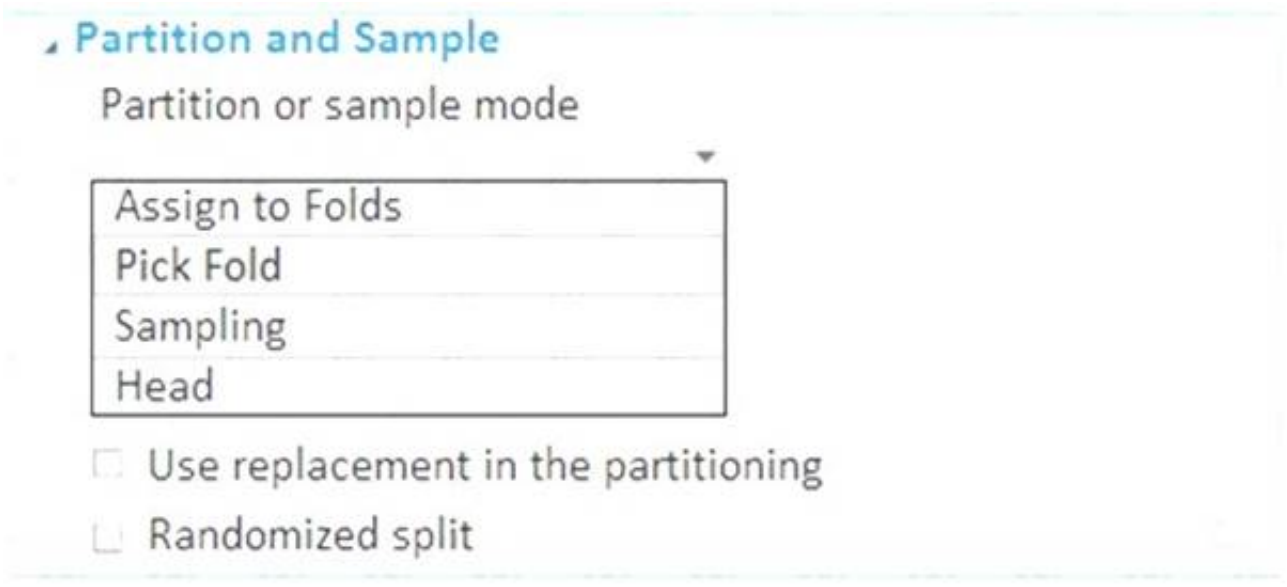
Answer: A

Explanation:
Box 1: Remove stop words
Remove words to optimize information retrieval.
Remove stop words: Select this option if you want to apply a predefined stopword list to the text column. Stop word removal is performed before any other processes.
Box 2: Lemmatization
Ensure that multiple related words from a single canonical form. Lemmatization converts multiple related words to a single canonical form Box 3: Remove special characters
Remove special characters: Use this option to replace any non-alphanumeric special characters with the pipe | character.
References:
https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/preprocess-text

NEW QUESTION 123
- (Exam Topic 3)
You have a dataset that contains 2,000 rows. You are building a machine learning classification model by using Azure Learning Studio. You add a Partition and Sample module to the experiment.
You need to configure the module. You must meet the following requirements:

- Divide the data into subsets
- Assign the rows into folds using a round-robin method
- Allow rows in the dataset to be reused

How should you configure the module? To answer, select the appropriate options in the dialog box in the answer area.
NOTE: Each correct selection is worth one point.



