

Exam Questions DP-100

Designing and Implementing a Data Science Solution on Azure

<https://www.2passeasy.com/dumps/DP-100/>



NEW QUESTION 1

- (Exam Topic 3)

HOTSPOT

You register the following versions of a model.

Model name	Model version	Tags	Properties
healthcare_model	3	'Training context': 'CPU Compute'	value:87.43
healthcare_model	2	'Training context': 'CPU Compute'	value:54.98
healthcare_model	1	'Training context': 'CPU Compute'	value:23.56

You use the Azure ML Python SDK to run a training experiment. You use a variable named run to reference the experiment run. After the run has been submitted and completed, you run the following code:

```
run.register_model(model_path='outputs/model.pkl',
    model_name='healthcare_model',
    tags={'Training context': 'CPU Compute'})
```

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 code will cause a previous version of the saved model to be overwritten.	<input type="radio"/>	<input type="radio"/>
The version number will now be 4.	<input type="radio"/>	<input type="radio"/>
The latest version of the stored model will have a property of value: 87.43.	<input type="radio"/>	<input type="radio"/>

- A. Mastered
 B. Not Mastered

Answer: A

Explanation:

Reference:

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

NEW QUESTION 2

- (Exam Topic 3)

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

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

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

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

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

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

Does the solution meet the goal?

- A. Yes
 B. No

Answer: B

Explanation:

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

Reference:

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

NEW QUESTION 3

- (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 4

- (Exam Topic 3)

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

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

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

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

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

Step 1: Use the Filter Based Feature Selection module

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

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

Step 2: Build a counting transform

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

Step 3: Test the hypothesis using t-Test References:

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

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

NEW QUESTION 5

- (Exam Topic 3)

You use an Azure Machine Learning workspace. You create the following Python code:

```
from azureml.core import ScriptRunConfig
src = ScriptRunConfig(source_directory=project_folder,
                      script='train.py'
                      environment=myenv)
```

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 default environment will be created	<input type="radio"/>	<input type="radio"/>
The training script will run on local compute	<input type="radio"/>	<input type="radio"/>
A script run configuration runs a training script named <code>train.py</code> located in a directory defined by the <code>project_folder</code> variable	<input type="radio"/>	<input type="radio"/>

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

Graphical user interface, text, application Description automatically generated

Box 1: No

Environment is a required parameter. The environment to use for the run. If no environment is specified, `azureml.core.runconfig.DEFAULT_CPU_IMAGE` will be used as the Docker image for the run.

The following example shows how to instantiate a new environment. `from azureml.core import Environment`

`myenv = Environment(name="myenv")` Box 2: Yes

Parameter `compute_target`: The compute target where training will happen. This can either be a `ComputeTarget` object, the name of an existing `ComputeTarget`, or the string "local". If no compute target is specified, your local machine will be used.

Box 3: Yes

Parameter `source_directory`. A local directory containing code files needed for a run. Parameter `script`. The file path relative to the `source_directory` of the script to be run. Reference:

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

NEW QUESTION 6

- (Exam Topic 3)

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

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

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

NOTE: Each correct selection is worth one point.

Answer Area

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

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

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

- A. Mastered
 B. Not Mastered

Answer: A

Explanation:

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

Box 1: source_directory

source_directory: A local directory containing code files needed for a run. Box 2: script

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

Reference:

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

NEW QUESTION 7

- (Exam Topic 3)

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

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

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

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

Does the solution meet the goal?

