

# Microsoft

## Exam Questions DP-100

Designing and Implementing a Data Science Solution on Azure



**NEW QUESTION 1**

- (Exam Topic 3)

You have an Azure Machine Learning workspace that contains a training cluster and an inference cluster. You plan to create a classification model by using the Azure Machine Learning designer.

You need to ensure that client applications can submit data as HTTP requests and receive predictions as responses.

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

Actions

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

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

Deploy a service to the compute cluster.

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

Deploy a service to the inference cluster.

Answer area

- A. Mastered
- B. Not Mastered

**Answer:** A

**Explanation:**

Actions

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

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

Deploy a service to the compute cluster.

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

Deploy a service to the inference cluster.

Answer area

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

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

Deploy a service to the inference cluster.

**NEW QUESTION 2**

- (Exam Topic 3)

HOTSPOT

You create a script for training a machine learning model in Azure Machine Learning service. You create an estimator by running the following code:

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

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

NOTE: Each correct selection is worth one point.

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

- A. Mastered  
 B. Not Mastered

**Answer:** A

**Explanation:**

Box 1: Yes

Parameter source\_directory is a local directory containing experiment configuration and code files needed for a training job.

Box 2: Yes

script\_params is a dictionary of command-line arguments to pass to the training script specified in entry\_script.

Box 3: No

Box 4: Yes

The conda\_packages parameter is a list of strings representing conda packages to be added to the Python environment for the experiment.

**NEW QUESTION 3**

- (Exam Topic 3)

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

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

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

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

Solution: Run the following code:

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

Does the solution meet the goal?

- A. Yes  
 B. No

**Answer:** B

**Explanation:**

train\_step is missing. Reference:

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

**NEW QUESTION 4**

- (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)
model1 = 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	<div>Automatically select the performance metrics for the classification.</div> <div>Automatically adjust weights directly proportional to class frequencies in the input data.</div> <div>Automatically adjust weights inversely proportional to class frequencies in the input data.</div>
C parameter	<div>Penalty parameter</div> <div>Degree of polynomial kernel function</div> <div>Size of the kernel cache</div>

- 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  $n_{\text{samples}} / (n_{\text{classes}} * \text{np.bincount}(y))$ .  
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 5**

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

Action	Answer area
Evaluate the probability function	
Remove duplicate rows	
Use the Filter Based Feature Selection module	<div>⬅️</div> <div>➡️</div>
Test the hypothesis using t-Test	<div>⬆️</div> <div>⬇️</div>
Compute linear correlation	
Build a counting transform	

- A. Mastered
- B. Not Mastered

Answer: A

**Explanation:**

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

**NEW QUESTION 6**

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

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

**Answer: C**

**Explanation:**

Differential privacy tries to protect against the possibility that a user can produce an indefinite number of reports to eventually reveal sensitive data. A value known as epsilon measures how noisy, or private, a report is. Epsilon has an inverse relationship to noise or privacy. The lower the epsilon, the more noisy (and private) the data is.

Reference:

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

**NEW QUESTION 7**

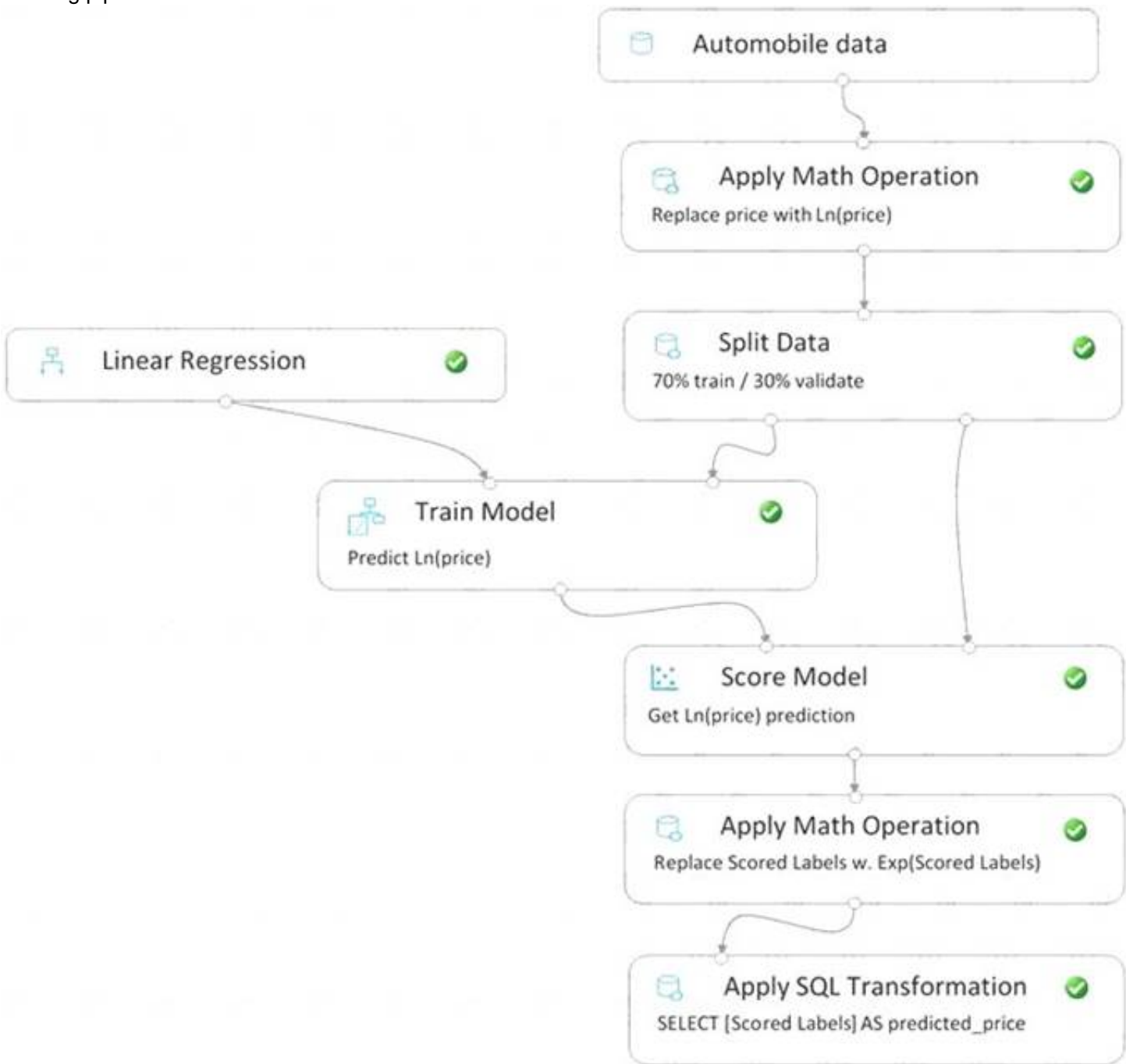
- (Exam Topic 3)

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.

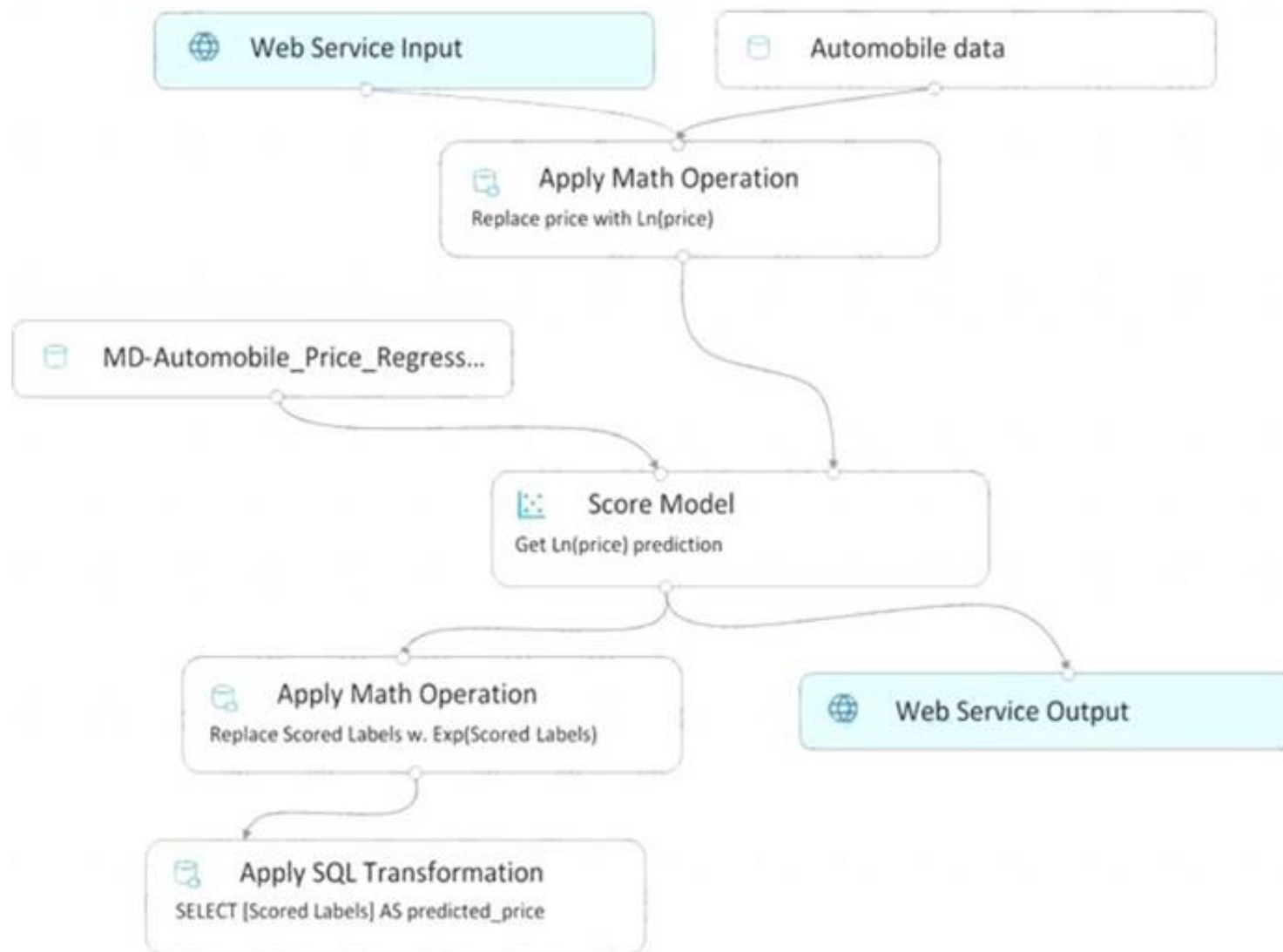
The training pipeline is shown in the exhibit. (Click the Training pipeline tab.)

Training pipeline



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

Real-time pipeline



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 8

- (Exam Topic 3)

You create a Python script named train.py and save it in a folder named scripts. The script uses the scikit-learn framework to train a machine learning model.

