

## Exam Questions DP-100

Designing and Implementing a Data Science Solution on Azure

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



NEW QUESTION 1

- (Exam Topic 3)

You have an Azure Machine Learning workspace that contains a training cluster and an inference cluster. You plan to create a classification model by using the Azure Machine Learning designer. You need to ensure that client applications can submit data as HTTP requests and receive predictions as responses. 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

Create a real-time inference pipeline and run the pipeline on the compute cluster.

Create a batch inference pipeline and run the pipeline on the compute cluster.

Deploy a service to the compute cluster.

Create a pipeline that trains a classification model and run the pipeline on the compute cluster.

Deploy a service to the inference cluster.

Answer area

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

Actions

Create a real-time inference pipeline and run the pipeline on the compute cluster.

Create a batch inference pipeline and run the pipeline on the compute cluster.

Deploy a service to the compute cluster.

Create a pipeline that trains a classification model and run the pipeline on the compute cluster.

Deploy a service to the inference cluster.

Answer area

Create a pipeline that trains a classification model and run the pipeline on the compute cluster.

Create a batch inference pipeline and run the pipeline on the compute cluster.

Deploy a service to the inference cluster.

NEW QUESTION 2

- (Exam Topic 3)

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

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

Does the solution meet the goal?

- A. Yes
- B. No

Answer: B

**Explanation:**

train\_step is missing. Reference:

<https://docs.microsoft.com/en-us/python/api/azureml-pipeline-core/azureml.pipeline.core.pipelinedata?view=azu>

**NEW QUESTION 3**

- (Exam Topic 3)

Note: This question is part of a series of questions that present the same scenario. Each question in the series contains a unique solution that might meet the stated goals. Some question sets might have more than one correct solution, while others might not have a correct solution.

After you answer a question in this section, you will NOT be able to return to it. As a result, these questions will not appear in the review screen.

You are creating a new experiment in Azure Machine Learning Studio.

One class has a much smaller number of observations than the other classes in the training set. You need to select an appropriate data sampling strategy to compensate for the class imbalance. Solution: You use the Stratified split for the sampling mode.

Does the solution meet the goal?

A. Yes

B. No

**Answer: B**

**Explanation:**

Instead use the Synthetic Minority Oversampling Technique (SMOTE) sampling mode.

Note: SMOTE is used to increase the number of underrepresented cases in a dataset used for machine learning. SMOTE is a better way of increasing the number of rare cases than simply duplicating existing cases.

References:

<https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/sMOTE>

**NEW QUESTION 4**

- (Exam Topic 3)

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

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

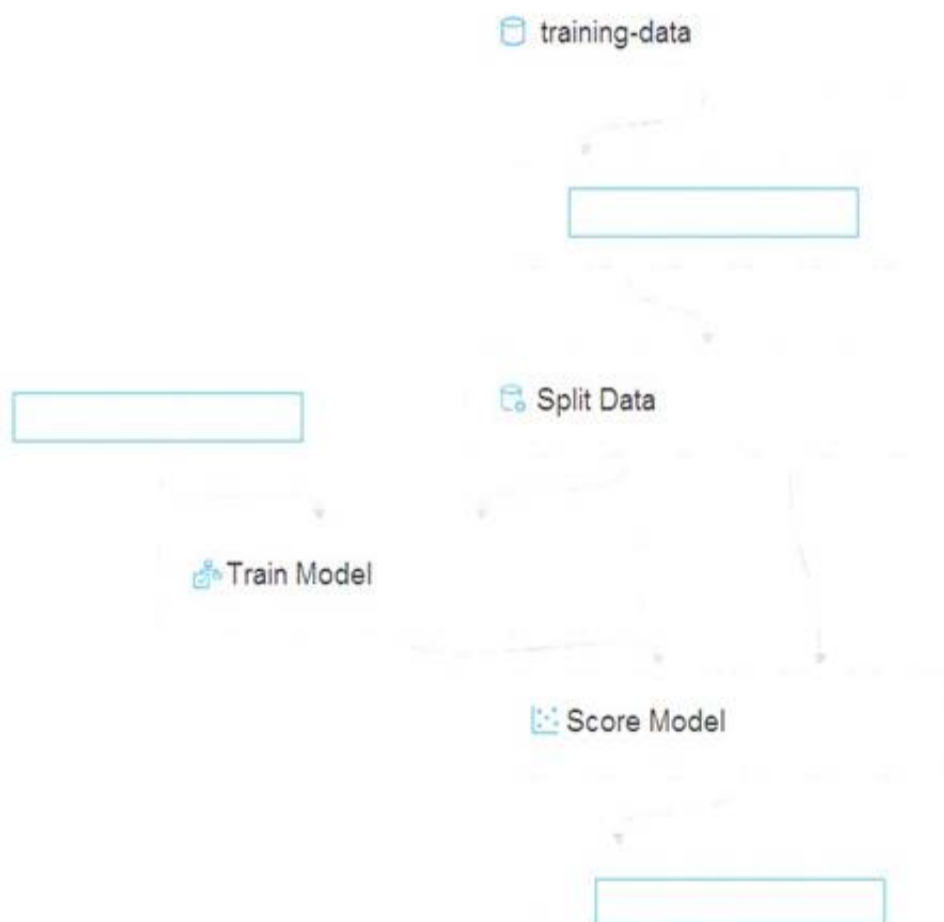
- Select the training features using the pandas filter method.
- Train a model based on the naive\_bayes.GaussianNB algorithm.
- Return only the Scored Labels column by using the query SELECT [Scored Labels] FROM t1; Which modules should you use? To answer, drag the appropriate modules to the appropriate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

NOTE: Each correct selection is worth one point.

**Modules**

**Answer Area**

- Create Python Model
- Train Model
- Two Class Neural Network
- Execute Python Script
- Apply SQL Transformation
- Select Columns in Dataset

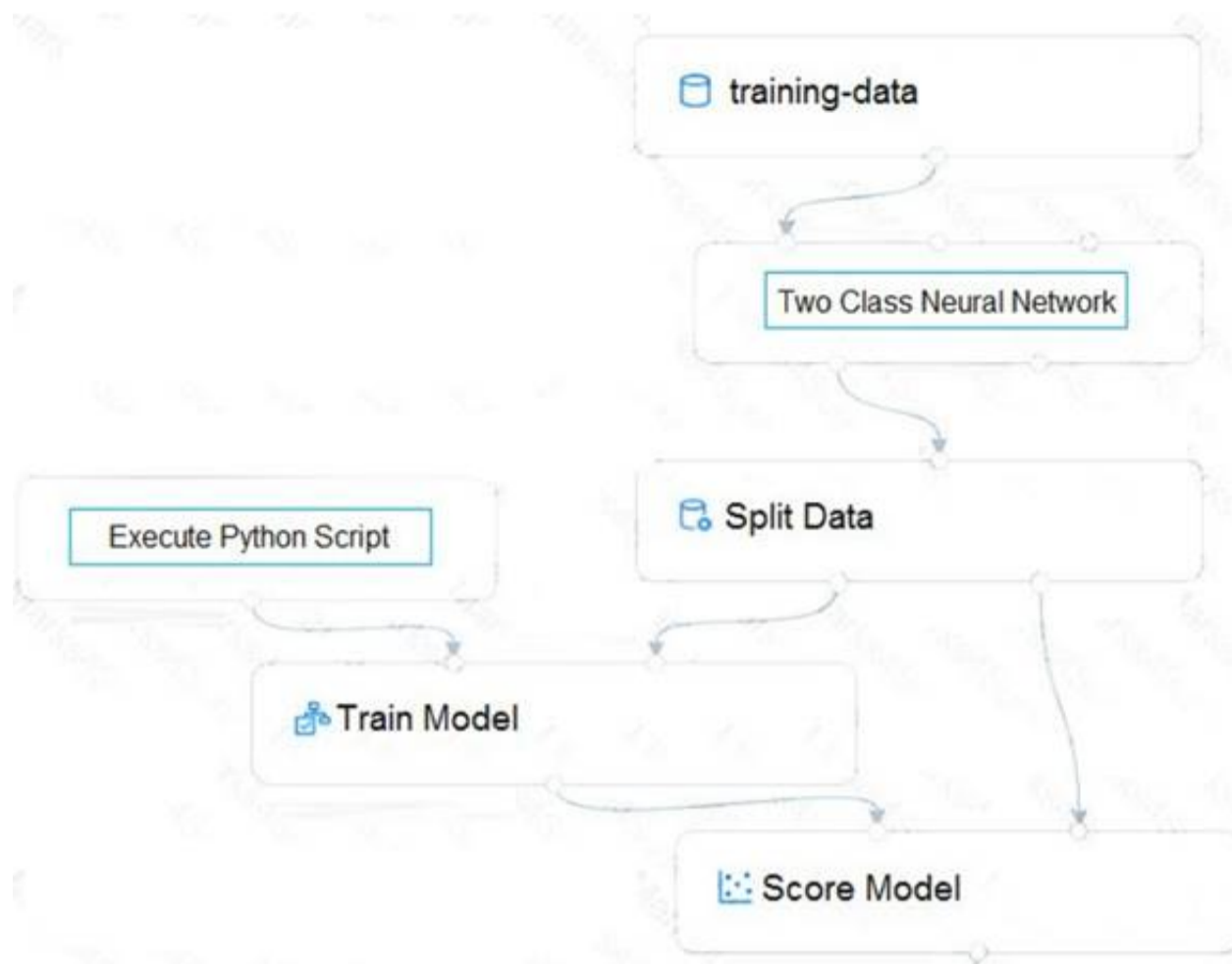


A. Mastered

B. Not Mastered

**Answer: A**

**Explanation:**



### NEW QUESTION 5

- (Exam Topic 3)

Your Azure Machine Learning workspace has a dataset named `real_estate_data`. A sample of the data in the dataset follows.

