Up-to-date Questions and Answers from authentic resources to improve knowledge and pass the exam at very first attempt. ----- Guaranteed.

killexams.com

Microsoft

DP-100

Designing and Implementing a Data Science Solution on Azure

Question: 98

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 contain 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: Use the last Observation Carried Forward (IOCF) method to impute the missing data points.

Does the solution meet the goal?
A. Yes
B. No

Answer: B

Explanation:

Instead use the Multiple Imputation by Chained Equations (MICE) method.

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: Last observation carried forward (LOCF) is a method of imputing missing data in longitudinal studies. If a person drops out of a study before it ends, then his or her last observed score on the dependent variable is used for all subsequent (i.e., missing) observation points. LOCF is used to maintain the sample size and to reduce the bias caused by the attrition of participants in a study.

References:

https://methods.sagepub.com/reference/encyc-of-research-design/n211.xml https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3074241/

Question: 99

You deploy a real-time inference service for a trained model.

The deployed model supports a business-critical application, and it is important to be able to monitor the data submitted to the web service and the predictions the data generates.

You need to implement a monitoring solution for the deployed model using minimal administrative effort.

What should you do?
A. View the explanations for the registered model in Azure ML studio.
B. Enable Azure Application Insights for the service endpoint and view logged data in the Azure portal.
C. Create an ML Flow tracking URI that references the endpoint, and view the data logged by ML Flow.
D. View the log files generated by the experiment used to train the model.

Answer: B

Explanation:

Configure logging with Azure Machine Learning studio

You can also enable Azure Application Insights from Azure Machine Learning studio. When you're ready to deploy your model as a web service, use the following steps to enable Application Insights:

Question: 100

You are solving a classification task.

You must evaluate your model on a limited data sample by using k-fold cross validation. You start by

configuring a k parameter as the number of splits.

You need to configure the k parameter for the cross-validation.

Which value should you use?
A. k=0.5
B. k=0
C. k=5
D. k=1

Answer: C

Explanation:

Leave One Out (LOO) cross-validation

Setting K = n (the number of observations) yields n-fold and is called leave-one out cross-validation (LOO), a special case of the K-fold approach.

LOO CV is sometimes useful but typically doesn't shake up the data enough. The estimates from each fold are highly correlated and hence their average can have high variance.

This is why the usual choice is K=5 or 10. It provides a good compromise for the bias-variance tradeoff.

Question: 101

DRAG DROP

You create an Azure Machine Learning workspace.

You must implement dedicated compute for model training in the workspace by using Azure Synapse compute resources. The solution must attach the dedicated compute and start an Azure Synapse session.

You need to implement the compute resources.

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

Actions		Answer area
Create compute clusters by using Azure Machine Learning studio.	1	
Create a linked service by using Azure Synapse studio.	2	
Create a linked service by using Azure Machine Learning studio.	3	
Create an Azure Synapse workspace by using the Azure portal.		
Create an Apache Spark pool by using the Azure portal.		

Answer:

Actions		Answer area
Create compute clusters by using Azure Machine Learning studio.	1	Create a linked service by using Azure Machine Learning studio.
Create a linked service by using Azure Synapse studio.	2	Create an Azure Synapse workspace by using the Azure portal.
Create a linked service by using Azure Machine Learning studio.	3	Create an Apache Spark pool by using the Azure portal.
Create an Azure Synapse workspace by using the Azure portal.		
Create an Apache Spark pool by using the Azure portal.		

Explanation:

Question: 102

You deploy a real-time inference service for a trained model.

The deployed model supports a business-critical application, and it is important to be able to monitor the data submitted to the web service and the predictions the data generates.

You need to implement a monitoring solution for the deployed model using minimal administrative effort.

What should you do?
A. View the explanations for the registered model in Azure ML studio.
B. Enable Azure Application Insights for the service endpoint and view logged data in the Azure portal.

C. Create an ML Flow tracking URI that references the endpoint, and view the data logged by ML Flow.
D. View the log files generated by the experiment used to train the model.

