

Microsoft

Exam Questions DP-100

Designing and Implementing a Data Science Solution on Azure



NEW QUESTION 1

- (Exam Topic 3)

HOTSPOT

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

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

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

NOTE: Each correct selection is worth one point.

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

A. Mastered

B. Not Mastered

Answer: A

Explanation:

Box 1: Yes

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

Box 2: Yes

script_params is a dictionary of command-line arguments to pass to the training script specified in entry_script.

Box 3: No

Box 4: Yes

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

NEW QUESTION 2

- (Exam Topic 3)

You train a machine learning model.

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

Which compute target should you use?

A. Azure Kubernetes Service (AKS) inference cluster

B. Azure Machine Learning compute cluster

C. Azure Container Instance (ACI)

D. attached Azure Databricks cluster

Answer: C

Explanation:

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

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

Reference:

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

NEW QUESTION 3

- (Exam Topic 3)

You have a dataset that includes confidential data. You use the dataset to train a model.

You must use a differential privacy parameter to keep the data of individuals safe and private. You need to reduce the effect of user data on aggregated results. What should you do?

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

Answer: C

Explanation:

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

Reference:

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

NEW QUESTION 4

- (Exam Topic 3)

You are creating a classification model for a banking company to identify possible instances of credit card fraud. You plan to create the model in Azure Machine Learning by using automated machine learning.

The training dataset that you are using is highly unbalanced. You need to evaluate the classification model.

Which primary metric should you use?

- A. normalized_mean_absolute_error
- B. [spearman_correlation
- C. AUC.weighted
- D. accuracy
- E. normalized_root_mean_squared_error

Answer: C

Explanation:

AUC_weighted is a Classification metric.

Note: AUC is the Area under the Receiver Operating Characteristic Curve. Weighted is the arithmetic mean of the score for each class, weighted by the number of true instances in each class.

Reference:

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

NEW QUESTION 5

- (Exam Topic 3)

You create an Azure Machine Learning workspace.

You need to detect data drift between a baseline dataset and a subsequent target dataset by using the

DataDriftDetector class.

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

NOTE: Each correct selection is worth one point.

```
from azureml.core import Workspace, Dataset
from datetime import datetime

ws = Workspace.from_config()
dset = Dataset.get_by_name(ws, 'target')
baseline = target.time_before(datetime(2021, 2, 1))
features = ['windAngle', 'windSpeed', 'temperature', 'stationName']

monitor = DataDriftDetector.          (ws, 'drift-monitor', baseline,

target, compute_target='cpu-cluster', frequency='Week', feature_list=None,
drift_threshold=.6, latency=24)

monitor = DataDriftDetector.get_by_name(ws, 'drift-monitor')
monitor = monitor.update(feature_list=features)
complete = monitor.          (datetime(2021, 1, 1), datetime.today())
```

backfill
create_from_datasets
create_from_model

backfill
list
update

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

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

Box 1: create_from_datasets

The create_from_datasets method creates a new DataDriftDetector object from a baseline tabular dataset and a target time series dataset.

Box 2: backfill

The backfill method runs a backfill job over a given specified start and end date.

Syntax: backfill(start_date, end_date, compute_target=None, create_compute_target=False) Reference:

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

NEW QUESTION 6

- (Exam Topic 3)

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

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

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

What should you do?

A. Register a new version of the model and update the entry script to load the new version of the model from its registered path.

B. Modify the AKS service deployment configuration to enable application insights and re-deploy to AKS.

C. Create an Azure Container Instances (ACI) web service deployment configuration and deploy the model on ACI.

D. Add a breakpoint to the first line of the entry script and redeploy the service to AKS.

E. Create a local web service deployment configuration and deploy the model to a local Docker container.

Answer: C

Explanation:

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

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

* 1. Register the model in the workspace model registry.

* 2. Define Inference Configuration:

* a. Create an Environment object based on the dependencies you specify in the environment yaml file or use one of our procured environments.

* b. Create an inference configuration (InferenceConfig object) based on the environment and the scoring script.

* 3. Deploy the model to Azure Container Instance (ACI) service or to Azure Kubernetes Service (AKS).

NEW QUESTION 7

- (Exam Topic 3)

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

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

➤ a Microsoft Surface device on which Microsoft Office has been installed. Corporate IT policies prevent the installation of additional software

➤ a Compute Instance named ds-workstation in the workspace with 2 CPUs and 8 GB of memory

➤ an Azure Machine Learning compute target named cpu-cluster with eight CPU-based nodes

➤ an Azure Machine Learning compute target named gpu-cluster with four CPU and GPU-based nodes

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

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

NOTE: Each correct selection is worth one point.

Resource type

Option

Run code to submit the experiment

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

Run the training script

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

A. Mastered

B. Not Mastered

Answer: A

Explanation:

Resource type

Option

Run code to submit the experiment

▼

the Microsoft Surface device

the ds-workstation notebook VM

the cpu-cluster compute target

the gpu-cluster compute target

Run the training script

▼

the ds-workstation notebook VM

the cpu-compute target

the gpu-compute target

the Microsoft Surface device

NEW QUESTION 8

- (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.dnn import TensorFlow
sk_est = TensorFlow(source_directory='./scripts',
    compute_target=aml-compute,
    entry_script='train.py')
```

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 9

- (Exam Topic 3)

You are developing deep learning models to analyze semi-structured, unstructured, and structured data types. You have the following data available for model building:

- Video recordings of sporting events
 - Transcripts of radio commentary about events
 - Logs from related social media feeds captured during sporting events
- You need to select an environment for creating the model.

Which environment should you use?

- A. Azure Cognitive Services
- B. Azure Data Lake Analytics
- C. Azure HDInsight with Spark MLlib
- D. Azure Machine Learning Studio

Answer: A

Explanation:

Azure Cognitive Services expand on Microsoft's evolving portfolio of machine learning APIs and enable developers to easily add cognitive features – such as emotion and video detection; facial, speech, and vision recognition; and speech and language understanding – into their applications. The goal of Azure Cognitive

Services is to help developers create applications that can see, hear, speak, understand, and even begin to reason. The catalog of services within Azure Cognitive Services can be categorized into five main pillars - Vision, Speech, Language, Search, and Knowledge.

References:

<https://docs.microsoft.com/en-us/azure/cognitive-services/welcome>

NEW QUESTION 10

- (Exam Topic 3)

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

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

You 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_list('Label Values', label_vals)` Does the solution meet the goal?

- A. Yes
- B. No

Answer: A

Explanation:

`run.log_list` log a list of values to the run with the given name using `log_list`. Example: `run.log_list("accuracies", [0.6, 0.7, 0.87])`

Note:

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

Data is read into a `pandas.DataFrame`, which is a two-dimensional, size-mutable, potentially heterogeneous tabular data.

`label_vals =data['label'].unique`

`label_vals` contains a list of unique label values. Reference:

<https://www.element61.be/en/resource/azure-machine-learning-services-complete-toolbox-ai> [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)) <https://pandas.pydata.org/docs/reference/api/pandas.DataFrame.html>

NEW QUESTION 10

- (Exam Topic 3)

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

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

NOTE: Each correct selection is worth one point.

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

Answer: DE

Explanation:

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

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

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

References:

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

NEW QUESTION 14

- (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 * ⓘ

Resource group * ⓘ [Create new](#)

Workspace details


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

Workspace name * ⓘ

Region * ⓘ

Workspace edition * ⓘ

- Basic
- Basic
- Enterprise

 For your convenience, these resources are available in the 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 18

- (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 21

- (Exam Topic 3)

You plan to use a Deep Learning Virtual Machine (DLVM) to train deep learning models using Compute Unified Device Architecture (CUDA) computations. You need to configure the DLVM to support CUDA. What should you implement?

- A. Intel Software Guard Extensions (Intel SGX) technology
- B. Solid State Drives (SSD)
- C. Graphic Processing Unit (GPU)
- D. Computer Processing Unit (CPU) speed increase by using overclocking
- E. High Random Access Memory (RAM) configuration

Answer: C

Explanation:

A Deep Learning Virtual Machine is a pre-configured environment for deep learning using GPU instances.

References:

<https://azuremarketplace.microsoft.com/en-au/marketplace/apps/microsoft-ads.dsvm-deep-learning>

NEW QUESTION 25

- (Exam Topic 3)

You are solving a classification task.

You must evaluate your model on a limited data sample by using k-fold cross validation. You start by configuring a k parameter as the number of splits.

You need to configure the k parameter for the cross-validation. Which value should you use?

- A. k=0.5
- B. k=0
- C. k=5
- D. k=1

Answer: C

Explanation:

Leave One Out (LOO) cross-validation

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

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

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

NEW QUESTION 26

- (Exam Topic 3)

You create an Azure Machine Learning workspace.

You must create a custom role named DataScientist that meets the following requirements:

- Role members must not be able to delete the workspace.
- Role members must not be able to create, update, or delete compute resource in the workspace.
- Role members must not be able to add new users to the workspace.

You need to create a JSON file for the DataScientist role in the Azure Machine Learning workspace. The custom role must enforce the restrictions specified by the IT Operations team.

Which JSON code segment should you use?

A)


```
{
  "Name": "DataScientist",
  "IsCustom": true,
  "Description": "Project Data Scientist role",
  "Actions": ["*"],
  "NotActions": [
    "Microsoft.MachineLearningServices/workspaces/*/delete",
    "Microsoft.MachineLearningServices/workspaces/computes/*/write",
    "Microsoft.MachineLearningServices/workspaces/computes/*/delete",
    "Microsoft.Authorization/*/write"
  ],
  "AssignableScopes": [
    "/subscriptions/<id>/resourceGroups/ml-rg/providers/Microsoft.MachineLearningServices/workspaces/ml-ws"
  ]
}
```

B)

```
{
  "Name": "DataScientist",
  "IsCustom": true,
  "Description": "Project Data Scientist role",
  "Actions": ["*"],
  "NotActions": [],
  "AssignableScopes": [
    "/subscriptions/<id>/resourceGroups/ml-rg/providers/Microsoft.MachineLearningServices/workspaces/ml-ws"
  ]
}
```

C)

```
{
  "Name": "DataScientist",
  "IsCustom": true,
  "Description": "Project Data Scientist role",
  "Actions": [
    "Microsoft.MachineLearningServices/workspaces/*/delete",
    "Microsoft.MachineLearningServices/workspaces/computes/*/write",
    "Microsoft.MachineLearningServices/workspaces/computes/*/delete",
    "Microsoft.Authorization/*/write"
  ],
  "NotActions": [],
  "AssignableScopes": [
    "/subscriptions/<id>/resourceGroups/ml-rg/providers/Microsoft.MachineLearningServices/workspaces/ml-ws"
  ]
}
```

D)

```
{
  "Name": "DataScientist",
  "IsCustom": true,
  "Description": "Project Data Scientist role",
  "Actions": [],
  "NotActions": ["*"],
  "AssignableScopes": [
    "/subscriptions/<id>/resourceGroups/ml-rg/providers/Microsoft.MachineLearningServices/workspaces/ml-ws"
  ]
}
```

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

Answer: A**Explanation:**

The following custom role can do everything in the workspace except for the following actions:

- It can't create or update a compute resource.
- It can't delete a compute resource.
- It can't add, delete, or alter role assignments.
- It can't delete the workspace.

To create a custom role, first construct a role definition JSON file that specifies the permission and scope for the role. The following example defines a custom role named "Data Scientist Custom" scoped at a specific workspace level:

data_scientist_custom_role.json :