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

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

NOTE: Each correct selection is worth one point.

#### Answer Area

```

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

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






```

- A. Mastered
- B. Not Mastered

**Answer:** A

**Explanation:**

Box 1: training\_data

The training data to be used within the experiment. It should contain both training features and a label column (optionally a sample weights column). If training\_data is specified, then the label\_column\_name parameter must also be specified.

Box 2: validation\_data

Provide validation data: In this case, you can either start with a single data file and split it into training and validation sets or you can provide a separate data file for the validation set. Either way, the validation\_data parameter in your AutoMLConfig object assigns which data to use as your validation set.

Example, the following code example explicitly defines which portion of the provided data in dataset to use for training and validation.

```
dataset = Dataset.Tabular.from_delimited_files(data)
```

```
training_data, validation_data = dataset.random_split(percentage=0.8, seed=1) automl_config = AutoMLConfig(compute_target = aml_remote_compute, task = 'classification',
```

```
primary_metric = 'AUC_weighted', training_data = training_data,
```

```
validation_data = validation_data, label_column_name = 'Class'
```

```
)
```

Box 3: label\_column\_name label\_column\_name:

The name of the label column. If the input data is from a pandas.DataFrame which doesn't have column names, column indices can be used instead, expressed as integers.

This parameter is applicable to training\_data and validation\_data parameters. Reference:

<https://docs.microsoft.com/en-us/python/api/azureml-train-automl-client/azureml.train.automl.automlconfig.auto>

**NEW QUESTION 6**

- (Exam Topic 3)

You use the designer to create a training pipeline for a classification model. The pipeline uses a dataset that includes the features and labels required for model training.

You create a real-time inference pipeline from the training pipeline. You observe that the schema for the generated web service input is based on the dataset and includes the label column that the model predicts. Client applications that use the service must not be required to submit this value.

You need to modify the inference pipeline to meet the requirement. What should you do?

- A. Add a Select Columns in Dataset module to the inference pipeline after the dataset and use it to select all columns other than the label.
- B. Delete the dataset from the training pipeline and recreate the real-time inference pipeline.
- C. Delete the Web Service Input module from the inference pipeline.
- D. Replace the dataset in the inference pipeline with an Enter Data Manually module that includes data for the feature columns but not the label column.

**Answer:** A

**Explanation:**

By default, the Web Service Input will expect the same data schema as the module output data which connects to the same downstream port as it. You can remove the target variable column in the inference pipeline using Select Columns in Dataset module. Make sure that the output of Select Columns in Dataset removing target variable column is connected to the same port as the output of the Web Service Input module.

Reference:

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

**NEW QUESTION 7**

- (Exam Topic 3)

You retrain an existing model.

You need to register the new version of a model while keeping the current version of the model in the registry.

What should you do?

- A. Register a model with a different name from the existing model and a custom property named version with the value 2.
- B. Register the model with the same name as the existing model.
- C. Save the new model in the default datastore with the same name as the existing model.
- D. Do not register the new model.
- E. Delete the existing model and register the new one with the same name.

**Answer:** B

**Explanation:**

Model version: A version of a registered model. When a new model is added to the Model Registry, it is added as Version 1. Each model registered to the same model name increments the version number.

Reference:

<https://docs.microsoft.com/en-us/azure/databricks/applications/mlflow/model-registry>

**NEW QUESTION 8**

- (Exam Topic 3) You are solving a classification task. The dataset is imbalanced.

You need to select an Azure Machine Learning Studio module to improve the classification accuracy. Which module should you use?

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

**Answer:** C

**Explanation:**

Use the SMOTE module in Azure Machine Learning Studio (classic) to increase the number of underrepresented cases in a dataset used for machine learning.

SMOTE is a better way of increasing the number of rare cases than simply duplicating existing cases.

You connect the SMOTE module to a dataset that is imbalanced. There are many reasons why a dataset might be imbalanced: the category you are targeting might be very rare in the population, or the data might simply be difficult to collect. Typically, you use SMOTE when the class you want to analyze is under-represented.

Reference:

<https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/smote>

### NEW QUESTION 9

- (Exam Topic 3)

You are a lead data scientist for a project that tracks the health and migration of birds. You create a multi-image classification deep learning model that uses a set of labeled bird photos collected by experts. You plan to use the model to develop a cross-platform mobile app that predicts the species of bird captured by app users.

You must test and deploy the trained model as a web service. The deployed model must meet the following requirements:

- > An authenticated connection must not be required for testing.
- > The deployed model must perform with low latency during inferencing.
- > The REST endpoints must be scalable and should have a capacity to handle large number of requests when multiple end users are using the mobile application.

You need to verify that the web service returns predictions in the expected JSON format when a valid REST request is submitted.

Which compute resources should you use? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.

### Context

### Resource

Test

	▼
ds-workstation notebook VM	
aks-compute cluster	
cpu-compute cluster	
gpu-compute cluster	

Production

	▼
ds-workstation notebook VM	
aks-compute cluster	
cpu-compute cluster	
gpu-compute cluster	

- A. Mastered
- B. Not Mastered

**Answer:** A

#### Explanation:

Box 1: ds-workstation notebook VM

An authenticated connection must not be required for testing.

On a Microsoft Azure virtual machine (VM), including a Data Science Virtual Machine (DSVM), you create local user accounts while provisioning the VM. Users then authenticate to the VM by using these credentials.

Box 2: gpu-compute cluster

Image classification is well suited for GPU compute clusters

Reference:

<https://docs.microsoft.com/en-us/azure/machine-learning/data-science-virtual-machine/dsvm-common-identity> <https://docs.microsoft.com/en-us/azure/architecture/reference-architectures/ai/training-deep-learning>

### NEW QUESTION 10

- (Exam Topic 3)

You deploy a model in Azure Container Instance.

You must use the Azure Machine Learning SDK to call the model API.

You need to invoke the deployed model using native SDK classes and methods.

How should you complete the command? To answer, select the appropriate options in the answer areas.

NOTE: Each correct selection is worth one point.

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

```
from azureml.core.webservice import requests
from azureml.core.webservice import Webservice
from azureml.core.webservice import LocalWebservice
```

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

```
predictions = service.run(input_json)
predictions = requests.post(service.scoring_uri, input_json)
predictions = service.deserialize(ws, input_json)
```

- A. Mastered  
 B. Not Mastered

**Answer:** A

**Explanation:**

Box 1: from azureml.core.webservice import Webservice

The following code shows how to use the SDK to update the model, environment, and entry script for a web service to Azure Container Instances:

```
from azureml.core import Environment
```

```
from azureml.core.webservice import Webservice
```

```
from azureml.core.model import Model, InferenceConfig Box 2: predictions = service.run(input_json)
```

Example: The following code demonstrates sending data to the service: import json

```
test_sample = json.dumps({'data': [ [1, 2, 3, 4, 5, 6, 7, 8, 9, 10],
```

```
[10, 9, 8, 7, 6, 5, 4, 3, 2, 1]
```

```
]])
```

```
test_sample = bytes(test_sample, encoding='utf8') prediction = service.run(input_data=test_sample)
```

```
print(prediction) Reference:
```

<https://docs.microsoft.com/bs-latn-ba/azure/machine-learning/how-to-deploy-azure-container-instance> <https://docs.microsoft.com/en-us/azure/machine-learning/how-to-troubleshoot-deployment>

**NEW QUESTION 10**

- (Exam Topic 3)

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

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

For each of the following statements, select Yes if the statement is true. Otherwise, select No.

NOTE: Each correct selection is worth one point.

**Yes**

**No**

Hyperparameter combinations for the runs are selected based on how previous samples performed in the previous experiment run.

☐
☐

The learning rate value 0.09 might be used during model training.

☐
☐

You can define an early termination policy for this hyperparameter tuning run.

☐
☐

- A. Mastered  
 B. Not Mastered

**Answer:** A

**Explanation:**

Box 1: Yes

Hyperparameters are adjustable parameters you choose to train a model that govern the training process itself. Azure Machine Learning allows you to automate hyperparameter exploration in an efficient manner, saving you significant time and resources. You specify the range of hyperparameter values and a maximum number of training runs. The system then automatically launches multiple simultaneous runs with different parameter configurations and finds the configuration that results in the best performance, measured by the metric you choose. Poorly performing training runs are automatically early terminated, reducing wastage of compute resources. These resources are instead used to explore other hyperparameter configurations.

