DP-100 Q&As

Designing and Implementing a Data Science Solution on Azure

Pass Microsoft DP-100 Exam with 100% Guarantee

Free Download Real Questions & Answers **PDF** and **VCE** file from:

https://www.pass4itsure.com/dp-100.html

100% Passing Guarantee
100% Money Back Assurance

Following Questions and Answers are all new published by Microsoft Official Exam Center

- **Instant Download** After Purchase
- **100% Money Back** Guarantee
- **365 Days** Free Update
- **800,000+** Satisfied Customers

QUESTION 1

You are developing a two-step Azure Machine Learning pipeline by using the Azure Machine Learning SDK for Python. You need to register the output of the pipeline as a new version of a named dataset after the run has been completed. What should you implement?

A. the as_input method of the OutputDatasetConfig class

B. the register_on_complete method of the OutputDatasetConfig class

C. the as_mount method of the DatasetConsumptionConfig class

D. the as_download method of the DatasetConsumptionConfig class

Correct Answer: A

The OutputDatasetConfig Class as_input method registers the output as a new version of a named Dataset after the run has ran.

If there are no datasets registered under the specified name, a new Dataset with the specified name will be registered. If there is a dataset registered under the specified name, then a new version will be added to this dataset.

Incorrect:

*

The OutputDatasetConfig Class as_input method specifies how to consume the output as an input in subsequent pipeline steps.

*

as_mount sets the mode to mount.

In the submitted run, files in the datasets will be mounted to local path on the compute target. The mount point can be retrieved from argument values and the input_datasets field of the run context.

*

as_download Set the mode to download.

In the submitted run, files in the dataset will be downloaded to local path on the compute target. The download location can be retrieved from argument values and the input_datasets field of the run context.

Reference: https://learn.microsoft.com/en-us/python/api/azureml-core/azureml.data.output_dataset_config.outputdatasetconfig https://learn.microsoft.com/en-us/python/api/azureml-core/azureml.data.dataset_consumption_config.datasetconsumptionconfig

QUESTION 2

DRAG DROP

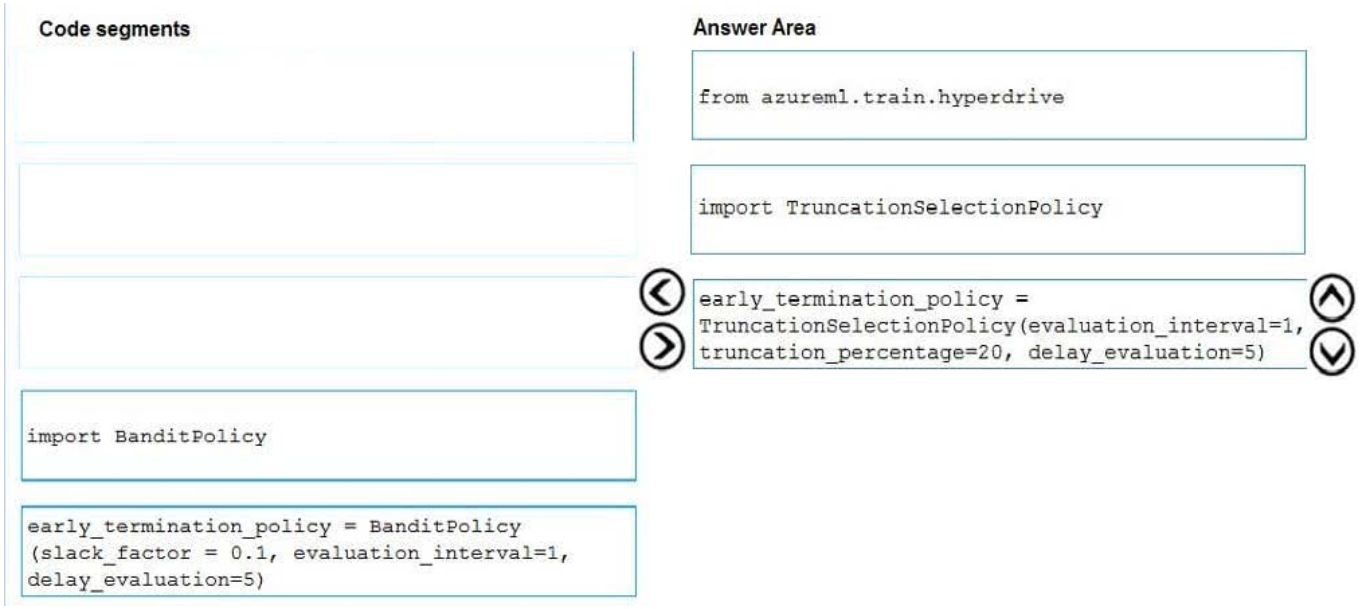
You need to implement an early stopping criteria policy for model training.

