# Microsoft

## Exam Questions DP-100

Designing and Implementing a Data Science Solution on Azure

**NEW QUESTION 1**
- (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.

| Action | Answer area |
|---|---|
| Evaluate the probability function | |
| Remove duplicate rows | |
| Use the Filter Based Feature Selection module | |
| Test the hypothesis using t-Test | |
| Compute linear correlation | |
| Build a counting transform | |

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 2**
- (Exam Topic 3)
You have a dataset that includes confidential data. You use the dataset to train a model.
You must use a differential privacy parameter to keep the data of individuals safe and private. You need to reduce the effect of user data on aggregated results.
What should you do?

A. Decrease the value of the epsilon parameter to reduce the amount of noise added to the data
B. Increase the value of the epsilon parameter to decrease privacy and increase accuracy
C. Decrease the value of the epsilon parameter to increase privacy and reduce accuracy
D. Set the value of the epsilon parameter to 1 to ensure maximum privacy

**Answer:** C

**Explanation:**
Differential privacy tries to protect against the possibility that a user can produce an indefinite number of reports to eventually reveal sensitive data. A value known as epsilon measures how noisy, or private, a report is. Epsilon has an inverse relationship to noise or privacy. The lower the epsilon, the more noisy (and private) the data is.
Reference:
https://docs.microsoft.com/en-us/azure/machine-learning/concept-differential-privacy

**NEW QUESTION 3**
- (Exam Topic 3)
You are creating a classification model for a banking company to identify possible instances of credit card fraud. You plan to create the model in Azure Machine Learning by using automated machine learning.
The training dataset that you are using is highly unbalanced. You need to evaluate the classification model.
Which primary metric should you use?

A. normalized_mean_absolute_error
B. [spearman_correlation
C. AUC.weighted
D. accuracy
E. normalized_root_mean_squared_error

**Answer:** C

**Explanation:**
AUC_weighted is a Classification metric.
Note: AUC is the Area under the Receiver Operating Characteristic Curve. Weighted is the arithmetic mean of the score for each class, weighted by the number of true instances in each class.
Reference:
https://docs.microsoft.com/en-us/azure/machine-learning/how-to-understand-automated-ml

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

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

**Answer:** C

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

**NEW QUESTION 5**
- (Exam Topic 3)
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: Relative Squared Error, Coefficient of Determination, Accuracy, Precision, Recall, F1 score, and AUC.
Does the solution meet the goal?

A. Yes
B. No

**Answer:** B

**Explanation:**
Relative Squared Error, Coefficient of Determination are good metrics to evaluate the linear regression model, but the others are metrics for classification models.
References:
https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/evaluate-model

**NEW QUESTION 6**
- (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 7**
- (Exam Topic 3)
You have a dataset that contains over 150 features. You use the dataset to train a Support Vector Machine (SVM) binary classifier.
You need to use the Permutation Feature Importance module in Azure Machine Learning Studio to compute a set of feature importance scores for the dataset.
In which order should you perform the actions? To answer, move all actions from the list of actions to the answer area and arrange them in the correct order.

**Actions**

| Add a Two-Class Support Vector Machine module to initialize the SVM classifier. |

| Set the Metric for measuring performance property to **Classification - Accuracy** and then run the experiment. |

| Add a Permutation Feature Importance module and connect the trained model and test dataset. |

| Add a dataset to the experiment. |

| Add a Split Data module to create training and test datasets. |

**Answer Area**

A. Mastered
B. Not Mastered

**Answer:** A

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

**NEW QUESTION 8**
- (Exam Topic 3)
You are developing deep learning models to analyze semi-structured, unstructured, and structured data types. You have the following data available for model building:
≽ Video recordings of sporting events
≽ Transcripts of radio commentary about events
≽ Logs from related social media feeds captured during sporting events You need to select an environment for creating the model.
Which environment should you use?

A. Azure Cognitive Services
B. Azure Data Lake Analytics
C. Azure HDInsight with Spark MLib
D. Azure Machine Learning Studio

**Answer:** A

**Explanation:**
Azure Cognitive Services expand on Microsoft's evolving portfolio of machine learning APIs and enable developers to easily add cognitive features – such as emotion and video detection; facial, speech, and vision recognition; and speech and language understanding – into their applications. The goal of Azure Cognitive Services is to help developers create applications that can see, hear, speak, understand, and even begin to reason. The catalog of services within Azure Cognitive Services can be categorized into five main pillars - Vision, Speech, Language, Search, and Knowledge.
References:
https://docs.microsoft.com/en-us/azure/cognitive-services/welcome

**NEW QUESTION 9**
- (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 10**
- (Exam Topic 3)
You use the Azure Machine Learning designer to create and run a training pipeline.
The pipeline must be run every night to inference predictions from a large volume of files. The folder where the files will be stored is defined as a dataset.
You need to publish the pipeline as a REST service that can be used for the nightly inferencing run. What should you do?

A. Create a batch inference pipeline
B. Set the compute target for the pipeline to an inference cluster
C. Create a real-time inference pipeline
D. Clone the pipeline

**Answer:** A

**Explanation:**
Azure Machine Learning Batch Inference targets large inference jobs that are not time-sensitive. Batch Inference provides cost-effective inference compute scaling, with unparalleled throughput for asynchronous applications. It is optimized for high-throughput, fire-and-forget inference over large collections of data.
You can submit a batch inference job by pipeline_run, or through REST calls with a published pipeline. Reference:
https://github.com/Azure/MachineLearningNotebooks/blob/master/how-to-use-azureml/machine-learning-pipeli

**NEW QUESTION 10**
- (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 15**
- (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 20**
- (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 23**
- (Exam Topic 3)
You are with a time series dataset in Azure Machine Learning Studio.
You need to split your dataset into training and testing subsets by using the Split Data module. Which splitting mode should you use?

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

**Answer:** B

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

**NEW QUESTION 26**
- (Exam Topic 3)
You create a binary classification model by using Azure Machine Learning Studio.
You must tune hyperparameters by performing a parameter sweep of the model. The parameter sweep must meet the following requirements:

≫ iterate all possible combinations of hyperparameters

≫ minimize computing resources required to perform the sweep

≫ You need to perform a parameter sweep of the model.
Which parameter sweep mode should you use?

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

**Answer:** D

**Explanation:**
Maximum number of runs on random grid: This option also controls the number of iterations over a random sampling of parameter values, but the values are not generated randomly from the specified range; instead, a matrix is created of all possible combinations of parameter values and a random sampling is taken over the matrix. This method is more efficient and less prone to regional oversampling or undersampling.
If you are training a model that supports an integrated parameter sweep, you can also set a range of seed values to use and iterate over the random seeds as well. This is optional, but can be useful for avoiding bias introduced by seed selection.

**NEW QUESTION 29**
- (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 30**
- (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 32**
- (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 33**
- (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.
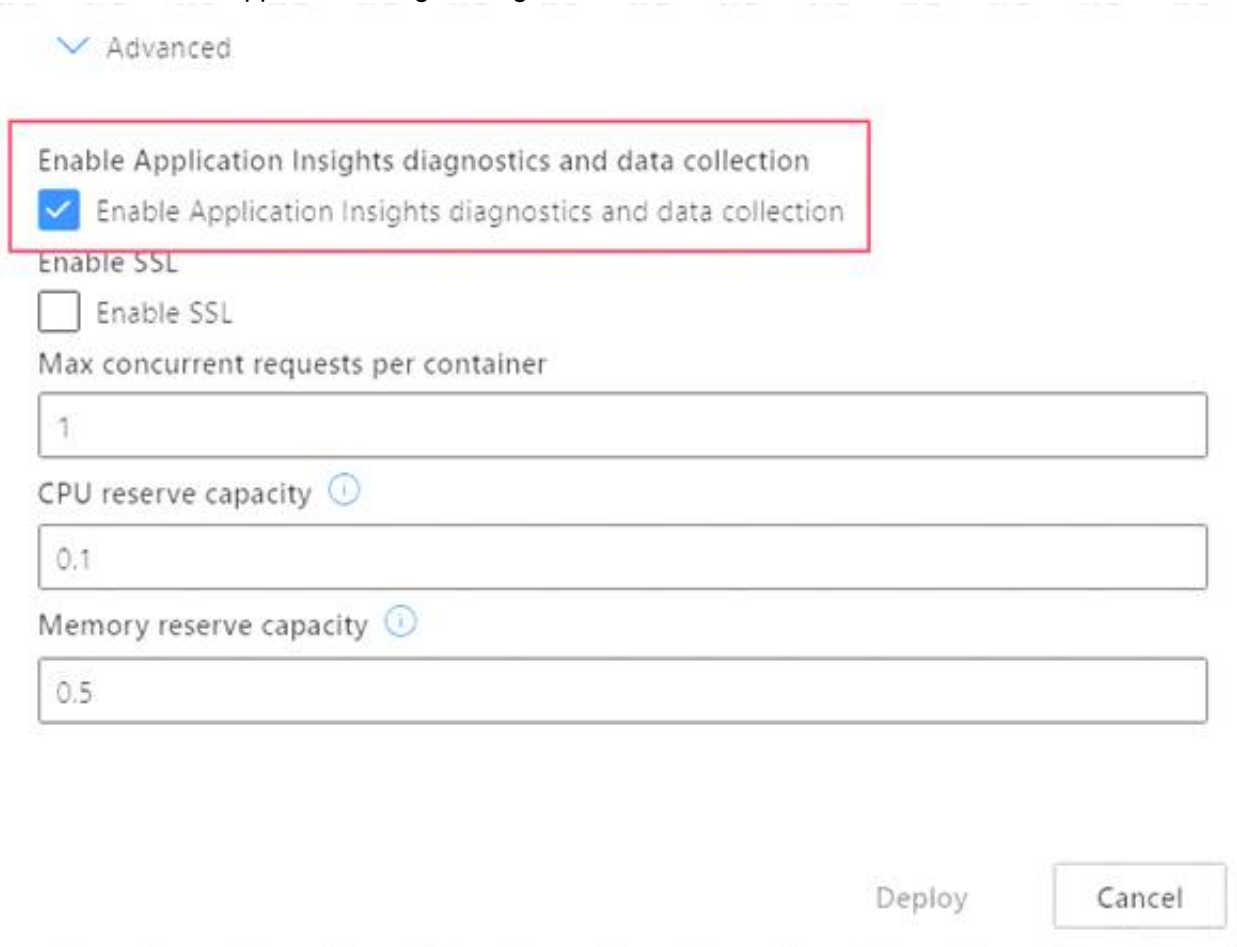
**Answer:** B

**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.

```
∨ Advanced

Enable Application Insights diagnostics and data collection
☑ Enable Application Insights diagnostics and data collection
Enable SSL
☐ Enable SSL
Max concurrent requests per container
1
CPU reserve capacity ⓘ
0.1
Memory reserve capacity ⓘ
0.5

                                    Deploy      Cancel
```

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


**NEW QUESTION 34**
- (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 36**
- (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<br>Microsoft.MachineLearningServices/workspaces/computes/*/write<br>Microsoft.MachineLearningServices/workspaces/delete |

| ▼ |
|---|
| Microsoft.MachineLearningServices/workspaces/*/write<br>Microsoft.MachineLearningServices/workspaces/computes/*/write<br>Microsoft.MachineLearningServices/workspaces/delete |

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

| ▼ |
|---|
| Microsoft.MachineLearningServices/workspaces/*/read<br>Microsoft.MachineLearningServices/workspaces/*/write<br>Microsoft.MachineLearningServices/workspaces/computes/*/delete |

| ▼ |
|---|
| Microsoft.MachineLearningServices/workspaces/*/read<br>Microsoft.MachineLearningServices/workspaces/*/write<br>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 39**
- (Exam Topic 3)
You create a multi-class image classification deep learning experiment by using the PyTorch framework. You plan to run the experiment on an Azure Compute cluster that has nodes with GPU's.
You need to define an Azure Machine Learning service pipeline to perform the monthly retraining of the image classification model. The pipeline must run with minimal cost and minimize the time required to train the model.
Which three pipeline steps should you run in sequence? To answer, move the appropriate actions from the list of actions to the answer area and arrange them in the correct order.