Box 2: Yes

uniform(low, high) - Returns a value uniformly distributed between low and high Box 3: No

Bayesian sampling does not currently support any early termination policy. Reference:

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

### NEW QUESTION 13

- (Exam Topic 3)

You create a batch inference pipeline by using the Azure ML SDK. You run the pipeline by using the following code:

```
from azureml.pipeline.core import Pipeline
```

```
from azureml.core.experiment import Experiment
```

```
pipeline = Pipeline(workspace=ws, steps=[parallelrun_step]) pipeline_run = Experiment(ws, 'batch_pipeline').submit(pipeline)
```

You need to monitor the progress of the pipeline execution.

What are two possible ways to achieve this goal? Each correct answer presents a complete solution.

NOTE: Each correct selection is worth one point.

A. Run the following code in a notebook:

```
from azureml.contrib.interpret.explanation.explanation_client import ExplanationClient
client = ExplanationClient.from_run(pipeline_run)
explanation = client.download_model_explanation()
explanation = client.download_model_explanation(top_k=4)
global_importance_values = explanation.get_ranked_global_values()
global_importance_names = explanation.get_ranked_global_names()
print('global importance values: {}'.format(global_importance_values))
print('global importance names: {}'.format(global_importance_names))
```

B. Use the Inference Clusters tab in Machine Learning Studio.

C. Use the Activity log in the Azure portal for the Machine Learning workspace.

D. Run the following code in a notebook:

```
from azureml.widgets import RunDetails
RunDetails(pipeline_run).show()
```

E. Run the following code and monitor the console output from the PipelineRun object:

```
pipeline_run.wait_for_completion(show_output=True)
```

A. Option A

B. Option B

C. Option C

D. Option D

E. Option E

**Answer:** DE

#### Explanation:

A batch inference job can take a long time to finish. This example monitors progress by using a Jupyter widget. You can also manage the job's progress by using:

➤ Azure Machine Learning Studio.

➤ Console output from the PipelineRun object. `from azureml.widgets import RunDetails RunDetails(pipeline_run).show()`

`pipeline_run.wait_for_completion(show_output=True)` Reference:

<https://docs.microsoft.com/en-us/azure/machine-learning/how-to-use-parallel-run-step#monitor-the-parallel-run>

### NEW QUESTION 18

- (Exam Topic 3)

You are building a machine learning model for translating English language textual content into French language textual content.

You need to build and train the machine learning model to learn the sequence of the textual content. Which type of neural network should you use?

A. Multilayer Perceptions (MLPs)

B. Convolutional Neural Networks (CNNs)

C. Recurrent Neural Networks (RNNs)

D. Generative Adversarial Networks (GANs)

**Answer:** C

#### Explanation:

To translate a corpus of English text to French, we need to build a recurrent neural network (RNN).

Note: RNNs are designed to take sequences of text as inputs or return sequences of text as outputs, or both. They're called recurrent because the network's hidden layers have a loop in which the output and cell state from each time step become inputs at the next time step. This recurrence serves as a form of memory. It allows contextual information to flow through the network so that relevant outputs from previous time steps can be applied to network operations at the current time step.

References:

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

### NEW QUESTION 21

- (Exam Topic 3)

You create and register a model in an Azure Machine Learning workspace.

You must use the Azure Machine Learning SDK to implement a batch inference pipeline that uses a ParallelRunStep to score input data using the model. You must specify a value for the ParallelRunConfig compute\_target setting of the pipeline step.

You need to create the compute target. Which class should you use?

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

**Answer: C**

**Explanation:**

Compute target to use for ParallelRunStep. This parameter may be specified as a compute target object or the string name of a compute target in the workspace. The compute\_target target is of AmlCompute or string.

Note: An Azure Machine Learning Compute (AmlCompute) is a managed-compute infrastructure that allows you to easily create a single or multi-node compute. The compute is created within your workspace region as a resource that can be shared with other users

Reference:

<https://docs.microsoft.com/en-us/python/api/azureml-contrib-pipeline-steps/azureml.contrib.pipeline.steps.parall> [https://docs.microsoft.com/en-us/python/api/azureml-core/azureml.core.compute.amlcompute\(class\)](https://docs.microsoft.com/en-us/python/api/azureml-core/azureml.core.compute.amlcompute(class))

**NEW QUESTION 26**

- (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 27**

- (Exam Topic 3)

You are creating an experiment by using Azure Machine Learning Studio.

You must divide the data into four subsets for evaluation. There is a high degree of missing values in the data. You must prepare the data for analysis.

You need to select appropriate methods for producing the experiment.

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

NOTE: More than one order of answer choices is correct. You will receive credit for any of the correct orders you select.

**Actions**

**Answer Area**

Build Counting Transform

Missing Values Scrubber

Feature Hashing

Clean Missing Data

Replace Discrete Values

Import Data

Latent Dirichlet Transformation

Partition and Sample



- A. Mastered
- B. Not Mastered

**Answer: A**

**Explanation:**

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

#### NEW QUESTION 28

- (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 33

- (Exam Topic 3)

Note: This question is part of a series of questions that present the same scenario. Each question in the series contains a unique solution that might meet the stated goals. Some question sets might have more than one correct solution, while others might not have a correct solution.

After you answer a question in this section, you will NOT be able to return to it. As a result, these questions will not appear in the review screen.

You are creating a new experiment in Azure Learning learning Studio.

One class has a much smaller number of observations than the other classes in the training

You need to select an appropriate data sampling strategy to compensate for the class imbalance. Solution: You use the Synthetic Minority Oversampling

Technique (SMOTE) sampling mode. Does the solution meet the goal?

A. Yes

B. No

**Answer:** A

#### Explanation:

SMOTE is used to increase the number of underrepresented cases in a dataset used for machine learning. SMOTE is a better way of increasing the number of rare cases than simply duplicating existing cases.

References:

<https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/smote>

#### NEW QUESTION 36

- (Exam Topic 3)

You need to select a pre built development environment for a series of data science experiments. You must use the R language for the experiments.

Which three environments can you use? Each correct answer presents a complete solution. NOTE: Each correct selection is worth one point.

A. MI.NET Library on a local environment

B. Azure Machine Learning Studio

C. Data Science Virtual Machine (DSVM)

D. Azure Data bricks

E. Azure Cognitive Services

**Answer:** ABD

#### NEW QUESTION 37

- (Exam Topic 3)

You are implementing a machine learning model to predict stock prices. The model uses a PostgreSQL database and requires GPU processing.

You need to create a virtual machine that is pre-configured with the required tools. What should you do?

A. Create a Data Science Virtual Machine (DSVM) Windows edition.

B. Create a Geo AI Data Science Virtual Machine (Geo-DSVM) Windows edition.

C. Create a Deep Learning Virtual Machine (DLVM) Linux edition.

D. Create a Deep Learning Virtual Machine (DLVM) Windows edition.

E. Create a Data Science Virtual Machine (DSVM) Linux edition.

**Answer:** E

#### NEW QUESTION 41

- (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'
resource_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.	<input type="radio"/>	<input type="radio"/>
An Azure Databricks experiment will be tracked only in the Azure Machine Learning workspace.	<input type="radio"/>	<input type="radio"/>
The epoch loss metric is set to be tracked.	<input type="radio"/>	<input type="radio"/>

