

## 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 are using C-Support Vector classification to do a multi-class classification with an unbalanced training dataset. The C-Support Vector classification using Python code shown below:

```
from sklearn.svm import svc
import numpy as np
svc = SVC(kernel= 'linear', class_weight= 'balanced', C=1.0, random_state=0)
modell = svc.fit(X_train, y)
```

You need to evaluate the C-Support Vector classification code.

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

Code Segment	Evaluation Statement
class_weight=balanced	<ul style="list-style-type: none"> <li>Automatically select the performance metrics for the classification.</li> <li>Automatically adjust weights directly proportional to class frequencies in the input data.</li> <li>Automatically adjust weights inversely proportional to class frequencies in the input data.</li> </ul>
C parameter	<ul style="list-style-type: none"> <li>Penalty parameter</li> <li>Degree of polynomial kernel function</li> <li>Size of the kernel cache</li> </ul>

- A. Mastered
- B. Not Mastered

**Answer: A**

**Explanation:**

Box 1: Automatically adjust weights inversely proportional to class frequencies in the input data

The "balanced" mode uses the values of y to automatically adjust weights inversely proportional to class frequencies in the input data as  $n\_samples / (n\_classes * np.bincount(y))$ .

Box 2: Penalty parameter

Parameter: C : float, optional (default=1.0) Penalty parameter C of the error term. References:

<https://scikit-learn.org/stable/modules/generated/sklearn.svm.SVC.html>

**NEW QUESTION 2**

- (Exam Topic 3)

HOTSPOT

You register the following versions of a model.

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

You use the Azure ML Python SDK to run a training experiment. You use a variable named run to reference the experiment run.

After the run has been submitted and completed, you run the following code:

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

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

NOTE: Each correct selection is worth one point.

	Yes	No
The code will cause a previous version of the saved model to be overwritten.	<input type="radio"/>	<input type="radio"/>
The version number will now be 4.	<input type="radio"/>	<input type="radio"/>
The latest version of the stored model will have a property of value: 87.43.	<input type="radio"/>	<input type="radio"/>

- A. Mastered
- B. Not Mastered

**Answer:** A

**Explanation:**

Reference:

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

**NEW QUESTION 3**

- (Exam Topic 3)

You train a machine learning model.

You must deploy the model as a real-time inference service for testing. The service requires low CPU utilization and less than 48 MB of RAM. The compute target for the deployed service must initialize automatically while minimizing cost and administrative overhead.

Which compute target should you use?

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

**Answer:** C

**Explanation:**

Azure Container Instances (ACI) are suitable only for small models less than 1 GB in size. Use it for low-scale CPU-based workloads that require less than 48 GB of RAM.

Note: Microsoft recommends using single-node Azure Kubernetes Service (AKS) clusters for dev-test of larger models.

Reference:

<https://docs.microsoft.com/id-id/azure/machine-learning/how-to-deploy-and-where>

**NEW QUESTION 4**

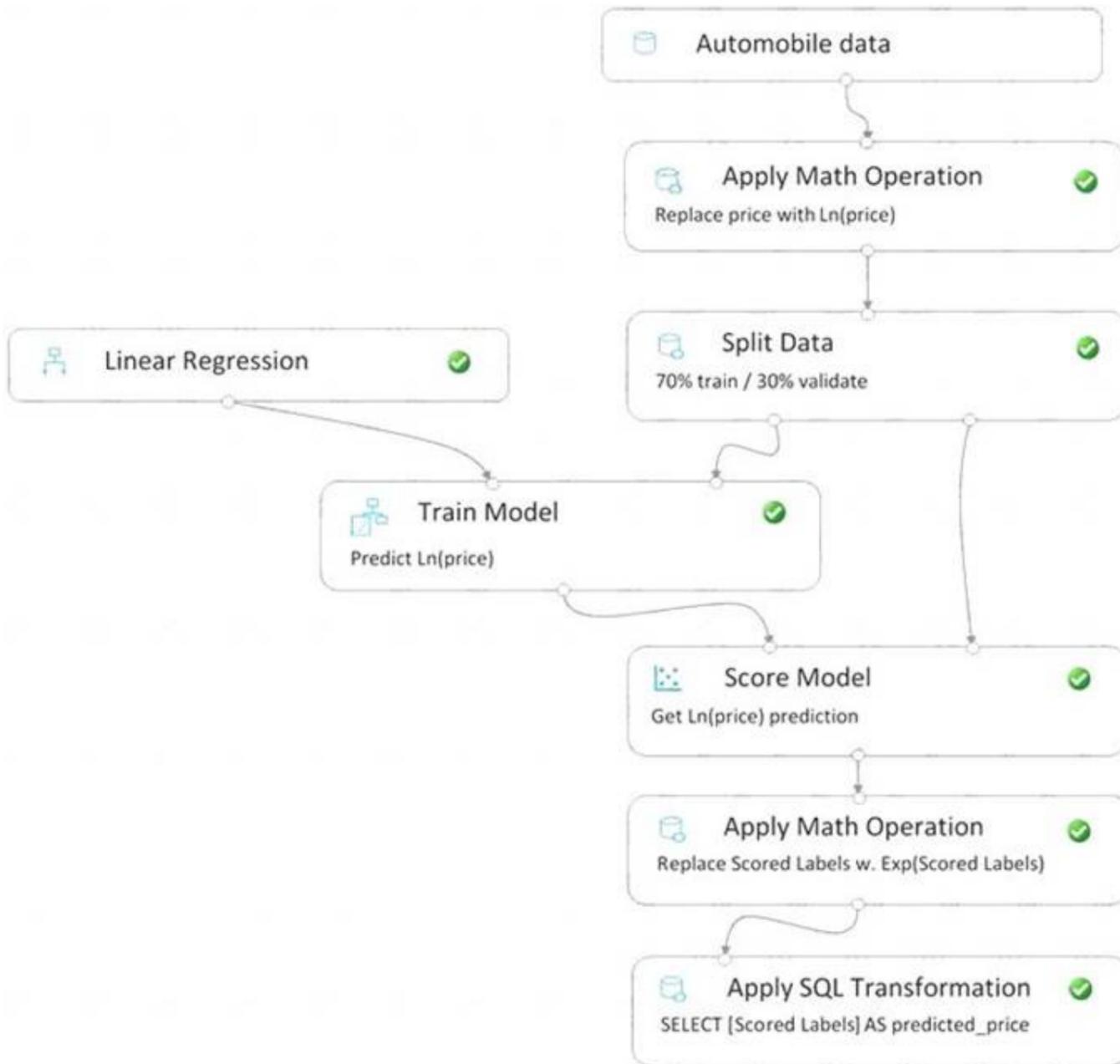
- (Exam Topic 3)

You create a pipeline in designer to train a model that predicts automobile prices.

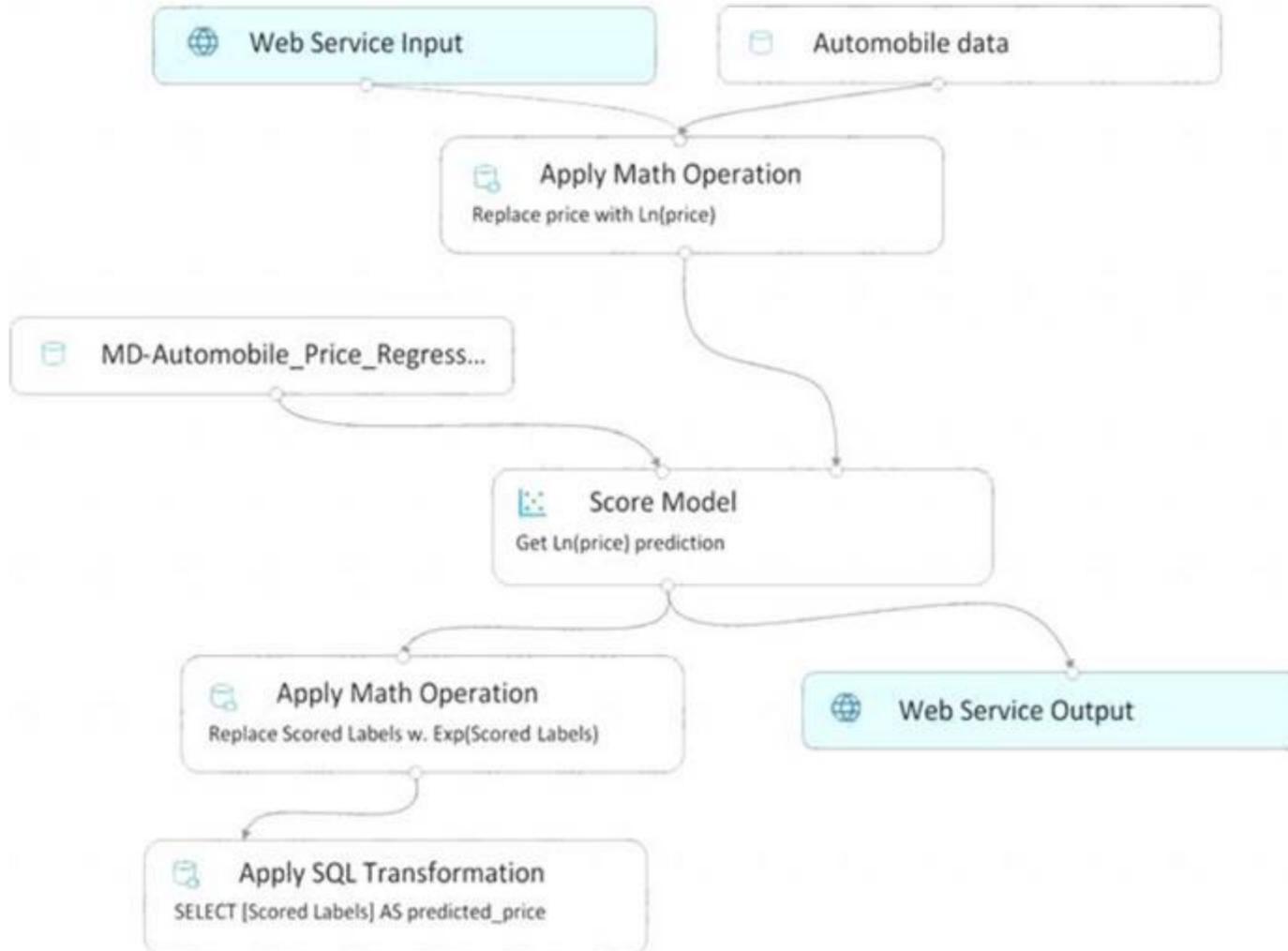
Because of non-linear relationships in the data, the pipeline calculates the natural log (Ln) of the prices in the training data, trains a model to predict this natural log of price value, and then calculates the exponential of the scored label to get the predicted price.