Answer: B

Explanation:

Configure logging with Azure Machine Learning studio

You can also enable Azure Application Insights from Azure Machine Learning studio. When you're ready to deploy your model as a web service, use the following steps to enable Application Insights:

Question: 103

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

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

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

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

Answer: C

Explanation:

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

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

The basic deployment tasks are:

Question: 104

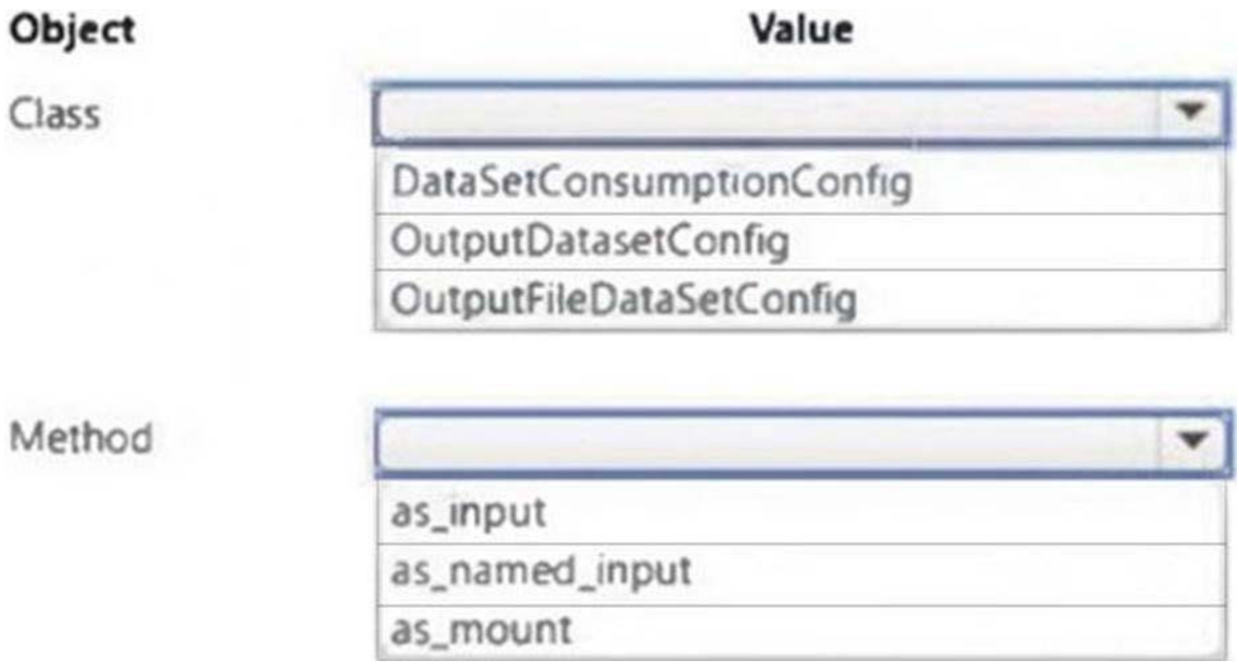
HOTSPOT

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

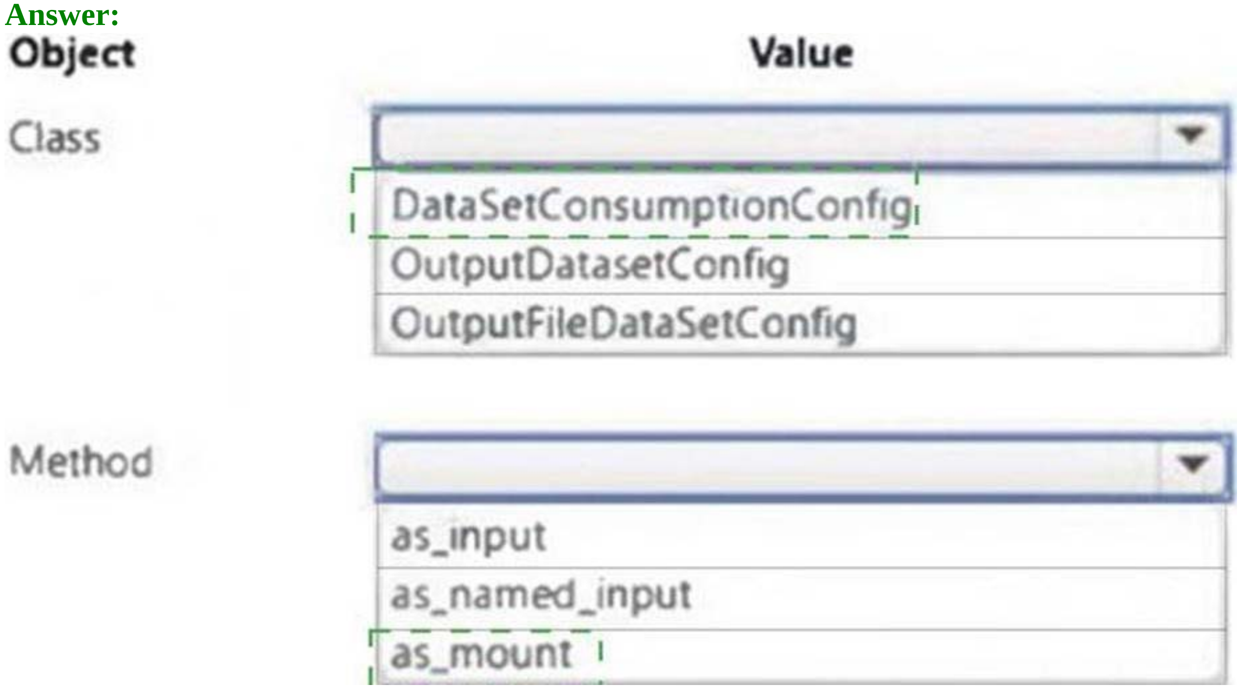
The pipeline will pass temporary data from the first step to the second step.