```
{
  "Name": "Data Scientist Custom", "IsCustom": true,
  "Description": "Can run experiment but can't create or delete compute.", "Actions": ["*"],
  "NotActions": [
    "Microsoft.MachineLearningServices/workspaces/*/delete", "Microsoft.MachineLearningServices/workspaces/write",
    "Microsoft.MachineLearningServices/workspaces/computes/*/write", "Microsoft.MachineLearningServices/workspaces/computes/*/delete",
    "Microsoft.Authorization/*/write"
  ]
}
```

```
],  
"AssignableScopes": [ "/subscriptions/<subscription_id>/resourceGroups/<resource_group_name>/providers/Microsoft.MachineLearni  
]  
}  
}
```

Reference:

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

NEW QUESTION 27

- (Exam Topic 3)

You are a lead data scientist for a project that tracks the health and migration of birds. You create a multi-image classification deep learning model that uses a set of labeled bird photos collected by experts. You plan to use the model to develop a cross-platform mobile app that predicts the species of bird captured by app users.

You must test and deploy the trained model as a web service. The deployed model must meet the following requirements:

- An authenticated connection must not be required for testing.
- The deployed model must perform with low latency during inferencing.
- The REST endpoints must be scalable and should have a capacity to handle large number of requests when multiple end users are using the mobile application.

You need to verify that the web service returns predictions in the expected JSON format when a valid REST request is submitted.

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

NOTE: Each correct selection is worth one point.

Context

Resource

Test

	▼
ds-workstation notebook VM	
aks-compute cluster	
cpu-compute cluster	
gpu-compute cluster	

Production

	▼
ds-workstation notebook VM	
aks-compute cluster	
cpu-compute cluster	
gpu-compute cluster	

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

Box 1: ds-workstation notebook VM

An authenticated connection must not be required for testing.

On a Microsoft Azure virtual machine (VM), including a Data Science Virtual Machine (DSVM), you create local user accounts while provisioning the VM. Users then authenticate to the VM by using these credentials.

Box 2: gpu-compute cluster

Image classification is well suited for GPU compute clusters

Reference:

<https://docs.microsoft.com/en-us/azure/machine-learning/data-science-virtual-machine/dsvm-common-identity> <https://docs.microsoft.com/en-us/azure/architecture/reference-architectures/ai/training-deep-learning>

NEW QUESTION 30

- (Exam Topic 3)

You are creating a machine learning model. You need to identify outliers in the data.

Which two visualizations can you use? Each correct answer presents a complete solution.

NOTE: Each correct selection is worth one point. NOTE: Each correct selection is worth one point.

- A. box plot
- B. scatter
- C. random forest diagram
- D. Venn diagram
- E. ROC curve

Answer: AB

Explanation:

The box-plot algorithm can be used to display outliers.

One other way to quickly identify Outliers visually is to create scatter plots. References:

<https://blogs.msdn.microsoft.com/azuredev/2017/05/27/data-cleansing-tools-in-azure-machine-learning/>

NEW QUESTION 34

- (Exam Topic 3)

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

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

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

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

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

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

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

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

Does the solution meet the goal?

- A. Yes
- B. No

Answer: B

Explanation:

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

Destination: Driver logs, Azure Machine Learning designer Reference:

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

NEW QUESTION 37

- (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 39

- (Exam Topic 3)
You use Azure Machine Learning designer to create a training pipeline for a regression model.
You need to prepare the pipeline for deployment as an endpoint that generates predictions asynchronously for a dataset of input data values.
What should you do?

- A. Clone the training pipeline.
- B. Create a batch inference pipeline from the training pipeline.
- C. Create a real-time inference pipeline from the training pipeline.
- D. Replace the dataset in the training pipeline with an Enter Data Manually module.

Answer: C

Explanation:
You must first convert the training pipeline into a real-time inference pipeline. This process removes training modules and adds web service inputs and outputs to handle requests.
Reference:
<https://docs.microsoft.com/en-us/azure/machine-learning/tutorial-designer-automobile-price-deploy> <https://docs.microsoft.com/en-us/azure/machine-learning/algorithm-module-reference/enter-data-manually>

NEW QUESTION 40

- (Exam Topic 3)
You create an Azure Machine Learning workspace and a new Azure DevOps organization. You register a model in the workspace and deploy the model to the target environment.
All new versions of the model registered in the workspace must automatically be deployed to the target environment.
You need to configure Azure Pipelines to deploy the model.
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

Create a service connection

Create a release pipeline

Create a build pipeline

Create an Azure DevOps project

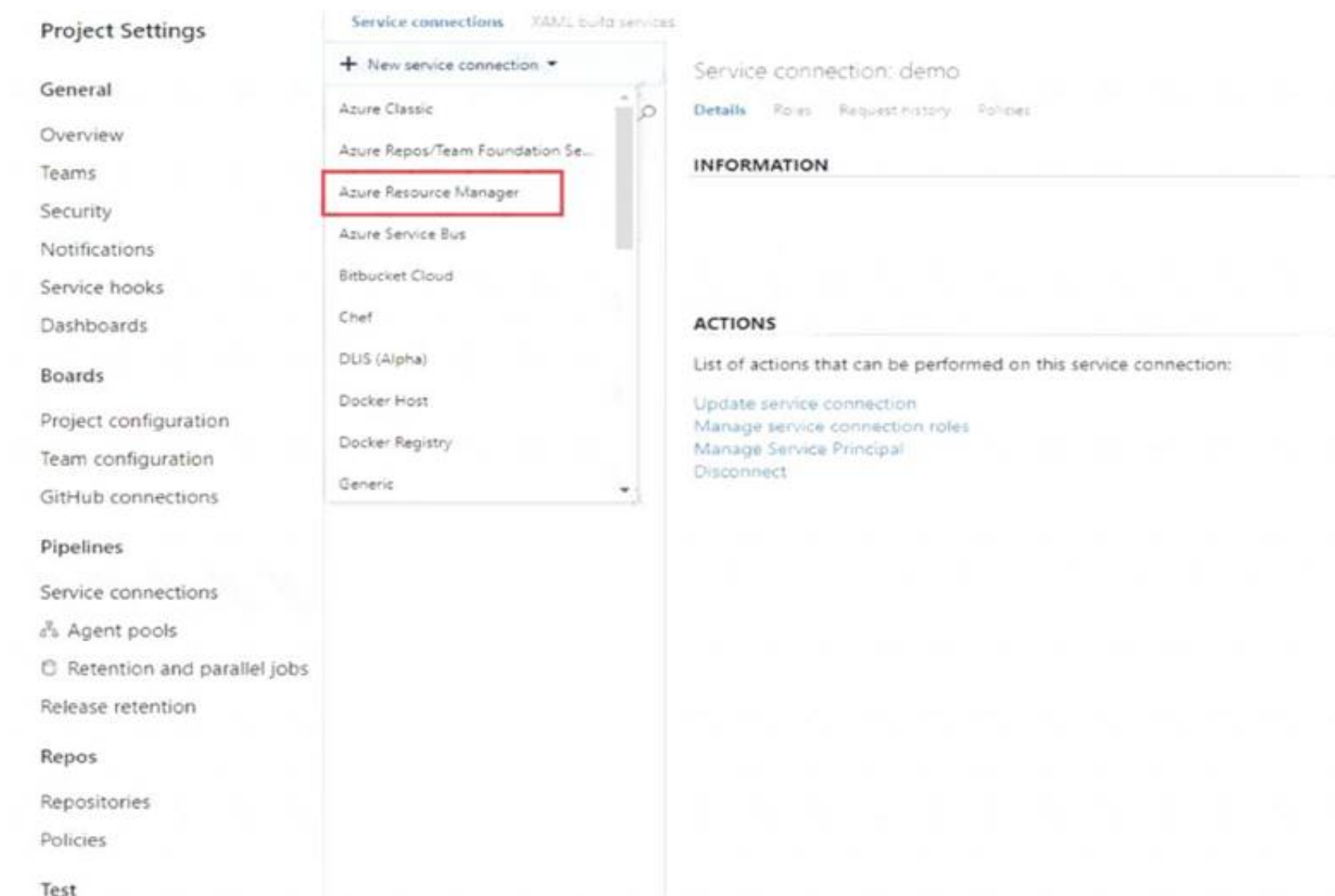
Install the Machine Learning extension for Azure Pipelines

Answer Area

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:
Graphical user interface, text, application, email Description automatically generated
Step 1: Create an Azure DevOps project
Step 2: Create a release pipeline
> Sign in to your Azure DevOps organization and navigate to your project.
> Go to Pipelines, and then select New pipeline.
Step 3: Install the Machine Learning extension for Azure Pipelines
You must install and configure the Azure CLI and ML extension.
Step 4: Create a service connection
How to set up your service connection
Graphical user interface, text, application, email Description automatically generated



Select AzureMLWorkspace for the scope level, then fill in the following subsequent parameters. Graphical user interface, text, application Description automatically generated

Note: How to enable model triggering in a release pipeline

- Go to your release pipeline and add a new artifact. Click on AzureML Model artifact then select the appropriate AzureML service connection and select from the available models in your workspace.
- Enable the deployment trigger on your model artifact as shown here. Every time a new version of that model is registered, a release pipeline will be triggered.

Reference:

<https://marketplace.visualstudio.com/items?itemName=ms-air-aiagility.vss-services-azureml> <https://docs.microsoft.com/en-us/azure/devops/pipelines/targets/azure-machine-learning>

NEW QUESTION 41

- (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.

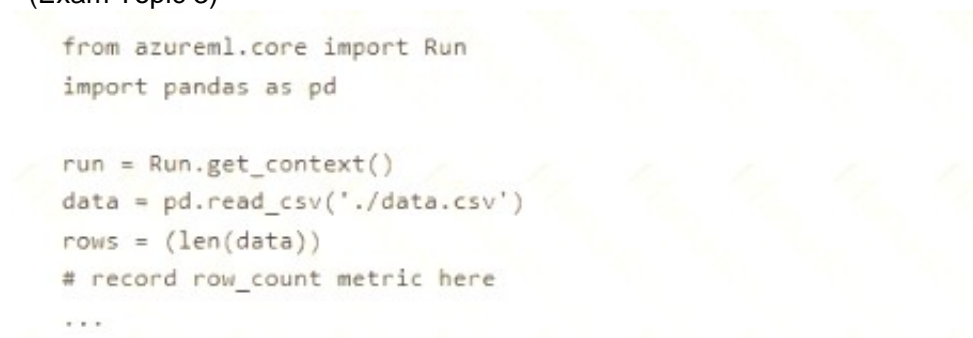
The two things you need to accomplish in your entry script are: Loading your model (using a function called init())

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 44

- (Exam Topic 3)



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 46

- (Exam Topic 3)

You use the following code to run a script as an experiment in Azure Machine Learning:

```
from azureml.core import Workspace, Experiment, Run
from azureml.core import RunConfig, ScriptRunConfig
ws = Workspace.from_config()
run_config = RunConfiguration()
run_config.target='local'
script_config = ScriptRunConfig(source_directory='./script', script='experiment.py', run_config=run_config)
experiment = Experiment(workspace=ws, name='script experiment')
run = experiment.submit(config=script_config)
run.wait_for_completion()
```

You must identify the output files that are generated by the experiment run. You need to add code to retrieve the output file names.

Which code segment should you add to the script?

- A. files = run.get_properties()
- B. files= run.get_file_names()
- C. files = run.get_details_with_logs()
- D. files = run.get_metrics()
- E. files = run.get_details()

Answer: B

Explanation:

You can list all of the files that are associated with this run record by called run.get_file_names() Reference:

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

NEW QUESTION 48

- (Exam Topic 3)

You are the owner of an Azure Machine Learning workspace.

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

You need to configure the custom role.

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

NOTE: Each correct selection is worth one point.

Answer Area

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

- A. Mastered
B. Not Mastered

Answer: A

Explanation:

Graphical user interface, application Description automatically generated

Graphical user interface, application Description automatically generated

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

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

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

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

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

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

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

NEW QUESTION 51

- (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 tin- 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 Principal Components Analysis (PCA) 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.

NEW QUESTION 55

- (Exam Topic 3)

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

You must use the Azure Machine Learning SDK to implement a batch inference pipeline that uses a ParallelRunStep to score input data using the model. You

must specify a value for the ParallelRunConfig compute_target setting of the pipeline step.
 You need to create the compute target. Which class should you use?

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

Answer: C

Explanation:

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

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

Reference:

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

NEW QUESTION 60

- (Exam Topic 3)

You plan to deliver a hands-on workshop to several students. The workshop will focus on creating data visualizations using Python. Each student will use a device that has internet access.

Student devices are not configured for Python development. Students do not have administrator access to install software on their devices. Azure subscriptions are not available for students.

You need to ensure that students can run Python-based data visualization code. Which Azure tool should you use?

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

Answer: C

Explanation:

References:

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

NEW QUESTION 64

- (Exam Topic 3)

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

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

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

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

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

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

- A. Yes
- B. NO

Answer: A

Explanation:

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

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

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

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

NEW QUESTION 69

- (Exam Topic 3)

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

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

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

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

NOTE: Each correct selection is worth one point.

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

▼

-0.106276
0.149676
0.859122
1

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

▼

a positive linear relationship
a negative linear relationship
no linear relationship

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

Box 1: 0.859122

Box 2: a positively linear relationship

+1 indicates a strong positive linear relationship

-1 indicates a strong negative linear correlation

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

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

NEW QUESTION 74

- (Exam Topic 3)

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

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

You have the following code:

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

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

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

A)

B)

C)

D)

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

Answer: D

Explanation:

Example:

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

```
}  
est = Estimator(source_directory=script_folder, script_params=script_params, compute_target=compute_target, environment_definition=env,  
entry_script='train.py')  
Reference:  
https://docs.microsoft.com/en-us/azure/machine-learning/tutorial-train-models-with-aml
```

NEW QUESTION 77

- (Exam Topic 3)

You are developing a data science workspace that uses an Azure Machine Learning service. You need to select a compute target to deploy the workspace. What should you use?

- A. Azure Data Lake Analytics
- B. Azure Databrick .
- C. Apache Spark for HDInsight.
- D. Azure Container Service

Answer: D

Explanation:

Azure Container Instances can be used as compute target for testing or development. Use for low-scale CPU-based workloads that require less than 48 GB of RAM.

Reference:

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

NEW QUESTION 78

- (Exam Topic 3)

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

Which compute target should you use?

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

Answer: B

Explanation:

Azure Machine Learning compute clusters is used for Batch inference. Run batch scoring on serverless compute. Supports normal and low-priority VMs. No support for real-time inference.

Reference:

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

NEW QUESTION 83

- (Exam Topic 3)

You are implementing a machine learning model to predict stock prices. The model uses a PostgreSQL database and requires GPU processing. You need to create a virtual machine that is pre-configured with the required tools. What should you do?

- A. Create a Data Science Virtual Machine (DSVM) Windows edition.
- B. Create a Geo AI Data Science Virtual Machine (Geo-DSVM) Windows edition.
- C. Create a Deep Learning Virtual Machine (DLVM) Linux edition.
- D. Create a Deep Learning Virtual Machine (DLVM) Windows edition.
- E. Create a Data Science Virtual Machine (DSVM) Linux edition.

Answer: E

NEW QUESTION 88

- (Exam Topic 3)

You use the Azure Machine Learning Python SDK to define a pipeline to train a model.

The data used to train the model is read from a folder in a datastore.

You need to ensure the pipeline runs automatically whenever the data in the folder changes. What should you do?

- A. Set the regenerate_outputs property of the pipeline to True
- B. Create a ScheduleRecurrence object with a Frequency of aut
- C. Use the object to create a Schedule for the pipeline
- D. Create a PipelineParameter with a default value that references the location where the training data is stored
- E. Create a Schedule for the pipelin
- F. Specify the datastore in the datastore property, and the folder containing the training data in the path_on_datascor property

Answer: D

Explanation:

Reference:

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

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 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: $\text{ZeroOneLoss}(x,y) = 1$ when $x \neq y$; otherwise 0.

Coefficient of determination, often referred to as R^2 , represents the predictive power of the model as a value between 0 and 1. Zero means the model is random (explains nothing); 1 means there is a perfect fit. However, caution should be used in interpreting R^2 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 96

- (Exam Topic 3)

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

The experiment fails.

You need to troubleshoot the failed experiment.

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

- A. Use the `get_metrics()` method of the run object to retrieve the experiment run logs.
- B. Use the `get_details_with_logs()` method of the run object to display the experiment run logs.
- C. View the log files for the experiment run in the experiment folder.
- D. View the logs for the experiment run in Azure Machine Learning studio.
- E. Use the `get_output()` method of the run object to retrieve the experiment run logs.

Answer: BD

Explanation:

Use `get_details_with_logs()` to fetch the run details and logs created by the run.

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

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

NEW QUESTION 98

- (Exam Topic 3)

You are determining if two sets of data are significantly different from one another by using Azure Machine Learning Studio.

Estimated values in one set of data may be more than or less than reference values in the other set of data. You must produce a distribution that has a constant Type I error as a function of the correlation.

You need to produce the distribution.

Which type of distribution should you produce?

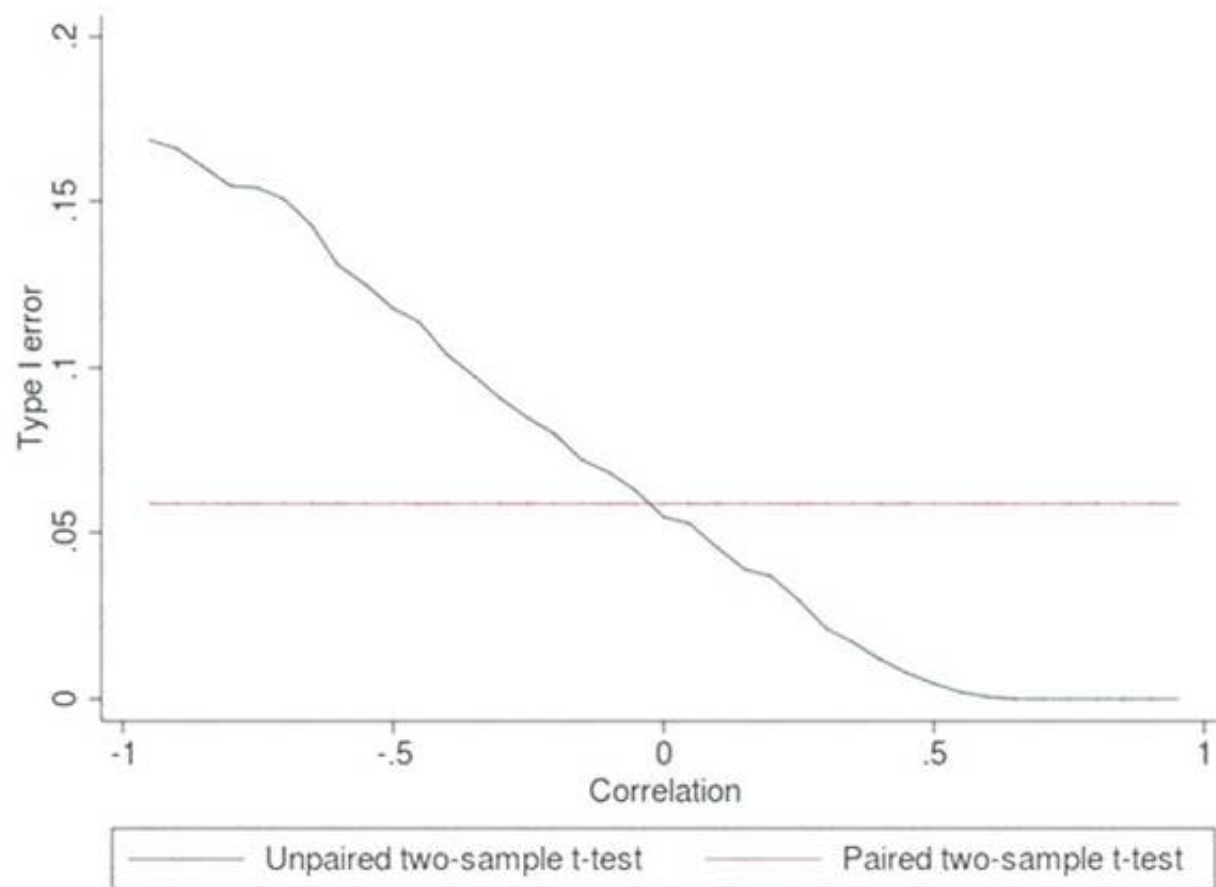
- A. Paired t-test with a two-tail option
- B. Unpaired t-test with a two tail option
- C. Paired t-test with a one-tail option
- D. Unpaired t-test with a one-tail option

Answer: A

Explanation:

Choose a one-tail or two-tail test. The default is a two-tailed test. This is the most common type of test, in which the expected distribution is symmetric around zero.

Example: Type I error of unpaired and paired two-sample t-tests as a function of the correlation. The simulated random numbers originate from a bivariate normal distribution with a variance of 1.



Reference:

<https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/test-hypothesis-using-t-test> https://en.wikipedia.org/wiki/Student%27s_t-test

NEW QUESTION 99

- (Exam Topic 3)

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

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

You train a classification model by using a logistic regression algorithm.

You must be able to explain the model's predictions by calculating the importance of each feature, both as an overall global relative importance value and as a measure of local importance for a specific set of predictions.

You need to create an explainer that you can use to retrieve the required global and local feature importance values.

Solution: Create a MimicExplainer. Does the solution meet the goal?

- A. Yes
- B. No

Answer: B

Explanation:

Instead use Permutation Feature Importance Explainer (PFI).

Note 1: Mimic explainer is based on the idea of training global surrogate models to mimic blackbox models. A global surrogate model is an intrinsically interpretable model that is trained to approximate the predictions of any black box model as accurately as possible. Data scientists can interpret the surrogate model to draw conclusions about the black box model.

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

Reference:

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

NEW QUESTION 103

- (Exam Topic 3)

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 105

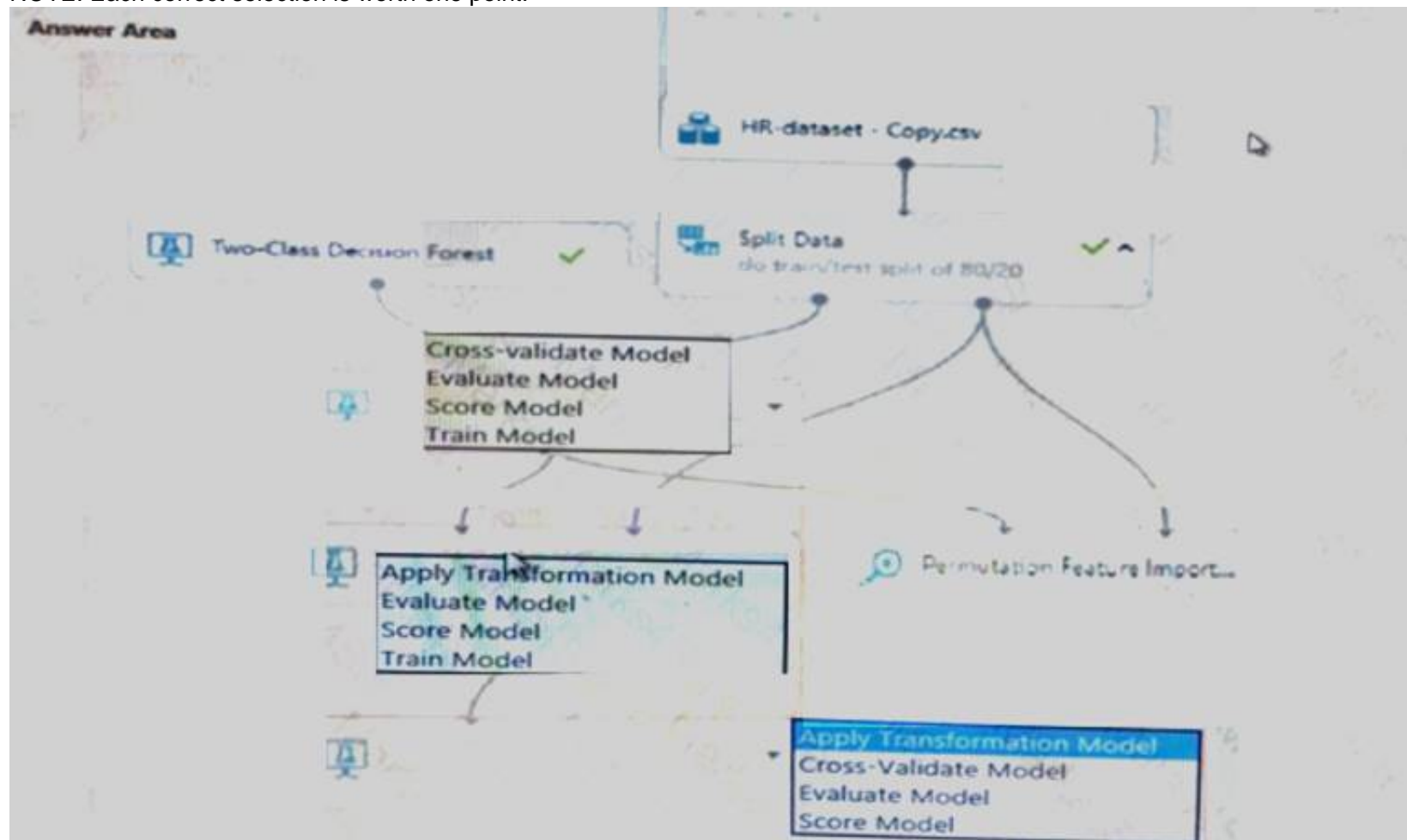
- (Exam Topic 3)

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

You must use a Receiver Operating Characteristic (RO C) curve and an F1 score to evaluate the model. You need to create the required business metrics.

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

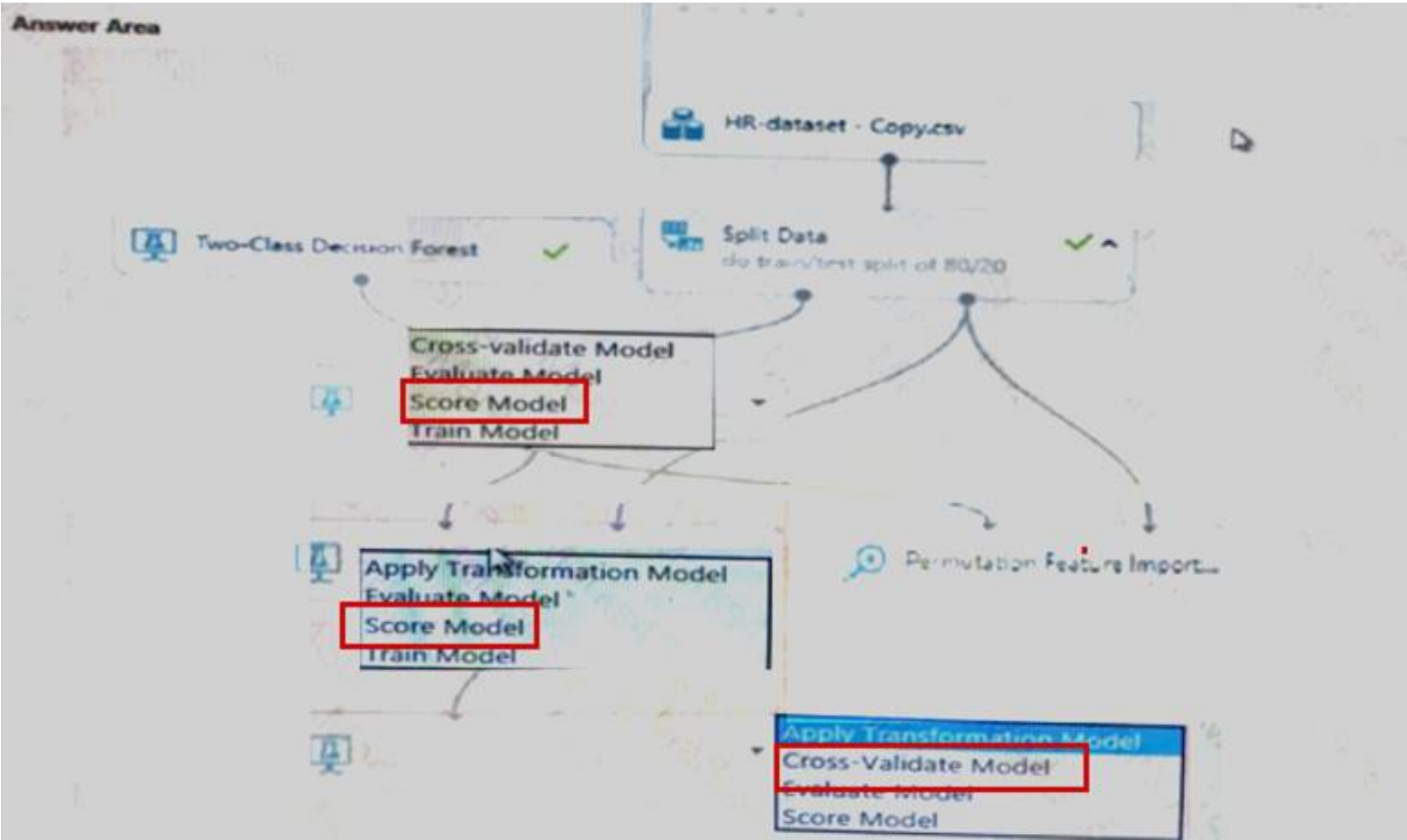
NOTE: Each correct selection is worth one point.



- A. Mastered
- B. Not Mastered

Answer: A

Explanation:



NEW QUESTION 106

- (Exam Topic 3)

You plan to explore demographic data for home ownership in various cities. The data is in a CSV file with the following format:

age,city,income,home_owner 21,Chicago,50000,0 35,Seattle,120000,1 23,Seattle,65000,0 45,Seattle,130000,1 18,Chicago,48000,0

You need to run an experiment in your Azure Machine Learning workspace to explore the data and log the results. The experiment must log the following information:

- > the number of observations in the dataset
- > a box plot of income by home_owner
- > a dictionary containing the city names and the average income for each city

You need to use the appropriate logging methods of the experiment's run object to log the required information.

How should you complete the code? To answer, drag the appropriate code segments to the correct locations. Each code segment 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.

Code segments	Answer Area
log	<pre> from azureml.core import Experiment, Run import pandas as pd import matplotlib.pyplot as plt # Create an Azure ML experiment in workspace experiment = Experiment(workspace = ws, name = "demo-experiment") # Start logging data from the experiment run = experiment.start_logging() # load the dataset data = pd.read_csv('research/demographics.csv') # Log the number of observations row_count = (len(data)) run. Segment ("observations", row_count) # Log box plot for income by home_owner fig = plt.figure(figsize=(9, 6)) ax = fig.gca() data.boxplot(column = 'income', by = "home_owner", ax = ax) ax.set_title('income by home_owner') ax.set_ylabel('income') run. Segment (name = 'income_by_home_owner', plot = fig) # Create a dataframe of mean income per city mean_inc_df = data.groupby('city')['income'].agg(np.mean).to_frame().reset_index() # Convert to a dictionary mean_inc_dict = mean_inc_df.to_dict('dict') # Log city names and average income dictionary run. Segment (name="mean_income_by_city", value= mean_inc_dict) # Complete tracking and get link to details run.complete() </pre>
log_list	
log_row	
log_table	
log_image	

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

Box 1: log

The number of observations in the dataset. `run.log(name, value, description=)`

Scalar values: Log a numerical or string value to the run with the given name. 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)`

Box 2: log_image

A box plot of income by home_owner.

log_image Log an image to the run record. Use log_image to log a .PNG image file or a matplotlib plot to the run. These images will be visible and comparable in the run record.

Example: `run.log_image("ROC", plot=plt)` Box 3: log_table

A dictionary containing the city names and the average income for each city. log_table: Log a dictionary object to the run with the given name.

NEW QUESTION 111

- (Exam Topic 3)

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