The training pipeline is shown in the exhibit. (Click the Training pipeline tab.)

Training pipeline



You create a real-time inference pipeline from the training pipeline, as shown in the exhibit. (Click the Real-time pipeline tab.)



You need to modify the inference pipeline to ensure that the web service returns the exponential of the scored label as the predicted automobile price and that client applications are not required to include a price value in the input values.

Which three modifications must you make to the inference pipeline? Each correct answer presents part of the solution.

NOTE: Each correct selection is worth one point.

- A. Connect the output of the Apply SQL Transformation to the Web Service Output module.
- B. Replace the Web Service Input module with a data input that does not include the price column.
- C. Add a Select Columns module before the Score Model module to select all columns other than price.
- D. Replace the training dataset module with a data input that does not include the price column.
- E. Remove the Apply Math Operation module that replaces price with its natural log from the data flow.

F. Remove the Apply SQL Transformation module from the data flow.

**Answer:** ACE

#### NEW QUESTION 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 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: Calculate the column median value and use the median value as the replacement for any missing value in the column.

Does the solution meet the goal?

A. Yes

B. No

**Answer:** B

#### Explanation:

Use the Multiple Imputation by Chained Equations (MICE) method. 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 6

- (Exam Topic 3)

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

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

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

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

Does the solution meet the goal?

A. Yes

B. No

**Answer:** B

#### Explanation:

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

Note: SMOTE is used to increase the number of 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 7

- (Exam Topic 3)

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

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

> Select the training features using the pandas filter method.

> Train a model based on the naive\_bayes.GaussianNB algorithm.

> Return only the Scored Labels column by using the query `SELECT [Scored Labels] FROM t1`; Which modules should you use? To answer, drag the appropriate modules to the appropriate locations. Each

module name may be used once, more than once, or not at all. You may need to drag the split bar between panes or scroll to view content.

NOTE: Each correct selection is worth one point.

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 8**

- (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 9**

- (Exam Topic 3)

You plan to use automated machine learning to train a regression model. You have data that has features which have missing values, and categorical features with few distinct values.

You need to configure automated machine learning to automatically impute missing values and encode categorical features as part of the training task.

Which parameter and value pair should you use in the AutoMLConfig class?

- A. featurization = 'auto'
- B. enable\_voting\_ensemble = True
- C. task = 'classification'
- D. exclude\_nan\_labels = True
- E. enable\_tf = True

**Answer:** A

**Explanation:**

Featurization str or FeaturizationConfig Values: 'auto' / 'off' / FeaturizationConfig

Indicator for whether featurization step should be done automatically or not, or whether customized featurization should be used.

Column type is automatically detected. Based on the detected column type preprocessing/featurization is done as follows:

Categorical: Target encoding, one hot encoding, drop high cardinality categories, impute missing values. Numeric: Impute missing values, cluster distance, weight of evidence.

DateTime: Several features such as day, seconds, minutes, hours etc. Text: Bag of words, pre-trained Word embedding, text target encoding. Reference:

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

**NEW QUESTION 10**

- (Exam Topic 3)

You train a model and register it in your Azure Machine Learning workspace. You are ready to deploy the model as a real-time web service.

You deploy the model to an Azure Kubernetes Service (AKS) inference cluster, but the deployment fails because an error occurs when the service runs the entry script that is associated with the model deployment.

You need to debug the error by iteratively modifying the code and reloading the service, without requiring a re-deployment of the service for each code update.

What should you do?

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

**Answer:** C

**Explanation:**

How to work around or solve common Docker deployment errors with Azure Container Instances (ACI) and Azure Kubernetes Service (AKS) using Azure Machine Learning.

The recommended and the most up to date approach for model deployment is via the Model.deploy() API using an Environment object as an input parameter. In this case our service will create a base docker image for you during deployment stage and mount the required models all in one call. The basic deployment tasks are:

- \* 1. Register the model in the workspace model registry.
- \* 2. Define Inference Configuration:
  - \* a. Create an Environment object based on the dependencies you specify in the environment yaml file or use one of our procured environments.
  - \* b. Create an inference configuration (InferenceConfig object) based on the environment and the scoring script.
- \* 3. Deploy the model to Azure Container Instance (ACI) service or to Azure Kubernetes Service (AKS).

**NEW QUESTION 10**

- (Exam Topic 3)

You are preparing to build a deep learning convolutional neural network model for image classification. You create a script to train the model using CUDA devices.

You must submit an experiment that runs this script in the Azure Machine Learning workspace. The following compute resources are available:

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

You need to specify the compute resources to be used for running the code to submit the experiment, and for running the script in order to minimize model training time.

Which resources should the data scientist use? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.

**Resource type**

**Option**

Run code to submit the experiment

	▼
the Microsoft Surface device	
the ds-workstation notebook VM	
the cpu-cluster compute target	
the gpu-cluster compute target	

Run the training script

	▼
the ds-workstation notebook VM	
the cpu-compute target	
the gpu-compute target	
the Microsoft Surface device	

- A. Mastered
- B. Not Mastered

**Answer:** A

**Explanation:**

**Resource type**

**Option**

Run code to submit the experiment

	▼
the Microsoft Surface device	
the ds-workstation notebook VM	
the cpu-cluster compute target	
the gpu-cluster compute target	

Run the training script

	▼
the ds-workstation notebook VM	
the cpu-compute target	
the gpu-compute target	
the Microsoft Surface device	

**NEW QUESTION 13**

- (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 15**

- (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 MLlib
- 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 20**

- (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-pipeline>

**NEW QUESTION 22**

- (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 overclocking
- 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 27**

- (Exam Topic 3)

You are hired as a data scientist at a winery. The previous data scientist used Azure Machine Learning. You need to review the models and explain how each model makes decisions.

Which explainer modules should you use? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.

Model type	Explainer
A random forest model for predicting the alcohol content in wine given a set of covariates	<div style="border: 1px solid black; padding: 5px;"> <div style="text-align: right; margin-bottom: 5px;">▼</div> <div style="border: 1px solid black; padding: 2px;">                     Tabular                      HAN                      Text                      Image                 </div> </div>
A natural language processing model for analyzing field reports	<div style="border: 1px solid black; padding: 5px;"> <div style="text-align: right; margin-bottom: 5px;">▼</div> <div style="border: 1px solid black; padding: 2px;">                     Tree                      HAN                      Text                      Image                 </div> </div>
An image classifier that determines the quality of the grape based upon its physical characteristics.	<div style="border: 1px solid black; padding: 5px;"> <div style="text-align: right; margin-bottom: 5px;">▼</div> <div style="border: 1px solid black; padding: 2px;">                     Kernel                      HAN                      Text                      Image                 </div> </div>

- A. Mastered
- B. Not Mastered

**Answer:** A

**Explanation:**

info based on the given model and data sets. The meta explainers leverage all the libraries (SHAP, LIME, Mimic, etc.) that we have integrated or developed. The following are the meta explainers available in the SDK:

Tabular Explainer: Used with tabular datasets. Text Explainer: Used with text datasets. Image Explainer: Used with image datasets. Box 1: Tabular

Box 2: Text

Box 3: Image Reference:

<https://medium.com/microsoftazure/automated-and-interpretable-machine-learning-d07975741298>

**NEW QUESTION 29**

- (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 32**

- (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 37**

- (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 41**

- (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 numpy as np
from sklearn.metrics import roc_auc_score
# code to train model omitted
auc = roc_auc_score(y_test, y_predicted)
print(np.float(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 43**

- (Exam Topic 3)

You are using the Hyperdrive feature in Azure Machine Learning to train a model. You configure the Hyperdrive experiment by running the following code:

```
from azureml.train.hyperdrive import RandomParameterSampling
param_sampling = RandomParameterSampling( {
    "learning_rate": normal(10, 3),
    "keep_probability": uniform(0.05, 0.1),
    "batch_size": choice(16, 32, 64, 128)
    "number_of_hidden_layers": choice(range(3,5))
})
```

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
By defining sampling in this manner, every possible combination of the parameters will be tested.	<input type="radio"/>	<input type="radio"/>
Random values of the learning_rate parameter will be selected from a normal distribution with a mean of 10 and a standard deviation of 3.	<input type="radio"/>	<input type="radio"/>
The keep_probability parameter value will always be either <b>0.05</b> or <b>0.1</b> .	<input type="radio"/>	<input type="radio"/>
Random values for the number_of_hidden_layers parameter will be selected from a normal distribution with a mean of 3 and a standard deviation of 5.	<input type="radio"/>	<input type="radio"/>

- A. Mastered
- B. Not Mastered

Answer: A

**Explanation:**

Box 1: Yes  
 In random sampling, hyperparameter values are randomly selected from the defined search space. Random sampling allows the search space to include both discrete and continuous hyperparameters.  
 Box 2: Yes  
 learning\_rate has a normal distribution with mean value 10 and a standard deviation of 3.  
 Box 3: No  
 keep\_probability has a uniform distribution with a minimum value of 0.05 and a maximum value of 0.1.  
 Box 4: No  
 number\_of\_hidden\_layers takes on one of the values [3, 4, 5].  
 Reference:  
<https://docs.microsoft.com/en-us/azure/machine-learning/how-to-tune-hyperparameters>

**NEW QUESTION 47**

- (Exam Topic 3)

You have a dataset that includes home sales data for a city. The dataset includes the following columns.