- 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 44**

- (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 47**

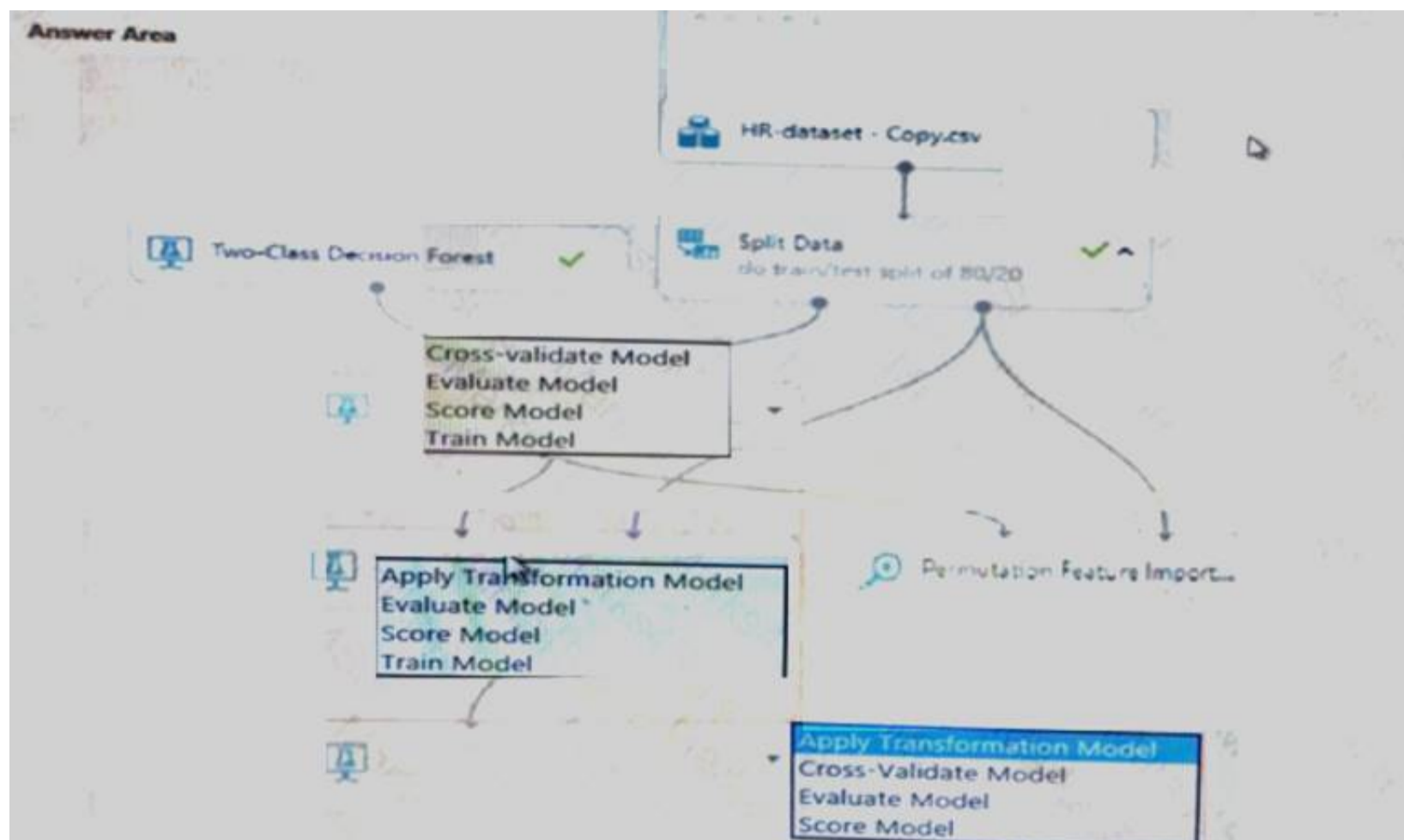
- (Exam Topic 3)

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.

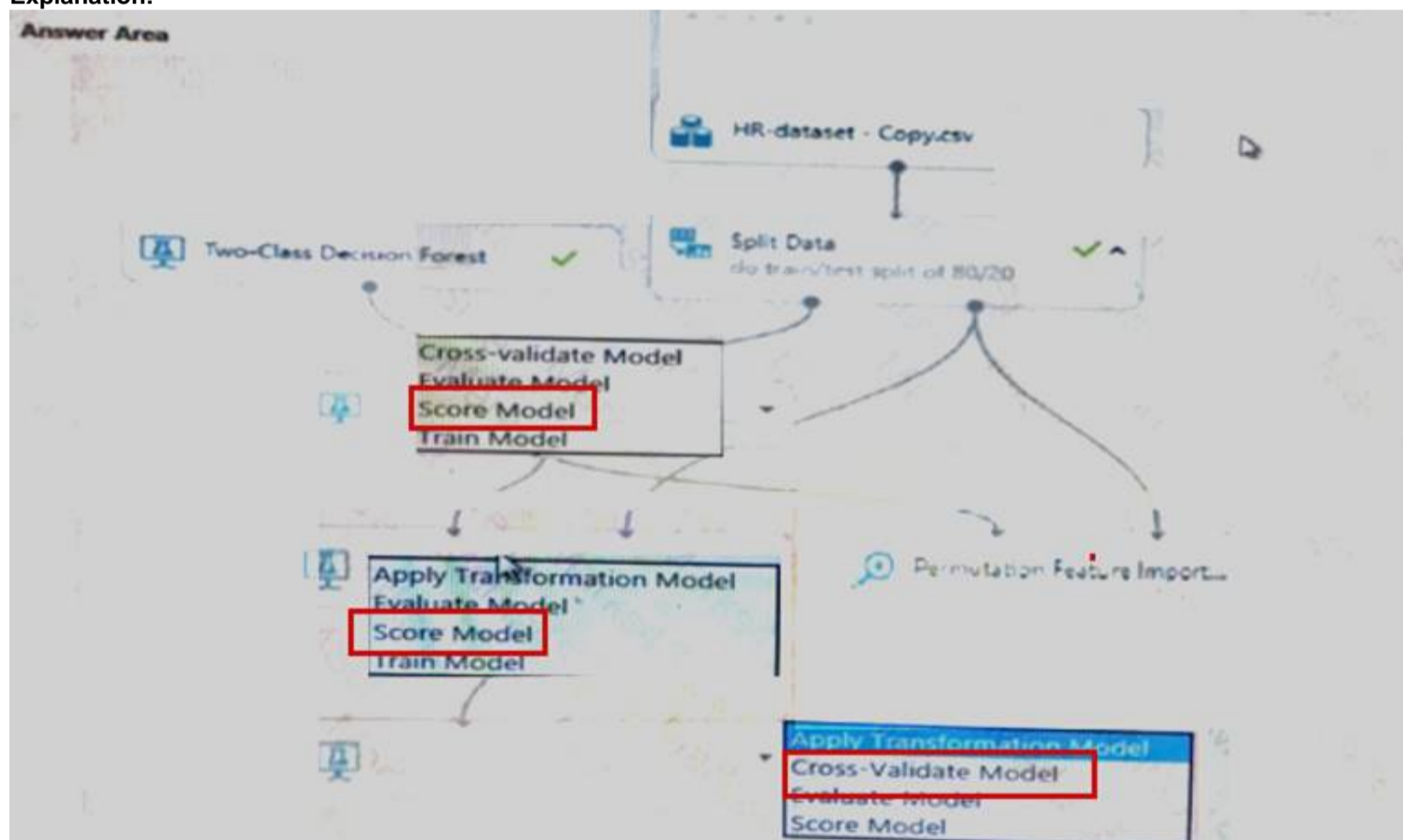
NOTE: Each correct selection is worth one point.



- A. Mastered
- B. Not Mastered

Answer: A

Explanation:



#### NEW QUESTION 51

- (Exam Topic 3)

You plan to explore demographic data for home ownership in various cities. The data is in a CSV file with the following format:

age,city,income,home\_owner 21,Chicago,50000,0 35,Seattle,120000,1 23,Seattle,65000,0 45,Seattle,130000,1 18,Chicago,48000,0

You need to run an experiment in your Azure Machine Learning workspace to explore the data and log the results. The experiment must log the following information:

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

You need to use the appropriate logging methods of the experiment's run object to log the required information.

How should you complete the code? To answer, drag the appropriate code segments to the correct locations. Each code segment may be used once, more than once, or not at all. You may need to drag the split bar between panes or scroll to view content.

NOTE: Each correct selection is worth one point.

## Code segments

log  
log\_list  
log\_row  
log\_table  
log\_image

## Answer Area

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

- A. Mastered  
B. Not Mastered

Answer: A

**Explanation:**

Box 1: log

The number of observations in the dataset. run.log(name, value, description="")

Scalar values: Log a numerical or string value to the run with the given name. Logging a metric to a run causes that metric to be stored in the run record in the experiment. You can log the same metric multiple times within a run, the result being considered a vector of that metric.

Example: run.log("accuracy", 0.95)

Box 2: log\_image

A box plot of income by home\_owner.

log\_image Log an image to the run record. Use log\_image to log a .PNG image file or a matplotlib plot to the run. These images will be visible and comparable in the run record.

Example: run.log\_image("ROC", plot=plt)

Box 3: log\_table

A dictionary containing the city names and the average income for each city. log\_table: Log a dictionary object to the run with the given name.

**NEW QUESTION 52**

- (Exam Topic 3)

Note: This question is part of a series of questions that present the same scenario. Each question in the series contains a unique solution that might meet the stated goals. Some question sets might have more than one correct solution, while others might not have a correct solution.

After you answer a question in this section, you will NOT be able to return to it. As a result, these questions will not appear in the review screen.

You are using Azure Machine Learning to run an experiment that trains a classification model.

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

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

You plan to use this configuration to run a script that trains a random forest model and then tests it with validation data. The label values for the validation data are stored in a variable named y\_test variable, and the predicted probabilities from the model are stored in a variable named y\_predicted.

You need to add logging to the script to allow Hyperdrive to optimize hyperparameters for the AUC metric. Solution: Run the following code:

```
import json, os
from sklearn.metrics import roc_auc_score
# code to train model omitted
auc = roc_auc_score(y_test, y_predicted)
os.makedirs("outputs", exist_ok = True)
with open("outputs/AUC.txt", "w") as file_cur:
    file_cur.write(auc)
```

Does the solution meet the goal?