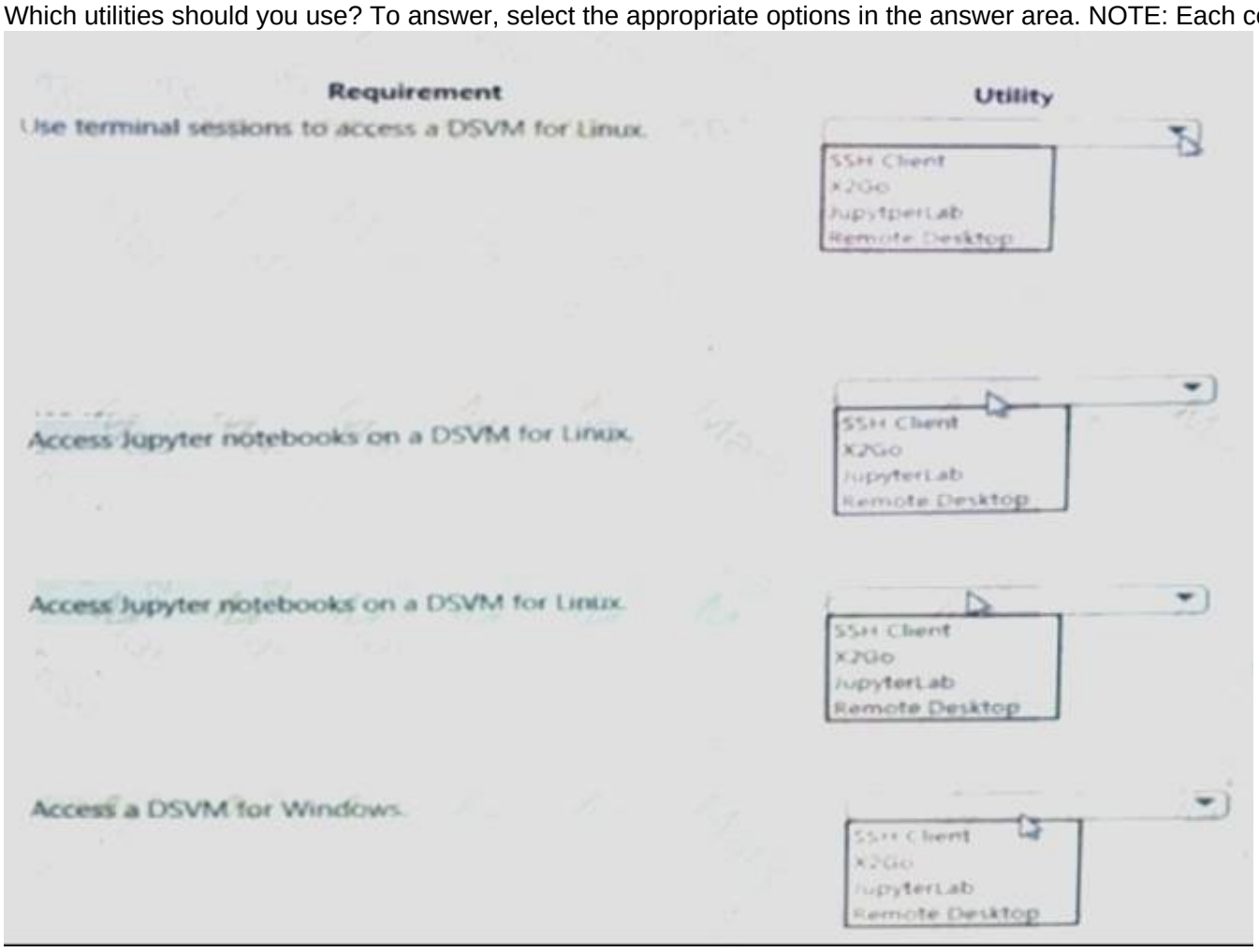
A. Mastered
B. Not Mastered

Answer: A

Explanation:
Use the Split data into partitions option when you want to divide the dataset into subsets of the data. This option is also useful when you want to create a custom number of folds for cross-validation, or to split rows into several groups.

- For Partition or sample mode, select Assign to Folds.
- Use replacement in the partitioning: Select this option if you want the sampled row to be put back into the pool of rows for potential reuse. As a result, the same row might be assigned to several folds.
- If you do not use replacement (the default option), the sampled row is not put back into the pool of rows for potential reuse. As a result, each row can be assigned to only one fold.
- Randomized split: Select this option if you want rows to be randomly assigned to folds. If you do not select this option, rows are assigned to folds using the round-robin method. References:
https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/partition-and-sample

NEW QUESTION 128
- (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 129
- (Exam Topic 3)
Note: This question is part of a series of questions that present the same scenario. Each question in the series contains a unique solution that might meet the stated goals. Some question sets might have more than one correct solution, while others might not have a correct solution.
After you answer a question in this section, you will NOT be able to return to it. As a result, these questions will not appear in the review screen.
You are using Azure Machine Learning Studio to perform feature engineering on a dataset. You need to normalize values to produce a feature column grouped into bins.
Solution: Apply an Entropy Minimum Description Length (MDL) binning mode. Does the solution meet the goal?