```
# ws, model, inference_config, and deployment_config defined previously
service = Model.deploy(ws, 'classification-service', [model], inference_config, deployment_config)
service.wait_for_deployment(True)
```

The deployment fails.

You need to troubleshoot the deployment failure by determining the actions that were performed during deployment and identifying the specific action that failed.

Which code segment should you run?

- A. `service.get_logs()`
- B. `service.state`
- C. `service.serialize()`
- D. `service.update_deployment_state()`

Answer: A

Explanation:

You can print out detailed Docker engine log messages from the service object. You can view the log for ACI, AKS, and Local deployments. The following example demonstrates how to print the logs.

if you already have the service object handy `print(service.get_logs())`

if you only know the name of the service (note there might be multiple services with the same name but different version number)

`print(ws.webservices['mysvc'].get_logs())` Reference:

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

NEW QUESTION 116

- (Exam Topic 3)

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

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

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

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

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

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

NOTE: Each correct selection is worth one point.


```
# Assume required modules are imported

ws = Workspace.from_config()
feature_importances = explanation.

explanation = client.

feature_importances = explanation.

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

from_run

list_model_explanations

from_run_id

download_model_explanation

(workspace = ws,
experiment_name='train_and_explain',
run_id='train_and_explain_12345')

upload_model_explanation

list_model_explanations

run

download_model_explanation

()

explanation

explanation_client

get_feature_important_dict

download_model_explanation

()

- A. Mastered
B. Not Mastered

Answer: A

Explanation:

Box 1: from_run_id

from_run_id(workspace, experiment_name, run_id) Create the client with factory method given a run ID. Returns an instance of the explanations Client.

Parameters

- > Workspace Workspace An object that represents a workspace.
- > experiment_name str The name of an experiment.
- > run_id str A GUID that represents a run.

Box 2: list_model_explanations

list_model_explanations returns a dictionary of metadata for all model explanations available.

Returns

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

Box 3: explanation:

Reference:

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

NEW QUESTION 117

- (Exam Topic 3)

You have a Python data frame named salesData in the following format:

	shop	2017	2018
0	Shop X	34	25
1	Shop Y	65	76
2	Shop Z	48	55

The data frame must be unpivoted to a long data format as follows:

	shop	year	value
0	Shop X	2017	34
1	Shop Y	2017	65
2	Shop Z	2017	48
3	Shop X	2018	25
4	Shop Y	2018	76
5	Shop Z	2018	55

You need to use the pandas.melt() function in Python to perform the transformation.