- A. Yes

B. No

**Answer:** B

**Explanation:**

Use a solution with logging.info(message) instead. Note: Python printing/logging example: logging.info(message)

Destination: Driver logs, Azure Machine Learning designer Reference:

<https://docs.microsoft.com/en-us/azure/machine-learning/how-to-debug-pipelines>

**NEW QUESTION 53**

- (Exam Topic 3)

Note: This question is part of a series of questions that present the same scenario. Each question in the series contains a unique solution that might meet the stated goals. Some question sets might have more than one correct solution, while others might not have a correct solution.

After you answer a question in this section, you will NOT be able to return to it. As a result, these questions will not appear in the review screen.

You create a model to forecast weather conditions based on historical data.

You need to create a pipeline that runs a processing script to load data from a datastore and pass the processed data to a machine learning model training script.

Solution: Run the following code:

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

Does the solution meet the goal?

A. Yes

B. No

**Answer:** A

**Explanation:**

The two steps are present: process\_step and train\_step Note:

Data used in pipeline can be produced by one step and consumed in another step by providing a PipelineData object as an output of one step and an input of one or more subsequent steps.

PipelineData objects are also used when constructing Pipelines to describe step dependencies. To specify that a step requires the output of another step as input, use a PipelineData object in the constructor of both steps.

For example, the pipeline train step depends on the process\_step\_output output of the pipeline process step: from azureml.pipeline.core import Pipeline, PipelineData

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

Reference:

<https://docs.microsoft.com/en-us/python/api/azureml-pipeline-core/azureml.pipeline.core.pipelinedata?view=azu>

**NEW QUESTION 54**

- (Exam Topic 3)

Note: This question is part of a series of questions that present the same scenario. Each question in the series contains a unique solution that might meet the stated goals. Some question sets might have more than one correct solution, while others might not have a correct solution.

After you answer a question in this section, you will NOT be able to return to it. As a result, these questions will not appear in the review screen.

You are creating a new experiment in Azure Machine Learning Studio.

One class has a much smaller number of observations than the other classes in the training set. You need to select an appropriate data sampling strategy to compensate for the class imbalance. Solution: You use the Scale and Reduce sampling mode.

Does the solution meet the goal?

A. Yes

B. No

**Answer:** B

**Explanation:**

Instead use the Synthetic Minority Oversampling Technique (SMOTE) sampling mode.

Note: SMOTE is used to increase the number of underrepresented cases in a dataset used for machine learning. SMOTE is a better way of increasing the number of rare cases than simply duplicating existing cases.

References:

<https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/smote>

**NEW QUESTION 57**

- (Exam Topic 3)

You 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
- B. Replace using Probabilistic PAC
- C. Replace using MICE
- D. Normalization

**Answer: B**

#### NEW QUESTION 61

- (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=1
2
3
6, shuffle = True, random_state=1)

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

print('train: %s, test: %5s' % (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 65

- (Exam Topic 3)

You are performing sentiment analysis using a CSV file that includes 12,000 customer reviews written in a short sentence format. You add the CSV file to Azure Machine Learning Studio and configure it as the starting point dataset of an experiment. You add the Extract N-Gram Features from Text module to the experiment to extract key phrases from the customer review column in the dataset.

You must create a new n-gram dictionary from the customer review text and set the maximum n-gram size to trigrams.

What should you select? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.

Properties

Project

Extract N-Gram Features from Text

Text column

Selected columns

Column type: String Feature

Launch column selector

Vocabulary mode

▼

Create

ReadOnly

Update

Merge

N-Grams size

▼

3

4

4,000

12,000

0

Weighting function

▼

Minimum word length

3

Maximum word length

25

Minimum n-gram document absolu...

5

Maximum n-gram document ratio

1

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

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

NEW QUESTION 66

- (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-learningpipeline>

**NEW QUESTION 67**

- (Exam Topic 3)

**HOTSPOT**

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

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

The data is collected every 12 hours: noon and midnight.

You plan to use automated machine learning to create a time-series model that predicts temperature over the next seven days. For the initial round of training, you want to train a maximum of 50 different models.

You must use the Azure Machine Learning SDK to run an automated machine learning experiment to train these models.

You need to configure the automated machine learning run.

How should you complete the `AutoMLConfig` definition? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.

```

automl_config = AutoMLConfig(task="
                                regression
                                forecasting
                                classification
                                deep learning

                                training_data=weather_df,
                                label_column_name="
                                humidity
                                pressure
                                visibility
                                temperature
                                days_since_last
                                observation_time

                                time_column_name="
                                humidity
                                pressure
                                visibility
                                temperature
                                days_since_last
                                observation_time

                                max_horizon=
                                2
                                6
                                7
                                12
                                14
                                50

                                iterations=
                                2
                                6
                                7
                                12
                                14
                                50

                                iteration_timeout_minutes=5,
                                primary_metric="r2_score")

```

- A. Mastered  
B. Not Mastered

**Answer:** A

**Explanation:**

Box 1: forecasting

Task: The type of task to run. Values can be 'classification', 'regression', or 'forecasting' depending on the type of automated ML problem to solve.

Box 2: temperature

The training data to be used within the experiment. It should contain both training features and a label column (optionally a sample weights column).

Box 3: observation\_time

time\_column\_name: The name of the time column. This parameter is required when forecasting to specify the datetime column in the input data used for building the time series and inferring its frequency. This setting is being deprecated. Please use forecasting\_parameters instead.

Box 4: 7

"predicts temperature over the next seven days"

max\_horizon: The desired maximum forecast horizon in units of time-series frequency. The default value is 1. Units are based on the time interval of your training data, e.g., monthly, weekly that the forecaster should

predict out. When task type is forecasting, this parameter is required.

Box 5: 50

"For the initial round of training, you want to train a maximum of 50 different models."

Iterations: The total number of different algorithm and parameter combinations to test during an automated ML experiment. Reference:

https://docs.microsoft.com/en-us/python/api/azureml-train-automl-client/azureml.train.automl.automlconfig.auto

**NEW QUESTION 68**

- (Exam Topic 3)

You are tuning a hyperparameter for an algorithm. The following table shows a data set with different hyperparameter, training error, and validation errors.

Hyperparameter (H)	Training error (TE)	Validation error (VE)
1	105	95
2	200	85
3	250	100
4	105	100
5	400	50

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

**Question**

**Answer Choise**

Which H value should you select based on the data?

▼

1

2

3

4

5

What H value displays the poorest training result?

▼

1

2

3

4

5

- A. Mastered  
B. Not Mastered

**Answer:** A

**Explanation:**

Box 1: 4

Choose the one which has lower training and validation error and also the closest match. Minimize variance (difference between validation error and train error).

Box 2: 5

Minimize variance (difference between validation error and train error). Reference:

https://medium.com/comet-ml/organizing-machine-learning-projects-project-management-guidelines-2d2b8565

**NEW QUESTION 73**

- (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

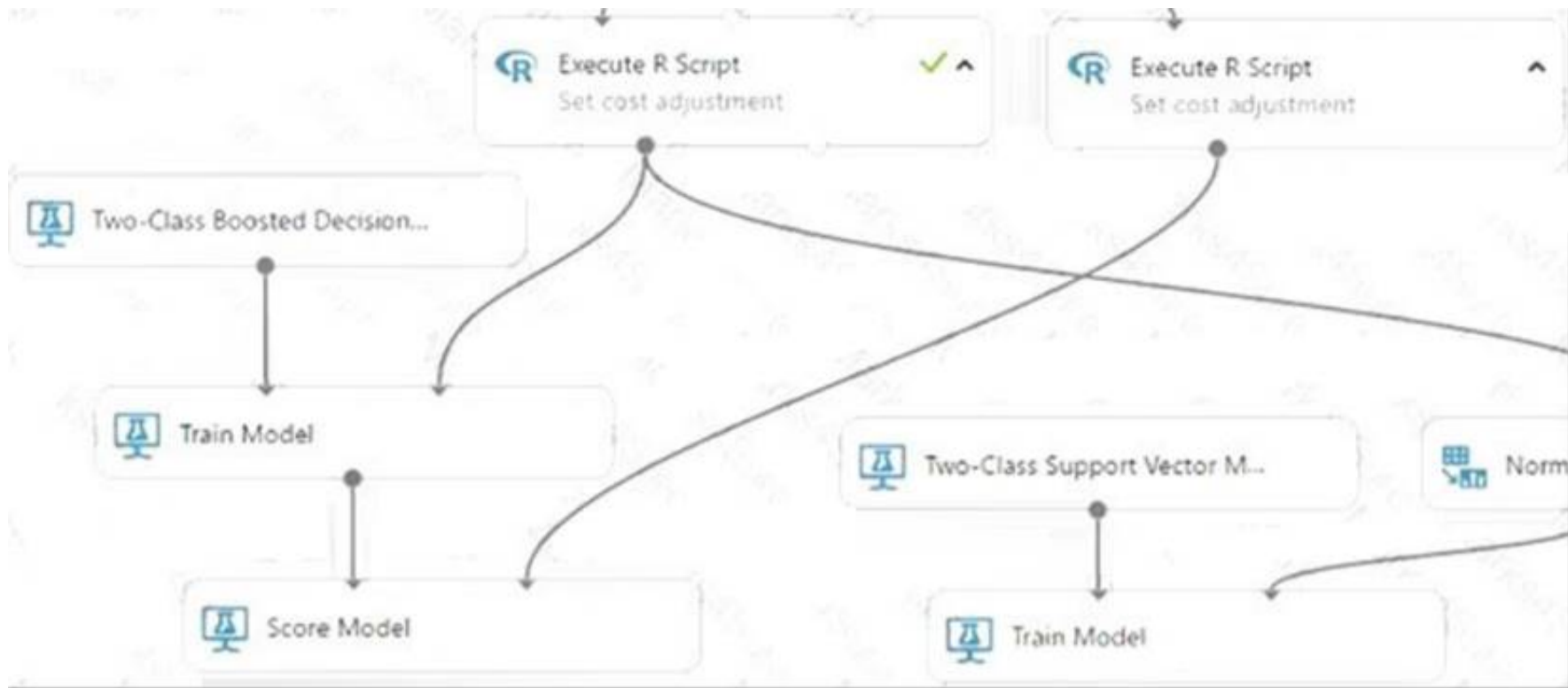
You use the testing data that was separated out by the Split Data module to score our trained models. You can then compare the results of the two models to see which generated better results.

