

# Microsoft

## Exam Questions DP-100

Designing and Implementing a Data Science Solution on Azure



### NEW QUESTION 1

- (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.

An IT department creates the following Azure resource groups and resources:

Resource group	Resources
ml_resources	<ul style="list-style-type: none"><li>• an Azure Machine Learning workspace named amlworkspace</li><li>• an Azure Storage account named amlworkspace12345</li><li>• an Application Insights instance named amlworkspace54321</li><li>• an Azure Key Vault named amlworkspace67890</li><li>• an Azure Container Registry named amlworkspace09876</li></ul>
general_compute	<p>A virtual machine named mlvm with the following configuration:</p> <ul style="list-style-type: none"><li>• Operating system: Ubuntu Linux</li><li>• Software installed: Python 3.6 and Jupyter Notebooks</li><li>• Size: NC6 (6 vCPUs, 1 vGPU, 56 Gb RAM)</li></ul>

The IT department creates an Azure Kubernetes Service (AKS)-based inference compute target named aks-cluster in the Azure Machine Learning workspace. You have a Microsoft Surface Book computer with a GPU. Python 3.6 and Visual Studio Code are installed. You need to run a script that trains a deep neural network (DNN) model and logs the loss and accuracy metrics.

Solution: Install the Azure ML SDK on the Surface Book. Run Python code to connect to the workspace. Run the training script as an experiment on the aks-cluster compute target.

Does the solution meet the goal?

- A. Yes  
B. No

**Answer: B**

#### Explanation:

Need to attach the mlvm virtual machine as a compute target in the Azure Machine Learning workspace.

Reference:

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

### NEW QUESTION 2

- (Exam Topic 3)

You are producing a multiple linear regression model in Azure Machine Learning Studio. Several independent variables are highly correlated.

You need to select appropriate methods for conducting effective feature engineering on all the data.

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

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

- A. Mastered  
B. Not Mastered

**Answer: A**

#### Explanation:

Step 1: Use the Filter Based Feature Selection module

Filter Based Feature Selection identifies the features in a dataset with the greatest predictive power.

The module outputs a dataset that contains the best feature columns, as ranked by predictive power. It also outputs the names of the features and their scores from the selected metric.

Step 2: Build a counting transform

A counting transform creates a transformation that turns count tables into features, so that you can apply the transformation to multiple datasets.

Step 3: Test the hypothesis using t-Test

References:  
<https://docs.microsoft.com/bs-latn-ba/azure/machine-learning/studio-module-reference/filter-based-feature-selec>

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

**NEW QUESTION 3**

- (Exam Topic 3)

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

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

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

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

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

Solution: Calculate the column median value and use the median value as the replacement for any missing value in the column.

Does the solution meet the goal?

A. Yes

B. No

**Answer: B**

**Explanation:**

Use the Multiple Imputation by Chained Equations (MICE) method. References: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3074241/>

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

**NEW QUESTION 4**

- (Exam Topic 3)

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

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

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

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

Does the solution meet the goal?

A. Yes

B. No

**Answer: B**

**Explanation:**

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

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

References:

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

**NEW QUESTION 5**

- (Exam Topic 3)

Your Azure Machine Learning workspace has a dataset named real\_estate\_data. A sample of the data in the dataset follows.

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

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

NOTE: Each correct selection is worth one point.

## Answer Area

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

ws = Workspace.from_config()
training_cluster = ComputeTarget(workspace=ws, name= 'aml-cluster1')
real_estate_ds = ws.datasets.get('real_estate_data')
split1_ds, split2_ds = real_estate_ds.random_split(percentage=0.7, seed=123)
automl_run_config = RunConfiguration(framework= "python")
automl_config = AutoMLConfig(
    task= 'regression',
    compute_target= training_cluster,
    run_configuration=automl_run_config,
    primary_metric='r2_score',
    =split1_ds,
    X
    Y
    X_valid
    Y_valid
    training_data
    =split2_ds
    X
    Y
    X_valid
    Y_valid
    validation_data
    training_data
    ='price')
    y
    y_valid
    y_max
    label_column_name
    exclude_nan_labels
```

- A. Mastered  
 B. Not Mastered

**Answer: A**

### Explanation:

Box 1: training\_data

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

Box 2: validation\_data

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

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

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

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

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

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

```
)
```

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

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

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

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

### NEW QUESTION 6

- (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 <b>Classification - Accuracy</b> 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 7**

- (Exam Topic 3)

You use the Azure Machine Learning designer to create and run a training pipeline.

The pipeline must be run every night to inference predictions from a large volume of files. The folder where the files will be stored is defined as a dataset.

You need to publish the pipeline as a REST service that can be used for the nightly inferencing run. What should you do?

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

**Answer:** A

**Explanation:**

Azure Machine Learning Batch Inference targets large inference jobs that are not time-sensitive. Batch Inference provides cost-effective inference compute scaling, with unparalleled throughput for asynchronous applications. It is optimized for high-throughput, fire-and-forget inference over large collections of data.

You can submit a batch inference job by pipeline\_run, or through REST calls with a published pipeline. Reference:

<https://github.com/Azure/MachineLearningNotebooks/blob/master/how-to-use-azureml/machine-learning-pipeli>

**NEW QUESTION 8**

- (Exam Topic 3)

You create a new Azure subscription. No resources are provisioned in the subscription. You need to create an Azure Machine Learning workspace.

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. Run Python code that uses the Azure ML SDK library and calls the Workspace.create method with name, subscription\_id, resource\_group, and location parameters.
- B. Use an Azure Resource Management template that includes a Microsoft.MachineLearningServices/ workspaces resource and its dependencies.
- C. Use the Azure Command Line Interface (CLI) with the Azure Machine Learning extension to call the az group create function with --name and --location parameters, and then the az ml workspace create function, specifying -w and -g parameters for the workspace name and resource group.
- D. Navigate to Azure Machine Learning studio and create a workspace.
- E. Run Python code that uses the Azure ML SDK library and calls the Workspace.get method with name, subscription\_id, and resource\_group parameters.

**Answer:** BCD

**Explanation:**

B: You can use an Azure Resource Manager template to create a workspace for Azure Machine Learning. Example:

```
{
  "type": "Microsoft.MachineLearningServices/workspaces",
  ...
}
```

C: You can create a workspace for Azure Machine Learning with Azure CLI Install the machine learning extension.  
 Create a resource group: az group create --name <resource-group-name> --location <location>



To create a new workspace where the services are automatically created, use the following command: `az ml workspace create -w <workspace-name> -g <resource-group-name>`

D: You can create and manage Azure Machine Learning workspaces in the Azure portal.

- Sign in to the Azure portal by using the credentials for your Azure subscription.
- In the upper-left corner of Azure portal, select + Create a resource.
- Use the search bar to find Machine Learning.
- Select Machine Learning.
- In the Machine Learning pane, select Create to begin.

Home > New > Machine Learning >

## Machine Learning

Create a machine learning workspace

Basics Networking Advanced Tags Review + create

### Project details

Select the subscription to manage deployed resources and costs. Use resource groups like folders to organize and manage all your resources.

Subscription \* ⓘ Documentation-team ▼

Resource group \* ⓘ docs-ws ▼

[Create new](#)

### Workspace details


Specify the name, region, and edition for the workspace.

Workspace name \* ⓘ docs-mlw ✓

Region \* ⓘ West Central US ▼

Workspace edition \* ⓘ

- Basic
- Basic
- Enterprise

 For your convenience, these resources are available in your workspace: Application Insights, Azure Key Vault

Reference:

<https://docs.microsoft.com/en-us/azure/machine-learning/how-to-create-workspace-template> <https://docs.microsoft.com/en-us/azure/machine-learning/how-to-manage-workspace-cli> <https://docs.microsoft.com/en-us/azure/machine-learning/how-to-manage-workspace>

### NEW QUESTION 9

- (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 10**

- (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 10**

- (Exam Topic 3)

You are moving a large dataset from Azure Machine Learning Studio to a Weka environment. You need to format the data for the Weka environment.

Which module should you use?

- A. Convert to CSV
- B. Convert to Dataset
- C. Convert to ARFF
- D. Convert to SVMLight

**Answer:** C

**Explanation:**

Use the Convert to ARFF module in Azure Machine Learning Studio, to convert datasets and results in Azure Machine Learning to the attribute-relation file format used by the Weka toolset. This format is known as ARFF.

The ARFF data specification for Weka supports multiple machine learning tasks, including data preprocessing, classification, and feature selection. In this format, data is organized by entites and their attributes, and is contained in a single text file.

References:

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

**NEW QUESTION 13**

- (Exam Topic 3)

You use the Azure Machine Learning service to create a tabular dataset named training.data. You plan to use this dataset in a training script.

You create a variable that references the dataset using the following code: training\_ds = workspace.datasets.get("training\_data")

You define an estimator to run the script.

You need to set the correct property of the estimator to ensure that your script can access the training.data dataset

Which property should you set?

A)

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

B)

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

C)

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

D)

```
source_directory = training_ds
```

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

**Answer:** A

**Explanation:**

Example:

# Get the training dataset

diabetes\_ds = ws.datasets.get("Diabetes Dataset")

# Create an estimator that uses the remote compute hyper\_estimator = SKLearn(source\_directory=experiment\_folder,

inputs=[diabetes\_ds.as\_named\_input('diabetes')], # Pass the dataset as an input compute\_target = cpu\_cluster, conda\_packages=['pandas','ipykernel','matplotlib'],

`pip_packages=['azureml-sdk','argparse','pyarrow'], entry_script='diabetes_training.py')`

Reference:

<https://notebooks.azure.com/GraemeMalcolm/projects/azureml-primers/html/04%20-%20Optimizing%20Model>

#### NEW QUESTION 14

- (Exam Topic 3)

You are conducting feature engineering to prepuce data for further analysis. The data includes seasonal patterns on inventory requirements. You need to select the appropriate method to conduct feature engineering on the data. Which method should you use?

- A. Exponential Smoothing (ETS) function.
- B. One Class Support Vector Machine module
- C. Time Series Anomaly Detection module
- D. Finite Impulse Response (FIR) Filter module.

**Answer:** D

#### NEW QUESTION 16

- (Exam Topic 3)

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

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

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

NOTE: Each correct selection is worth one point.

	Yes	No
By defining sampling in this manner, every possible combination of the parameters will be tested.	<input type="radio"/>	<input type="radio"/>
Random values of the learning_rate parameter will be selected from a normal distribution with a mean of 10 and a standard deviation of 3.	<input type="radio"/>	<input type="radio"/>
The keep_probability parameter value will always be either <b>0.05</b> or <b>0.1</b> .	<input type="radio"/>	<input type="radio"/>
Random values for the number_of_hidden_layers parameter will be selected from a normal distribution with a mean of 3 and a standard deviation of 5.	<input type="radio"/>	<input type="radio"/>

- A. Mastered
- B. Not Mastered

**Answer:** A

#### Explanation:

Box 1: Yes

In random sampling, hyperparameter values are randomly selected from the defined search space. Random sampling allows the search space to include both discrete and continuous hyperparameters.

Box 2: Yes

learning\_rate has a normal distribution with mean value 10 and a standard deviation of 3.

Box 3: No

keep\_probability has a uniform distribution with a minimum value of 0.05 and a maximum value of 0.1.

Box 4: No

number\_of\_hidden\_layers takes on one of the values [3, 4, 5].

Reference:

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

#### NEW QUESTION 20

- (Exam Topic 3)

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

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

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



NOTE: Each correct selection is worth one point.

	Yes	No
Hyperparameter combinations for the runs are selected based on how previous samples performed in the previous experiment run.	<input type="radio"/>	<input type="radio"/>
The learning rate value 0.09 might be used during model training.	<input type="radio"/>	<input type="radio"/>
You can define an early termination policy for this hyperparameter tuning run.	<input type="radio"/>	<input type="radio"/>

- A. Mastered
- B. Not Mastered

**Answer:** A

**Explanation:**

Box 1: Yes

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

Box 2: Yes

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

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

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

**NEW QUESTION 22**

- (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 26**

- (Exam Topic 3)

You are a data scientist working for a hotel booking website company. You use the Azure Machine Learning service to train a model that identifies fraudulent transactions.

You must deploy the model as an Azure Machine Learning real-time web service using the Model.deploy method in the Azure Machine Learning SDK. The deployed web service must return real-time predictions of fraud based on transaction data input.

You need to create the script that is specified as the entry\_script parameter for the InferenceConfig class used to deploy the model.

What should the entry script do?

- A. Start a node on the inference cluster where the web service is deployed.
- B. Register the model with appropriate tags and properties.
- C. Create a Conda environment for the web service compute and install the necessary Python packages.
- D. Load the model and use it to predict labels from input data.
- E. Specify the number of cores and the amount of memory required for the inference compute.

**Answer:** D

**Explanation:**

The entry script receives data submitted to a deployed web service and passes it to the model. It then takes the response returned by the model and returns that to the client. The script is specific to your model. It must understand the data that the model expects and returns.

The two things you need to accomplish in your entry script are: Loading your model (using a function called init())

Running your model on input data (using a function called run()) Reference:

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

**NEW QUESTION 30**

- (Exam Topic 3)

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

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

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

- A. run.upload\_file('row\_count', './data.csv')
- B. run.log('row\_count', rows)
- C. run.tag('row\_count', rows)
- D. run.log\_table('row\_count', rows)
- E. run.log\_row('row\_count', rows)

**Answer:** B

**Explanation:**

Log a numerical or string value to the run with the given name using log(name, value, description="). Logging a metric to a run causes that metric to be stored in the run record in the experiment. You can log the same metric multiple times within a run, the result being considered a vector of that metric.

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

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

**NEW QUESTION 34**

- (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 37**

- (Exam Topic 3)

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

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

You have a Python script named train.py in a local folder named scripts. The script trains a regression model by using scikit-learn. The script includes code to load a training data file which is also located in the scripts folder.

You must run the script as an Azure ML experiment on a compute cluster named aml-compute.

You need to configure the run to ensure that the environment includes the required packages for model training. You have instantiated a variable named aml-compute that references the target compute cluster.

Solution: Run the following code:

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

Does the solution meet the goal?

- A. Yes
- B. No

**Answer: B**



**Explanation:**

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

Example:

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

Reference:

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

**NEW QUESTION 38**

- (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 Microsoft 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 43**

- (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 47**

- (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 49**

- (Exam Topic 3)

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

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

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

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

Solution: Run the following code:

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

Does the solution meet the goal?

A. Yes

B. No

**Answer:** B

**Explanation:**

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

Compare with this example, the pipeline train step depends on the process\_step\_output output of the pipeline process step:

```
from azureml.pipeline.core import Pipeline, PipelineData
from azureml.pipeline.steps import PythonScriptStep

datastore = ws.get_default_datastore()
process_step_output = PipelineData("processed_data", datastore=datastore)
process_step = PythonScriptStep(script_name="process.py",
                                arguments=["--data_for_train", process_step_output],
                                outputs=[process_step_output],
                                compute_target=aml_compute,
                                source_directory=process_directory)

train_step = PythonScriptStep(script_name="train.py",
                                arguments=["--data_for_train", process_step_output],
                                inputs=[process_step_output],
                                compute_target=aml_compute,
                                source_directory=train_directory)

pipeline = Pipeline(workspace=ws, steps=[process_step, train_step])
```

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

### NEW QUESTION 53

- (Exam Topic 3)

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

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

You create 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 55

- (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)
```

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:

```
from sklearn.metrics import roc_auc_score
import logging
# code to train model omitted
auc = roc_auc_score(y_test, y_predicted)
logging.info("AUC: " + str(auc))
```

Does the solution meet the goal?

- A. Yes
- B. No

**Answer: A**

**Explanation:**

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 57**

- (Exam Topic 3)

You are analyzing a dataset containing historical data from a local taxi company. You are developing a regression model.

You must predict the fare of a taxi trip.

You need to select performance metrics to correctly evaluate the regression model. Which two metrics can you use? Each correct answer presents a complete solution. NOTE: Each correct selection is worth one point.

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

**Answer:** BE

**Explanation:**

References:

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

**NEW QUESTION 58**

- (Exam Topic 3)

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

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

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

You define a DataReference object by running the following code:

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

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

- A. 

```
import os
import argparse
import pandas as pd

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

```
import os
import argparse
import pandas as pd

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

```
import pandas as pd

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



```
D. import os
import argparse
import pandas as pd

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

E. import os
import argparse
import pandas as pd

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

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

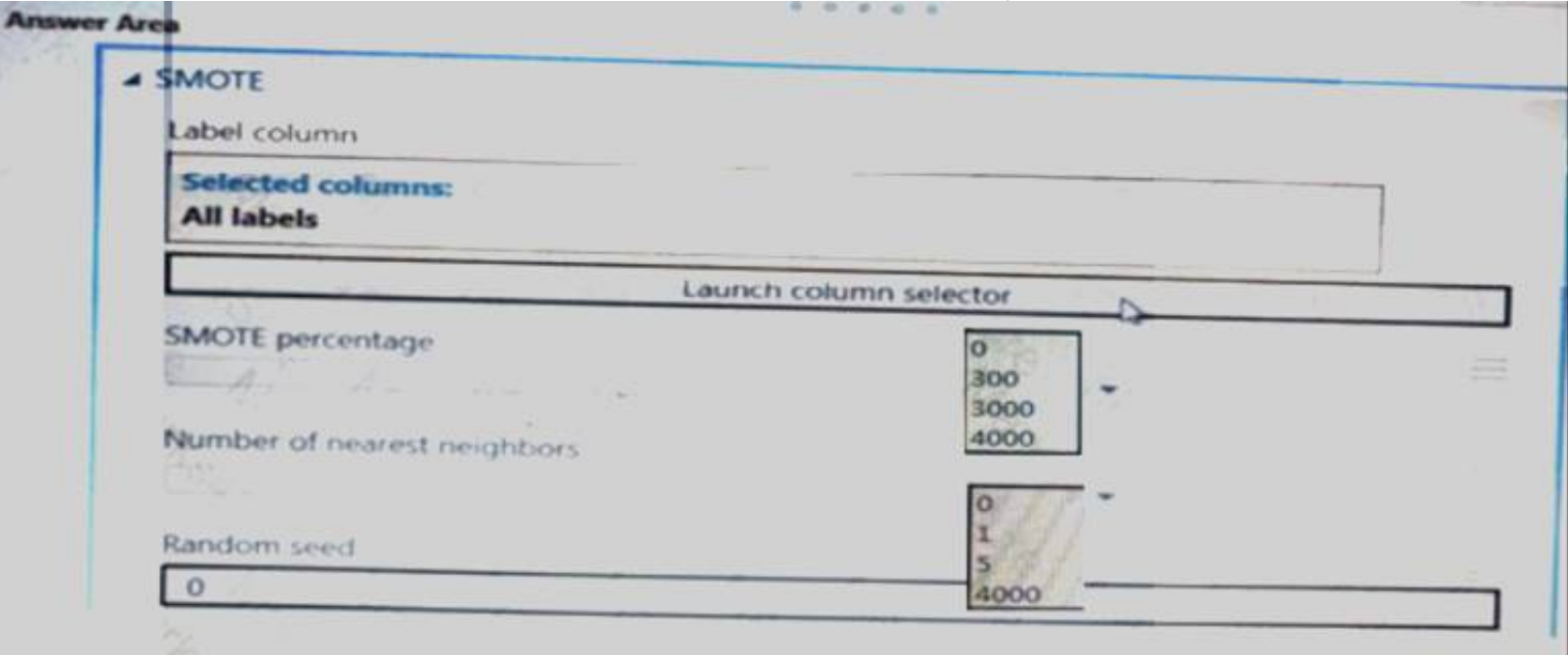
**Answer:** E

**Explanation:**

Example:  
data\_folder = args.data\_folder  
# Load Train and Test data  
train\_data = pd.read\_csv(os.path.join(data\_folder, 'data.csv')) Reference:  
<https://www.element61.be/en/resource/azure-machine-learning-services-complete-toolbox-ai>

**NEW QUESTION 60**

- (Exam Topic 3)  
You create an experiment in Azure Machine Learning Studio- You add a training dataset that contains 10.000 rows. The first 9.000 rows represent class 0 (90 percent). The first 1.000 rows represent class 1 (10 percent).  
The training set is unbalanced between two Classes. You must increase the number of training examples for class 1 to 4,000 by using data rows. You add the Synthetic Minority Oversampling Technique (SMOTE) module to the experiment.  
You need to configure the module.  
Which values should you use? To answer, select the appropriate options in the dialog box in the answer area. NOTE: Each correct selection is worth one point.