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.

Select and Place:

Correct Answer:



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)

Incorrect Answers:

Bandit is a termination policy based on slack factor/slack amount and evaluation interval. The policy early terminates any runs where the primary metric is not within the specified slack factor / slack amount with respect to the best performing

training run.

Example:

from azureml.train.hyperdrive import BanditPolicy

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

References:

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

QUESTION 3

You have a dataset created for multiclass classification tasks that contains a normalized numerical feature set with 10,000 data points and 150 features.

You use 75 percent of the data points for training and 25 percent for testing. You are using the scikit-learn machine learning library in Python. You use X to denote the feature set and Y to denote class labels.

You create the following Python data frames:

Name	Description
X_train	training feature set
Y_train	training class labels
x_train	testing feature set
y_train	testing class labels

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

NOTE: Each correct selection is worth one point.

Hot Area:

Answer Area

```
from sklearn.decomposition import PCA
pca = [ ▼ ]
      PCA()
      PCA(n_components = 150)
      PCA(n_components = 10)
      PCA(n_components = 10000)
X_train= [ ▼ ].fit_transform(X_train)
      pca
      model
      sklearn.decomposition
x_test = pca.[ ▼ ]
      x_test
      X_train
      fit(x_test)
      transform(x_test)
```

Correct Answer:

Answer Area

```
from sklearn.decomposition import PCA
pca = [ ▼ ]
      PCA()
      PCA(n_components = 150)
      PCA(n_components = 10)
      PCA(n_components = 10000)
X_train= [ ▼ ].fit_transform(X_train)
      pca
      model
      sklearn.decomposition
x_test = pca.[ ▼ ]
      x_test
      X_train
      fit(x_test)
      transform(x_test)
```

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

QUESTION 4

You are implementing hyperparameter tuning by using Bayesian sampling for a model training from a notebook. The notebook is in an Azure Machine Learning workspace that uses a compute cluster with 20 nodes.

The code implements Bandit termination policy with slack factor set to 0.2 and the HyperDriveConfig class instance with max_concurrent_runs set to 10.

You must increase effectiveness of the tuning process by improving sampling convergence.

You need to select which sampling convergence to use.

What should you select?

A. Set the value of slack factor of early_termination_policy to 0.9.

B. Set the value of max_concurrent_runs of HyperDriveConfig to 4.

C. Set the value of slack factor of early_termination_policy to 0.1.

D. Set the value of max_concurrent_runs of HyperDriveConfig to 20.

Correct Answer: B

Bayesian sampling

Bayesian sampling is based on the Bayesian optimization algorithm. It picks samples based on how previous samples did, so that new samples improve the primary metric.

The number of concurrent jobs has an impact on the effectiveness of the tuning process. A smaller number of concurrent jobs may lead to better sampling convergence, since the smaller degree of parallelism increases the number

of jobs

that benefit from previously completed jobs.

Reference: https://learn.microsoft.com/en-us/azure/machine-learning/how-to-tune-hyperparameters

QUESTION 5

HOTSPOT

You plan to implement a two-step pipeline by using the Azure Machine Learning SDK for Python.