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



Answer:



Question: 105

HOTSPOT

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)
```

Answer Area

	Yes	No
The compute is created in the same region as the Machine Learning service workspace.	○	○
The compute resource created by the code is displayed as a compute cluster in Azure Machine Learning studio.	○	○
The minimum number of nodes will be zero.	○	○

Answer:

Answer Area

	Yes	No
The compute is created in the same region as the Machine Learning service workspace.	○	○
The compute resource created by the code is displayed as a compute cluster in Azure Machine Learning studio.	○	○
The minimum number of nodes will be zero.	○	○

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

Question: 106

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 TabularExplainer.

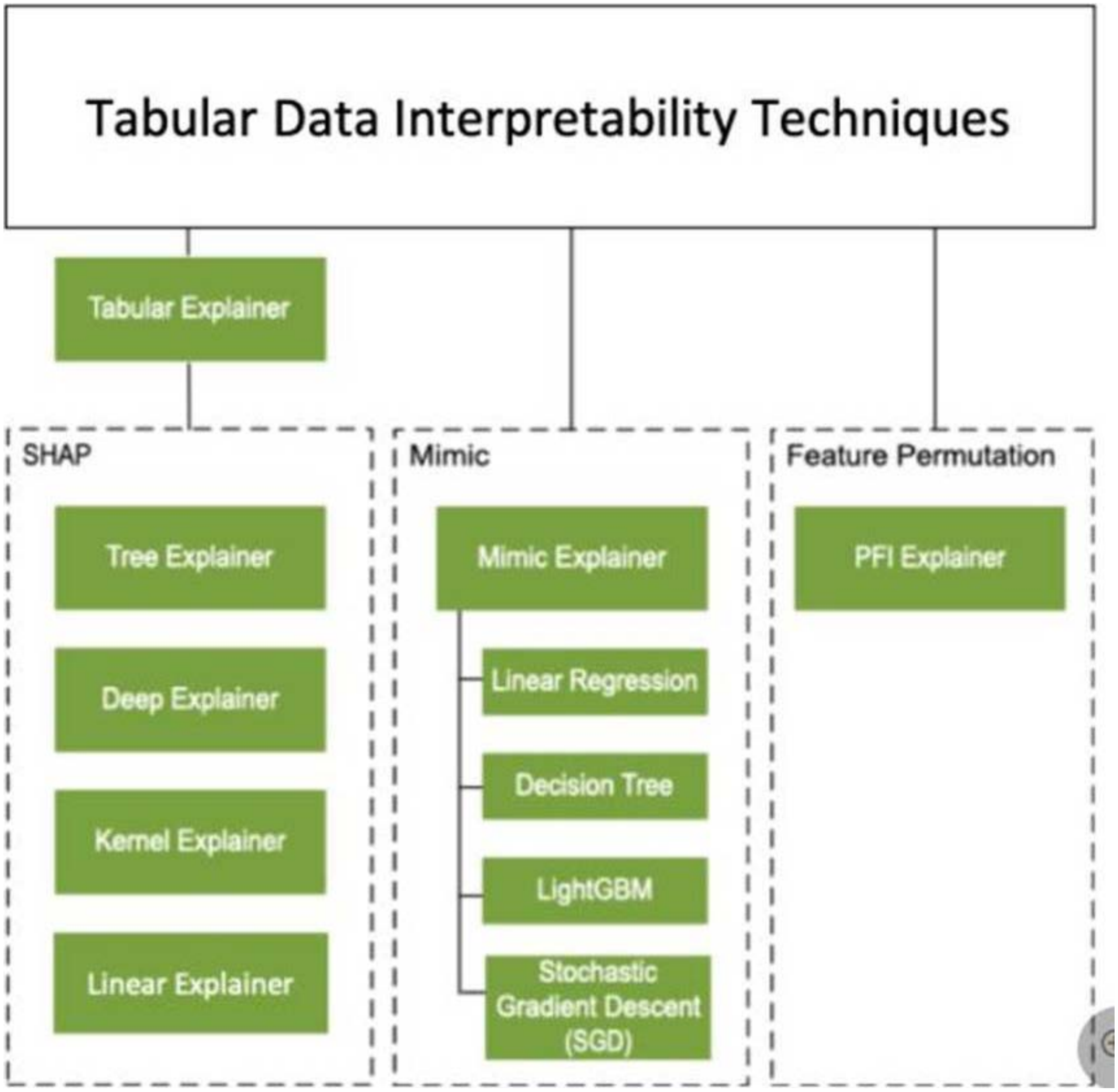
Does the solution meet the goal?
A. Yes
B. No

Answer: B

Explanation:

Instead use Permutation Feature Importance Explainer (PFI).

Note 1:



Note 2: Permutation Feature Importance Explainer (PFI): Permutation Feature Importance is a technique used to explain classification and regression models. At a high level, the way it works is by randomly shuffling data one feature at a time for the entire dataset and calculating how much the performance metric of interest changes. The larger the change, the more important that feature is. PFI can explain the overall behavior of any underlying model but does not explain individual predictions.

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

Question: 107

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 underepresented 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

Question: 108

You use the following code to define the steps for a pipeline:

from azureml.core import Workspace, Experiment, Run

from azureml.pipeline.core import Pipeline

from azureml.pipeline.steps import PythonScriptStep

ws = Workspace.from_config()

. . .

step1 = PythonScriptStep(name="step1", …)

step2 = PythonScriptsStep(name="step2", …)