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.

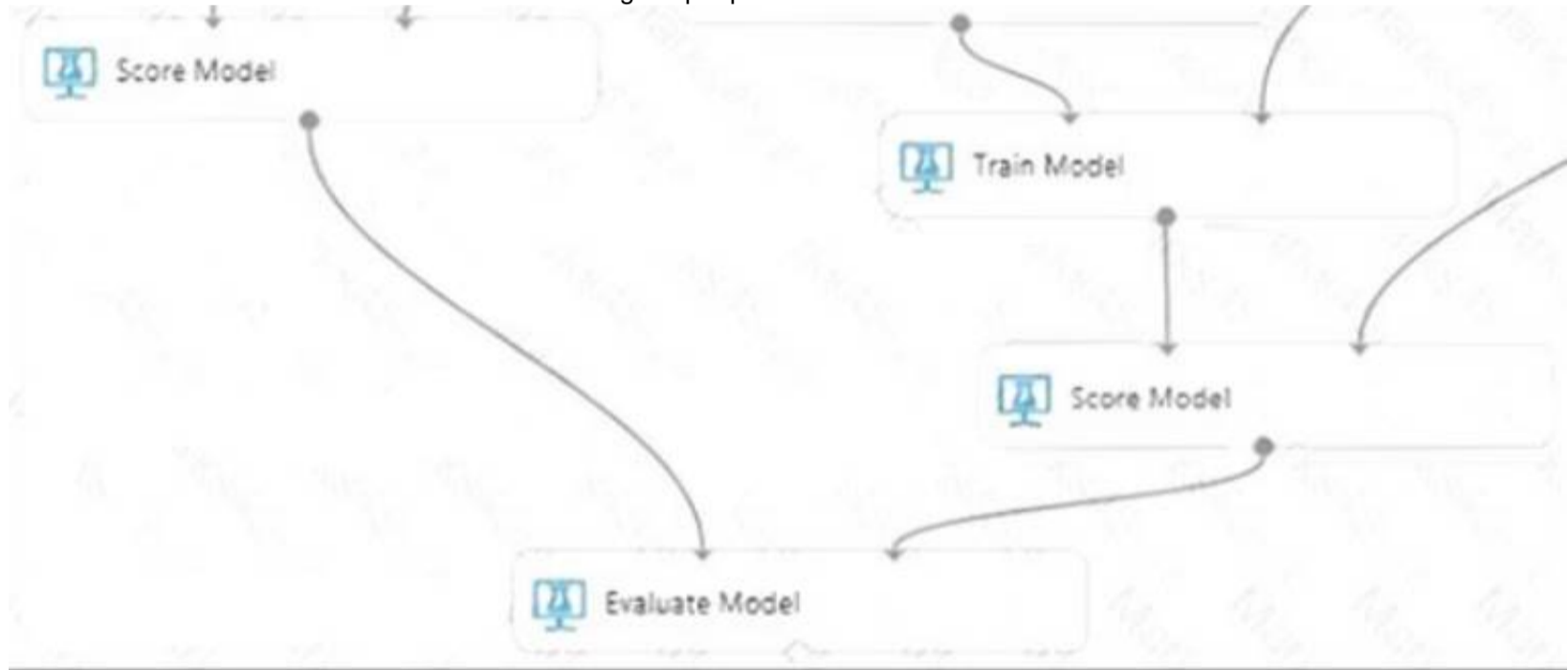
\* 3. Connect the right Execute R Script module (our testing data) to the right input port of the Score Model module.



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 77

- (Exam Topic 3)

You configure a Deep Learning Virtual Machine for Windows.

You need to recommend tools and frameworks to perform the following:

- > Build deep neural network (DNN) models
- > Perform interactive data exploration and visualization

Which tools and frameworks should you recommend? To answer, drag the appropriate tools to the correct tasks. Each tool may be used once, more than once, or not at all. You may need to drag the split bar between panes or scroll to view content.

NOTE: Each correct selection is worth one point.

Tools	Answer Area						
<div>Vowpal Wabbit</div> <div>PowerBI Desktop</div> <div>Azure Data Factory</div> <div>Microsoft Cognitive Toolkit</div>	<table border="1"><thead><tr><th>Task</th><th>Tool</th></tr></thead><tbody><tr><td>Build DNN models</td><td><div>Tool</div></td></tr><tr><td>Enable interactive data exploration and visualization</td><td><div>Tool</div></td></tr></tbody></table>	Task	Tool	Build DNN models	<div>Tool</div>	Enable interactive data exploration and visualization	<div>Tool</div>
Task	Tool						
Build DNN models	<div>Tool</div>						
Enable interactive data exploration and visualization	<div>Tool</div>						

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

Box 1: Vowpal Wabbit

Use the Train Vowpal Wabbit Version 8 module in Azure Machine Learning Studio (classic), to create a machine learning model by using Vowpal Wabbit.

Box 2: PowerBI Desktop

Power BI Desktop is a powerful visual data exploration and interactive reporting tool

BI is a name given to a modern approach to business decision making in which users are empowered to find, explore, and share insights from data across the enterprise.

References:

<https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/train-vowpal-wabbit-version-> <https://docs.microsoft.com/en-us/azure/architecture/data-guide/scenarios/interactive-data-exploration>

### NEW QUESTION 79

- (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 digitidentification.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 82

- (Exam Topic 3)

You plan to use a Data Science Virtual Machine (DSVM) with the open source deep learning frameworks Caffe2 and Theano. You need to select a pre configured DSVM to support the framework.

What should you create?

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

**Answer:** E

### NEW QUESTION 87

- (Exam Topic 3)

You create a classification model with a dataset that contains 100 samples with Class A and 10,000 samples with Class B

The variation of Class B is very high. You need to resolve imbalances. Which method should you use?

- A. Partition and Sample
- B. Cluster Centroids
- C. Tomek links
- D. Synthetic Minority Oversampling Technique (SMOTE)

**Answer:** D

### NEW QUESTION 91

- (Exam Topic 3)

You use the Azure Machine Learning designer to create and run a training pipeline. You then create a real-time inference pipeline.

You must deploy the real-time inference pipeline as a web service.

What must you do before you deploy the real-time inference pipeline?

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

**Answer:** D

**Explanation:**

You need to create an inferencing cluster. Deploy the real-time endpoint

After your AKS service has finished provisioning, return to the real-time inferencing pipeline to complete deployment.

- > Select Deploy above the canvas.
- > Select Deploy new real-time endpoint.
- > Select the AKS cluster you created.
- > Select Deploy. Reference:

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

**NEW QUESTION 92**

- (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.

**Compute types**

Attached compute

Inference cluster

Training cluster

**Answer Area**

Requirement	Compute type
Train models by using the Azure Machine Learning designer.	<div style="border: 1px solid #ccc; padding: 5px; text-align: center;">Compute type</div>
Score new data through a trained model published as a real-time web service.	<div style="border: 1px solid #ccc; padding: 5px; text-align: center;">Compute type</div>
Train models by using an Azure Databricks cluster.	<div style="border: 1px solid #ccc; padding: 5px; text-align: center;">Compute type</div>
Deploy models by using the Azure Machine Learning designer.	<div style="border: 1px solid #ccc; padding: 5px; text-align: center;">Compute type</div>

- 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 95**

- (Exam Topic 3)

You are performing a classification task in Azure Machine Learning Studio.

You must prepare balanced testing and training samples based on a provided data set. You need to split the data with a 0.75:0.25 ratio.

Which value should you use for each parameter? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.

Parameter	Value
Splitting mode	<div>▼</div> <div>Split rows Recommender Split Regular Expression Split Relative Expression Split</div>
Fraction of rows in the first output dataset	<div>▼</div> <div>0.75 0.25 0.5 1</div>
Randomized split	<div>▼</div> <div>True False</div>
Stratified split	<div>▼</div> <div>True False</div>

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

Box 1: Split rows

Use the Split Rows 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.