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

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

Setting	Value
Prediction task	<div style="border: 1px solid black; padding: 5px;">                     Classification                      Forecasting                      Regression                      Outlier                 </div>
Target column	<div style="border: 1px solid black; padding: 5px;">                     Price                      Bedrooms                      Size                      HasGarage                      HomeType                 </div>

- A. Mastered
- B. Not Mastered

**Answer:** A

**Explanation:**

Box 1: Regression

Regression is a supervised machine learning technique used to predict numeric values. Box 2: Price

Reference:

<https://docs.microsoft.com/en-us/learn/modules/create-regression-model-azure-machine-learning-designer>

**NEW QUESTION 48**

- (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 49**

- (Exam Topic 3)

You plan to build a team data science environment. Data for training models in machine learning pipelines will be over 20 GB in size.

You have the following requirements:

- > Models must be built using Caffe2 or Chainer frameworks.
- > Data scientists must be able to use a data science environment to build the machine learning pipelines and train models on their personal devices in both connected and disconnected network environments.
- > Personal devices must support updating machine learning pipelines when connected to a network. You need to select a data science environment.

Which environment should you use?

- A. Azure Machine Learning Service
- B. Azure Machine Learning Studio
- C. Azure Databricks
- D. Azure Kubernetes Service (AKS)

**Answer:** A

**Explanation:**

The Data Science Virtual Machine (DSVM) is a customized VM image on Microsoft's Azure cloud built specifically for doing data science. Caffe2 and Chainer are supported by DSVM.

DSVM integrates with Azure Machine Learning.

**NEW QUESTION 50**

- (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 54**

- (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 55**

- (Exam Topic 3)

You use Azure Machine Learning to train and register a model.

You must deploy the model into production as a real-time web service to an inference cluster named service-compute that the IT department has created in the Azure Machine Learning workspace.

Client applications consuming the deployed web service must be authenticated based on their Azure Active Directory service principal.

You need to write a script that uses the Azure Machine Learning SDK to deploy the model. The necessary modules have been imported.

How should you complete the code? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.

```
# Assume the necessary modules have been imported
deploy_target = (ws, "service-compute")
AksCompute
AmlCompute
RemoteCompute
BatchCompute
deployment_config = .deploy_configuration(cpu_cores=1, memory_gb=1,
AksWebservice
AciWebservice
LocalWebService
)
token_auth_enabled=True
token_auth_enabled=False
auth_enabled=True
auth_enabled=False
service = Model.deploy(ws, "ml-service",
[model], inference_config, deployment_config, deploy_target)
service.wait_for_deployment(show_output = True)
```

- A. Mastered
- B. Not Mastered

**Answer:** A

**Explanation:**

Box 1: AksCompute Example:

```
aks_target = AksCompute(ws,"myaks")
```

# If deploying to a cluster configured for dev/test, ensure that it was created with enough  
 # cores and memory to handle this deployment configuration. Note that memory is also used by  
 # things such as dependencies and AML components.

```
deployment_config = AksWebservice.deploy_configuration(cpu_cores = 1, memory_gb = 1)
```

```
service = Model.deploy(ws, "myservice", [model], inference_config, deployment_config, aks_target)
```

Box 2: AksWebservice  
 Box 3: token\_auth\_enabled=Yes

Whether or not token auth is enabled for the Webservice.

Note: A Service principal defined in Azure Active Directory (Azure AD) can act as a principal on which authentication and authorization policies can be enforced in Azure Databricks.

The Azure Active Directory Authentication Library (ADAL) can be used to programmatically get an Azure AD access token for a user.

Reference:

<https://docs.microsoft.com/en-us/azure/machine-learning/how-to-deploy-azure-kubernetes-service> <https://docs.microsoft.com/en-us/azure/databricks/dev-tools/api/latest/aad/service-prin-aad-token>

**NEW QUESTION 56**

- (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 59**

- (Exam Topic 3)

You are using a Git repository to track work in an Azure Machine Learning workspace. You need to authenticate a Git account by using SSH.

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

**Actions**

**Answer Area**

- Generate a public/private key pair
- Add the private key to the Git account
- Clone the Git repository by using an SSH repository URL
- Add the public key to the Git account
- Create a new Azure Key Vault resource



- A. Mastered
- B. Not Mastered

**Answer:** A

**Explanation:**

Graphical user interface, text, application, chat or text message Description automatically generated  
 Authenticate your Git Account with SSH: Step 1: Generating a public/private key pair Generate a new SSH key  
 \* 1. Open the terminal window in the Azure Machine Learning Notebook Tab.  
 \* 2. Paste the text below, substituting in your email address. `ssh-keygen -t rsa -b 4096 -C "your_email@example.com"`  
 This creates a new ssh key, using the provided email as a label.  
 > Generating public/private rsa key pair.  
 Step 2: Add the public key to the Git Account  
 In your terminal window, copy the contents of your public key file. Step 3: Clone the Git repository by using an SSH repository URL  
 \* 1. Copy the SSH Git clone URL from the Git repo.  
 \* 2. Paste the url into the git clone command below, to use your SSH Git repo URL. This will look something like:  
`git clone git@example.com:GitUser/azureml-example.git Cloning into 'azureml-example'.`  
 Reference:  
<https://docs.microsoft.com/en-us/azure/machine-learning/concept-train-model-git-integration>

**NEW QUESTION 63**

- (Exam Topic 3)  
 You are performing clustering by using the K-means algorithm. You need to define the possible termination conditions.  
 Which three conditions can you use? Each correct answer presents a complete solution. NOTE: Each correct selection is worth one point.

- A. A fixed number of iterations is executed.
- B. The residual sum of squares (RSS) rises above a threshold.
- C. The sum of distances between centroids reaches a maximum.
- D. The residual sum of squares (RSS) falls below a threshold.
- E. Centroids do not change between iterations.

**Answer:** ADE

**Explanation:**

References:  
<https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/k-means-clustering> <https://nlp.stanford.edu/IR-book/html/htmledition/k-means-1.html>

**NEW QUESTION 64**

- (Exam Topic 3)  
 You plan to deliver a hands-on workshop to several students. The workshop will focus on creating data visualizations using Python. Each student will use a device that has internet access.  
 Student devices are not configured for Python development. Students do not have administrator access to install software on their devices. Azure subscriptions are not available for students.  
 You need to ensure that students can run Python-based data visualization code. Which Azure tool should you use?

- A. Anaconda Data Science Platform
- B. Azure BatchAI
- C. Azure Notebooks
- D. Azure Machine Learning Service

**Answer:** C

**Explanation:**

References:  
<https://notebooks.azure.com/>

**NEW QUESTION 69**

- (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 71**

- (Exam Topic 3)

You previously deployed a model that was trained using a tabular dataset named training-dataset, which is based on a folder of CSV files. Over time, you have collected the features and predicted labels generated by the model in a folder containing a CSV file for each month. You have created two tabular datasets based on the folder containing the inference data: one named predictions-dataset with a schema that matches the training data exactly, including the predicted label; and another named features-dataset with a schema containing all of the feature columns and a timestamp column based on the filename, which includes the day, month, and year. You need to create a data drift monitor to identify any changing trends in the feature data since the model was trained. To accomplish this, you must define the required datasets for the data drift monitor. Which datasets should you use to configure the data drift monitor? To answer, drag the appropriate datasets to the correct data drift monitor options. Each source 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.

Target datasets	Answer Area
training-dataset	Baseline dataset
predictions-dataset	Target dataset
features-dataset	Target dataset

- A. Mastered
- B. Not Mastered

**Answer: A**

**Explanation:**

Text Description automatically generated with medium confidence  
 Box 1: training-dataset  
 Baseline dataset - usually the training dataset for a model. Box 2: predictions-dataset  
 Target dataset - usually model input data - is compared over time to your baseline dataset. This comparison means that your target dataset must have a timestamp column specified.  
 The monitor will compare the baseline and target datasets. Reference:  
<https://docs.microsoft.com/en-us/azure/machine-learning/how-to-monitor-datasets>

**NEW QUESTION 73**

- (Exam Topic 3)

You are using a decision tree algorithm. You have trained a model that generalizes well at a tree depth equal to 10. You need to select the bias and variance properties of the model with varying tree depth values. Which properties should you select for each tree depth? To answer, select the appropriate options in the answer area.

**Answer Area**

Tree Depth	Bias	Variance
5	<input type="text"/> High Low Identical	<input type="text"/> High Low Identical
15	<input type="text"/> High Low Identical	<input type="text"/> High Low Identical

- A. Mastered
- B. Not Mastered

Answer: A

**Explanation:**

In decision trees, the depth of the tree determines the variance. A complicated decision tree (e.g. deep) has low bias and high variance.  
 Note: In statistics and machine learning, the bias–variance tradeoff is the property of a set of predictive models whereby models with a lower bias in parameter estimation have a higher variance of the parameter estimates across samples, and vice versa. Increasing the bias will decrease the variance. Increasing the variance will decrease the bias.  
 References:  
<https://machinelearningmastery.com/gentle-introduction-to-the-bias-variance-trade-off-in-machine-learning/>

**NEW QUESTION 74**

- (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 79**

- (Exam Topic 3)

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

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

Does the solution meet the goal?

- A. Yes
- B. No

Answer: B

**Explanation:**

The scikit-learn estimator provides a simple way of launching a scikit-learn training job on a compute target. It is implemented through the SKLearn class, which can be used to support single-node CPU training.