How should you complete the code segment? To answer, select the appropriate options in the answer area. NOTE: Each correct selection is worth one point.

Answer Area

```
import pandas as pd
salesData = pd.melt(
```

▼

dataFrame
pandas
salesData
year

```
, id_vars='
```

▼

shop
year
value
Shop X, Shop Y, Shop
Z

```
', value_vars=
```

▼

'shop'
'year'
['year']
['2017', '2018']

```
)
```

- A. Mastered
 B. Not Mastered

Answer: A

Explanation:

Box 1: dataFrame

Syntax: pandas.melt(frame, id_vars=None, value_vars=None, var_name=None, value_name='value', col_level=None)[source]

Where frame is a DataFrame Box 2: shop

Paramter id_vars id_vars : tuple, list, or ndarray, optional Column(s) to use as identifier variables.

Box 3: ['2017','2018']

value_vars : tuple, list, or ndarray, optional

Column(s) to unpivot. If not specified, uses all columns that are not set as id_vars. Example:

df = pd.DataFrame({'A': {0: 'a', 1: 'b', 2: 'c'},

'B': {0: 1, 1: 3, 2: 5},

'C': {0: 2, 1: 4, 2: 6}})

pd.melt(df, id_vars=['A'], value_vars=['B', 'C']) A variable value

0 a B 1

1 b B 3

2 c B 5

3 a C 2

4 b C 4

5 c C 6

References:

<https://pandas.pydata.org/pandas-docs/stable/reference/api/pandas.melt.html>

NEW QUESTION 121

- (Exam Topic 3)

You have a Python script that executes a pipeline. The script includes the following code: from azureml.core import Experiment

pipeline_run = Experiment(ws, 'pipeline_test').submit(pipeline) You want to test the pipeline before deploying the script.

You need to display the pipeline run details written to the STDOUT output when the pipeline completes. Which code segment should you add to the test script?

- A. pipeline_run.get.metrics()
 B. pipeline_run.wait_for_completion(show_output=True)
 C. pipeline_param = PipelineParameter(name="stdout", default_value="console")
 D. pipeline_run.get_status()

Answer: B

Explanation:

wait_for_completion: Wait for the completion of this run. Returns the status object after the wait. Syntax: wait_for_completion(show_output=False,

wait_post_processing=False, raise_on_error=True) Parameter: show_output

Indicates whether to show the run output on sys.stdout.

NEW QUESTION 122

- (Exam Topic 3)

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

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

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

You need to configure the module.

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

SMOTE

Label column

Selected columns:

All labels

Launch column selector

SMOTE percentage

0

300

3000

4000

Number of nearest neighbors

0

1

5

4000

Random seed

0

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

Box 1: 300

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

Box 2: 5

We should use 5 data rows.

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

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

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

References:

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

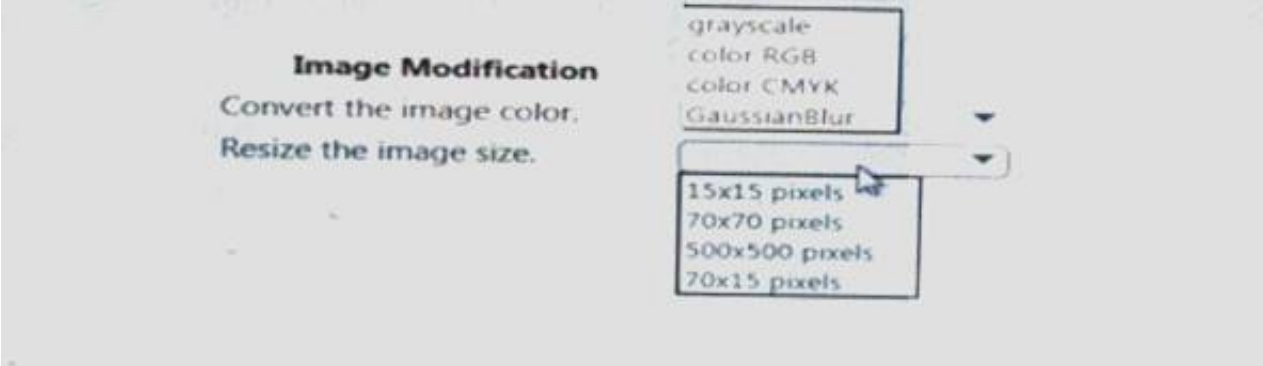
NEW QUESTION 127

- (Exam Topic 3)

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

- Reduce the number of training epochs.
 - Reduce the size of the neural network.
 - Reduce over-fitting of the neural network.
- You need to select the image modification values.

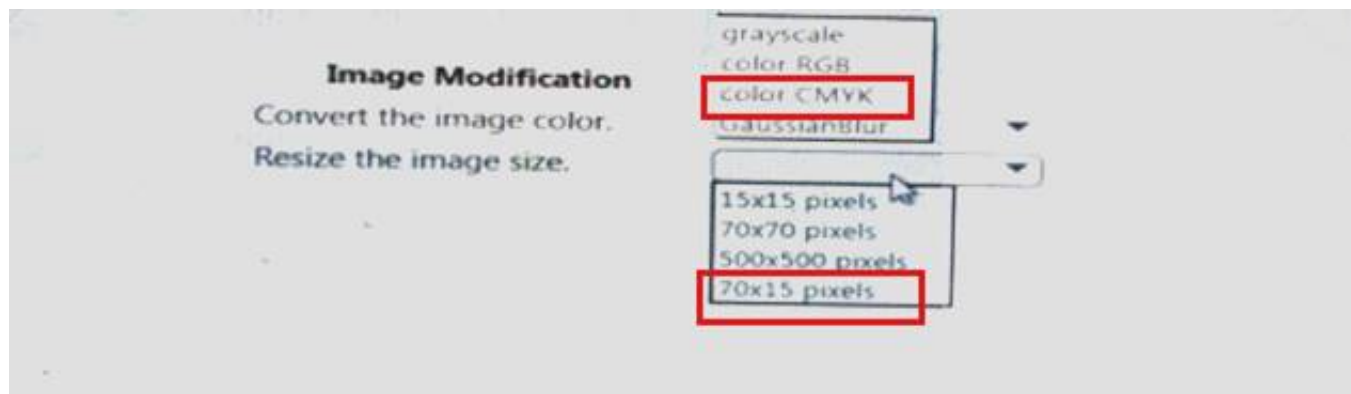
Which value should you use? To answer, select the appropriate Options in the answer area. NOTE: Each correct selection is worth one point.



- A. Mastered
- B. Not Mastered

Answer: A

Explanation:



NEW QUESTION 131

- (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 135

- (Exam Topic 3)

A biomedical research company plans to enroll people in an experimental medical treatment trial.

You create and train a binary classification model to support selection and admission of patients to the trial. The model includes the following features: Age, Gender, and Ethnicity.

The model returns different performance metrics for people from different ethnic groups.

You need to use Fairlearn to mitigate and minimize disparities for each category in the Ethnicity feature. Which technique and constraint should you use? To answer, select the appropriate options in the answer area. NOTE: Each correct selection is worth one point.

Option	Value
Technique	<div><div>Grid search</div><div>Outlier detection</div><div>Dimensionality reduction</div></div>
Constraint	<div><div>Demographic parity</div><div>False-positive rate parity</div></div>

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

Graphical user interface, text, application, chat or text message Description automatically generated

Box 1: Grid Search

Fairlearn open-source package provides postprocessing and reduction unfairness mitigation algorithms: ExponentiatedGradient, GridSearch, and ThresholdOptimizer.

Note: The Fairlearn open-source package provides postprocessing and reduction unfairness mitigation algorithms types:

- Reduction: These algorithms take a standard black-box machine learning estimator (e.g., a LightGBM model) and generate a set of retrained models using a sequence of re-weighted training datasets.
- Post-processing: These algorithms take an existing classifier and the sensitive feature as input.

Box 2: Demographic parity

The Fairlearn open-source package supports the following types of parity constraints: Demographic parity, Equalized odds, Equal opportunity, and Bounded group loss.

Reference:

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

NEW QUESTION 137

- (Exam Topic 3)

You use the Azure Machine Learning SDK to run a training experiment that trains a classification model and calculates its accuracy metric.

The model will be retrained each month as new data is available. You must register the model for use in a batch inference pipeline.

You need to register the model and ensure that the models created by subsequent retraining experiments are registered only if their accuracy is higher than the currently registered model.

What are two possible ways to achieve this goal? Each correct answer presents a complete solution. NOTE: Each correct selection is worth one point.

- A. Specify a different name for the model each time you register it.
- B. Register the model with the same name each time regardless of accuracy, and always use the latest version of the model in the batch inferencing pipeline.
- C. Specify the model framework version when registering the model, and only register subsequent models if this value is higher.
- D. Specify a property named accuracy with the accuracy metric as a value when registering the model, and only register subsequent models if their accuracy is higher than the accuracy property value of the currently registered model.
- E. Specify a tag named accuracy with the accuracy metric as a value when registering the model, and only register subsequent models if their accuracy is higher than the accuracy tag value of the currently registered mode

Answer: CE

Explanation:

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

Reference:

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

NEW QUESTION 142

- (Exam Topic 3)

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

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

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

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

Solution: Run the following code:

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

Does the solution meet the goal?

- A. Yes
- B. No

Answer: A

Explanation:

The two steps are present: process_step and train_step Note:

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

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

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

from azureml.pipeline.steps import PythonScriptStep datastore = ws.get_default_datastore()

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

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

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

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

NEW QUESTION 147

- (Exam Topic 3)

You use the following code to define the steps for a pipeline: from azureml.core import Workspace, Experiment, Run from azureml.pipeline.core import Pipeline from azureml.pipeline.steps import PythonScriptStep ws = Workspace.from_config()

```
...
step1 = PythonScriptStep(name="step1", ...) step2 = PythonScriptsStep(name="step2", ...) pipeline_steps = [step1, step2]
```

You need to add code to run the steps.

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

NOTE: Each correct selection is worth one point.

- A. experiment = Experiment(workspace=ws, name='pipeline-experiment')run = experiment.submit(config=pipeline_steps)
- B. run = Run(pipeline_steps)
- C. pipeline = Pipeline(workspace=ws, steps=pipeline_steps) experiment = Experiment(workspace=ws, name='pipeline-experiment')run = experiment.submit(pipeline)
- D. pipeline = Pipeline(workspace=ws, steps=pipeline_steps)run = pipeline.submit(experiment_name='pipeline-experiment')

Answer: CD

Explanation:

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

Build the pipeline. Example:

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

Submit the pipeline to be run

pipeline_run1 = Experiment(ws, 'Compare_Models_Exp').submit(pipeline1) Reference:

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

NEW QUESTION 149

- (Exam Topic 3)

You are analyzing a dataset by using Azure Machine Learning Studio.

YOU need to generate a statistical summary that contains the p value and the unique value count for each feature column.

Which two modules can you users? Each correct answer presents a complete solution. NOTE: Each correct selection is worth one point.

- A. Execute Python Script
- B. Export Count Table
- C. Convert to Indicator Values
- D. Summarize Data
- E. Compute linear Correlation

Answer: BE

Explanation:

The Export Count Table module is provided for backward compatibility with experiments that use the Build Count Table (deprecated) and Count Featurizer (deprecated) modules.

E: Summarize Data statistics are useful when you want to understand the characteristics of the complete dataset. For example, you might need to know:

How many missing values are there in each column? How many unique values are there in a feature column?

What is the mean and standard deviation for each column?

The module calculates the important scores for each column, and returns a row of summary statistics for each variable (data column) provided as input.