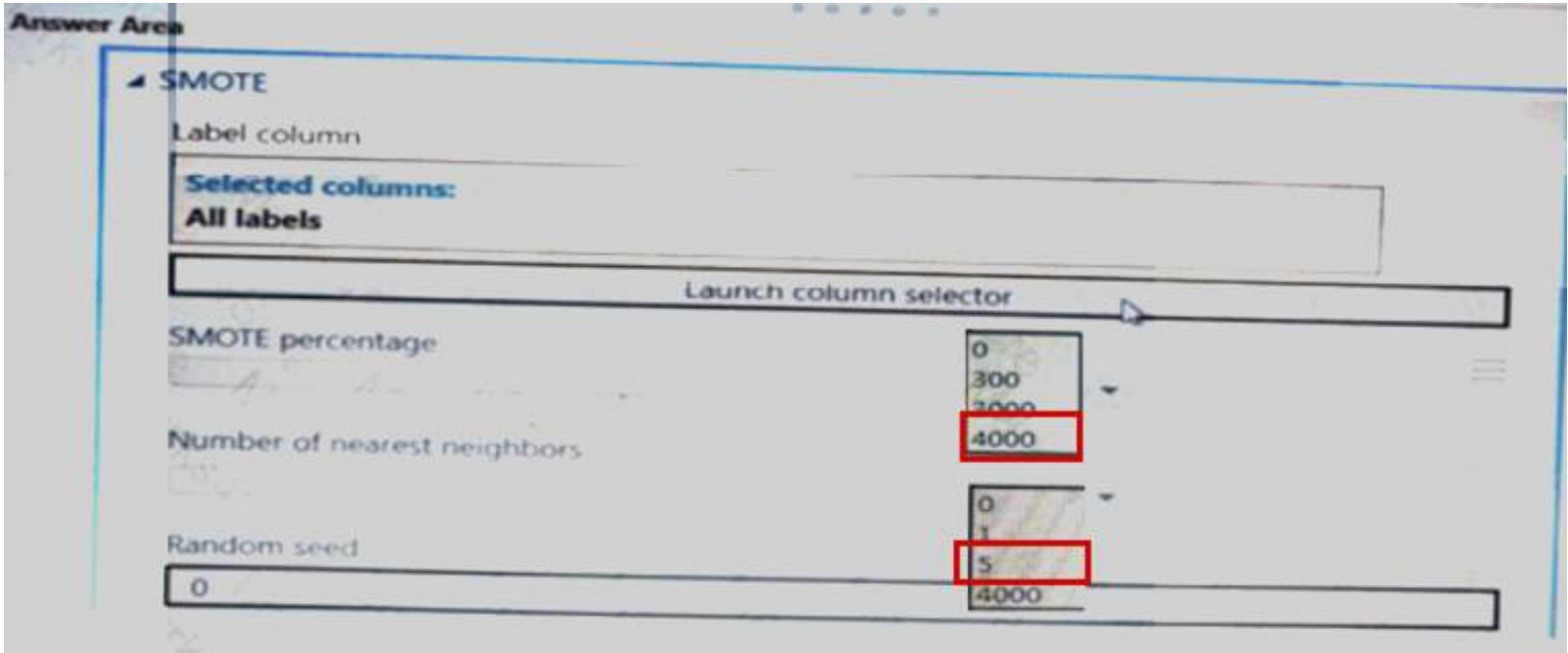


- A. Mastered
- B. Not Mastered

**Answer:** A

**Explanation:**





**NEW QUESTION 65**

- (Exam Topic 3)

You train and register a model in your Azure Machine Learning workspace.

You must publish a pipeline that enables client applications to use the model for batch inferencing. You must use a pipeline with a single ParallelRunStep step that runs a Python inferencing script to get predictions from the input data.

You need to create the inferencing script for the ParallelRunStep pipeline step.

Which two functions should you include? Each correct answer presents part of the solution.

NOTE: Each correct selection is worth one point.

- A. run(mini\_batch) D
- B. main()
- C. batch()
- D. init()
- E. score(mini\_batch)

**Answer:** AD

**Explanation:**

Reference:

<https://github.com/Azure/MachineLearningNotebooks/tree/master/how-to-use-azureml/machine-learningpipelin>

**NEW QUESTION 70**

- (Exam Topic 3)