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

A. Mastered
B. Not Mastered

**Answer:** A

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

**NEW QUESTION 40**
- (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 tin- other classes in the training set. You need to select an appropriate data sampling strategy to compensate for the class imbalance. Solution: You use the Principal Components Analysis (PCA) sampling mode.
Does the solution meet the goal?

A. Yes
B. No

**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.

**NEW QUESTION 41**
- (Exam Topic 3)
You create an Azure Machine Learning workspace named ML-workspace. You also create an Azure Databricks workspace named DB-workspace. DB-workspace contains a cluster named DB-cluster.
You must use DB-cluster to run experiments from notebooks that you import into DB-workspace.
You need to use ML-workspace to track MLflow metrics and artifacts generated by experiments running on DB-cluster. The solution must minimize the need for custom code.
What should you do?

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

**Answer:** B

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

**NEW QUESTION 42**
- (Exam Topic 3)
Note: This question is part of a series of questions that present the same scenario. Each question in the series contains a unique solution that might meet the stated goals. Some question sets might have more than one correct solution, while others might not have a correct solution.
After you answer a question in this section, you will NOT be able to return to it. As a result, these questions will not appear in the review screen.
You are analyzing a numerical dataset which contains missing values in several columns.
You must clean the missing values using an appropriate operation without affecting the dimensionality of the feature set.
You need to analyze a full dataset to include all values.
Solution: Replace each missing value using the Multiple Imputation by Chained Equations (MICE) method. Does the solution meet the goal?

A. Yes
B. NO

**Answer:** A

**Explanation:**
Replace using MICE: For each missing value, this option assigns a new value, which is calculated by using a method described in the statistical literature as "Multivariate Imputation using Chained Equations" or "Multiple Imputation by Chained Equations". With a multiple imputation method, each variable with missing data is modeled conditionally using the other variables in the data before filling in the missing values.
Note: Multivariate imputation by chained equations (MICE), sometimes called "fully conditional specification" or "sequential regression multiple imputation" has emerged in the statistical literature as one principled method of addressing missing data. Creating multiple imputations, as opposed to single imputations, accounts for the statistical uncertainty in the imputations. In addition, the chained equations approach is very flexible and can handle variables of varying types (e.g., continuous or binary) as well as complexities such as bounds or survey skip patterns.
References: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3074241/
https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/clean-missing-data

**NEW QUESTION 45**
- (Exam Topic 3)
You plan to run a Python script as an Azure Machine Learning experiment.
The script must read files from a hierarchy of folders. The files will be passed to the script as a dataset argument.
You must specify an appropriate mode for the dataset argument.
Which two modes can you use? Each correct answer presents a complete solution.
NOTE: Each correct selection is worth one point.

A. to_pandas_dataframe ()
B. as_download()
C. as_upload()
D. as mount ()

**Answer:** B

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

**NEW QUESTION 46**
- (Exam Topic 3)
You use an Azure Machine Learning workspace.
You have a trained model that must be deployed as a web service. Users must authenticate by using Azure Active Directory.
What should you do?

A. Deploy the model to Azure Kubernetes Service (AKS). During deployment, set the token_auth_enabled parameter of the target configuration object to true
B. Deploy the model to Azure Container Instance
C. During deployment, set the auch_enabled parameter of the target configuration object to true
D. Deploy the model to Azure Container Instance
E. During deployment, set the coken_auch_enabled parameter of the target configuration object to true
F. Deploy the model to Azure Kubernetes Service (AKS). During deployment, set the auc
G. enabled parameter of the target configuration object to true

**Answer:** A

**Explanation:**
To control token authentication, use the token_auth_enabled parameter when you create or update a deployment
Token authentication is disabled by default when you deploy to Azure Kubernetes Service.
Note: The model deployments created by Azure Machine Learning can be configured to use one of two authentication methods:
key-based: A static key is used to authenticate to the web service.
token-based: A temporary token must be obtained from the Azure Machine Learning workspace (using Azure Active Directory) and used to authenticate to the web service.
Reference:
https://docs.microsoft.com/en-us/azure/machine-learning/how-to-authenticate-web-service

**NEW QUESTION 49**
- (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 True.
Deploy the model to the service. Does the solution meet the goal?

A. Yes
B. No

**Answer:** A

**Explanation:**
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 51**
- (Exam Topic 3)
You are developing a data science workspace that uses an Azure Machine Learning service. You need to select a compute target to deploy the workspace.
What should you use?

A. Azure Data Lake Analytics
B. Azure Databrick .
C. Apache Spark for HDInsight.
D. Azure Container Service

**Answer:** D

**Explanation:**
Azure Container Instances can be used as compute target for testing or development. Use for low-scale CPU-based workloads that require less than 48 GB of RAM.
Reference:
https://docs.microsoft.com/en-us/azure/machine-learning/service/how-to-deploy-and-where


**NEW QUESTION 53**
- (Exam Topic 3)
You are building an intelligent solution using machine learning models. The environment must support the following requirements:

≫ Data scientists must build notebooks in a cloud environment

≫ Data scientists must use automatic feature engineering and model building in machine learning pipelines.

≫ Notebooks must be deployed to retrain using Spark instances with dynamic worker allocation.

≫ Notebooks must be exportable to be version controlled locally.
You need to create the environment.
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:**
Step 1: Create an Azure HDInsight cluster to include the Apache Spark Mlib library

Step 2: Install Microsot Machine Learning for Apache Spark You install AzureML on your Azure HDInsight cluster.

Microsoft Machine Learning for Apache Spark (MMLSpark) provides a number of deep learning and data science tools for Apache Spark, including seamless integration of Spark Machine Learning pipelines with Microsoft Cognitive Toolkit (CNTK) and OpenCV, enabling you to quickly create powerful, highly-scalable predictive and analytical models for large image and text datasets.

Step 3: Create and execute the Zeppelin notebooks on the cluster

Step 4: When the cluster is ready, export Zeppelin notebooks to a local environment. Notebooks must be exportable to be version controlled locally.

References:

https://docs.microsoft.com/en-us/azure/hdinsight/spark/apache-spark-zeppelin-notebook https://azuremlbuild.blob.core.windows.net/pysparkapi/intro.html

**NEW QUESTION 57**

- (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 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 60**

- (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 61**
- (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 63**
- (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 MimicExplainer. Does the solution meet the goal?

A. Yes
B. No

**Answer:** B

**Explanation:**

Instead use Permutation Feature Importance Explainer (PFI).
Note 1: Mimic explainer is based on the idea of training global surrogate models to mimic blackbox models. A global surrogate model is an intrinsically interpretable model that is trained to approximate the predictions of any black box model as accurately as possible. Data scientists can interpret the surrogate model to draw conclusions about the black box model.
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 68**
- (Exam Topic 3)
You create an Azure Databricks workspace and a linked Azure Machine Learning workspace. You have the following Python code segment in the Azure Machine Learning workspace: import mlflow
import mlflow.azureml import azureml.mlflow import azureml.core
from azureml.core import Workspace subscription_id = 'subscription_id' resourse_group = 'resource_group_name' workspace_name = 'workspace_name'
ws = Workspace.get(name=workspace_name, subscription_id=subscription_id, resource_group=resource_group)
experimentName = "/Users/{user_name}/{experiment_folder}/{experiment_name}" mlflow.set_experiment(experimentName)
uri = ws.get_mlflow_tracking_uri() mlflow.set_tracking_uri(uri)
Instructions: For each of the following statements, select Yes if the statement is true. Otherwise, select No.
NOTE: Each correct selection is worth one point.

| | Yes | No |
|---|---|---|
| A resource group and Azure Machine Learning workspace will be created. | ○ | ○ |
| An Azure Databricks experiment will be tracked only in the Azure Machine Learning workspace. | ○ | ○ |
| The epoch loss metric is set to be tracked. | ○ | ○ |

A. Mastered
B. Not Mastered

**Answer:** A

**Explanation:**
A screenshot of a computer Description automatically generated with medium confidence
Box 1: No
The Workspace.get method loads an existing workspace without using configuration files. ws = Workspace.get(name="myworkspace", subscription_id='<azure-subscription-id>', resource_group='myresourcegroup')
Box 2: Yes
MLflow Tracking with Azure Machine Learning lets you store the logged metrics and artifacts from your local runs into your Azure Machine Learning workspace. The get_mlflow_tracking_uri() method assigns a unique tracking URI address to the workspace, ws, and set_tracking_uri() points the MLflow tracking URI to that address.
Box 3: Yes
Note: In Deep Learning, epoch means the total dataset is passed forward and backward in a neural network once.
Reference:
https://docs.microsoft.com/en-us/python/api/azureml-core/azureml.core.workspace.workspace https://docs.microsoft.com/en-us/azure/machine-learning/how-to-use-mlflow

**NEW QUESTION 71**
- (Exam Topic 3)
You are building a recurrent neural network to perform a binary classification. You review the training loss, validation loss, training accuracy, and validation accuracy for each training epoch.
You need to analyze model performance.
Which observation indicates that the classification model is over fitted?