Example:

```
from azureml.train.sklearn import SKLearn
}
estimator = SKLearn(source_directory=project_folder, compute_target=compute_target, entry_script='train_iris.py'
)
```

Reference:

<https://docs.microsoft.com/en-us/azure/machine-learning/how-to-train-scikit-learn>

**NEW QUESTION 83**

- (Exam Topic 3)

You register a file dataset named csvfolder that references a folder. The folder includes multiple comma-separated values (CSV) files in an Azure storage blob container. You plan to use the following code to run a script that loads data from the file dataset. You create and instantiate the following variables:

Variable	Description
remote_cluster	References the Azure Machine Learning compute cluster
ws	References the Azure Machine Learning workspace

You have the following code:

```
from azureml.train.estimator import Estimator
file_dataset = ws.datasets.get('csv_folder')
estimator = Estimator(source_directory=script_folder,

compute_target = remote_cluster,
entry_script='script.py')
run = experiment.submit(config=estimator)
run.wait_for_completion(show_output=True)
```

You need to pass the dataset to ensure that the script can read the files it references. Which code segment should you insert to replace the code comment?

A)

```
inputs=[file_dataset.as_named_input('training_files').to_pandas_dataframe()],
```

B)

```
inputs=[file_dataset.as_named_input('training_files').as_mount()],
```

C)

```
script_params={'--training_files': file_dataset},
```

D)

```
inputs=[file_dataset.as_named_input('training_files')],
```

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

**Answer: D**

**Explanation:**

Example:

```
from azureml.train.estimator import Estimator
script_params = {
# to mount files referenced by mnist dataset
'--data-folder': mnist_file_dataset.as_named_input('mnist_opendataset').as_mount(),
'--regularization': 0.5
}
est = Estimator(source_directory=script_folder, script_params=script_params, compute_target=compute_target, environment_definition=env,
entry_script='train.py')
```

Reference:

<https://docs.microsoft.com/en-us/azure/machine-learning/tutorial-train-models-with-aml>

**NEW QUESTION 84**

- (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 89**

- (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 92**

- (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(n_components=10)
X_train = pca.fit_transform(X_train)
x_test = pca.transform(x_test)

```

The screenshot shows a code editor with the following code and three dropdown menus highlighting parts of it:

- Box 1:** Highlights the PCA class instantiation: `PCA(n_components = 10)`.
- Box 2:** Highlights the `.fit_transform(X_train)` method call.
- Box 3:** Highlights the `transform(x_test)` method call.

- 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 96**

- (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 98**

- (Exam Topic 3)

Note: This question is part of a series of questions that present the same scenario. Each question in the series contains a unique solution that might meet the stated goals. Some question sets might have more than one correct solution, while others might not have a correct solution. After you answer a question in this section, you will NOT be able to return to it. As a result, these questions will not appear in the review screen. You are creating a model to predict the price of a student’s artwork depending on the following variables: the student’s length of education, degree type, and art form. You start by creating a linear regression model. You need to evaluate the linear regression model. Solution: Use the following metrics: Mean Absolute Error, Root Mean Absolute Error, Relative Absolute Error, Relative Squared Error, and the Coefficient of Determination. Does the solution meet the goal?

- A. Yes
- B. No

Answer: A

**Explanation:**

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

**NEW QUESTION 101**

- (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 103**

- (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.

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 108**

- (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,M,f2,l

1,1,2,0  
 2,1,1,1  
 32,10

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/*/*.csv')
training_data = Dataset.Tabular.from_delimited_files(paths)
```

Does the solution meet the goal?

- A. Yes
- B. No

Answer: A

**NEW QUESTION 113**

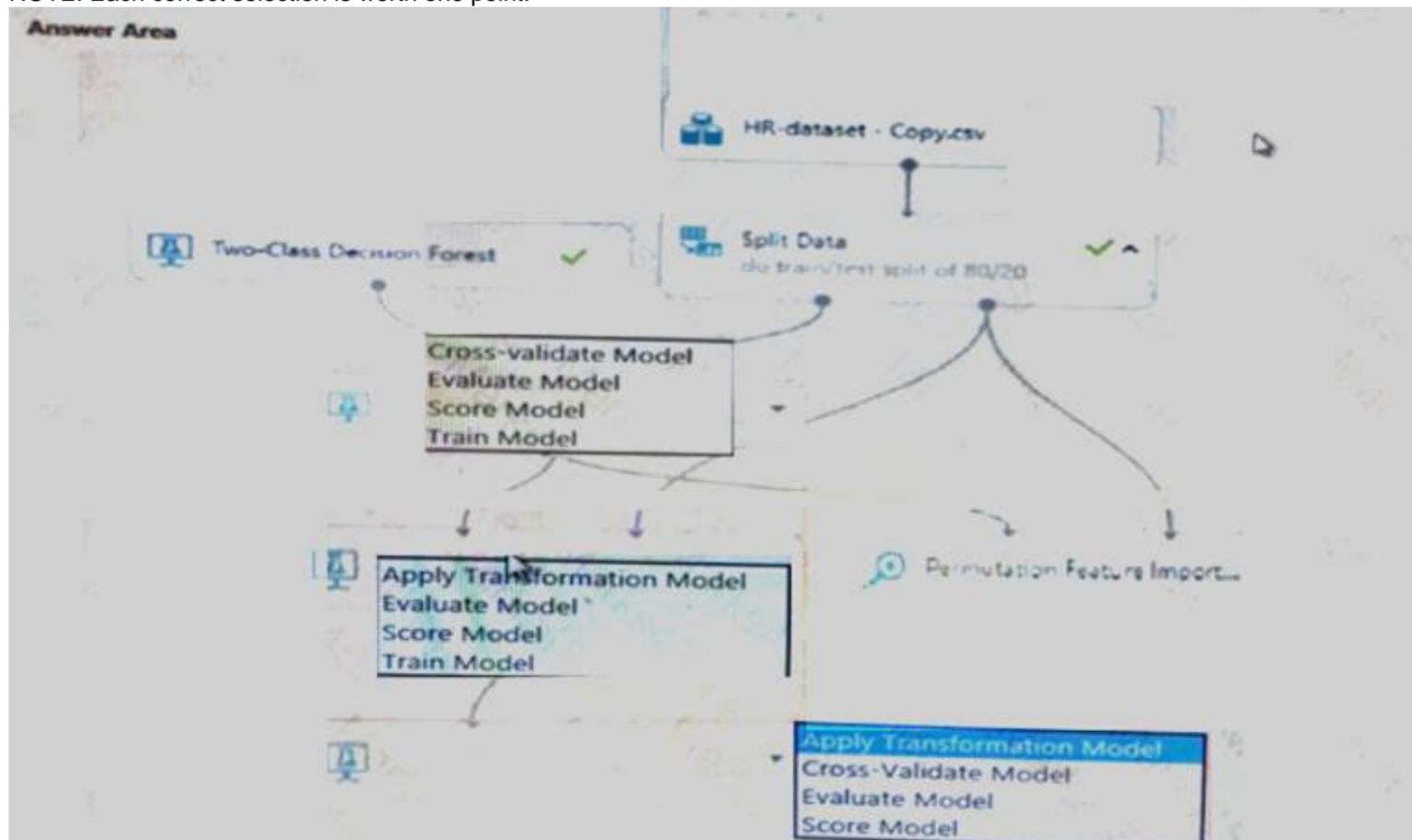
- (Exam Topic 3)

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.

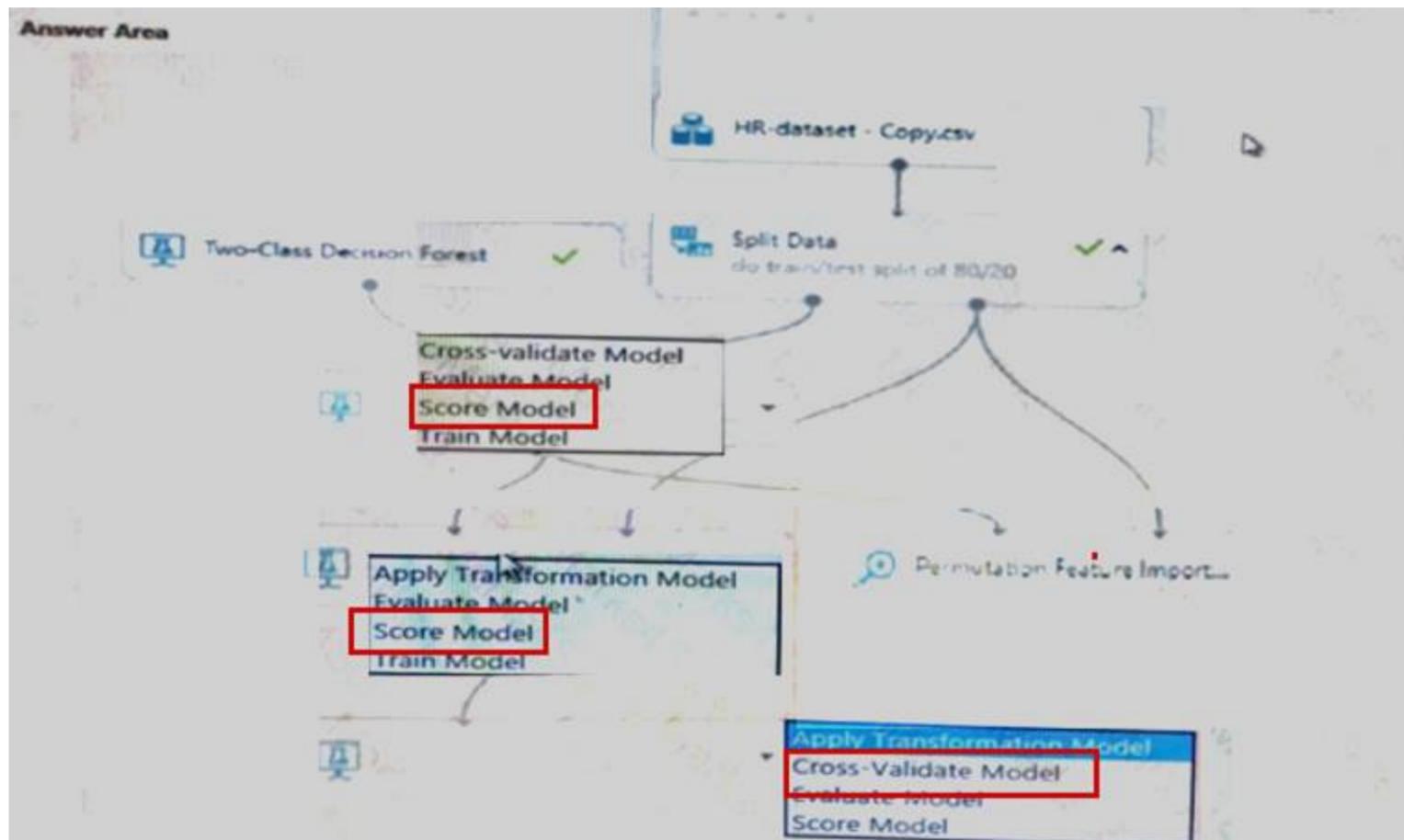
NOTE: Each correct selection is worth one point.