- A. Yes
 B. No

Answer: B

Explanation:

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

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

References:

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

NEW QUESTION 8

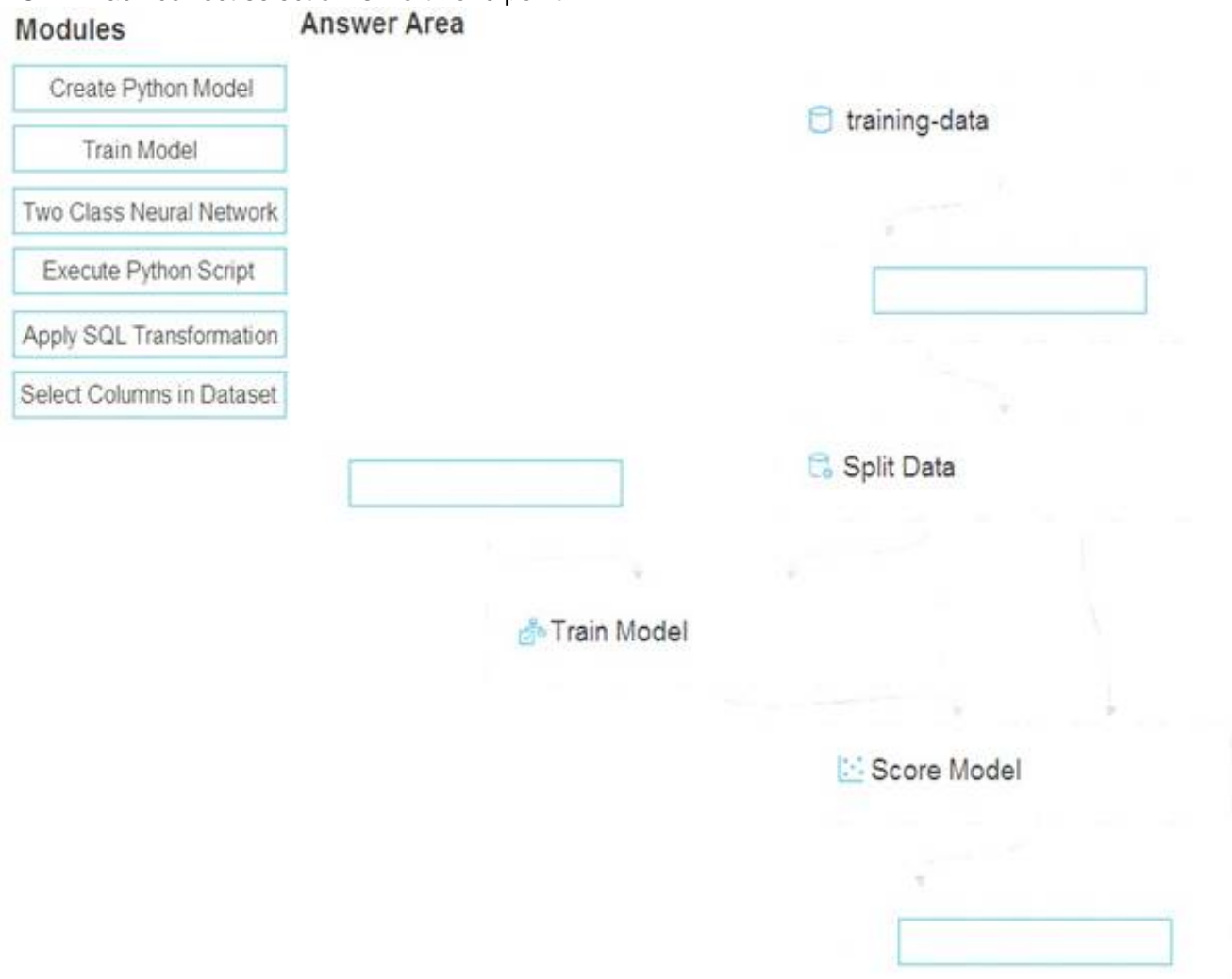
- (Exam Topic 3)

You create a training pipeline using the Azure Machine Learning designer. You upload a CSV file that contains the data from which you want to train your model.

You need to use the designer to create a pipeline that includes steps to perform the following tasks:

- > Select the training features using the pandas filter method.
- > Train a model based on the naive_bayes.GaussianNB algorithm.