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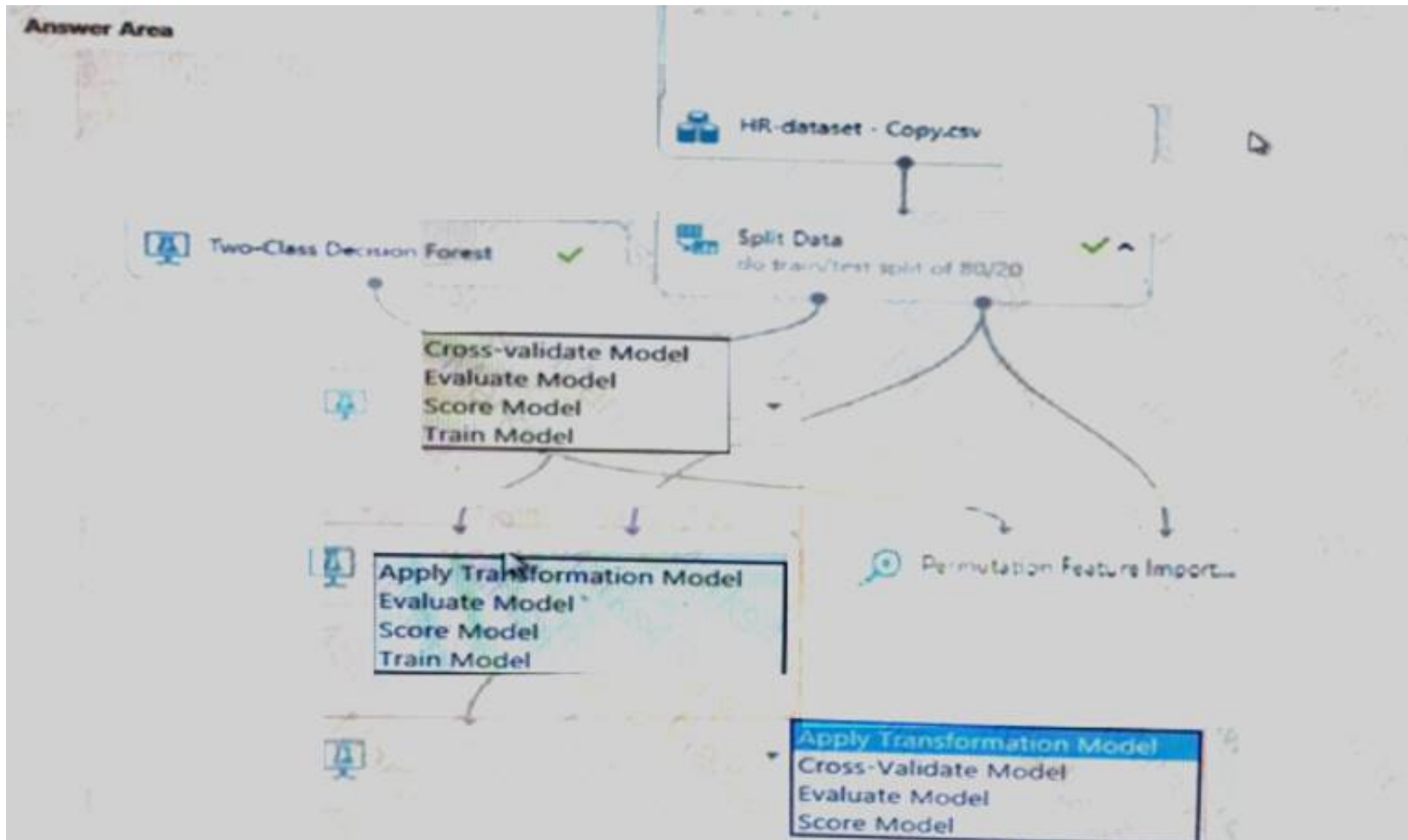
**Answer:** B

**NEW QUESTION 72**
- (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.
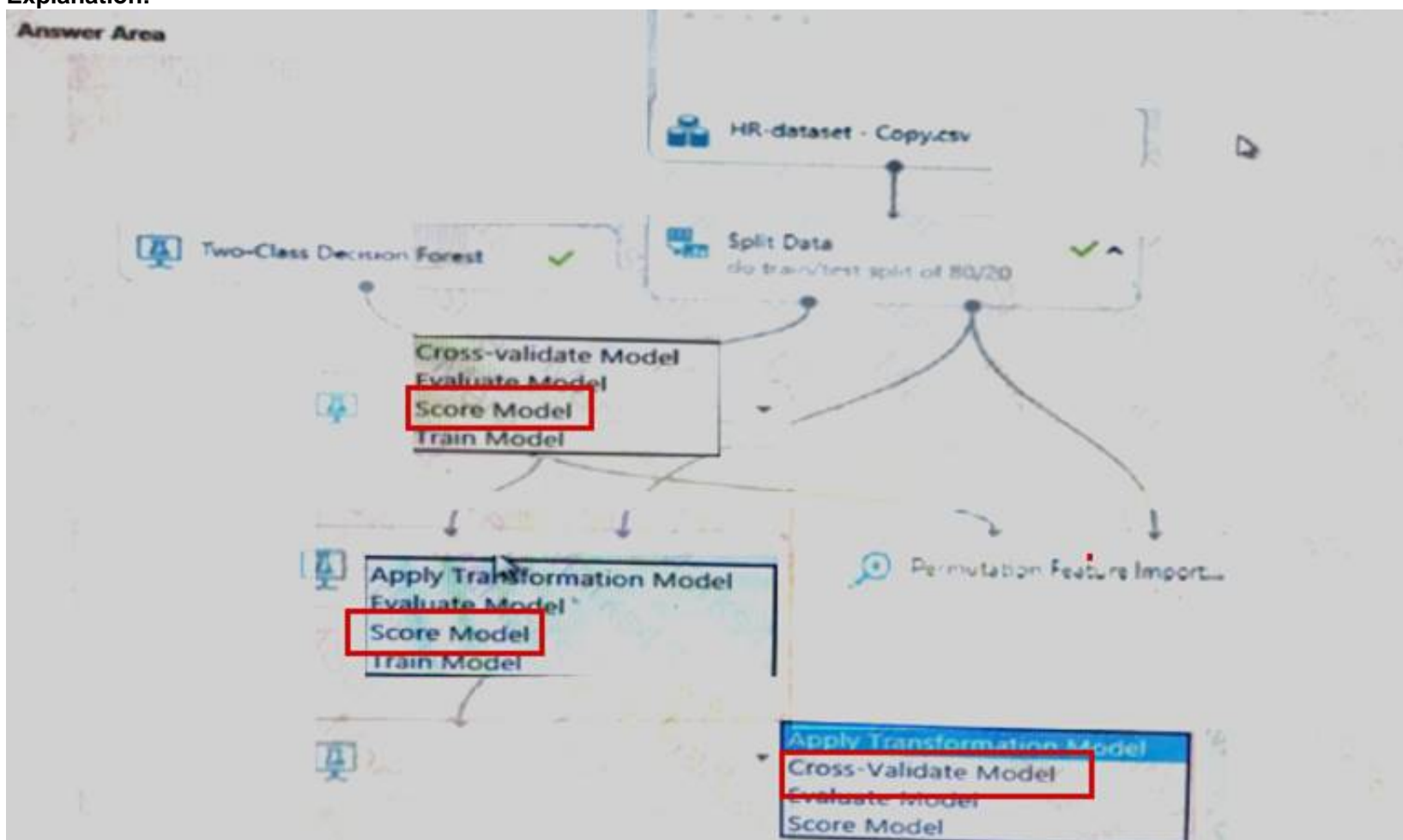NOTE: Each correct selection is worth one point.

A. Mastered
B. Not Mastered

**Answer:** A

**Explanation:**



**NEW QUESTION 74**
- (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 75
- (Exam Topic 3)
You plan to provision an Azure Machine Learning Basic edition workspace for a data science project. You need to identify the tasks you will be able to perform in the workspace.
Which three tasks will you be able to perform? Each correct answer presents a complete solution.
NOTE: Each correct selection is worth one point. D

A. Create a Compute Instance and use it to run code in Jupyter notebooks.
B. Create an Azure Kubernetes Service (AKS) inference cluster.
C. Use the designer to train a model by dragging and dropping pre-defined modules.
D. Create a tabular dataset that supports versioning.
E. Use the Automated Machine Learning user interface to train a model.

**Answer:** ABD

**Explanation:**
Reference:
https://azure.microsoft.com/en-us/pricing/details/machine-learning/

## NEW QUESTION 78
- (Exam Topic 3)
You have a Python script that executes a pipeline. The script includes the following code:
from azureml.core import Experiment
pipeline_run = Experiment(ws, 'pipeline_test').submit(pipeline) You want to test the pipeline before deploying the script.
You need to display the pipeline run details written to the STDOUT output when the pipeline completes. Which code segment should you add to the test script?

A. pipeline_run.get.metrics()
B. pipeline_run.wait_for_completion(show_output=True)
C. pipeline_param = PipelineParameter(name="stdout", default_value="console")
D. pipeline_run.get_status()

**Answer:** B

**Explanation:**
wait_for_completion: Wait for the completion of this run. Returns the status object after the wait. Syntax: wait_for_completion(show_output=False, wait_post_processing=False, raise_on_error=True) Parameter: show_output
Indicates whether to show the run output on sys.stdout.

## NEW QUESTION 82
- (Exam Topic 3)
You have a Python data frame named salesData in the following format:

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

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

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

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

Answer Area

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

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 84
- (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 85
- (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.

| Option | Value |
|--------|-------|
| Technique | ⌄ |
| | Grid search |
| | Outlier detection |
| | Dimensionality reduction |
| Constraint | ⌄ |
| | Demographic parity |
| | False-positive rate parity |

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 86**
- (Exam Topic 3)
You use the following code to define the steps for a pipeline: from azureml.core import Workspace, Experiment, Run from azureml.pipeline.core import Pipeline from azureml.pipeline.steps import PythonScriptStep ws = Workspace.from_config()
. . .
step1 = PythonScriptStep(name="step1", ...) step2 = PythonScriptsStep(name="step2", ...) pipeline_steps = [step1, step2]
You need to add code to run the steps.
Which two code segments can you use to achieve this goal? Each correct answer presents a complete solution.
NOTE: Each correct selection is worth one point.