- A. Mastered
- B. Not Mastered

Answer: A

Explanation:



**NEW QUESTION 114**

- (Exam Topic 3)

You deploy a model as an Azure Machine Learning real-time web service using the following code.

```
# ws, model, inference_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 119**

- (Exam Topic 3)

You write code to retrieve an experiment that is run from your Azure Machine Learning workspace.

The run used the model interpretation support in Azure Machine Learning to generate and upload a model explanation.

Business managers in your organization want to see the importance of the features in the model.

You need to print out the model features and their relative importance in an output that looks similar to the following.

Feature	Importance
0	1.5627435610083558
2	0.6077689312583112
4	0.5574002432900718
3	0.42858759955671777
1	0.3501361539771977

How should you complete the code? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.

```
# Assume required modules are imported
```

```
ws = Workspace.from_config()
feature_importances = explanation.
```

	( workspace = ws,
from_run	experiment_name='train_and_explain',
list_model_explanations	run_id='train_and_explain_12345')
from_run_id	
download_model_explanation	

```
explanation = client.
```

	( )
upload_model_explanation	
list_model_explanations	
run	
download_model_explanation	

```
feature_importances = explanation.
```

	( )
explanation	
explanation_client	
get_feature_important_dict	
download_model_explanation	

```
for key, value in feature_importances.items():
    print(key, "\t", value)
```

- A. Mastered
- B. Not Mastered

**Answer:** A

**Explanation:**

Box 1: from\_run\_id

from\_run\_id(workspace, experiment\_name, run\_id) Create the client with factory method given a run ID. Returns an instance of the explanations Client.

Parameters

- > Workspace Workspace An object that represents a workspace.
- > experiment\_name str The name of an experiment.
- > run\_id str A GUID that represents a run.

Box 2: list\_model\_explanations

list\_model\_explanations returns a dictionary of metadata for all model explanations available.

Returns

A dictionary of explanation metadata such as id, data type, explanation: method, model type, and upload time, sorted by upload time

Box 3: explanation:

Reference:

<https://docs.microsoft.com/en-us/python/api/azureml-contrib-interpret/azureml.contrib.interpret>.

**NEW QUESTION 123**

- (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 127**

- (Exam Topic 3)

You are planning to register a trained model in an Azure Machine Learning workspace.

You must store additional metadata about the model in a key-value format. You must be able to add new metadata and modify or delete metadata after creation.

You need to register the model. Which parameter should you use?

- A. description
- B. model\_framework
- C. cags
- D. properties

Answer: D

**Explanation:**

azureml.core.Model.properties:

Dictionary of key value properties for the Model. These properties cannot be changed after registration, however new key value pairs can be added.

Reference:

<https://docs.microsoft.com/en-us/python/api/azureml-core/azureml.core.model.model>

**NEW QUESTION 128**

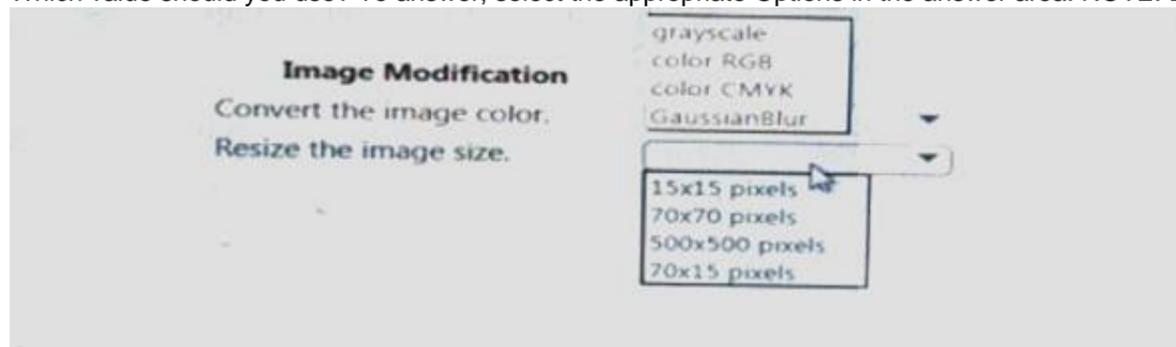
- (Exam Topic 3)

You are training a deep learning model to identify cats and dogs. You have 25,000 color images. You must meet the following requirements:

- Reduce the number of training epochs.
- Reduce the size of the neural network.
- Reduce over-fitting of the neural network.

You need to select the image modification values.

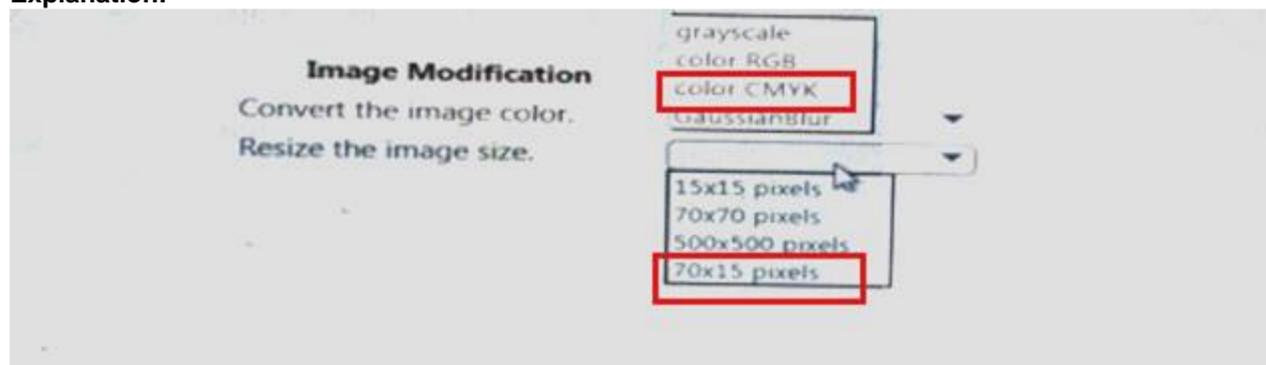
Which value should you use? To answer, select the appropriate Options in the answer area. NOTE: Each correct selection is worth one point.



- A. Mastered
- B. Not Mastered

Answer: A

**Explanation:**



**NEW QUESTION 129**

- (Exam Topic 3)

You are analyzing a dataset containing historical data from a local taxi company. You are developing 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 close 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 130**

- (Exam Topic 3)

You define a datastore named ml-data for an Azure Storage blob container. In the container, you have a folder named train that contains a file named data.csv.

You plan to use the file to train a model by using the Azure Machine Learning SDK.

You plan to train the model by using the Azure Machine Learning SDK to run an experiment on local compute.

You define a DataReference object by running the following code:

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

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

- A. 

```
import os
import argparse
import pandas as pd

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

```
import os
import argparse
import pandas as pd

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

```
import pandas as pd

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

```
import os
import argparse
import pandas as pd

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

```
import os
import argparse
import pandas as pd

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

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

**Answer:** E

**Explanation:**

Example:  

```
data_folder = args.data_folder
# Load Train and Test data
train_data = pd.read_csv(os.path.join(data_folder, 'data.csv'))
```

Reference:  
<https://www.element61.be/en/resource/azure-machine-learning-services-complete-toolbox-ai>

**NEW QUESTION 134**

- (Exam Topic 3)

You use the Azure Machine Learning SDK to run a training experiment that trains a classification model and calculates its accuracy metric. The model will be retrained each month as new data is available. You must register the model for use in a batch inference pipeline. You need to register the model and ensure that the models created by subsequent retraining experiments are registered only if their accuracy is higher than the currently registered model. 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. Specify a different name for the model each time you register it.
- B. Register the model with the same name each time regardless of accuracy, and always use the latest version of the model in the batch inferencing pipeline.
- C. Specify the model framework version when registering the model, and only register subsequent models if this value is higher.
- D. Specify a property named accuracy with the accuracy metric as a value when registering the model, and only register subsequent models if their accuracy is higher than the accuracy property value of the currently registered model.
- E. Specify a tag named accuracy with the accuracy metric as a value when registering the model, and only register subsequent models if their accuracy is higher than the accuracy tag value of the currently registered mode

**Answer:** CE

**Explanation:**

E: Using tags, you can track useful information such as the name and version of the machine learning library used to train the model. Note that tags must be alphanumeric.

Reference:

<https://notebooks.azure.com/xavierheriat/projects/azureml-getting-started/html/how-to-use-azureml/deployment/>

**NEW QUESTION 139**

- (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 140**

- (Exam Topic 3)

You must store data in Azure Blob Storage to support Azure Machine Learning. You need to transfer the data into Azure Blob Storage.