You must run the script as an Azure Machine Learning experiment on your local workstation. You need to write Python code to initiate an experiment that runs the train.py script.

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

NOTE: Each correct selection is worth one point.

## Answer Area

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

ws = Workspace.from_config()
py_sk = Environment('sklearn-training')
pkgs = CondaDependencies.create(pip_packages=['scikit-learn', 'azureml-defaults'])
py_sk.python.conda_dependencies = pkgs
script_config = ScriptRunConfig (
    script = 'scripts',
    source_directory = 'train.py',
    resume_from = py_sk,
    arguments = ['train.py'],
    environment = py_sk,
    compute_target = 'cpu-cluster'

experiment = Experiment(workspace=ws, name='training-experiment')
run = experiment.submit(config=script_config)
```

- A. Mastered  
 B. Not Mastered

**Answer:** A

### Explanation:

Graphical user interface, text, application, table, Word Description automatically generated

Box 1: source\_directory

source\_directory: A local directory containing code files needed for a run. Box 2: script

Script: The file path relative to the source\_directory of the script to be run. Box 3: environment

Reference:

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

## NEW QUESTION 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.config.auto>

## NEW QUESTION 10

- (Exam Topic 3)

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



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

You are creating a model to predict the price of a student's artwork depending on the following variables: the student's length of education, degree type, and art form.

You start by creating a linear regression model. You need to evaluate the linear regression model.

Solution: Use the following metrics: Relative Squared Error, Coefficient of Determination, Accuracy, Precision, Recall, F1 score, and AUC.

Does the solution meet the goal?

- A. Yes
- B. No

**Answer: B**

**Explanation:**

Relative Squared Error, Coefficient of Determination are good metrics to evaluate the linear regression model, but the others are metrics for classification models.

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

**NEW QUESTION 10**

- (Exam Topic 3)

You have a dataset that contains over 150 features. You use the dataset to train a Support Vector Machine (SVM) binary classifier.

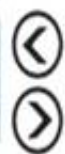
You need to use the Permutation Feature Importance module in Azure Machine Learning Studio to compute a set of feature importance scores for the dataset.

In which order should you perform the actions? To answer, move all actions from the list of actions to the answer area and arrange them in the correct order.

**Actions**

**Answer Area**

- Add a Two-Class Support Vector Machine module to initialize the SVM classifier.
- Set the Metric for measuring performance property to **Classification - Accuracy** and then run the experiment.
- Add a Permutation Feature Importance module and connect the trained model and test dataset.
- Add a dataset to the experiment.
- Add a Split Data module to create training and test datasets.



- A. Mastered
- B. Not Mastered

**Answer: A**

**Explanation:**

Step 1: Add a Two-Class Support Vector Machine module to initialize the SVM classifier.

Step 2: Add a dataset to the experiment

Step 3: Add a Split Data module to create training and test dataset.

To generate a set of feature scores requires that you have an already trained model, as well as a test dataset. Step 4: Add a Permutation Feature Importance module and connect to the trained model and test dataset. Step 5: Set the Metric for measuring performance property to Classification - Accuracy and then run the experiment.

Reference:

<https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/two-class-support-vector-mac> <https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/permutation-feature-importan>

**NEW QUESTION 14**

- (Exam Topic 3)

You use the Two-Class Neural Network module in Azure Machine Learning Studio to build a binary classification model. You use the Tune Model Hyperparameters module to tune accuracy for the model.

You need to select the hyperparameters that should be tuned using the Tune Model Hyperparameters module. Which two hyperparameters should you use? Each correct answer presents part of the solution.

NOTE: Each correct selection is worth one point.

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

**Answer: DE**

**Explanation:**

D: For Number of learning iterations, specify the maximum number of times the algorithm should process the training cases.



E: For Hidden layer specification, select the type of network architecture to create.

Between the input and output layers you can insert multiple hidden layers. Most predictive tasks can be accomplished easily with only one or a few hidden layers.

References:

<https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/two-class-neural-network>

### NEW QUESTION 19

- (Exam Topic 3)

You retrain an existing model.

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

What should you do?

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

**Answer: B**

#### Explanation:

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

Reference:

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

### NEW QUESTION 24

- (Exam Topic 3)

You are training machine learning models in Azure Machine Learning. You use Hyperdrive to tune the hyperparameters. In previous model training and tuning runs, many models showed similar performance. You need to select an early termination policy that meets the following requirements:

- accounts for the performance of all previous runs when evaluating the current run
- avoids comparing the current run with only the best performing run to date

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

- A. Bandit
- B. Median stopping
- C. Default
- D. Truncation selection

**Answer: BC**

#### Explanation:

The Median Stopping policy computes running averages across all runs and cancels runs whose best performance is worse than the median of the running averages.

If no policy is specified, the hyperparameter tuning service will let all training runs execute to completion. Reference:

<https://docs.microsoft.com/en-us/python/api/azureml-train-core/azureml.train.hyperdrive.medianstoppingpolicy> <https://docs.microsoft.com/en-us/python/api/azureml-train-core/azureml.train.hyperdrive.truncationselectionpolicy> <https://docs.microsoft.com/en-us/python/api/azureml-train-core/azureml.train.hyperdrive.banditpolicy>

### NEW QUESTION 25

- (Exam Topic 3)

You run an experiment that uses an AutoMLConfig class to define an automated machine learning task with a maximum of ten model training iterations. The task will attempt to find the best performing model based on a metric named accuracy.

You submit the experiment with the following code:

You need to create Python code that returns the best model that is generated by the automated machine learning task. Which code segment should you use?

A)

```
best_model = automl_run.get_details()
```

B)

```
best_model = automl_run.get_output()[1]
```

C)

```
best_model = automl_run.get_file_names()[1]
```

D)

```
best_model = automl_run.get_metrics()
```

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

**Answer: B**

#### Explanation:

The get\_output method returns the best run and the fitted model. Reference:

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

**NEW QUESTION 27**

- (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 30**

- (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><div></div><div>Tabular</div><div>HAN</div><div>Text</div><div>Image</div></div>
A natural language processing model for analyzing field reports	<div><div></div><div>Tree</div><div>HAN</div><div>Text</div><div>Image</div></div>
An image classifier that determines the quality of the grape based upon its physical characteristics.	<div><div></div><div>Kernel</div><div>HAN</div><div>Text</div><div>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 31**

- (Exam Topic 3)

You are creating a binary classification by using a two-class logistic regression model. You need to evaluate the model results for imbalance. Which evaluation metric should you use?

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

**Answer:** B

**Explanation:**

One can inspect the true positive rate vs. the false positive rate in the Receiver Operating Characteristic (ROC) curve and the corresponding Area Under the Curve (AUC) value. The closer this curve is to the upper left corner, the better the classifier's performance is (that is maximizing the true positive rate while minimizing the false positive rate). Curves that are close to the diagonal of the plot, result from classifiers that tend to make predictions that are close to random guessing.

References:

<https://docs.microsoft.com/en-us/azure/machine-learning/studio/evaluate-model-performance#evaluating-a-bina>

**NEW QUESTION 35**

- (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 40**

- (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 43**

- (Exam Topic 3)

You are evaluating a completed binary classification machine learning model. You need to use the precision as the valuation metric.

Which visualization should you use?



- A. Binary classification confusion matrix
- B. box plot
- C. Gradient descent
- D. coefficient of determination

**Answer:** A

**Explanation:**

References:  
<https://machinelearningknowledge.ai/confusion-matrix-and-performance-metrics-machine-learning/>

**NEW QUESTION 47**

- (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 51**

- (Exam Topic 3)

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

```
from azureml.pipeline.core import Pipeline
from azureml.core.experiment import Experiment
pipeline = Pipeline(workspace=ws, steps=[parallelrun_step]) pipeline_run = Experiment(ws, 'batch_pipeline').submit(pipeline)
You need to monitor the progress of the pipeline execution.
```

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

NOTE: Each correct selection is worth one point.

- A. Run the following code in a notebook:

```
from azureml.contrib.interpret.explanation.explanation_client import ExplanationClient
client = ExplanationClient.from_run(pipeline_run)
explanation = client.download_model_explanation()
explanation = client.download_model_explanation(top_k=4)
global_importance_values = explanation.get_ranked_global_values()
global_importance_names = explanation.get_ranked_global_names()
print('global importance values: {}'.format(global_importance_values))
print('global importance names: {}'.format(global_importance_names))
```
- B. Use the Inference Clusters tab in Machine Learning Studio.
- C. Use the Activity log in the Azure portal for the Machine Learning workspace.
- D. Run the following code in a notebook:

```
from azureml.widgets import RunDetails
RunDetails(pipeline_run).show()
```
- E. Run the following code and monitor the console output from the PipelineRun object:

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

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

**Answer:** DE

**Explanation:**

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

- Azure Machine Learning Studio.
- Console output from the PipelineRun object. `from azureml.widgets import RunDetails RunDetails(pipeline_run).show()`

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

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



**NEW QUESTION 54**

- (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 59**

- (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 63**

- (Exam Topic 3)

You create a datastore named training\_data that references a blob container in an Azure Storage account. The blob container contains a folder named csv\_files in which multiple comma-separated values (CSV) files are stored.

You have a script named train.py in a local folder named ./script that you plan to run as an experiment using an estimator. The script includes the following code to read data from the csv\_files folder:

```
import os
import argparse
import pandas as pd

from sklearn.model_selection import train_test_split
from sklearn.linear_model import LogisticRegression
from azureml.core import Run

run = Run.get_context()
parser = argparse.ArgumentParser()
parser.add_argument('--data-folder', type=str, dest='data_folder', help='data reference')
args = parser.parse_args()

data_folder = args.data_folder
csv_files = os.listdir(data_folder)
training_data = pd.concat((pd.read_csv(os.path.join(data_folder, csv_file)) for csv_file in csv_files))

# Code goes on to split the training data and train a logistic regression model
```

You have the following script.

```
from azureml.core import Workspace, Datastore, Experiment
from azureml.train.sklearn import SKLearn

ws = Workspace.from_config()
exp = Experiment(workspace=ws, name='csv_training')
ds = Datastore.get(ws, datastore_name='training_data')
data_ref = ds.path('csv_files')

# Code to define estimator goes here

run = exp.submit(config=estimator)
run.wait_for_completion(show_output=True)
```

You need to configure the estimator for the experiment so that the script can read the data from a data reference named data\_ref that references the csv\_files folder in the training\_data datastore.

Which code should you use to configure the estimator?

- A. 

```
estimator = SKLearn(source_directory='./script',
    inputs=[data_ref.as_named_input('data-folder').to_pandas_dataframe()],
    compute_target='local',
    entry_script='train.py')
```
- B. 

```
script_params = {
    '--data-folder': data_ref.as_mount()
}
estimator = SKLearn(source_directory='./script',
    script_params=script_params,
    compute_target='local',
    entry_script='train.py')
```
- C. 

```
estimator = SKLearn(source_directory='./script',
    inputs=[data_ref.as_named_input('data-folder').as_mount()],
    compute_target='local',
    entry_script='train.py')
```
- D. 