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

**Answer:** CD

**Explanation:**
After you define your steps, you build the pipeline by using some or all of those steps.
# Build the pipeline. Example:
pipeline1 = Pipeline(workspace=ws, steps=[compare_models])
# Submit the pipeline to be run
pipeline_run1 = Experiment(ws, 'Compare_Models_Exp').submit(pipeline1) Reference:
https://docs.microsoft.com/en-us/azure/machine-learning/how-to-create-machine-learning-pipelines
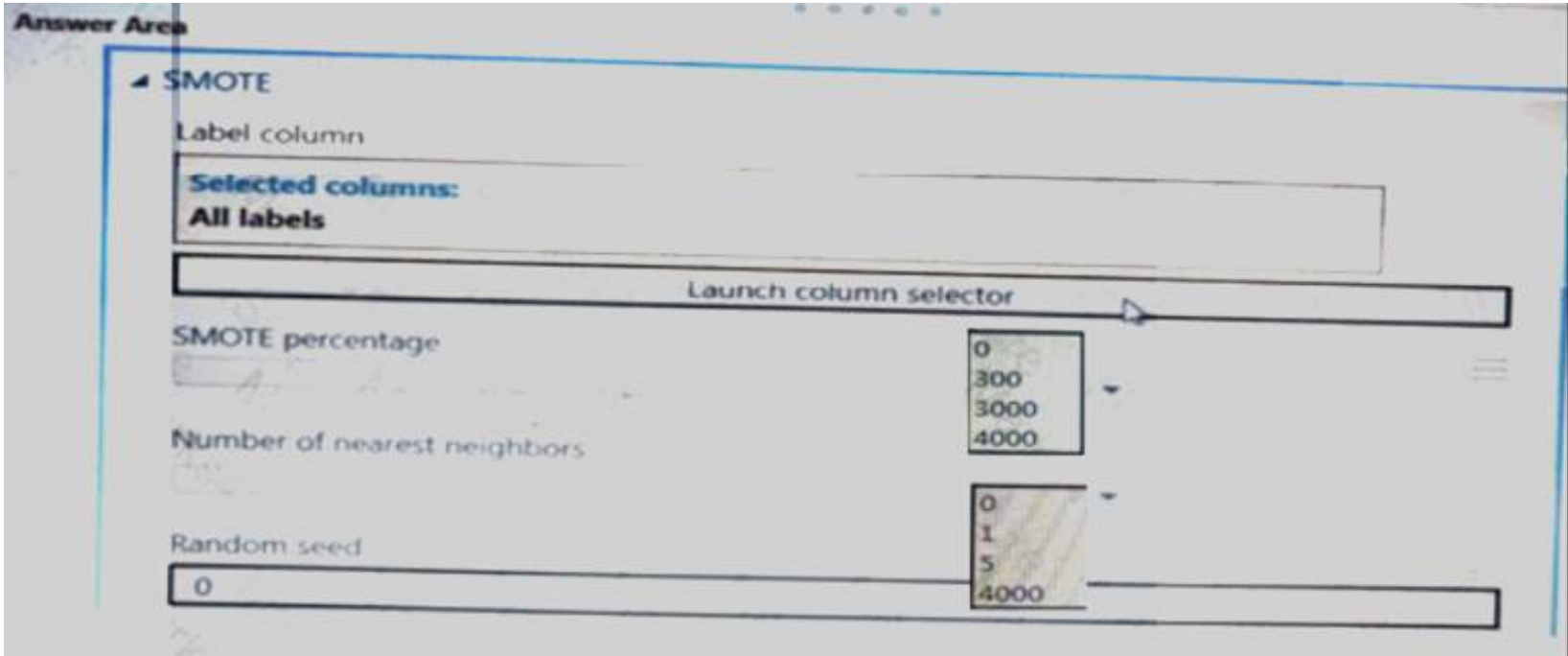
**NEW QUESTION 89**
- (Exam Topic 3)
You create an experiment in Azure Machine Learning Studio- You add a training dataset that contains 10.000 rows. The first 9.000 rows represent class 0 (90 percent). The first 1.000 rows represent class 1 (10 percent).
The training set is unbalanced between two Classes. You must increase the number of training examples for class 1 to 4,000 by using data rows. You add the Synthetic Minority Oversampling Technique (SMOTE) module to the experiment.
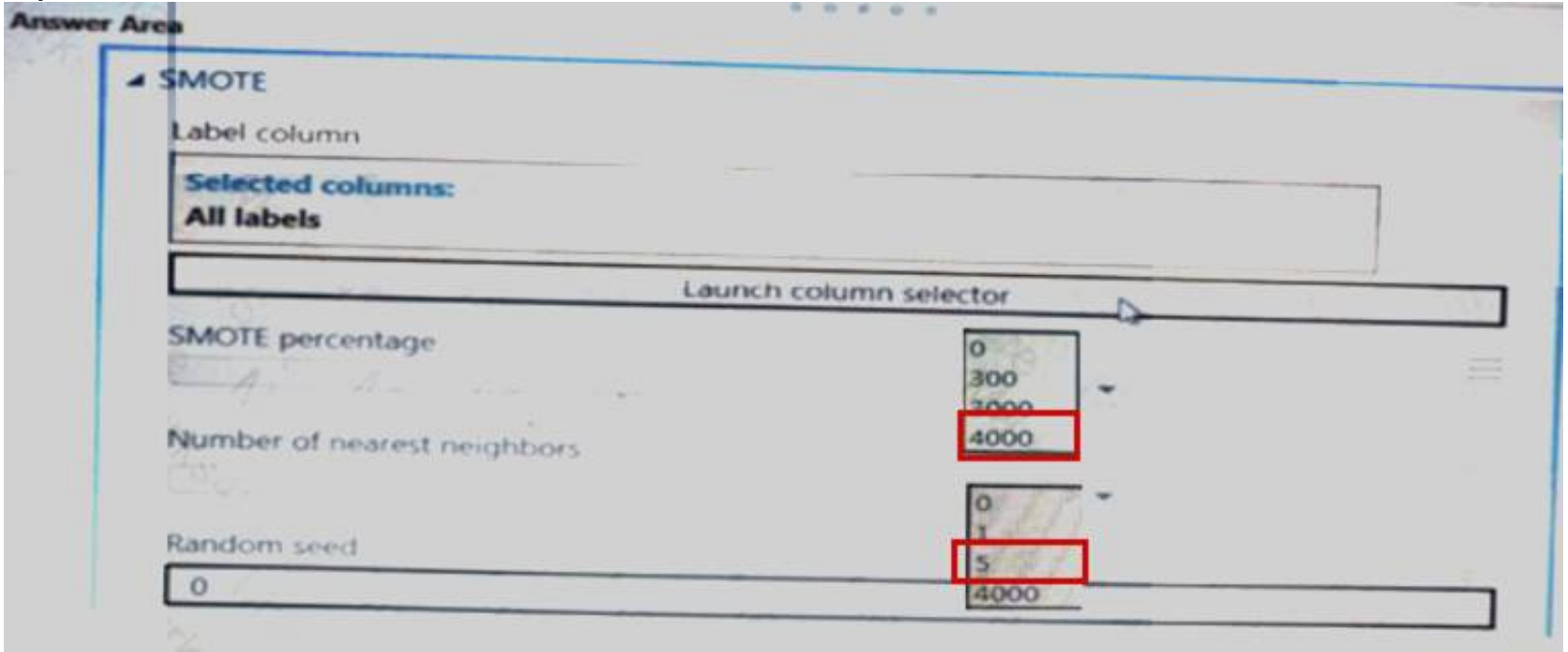You need to configure the module.
Which values should you use? To answer, select the appropriate options in the dialog box in the answer area. NOTE: Each correct selection is worth one point.

A. Mastered
B. Not Mastered
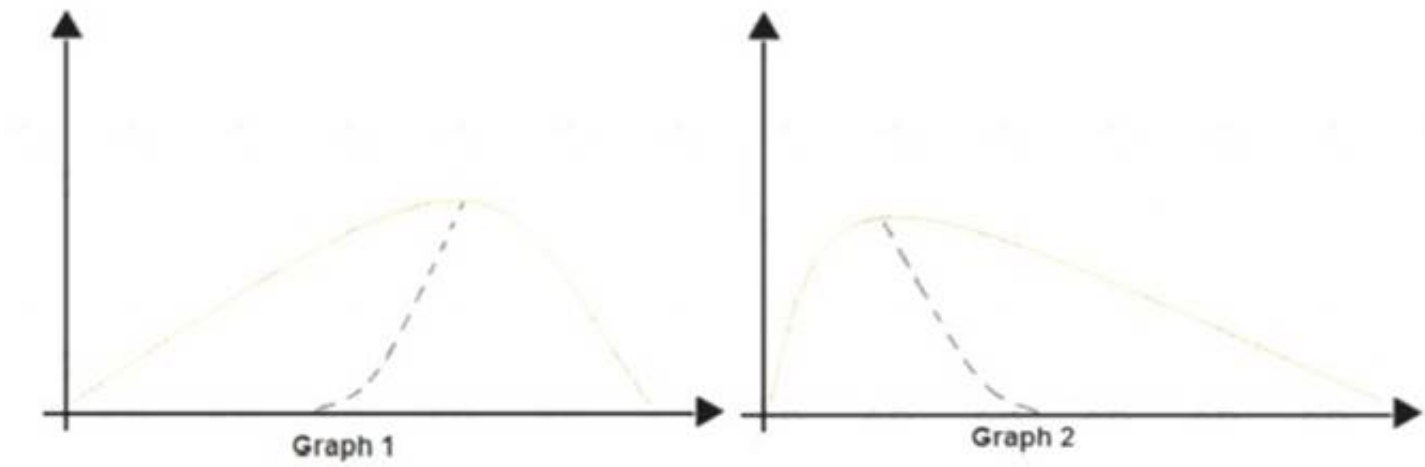
**Answer:** A

**Explanation:**



**NEW QUESTION 93**
- (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? | ▼ <br> Negative skew <br> Positive skew <br> Normal distribution <br> Bimodal distribution |
| Which type of distribution is shown for the dataset density curve of Graph 2? | ▼ <br> Negative skew <br> Positive skew <br> Normal distribution <br> 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 98**
- (Exam Topic 3)
You are evaluating a Python NumPy array that contains six data points defined as follows: data = [10, 20, 30, 40, 50, 60]
You must generate the following output by using the k-fold algorithm implantation in the Python Scikit-learn machine learning library:
train: [10 40 50 60], test: [20 30]
train: [20 30 40 60], test: [10 50]
train: [10 20 30 50], test: [40 60]
You need to implement a cross-validation to generate the output.
How should you complete the code segment? To answer, select the appropriate code segment in the dialog box in the answer area.
NOTE: Each correct selection is worth one point.

```
from numpy import array
from sklearn.model_selection import  ▼
                                       K-Means
                                       k-fold
                                       CrossValidation
                                       ModelSelection

data - array([10, 20, 30, 40, 50, 60])
kfold - Kfold(n_splits-  ▼      ,shuffle - True, random_state-1)
                         1
                         2
                         3
                         6

for train, test in kFold, split(    ▼   ):
                                   data
                                   k-fold
                                   array
                                   train, test

print('train: %s, test: %5' % (data[train], data[test])
```

A. Mastered
B. Not Mastered

**Answer:** A

**Explanation:**
Box 1: k-fold
Box 2: 3
K-F olds cross-validator provides train/test indices to split data in train/test sets. Split dataset into k consecutive folds (without shuffling by default).
The parameter n_splits ( int, default=3) is the number of folds. Must be at least 2. Box 3: data
Example: Example:
>>>
>>> from sklearn.model_selection import KFold
>>> X = np.array([[1, 2], [3, 4], [1, 2], [3, 4]])
>>> y = np.array([1, 2, 3, 4])
>>> kf = KFold(n_splits=2)
>>> kf.get_n_splits(X) 2

```
>>> print(kf)
KFold(n_splits=2, random_state=None, shuffle=False)
>>> for train_index, test_index in kf.split(X): print("TRAIN:", train_index, "TEST:", test_index) X_train, X_test = X[train_index], X[test_index] y_train, y_test =
y[train_index], y[test_index] TRAIN: [2 3] TEST: [0 1]
TRAIN: [0 1] TEST: [2 3]
```

References:
https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.KFold.html

**NEW QUESTION 100**
- (Exam Topic 3)
You train and register a model in your Azure Machine Learning workspace.
You must publish a pipeline that enables client applications to use the model for batch inferencing. You must use a pipeline with a single ParallelRunStep step that runs a Python inferencing script to get predictions from the input data.
You need to create the inferencing script for the ParallelRunStep pipeline step.
Which two functions should you include? Each correct answer presents part of the solution.
NOTE: Each correct selection is worth one point.