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

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



- A. Mastered
- B. Not Mastered

**Answer:** A

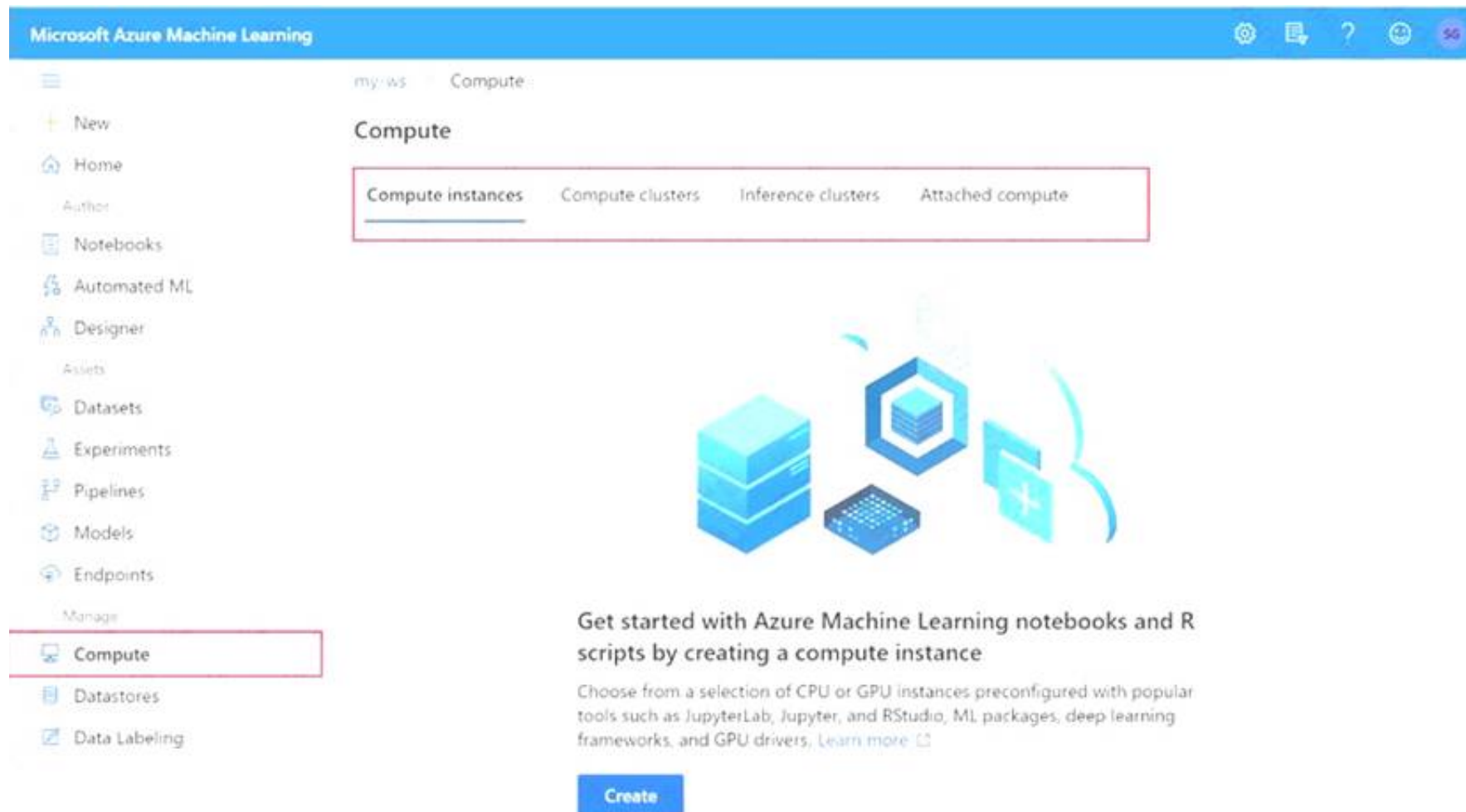
**Explanation:**

Box 1: Yes

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

It is displayed as a compute cluster. View compute targets

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



Box 3: Yes

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

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

## NEW QUESTION 72

- (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. You must not apply an early termination policy.

learning\_rate: any value between 0.001 and 0.1

• batch\_size: 16, 32, or 64

You need to configure the sampling method for the Hyperdrive experiment

Which two sampling methods can you use? Each correct answer is a complete solution. NOTE: Each correct selection is worth one point.

- A. Grid sampling
- B. No sampling
- C. Bayesian sampling
- D. Random sampling

**Answer:** CD

### Explanation:

C: Bayesian sampling is based on the Bayesian optimization algorithm and makes intelligent choices on the hyperparameter values to sample next. It picks the sample based on how the previous samples performed, such that the new sample improves the reported primary metric.

Bayesian sampling does not support any early termination policy Example:

```
from azureml.train.hyperdrive import BayesianParameterSampling from azureml.train.hyperdrive import uniform, choice param_sampling =
```

```
BayesianParameterSampling( { "learning_rate": uniform(0.05, 0.1),
```

```
"batch_size": choice(16, 32, 64, 128)
```

```
}
```

```
)
```

D: In random sampling, hyperparameter values are randomly selected from the defined search space. Random sampling allows the search space to include both discrete and continuous hyperparameters.

Reference:

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

## NEW QUESTION 77

- (Exam Topic 3)

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

What should you create?

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

**Answer:** E

## NEW QUESTION 82

- (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 84

- (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 88

- (Exam Topic 3)

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

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

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

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

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

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

NOTE: Each correct selection is worth one point.

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

- A. Mastered
- B. Not Mastered

**Answer:** A

#### Explanation:

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

Box 3: No

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

Reference:

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

#### NEW QUESTION 89

- (Exam Topic 3)

You create machine learning models by using Azure Machine Learning.

You plan to train and score models by using a variety of compute contexts. You also plan to create a new compute resource in Azure Machine Learning studio.

You need to select the appropriate compute types.

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



NOTE: Each correct selection is worth one point.

**Compute types**

Attached compute

Inference cluster

Training cluster

**Answer Area**

Requirement	Compute type
Train models by using the Azure Machine Learning designer.	<div>Compute type</div>
Score new data through a trained model published as a real-time web service.	<div>Compute type</div>
Train models by using an Azure Databricks cluster.	<div>Compute type</div>
Deploy models by using the Azure Machine Learning designer.	<div>Compute type</div>

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

Box 1: Attached compute

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

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

NEW QUESTION 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 AciWeb service 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_web service_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 95

- (Exam Topic 3)

You are performing feature engineering on a dataset.

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

You need to add the new feature to the dataset.

Which Azure Machine Learning Studio module should you use?

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

Answer: A



**Explanation:**

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

**NEW QUESTION 98**

- (Exam Topic 3)

You plan to preprocess text from CSV files. You load the Azure Machine Learning Studio default stop words list.

You need to configure the Preprocess Text module to meet the following requirements:

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

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

Preprocess Text

Language  
English

Remove by part of speech  
False

Text column to clean  
Selected columns:  
Column names: String, Feature  
Launch column selector

☐ Remove stop words

☐ Lemmatization

☐ Detect sentences

☐ Normalize case to lowercase

☐ Remove numbers

☐ Remove special characters

☐ Remove duplicate characters

☐ Remove email addresses

☐ Remove URLs

☐ Expand verb contractions

☐ Normalize backslashes to slashes

☐ Split tokens on special characters

- A. Mastered
- B. Not Mastered

**Answer: A**

**Explanation:**

Box 1: Remove stop words

Remove words to optimize information retrieval.

Remove stop words: Select this option if you want to apply a predefined stopword list to the text column. Stop word removal is performed before any other processes.

Box 2: Lemmatization

Ensure that multiple related words from a single canonical form. Lemmatization converts multiple related words to a single canonical form Box 3: Remove special characters

Remove special characters: Use this option to replace any non-alphanumeric special characters with the pipe | character.

References:

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

**NEW QUESTION 101**

- (Exam Topic 3)

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

The custom role has the following JSON definition:

```
{
  "Name": "MyRole",
  "IsCustom": true,
  "Description": "New custom role description.",
  "Actions": ["*"],
  "NotActions": [
    "Microsoft.MachineLearningServices/workspaces/write",
    "Microsoft.MachineLearningServices/workspaces/computes/*/write",
    "Microsoft.MachineLearningServices/workspaces/computes/*/delete",
    "Microsoft.Authorization/*/write"
  ],
  "AssignableScopes": [
    "/subscriptions/<subscription_id>/resourceGroups/resourcegroup1/providers/
    Microsoft.MachineLearningServices/workspaces/workspace1"
  ]
}
```

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

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

A.