You can also randomize the selection of rows in each group, and use stratified sampling. In stratified sampling, you must select a single column of data for which you want values to be apportioned equally among the two result datasets.

Box 2: 0.75

If you specify a number as a percentage, or if you use a string that contains the "%" character, the value is interpreted as a percentage. All percentage values must be within the range (0, 100), not including the values 0 and 100.

Box 3: Yes

To ensure splits are balanced.

Box 4: No

If you use the option for a stratified split, the output datasets can be further divided by subgroups, by selecting a strata column.

Reference:

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

NEW QUESTION 98

- (Exam Topic 3)

You are performing feature engineering on a dataset.

You must add a feature named CityName and populate the column value with the text London.

You need to add the new feature to the dataset.

Which Azure Machine Learning Studio module should you use?

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

Answer: A

Explanation:

Typical metadata changes might include marking columns as features. References:

<https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/edit-metadata>

NEW QUESTION 99

- (Exam Topic 3)

You use Data Science Virtual Machines (DSVMs) for Windows and Linux in Azure. You need to access the DSVMs.

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

Requirement	Utility
Use terminal sessions to access a DSVM for Linux.	<div> SSH Client  X2Go  JupyterLab  Remote Desktop </div>
Access Jupyter notebooks on a DSVM for Linux.	<div> SSH Client  X2Go  JupyterLab  Remote Desktop </div>
Access Jupyter notebooks on a DSVM for Linux.	<div> SSH Client  X2Go  JupyterLab  Remote Desktop </div>
Access a DSVM for Windows.	<div> SSH Client  X2Go  JupyterLab  Remote Desktop </div>

- A. Mastered  
B. Not Mastered

Answer: A

Explanation:

Requirement	Utility
Use terminal sessions to access a DSVM for Linux.	<div> SSH Client  X2Go  JupyterLab  Remote Desktop </div>
Access Jupyter notebooks on a DSVM for Linux.	<div> SSH Client  X2Go  JupyterLab  Remote Desktop </div>
Access Jupyter notebooks on a DSVM for Linux.	<div> SSH Client  X2Go  JupyterLab  Remote Desktop </div>
Access a DSVM for Windows.	<div> SSH Client  X2Go  JupyterLab  Remote Desktop </div>

#### NEW QUESTION 100

- (Exam Topic 3)

You are creating a machine learning model. You have a dataset that contains null rows.

You need to use the Clean Missing Data module in Azure Machine Learning Studio to identify and resolve the null and missing data in the dataset.

Which parameter should you use?

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

Answer: B

Explanation:

Remove entire row: Completely removes any row in the dataset that has one or more missing values. This is useful if the missing value can be considered randomly missing.

References:

<https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/clean-missing-data>

### NEW QUESTION 103

- (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 107

- (Exam Topic 3)

You are building a regression model to estimate the number of calls during an event.

You need to determine whether the feature values achieve the conditions to build a Poisson regression model. Which two conditions must the feature set contain?

Each correct answer presents part of the solution. NOTE:

Each correct selection is worth one point.

- A. The label data must be a negative value.
- B. The label data 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 112

- (Exam Topic 3)

You are performing a filter-based feature selection for a dataset to build a multi-class classifier by using Azure Machine Learning Studio.

The dataset contains categorical features that are highly correlated to the output label column.

You need to select the appropriate feature scoring statistical method to identify the key predictors. Which method should you use?

- A. Chi-squared
- B. Spearman correlation
- C. Kendall correlation
- D. Pearson 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 113

- (Exam Topic 3)

You are developing a linear regression model in Azure Machine Learning Studio. You run an experiment to compare different algorithms.

The following image displays the results dataset output:

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

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

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

- A. Mastered  
 B. Not Mastered

Answer: A

#### Explanation:

Box 1: Boosted Decision Tree Regression

Mean absolute error (MAE) measures how close the predictions are to the actual outcomes; thus, a lower score is better.

Box 2:

Online Gradient Descent: If you want the algorithm to find the best parameters for you, set Create trainer mode option to Parameter Range. You can then specify multiple values for the algorithm to try.

References:

<https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/evaluate-model> <https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/linear-regression>

#### NEW QUESTION 114

- (Exam Topic 3)

Note: This question is part of a series of questions that present the same scenario. Each question in the series contains a unique solution that might meet the stated goals. Some question sets might have more than one correct solution, while others might not have a correct solution.

After you answer a question in this section, you will NOT be able to return to it. As a result, these questions will not appear in the review screen.

You create an Azure Machine Learning service datastore in a workspace. The datastore contains the following files:

- /data/2018/Q1 .csv
- /data/2018/Q2.csv
- /data/2018/Q3.csv
- /data/2018/Q4.csv
- /data/2019/Q1.csv

All files store data in the following format:

- id,f1,f2,l
- 1,1,2,0
- 2,1,1,1
- 3.2.1.0

You run the following code:

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

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

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

Solution: Run the following code:

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

Does the solution meet the goal?

- A. Yes
- B. No

**Answer:** A

**Explanation:**

Use two file paths.

Use Dataset.Tabular\_from\_delimited as the data isn't cleansed. Note:

A TabularDataset represents data in a tabular format by parsing the provided file or list of files. This provides you with the ability to materialize the data into a pandas or Spark DataFrame so you can work with familiar data preparation and training libraries without having to leave your notebook. You can create a TabularDataset object from .csv, .tsv, .parquet, .jsonl files, and from SQL query results.

Reference:

<https://docs.microsoft.com/en-us/azure/machine-learning/how-to-create-register-datasets>

**NEW QUESTION 116**

- (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 118**

- (Exam Topic 3)

You create an Azure Machine Learning compute resource to train models. The compute resource is configured as follows:

- Minimum nodes: 2
- Maximum nodes: 4

You must decrease the minimum number of nodes and increase the maximum number of nodes to the following values:

- Minimum nodes: 0
- Maximum nodes: 8

You need to reconfigure the compute resource.

What are three possible ways to achieve this goal? Each correct answer presents a complete solution.

NOTE: Each correct selection is worth one point.

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

**Answer:** ABC

**Explanation:**

Reference:

[https://docs.microsoft.com/en-us/python/api/azureml-core/azureml.core.compute.amlcompute\(class\)](https://docs.microsoft.com/en-us/python/api/azureml-core/azureml.core.compute.amlcompute(class))

**NEW QUESTION 123**

- (Exam Topic 3)

You are working on a classification task. You have a dataset indicating whether a student would like to play soccer and associated attributes. The dataset includes the following columns:

Name	Description
IsPlaySoccer	Values can be 1 and 0.
Gender	Values can be M or F.
PrevExamMarks	Stores values from 0 to 100
Height	Stores values in centimeters
Weight	Stores values in kilograms

You need to classify variables by type.

Which variable should you add to each category? To answer, select the appropriate options in the answer area. NOTE: Each correct selection is worth one point.

Category	Variables
Categorical variables	<div> Gender, IsPlaySoccer  Gender, PrevExamMarks, Height, Weight  PrevExamMarks, Height, Weight  IsPlaySoccer </div>
Continuous variables	<div> Gender, IsPlaySoccer  Gender, PrevExamMarks, Height, Weight  PrevExamMarks, Height, Weight  IsPlaySoccer </div>

- A. Mastered
- B. Not Mastered

**Answer:** A

**Explanation:**

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

#### NEW QUESTION 127

- (Exam Topic 3)

You create a binary classification model. The model is registered in an Azure Machine Learning workspace. You use the Azure Machine Learning Fairness SDK to assess the model fairness.

You develop a training script for the model on a local machine.

You need to load the model fairness metrics into Azure Machine Learning studio. What should you do?

- A. Implement the download\_dashboard\_by\_upload\_id function
- B. Implement the create\_group\_metric\_sec function
- C. Implement the upload\_dashboard\_dictionary function
- D. Upload the training script

**Answer:** C

**Explanation:**

import azureml.contrib.fairness package to perform the upload:

from azureml.contrib.fairness import upload\_dashboard\_dictionary, download\_dashboard\_by\_upload\_id Reference:

<https://docs.microsoft.com/en-us/azure/machine-learning/how-to-machine-learning-fairness-aml>

#### NEW QUESTION 130

- (Exam Topic 3)

You run a script as an experiment in Azure Machine Learning.

You have a Run object named run that references the experiment run. You must review the log files that were generated during the experiment run.

You need to download the log files to a local folder for review.

Which two code segments can you run to achieve this goal? Each correct answer presents a complete solution.

NOTE: Each correct selection is worth one point.

- A. run.get\_details()
- B. run.get\_file\_names()
- C. run.get\_metrics()
- D. run.download\_files(output\_directory='./runfiles')
- E. run.get\_all\_logs(destination='./runlogs')

**Answer:** AE

**Explanation:**

The run Class get\_all\_logs method downloads all logs for the run to a directory.