A. run(mini_batch) D
B. main()
C. batch()
D. init()
E. score(mini_batch)

**Answer:** AD

**Explanation:**
Reference:
https://github.com/Azure/MachineLearningNotebooks/tree/master/how-to-use-azureml/machine-learningpipelin

**NEW QUESTION 105**
- (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=1
B. k=10
C. k=0.5
D. k=0.9

**Answer:** B

**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 109**
- (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 112**
- (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 117**
- (Exam Topic 3)
You have an Azure Machine Learning workspace named workspace1 that is accessible from a public endpoint. The workspace contains an Azure Blob storage datastore named store1 that represents a blob container in an Azure storage account named account1. You configure workspace1 and account1 to be accessible by using private endpoints in the same virtual network.
You must be able to access the contents of store1 by using the Azure Machine Learning SDK for Python. You must be able to preview the contents of store1 by using Azure Machine Learning studio.
You need to configure store1.
What should you do? To answer, select the appropriate options in the answer area.
NOTE: Each correct selection is worth one point.

| Requirement | Action |
| --- | --- |
| Access the contents of store1 by using the Azure Machine Learning SDK for Python. | Set store1 as the default datastore.<br>Disable data validation for store1.<br>Update authentication for store1.<br>Regenerate the keys of account1. |
| Preview the contents of store1 by using Azure Machine Learning studio. | Set store1 as the default datastore.<br>Disable data validation for store1.<br>Update authentication for store1.<br>Regenerate the keys of account1. |

A. Mastered
B. Not Mastered

**Answer:** A

**Explanation:**
Text, table Description automatically generated
Box 1: Regenerate the keys of account1.
Azure Blob Storage support authentication through Account key or SAS token.
To authenticate your access to the underlying storage service, you can provide either your account key, shared access signatures (SAS) tokens, or service principal
Box 2: Update the authentication for store1.
For Azure Machine Learning studio users, several features rely on the ability to read data from a dataset; such as dataset previews, profiles and automated machine learning. For these features to work with storage behind virtual networks, use a workspace managed identity in the studio to allow Azure Machine Learning to access the storage account from outside the virtual network.
Note: Some of the studio's features are disabled by default in a virtual network. To re-enable these features, you must enable managed identity for storage accounts you intend to use in the studio.
The following operations are disabled by default in a virtual network:

> Preview data in the studio.
Reference:
https://docs.microsoft.com/en-us/azure/machine-learning/how-to-access-data

**NEW QUESTION 122**
- (Exam Topic 3)
An organization uses Azure Machine Learning service and wants to expand their use of machine learning. You have the following compute environments. The organization does not want to create another compute
environment.

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

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

**Environments**

| nb_server |
| aks_cluster |
| mlc_cluster |

**Answer Area**

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

A. Mastered
B. Not Mastered

**Answer:** A

**Explanation:**
Box 1: nb_server

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

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

**NEW QUESTION 126**
- (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.

## Modules

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

## Answer Area

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:
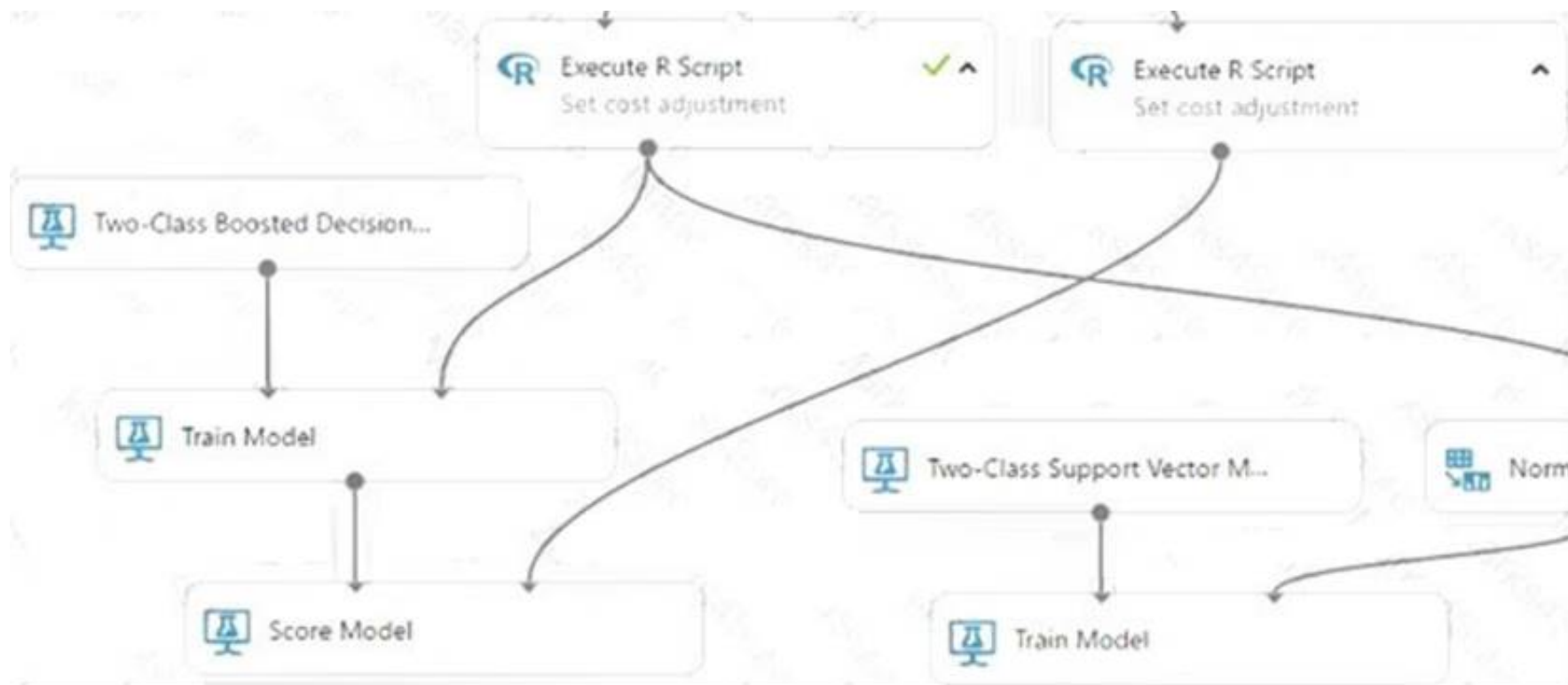


Step 2: Score Model
Score and evaluate the models
You use the testing data that was separated out by the Split Data module to score our trained models. You can then compare the results of the two models to see which generated better results.
Add the Score Model modules
* 1. Find the Score Model module 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.
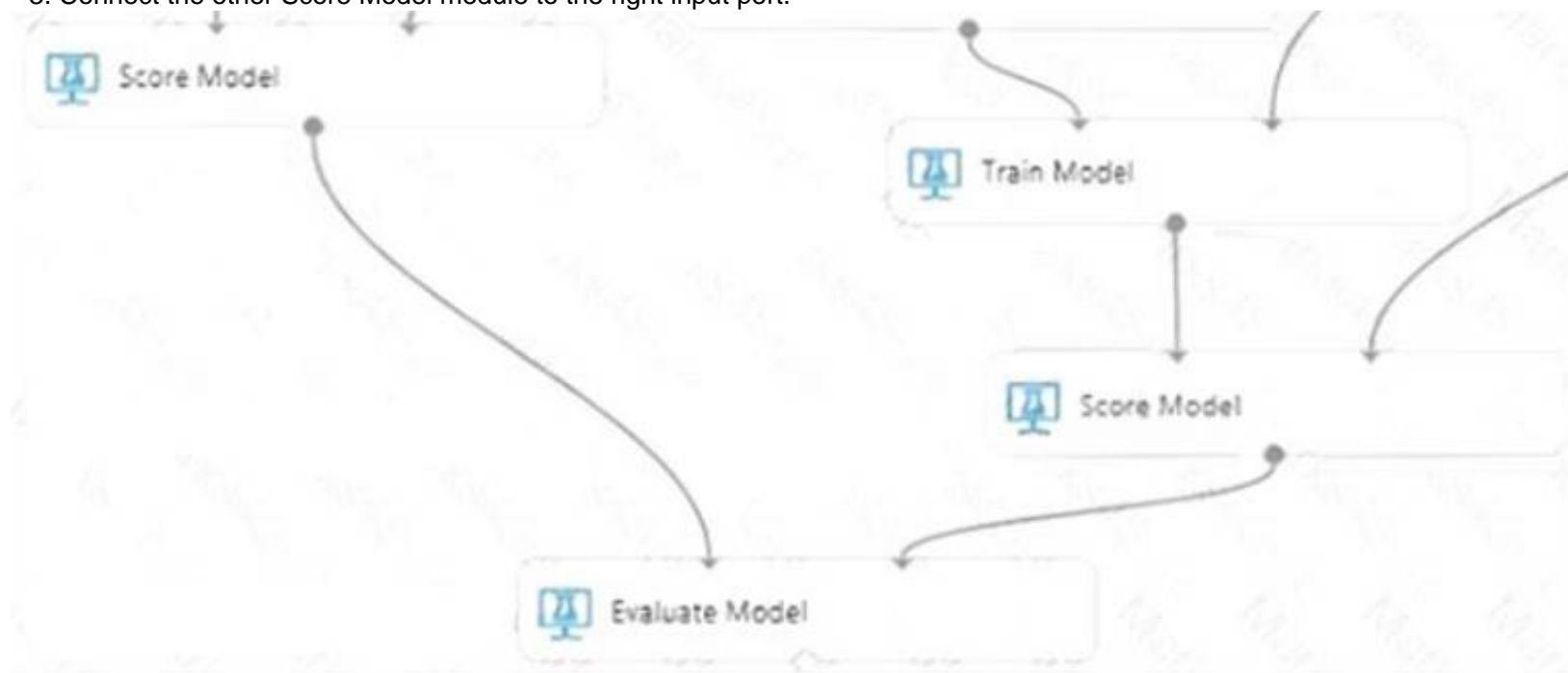* 3. Connect the right Execute R Script module (our testing data) to the right input port of the Score Model module.

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 127**
- (Exam Topic 3)
You create a new Azure Databricks workspace.
You configure a new cluster for long-running tasks with mixed loads on the compute cluster as shown in the image below.

**Microsoft Azure**

Create Cluster

New Cluster    Cancel    Create Cluster    2-8 Workers: 28.0-112.0 GB Memory, 8-32 Cores, 1.5-6 DBU
1 Driver: 14.0 GB Memory, 4 Cores, 0.75 DBU

Azure Databricks

Home

Cluster Name
mysparkcluster

Workspace

Cluster Mode
Standard

Recents

Pool
None

Data

Databricks Runtime Version    Learn more
Runtime: 6.4 (Scala 2.11, Spark 2.4.5)

Clusters

New This Runtime version supports only Python 3.

Autopilot Options
☑ Enable autoscaling
☑ Terminate after 120 minutes of inactivity

Jobs

Worker Type    Min Workers    Max Workers
Standard_DS3_v2    14.0 GB Memory, 4 Cores, 0.75 DBU    2    8

Models

Driver Type
Same as worker    14.0 GB Memory, 4 Cores, 0.75 DBU

Search

▸ Advanced Options

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

Code for each user runs as a separate process    ▼
Yes
No

The number of workers is fixed for the entire duration of the job    ▼
Yes
No

A. Mastered
B. Not Mastered

**Answer:** A

**Explanation:**
Box 1: No
Running user code in separate processes is not possible in Scala. Box 2: No
Autoscaling is enabled. Minimum 2 workers, Maximum 8 workers. Reference:
https://docs.databricks.com/clusters/configure.html

**NEW QUESTION 132**
- (Exam Topic 3)

```
source_directory=scripts_folder,
entry_script="batch_pipeline.py",
mini_batch_size="5",
error_threshold=10,
output_action="append_row",
environment=batch_env,
compute_target=compute_target,
logging_level="DEBUG",
node_count=4)
```

You need to obtain the output from the pipeline execution. Where will you find the output?