**Answer:**

**Explanation:**

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

Box 1: No

The actions listed in NotActions are prohibited.

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

Box 2: No

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

Writing metrics is not listed in NotActions. Reference:

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

#### NEW QUESTION 105

- (Exam Topic 3)

You have a Jupyter Notebook that contains Python code that is used to train a model.

You must create a Python script for the production deployment. The solution must minimize code maintenance.

Which two actions should you perform? Each correct answer presents part of the solution.

NOTE: Each correct selection is worth one point.

- A. Refactor the Jupyter Notebook code into functions
- B. Save each function to a separate Python file
- C. Define a main() function in the Python script
- D. Remove all comments and functions from the Python script

**Answer:** AC

**Explanation:**

Reference:

<https://www.guru99.com/learn-python-main-function-with-examples-understand-main.html> <https://towardsdatascience.com/from-jupyter-notebook-to-deployment-a-straightforward-example-1838c203a43>

#### NEW QUESTION 108

- (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 109**

- (Exam Topic 3)

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

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

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

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

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

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

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

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

Does the solution meet the goal?

- A. Yes
- B. No

**Answer:** B

**Explanation:**

Instead use the `run_log` function to log the contents in `label_vals`: `for label_val in label_vals:`

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

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

**NEW QUESTION 114**

- (Exam Topic 3)

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

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

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

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

All files store data in the following format:

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

You run the following code:

```
data_store = Datastore.register_azure_blob_container(workspace=ws,
    datastore_name='data_store',
    container_name='quarterly_data',
    account_name='companydata',
    account_key='NRPxk8duxbm3...',
    create_if_not_exists=False)
```

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

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

Solution: Run the following code:

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

Does the solution meet the goal?

- A. Yes

B. No

**Answer:** A

**Explanation:**

Use two file paths.

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

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

Reference:

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

**NEW QUESTION 115**

- (Exam Topic 3)

You create a multi-class image classification deep learning model. You train the model by using PyTorch version 1.2.

You need to ensure that the correct version of PyTorch can be identified for the inferencing environment when the model is deployed.

What should you do?

- A. Save the model locally as a.pt file, and deploy the model as a local web service.
- B. Deploy the model on computer that is configured to use the default Azure Machine Learning conda environment.
- C. Register the model with a .pt file extension and the default version property.
- D. Register the model, specifying the model\_framework and model\_framework\_version properties.

**Answer:** D

**Explanation:**

framework\_version: The PyTorch version to be used for executing training code. Reference:

<https://docs.microsoft.com/en-us/python/api/azureml-train-core/azureml.train.dnn.pytorch?view=azure-ml-py>

**NEW QUESTION 116**

- (Exam Topic 3)

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

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

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

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

NOTE: Each correct selection is worth one point.

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

**Answer:** AE

**Explanation:**

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

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

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

**NEW QUESTION 120**

- (Exam Topic 3)

An organization creates and deploys a multi-class image classification deep learning model that uses a set of labeled photographs.

The software engineering team reports there is a heavy inferencing load for the prediction web services during the summer. The production web service for the model fails to meet demand despite having a fully-utilized compute cluster where the web service is deployed.

You need to improve performance of the image classification web service with minimal downtime and minimal administrative effort.

What should you advise the IT Operations team to do?

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

**Answer:** D

**Explanation:**

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

Reference:

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

**NEW QUESTION 124**

- (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 129

- (Exam Topic 3)

You register a model that you plan to use in a batch inference pipeline.

The batch inference pipeline must use a `ParallelRunStep` step to process files in a file dataset. The script has the `ParallelRunStep` step runs must process six input files each time the inferencing function is called.

You need to configure the pipeline.

Which configuration setting should you specify in the `ParallelRunConfig` object for the `ParallelRunStep` step?

- A. `process_count_per_node= "6"`
- B. `node_count= "6"`
- C. `mini_batch_size= "6"`
- D. `error_threshold= "6"`

**Answer: B**

### Explanation:

`node_count` is the number of nodes in the compute target used for running the `ParallelRunStep`. Reference:

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

### NEW QUESTION 131

- (Exam Topic 2)