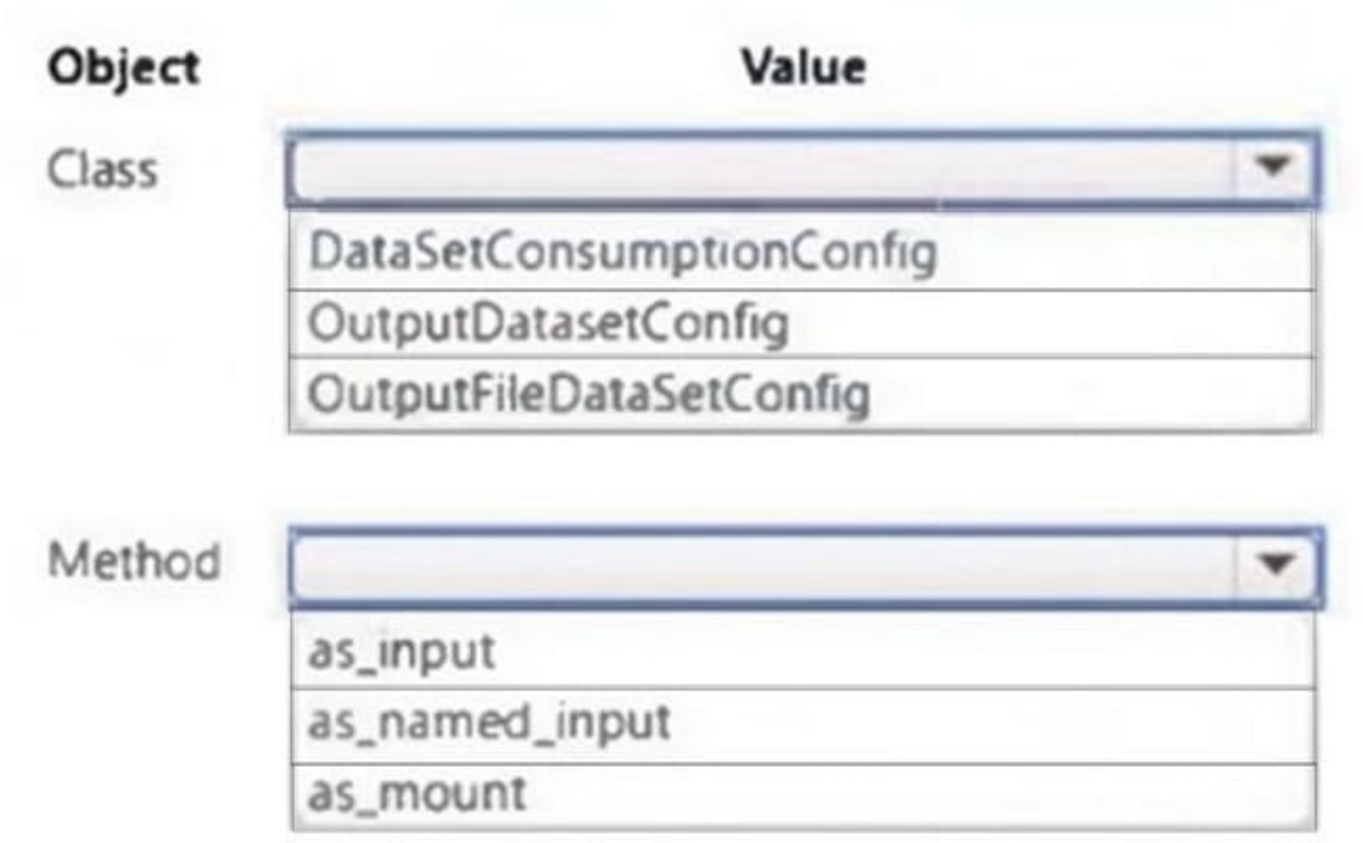
The pipeline will pass temporary data from the first step to the second step.