A. the Activity Log in the Azure portal for the Machine Learning workspace
B. a file named parallel_run_step.txt located in the output folder
C. the digitjdentification.py script
D. the Inference Clusters tab in Machine Learning studio
E. the debug log

**Answer:** B

**Explanation:**
output_action (str): How the output is to be organized. Currently supported values are 'append_row' and 'summary_only'.

> 'append_row' – All values output by run() method invocations will be aggregated into one unique file named parallel_run_step.txt that is created in the output location.

> 'summary_only' Reference:
https://docs.microsoft.com/en-us/python/api/azureml-contrib-pipeline-steps/azureml.contrib.pipeline.steps.parall


**NEW QUESTION 133**
- (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.
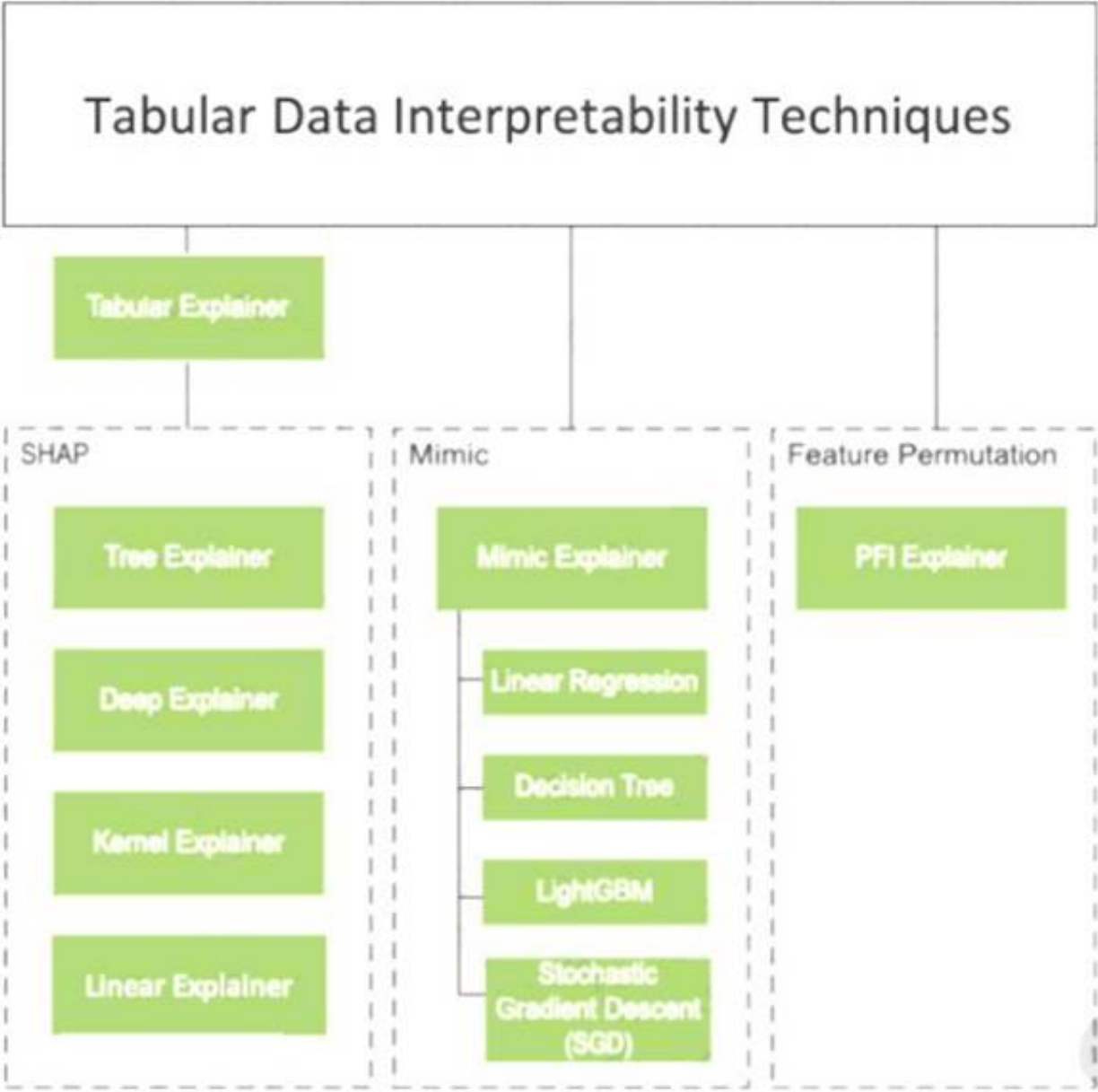Solution: Create a TabularExplainer. Does the solution meet the goal?

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 137**
- (Exam Topic 3)
You create machine learning models by using Azure Machine Learning.
You plan to train and score models by using a variety of compute contexts. You also plan to create a new compute resource in Azure Machine Learning studio.
You need to select the appropriate compute types.
Which compute types should you select? To answer, drag the appropriate compute types to the correct requirements. Each compute type may be used once, more than once, or not at all. You may need to drag the split bar between panes or scroll to view content.
NOTE: Each correct selection is worth one point.



A. Mastered
B. Not Mastered

**Answer:** A

**Explanation:**
Box 1: Attached compute

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

Box 2: Inference cluster Box 3: Training cluster Box 4: Attached compute

## NEW QUESTION 141
- (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 146
- (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 AciWebservice instance.
Set the value of the ssl_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 147
- (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.

◢ **Partition and Sample**

Partition or sample mode

| ▼ |
|---|
| Assign to Folds |
| Pick Fold |
| Sampling |
| Head |

Rate of sampling

| .2 |
|---|

Random seed for sampling

| ▼ |
|---|
| 0 |
| 1 |
| time.clock() |
| utcNow() |

Stratified split for sampling

| False | ⌄ |
|---|---|

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 151**
- (Exam Topic 3)
You use Azure Machine Learning Studio to build a machine learning experiment. You need to divide data into two distinct datasets.
Which module should you use?

A. Split Data
B. Load Trained Model
C. Assign Data to Clusters
D. Group Data into Bins

**Answer:** D

**Explanation:**
The Group Data into Bins module supports multiple options for binning data. You can customize how the bin edges are set and how values are apportioned into the bins.
References:
https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/group-data-into-bins

**NEW QUESTION 156**
- (Exam Topic 3)
You have a comma-separated values (CSV) file containing data from which you want to train a classification model.
You are using the Automated Machine Learning interface in Azure Machine Learning studio to train the classification model. You set the task type to Classification.
You need to ensure that the Automated Machine Learning process evaluates only linear models. What should you do?

A. Add all algorithms other than linear ones to the blocked algorithms list.
B. Set the Exit criterion option to a metric score threshold.
C. Clear the option to perform automatic featurization.
D. Clear the option to enable deep learning.
E. Set the task type to Regression.

**Answer:**

C

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

**NEW QUESTION 159**
- (Exam Topic 3)
You create a binary classification model. You need to evaluate the model performance.
Which two metrics can you use? Each correct answer presents a complete solution.
NOTE: Each correct selection is worth one point.

A. relative absolute error
B. precision
C. accuracy
D. mean absolute error
E. coefficient of determination

**Answer:** BC

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

**NEW QUESTION 163**
- (Exam Topic 3)
You are building a regression model tot estimating the number of calls during an event.
You need to determine whether the feature values achieve the conditions to build a Poisson regression model. Which two conditions must the feature set contain?
I ach correct answer presents part of the solution. NOTE:
Each correct selection is worth one point.

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

**Answer:** CE

**Explanation:**
Poisson regression is intended for use in regression models that are used to predict numeric values, typically counts. Therefore, you should use this module to create your regression model only if the values you are trying to predict fit the following conditions:

➤ The response variable has a Poisson distribution.

➤ Counts cannot be negative. The method will fail outright if you attempt to use it with negative labels.

➤ A Poisson distribution is a discrete distribution; therefore, it is not meaningful to use this method with non-whole numbers.
References:
https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/poisson-regression

**NEW QUESTION 165**
- (Exam Topic 3)
You are performing a filter based feature selection for a dataset 10 build a multi class classifies by using Azure Machine Learning Studio.
The dataset contains categorical features that are highly correlated to the output label column.
You need to select the appropriate feature scoring statistical method to identify the key predictors. Which method should you use?

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

**Answer:** D

**Explanation:**
Pearson's correlation statistic, or Pearson's correlation coefficient, is also known in statistical models as the r value. For any two variables, it returns a value that indicates the strength of the correlation
Pearson's correlation coefficient is the test statistics that measures the statistical relationship, or association, between two continuous variables. It is known as the best method of measuring the association between variables of interest because it is based on the method of covariance. It gives information about the magnitude of the association, or correlation, as well as the direction of the relationship.
Reference:
https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/filter-based-feature-selection https://www.statisticssolutions.com/pearsons-correlation-coefficient/

**NEW QUESTION 167**
- (Exam Topic 3)
You create an Azure Machine Learning compute target named ComputeOne by using the STANDARD_D1 virtual machine image.
You define a Python variable named was that references the Azure Machine Learning workspace. You run the following Python code:

```
from azureml.core.compute import ComputeTarget, AmlCompute
from azureml.core.compute_target import ComputeTargetException
the_cluster_name = "ComputeOne"
try:
    the_cluster = ComputeTarget(workspace=ws, name=the_cluster_name)
    print('Step1')
except ComputeTargetException:
    config = AmlCompute.provisioning_configuration(vm_size='STANDARD_DS12_v2', max_nodes=4)
    the_cluster = ComputeTarget.create(ws, the_cluster_name, config)
    print('Step2')
```

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 |
|---|---|---|
| A new machine learning compute resource is created with a virtual machine size of STANDARD_DS12_v2 and a maximum of four nodes. | O | O |
| Any experiments configured to use `the_cluster` will run on ComputeOne. | O | O |
| The text **Step1** will be printed to the screen. | O | O |

A. Mastered
B. Not Mastered

**Answer:** A

**Explanation:**
Box 1:Yes
ComputeTargetException class: An exception related to failures when creating, interacting with, or configuring a compute target. This exception is commonly raised for failures attaching a compute target,
missing headers, and unsupported configuration values. Create(workspace, name, provisioning_configuration)
Provision a Compute object by specifying a compute type and related configuration. This method creates a new compute target rather than attaching an existing one. Box 2: Yes
Box 3: No
The line before print('Step1') will fail. Reference:
https://docs.microsoft.com/en-us/python/api/azureml-core/azureml.core.compute.computetarget

**NEW QUESTION 172**
- (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 174**
- (Exam Topic 3)
You plan to run a script as an experiment using a Script Run Configuration. The script uses modules from the scipy library as well as several Python packages that are not typically installed in a default conda environment
You plan to run the experiment on your local workstation for small datasets and scale out the experiment by running it on more powerful remote compute clusters for larger datasets.
You need to ensure that the experiment runs successfully on local and remote compute with the least administrative effort.
What should you do?