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

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

#### Modules

Score Matchbox Recommender

Apply Transformation

Evaluate Recommender

Evaluate Model

Train Model

Sweep Clustering

Score Model

Load Trained Model

#### Answer Area



- A. Mastered
- B. Not Mastered

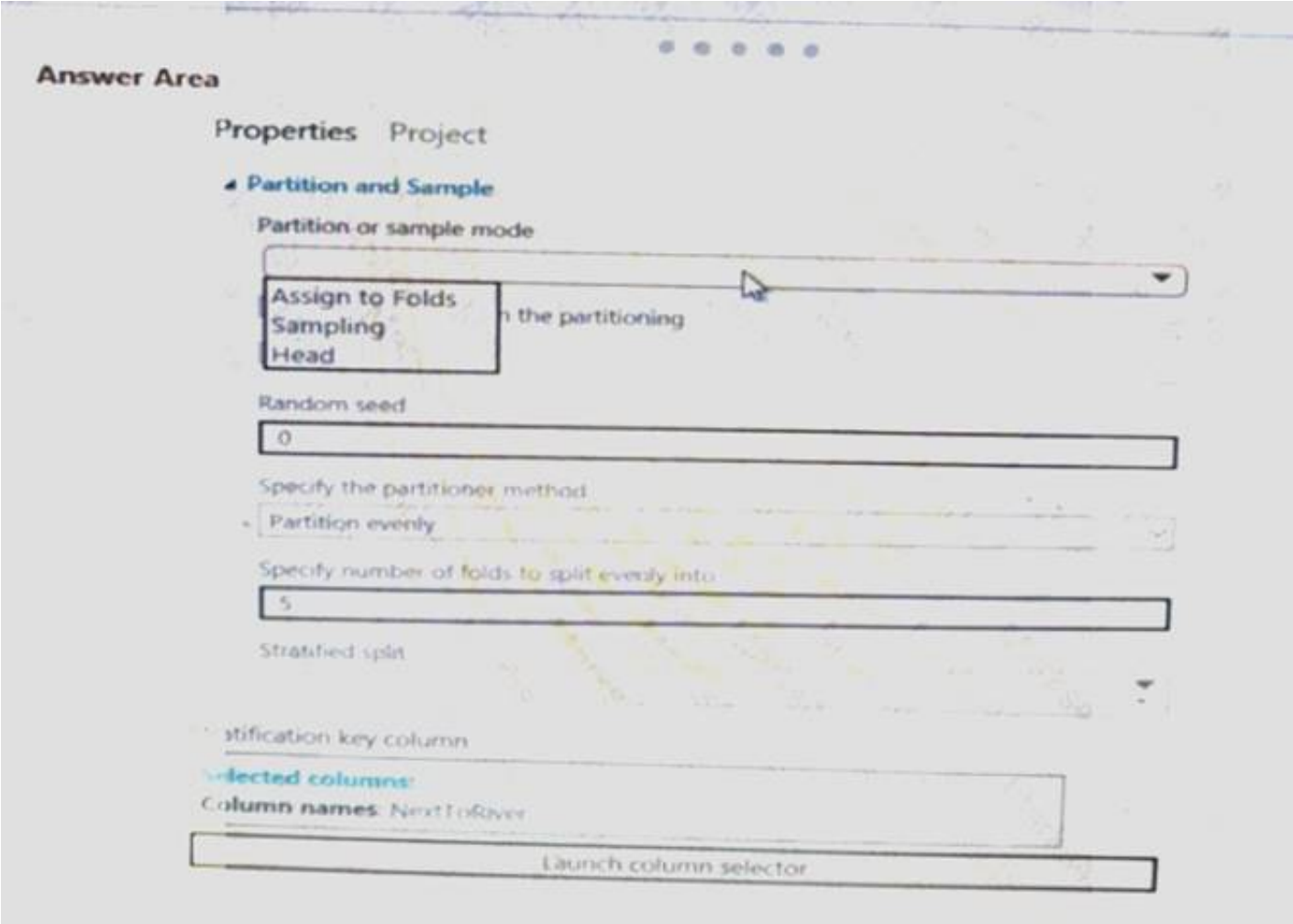
**Answer: A**

### Explanation:

Step 1: Sweep Clustering  
 Start by using the "Tune Model Hyperparameters" module to select the best sets of parameters for each of the models we're considering.  
 One of the interesting things about the "Tune Model Hyperparameters" module is that it not only outputs the results from the Tuning, it also outputs the Trained Model.  
 Step 2: Train Model Step 3: Evaluate Model  
 Scenario: You need to provide the test results to the Fabrikam Residences team. You create data visualizations to aid in presenting the results.  
 You must produce a Receiver Operating Characteristic (ROC) curve to conduct a diagnostic test evaluation of the model. You need to select appropriate methods for producing the ROC curve in Azure Machine Learning Studio to compare the Two-Class Decision Forest and the Two-Class Decision Jungle modules with one another.  
 References:  
<http://breaking-bi.blogspot.com/2017/01/azure-machine-learning-model-evaluation.html>

**NEW QUESTION 132**

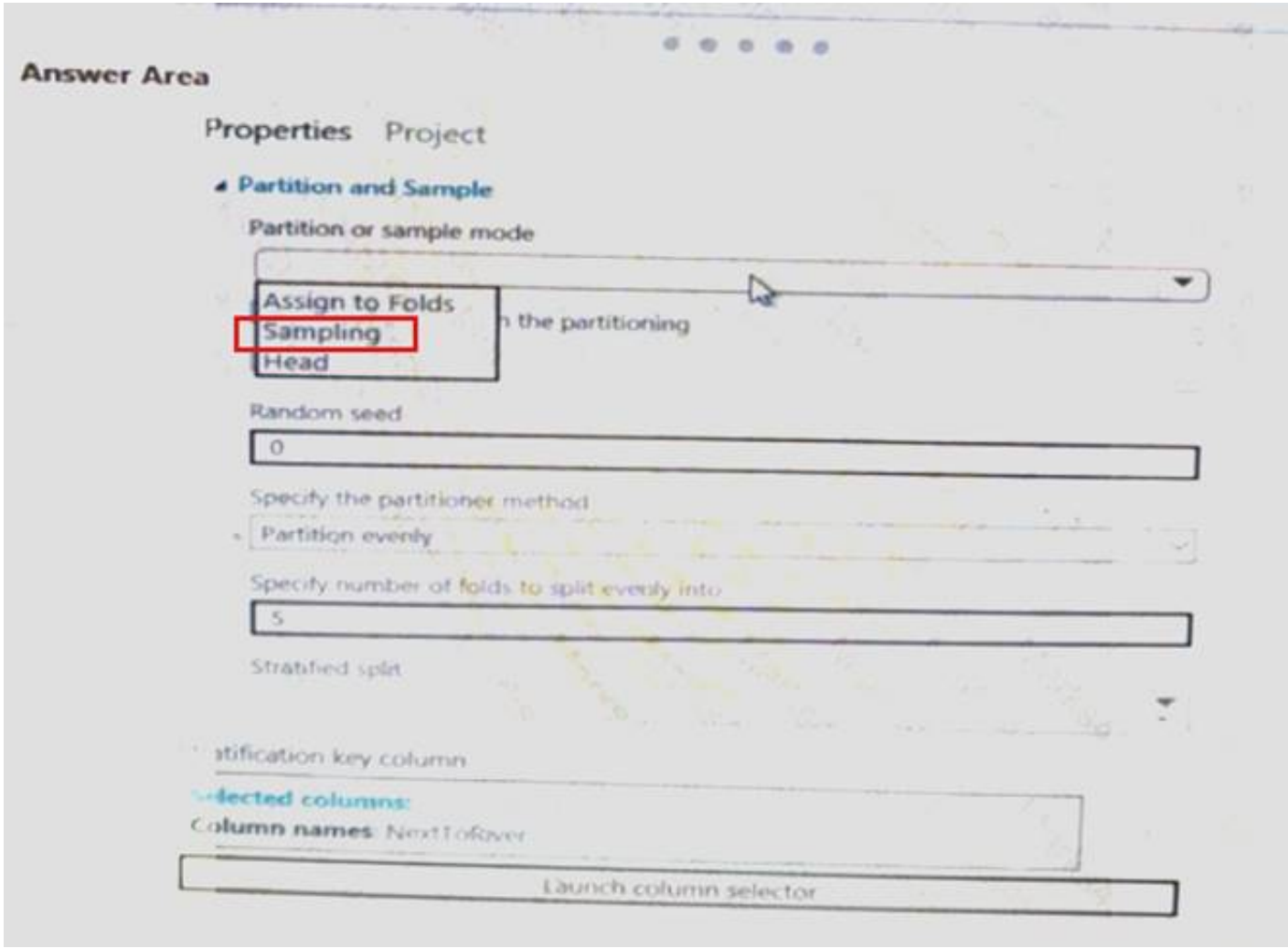
- (Exam Topic 2)  
 You need to identify the methods for dividing the data according, to the testing requirements.  
 Which properties should you select? To answer, select the appropriate option-, m the answer area. NOTE: Each correct selection is worth one point.