```
script_params = {
    '--data-folder': data_ref.as_download(path_on_compute='csv_files')
}
estimator = SKLearn(source_directory='./script',
    script_params=script_params,
    compute_target='local',
    entry_script='train.py')
```
- E. 

```
estimator = SKLearn(source_directory='./script',
    inputs=[data_ref.as_named_input('data-folder').as_download(path_on_compute='csv_files')],
    compute_target='local',
    entry_script='train.py')
```

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

**Answer:** B

#### Explanation:

Besides passing the dataset through the inputs parameter in the estimator, you can also pass the dataset through script\_params and get the data path (mounting point) in your training script via arguments. This way, you can keep your training script independent of azureml-sdk. In other words, you will be able use the same training script for local debugging and remote training on any cloud platform.

Example:

```
from azureml.train.sklearn import SKLearn
script_params = {
    # mount the dataset on the remote compute and pass the mounted path as an argument to the training script '--data-folder':
    mnist_ds.as_named_input('mnist').as_mount(),
    '--regularization': 0.5
}
est = SKLearn(source_directory=script_folder, script_params=script_params, compute_target=compute_target, environment_definition=env,
    entry_script='train_mnist.py')
# Run the experiment
run = experiment.submit(est)
run.wait_for_completion(show_output=True)
```

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

#### NEW QUESTION 64

- (Exam Topic 3)

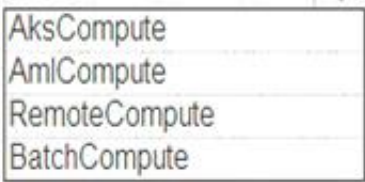
You use Azure Machine Learning to train and register a model.


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.

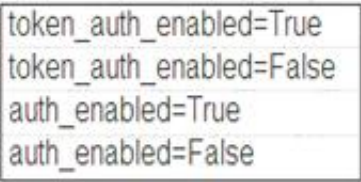
Client applications consuming the deployed web service must be authenticated based on their Azure Active Directory service principal.

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")

deployment_config = .deploy_configuration(cpu_cores=1, memory_gb=1,

 )

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 68**

- (Exam Topic 3)

A coworker registers a datastore in a Machine Learning services workspace by using the following code:

```
Datastore.register_azure_blob_container(workspace=ws,
    datastore_name='demo_datastore',
    container_name='demo_datacontainer',
    account_name='demo_account',
    account_key='0A0A0A-0A0A00A-0A00A0A0A0A',
    create_if_not_exists=True)
```

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

**Answer Area**

```
import azureml.core
from azureml.core import Workspace, Datastore
ws = Workspace.from_config()

datastore = .get( , 
```

- A. Mastered  
B. Not Mastered

**Answer:** A

**Explanation:**

Box 1: DataStore



To get a specific datastore registered in the current workspace, use the get() static method on the Datastore class:

# Get a named datastore from the current workspace

datastore = Datastore.get(ws, datastore\_name='your datastore name') Box 2: ws

Box 3: demo\_datastore Reference:

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

### NEW QUESTION 73

- (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 76

- (Exam Topic 3)

You are the owner of an Azure Machine Learning workspace.

You must prevent the creation or deletion of compute resources by using a custom role. You must allow all other operations inside the workspace.

You need to configure the custom role.

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

NOTE: Each correct selection is worth one point.

#### Answer Area

```
{
  "Name": "Data Scientist Custom",
  "IsCustom": true
  "Description": "Description"
  "Actions": [
    Microsoft.MachineLearningServices/workspaces/*/read
    Microsoft.MachineLearningServices/workspaces/computes/*/write
    Microsoft.MachineLearningServices/workspaces/delete
    Microsoft.MachineLearningServices/workspaces/*/write
    Microsoft.MachineLearningServices/workspaces/computes/*/write
    Microsoft.MachineLearningServices/workspaces/delete
  ],
  "NotActions": [
    Microsoft.MachineLearningServices/workspaces/*/read
    Microsoft.MachineLearningServices/workspaces/*/write
    Microsoft.MachineLearningServices/workspaces/computes/*/delete
    Microsoft.MachineLearningServices/workspaces/*/read
    Microsoft.MachineLearningServices/workspaces/*/write
    Microsoft.MachineLearningServices/workspaces/computes/*/write
  ],
  "AssignableScopes": [
    "/subscriptions/<subscription_id>"
  ]
}
```

- A. Mastered
- B. Not Mastered

**Answer: A**

#### Explanation:

Graphical user interface, application Description automatically generated



Graphical user interface, application Description automatically generated

Box 1: Microsoft.MachineLearningServices/workspaces/\*/read

Reader role: Read-only actions in the workspace. Readers can list and view assets, including datastore credentials, in a workspace. Readers can't create or update these assets.

Box 2: Microsoft.MachineLearningServices/workspaces/\*/write

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

Box 3: Box 2: Microsoft.MachineLearningServices/workspaces/computes/\*/delete

Box 4: Microsoft.MachineLearningServices/workspaces/computes/\*/write Reference:

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

#### NEW QUESTION 77

- (Exam Topic 3)

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

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

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

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

**Answer:** C

#### Explanation:

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

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

Reference:

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

#### NEW QUESTION 81

- (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 85

- (Exam Topic 3)

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

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

You are analyzing a numerical dataset which contains missing values in several columns.

You must clean the missing values using an appropriate operation without affecting the dimensionality of the feature set.

You need to analyze a full dataset to include all values.

Solution: Replace each missing value using the Multiple Imputation by Chained Equations (MICE) method. Does the solution meet the goal?

- A. Yes
- B. NO

**Answer:** A

#### Explanation:

Replace using MICE: For each missing value, this option assigns a new value, which is calculated by using a method described in the statistical literature as "Multivariate Imputation using Chained Equations" or "Multiple Imputation by Chained Equations". With a multiple imputation method, each variable with missing data is modeled conditionally using the other variables in the data before filling in the missing values.

Note: Multivariate imputation by chained equations (MICE), sometimes called “fully conditional specification” or “sequential regression multiple imputation” has emerged in the statistical literature as one principled method of addressing missing data. Creating multiple imputations, as opposed to single imputations, accounts for the statistical uncertainty in the imputations. In addition, the chained equations approach is very flexible and can handle variables of varying types (e.g., continuous or binary) as well as complexities such as bounds or survey skip patterns.

References: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3074241/>

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

#### NEW QUESTION 87

- (Exam Topic 3)

You have a feature set containing the following numerical features: X, Y, and Z.

The Poisson correlation coefficient (r-value) of X, Y, and Z features is shown in the following image:

	X	Y	Z
X	1	0.149676	-0.106276
Y	0.149676	1	0.859122
Z	-0.106276	0.859122	1

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

NOTE: Each correct selection is worth one point.

What is the r-value for the correlation of Y to Z?

▼

-0.106276

0.149676

0.859122

1

Which type of relationship exists between Z and Y in the feature set?

▼

a positive linear relationship

a negative linear relationship

no linear relationship

- A. Mastered  
 B. Not Mastered

**Answer: A**

**Explanation:**

Box 1: 0.859122

Box 2: a positively linear relationship

+1 indicates a strong positive linear relationship

-1 indicates a strong negative linear correlation

0 denotes no linear relationship between the two variables. References:

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

**NEW QUESTION 90**

- (Exam Topic 3)

You use an Azure Machine Learning workspace.

You have a trained model that must be deployed as a web service. Users must authenticate by using Azure Active Directory.

What should you do?

- A. Deploy the model to Azure Kubernetes Service (AKS). During deployment, set the token\_auth\_enabled parameter of the target configuration object to true  
 B. Deploy the model to Azure Container Instance  
 C. During deployment, set the auch\_enabled parameter of the target configuration object to true  
 D. Deploy the model to Azure Container Instance  
 E. During deployment, set the coken\_auch\_enabled parameter of the target configuration object to true  
 F. Deploy the model to Azure Kubernetes Service (AKS). During deployment, set the auc  
 G. enabled parameter of the target configuration object to true

**Answer: A**

**Explanation:**

To control token authentication, use the token\_auth\_enabled parameter when you create or update a deployment

Token authentication is disabled by default when you deploy to Azure Kubernetes Service.

Note: The model deployments created by Azure Machine Learning can be configured to use one of two authentication methods:

key-based: A static key is used to authenticate to the web service.

token-based: A temporary token must be obtained from the Azure Machine Learning workspace (using Azure Active Directory) and used to authenticate to the web service.

Reference:

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

**NEW QUESTION 92**

- (Exam Topic 3)

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

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

You train and register a machine learning model.

You plan to deploy the model as a real-time web service. Applications must use key-based authentication to use the model.

You need to deploy the web service. Solution:

Create an AksWebservice instance.

Set the value of the auth\_enabled property to True.

Deploy the model to the service. Does the solution meet the goal?

- A. Yes

B. No

**Answer:** A

**Explanation:**

Key-based authentication.

Web services deployed on AKS have key-based auth enabled by default. ACI-deployed services have key-based auth disabled by default, but you can enable it by setting `auth_enabled = TRUE` when creating the ACI web service. The following is an example of creating an ACI deployment configuration with key-based auth enabled.

```
deployment_config <- aci_webservice_deployment_config(cpu_cores = 1, memory_gb = 1,
```

```
auth_enabled = TRUE) Reference:
```

<https://azure.github.io/azureml-sdk-for-r/articles/deploying-models.html>

**NEW QUESTION 97**

- (Exam Topic 3)

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

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

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

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

You need to select an appropriate data sampling strategy to compensate for the class imbalance. Solution: You use the Synthetic Minority Oversampling Technique (SMOTE) sampling mode. Does the solution meet the goal?

A. Yes

B. No

**Answer:** A

**Explanation:**

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

References:

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

**NEW QUESTION 102**

- (Exam Topic 3)

You are building an intelligent solution using machine learning models. The environment must support the following requirements:

- > Data scientists must build notebooks in a cloud environment
- > Data scientists must use automatic feature engineering and model building in machine learning pipelines.
- > Notebooks must be deployed to retrain using Spark instances with dynamic worker allocation.
- > Notebooks must be exportable to be version controlled locally.

You need to create the environment.

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

Actions	Answer area
Install the Azure Machine Learning SDK for Python on the cluster.	
When the cluster is ready, export Zeppelin notebooks to a local environment.	
Create and execute a Jupyter notebook by using automated machine learning (AutoML) on the cluster.	
Install Microsoft Machine Learning for Apache Spark.	⬅
When the cluster is ready and has processed the notebook, export your Jupyter notebook to a local environment.	➡
Create an Azure HDInsight cluster to include the Apache Spark Mlib library.	⬆
Create and execute the Zeppelin notebooks on the cluster.	⬇
Create an Azure Databricks cluster.	

A. Mastered

B. Not Mastered

**Answer:** A

**Explanation:**

Step 1: Create an Azure HDInsight cluster to include the Apache Spark Mlib library

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



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

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

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

References:

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

### NEW QUESTION 103

- (Exam Topic 3)

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

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

- A. MI.NET Library on a local environment
- B. Azure Machine Learning Studio
- C. Data Science Virtual Machine (OSVM)
- D. Azure Data bricks
- E. Azure Cognitive Services

**Answer:** ABD

### NEW QUESTION 108

- (Exam Topic 3)

You plan to use Hyperdrive to optimize the hyperparameters selected when training a model. You create the following code to define options for the hyperparameter experiment