A. Create and register an Environment that includes the required package
B. Use this Environment for all experiment runs.

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

**Answer:** A

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

**NEW QUESTION 175**
- (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 177**
- (Exam Topic 3)
Note: This question is part of a series of questions that present the same scenario. Each question in the series contains a unique solution that might meet the stated goals. Some question sets might have more than one correct solution, while others might not have a correct solution.
After you answer a question in this section, you will NOT be able to return to it. As a result, these questions will not appear in the review screen.
You are a data scientist using Azure Machine Learning Studio.
You need to normalize values to produce an output column into bins to predict a target column. Solution: Apply a Quantiles binning mode with a PQuantile normalization.
Does the solution meet the goal?

A. Yes
B. No

**Answer:** B

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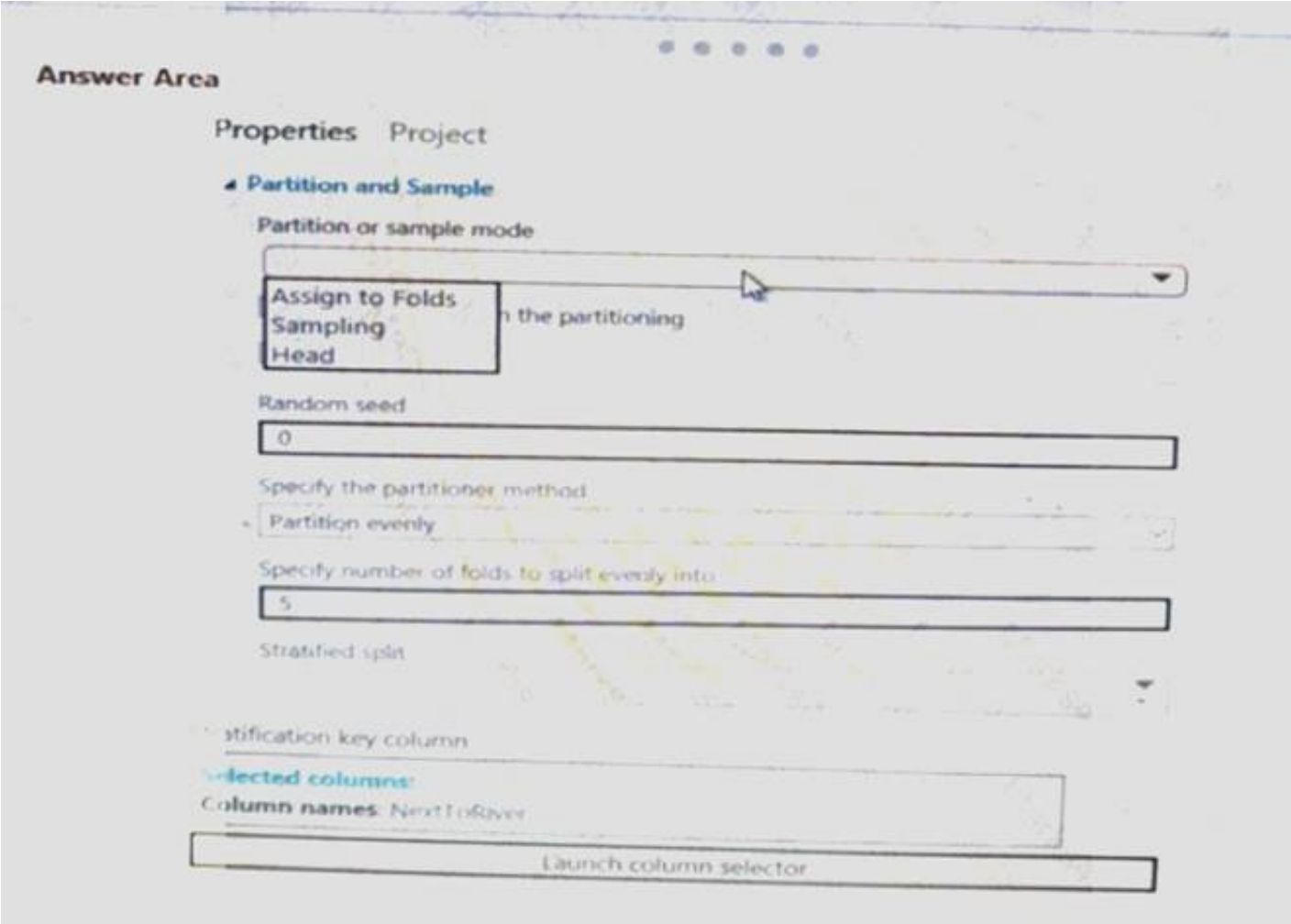Use the Entropy MDL binning mode which has a target column. References:
https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/group-data-into-bins

**NEW QUESTION 182**
- (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 184**
- (Exam Topic 2)
You need to configure the Edit Metadata module so that the structure of the datasets match.
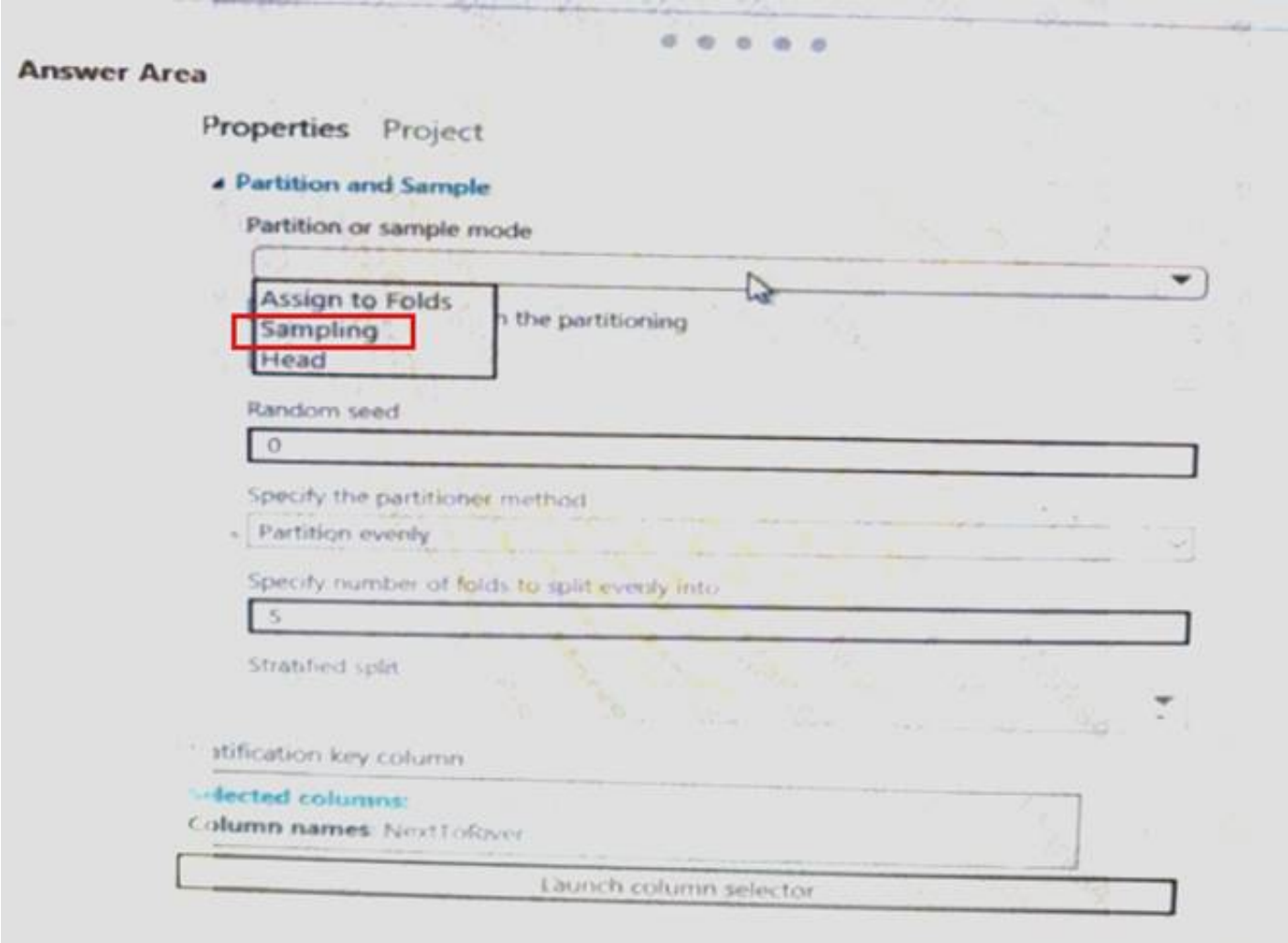Which configuration options should you select? To answer, select the appropriate options in the answer area.
NOTE: Each correct selection is worth one point.

Properties      Project

◢ **Edit Metadata**

Column

Selected columns:
Column names: MedianValue

Launch column selector

▼
Floating point
DateTime
TimeSpan
Integer

▼
Unchanged
Make Categorical
Make Uncategorical

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 186**
- (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 189**
- (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 190**
- (Exam Topic 2)

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

**Code segments**

```
early_termination_policy =
TruncationSelectionPolicy(evaluation_interval=1,
truncation_percentage=20, delay_evaluation=5)
```

```
import TruncationSelectionPolicy
```

```
from azureml.train.hyperdrive
```

```
import BanditPolicy
```

```
early_termination_policy = BanditPolicy
(slack_factor = 0.1, evaluation_interval=1,
delay_evaluation=5)
```

**Answer Area**

A. Mastered
B. Not Mastered

**Answer:** A

**Explanation:**
You need to implement an early stopping criterion on models that provides savings without terminating promising jobs.
Truncation selection cancels a given percentage of lowest performing runs at each evaluation interval. Runs are compared based on their performance on the primary metric and the lowest X% are terminated. Example:
from azureml.train.hyperdrive import TruncationSelectionPolicy
early_termination_policy = TruncationSelectionPolicy(evaluation_interval=1, truncation_percentage=20, delay_evaluation=5)

**NEW QUESTION 195**
- (Exam Topic 1)
You need to define an evaluation strategy for the crowd sentiment models.
Which three actions should you perform in sequence? To answer, move the appropriate actions from the list of actions to the answer area and arrange them in the correct order.

**Actions**

- Add new features for retraining supervised models.
- Filter labeled cases for retraining using the shortest distance from centroids.
- Evaluate the changes in correlation between model error rate and centroid distance
- Impute unavailable features with centroid aligned models
- Filter labeled cases for retraining using the longest distance from centroids.
- Remove features before retraining supervised models.