- A. Mastered
- B. Not Mastered

**Answer:** A

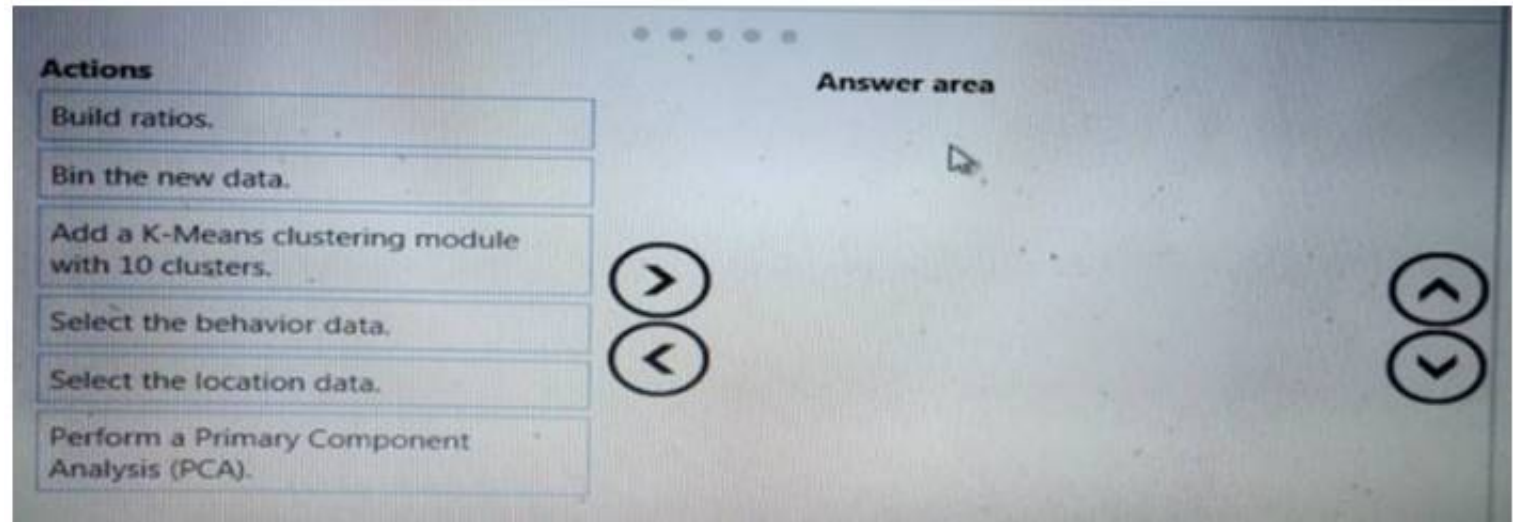
**Explanation:**



NEW QUESTION 135

- (Exam Topic 1)

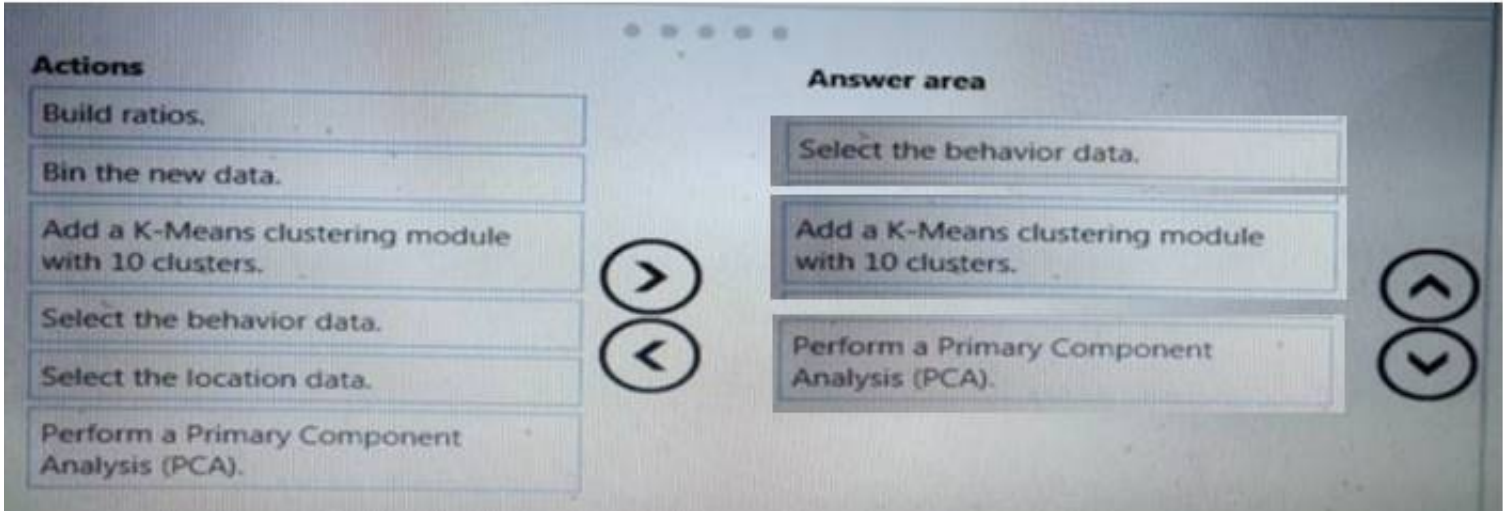
You need to modify the inputs for the global penalty event model to address the bias and variance issue. Which three actions should you perform in sequence? To answer, move the appropriate actions from the list of actions to the answer area and arrange them in the correct order.



- A. Mastered
- B. Not Mastered

Answer: A

Explanation:



NEW QUESTION 140

- (Exam Topic 1)

You need to define a modeling strategy for ad response.

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
Implement a K-Means Clustering model.	
Use the raw score as a feature in a Score Matchbox Recommender model.	
Use the cluster as a feature in a Decision Jungle model.	<div><div></div><div></div></div>
Use the raw score as a feature in a Logistic Regression model.	<div><div></div><div></div></div>
Implement a Sweep Clustering model.	

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

Step 1: Implement a K-Means Clustering model  
Step 2: Use the cluster as a feature in a Decision jungle model.  
Decision jungles are non-parametric models, which can represent non-linear decision boundaries. Step 3: Use the raw score as a feature in a Score Matchbox Recommender model  
The goal of creating a recommendation system is to recommend one or more "items" to "users" of the system. Examples of an item could be a movie, restaurant, book, or song. A user could be a person, group of persons, or other entity with item preferences.  
Scenario:  
Ad response rated declined.  
Ad response models must be trained at the beginning of each event and applied during the sporting event. Market segmentation models must optimize for similar ad response history.  
Ad response models must support non-linear boundaries of features. References:  
<https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/multiclass-decision-jungle> <https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/score-matchbox-recommende>

NEW QUESTION 145

.....

## 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](#)**