References:

<https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/export-count-table> <https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/summarize-data>

NEW QUESTION 154

- (Exam Topic 3)

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

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

NOTE: Each correct selection is worth one point.

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

Answer: BCD

Explanation:

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

AzCopy Python SSIS

References:

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

NEW QUESTION 155

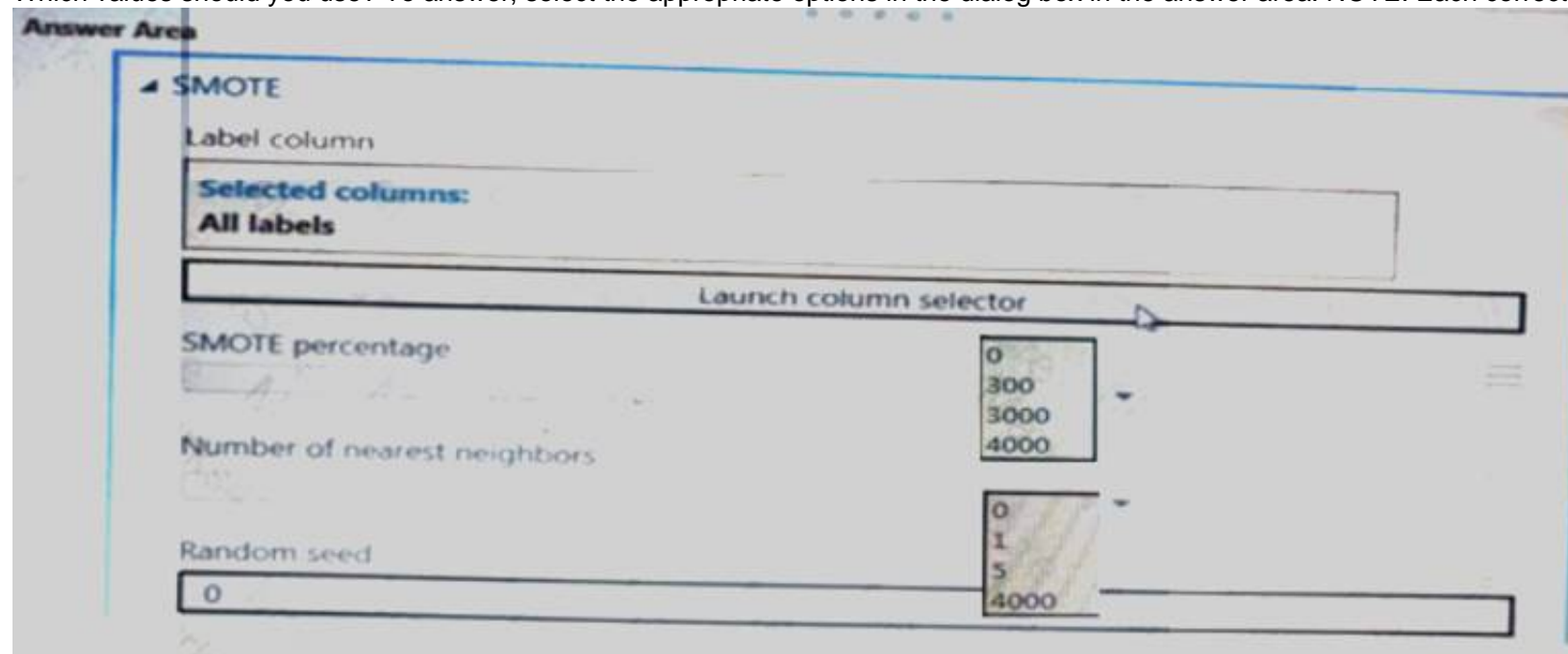
- (Exam Topic 3)

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.

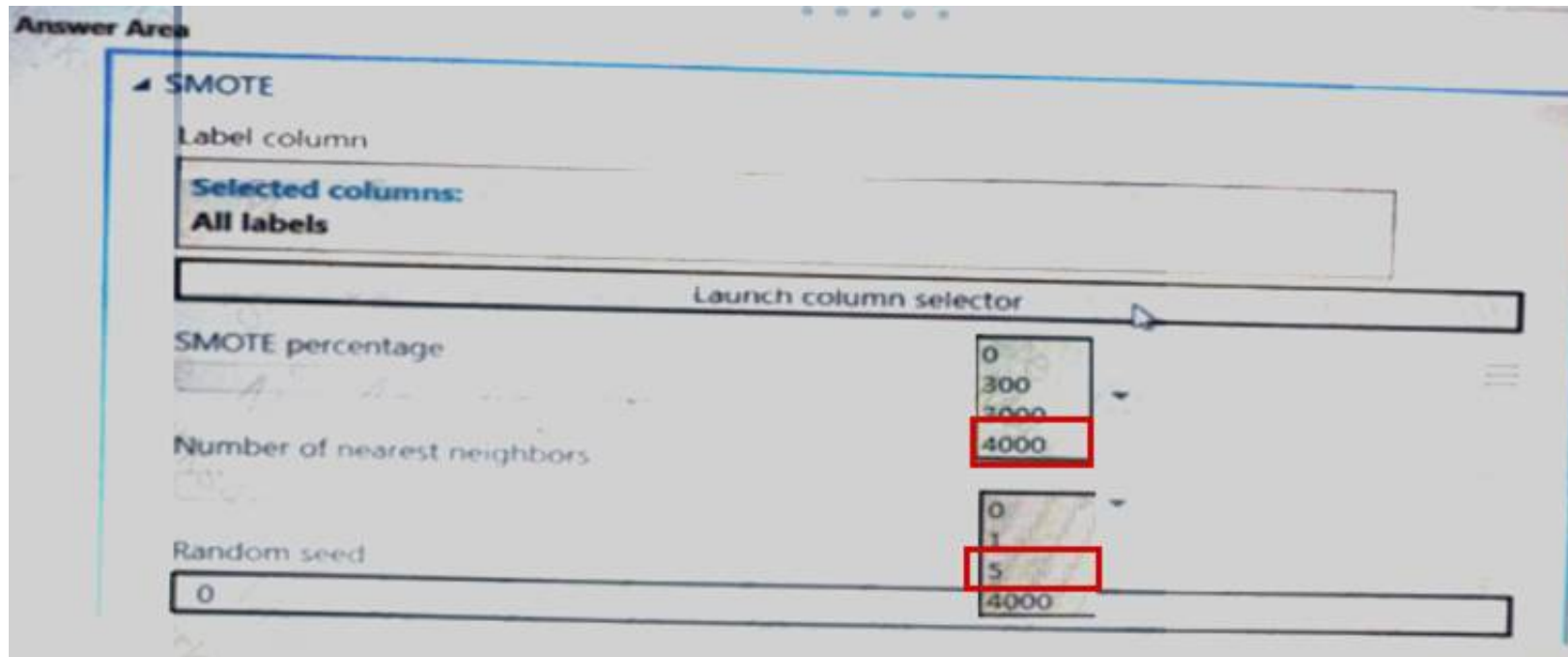
Which values should you use? To answer, select the appropriate options in the dialog box in the answer area. NOTE: Each correct selection is worth one point.



- A. Mastered
- B. Not Mastered

Answer: A

Explanation:



NEW QUESTION 160

- (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 Scale and Reduce 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 163

- (Exam Topic 3)

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

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

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

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

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

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

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

NOTE: Each correct selection is worth one point.

```
from numpy import array
from sklearn.model_selection import K-Means
k-fold
CrossValidation
ModelSelection

data = array([10, 20, 30, 40, 50, 60])
kfold = Kfold(n_splits=1
2
3
6, shuffle = True, random_state=1)

for train, test in kFold, split(data
k-fold
array
train, test):

print('train: %s, test: %5' % (data[train], data[test]))
```

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

Box 1: k-fold

Box 2: 3

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

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

Example: Example:

```
>>>
```

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

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

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

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

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

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

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

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

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

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

References:

https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.KFold.html

NEW QUESTION 166

- (Exam Topic 3)

You are performing sentiment analysis using a CSV file that includes 12,000 customer reviews written in a short sentence format. You add the CSV file to Azure Machine Learning Studio and configure it as the starting point dataset of an experiment. You add the Extract N-Gram Features from Text module to the experiment to extract key phrases from the customer review column in the dataset.

You must create a new n-gram dictionary from the customer review text and set the maximum n-gram size to trigrams.

What should you select? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.

Properties

Project

Extract N-Gram Features from Text

Text column

Selected columns

Column type: String Feature

Launch column selector

Vocabulary mode

▼

Create

ReadOnly

Update

Merge

N-Grams size

▼

3

4

4,000

12,000

0

Weighting function

▼

Minimum word length

3

Maximum word length

25

Minimum n-gram document absolu...

5

Maximum n-gram document ratio

1

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

Vocabulary mode: Create

For Vocabulary mode, select Create to indicate that you are creating a new list of n-gram features. N-Grams size: 3

For N-Grams size, type a number that indicates the maximum size of the n-grams to extract and store. For example, if you type 3, unigrams, bigrams, and trigrams will be created.

Weighting function: Leave blank

The option, Weighting function, is required only if you merge or update vocabularies. It specifies how terms in the two vocabularies and their scores should be weighted against each other.

References:

<https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/extract-n-gram-features-from>

NEW QUESTION 167

- (Exam Topic 3)

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

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

Which of the following is correct?

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

Answer: B

Explanation:

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

References:

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

NEW QUESTION 169

- (Exam Topic 3)

Your team is building a data engineering and data science development environment. The environment must support the following requirements:

- support Python and Scala
 - compose data storage, movement, and processing services into automated data pipelines
 - the same tool should be used for the orchestration of both data engineering and data science
 - support workload isolation and interactive workloads
 - enable scaling across a cluster of machines
- You need to create the environment.

What should you do?

- A. Build the environment in Apache Hive for HDInsight and use Azure Data Factory for orchestration.
- B. Build the environment in Azure Databricks and use Azure Data Factory for orchestration.
- C. Build the environment in Apache Spark for HDInsight and use Azure Container Instances for orchestration.
- D. Build the environment in Azure Databricks and use Azure Container Instances for orchestration.

Answer: B

Explanation:

In Azure Databricks, we can create two different types of clusters.

- Standard, these are the default clusters and can be used with Python, R, Scala and SQL
- High-concurrency

Azure Databricks is fully integrated with Azure Data Factory.

NEW QUESTION 173

- (Exam Topic 3)

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

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

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

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

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

1,1.2,0

2,1,1,

1 3,2.1,0

You run the following code:

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

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

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

Solution: Run the following code:

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

Does the solution meet the goal?

- A. Yes
- B. No

Answer: B

Explanation:

Use two file paths.

Use Dataset.Tabular_from_delimited, instead of Dataset.File.from_files as the data isn't cleansed. Reference:

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

NEW QUESTION 174

- (Exam Topic 3)

You are analyzing a raw dataset that requires cleaning.

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

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

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

Answer Area

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

- A. Mastered
B. Not Mastered

Answer: A

Explanation:

Box 1: Clean Missing Data Box 2: SMOTE

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

Box 3: Convert to Indicator Values

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

Box 4: Remove Duplicate Rows References:

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

NEW QUESTION 179

- (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: for label_val in label_vals:

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

Does the solution meet the goal?