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

NOTE: Each correct selection is worth one point.

- A. Bulk Insert SQL Query
- B. AzCopy
- C. Python script
- D. Azure Storage Explorer
- E. Bulk Copy Program (BCP)

**Answer:** BCD

**Explanation:**

You can move data to and from Azure Blob storage using different technologies: Azure Storage-Explorer

AzCopy Python SSIS

References:

<https://docs.microsoft.com/en-us/azure/machine-learning/team-data-science-process/move-azure-blob>

**NEW QUESTION 145**

- (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 150**

- (Exam Topic 3)

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

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

Answer: D

**Explanation:**

Azure Machine Learning compute cluster is a managed-compute infrastructure that allows you to easily create a single or multi-node compute. The compute is created within your workspace region as a resource that can be shared with other users in your workspace. The compute scales up automatically when a job is submitted, and can be put in an Azure Virtual Network.

Reference:

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

**NEW QUESTION 152**

- (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 156

- (Exam Topic 3)

You are building recurrent neural network to perform a binary classification.

The training loss, validation loss, training accuracy, and validation accuracy of each training epoch has been provided. You need to identify whether the classification model is over fitted.

Which of the following is correct?

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

**Answer:** B

#### Explanation:

An overfit model is one where performance on the train set is good and continues to improve, whereas performance on the validation set improves to a point and then begins to degrade.

References:

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

#### NEW QUESTION 159

- (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 162

- (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 165

- (Exam Topic 3)

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

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

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

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

All files store data in the following format: id,f1,f2i

```
1,1.2,0
2,1,1,
1 3,2.1,0
```

You run the following code:

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

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

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

Solution: Run the following code:

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

Does the solution meet the goal?

- A. Yes
- B. No

**Answer: B**

**Explanation:**

Use two file paths.

Use Dataset.Tabular\_from\_delimited, instead of Dataset.File.from\_files as the data isn't cleansed. Reference:  
<https://docs.microsoft.com/en-us/azure/machine-learning/how-to-create-register-datasets>

**NEW QUESTION 169**

- (Exam Topic 3)

You plan to use the Hyperdrive feature of Azure Machine Learning to determine the optimal hyperparameter values when training a model.

You must use Hyperdrive to try combinations of the following hyperparameter values:

- learning\_rate: any value between 0.001 and 0.1
- batch\_size: 16, 32, or 64

You need to configure the search space for the Hyperdrive experiment.

Which two parameter expressions should you use? Each correct answer presents part of the solution. NOTE: Each correct selection is worth one point.

- A. a choice expression for learning\_rate
- B. a uniform expression for learning\_rate
- C. a normal expression for batch\_size
- D. a choice expression for batch\_size
- E. a uniform expression for batch\_size

**Answer: BD**

**Explanation:**

B: Continuous hyperparameters are specified as a distribution over a continuous range of values. Supported distributions include:

➤ uniform(low, high) - Returns a value uniformly distributed between low and high

D: Discrete hyperparameters are specified as a choice among discrete values. choice can be:

➤ one or more comma-separated values

➤ a range object

➤ any arbitrary list object Reference:

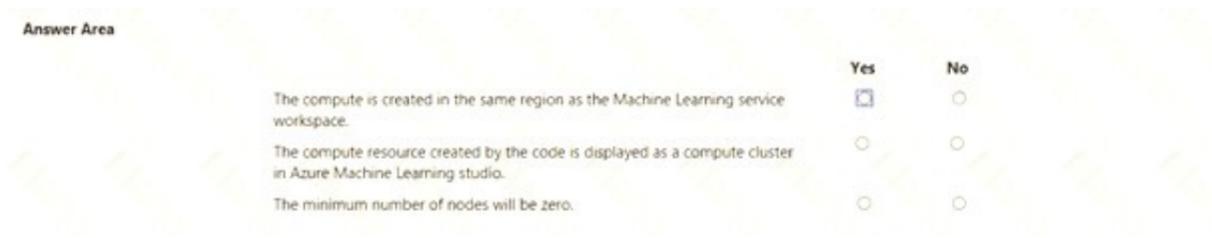
<https://docs.microsoft.com/en-us/azure/machine-learning/how-to-tune-hyperparameters>

**NEW QUESTION 170**

- (Exam Topic 3)

You are using Azure Machine Learning to train machine learning models. You need a compute target on which to remotely run the training script. You run the following Python code:

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



- A. Mastered
- B. Not Mastered

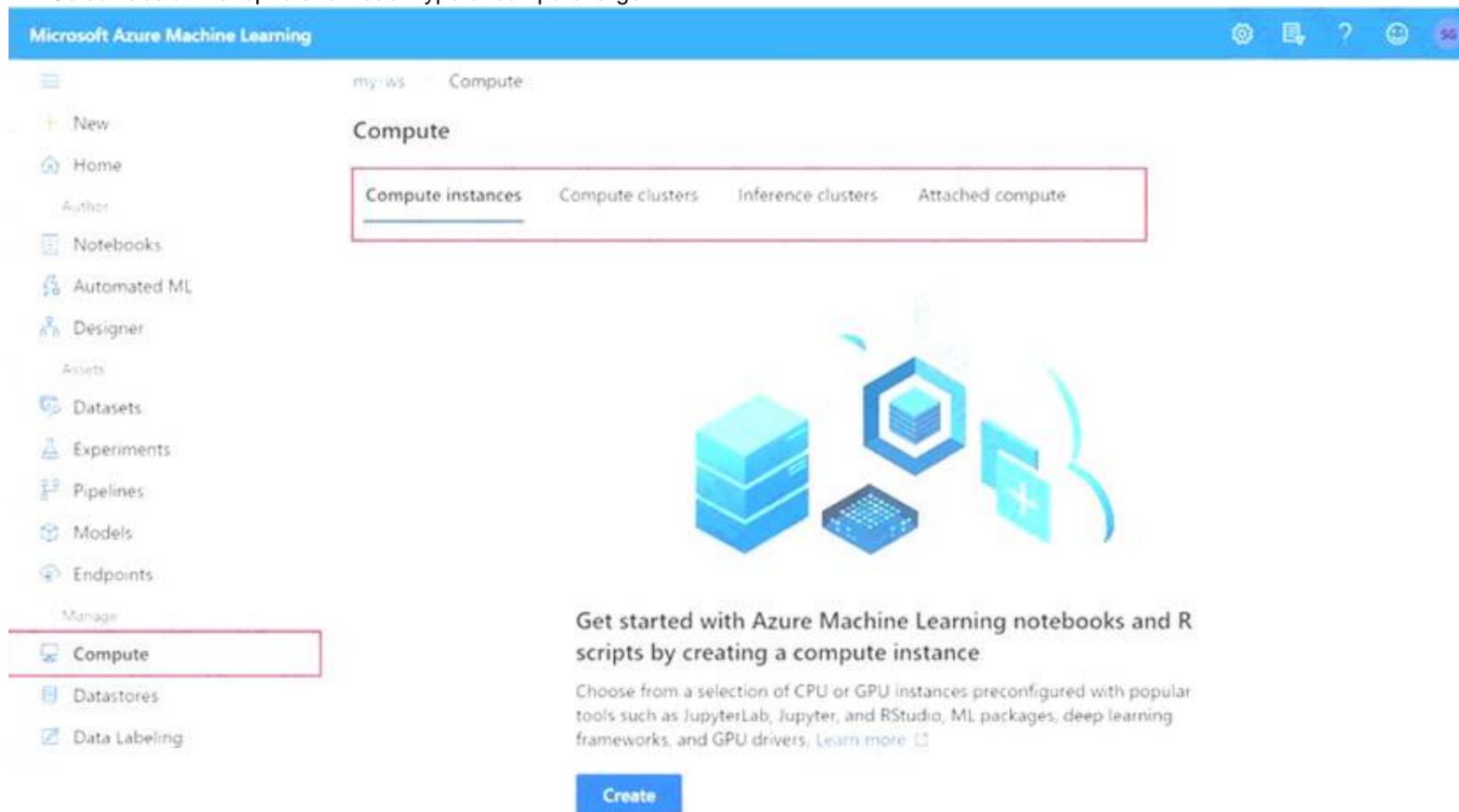
Answer: A

**Explanation:**

Box 1: Yes

The compute is created within your workspace region as a resource that can be shared with other users. Box 2: Yes  
 It is displayed as a compute cluster. View compute targets

- \* 1. To see all compute targets for your workspace, use the following steps:
- \* 2. Navigate to Azure Machine Learning studio.
- \* 3. Under Manage, select Compute.
- \* 4. Select tabs at the top to show each type of compute target.



Box 3: Yes

min\_nodes is not specified, so it defaults to 0. Reference:

<https://docs.microsoft.com/en-us/python/api/azureml-core/azureml.core.compute.amlcompute.amlcomputeprovi> <https://docs.microsoft.com/en-us/azure/machine-learning/how-to-create-attach-compute-studio>

**NEW QUESTION 174**

- (Exam Topic 3)

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

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

You plan to use a Python script to run an Azure Machine Learning experiment. The script creates a reference to the experiment run context, loads data from a file, identifies the set of unique values for the label column, and completes the experiment run:

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

The experiment must record the unique labels in the data as metrics for the run that can be reviewed later.

You must add code to the script to record the unique label values as run metrics at the point indicated by the comment.

Solution: Replace the comment with the following code:

```
run.upload_file('outputs/labels.csv', './data.csv')
```

Does the solution meet the goal?

- A. Yes
- B. No

Answer: B

**Explanation:**

label\_vals has the unique labels (from the statement label\_vals = data['label'].unique()), and it has to be logged.