A. Yes
B. No

Answer: A

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

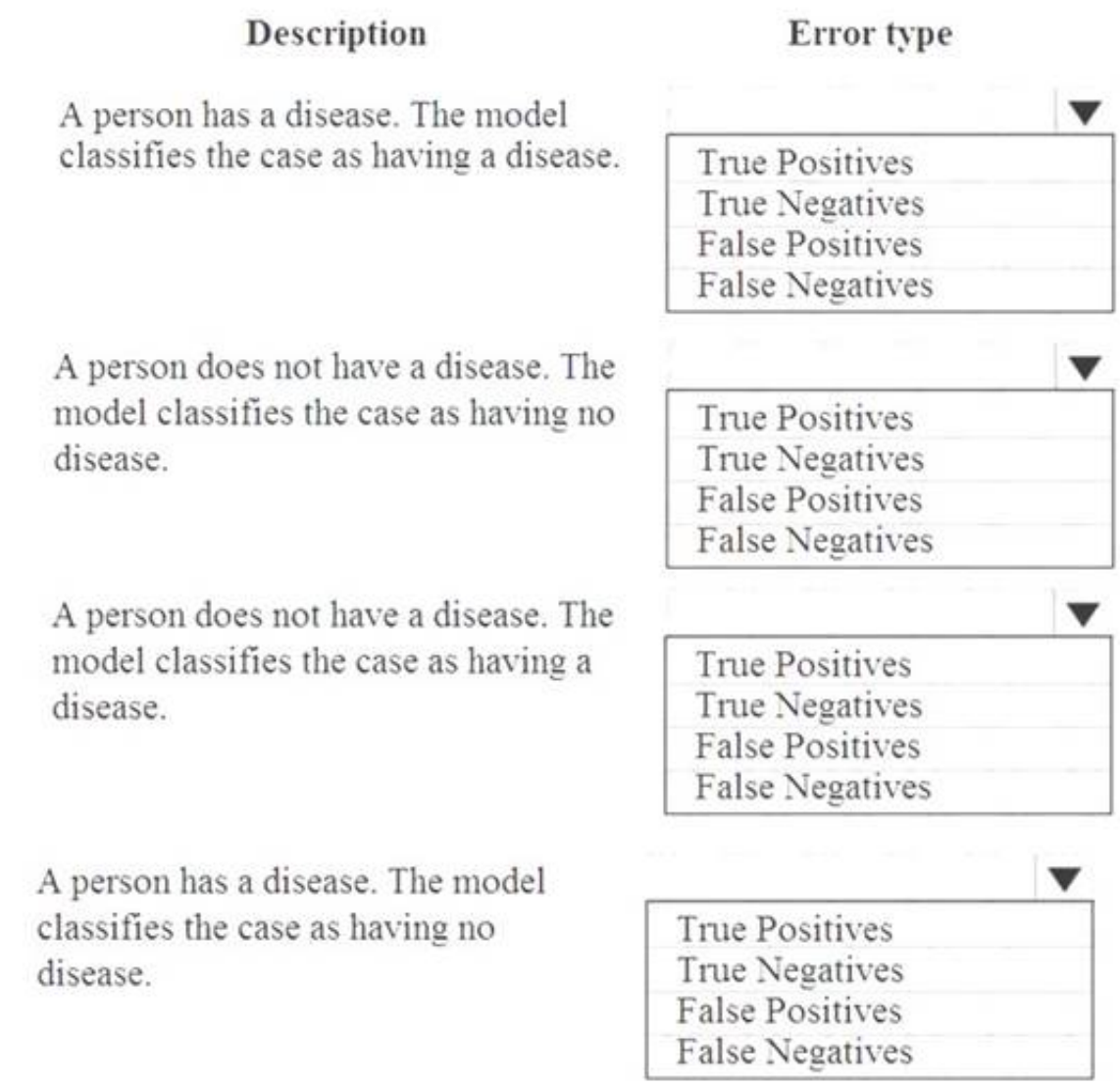
NEW QUESTION 133
- (Exam Topic 3)
You are creating a machine learning model. You have a dataset that contains null rows.
You need to use the Clean Missing Data module in Azure Machine Learning Studio to identify and resolve the null and missing data in the dataset.
Which parameter should you use?

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

Answer: B

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

NEW QUESTION 136
- (Exam Topic 3)
You create a binary classification model to predict whether a person has a disease. You need to detect possible classification errors.
Which error type should you choose for each description? To answer, select the appropriate options in the answer area.
NOTE: Each correct selection is worth one point.



A. Mastered
B. Not Mastered

Answer: A

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

NEW QUESTION 140
- (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, Accuracy, Precision, Recall, F1 score, and AUC.
Does the solution meet the goal?

A. Yes
B. No

Answer: B

Explanation:
Accuracy, Precision, Recall, F1 score, and AUC are metrics for evaluating classification models. Note: Mean Absolute Error, Root Mean Absolute Error, Relative Absolute Error are OK for the linear
regression model.
References:
https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/evaluate-model

NEW QUESTION 141
- (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)

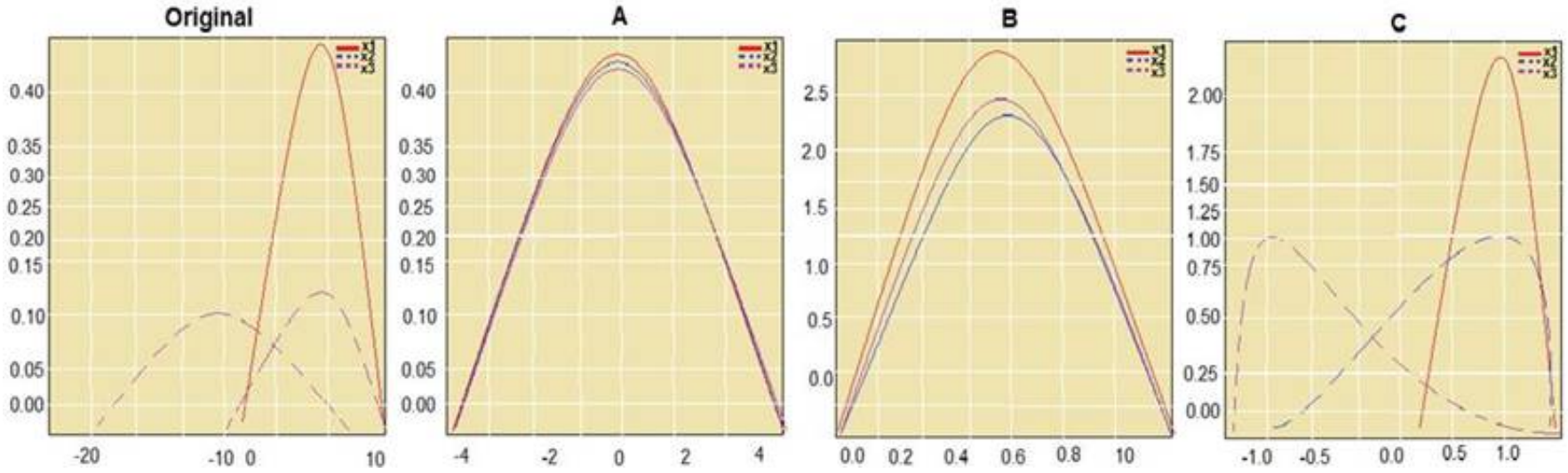
```
from azureml.core import Workspace
```

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 145
- (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.