```
import azureml.train.hyperdrive.parameter_expressions as pe
from azureml.train.hyperdrive import GridParameterSampling, HyperDriveConfig

param_sampling = GridParameterSampling({
    "max_depth" : pe.choice(6, 7, 8, 9),
    "learning_rate" : pe.choice(0.05, 0.1, 0.15)
})

hyperdrive_run_config = HyperDriveConfig(
    estimator = estimator,
    hyperparameter_sampling = param_sampling,
    policy = None,
    primary_metric_name = "auc",
    primary_metric_goal = PrimaryMetricGoal.MAXIMIZE,

    estimator = estimator,
    hyperparameter_sampling = param_sampling,
    policy = None,
    primary_metric_name = "auc",
    primary_metric_goal = PrimaryMetricGoal.MAXIMIZE,
    max_total_runs = 50,
    max_concurrent_runs = 4)
```

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

#### Answer Area

	Yes	No
There will be 50 runs for this hyperparameter tuning experiment.	<input type="radio"/>	<input type="radio"/>
You can use the policy parameter in the HyperDriveConfig class to specify a security policy.	<input type="radio"/>	<input type="radio"/>
The experiment will create a run for every possible value for the learning rate parameter between 0.05 and 0.15.	<input type="radio"/>	<input type="radio"/>

- A. Mastered
- B. Not Mastered

**Answer:** A

#### Explanation:

Box 1: No

max\_total\_runs (50 here)

The maximum total number of runs to create. This is the upper bound; there may be fewer runs when the sample space is smaller than this value.

Box 2: Yes

Policy EarlyTerminationPolicy



The early termination policy to use. If None - the default, no early termination policy will be used. Box 3: No  
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/python/api/azureml-train-core/azureml.train.hyperdrive.hyperdriveconfig> <https://docs.microsoft.com/en-us/azure/machine-learning/how-to-tune-hyperparameters>

#### NEW QUESTION 113

- (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 115

- (Exam Topic 3)

You train and register a machine learning model. You create a batch inference pipeline that uses the model to generate predictions from multiple data files. You must publish the batch inference pipeline as a service that can be scheduled to run every night. You need to select an appropriate compute target for the inference service. Which compute target should you use?

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

**Answer:** B

#### Explanation:

Azure Machine Learning compute clusters is used for Batch inference. Run batch scoring on serverless compute. Supports normal and low-priority VMs. No support for real-time inference.  
Reference:  
<https://docs.microsoft.com/en-us/azure/machine-learning/concept-compute-target>

#### NEW QUESTION 120

- (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
<code>X_train</code>	training feature set
<code>Y_train</code>	training class labels
<code>x_train</code>	testing feature set
<code>y_train</code>	testing class labels

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

```
from sklearn.decomposition import PCA
pca = PCA()
pca.fit_transform(X_train)
x_test = pca.transform(x_test)
```

Box 1: PCA()  
PCA(n\_components = 150)  
PCA(n\_components = 10)  
PCA(n\_components = 10000)

Box 2: pca  
model  
sklearn.decomposition

Box 3: x\_test  
X\_train  
fit(x\_test)  
transform(x\_test)

- A. Mastered  
B. Not Mastered

**Answer: A**

**Explanation:**

Box 1: PCA(n\_components = 10)

Need to reduce the dimensionality of the feature set to 10 features in both training and testing sets. Example:

from sklearn.decomposition import PCA pca = PCA(n\_components=2) ;2 dimensions principalComponents = pca.fit\_transform(x)

Box 2: pca

fit\_transform(X[, y])fits the model with X and apply the dimensionality reduction on X. Box 3: transform(x\_test)

transform(X) applies dimensionality reduction to X. References:

<https://scikit-learn.org/stable/modules/generated/sklearn.decomposition.PCA.html>

**NEW QUESTION 123**

- (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 128**

- (Exam Topic 3)

The finance team asks you to train a model using data in an Azure Storage blob container named finance-data. You need to register the container as a datastore in an Azure Machine Learning workspace and ensure that an error will be raised if the container does not exist.

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

NOTE: Each correct selection is worth one point.

```
datastore = Datastore.register_azure_blob_container(workspace = ws,
```

Box 1: register\_azure\_blob\_container  
register\_azure\_file\_share  
register\_azure\_data\_lake  
register\_azure\_sql\_database

```
datastore_name = 'finance_datastore',  
container_name = 'finance-data',  
account_name = 'fintrainingdatastorage',  
account_key = 'FWUYORRv3XoyNe...',  
create_if_not_exists = True  
create_if_not_exists = False  
overwrite = True  
overwrite = False
```

- A. Mastered

B. Not Mastered

**Answer:** A

**Explanation:**

Box 1: register\_azure\_blob\_container

Register an Azure Blob Container to the datastore.

Box 2: create\_if\_not\_exists = False

Create the file share if it does not exists, defaults to False. Reference:

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

**NEW QUESTION 129**

- (Exam Topic 3)

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

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

You are 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 131**

- (Exam Topic 3)

You use the Azure Machine Learning SDK in a notebook to run an experiment using a script file in an experiment folder.

The experiment fails.

You need to troubleshoot the failed experiment.

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

A. Use the get\_metrics() method of the run object to retrieve the experiment run logs.

B. Use the get\_details\_with\_logs() method of the run object to display the experiment run logs.

C. View the log files for the experiment run in the experiment folder.

D. View the logs for the experiment run in Azure Machine Learning studio.

E. Use the get\_output() method of the run object to retrieve the experiment run logs.

**Answer:** BD

**Explanation:**

Use get\_details\_with\_logs() to fetch the run details and logs created by the run.

You can monitor Azure Machine Learning runs and view their logs with the Azure Machine Learning studio. Reference:

<https://docs.microsoft.com/en-us/python/api/azureml-pipeline-core/azureml.pipeline.core.steprun> <https://docs.microsoft.com/en-us/azure/machine-learning/how-to-monitor-view-training-logs>

**NEW QUESTION 132**

- (Exam Topic 3)

You create an Azure Databricks workspace and a linked Azure Machine Learning workspace. You have the following Python code segment in the Azure Machine Learning workspace:

```
import mlflow
```

```
import mlflow.azureml import azureml.mlflow import azureml.core
```

```
from azureml.core import Workspace subscription_id = 'subscription_id' resource_group = 'resource_group_name' workspace_name = 'workspace_name'
```

```
ws = Workspace.get(name=workspace_name, subscription_id=subscription_id, resource_group=resource_group)
```

```
experimentName = "/Users/{user_name}/{experiment_folder}/{experiment_name}" mlflow.set_experiment(experimentName)
```