Note:

Instead use the run\_log function to log the contents in label\_vals: for label\_val in label\_vals:

```
run.log('Label Values', label_val)
```

Reference:

<https://www.element61.be/en/resource/azure-machine-learning-services-complete-toolbox-ai>

**NEW QUESTION 177**

- (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 178**

- (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**

**Answer Area**

Scenario	Environment
Run an Azure Machine Learning Designer training pipeline.	<input type="text" value="Environment"/>
Deploying a web service from the Azure Machine Learning designer.	<input type="text" value="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 182**

- (Exam Topic 3)

You plan to run a Python script as an Azure Machine Learning experiment. The script contains the following code:

```
import os, argparse, glob
from azureml.core import Run
parser = argparse.ArgumentParser()
parser.add_argument('--input-data', type=str, dest='data_folder')
args = parser.parse_args()
data_path = args.data_folder
file_paths = glob.glob(data_path + "/*.jpg")
```

You must specify a file dataset as an input to the script. The dataset consists of multiple large image files and must be streamed directly from its source.

You need to write code to define a ScriptRunConfig object for the experiment and pass the ds dataset as an argument.

Which code segment should you use?

- A. arguments = ['--input-data', ds.to\_pandas\_dataframe()]
- B. arguments = ['--input-data', ds.as\_mount()]
- C. arguments = ['--data-data', ds]
- D. arguments = ['--input-data', ds.as\_download()]

**Answer: A**

**Explanation:**

If you have structured data not yet registered as a dataset, create a TabularDataset and use it directly in your training script for your local or remote experiment.

To load the TabularDataset to pandas DataFrame df = dataset.to\_pandas\_dataframe()

Note: TabularDataset represents data in a tabular format created by parsing the provided file or list of files. Reference:

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

**NEW QUESTION 187**

- (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	Tool
Vowpal Wabbit	Build DNN models	Tool
PowerBI Desktop	Enable interactive data exploration and visualization	Tool
Azure Data Factory		
Microsoft Cognitive Toolkit		

- 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/machine-learning/studio-module-reference/train-vowpal-wabbit-version-8) <https://docs.microsoft.com/en-us/azure/architecture/data-guide/scenarios/interactive-data-exploration>

**NEW QUESTION 189**

- (Exam Topic 3)

You are a data scientist building a deep convolutional neural network (CNN) for image classification. The CNN model you built shows signs of overfitting. You need to reduce overfitting and converge the model to an optimal fit.

Which two actions should you perform? Each correct answer presents a complete solution.

NOTE: Each correct selection is worth one point.

- A. Reduce the amount of training data.
- B. Add an additional dense layer with 64 input units
- C. Add L1/L2 regularization.
- D. Use training data augmentation
- E. Add an additional dense layer with 512 input units.

**Answer:** AC

**Explanation:**

References:

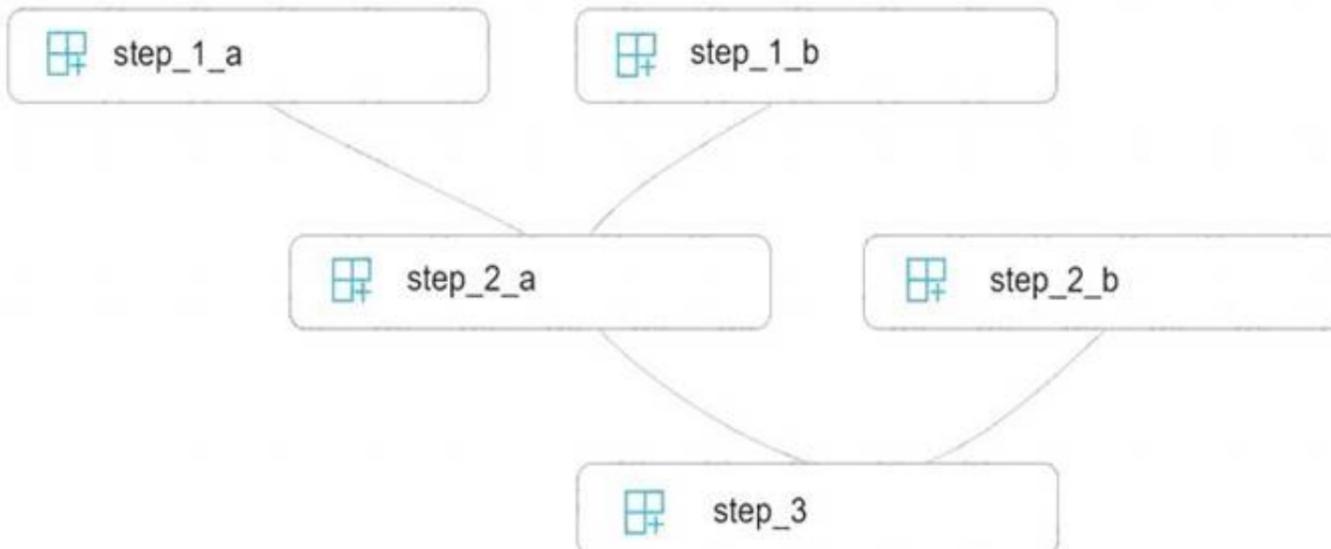
<https://machinelearningmastery.com/how-to-reduce-overfitting-in-deep-learning-with-weight-regularization/>  
[https://en.wikipedia.org/wiki/Convolutional\\_neural\\_network](https://en.wikipedia.org/wiki/Convolutional_neural_network)

**NEW QUESTION 193**

- (Exam Topic 3)

You write five Python scripts that must be processed in the order specified in Exhibit A – which allows the same modules to run in parallel, but will wait for modules with dependencies.

You must create an Azure Machine Learning pipeline using the Python SDK, because you want to script to create the pipeline to be tracked in your version control system. You have created five PythonScriptSteps and have named the variables to match the module names.



You need to create the pipeline shown. Assume all relevant imports have been done. Which Python code segment should you use?

- A. `p = Pipeline(ws, steps=[[[[step_1_a, step_1_b], step_2_a], step_2_b], step_3])`
- B. 

```
pipeline_steps = {
    "Pipeline": {
        "run": step_3,
        "run_after": [{
            {"run": step_2_a,
             "run_after":
                [{"run": step_1_a},
                 {"run": step_1_b}]
            },
            {"run": step_2_b}
        ]
    }
}
p = Pipeline(ws, steps=pipeline_steps)
```
- C. `step_2_a.run_after(step_1_b)`  
`step_2_a.run_after(step_1_a)`  
`step_3.run_after(step_2_b)`  
`step_3.run_after(step_2_a)`  
`p = Pipeline(ws, steps=[step_3])`
- D. `p = Pipeline(ws, steps=[step_1_a, step_1_b, step_2_a, step_2_b, step_3])`

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

**Answer:** A

**Explanation:**

The steps parameter is an array of steps. To build pipelines that have multiple steps, place the steps in order in this array.

Reference:

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

**NEW QUESTION 197**

- (Exam Topic 3)

You train a classification model by using a decision tree algorithm.

You create an estimator by running the following Python code. The variable `feature_names` is a list of all feature names, and `class_names` is a list of all class names.

```
from interpret.ext.blackbox import TabularExplainer
```

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

You need to explain the predictions made by the model for all classes by determining the importance of all features.

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

NOTE: Each correct selection is worth one point.

	Yes	No
The SHAP TreeExplainer will be used to interpret the model.	<input type="radio"/>	<input type="radio"/>
If you omit the features and classes parameters in the TabularExplainer instantiation, the explainer still works as expected.	<input type="radio"/>	<input type="radio"/>
You could interpret the model by using a MimicExplainer instead of a TabularExplainer.	<input type="radio"/>	<input type="radio"/>

- A. Mastered
- B. Not Mastered

**Answer:** A

**Explanation:**

and visualizations more informative, you can choose to pass in feature names and output class names if doing classification.

Box 3: No

TabularExplainer automatically selects the most appropriate one for your use case, but you can call each of its three underlying explainers underneath (TreeExplainer, DeepExplainer, or KernelExplainer) directly.

Reference:

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

**NEW QUESTION 198**

- (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 203**

- (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	Requirement	Compute type
Attached compute	Train models by using the Azure Machine Learning designer.	Compute type
Inference cluster	Score new data through a trained model published as a real-time web service.	Compute type
Training cluster	Train models by using an Azure Databricks cluster.	Compute type
	Deploy models by using the Azure Machine Learning designer.	Compute type

- 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
<b>Azure Machine Learning compute instance</b>	yes & hyperparameter tuning	yes	<b>yes</b>

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

**NEW QUESTION 206**

- (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 209**

- (Exam Topic 3)

You have a model with a large difference between the training and validation error values. You must create a new model and perform cross-validation. You need to identify a parameter set for the new model using Azure Machine Learning Studio.

Which module you should use for each step? To answer, drag the appropriate modules to the correct steps. Each module may be used once or 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	Step	Module
Two-Class Boosted Decision Tree	Define the parameter scope	
Partition and Sample	Define the cross-validation settings	
Tune Model Hyperparameters	Define the metric	
Split Data	Train, evaluate, and compare	

- A. Mastered
- B. Not Mastered

**Answer: A**

**Explanation:**

Box 1: Split data  
 Box 2: Partition and Sample  
 Box 3: Two-Class Boosted Decision Tree Box 4: Tune Model Hyperparameters  
 Integrated train and tune: You configure a set of parameters to use, and then let the module iterate over multiple combinations, measuring accuracy until it finds a "best" model. With most learner modules, you can choose which parameters should be changed during the training process, and which should remain fixed. We recommend that you use Cross-Validate Model to establish the goodness of the model given the specified parameters. Use Tune Model Hyperparameters to identify the optimal parameters. References: <https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/partition-and-sample>

**NEW QUESTION 214**

- (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 219**

- (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 223**

- (Exam Topic 3)

You create an Azure Machine Learning workspace and set up a development environment. You plan to train a deep neural network (DNN) by using the Tensorflow framework and by using estimators to submit training scripts.