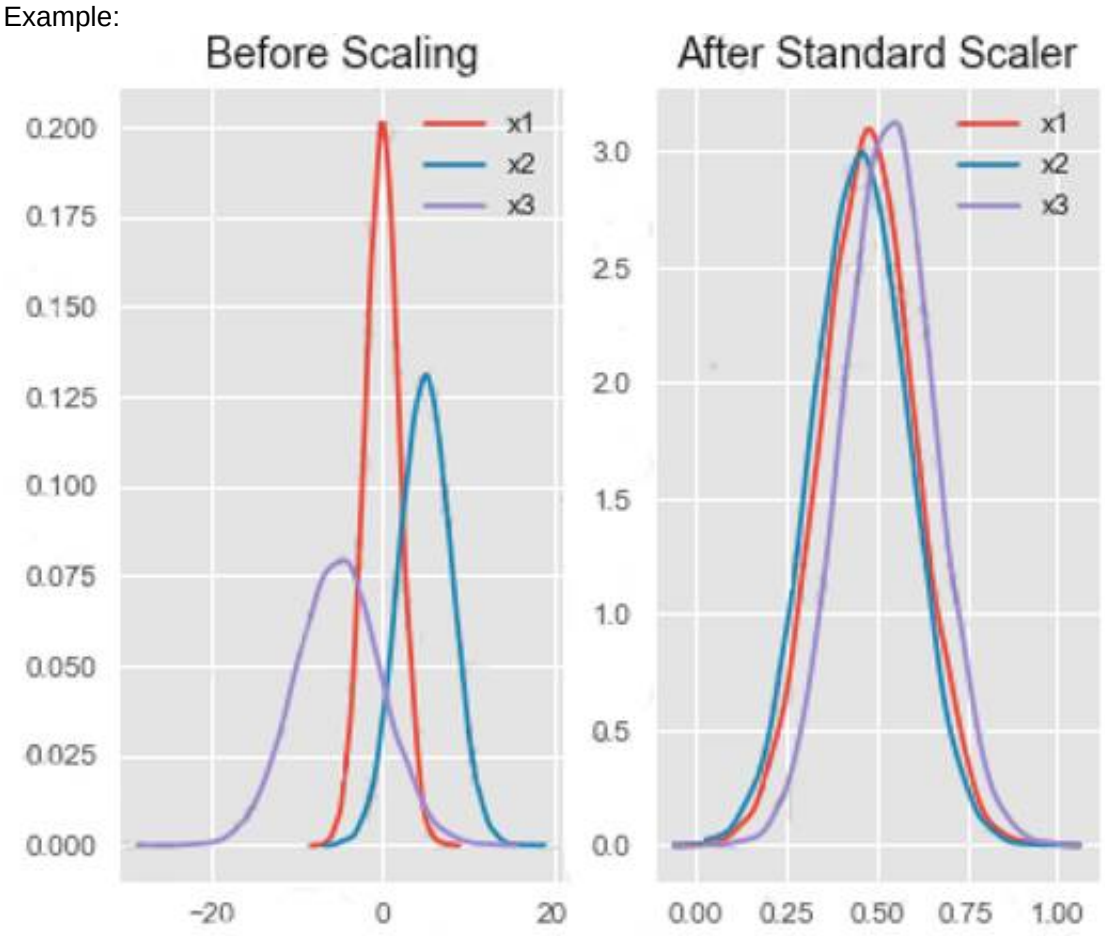
Use the drop-down menus to select the answer choice that answers each question based on the information presented in the graphic.
NOTE: Each correct selection is worth one point.