- A. Yes
B. No

Answer: A

Explanation:

The run_log function is used 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 184

- (Exam Topic 3)

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

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

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

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

run = Run.get_context()
data = pd.read_csv('data.csv')
label_vals = data['label'].unique()
```

Add code to record metrics here run.complete()
 The experiment must record the unique labels in the data as metrics for the run that can be reviewed later.
 You must add code to the script to record the unique label values as run metrics at the point indicated by the comment.
 Solution: Replace the comment with the following code:
 run.upload_file('outputs/labels.csv', './data.csv') Does the solution meet the goal?

- A. Yes
- B. No

Answer: B

Explanation:
 label_vals has the unique labels (from the statement label_vals = data['label'].unique()), and it has to be logged.
 Note:
 Instead use the run_log function to log the contents in label_vals: for label_val in label_vals:
 run.log('Label Values', label_val) Reference:
<https://www.element61.be/en/resource/azure-machine-learning-services-complete-toolbox-ai>

NEW QUESTION 185

- (Exam Topic 3)
 An organization uses Azure Machine Learning service and wants to expand their use of machine learning. You have the following compute environments. The organization does not want to create another compute environment.

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

You need to determine which compute environment to use for the following scenarios.
 Which compute types should you use? To answer, drag the appropriate compute environments to the correct scenarios. Each compute environment may be used once, more than once, or not at all. You may need to drag the split bar between panes or scroll to view content.
 NOTE: Each correct selection is worth one point.

Environments

nb_server

aks_cluster

mlc_cluster

Answer Area

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

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:
 Box 1: nb_server

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

Box 2: mlc_cluster

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

Reference:

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

NEW QUESTION 189

- (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 191

- (Exam Topic 3)

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

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

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

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

Which code segment should you use?

- A. arguments = ['--input-data', ds.to_pandas_dataframe()]
- B. arguments = ['--input-data', ds.as_mount()]
- C. arguments = ['--data-data', ds]
- D. arguments = ['--input-data', ds.as_download()]

Answer: A

Explanation:

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

To load the TabularDataset to pandas DataFrame df = dataset.to_pandas_dataframe()

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

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

NEW QUESTION 195

- (Exam Topic 3)

You are creating a machine learning model in Python. The provided dataset contains several numerical columns and one text column. The text column represents a product's category. The product category will always be one of the following:

- > Bikes
- > Cars
- > Vans
- > Boats

You are building a regression model using the scikit-learn Python package.

You need to transform the text data to be compatible with the scikit-learn Python package.

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

NOTE: Each correct selection is worth one point.

```
from sklearn import linear_model
import
    pandas as df
    numpy as df
    scipy as df

dataset = df.read_csv("data\\ProductSales.csv")
ProductCategoryMapping = {"Bikes":1, "Cars":2, "Boats": 3,
"Vans": 4}
dataset['ProductCategoryMapping'] =
dataset['ProductCategory'].
    map[ProductCategoryMapping]
    reduce[ProductCategoryMapping]
    transpose[ProductCategoryMapping]

regr = linear_model.LinearRegression()
X_train = dataset[['ProductCategoryMapping', 'ProductSize',
'ProductCost']]
y_train = dataset[['Sales']]
regr.fit(X_train, y_train)
```

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

Box 1: pandas as df

Pandas takes data (like a CSV or TSV file, or a SQL database) and creates a Python object with rows and columns called data frame that looks very similar to table in a statistical software (think Excel or SPSS for example).

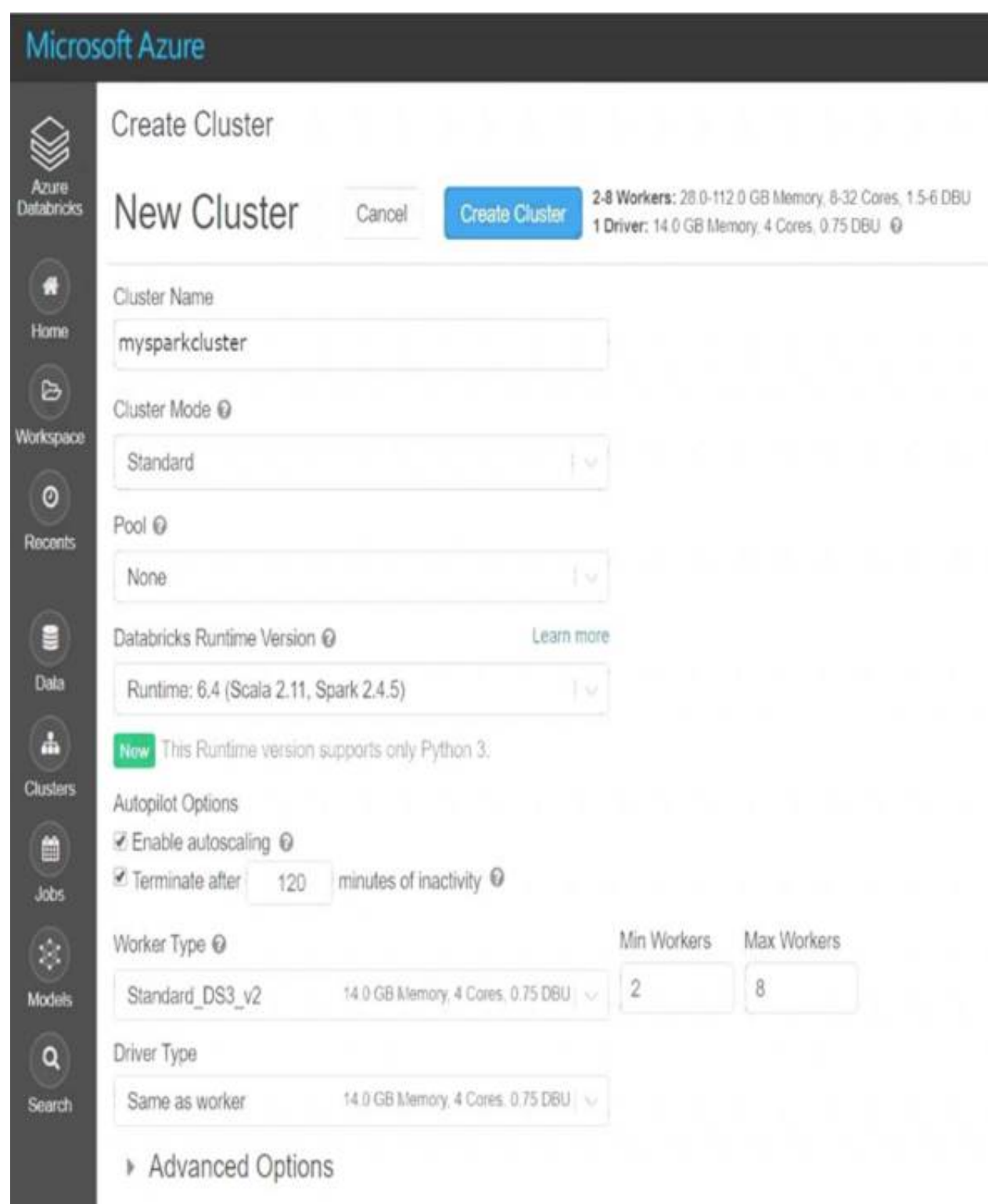
Box 2: transpose[ProductCategoryMapping] Reshape the data from the pandas Series to columns. Reference:
<https://datascienceplus.com/linear-regression-in-python/>

NEW QUESTION 196

- (Exam Topic 3)

You create a new Azure Databricks workspace.

You configure a new cluster for long-running tasks with mixed loads on the compute cluster as shown in the image below.



Microsoft Azure

Azure Databricks

Create Cluster

New Cluster Cancel Create Cluster 2-8 Workers: 28.0-112.0 GB Memory, 8-32 Cores, 1.5-6 DBU
 1 Driver: 14.0 GB Memory, 4 Cores, 0.75 DBU ⓘ

Cluster Name
 mysparkcluster

Cluster Mode ⓘ
 Standard

Pool ⓘ
 None

Databricks Runtime Version ⓘ [Learn more](#)
 Runtime: 6.4 (Scala 2.11, Spark 2.4.5)

Now This Runtime version supports only Python 3.

Autopilot Options
☒ Enable autoscaling ⓘ
☒ Terminate after 120 minutes of inactivity ⓘ

Worker Type ⓘ Min Workers Max Workers
 Standard_DS3_v2 14.0 GB Memory, 4 Cores, 0.75 DBU 2 8

Driver Type
 Same as worker 14.0 GB Memory, 4 Cores, 0.75 DBU

▶ Advanced Options

Use the drop-down menus to select the answer choice that completes each statement based on the information presented in the graphic.
 NOTE: Each correct selection is worth one point.

Code for each user runs as a separate process

	▼
Yes	
No	

The number of workers is fixed for the entire duration of the job

	▼
Yes	
No	

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

Box 1: No
 Running user code in separate processes is not possible in Scala. Box 2: No
 Autoscaling is enabled. Minimum 2 workers, Maximum 8 workers. Reference:
<https://docs.databricks.com/clusters/configure.html>

NEW QUESTION 197

- (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 201

- (Exam Topic 3)

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

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

You are a data scientist using Azure Machine Learning Studio.

You need to normalize values to produce an output column into bins to predict a target column. Solution: Apply a Quantiles normalization with a QuantileIndex normalization.

Does the solution meet the GOAL?

- A. Yes
- B. No

Answer: B

Explanation:

Use the Entropy MDL binning mode which has a target column. References:

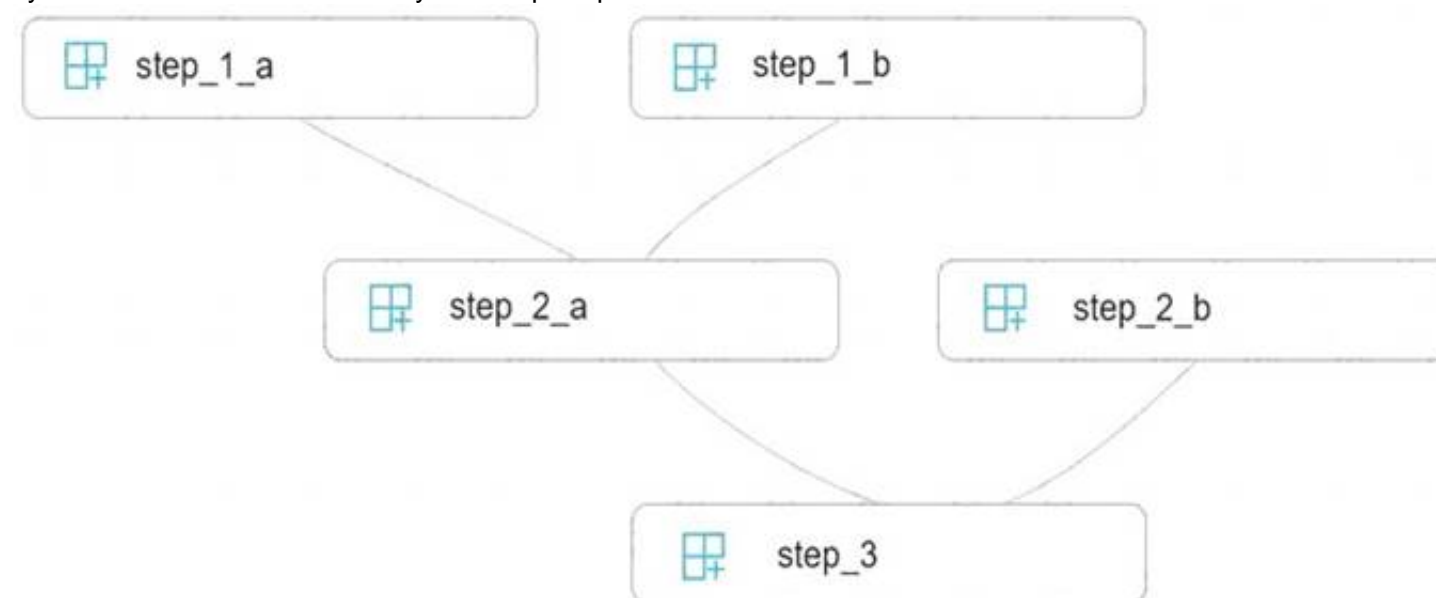
<https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/group-data-into-bins>

NEW QUESTION 202

- (Exam Topic 3)

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

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



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

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