```
uri = ws.get_mlflow_tracking_uri() mlflow.set_tracking_uri(uri)
```

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

NOTE: Each correct selection is worth one point.



	Yes	No
A resource group and Azure Machine Learning workspace will be created.	<input type="radio"/>	<input type="radio"/>
An Azure Databricks experiment will be tracked only in the Azure Machine Learning workspace.	<input type="radio"/>	<input type="radio"/>
The epoch loss metric is set to be tracked.	<input type="radio"/>	<input type="radio"/>

- A. Mastered  
 B. Not Mastered

**Answer: A**

**Explanation:**

A screenshot of a computer Description automatically generated with medium confidence

Box 1: No

The Workspace.get method loads an existing workspace without using configuration files. ws = Workspace.get(name="myworkspace", subscription\_id='<azure-subscription-id>', resource\_group='myresourcegroup')

Box 2: Yes

MLflow Tracking with Azure Machine Learning lets you store the logged metrics and artifacts from your local runs into your Azure Machine Learning workspace. The get\_mlflow\_tracking\_uri() method assigns a unique tracking URI address to the workspace, ws, and set\_tracking\_uri() points the MLflow tracking URI to that address.

Box 3: Yes

Note: In Deep Learning, epoch means the total dataset is passed forward and backward in a neural network once.

Reference:

<https://docs.microsoft.com/en-us/python/api/azureml-core/azureml.core.workspace.workspace> <https://docs.microsoft.com/en-us/azure/machine-learning/how-to-use-mlflow>

**NEW QUESTION 135**

- (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 137**

- (Exam Topic 3)

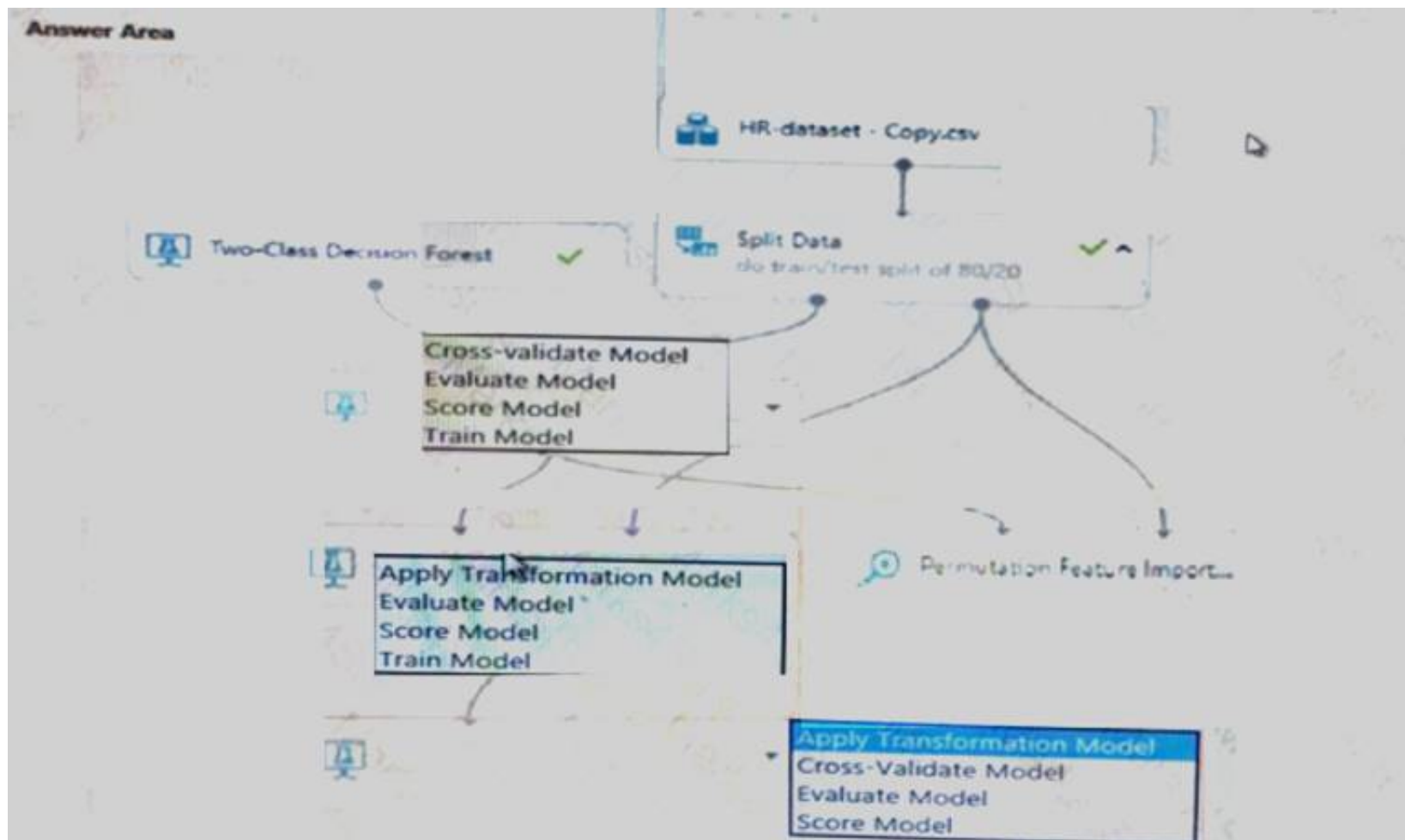
You create a binary classification model using Azure Machine Learning Studio.

You must use a Receiver Operating Characteristic (ROC) 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.

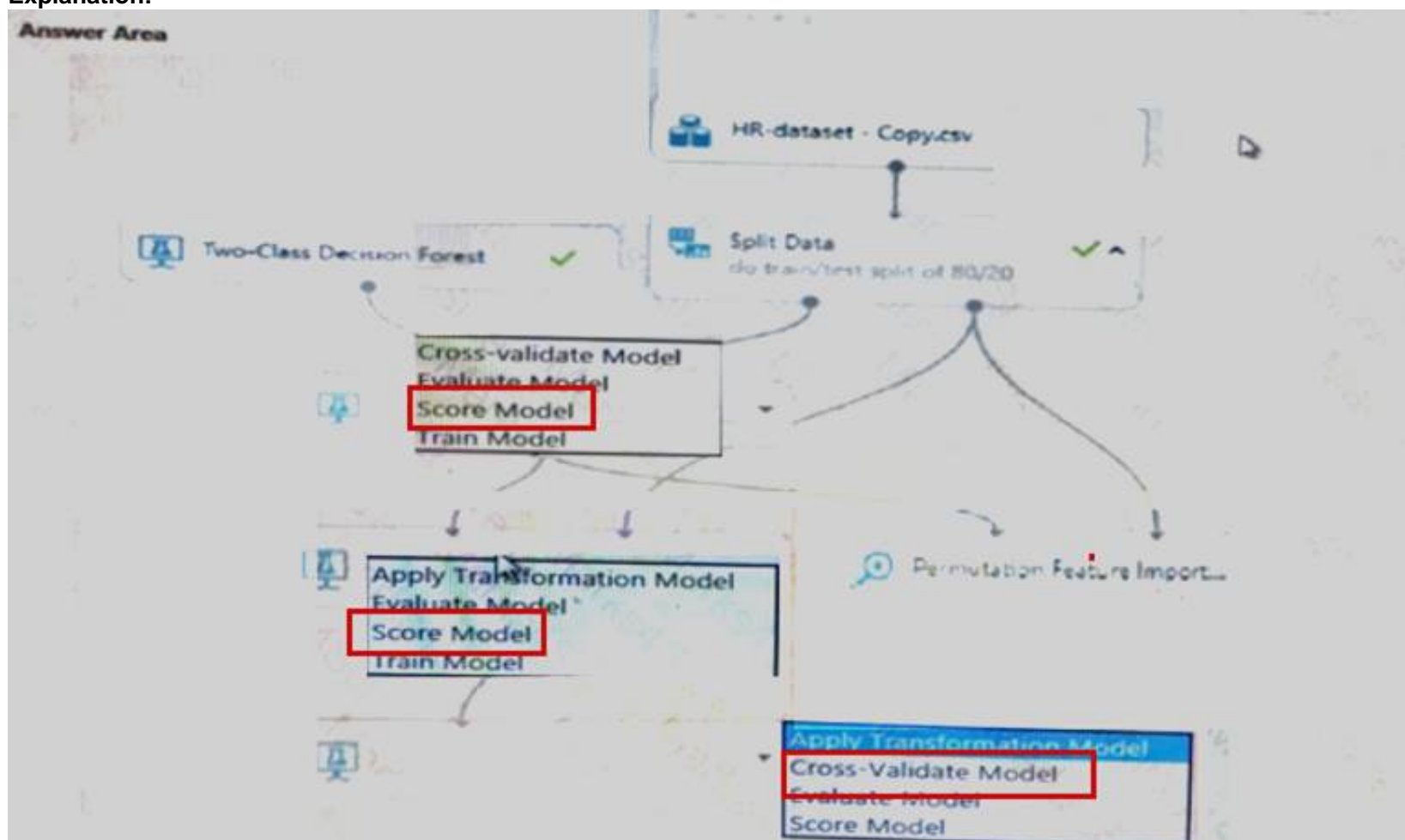




- A. Mastered
- B. Not Mastered

Answer: A

Explanation:



#### NEW QUESTION 141

- (Exam Topic 3)

#### HOTSPOT

You have an Azure blob container that contains a set of TSV files. The Azure blob container is registered as a datastore for an Azure Machine Learning service workspace. Each TSV file uses the same data schema.

You plan to aggregate data for all of the TSV files together and then register the aggregated data as a dataset in an Azure Machine Learning workspace by using the Azure Machine Learning SDK for Python.

You run the following code.

```
from azureml.core.workspace import Workspace
from azureml.core.datastore import Datastore
from azureml.core.dataset import Dataset
import pandas as pd
datastore_paths = (datastore, './data/*.tsv')
myDataset_1 = Dataset.File.from_files(path=datastore_paths)
myDataset_2 = Dataset.Tabular.from_delimited_files(path=datastore_paths, separator='\t')
```

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 myDataset_1 dataset can be converted into a pandas dataframe by using the following method: using myDataset_1.to_pandas_dataframe()	<input type="radio"/>	<input type="radio"/>
The myDataset_1.to_path() method returns an array of file paths for all of the TSV files in the dataset.	<input type="radio"/>	<input type="radio"/>
The myDataset_2 dataset can be converted into a pandas dataframe by using the following method: myDataset_2.to_pandas_dataframe()	<input type="radio"/>	<input type="radio"/>

- A. Mastered  
B. Not Mastered

**Answer:** A

**Explanation:**

Box 1: No

FileDataset references single or multiple files in datastores or from public URLs. The TSV files need to be parsed.

Box 2: Yes

to\_path() gets a list of file paths for each file stream defined by the dataset. Box 3: Yes

TabularDataset.to\_pandas\_dataframe loads all records from the dataset into a pandas DataFrame. TabularDataset represents data in a tabular format created by parsing the provided file or list of files.

Note: TSV is a file extension for a tab-delimited file used with spreadsheet software. TSV stands for Tab Separated Values. TSV files are used for raw data and can be imported into and exported from spreadsheet software. TSV files are essentially text files, and the raw data can be viewed by text editors, though they are often used when moving raw data between spreadsheets.

Reference:

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

**NEW QUESTION 144**

- (Exam Topic 3)

You plan to provision an Azure Machine Learning Basic edition workspace for a data science project. You need to identify the tasks you will be able to perform in the workspace.

Which three tasks will you be able to perform? Each correct answer presents a complete solution.

NOTE: Each correct selection is worth one point. D

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

**Answer:** ABD

**Explanation:**

Reference:

<https://azure.microsoft.com/en-us/pricing/details/machine-learning/>

**NEW QUESTION 147**

- (Exam Topic 3)

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

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

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

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

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

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

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

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

Does the solution meet the goal?

- A. Yes
- B. No

**Answer: B**

**Explanation:**

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

Destination: Driver logs, Azure Machine Learning designer Reference:

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

**NEW QUESTION 149**

- (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 150**

- (Exam Topic 3)

You create an experiment in Azure Machine Learning Studio. You add a training dataset that contains 10,000 rows. The first 9,000 rows represent class 0 (90 percent).

The remaining 1,000 rows represent class 1 (10 percent).

The training set is imbalanced between two classes. You must increase the number of training examples for class 1 to 4,000 by using 5 data rows. You add the Synthetic Minority Oversampling Technique (SMOTE) module to the experiment.

You need to configure the module.

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



SMOTE

Label column

Selected columns:

All labels

Launch column selector

SMOTE percentage

0

300

3000

4000

Number of nearest neighbors

0

1

5

4000

Random seed

0

- A. Mastered  
B. Not Mastered

Answer: A

Explanation:

Box 1: 300

You type 300 (%), the module triples the percentage of minority cases (3000) compared to the original dataset (1000).

Box 2: 5

We should use 5 data rows.

Use the Number of nearest neighbors option to determine the size of the feature space that the SMOTE algorithm uses when in building new cases. A nearest neighbor is a row of data (a case) that is very similar to some target case. The distance between any two cases is measured by combining the weighted vectors of all features.

By increasing the number of nearest neighbors, you get features from more cases.

By keeping the number of nearest neighbors low, you use features that are more like those in the original sample.

References:

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

NEW QUESTION 154

- (Exam Topic 3)

You use Azure Machine Learning to deploy a model as a real-time web service.

You need to create an entry script for the service that ensures that the model is loaded when the service starts and is used to score new data as it is received.

Which functions should you include in the script? To answer, drag the appropriate functions to the correct actions. Each function may be used once, more than once, or not at all. You may need to drag the split bar between panes or scroll to view content

NOTE: Each correct selection is worth one point.

Functions

main()

init()

score()

predict()

run()

Answer Area

Action

Load the model when the service starts.  
Use the model to score new data.

Function

- A. Mastered  
B. Not Mastered

Answer: A

Explanation:

Box 1: init()

The entry script has only two required functions, init() and run(data). These functions are used to initialize the service at startup and run the model using request data passed in by a client. The rest of the script handles loading and running the model(s).

Box 2: run() Reference:

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

### NEW QUESTION 155

- (Exam Topic 3)

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

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

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

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

Solution: Run the following code:

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

Does the solution meet the goal?

- A. Yes
- B. No

**Answer: A**

#### Explanation:

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

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

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

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

from azureml.pipeline.steps import PythonScriptStep datastore = ws.get\_default\_datastore()

process\_step\_output = PipelineData("processed\_data", datastore=datastore) process\_step = PythonScriptStep(script\_name="process.py", arguments=["--data\_for\_train", process\_step\_output], outputs=[process\_step\_output], compute\_target=aml\_compute, source\_directory=process\_directory)

train\_step = PythonScriptStep(script\_name="train.py", arguments=["--data\_for\_train", process\_step\_output], inputs=[process\_step\_output], compute\_target=aml\_compute, source\_directory=train\_directory)

pipeline = Pipeline(workspace=ws, steps=[process\_step, train\_step]) Reference:

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

### NEW QUESTION 157

- (Exam Topic 3)

You are evaluating a Python NumPy array that contains six data points defined as follows: data = [10, 20, 30, 40, 50, 60]

You must generate the following output by using the k-fold algorithm implantation in the Python Scikit-learn machine learning library:

train: [10 40 50 60], test: [20 30]

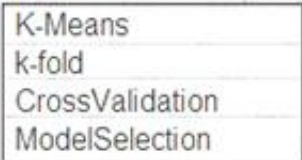
train: [20 30 40 60], test: [10 50]


train: [10 20 30 50], test: [40 60]

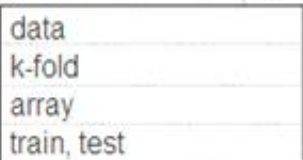
You need to implement a cross-validation to generate the output.

How should you complete the code segment? To answer, select the appropriate code segment in the dialog box in the answer area.

NOTE: Each correct selection is worth one point.