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

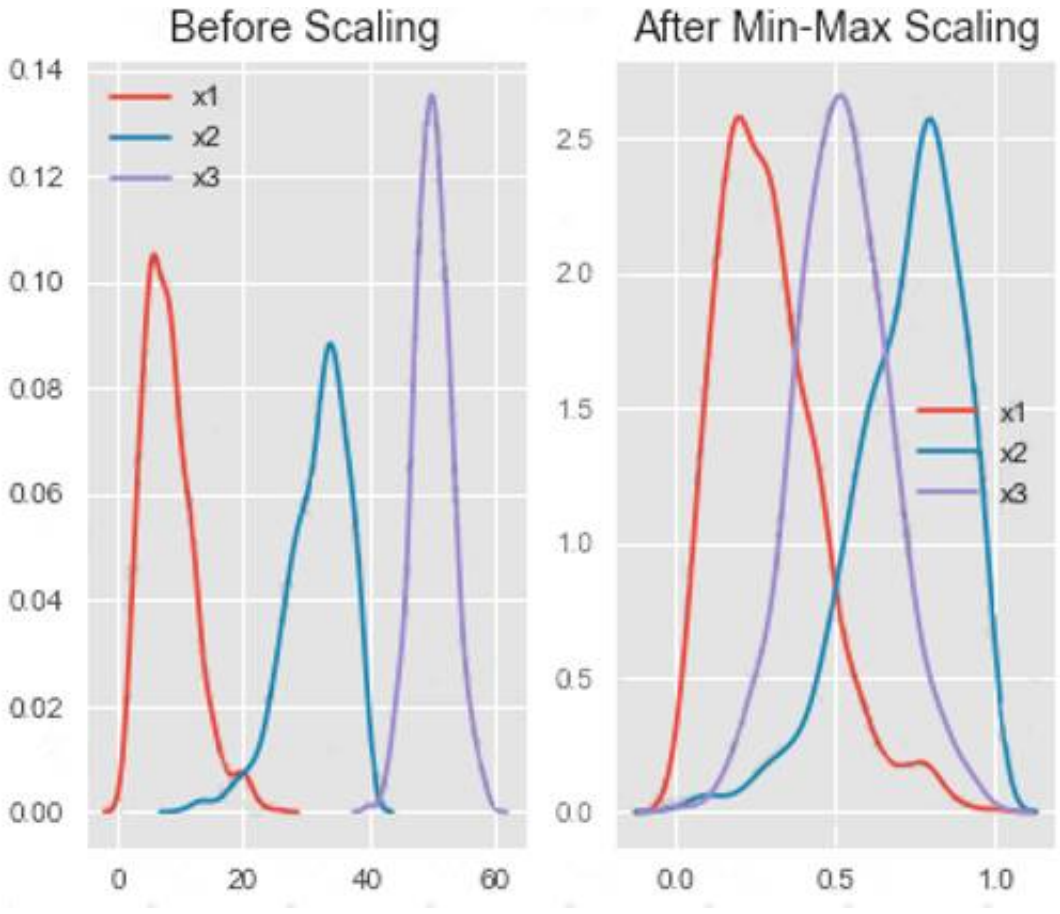
A. Mastered
B. Not Mastered

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 147
- (Exam Topic 3)
An organization creates and deploys a multi-class image classification deep learning model that uses a set of labeled photographs.
The software engineering team reports there is a heavy inferencing load for the prediction web services during the summer. The production web service for the model fails to meet demand despite having a fully-utilized compute cluster where the web service is deployed.
You need to improve performance of the image classification web service with minimal downtime and minimal administrative effort.
What should you advise the IT Operations team to do?

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

Answer: D

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

NEW QUESTION 152
- (Exam Topic 3)
Note: This question is part of a series of questions that present the same scenario. Each question in the series contains a unique solution that might meet the stated goals. Some question sets might have more than one correct solution, while others might not have a correct solution.
After you answer a question in this section, you will NOT be able to return to it. As a result, these questions will not appear in the review screen.
You 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

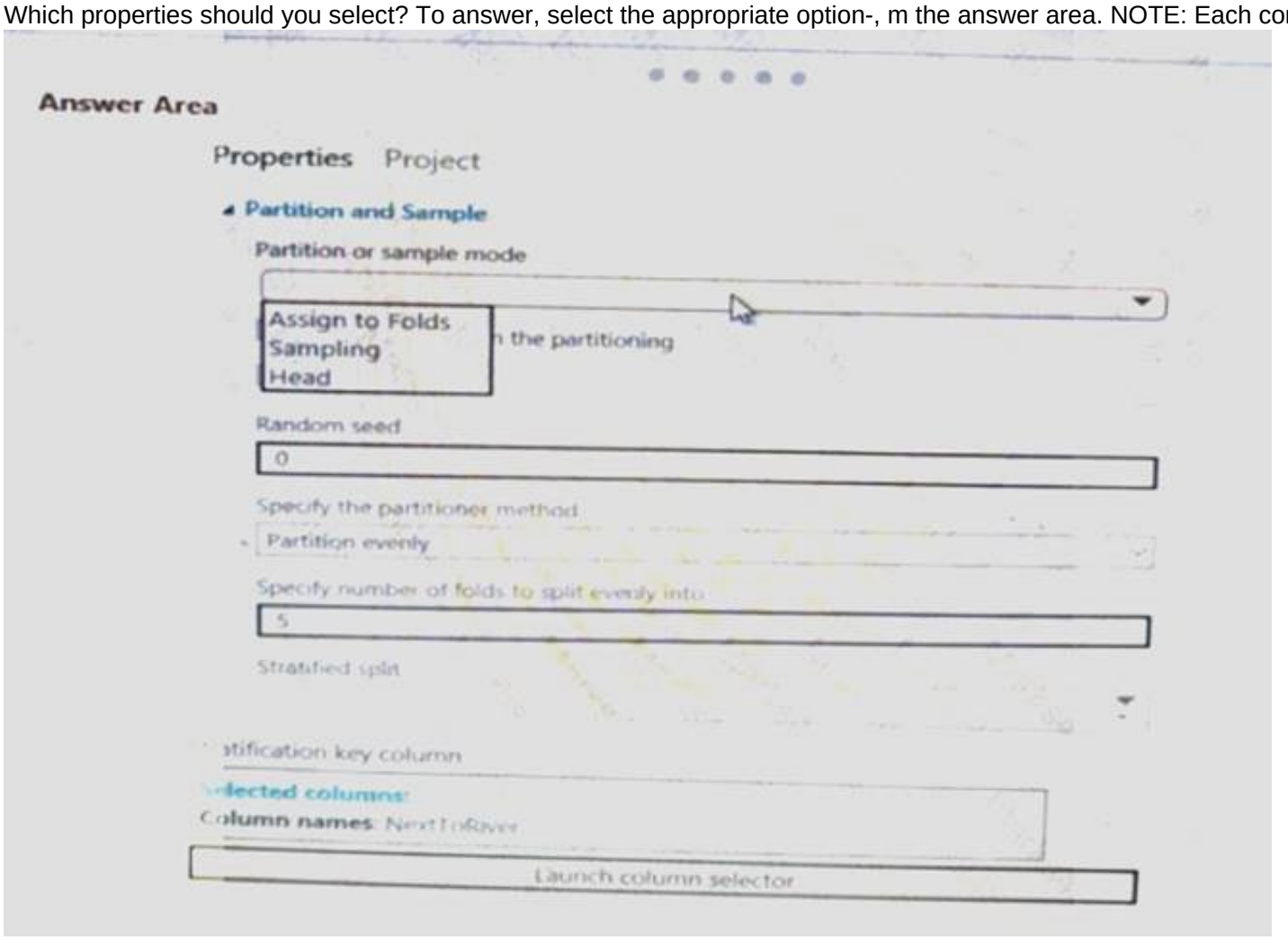
NEW QUESTION 157
- (Exam Topic 2)
You need to select a feature extraction method. Which method should you use?