You must optimize computation speed for training runs.

You need to choose the appropriate estimator to use as well as the appropriate training compute target configuration.

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

Answer Area

Parameter	Value
Estimator	<div style="border: 1px solid black; padding: 5px;">                     Estimator                      SKLearn                      PyTorch                      Tensorflow                      Chainer                 </div>
Training compute	<div style="border: 1px solid black; padding: 5px;">                     12 vCPU, 48 GB memory, 96 GB SSD                      12 vCPU, 112 GB memory, 680 GB SSD, 2 GPU, 24 GB GPU memory                      16 vCPU, 128 GB memory, 160 GB HDD, 80 GB NVME disk (4000 MBps)                      44 vCPU, 352 GB memory, 3.4 GHz CPU frequency all cores                 </div>

- A. Mastered
- B. Not Mastered

**Answer: A**

**Explanation:**

Box 1: Tensorflow

TensorFlow represents an estimator for training in TensorFlow experiments. Box 2: 12 vCPU, 112 GB memory...,2 GPU,..

Use GPUs for the deep neural network. Reference:

<https://docs.microsoft.com/en-us/python/api/azureml-train-core/azureml.train.dnn>

**NEW QUESTION 224**

- (Exam Topic 3)

You create an Azure Machine Learning workspace named workspace1. You assign a custom role to a user of workspace1.

The custom role has the following JSON definition:

```
{
  "Name": "MyRole",
  "IsCustom": true,
  "Description": "New custom role description.",
  "Actions": ["*"],
  "NotActions": [
    "Microsoft.MachineLearningServices/workspaces/write",
    "Microsoft.MachineLearningServices/workspaces/computes/*/write",
    "Microsoft.MachineLearningServices/workspaces/computes/*/delete",
    "Microsoft.Authorization/*/write"
  ],
  "AssignableScopes": [
    "/subscriptions/<subscription_id>/resourceGroups/resourcegroup1/providers/Microsoft.MachineLearningServices/workspaces/workpace1"
  ]
}
```

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.

Statements	Yes	No
The user can perform all actions in the workspace	<input type="radio"/>	<input type="radio"/>
The user can delete a compute resource in the workspace	<input type="radio"/>	<input type="radio"/>
The user can write metrics to the workspace	<input type="radio"/>	<input type="radio"/>

A.

**Answer:**

**Explanation:**

Graphical user interface, text, application, email Description automatically generated

Box 1: No

The actions listed in NotActions are prohibited.

If the roles include Actions that have a wildcard (\*), the effective permissions are computed by subtracting the NotActions from the allowed Actions.

Box 2: No

Deleting compute resources in the workspace is in the NotActions list. Box 3: Yes

Writing metrics is not listed in NotActions. Reference:

<https://docs.microsoft.com/en-us/azure/role-based-access-control/overview#how-azure-rbac-determines-if-a-use>

**NEW QUESTION 227**

- (Exam Topic 3)

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

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

You are using Azure Machine Learning Studio to perform feature engineering on a dataset. You need to normalize values to produce a feature column grouped into bins.

Solution: Apply an Entropy Minimum Description Length (MDL) binning mode. Does the solution meet the goal?

A. Yes

B. No

**Answer:** A

**Explanation:**

Entropy MDL binning mode: This method requires that you select the column you want to predict and the column or columns that you want to group into bins. It then makes a pass over the data and attempts to determine the number of bins that minimizes the entropy. In other words, it chooses a number of bins that allows the data column to best predict the target column. It then returns the bin number associated with each row of your data in a column named <colname>quantized.

References:

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

**NEW QUESTION 231**

- (Exam Topic 3)

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

A. HDInsight

B. Azure Databricks

C. Azure Kubernetes Services

D. the existing Machine Learning Compute resource

E. a new Machine Learning Compute resource

**Answer:** C

**Explanation:**

Azure Kubernetes Service (AKS) can be used real-time inference. Reference:

<https://docs.microsoft.com/en-us/azure/machine-learning/concept-compute-target>

**NEW QUESTION 236**

- (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 237**

- (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 PFExplainer. Does the solution meet the goal?

A. Yes

B. No

**Answer:** A

**Explanation:**

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

Reference:

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

**NEW QUESTION 241**

- (Exam Topic 3)

You are performing a filter based feature selection for a dataset 10 build a multi class classifiers 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 242**

- (Exam Topic 3)

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

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

You plan to use a Python script to run an Azure Machine Learning experiment. The script creates a reference to the experiment run context, loads data from a file, identifies the set of unique values for the label column, and completes the experiment run:

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

The experiment must record the unique labels in the data as metrics for the run that can be reviewed later.

You must add code to the script to record the unique label values as run metrics at the point indicated by the comment.

Solution: Replace the comment with the following code: `run.log_table('Label Values', label_vals)`

Does the solution meet the goal?

A. Yes

B. No

**Answer:** B

**Explanation:**

Instead use the run\_log function to log the contents in label\_vals: for label\_val in label\_vals:  
 run.log('Label Values', label\_val) Reference:  
<https://www.element61.be/en/resource/azure-machine-learning-services-complete-toolbox-ai>

**NEW QUESTION 244**

- (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 245**

- (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 target
- 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 248**

- (Exam Topic 3)

You run an automated machine learning experiment in an Azure Machine Learning workspace. Information about the run is listed in the table below:

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

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

A)

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

B)

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

C)

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

D)

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

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

**Answer:** A

**Explanation:**

The get\_output method on automl\_classifier returns the best run and the fitted model for the last invocation. Overloads on get\_output allow you to retrieve the best run and fitted model for any logged metric or for a particular iteration.

In []:  
 best\_run, fitted\_model = local\_run.get\_output() Reference:  
<https://notebooks.azure.com/azureml/projects/azureml-getting-started/html/how-to-use-azureml/automated-mach>

**NEW QUESTION 250**

- (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 252**

- (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	<ul style="list-style-type: none"> <li>Gender, IsPlaySoccer</li> <li>Gender, PrevExamMarks, Height, Weight</li> <li>PrevExamMarks, Height, Weight</li> <li>IsPlaySoccer</li> </ul>
Continuous variables	<ul style="list-style-type: none"> <li>Gender, IsPlaySoccer</li> <li>Gender, PrevExamMarks, Height, Weight</li> <li>PrevExamMarks, Height, Weight</li> <li>IsPlaySoccer</li> </ul>

- A. Mastered
- B. Not Mastered

**Answer:** A

**Explanation:**

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

**NEW QUESTION 254**

- (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 257**

- (Exam Topic 3)

You are a data scientist creating a linear regression model.

You need to determine how closely the data fits the regression line. Which metric should you review?

- A. Coefficient of determination
- B. Recall
- C. Precision
- D. Mean absolute error
- E. Root Mean Square Error

**Answer:** A

**Explanation:**

Coefficient of determination, often referred to as R<sup>2</sup>, represents the predictive power of the model as a value between 0 and 1. Zero means the model is random (explains nothing); 1 means there is a perfect fit. However, caution should be used in interpreting R<sup>2</sup> values, as low values can be entirely normal and high values can be suspect.

References:

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

**NEW QUESTION 261**

- (Exam Topic 3)

You create a deep learning model for image recognition on Azure Machine Learning service using GPU-based training.

You must deploy the model to a context that allows for real-time GPU-based inferencing. You need to configure compute resources for model inferencing.

Which compute type should you use?

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

**Answer:** B

**Explanation:**

You can use Azure Machine Learning to deploy a GPU-enabled model as a web service. Deploying a model on Azure Kubernetes Service (AKS) is one option. The AKS cluster provides a GPU resource that is used by the model for inference. Inference, or model scoring, is the phase where the deployed model is used to make predictions. Using GPUs instead of CPUs offers performance advantages on highly parallelizable computation.

Reference:  
<https://docs.microsoft.com/en-us/azure/machine-learning/how-to-deploy-inferencing-gpus>

**NEW QUESTION 264**

- (Exam Topic 3)

You are using an Azure Machine Learning workspace. You set up an environment for model testing and an environment for production. The compute target for testing must minimize cost and deployment efforts. The compute target for production must provide fast response time, autoscaling of the deployed service, and support real-time inferencing.

You need to configure compute targets for model testing and production.

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

NOTE: Each correct selection is worth one point.

Environment	Compute target
Testing	<div style="border: 1px solid black; padding: 5px;"><div style="text-align: right; border-bottom: 1px solid black;">▼</div><div style="border-bottom: 1px solid black; padding: 2px;">Local web service</div><div style="border-bottom: 1px solid black; padding: 2px;">Azure Kubernetes Services (AKS)</div><div style="border-bottom: 1px solid black; padding: 2px;">Azure Container Instances</div><div style="padding: 2px;">Azure Machine Learning compute clusters</div></div>
Production	<div style="border: 1px solid black; padding: 5px;"><div style="text-align: right; border-bottom: 1px solid black;">▼</div><div style="border-bottom: 1px solid black; padding: 2px;">Local web service</div><div style="border-bottom: 1px solid black; padding: 2px;">Azure Kubernetes Services (AKS)</div><div style="border-bottom: 1px solid black; padding: 2px;">Azure Container Instances</div><div style="padding: 2px;">Azure Machine Learning compute clusters</div></div>

- A. Mastered
- B. Not Mastered

**Answer:** A

**Explanation:**

Text, application Description automatically generated

Box 1: Local web service

The Local web service compute target is used for testing/debugging. Use it for limited testing and troubleshooting. Hardware acceleration depends on use of libraries in the local system.

Box 2: Azure Kubernetes Service (AKS)

Azure Kubernetes Service (AKS) is used for Real-time inference. Recommended for production workloads.

Use it for high-scale production deployments. Provides fast response time and autoscaling of the deployed service

Reference:

<https://docs.microsoft.com/en-us/azure/machine-learning/concept-compute-target>

**NEW QUESTION 267**

.....

## 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