```
from numpy import array
from sklearn.model_selection import 

data = array([10, 20, 30, 40, 50, 60])
kfold = Kfold(n_splits=, shuffle = True, random_state=1)

for train, test in kFold, split():

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

- A. Mastered
- B. Not Mastered

**Answer:** A

**Explanation:**

Box 1: k-fold

Box 2: 3

K-F olds cross-validator provides train/test indices to split data in train/test sets. Split dataset into k consecutive folds (without shuffling by default).

The parameter n\_splits ( int, default=3) is the number of folds. Must be at least 2. Box 3: data

Example: Example:

```
>>>
```

```
>>> from sklearn.model_selection import KFold
```

```
>>> X = np.array([[1, 2], [3, 4], [1, 2], [3, 4]])
```

```
>>> y = np.array([1, 2, 3, 4])
```

```
>>> kf = KFold(n_splits=2)
```

```
>>> kf.get_n_splits(X) 2
```

```
>>> print(kf)
```

```
KFold(n_splits=2, random_state=None, shuffle=False)
```

```
>>> for train_index, test_index in kf.split(X): print("TRAIN:", train_index, "TEST:", test_index) X_train, X_test = X[train_index], X[test_index] y_train, y_test =
```

```
y[train_index], y[test_index] TRAIN: [2 3] TEST: [0 1]
```

```
TRAIN: [0 1] TEST: [2 3]
```

References:

[https://scikit-learn.org/stable/modules/generated/sklearn.model\\_selection.KFold.html](https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.KFold.html)

**NEW QUESTION 161**

- (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 163

- (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 164**

- (Exam Topic 3)

You are analyzing a raw dataset that requires cleaning.

You must perform transformations and manipulations by using Azure Machine Learning Studio. You need to identify the correct modules to perform the transformations.

Which modules should you choose? To answer, drag the appropriate modules to the correct scenarios. Each module may be used once, more than once, or not at all.

You may need to drag the split bar between panes or scroll to view content. NOTE: Each correct selection is worth one point.

**Answer Area**

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

- A. Mastered
- B. Not Mastered

**Answer:** A

**Explanation:**

Box 1: Clean Missing Data Box 2: SMOTE

Use the SMOTE module in Azure Machine Learning Studio to increase the number of 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.

Box 3: Convert to Indicator Values

Use the Convert to Indicator Values module in Azure Machine Learning Studio. The purpose of this module is to convert columns that contain categorical values into a series of binary indicator columns that can more easily be used as features in a machine learning model.

Box 4: Remove Duplicate Rows

References: <https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/smote> <https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/convert-to-indicator-values>

**NEW QUESTION 168**

- (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 170**

- (Exam Topic 3)

An organization uses Azure Machine Learning service and wants to expand their use of machine learning. You have the following compute environments. The organization does not want to create another compute environment.

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

You need to determine which compute environment to use for the following scenarios.

Which compute types should you use? To answer, drag the appropriate compute environments to the correct scenarios. Each compute environment may be used once, more than once, or not at all. You may need to drag the split bar between panes or scroll to view content.

NOTE: Each correct selection is worth one point.

**Environments**

nb\_server
aks\_cluster
mlc\_cluster

**Answer Area**

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

A. Mastered

B. Not Mastered

**Answer:** A

**Explanation:**

Box 1: nb\_server

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

Box 2: mlc\_cluster

With Azure Machine Learning, you can train your model on a variety of resources or environments, collectively referred to as compute targets. A compute target can be a local machine or a cloud resource, such as an Azure Machine Learning Compute, Azure HDInsight or a remote virtual machine.

Reference:

<https://docs.microsoft.com/en-us/azure/machine-learning/concept-compute-target> <https://docs.microsoft.com/en-us/azure/machine-learning/how-to-set-up-training-targets>



**NEW QUESTION 175**

- (Exam Topic 3)

You are evaluating a completed binary classification machine. You need to use the precision as the evaluation metric. Which visualization should you use?

- A. scatter plot
- B. coefficient of determination
- C. Receiver Operating Characteristic CROC) curve
- D. Gradient descent

**Answer: C**

**Explanation:**

Receiver operating characteristic (or ROC) is a plot of the correctly classified labels vs. the incorrectly classified labels for a particular model.

References:

<https://docs.microsoft.com/en-us/azure/machine-learning/how-to-understand-automated-ml#confusion-matrix>

**NEW QUESTION 179**

- (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 183**

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

- A. Yes
- B. No

**Answer: B**

**Explanation:**

Those are metrics for evaluating classification models, instead use: Mean Absolute Error, Root Mean Absolute Error, Relative Absolute Error, Relative Squared Error, and the Coefficient of Determination.

References:

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

**NEW QUESTION 188**

- (Exam Topic 3)

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

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

You train and register a machine learning model.

You plan to deploy the model as a real-time web service. Applications must use key-based authentication to use the model.

You need to deploy the web service.

Solution:

Create an AksWebService instance.

Set the value of the auth\_enabled property to False.

Set the value of the token\_auth\_enabled property to True.

Deploy the model to the service. Does the solution meet the goal?

- A. Yes
- B. No

**Answer: B**

**Explanation:**

Instead use only auth\_enabled = TRUE Note: Key-based authentication.

Web services deployed on AKS have key-based auth enabled by default. ACI-deployed services have key-based auth disabled by default, but you can enable it by setting auth\_enabled = TRUE when creating the ACI web service. The following is an example of creating an ACI deployment configuration with key-based auth enabled.

```
deployment_config <- aci_webservice_deployment_config(cpu_cores = 1, memory_gb = 1,
```

```
auth_enabled = TRUE) Reference:
```

<https://azure.github.io/azureml-sdk-for-r/articles/deploying-models.html>

**NEW QUESTION 191**

- (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						
	<table border="1"><thead><tr><th>Task</th><th>Tool</th></tr></thead><tbody><tr><td>Build DNN models</td><td>Tool</td></tr><tr><td>Enable interactive data exploration and visualization</td><td>Tool</td></tr></tbody></table>	Task	Tool	Build DNN models	Tool	Enable interactive data exploration and visualization	Tool
Task	Tool						
Build DNN models	Tool						
Enable interactive data exploration and visualization	Tool						
Vowpal Wabbit							
PowerBI Desktop							
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/architecture/data-guide/scenarios/interactive-data-exploration>

**NEW QUESTION 196**

- (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 201**

- (Exam Topic 3)

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

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

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

**Answer:** D

#### NEW QUESTION 203

- (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 an Equal Width with Custom Start and Stop binning mode.

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 204

- (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 207

- (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 212

- (Exam Topic 3)

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

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

You train and register a machine learning model.

You plan to deploy the model as a real-time web service. Applications must use key-based authentication to use the model.



You need to deploy the web service. Solution:  
Create an AciWebservice instance.  
Set the value of the ssl\_enabled property to True.  
Deploy the model to the service. Does the solution meet the goal?

- A. Yes
- B. No

Answer: B

Explanation:

Instead use only auth\_enabled = TRUE Note: Key-based authentication.  
Web services deployed on AKS have key-based auth enabled by default. ACI-deployed services have key-based auth disabled by default, but you can enable it by setting auth\_enabled = TRUE when creating the ACI web service. The following is an example of creating an ACI deployment configuration with key-based auth enabled.  
deployment\_config <- aci\_webservice\_deployment\_config(cpu\_cores = 1, memory\_gb = 1, auth\_enabled = TRUE) Reference:  
<https://azure.github.io/azureml-sdk-for-r/articles/deploying-models.html>

NEW QUESTION 215

- (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 217

- (Exam Topic 3)  
You have a dataset that contains 2,000 rows. You are building a machine learning classification model by using Azure Learning Studio. You add a Partition and Sample module to the experiment. You need to configure the module. You must meet the following requirements:  
> Divide the data into subsets  
> Assign the rows into folds using a round-robin method  
> Allow rows in the dataset to be reused  
How should you configure the module? To answer, select the appropriate options in the dialog box in the answer area.  
NOTE: Each correct selection is worth one point.

**Partition and Sample**

Partition or sample mode

Assign to Folds

Pick Fold

Sampling

Head

☐ Use replacement in the partitioning

☐ Randomized split

- A. Mastered  
 B. Not Mastered

**Answer:** A

**Explanation:**

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

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

**NEW QUESTION 221**

- (Exam Topic 3)