You need to identify the class and the corresponding method that should be used in the second step to access temporary data generated by the first step in the pipeline.

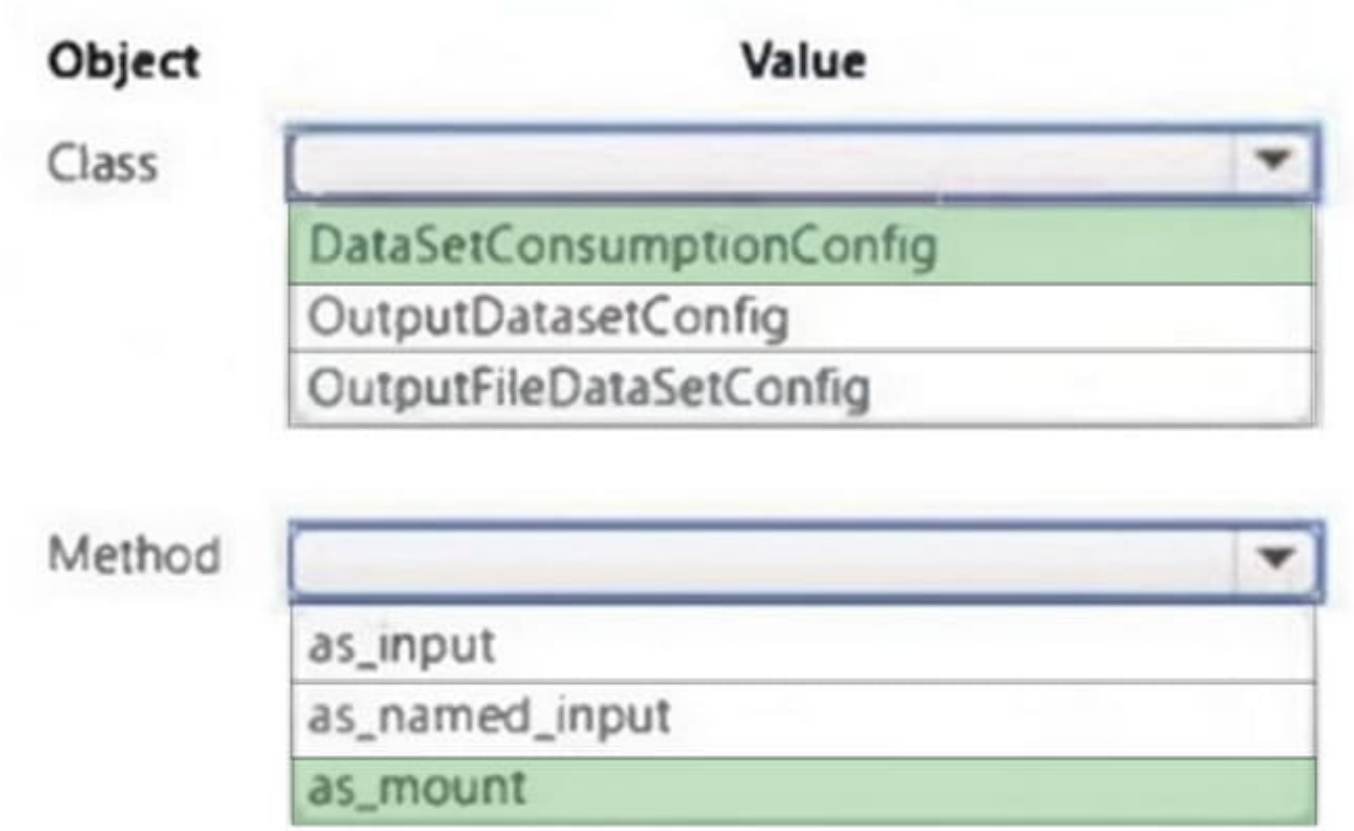
Which class and method should you identify? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point

Hot Area:



Correct Answer:



DP-100 PDF Dumps DP-100 Study Guide DP-100 Braindumps