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

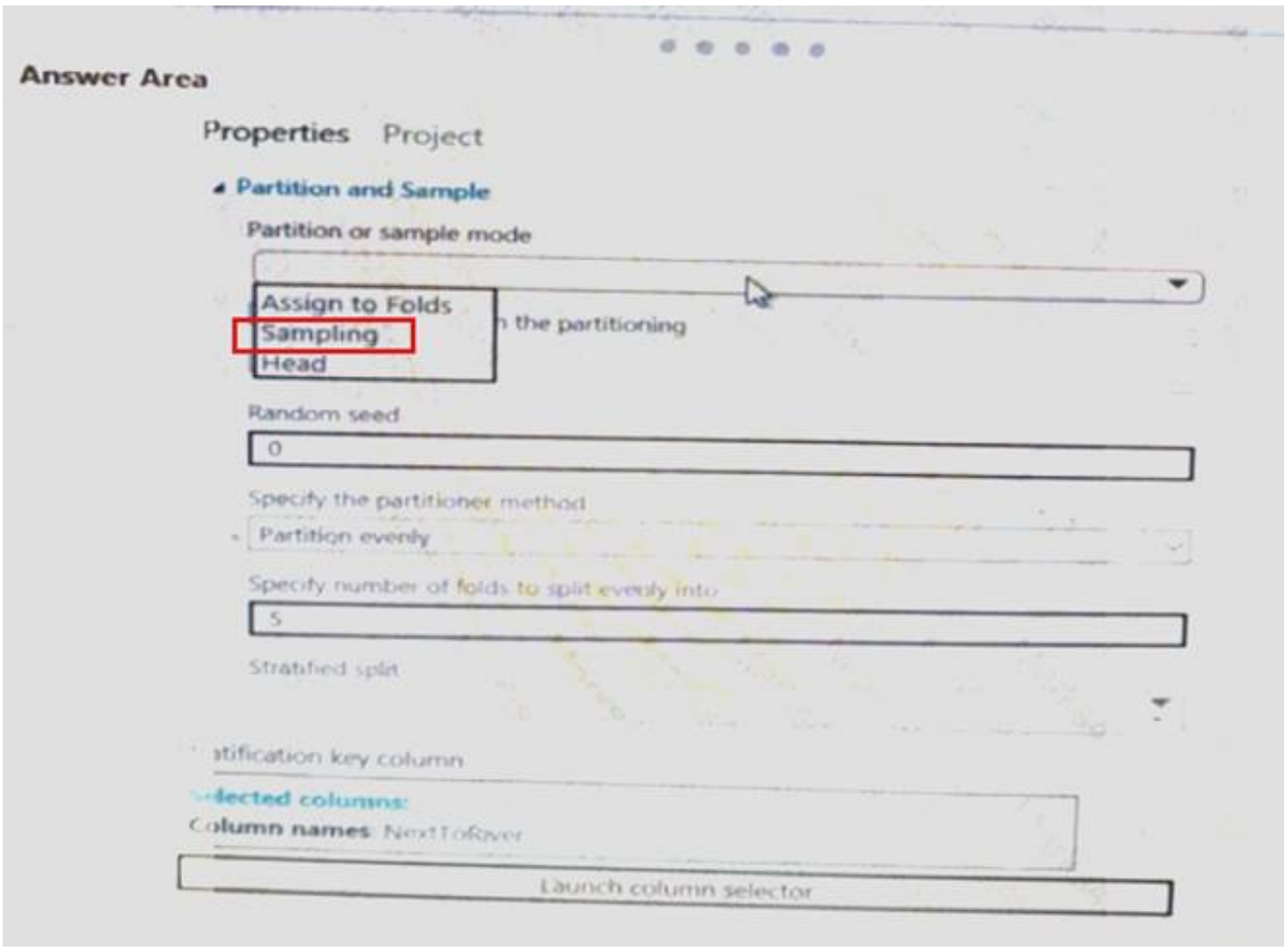
NEW QUESTION 158
- (Exam Topic 2)
You need to identify the methods for dividing the data according, to the testing requirements.
Which properties should you select? To answer, select the appropriate option-, m the answer area. NOTE: Each correct selection is worth one point.