**Answer Area**

A. Mastered
B. Not Mastered

**Answer:** A

**Explanation:**
Scenario:
Experiments for local crowd sentiment models must combine local penalty detection data.

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


## NEW QUESTION 197
- (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 198
- (Exam Topic 1)
You need to define a process for penalty event detection.
Which three actions should you perform in sequence? To answer, move the appropriate actions from the list of actions to the answer area and arrange them in the correct order.
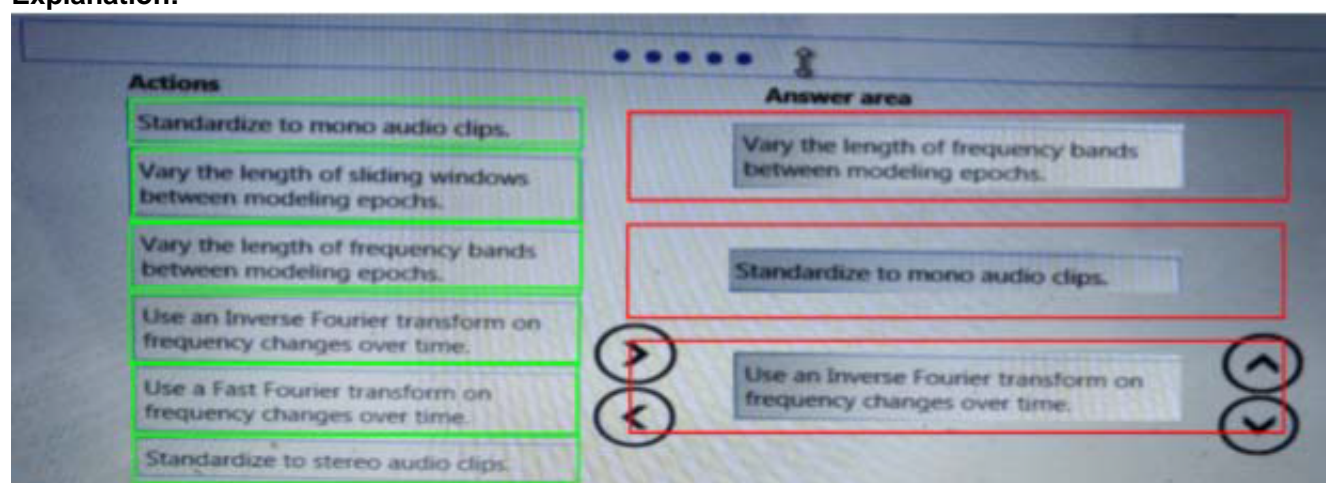
A. Mastered
B. Not Mastered

**Answer:** A

**Explanation:**



**NEW QUESTION 201**
- (Exam Topic 1)
You need to implement a new cost factor scenario for the ad response models as illustrated in the performance curve exhibit.
Which technique should you use?

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

**Answer:** A

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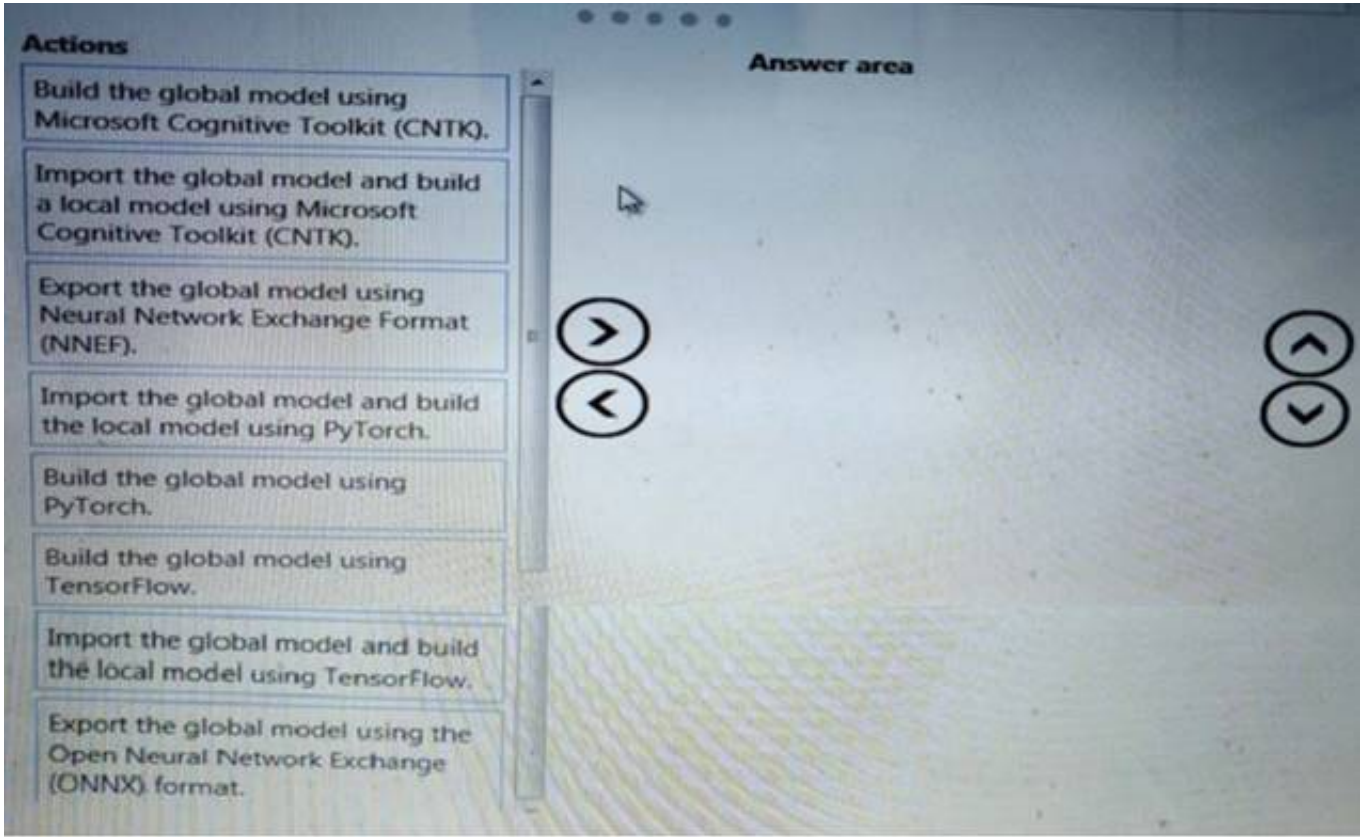


The ad propensity model uses a cut threshold is 0.45 and retrains occur if weighted Kappa deviated from 0.1 +/- 5%.

**NEW QUESTION 204**
- (Exam Topic 1)
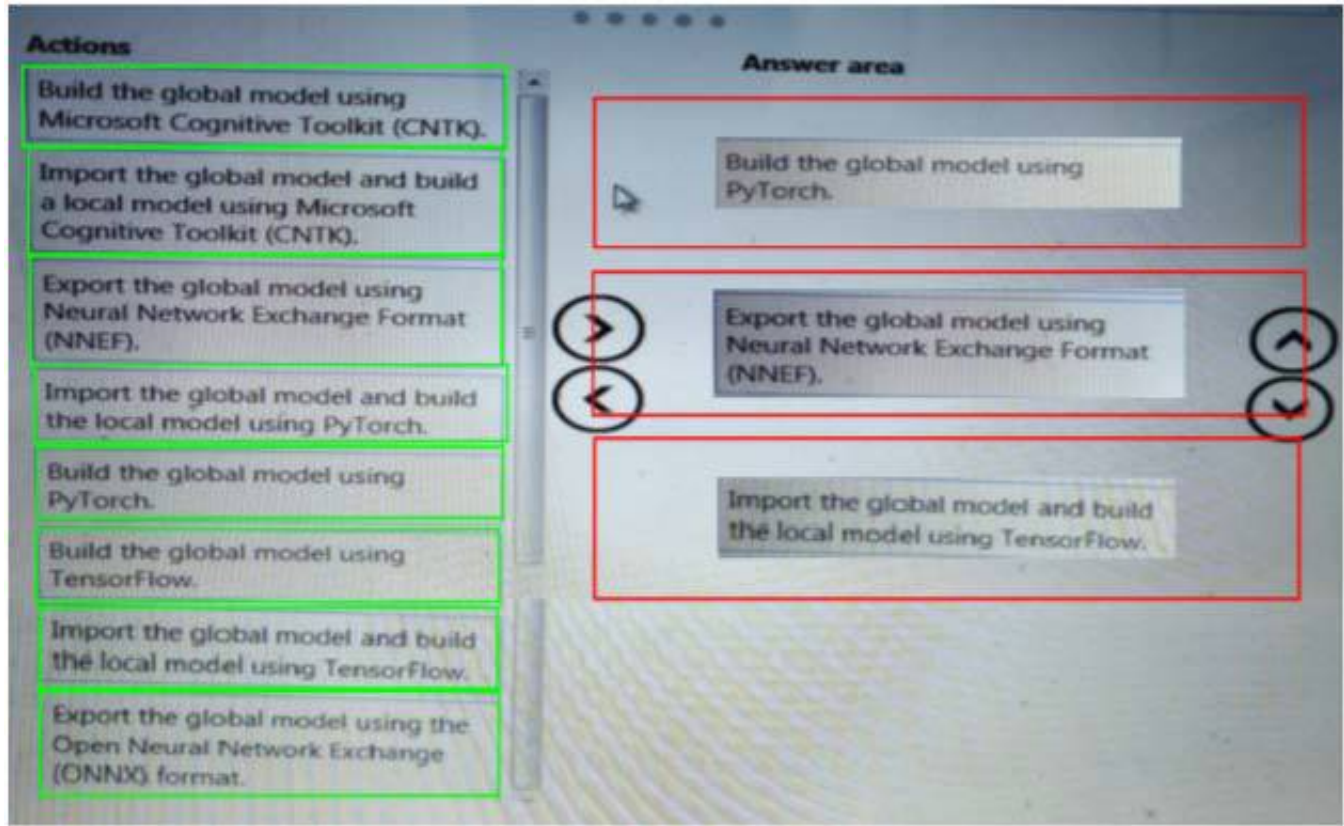You need to define a process for penalty event detection.
Which three actions should you perform in sequence? To answer, move the appropriate actions from the list of actions to the answer area and arrange them in the correct order.

A. Mastered
B. Not Mastered

**Answer:** A

**Explanation:**



**NEW QUESTION 209**
- (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.

```
import pytorch as deeplearninglib
import tensorflow as deeplearninglib
import cntk as deeplearninglib
```

```
train_smapler = deeplearminglib.DistributedSampler.(penalty_video_dataset)
train_sampler = deeplearminglib.log_uniform_candidate_sampler.(penalty_video_dataset)
train_sampler = deeplearminglib.WeightedRandomSampler.(penalty_video_dataset)
train_sampler = deeplearminglib.all_candidate_sampler.(penalty_video_dataset)
```

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

```
optimizer = deeplearninglib.optim.SGD(model.parameters().lr=0,01)
optimizer = deeplearninglib.train.GradientDescentOptimizer(learning_rate=0.10)
```

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

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

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 210**
......

# Thank You for Trying Our Product

## We offer two products:

1st - We have Practice Tests Software with Actual Exam Questions

2nd - Questons and Answers in PDF Format

## 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

## 100% Actual & Verified — Instant Download, Please Click
[Order The DP-100 Practice Test Here](https://www.surepassexam.com/DP-100-exam-dumps.html)