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

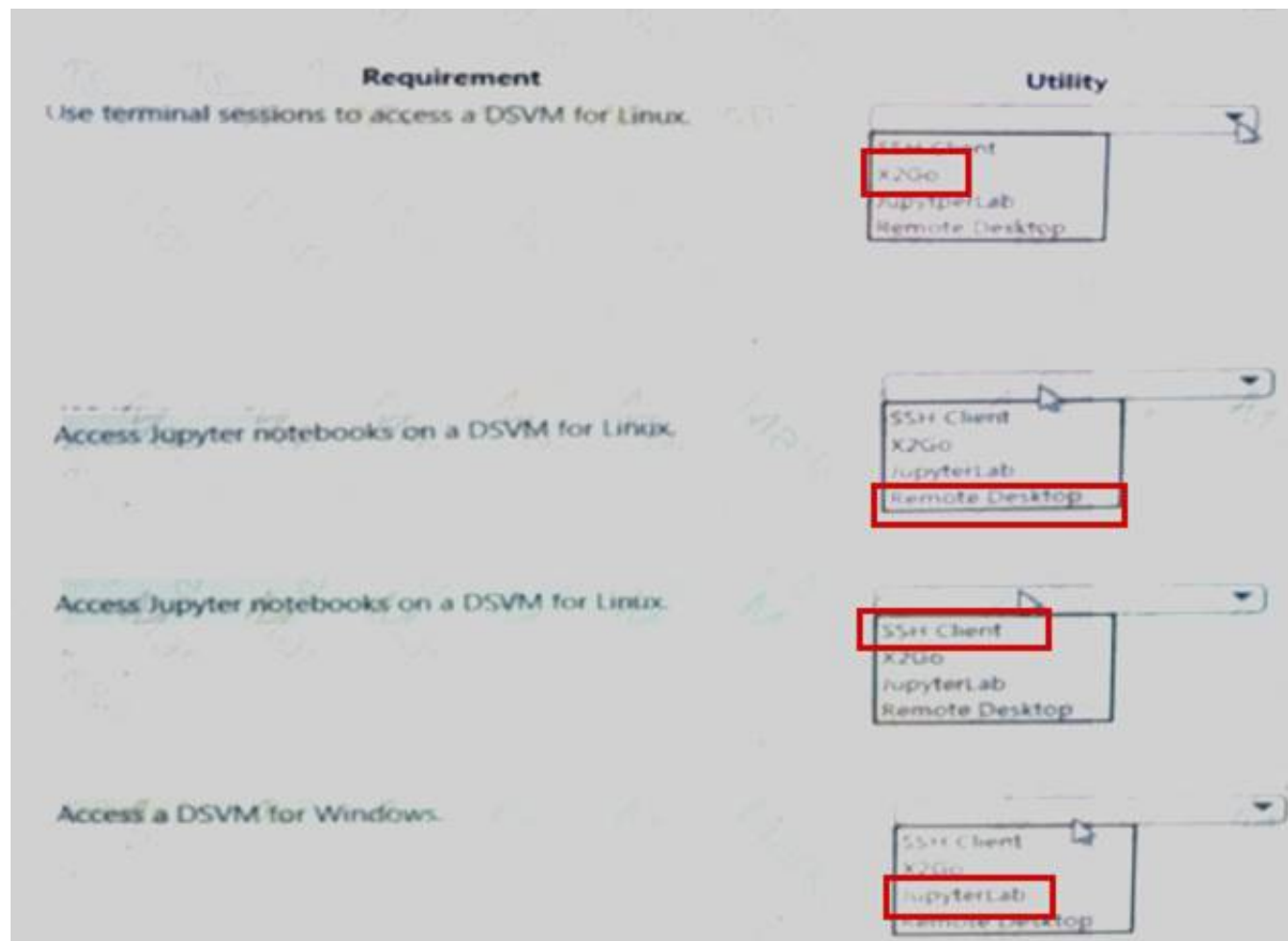
Which utilities should you use? To answer, select the appropriate options in the answer area. NOTE: Each correct selection is worth one point.

Requirement	Utility
Use terminal sessions to access a DSVM for Linux.	<input checked="" type="checkbox"/> SSH Client <input type="checkbox"/> X2Go <input type="checkbox"/> JupyterLab <input type="checkbox"/> Remote Desktop
Access Jupyter notebooks on a DSVM for Linux.	<input type="checkbox"/> SSH Client <input checked="" type="checkbox"/> X2Go <input checked="" type="checkbox"/> JupyterLab <input type="checkbox"/> Remote Desktop
Access Jupyter notebooks on a DSVM for Linux.	<input type="checkbox"/> SSH Client <input checked="" type="checkbox"/> X2Go <input checked="" type="checkbox"/> JupyterLab <input type="checkbox"/> Remote Desktop
Access a DSVM for Windows.	<input type="checkbox"/> SSH Client <input checked="" type="checkbox"/> X2Go <input type="checkbox"/> JupyterLab <input type="checkbox"/> Remote Desktop

- A. Mastered  
 B. Not Mastered

**Answer:** A

**Explanation:**



#### NEW QUESTION 225

- (Exam Topic 3)

You have a comma-separated values (CSV) file containing data from which you want to train a classification model.

You are using the Automated Machine Learning interface in Azure Machine Learning studio to train the classification model. You set the task type to Classification.

You need to ensure that the Automated Machine Learning process evaluates only linear models. What should you do?

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

**Answer: C**

#### Explanation:

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

#### NEW QUESTION 227

- (Exam Topic 3)

You create a binary classification model to predict whether a person has a disease. You need to detect possible classification errors.

Which error type should you choose for each description? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.



Description	Error type
A person has a disease. The model classifies the case as having a disease.	<div>▼</div> <div>True Positives</div> <div>True Negatives</div> <div>False Positives</div> <div>False Negatives</div>
A person does not have a disease. The model classifies the case as having no disease.	<div>▼</div> <div>True Positives</div> <div>True Negatives</div> <div>False Positives</div> <div>False Negatives</div>
A person does not have a disease. The model classifies the case as having a disease.	<div>▼</div> <div>True Positives</div> <div>True Negatives</div> <div>False Positives</div> <div>False Negatives</div>
A person has a disease. The model classifies the case as having no disease.	<div>▼</div> <div>True Positives</div> <div>True Negatives</div> <div>False Positives</div> <div>False Negatives</div>

- A. Mastered  
 B. Not Mastered

Answer: A

#### Explanation:

Box 1: True Positive

A true positive is an outcome where the model correctly predicts the positive class Box 2: True Negative

A true negative is an outcome where the model correctly predicts the negative class. Box 3: False Positive

A false positive is an outcome where the model incorrectly predicts the positive class. Box 4: False Negative

A false negative is an outcome where the model incorrectly predicts the negative class. Note: Let's make the following definitions:

"Wolf" is a positive class. "No wolf" is a negative class.

We can summarize our "wolf-prediction" model using a 2x2 confusion matrix that depicts all four possible outcomes:

Reference:

<https://developers.google.com/machine-learning/crash-course/classification/true-false-positive-negative>

#### NEW QUESTION 228

- (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 231

- (Exam Topic 3)

You are building a regression model tot estimating the number of calls during an event.

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

Each correct answer presents part of the solution. NOTE:

Each correct selection is worth one point.

- A. The label data must be a negative value.  
 B. The label data can be positive or negative,

- C. The label data must be a positive value
- D. The label data must be non discrete.
- E. The data must be whole numbers.

**Answer:** CE

**Explanation:**

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

- The response variable has a Poisson distribution.
- Counts cannot be negative. The method will fail outright if you attempt to use it with negative labels.
- A Poisson distribution is a discrete distribution; therefore, it is not meaningful to use this method with non-whole numbers.

References:

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

**NEW QUESTION 236**

- (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 classifies by using Azure Machine Learning Studio.

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

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

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

**Answer:** D

**Explanation:**

Pearson's correlation statistic, or Pearson's correlation coefficient, is also known in statistical models as the r value. For any two variables, it returns a value that indicates the strength of the correlation

Pearson's correlation coefficient is the test statistics that measures the statistical relationship, or association, between two continuous variables. It is known as the best method of measuring the association between variables of interest because it is based on the method of covariance. It gives information about the magnitude of the association, or correlation, as well as the direction of the relationship.

Reference:

<https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/filter-based-feature-selection> <https://www.statisticssolutions.com/pearsons-correlation-coefficient/>

**NEW QUESTION 245**

- (Exam Topic 3)

You are creating a new Azure Machine Learning pipeline using the designer.

The pipeline must train a model using data in a comma-separated values (CSV) file that is published on a website. You have not created a dataset for this file.

You need to ingest the data from the CSV file into the designer pipeline using the minimal administrative effort.

Which module should you add to the pipeline in Designer?

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

**Answer:** D

**Explanation:**

The preferred way to provide data to a pipeline is a Dataset object. The Dataset object points to data that lives in or is accessible from a datastore or at a Web URL. The Dataset class is abstract, so you will create an instance of either a FileDataset (referring to one or more files) or a TabularDataset that's created by from

one or more files with delimited columns of data.

Example:

```
from azureml.core import Dataset
```

```
iris_tabular_dataset = Dataset.Tabular.from_delimited_files([(def_blob_store, 'train-dataset/iris.csv')])
```

 Reference:

<https://docs.microsoft.com/en-us/azure/machine-learning/how-to-create-your-first-pipeline>

#### NEW QUESTION 246

- (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 247

- (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 251

- (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 254

- (Exam Topic 3)

You use Azure Machine Learning to train a model based on a dataset named dataset1. You define a dataset monitor and create a dataset named dataset2 that contains new data.

You need to compare dataset1 and dataset2 by using the Azure Machine Learning SDK for Python. Which method of the DataDriftDetector class should you use?



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

**Answer:** C

**Explanation:**

A backfill run is used to see how data changes over time. Reference:

<https://docs.microsoft.com/en-us/python/api/azureml-datadrift/azureml.datadrift.datadriftdetector.datadriftdetect>

**NEW QUESTION 259**

- (Exam Topic 3)

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

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

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

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

**Answer:** C

**Explanation:**

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

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

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

**NEW QUESTION 262**

- (Exam Topic 3)

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

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

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

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

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

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

Solution: Run the following code:

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

Does the solution meet the goal?

- A. Yes
- B. No

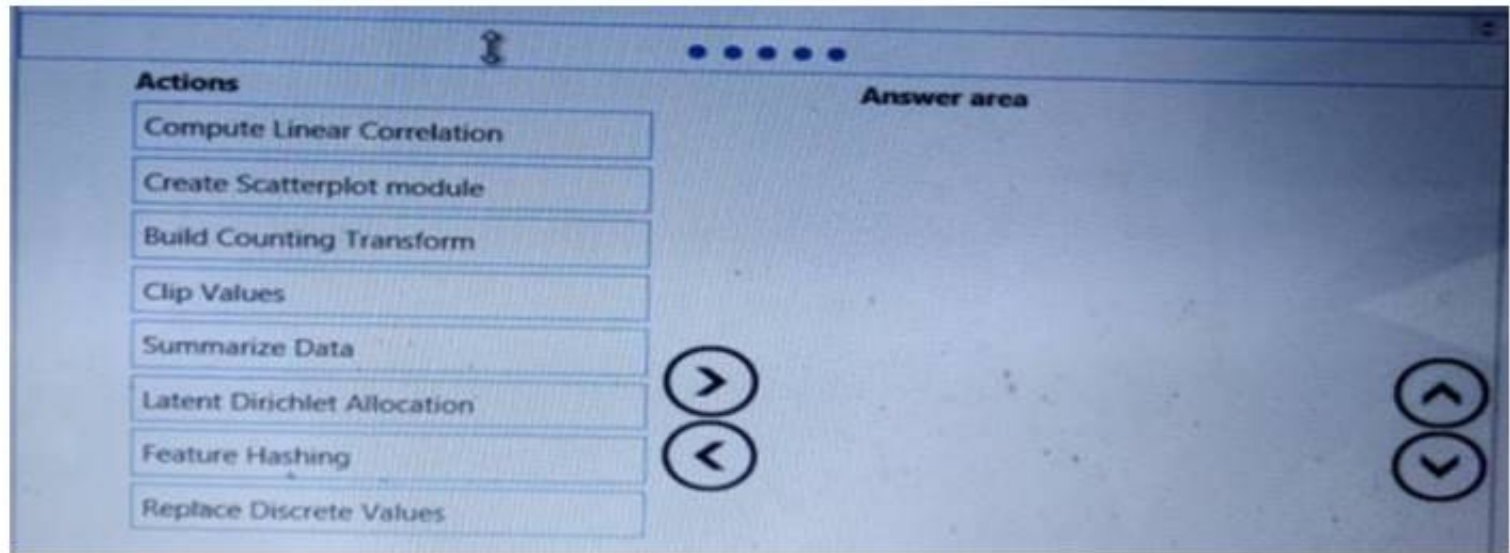
**Answer:** A

**NEW QUESTION 267**

- (Exam Topic 2)

You need to visually identify whether outliers exist in the Age column and quantify the outliers before the outliers are removed.