pipeline_steps = [step1, step2]

You need to add code to run the steps.

Which two code segments can you use to achieve this goal? Each correct answer presents a complete solution. NOTE: Each correct selection is worth one point.
A. experiment = Experiment(workspace=ws,
name='pipeline-experiment')
run = experiment.submit(config=pipeline_steps)
B. run = Run(pipeline_steps)
C. pipeline = Pipeline(workspace=ws, steps=pipeline_steps) experiment = Experiment(workspace=ws, name='pipeline-experiment') run = experiment.submit(pipeline)
D. pipeline = Pipeline(workspace=ws, steps=pipeline_steps)
run = pipeline.submit(experiment_name='pipeline-experiment')

Answer: C,D

Explanation:

After you define your steps, you build the pipeline by using some or all of those steps.

Build the pipeline. Example:

pipeline1 = Pipeline(workspace=ws, steps=[compare_models])

Submit the pipeline to be run

pipeline_run1 = Experiment(ws, ‘Compare_Models_Exp’).submit(pipeline1)

Reference: https://docs.microsoft.com/en-us/azure/machine-learning/how-to-create-machine-learning-pipelines

Question: 109

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

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

You create an Azure Machine Learning service datastore in a workspace.

The datastore contains the following files:

• /data/2018/Q1.csv

• /data/2018/Q2.csv

• /data/2018/Q3.csv

• /data/2018/Q4.csv

• /data/2019/Q1.csv

All files store data in the following format:

id,f1,f2i

1,1.2,0

2,1,1,

1 3,2.1,0

You run the following code:

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

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

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

Solution: Run the following code:

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

Does the solution meet the goal?
A. Yes
B. No

Answer: B

Explanation:

Use two file paths.

Use Dataset.Tabular_from_delimeted, instead of Dataset.File.from_files as the data isn't cleansed.

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