A. Mastered
B. Not Mastered

Answer: A

Explanation:

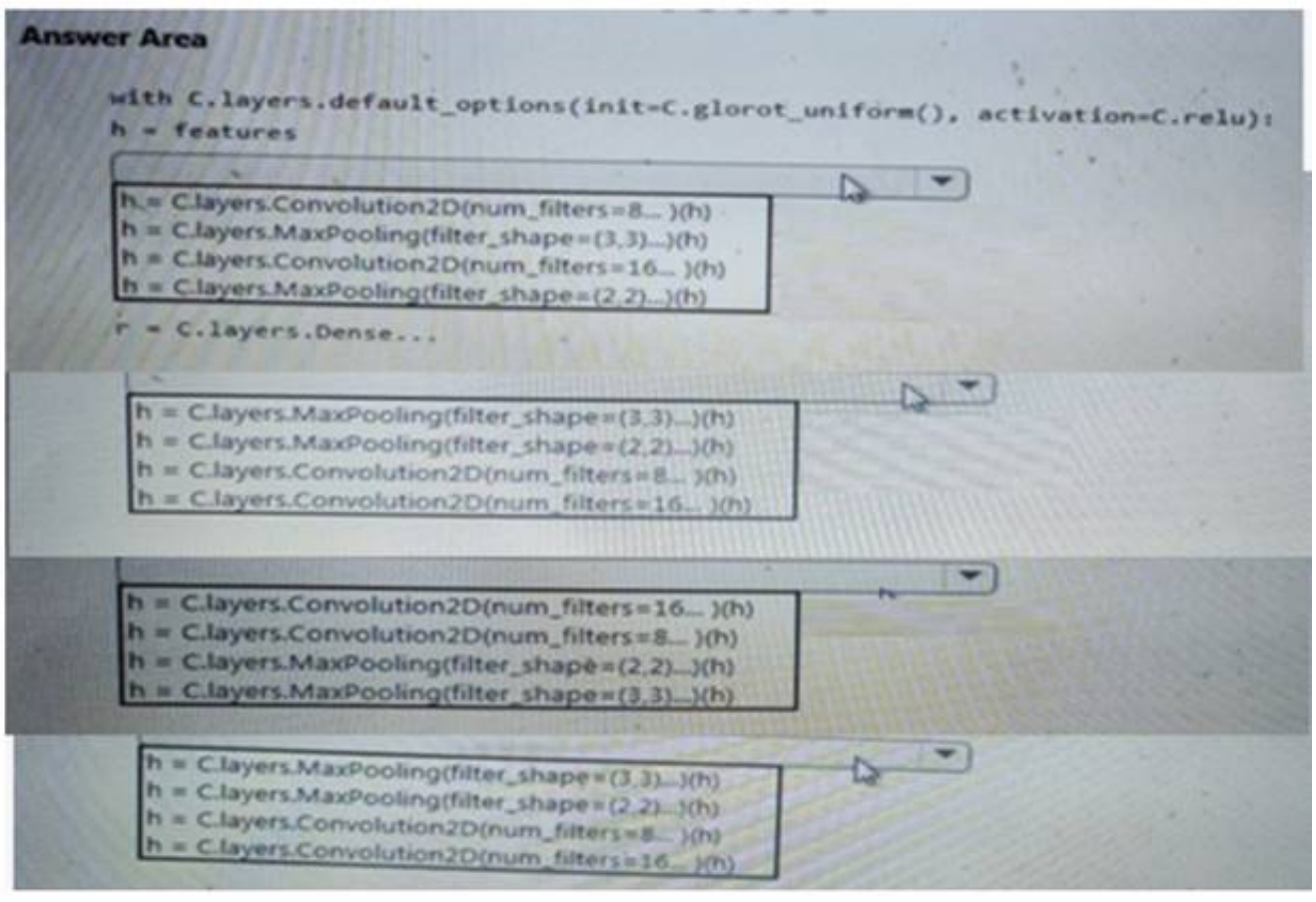


NEW QUESTION 162
- (Exam Topic 1)
You need to select an environment that will meet the business and data requirements. Which environment should you use?