The run Class get\_details gets the definition, status information, current log files, and other details of the run. Reference:

[https://docs.microsoft.com/en-us/python/api/azureml-core/azureml.core.run\(class\)](https://docs.microsoft.com/en-us/python/api/azureml-core/azureml.core.run(class))

#### NEW QUESTION 135

- (Exam Topic 3)

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

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

**Answer: D**

**Explanation:**

The Azure Machine Learning SDK does not provide support scaling an AKS cluster. To scale the nodes in the cluster, use the UI for your AKS cluster in the Azure Machine Learning studio. You can only change the node count, not the VM size of the cluster.

Reference:

<https://docs.microsoft.com/en-us/azure/machine-learning/how-to-create-attach-kubernetes>

**NEW QUESTION 140**

- (Exam Topic 3)

Note: This question is part of a series of questions that present the same scenario. Each question in the series contains a unique solution that might meet the stated goals. Some question sets might have more than one correct solution, while others might not have a correct solution.

After you answer a question in this section, you will NOT be able to return to it. As a result, these questions will not appear in the review screen.

You are 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 144**

- (Exam Topic 2)

You need to produce a visualization for the diagnostic test evaluation according to the data visualization requirements.

Which three modules should you recommend be used in sequence? To answer, move the appropriate modules from the list of modules to the answer area and arrange them in the correct order.

**Modules**

**Answer Area**

Score Matchbox Recommender

Apply Transformation

Evaluate Recommender

Evaluate Model

Train Model

Sweep Clustering

Score Model

Load Trained Model



- A. Mastered
- B. Not Mastered

**Answer:** A

**Explanation:**

Step 1: Sweep Clustering

Start by using the "Tune Model Hyperparameters" module to select the best sets of parameters for each of the models we're considering.

One of the interesting things about the "Tune Model Hyperparameters" module is that it not only outputs the results from the Tuning, it also outputs the Trained Model.

Step 2: Train Model Step 3: Evaluate Model

Scenario: You need to provide the test results to the Fabrikam Residences team. You create data visualizations to aid in presenting the results.

You must produce a Receiver Operating Characteristic (ROC) curve to conduct a diagnostic test evaluation of the model. You need to select appropriate methods for producing the ROC curve in Azure Machine Learning Studio to compare the Two-Class Decision Forest and the Two-Class Decision Jungle modules with one another.

References:

<http://breaking-bi.blogspot.com/2017/01/azure-machine-learning-model-evaluation.html>

**NEW QUESTION 147**

- (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  $\tau$ ), 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 150**

- (Exam Topic 2)

You need to configure the Feature Based Feature Selection module based on the experiment requirements and datasets.

How should you configure the module properties? To answer, select the appropriate options in the dialog box in the answer area.

NOTE: Each correct selection is worth one point.

Filter Based Feature Selection

Feature scoring method

Fisher Score  
Chi-squared  
Mutual information  
Counts

☒ Operate on feature columns only

Target column

MedianValue  
AvgRooms/nHouse

Launch column selector

Number of desired features

1

- A. Mastered
- B. Not Mastered

**Answer:** A

**Explanation:**

Box 1: Mutual Information.

The mutual information score is particularly useful in feature selection because it maximizes the mutual information between the joint distribution and target variables in datasets with many dimensions.

Box 2: MedianValue

MedianValue is the feature column, , it is the predictor of the dataset.  
Scenario: The MedianValue and AvgRoomsinHouse columns both hold data in numeric format. You need to select a feature selection algorithm to analyze the relationship between the two columns in more detail.  
References:  
<https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/filter-based-feature-selection>

NEW QUESTION 155

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

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

Answer: D

NEW QUESTION 160

- (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://en.wikipedia.org/wiki/Nearest_centroid_classifier)  
<https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/sweep-clustering>

NEW QUESTION 163

- (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

Define a cross-entropy function activation.

Add cost functions for each target state.

Evaluate the classification error metric.

Evaluate the distance error metric.

Add cost functions for each component metric.

Define a sigmoid loss function activation.

Answer Area

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

Step 1: Define a cross-entropy function activation  
When using a neural network to perform classification and prediction, it is usually better to use cross-entropy error than classification error, and somewhat better to use cross-entropy error than mean squared error to evaluate the quality of the neural network.  
Step 2: Add cost functions for each target state. Step 3: Evaluated the distance error metric. References:  
<https://www.analyticsvidhya.com/blog/2018/04/fundamentals-deep-learning-regularization-techniques/>

NEW QUESTION 168

- (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.

Actions

Build the global model using Microsoft Cognitive Toolkit (CNTK).

Import the global model and build a local model using Microsoft Cognitive Toolkit (CNTK).

Export the global model using Neural Network Exchange Format (NNEF).

Import the global model and build the local model using PyTorch.

Build the global model using PyTorch.

Build the global model using TensorFlow.

Import the global model and build the local model using TensorFlow.

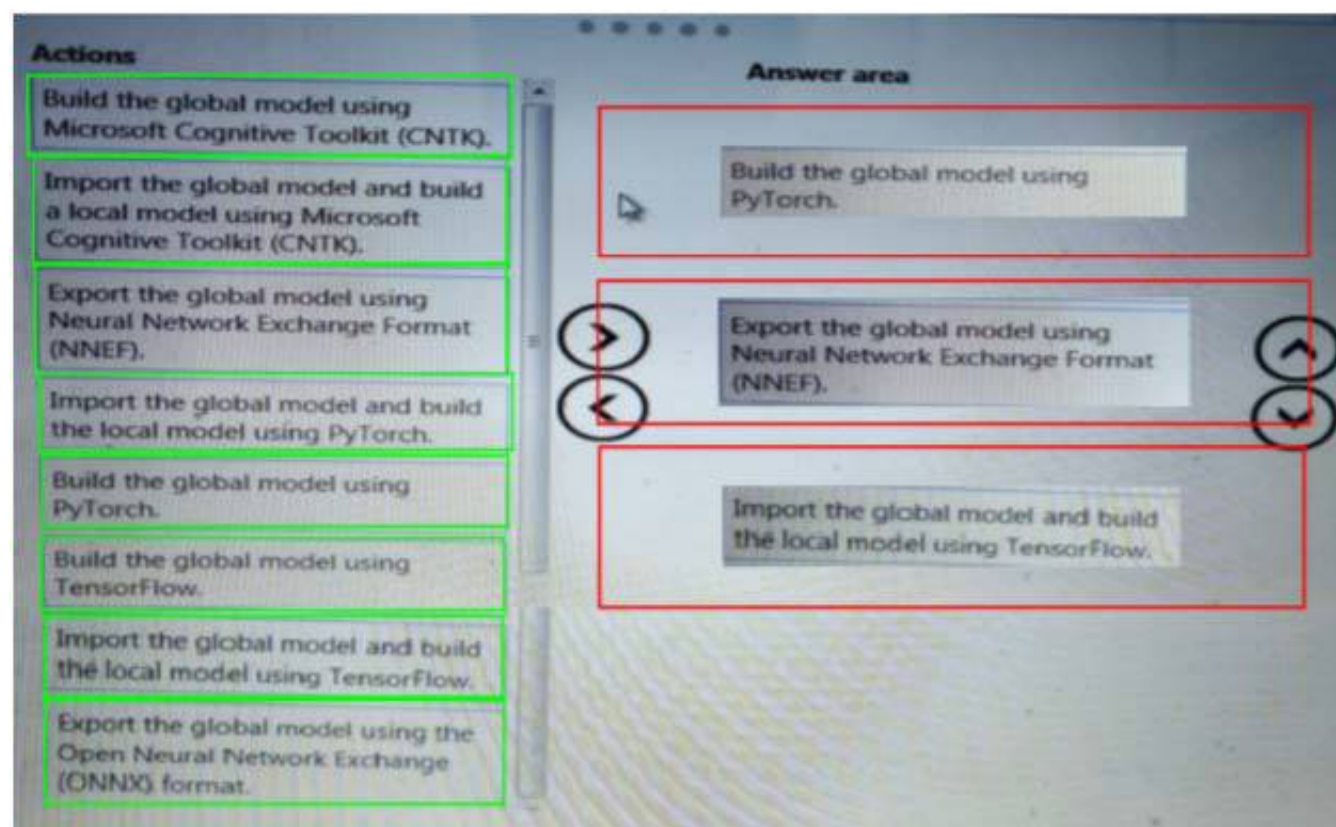
Export the global model using the Open Neural Network Exchange (ONNX) format.

Answer area

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:



NEW QUESTION 169

.....

## THANKS FOR TRYING THE DEMO OF OUR PRODUCT

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

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

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

## Money Back Guarantee

### DP-100 Practice Exam Features:

- \* DP-100 Questions and Answers Updated Frequently
- \* DP-100 Practice Questions Verified by Expert Senior Certified Staff
- \* DP-100 Most Realistic Questions that Guarantee you a Pass on Your FirstTry
- \* DP-100 Practice Test Questions in Multiple Choice Formats and Updatesfor 1 Year