```

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

```
- C. `step_2_a.run_after(step_1_b)`
`step_2_a.run_after(step_1_a)`
`step_3.run_after(step_2_b)`
`step_3.run_after(step_2_a)`
`p = Pipeline(ws, steps=[step_3])`
- D. `p = Pipeline(ws, steps=[step_1_a, step_1_b, step_2_a, step_2_b, step_3])`

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

Answer: A

Explanation:

The steps parameter is an array of steps. To build pipelines that have multiple steps, place the steps in order in this array.
Reference:
<https://docs.microsoft.com/en-us/azure/machine-learning/how-to-use-parallel-run-step>

NEW QUESTION 206

- (Exam Topic 3)

You are performing a classification task in Azure Machine Learning Studio.
You must prepare balanced testing and training samples based on a provided data set. You need to split the data with a 0.75:0.25 ratio.
Which value should you use for each parameter? To answer, select the appropriate options in the answer area.
NOTE: Each correct selection is worth one point.

Parameter	Value
Splitting mode	<div><div>▼</div><div>Split rows Recommender Split Regular Expression Split Relative Expression Split</div></div>
Fraction of rows in the first output dataset	<div><div>▼</div><div>0.75 0.25 0.5 1</div></div>
Randomized split	<div><div>▼</div><div>True False</div></div>
Stratified split	<div><div>▼</div><div>True False</div></div>

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

Box 1: Split rows
Use the Split Rows option if you just want to divide the data into two parts. You can specify the percentage of data to put in each split, but by default, the data is divided 50-50.
You can also randomize the selection of rows in each group, and use stratified sampling. In stratified sampling, you must select a single column of data for which you want values to be apportioned equally among the two result datasets.
Box 2: 0.75
If you specify a number as a percentage, or if you use a string that contains the "%" character, the value is interpreted as a percentage. All percentage values must be within the range (0, 100), not including the values 0 and 100.
Box 3: Yes
To ensure splits are balanced.
Box 4: No
If you use the option for a stratified split, the output datasets can be further divided by subgroups, by selecting a strata column.
Reference:
<https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/split-data>

NEW QUESTION 210

- (Exam Topic 3)

You are a data scientist working for a bank and have used Azure ML to train and register a machine learning model that predicts whether a customer is likely to repay a loan.
You want to understand how your model is making selections and must be sure that the model does not violate government regulations such as denying loans based on where an applicant lives.
You need to determine the extent to which each feature in the customer data is influencing predictions. What should you do?

- A. Enable data drift monitoring for the model and its training dataset.
- B. Score the model against some test data with known label values and use the results to calculate a confusion matrix.
- C. Use the Hyperdrive library to test the model with multiple hyperparameter values.
- D. Use the interpretability package to generate an explainer for the model.
- E. Add tags to the model registration indicating the names of the features in the training dataset.

Answer: D

Explanation:
for your model with different test data. The steps in this section show you how to compute and visualize engineered feature importance based on your test data.
Reference:
<https://docs.microsoft.com/en-us/azure/machine-learning/how-to-machine-learning-interpretability-automl>

NEW QUESTION 211
- (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 215
- (Exam Topic 3)
You create an Azure Machine Learning workspace and set up a development environment. You plan to train a deep neural network (DNN) by using the Tensorflow framework and by using estimators to submit training scripts.
You must optimize computation speed for training runs.
You need to choose the appropriate estimator to use as well as the appropriate training compute target configuration.
Which values should you use? To answer, select the appropriate options in the answer area. NOTE: Each correct selection is worth one point.

Answer Area

Parameter	Value
Estimator	<div>Estimator</div> <div>SkLearn</div> <div>PyTorch</div> <div>Tensorflow</div> <div>Chainer</div>
Training compute	<div>12 vCPU, 48 GB memory, 96 GB SSD</div> <div>12 vCPU, 112 GB memory, 680 GB SSD, 2 GPU, 24 GB GPU memory</div> <div>16 vCPU, 128 GB memory, 160 GB HDD, 80 GB NVME disk (4000 MBps)</div> <div>44 vCPU, 352 GB memory, 3.4 GHz CPU frequency all cores</div>

Click on the selection for training compute

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:
Box 1: Tensorflow
TensorFlow represents an estimator for training in TensorFlow experiments. Box 2: 12 vCPU, 112 GB memory...,2 GPU,...
Use GPUs for the deep neural network. Reference:
<https://docs.microsoft.com/en-us/python/api/azureml-train-core/azureml.train.dnn>

NEW QUESTION 218
- (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 223

- (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 226

- (Exam Topic 3)

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

- A. HDInsight

- B. Azure Databricks
- C. Azure Kubernetes Services
- D. the existing Machine Learning Compute resource
- E. a new Machine Learning Compute resource

Answer: C

Explanation:

Azure Kubernetes Service (AKS) can be used real-time inference. Reference:
<https://docs.microsoft.com/en-us/azure/machine-learning/concept-compute-target>

NEW QUESTION 231

- (Exam Topic 3)

You are creating a machine learning model. You have a dataset that contains null rows.

You need to use the Clean Missing Data module in Azure Machine Learning Studio to identify and resolve the null and missing data in the dataset.

Which parameter should you use?

- A. Replace with mean
- B. Remove entire column
- C. Remove entire row
- D. Hot Deck

Answer: B

Explanation:

Remove entire row: Completely removes any row in the dataset that has one or more missing values. This is useful if the missing value can be considered randomly missing.

References:

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

NEW QUESTION 232

- (Exam Topic 3)

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

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

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

A)

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

B)

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

C)

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

D)

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

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

Answer: A

Explanation:

The get_output method on automl_classifier returns the best run and the fitted model for the last invocation. Overloads on get_output allow you to retrieve the best run and fitted model for any logged metric or for a particular iteration.

In []:

best_run, fitted_model = local_run.get_output() Reference:

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

NEW QUESTION 235

- (Exam Topic 3)

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

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

You are a data scientist using Azure Machine Learning Studio.

You need to normalize values to produce an output column into bins to predict a target column. Solution: Apply a Quantiles binning mode with a PQuantile normalization.

Does the solution meet the goal?

- A. Yes
- B. No

Answer: B

Explanation:

Use the Entropy MDL binning mode which has a target column. References:

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

NEW QUESTION 239

- (Exam Topic 3)

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

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

Which compute type should you use?

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

Answer: B

Explanation:

You can use Azure Machine Learning to deploy a GPU-enabled model as a web service. Deploying a model on Azure Kubernetes Service (AKS) is one option.

The AKS cluster provides a GPU resource that is used by the model for inference.

Inference, or model scoring, is the phase where the deployed model is used to make predictions. Using GPUs instead of CPUs offers performance advantages on highly parallelizable computation.

Reference:

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

NEW QUESTION 240

- (Exam Topic 3)

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

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

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

- A. Implement the `download_dashboard_by_upload_id` function
- B. Implement the `create_group_metric_sec` function
- C. Implement the `upload_dashboard_dictionary` function
- D. Upload the training script

Answer: C

Explanation:

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

from `azureml.contrib.fairness` import `upload_dashboard_dictionary`, `download_dashboard_by_upload_id` Reference:

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

NEW QUESTION 241

- (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 243

- (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

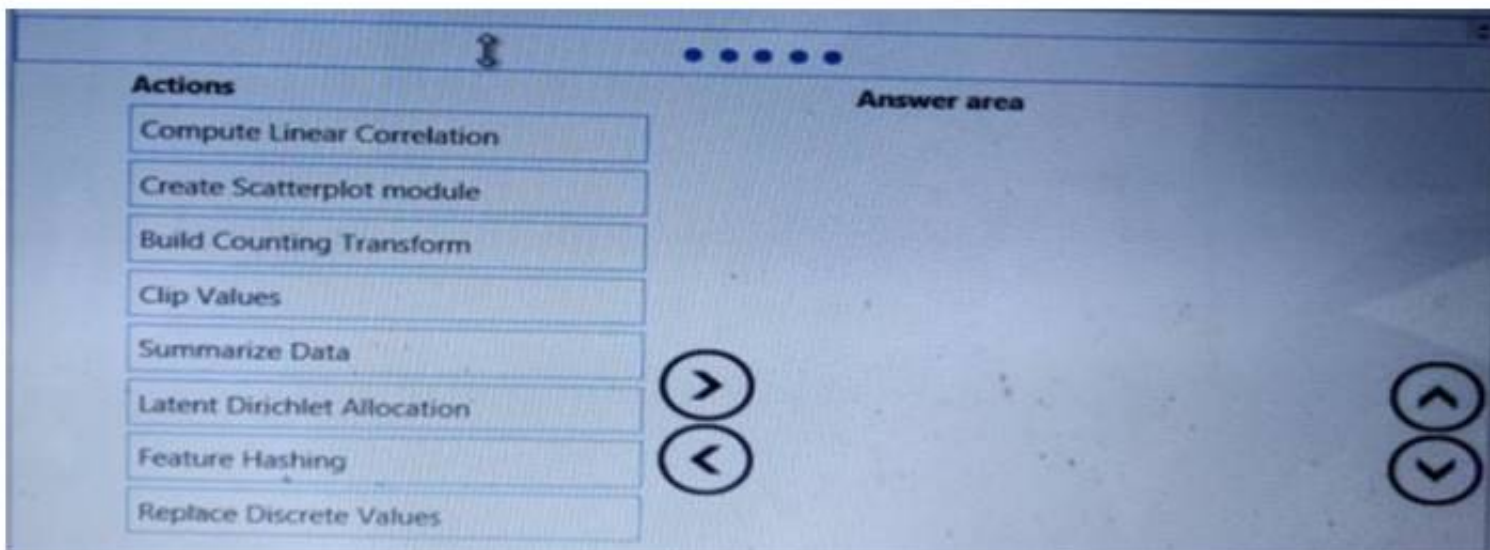
Answer: A

NEW QUESTION 246

- (Exam Topic 2)

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

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



- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

Create Scatterplot Summarize Data Clip Values

You can use the Clip Values module in Azure Machine Learning Studio, to identify and optionally replace data values that are above or below a specified threshold. This is useful when you want to remove outliers or replace them with a mean, a constant, or other substitute value.

References:

<https://blogs.msdn.microsoft.com/azuredev/2017/05/27/data-cleansing-tools-in-azure-machine-learning/> <https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/clip-values>

NEW QUESTION 249

- (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 252

- (Exam Topic 2)
You need to select a feature extraction method. Which method should you use?

- A. Mutual information
- B. Mood's median test
- C. Kendall correlation
- D. Permutation Feature Importance

Answer: C

Explanation:

In statistics, the Kendall rank correlation coefficient, commonly referred to as Kendall's tau coefficient (after the Greek letter τ), is a statistic used to measure the ordinal association between two measured quantities.
It is a supported method of the Azure Machine Learning Feature selection.
Scenario: When you train a Linear Regression module using a property dataset that shows data for property prices for a large city, you need to determine the best features to use in a model. You can choose standard metrics provided to measure performance before and after the feature importance process completes. You must ensure that the distribution of the features across multiple training models is consistent.
References:
<https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/feature-selection-modules>

NEW QUESTION 256

.....

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](#)