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

Answer: D

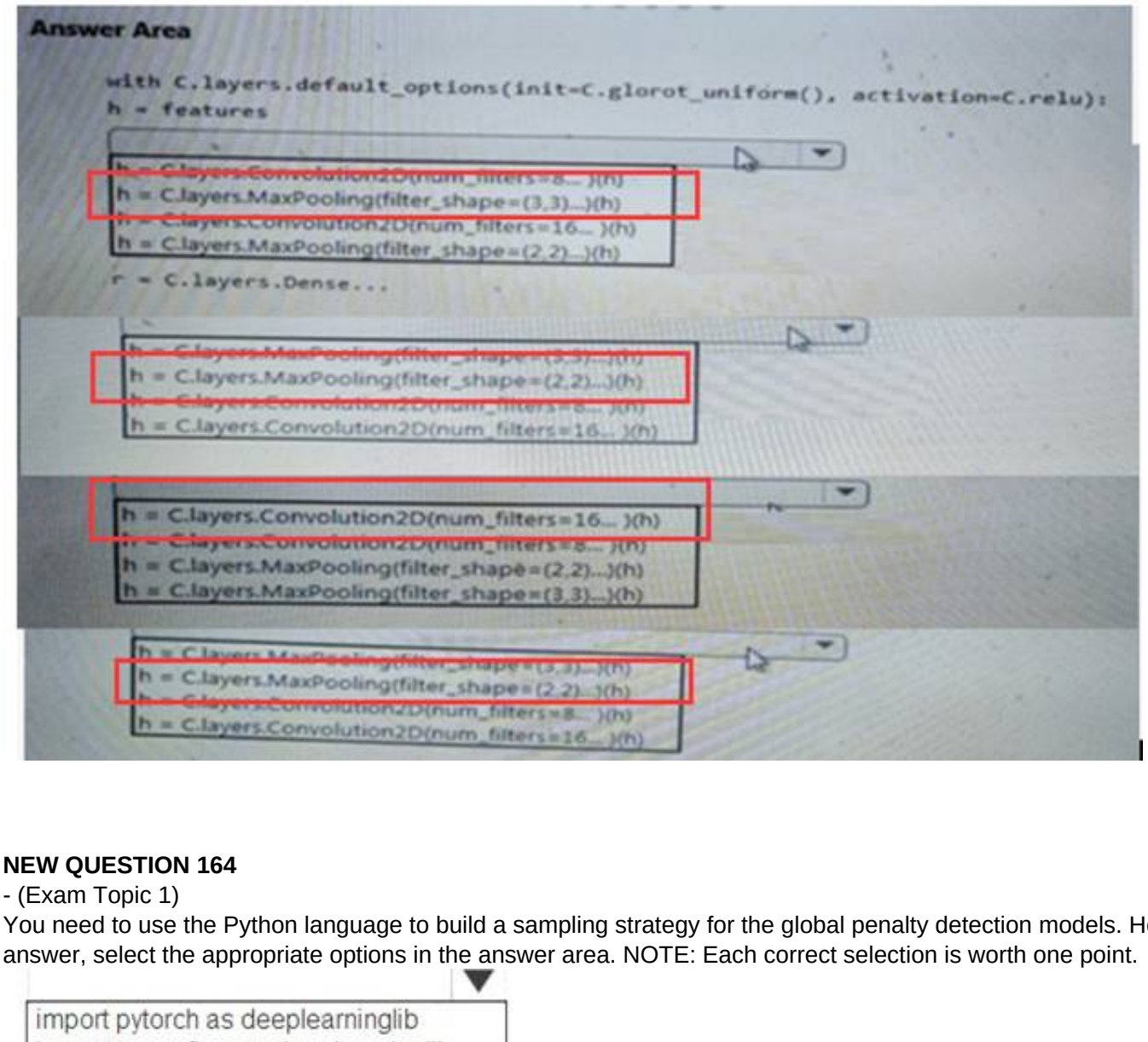
NEW QUESTION 163
- (Exam Topic 1)
You need to build a feature extraction strategy for the local models.
How should you complete the code segment? To answer, select the appropriate options in the answer area. NOTE: Each correct selection is worth one point.



A. Mastered
B. Not Mastered

Answer: A

Explanation:



NEW QUESTION 164
- (Exam Topic 1)
You need to use the Python language to build a sampling strategy for the global penalty detection models. How should you complete the code segment? To answer, select the appropriate options in the answer area. NOTE: Each correct selection is worth one point.

A. Mastered
B. Not Mastered

Answer: A

Explanation:
Box 1: import pytorch as deeplearninglib Box 2: ..DistributedSampler(Sampler).. DistributedSampler(Sampler):
Sampler that restricts data loading to a subset of the dataset.
It is especially useful in conjunction with class:`torch.nn.parallel.DistributedDataParallel`. In such case, each process can pass a DistributedSampler instance as a DataLoader sampler, and load a subset of the original dataset that is exclusive to it.
Scenario: Sampling must guarantee mutual and collective exclusively between local and global segmentation models that share the same features.
Box 3: optimizer = deeplearninglib.train. GradientDescentOptimizer(learning_rate=0.10)

NEW QUESTION 167
.....

Relate Links

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

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

Contact us

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

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