Which three Azure Machine Learning Studio modules should you use in sequence? To answer, move the appropriate modules from the list of modules to the answer area and arrange them in the correct order.



- A. Mastered
- B. Not Mastered

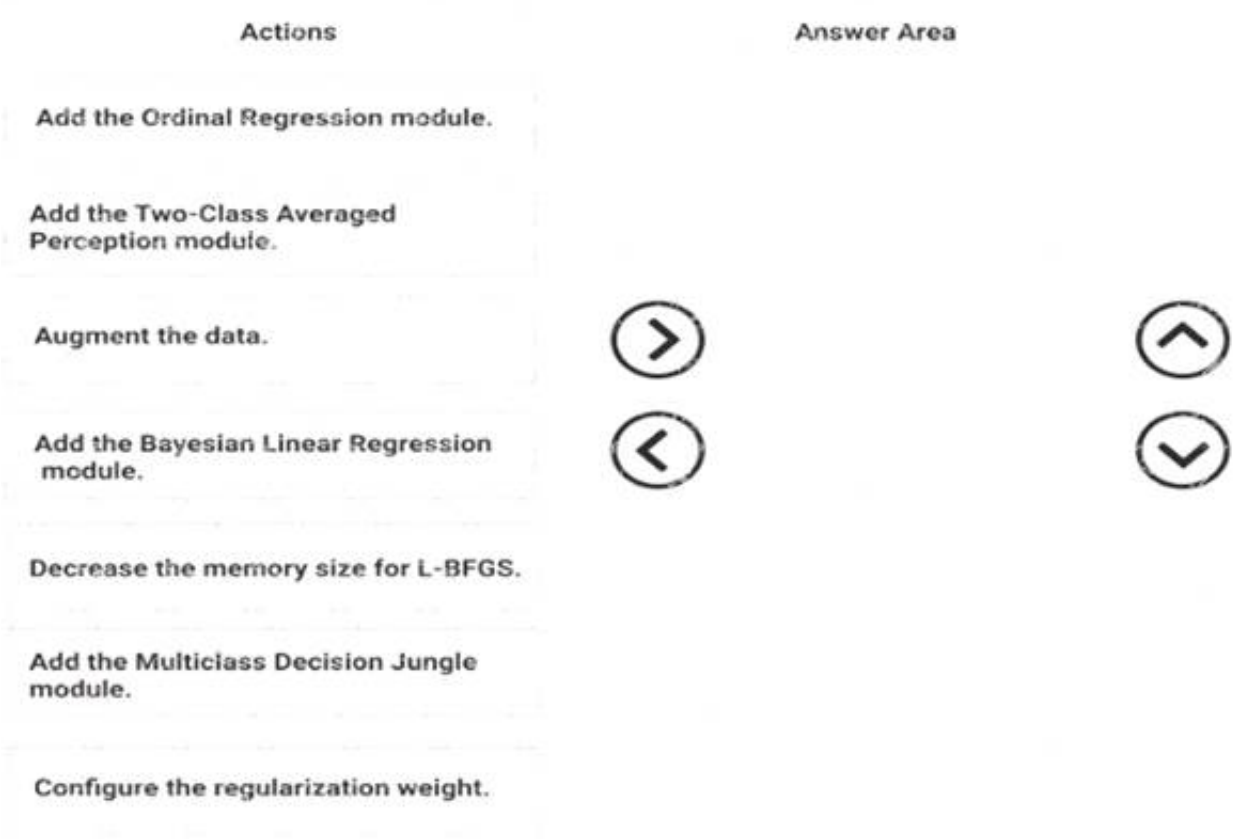
Answer: A

Explanation:

Create Scatterplot Summarize Data Clip Values  
You can use the Clip Values module in Azure Machine Learning Studio, to identify and optionally replace data values that are above or below a specified threshold. This is useful when you want to remove outliers or replace them with a mean, a constant, or other substitute value.  
References:  
<https://blogs.msdn.microsoft.com/azuredev/2017/05/27/data-cleansing-tools-in-azure-machine-learning/> <https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/clip-values>

NEW QUESTION 268

- (Exam Topic 2)  
You need to correct the model fit issue.  
Which three actions should you perform in sequence? To answer, move the appropriate actions from the list of actions to the answer area and arrange them in the correct order.



- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

Step 1: Augment the data  
Scenario: Columns in each dataset contain missing and null values. The datasets also contain many outliers.  
Step 2: Add the Bayesian Linear Regression module.  
Scenario: You produce a regression model to predict property prices by using the Linear Regression and Bayesian Linear Regression modules.  
Step 3: Configure the regularization weight.  
Regularization typically is used to avoid overfitting. For example, in L2 regularization weight, type the value to use as the weight for L2 regularization. We recommend that you use a non-zero value to avoid overfitting.  
Scenario:  
Model fit: The model shows signs of overfitting. You need to produce a more refined regression model that reduces the overfitting.

NEW QUESTION 272

- (Exam Topic 2)

You need to implement early stopping criteria as suited in the model training requirements.

Which three code segments should you use to develop the solution? To answer, move the appropriate code segments from the list of code segments to the answer area and arrange them in the correct order.

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

Code segments	Answer Area
<pre>early_termination_policy = TruncationSelectionPolicy(evaluation_interval=1, truncation_percentage=20, delay_evaluation=5)</pre>	
<pre>import TruncationSelectionPolicy</pre>	
<pre>from azureml.train.hyperdrive</pre>	⬅️ ⬆️
<pre>import BanditPolicy</pre>	⬆️ ⬇️
<pre>early_termination_policy = BanditPolicy (slack_factor = 0.1, evaluation_interval=1, delay_evaluation=5)</pre>	

- A. Mastered
- B. Not Mastered

**Answer:** A

**Explanation:**

You need to implement an early stopping criterion on models that provides savings without terminating promising jobs.

Truncation selection cancels a given percentage of lowest performing runs at each evaluation interval. Runs are compared based on their performance on the primary metric and the lowest X% are terminated. Example:

```
from azureml.train.hyperdrive import TruncationSelectionPolicy
early_termination_policy = TruncationSelectionPolicy(evaluation_interval=1, truncation_percentage=20, delay_evaluation=5)
```

**NEW QUESTION 273**

- (Exam Topic 1)

You need to implement a feature engineering strategy for the crowd sentiment local models. What should you do?

- A. Apply an analysis of variance (ANOVA).
- B. Apply a Pearson correlation coefficient.
- C. Apply a Spearman correlation coefficient.
- D. Apply a linear discriminant analysis.

**Answer:** D

**Explanation:**

The linear discriminant analysis method works only on continuous variables, not categorical or ordinal variables.

Linear discriminant analysis is similar to analysis of variance (ANOVA) in that it works by comparing the means of the variables.

Scenario:

Data scientists must build notebooks in a local environment using automatic feature engineering and model building in machine learning pipelines.

Experiments for local crowd sentiment models must combine local penalty detection data. All shared features for local models are continuous variables.

**NEW QUESTION 275**

- (Exam Topic 1)

You need to resolve the local machine learning pipeline performance issue. What should you do?

- A. Increase Graphic Processing Units (GPUs).
- B. Increase the learning rate.
- C. Increase the training iterations.
- D. Increase Central Processing Units (CPUs).

**Answer:** A

**NEW QUESTION 280**

- (Exam Topic 1)

You need to implement a new cost factor scenario for the ad response models as illustrated in the performance curve exhibit.

Which technique should you use?

- A. Set the threshold to 0.5 and retrain if weighted Kappa deviates +/- 5% from 0.45.



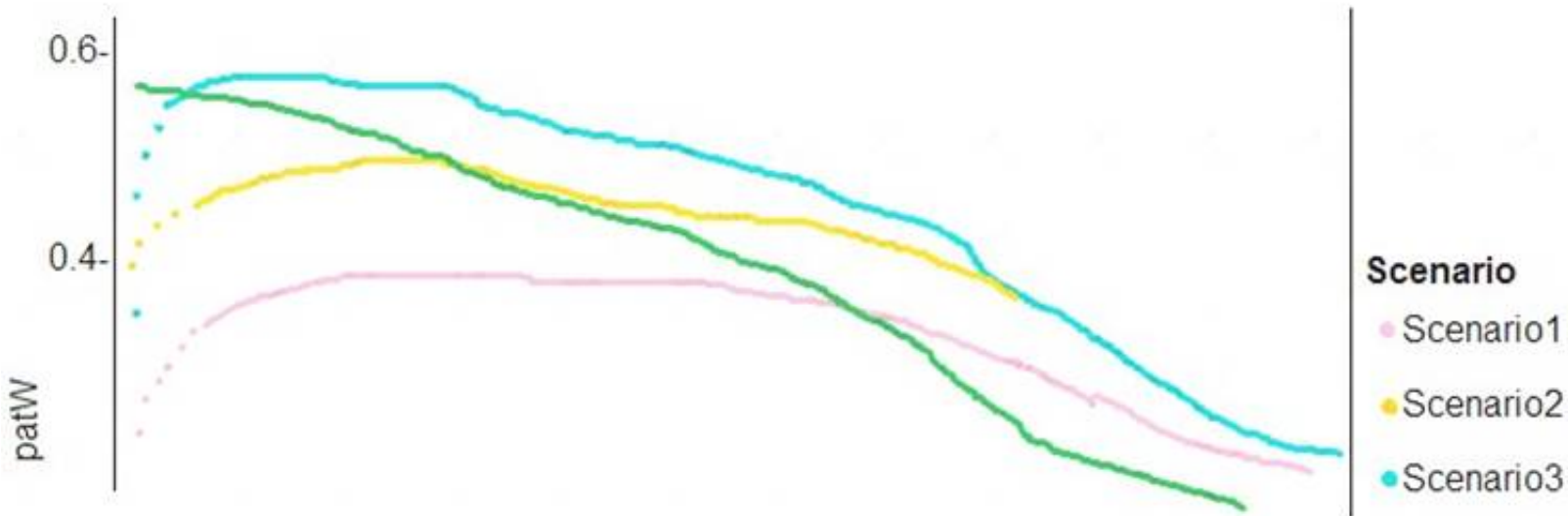
- B. Set the threshold to 0.05 and retrain if weighted Kappa deviates +/- 5% from 0.5.
- C. Set the threshold to 0.2 and retrain if weighted Kappa deviates +/- 5% from 0.6.
- D. Set the threshold to 0.75 and retrain if weighted Kappa deviates +/- 5% from 0.15.

**Answer:** A

**Explanation:**

Scenario:

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



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

**NEW QUESTION 284**

.....

## Thank You for Trying Our Product

### We offer two products:

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

2nd - Questions and Answers in PDF Format

### DP-100 Practice Exam Features:

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

**100% Actual & Verified — Instant Download, Please Click**  
**[Order The DP-100 Practice Test Here](#)**