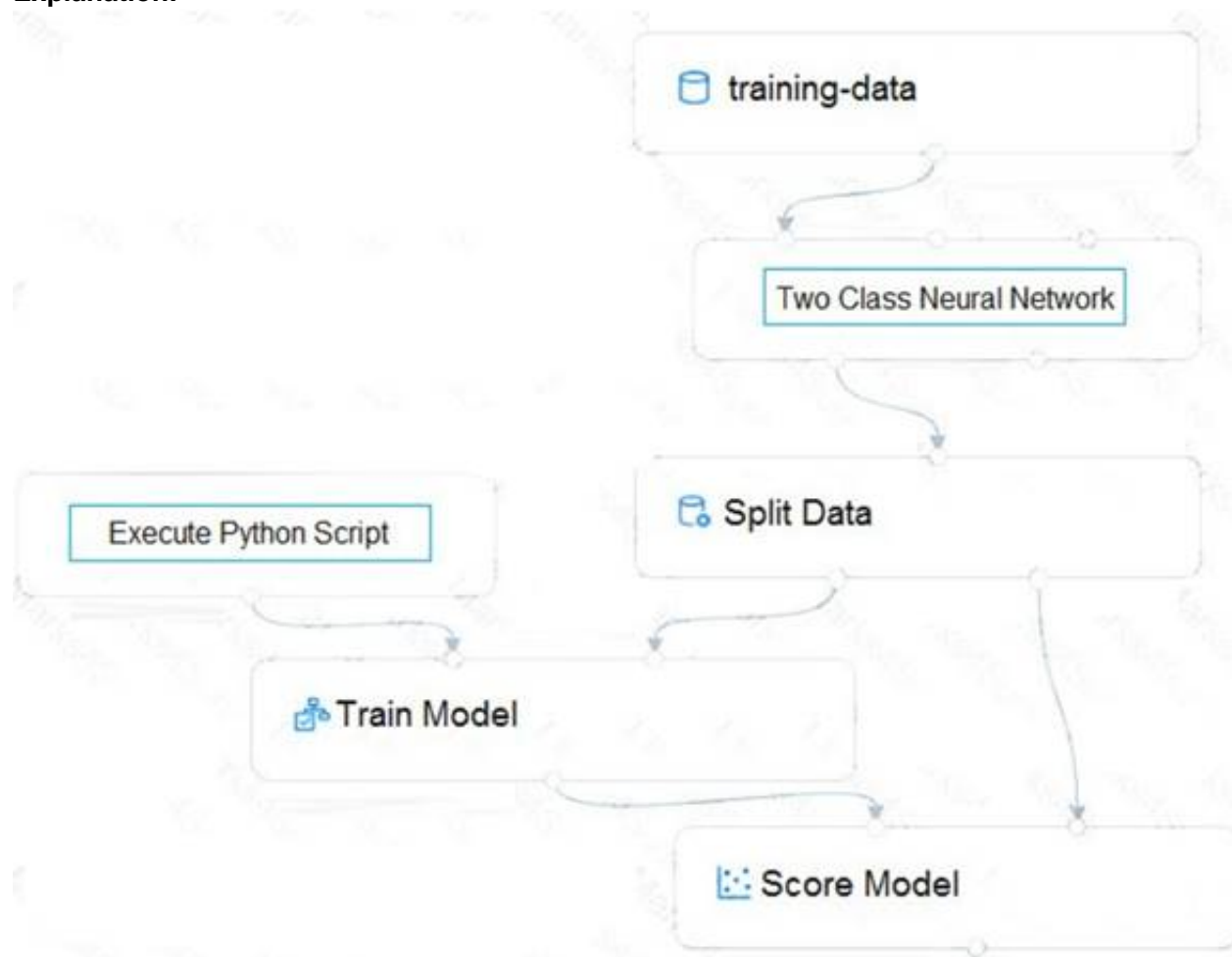
> Return only the Scored Labels column by using the query SELECT [Scored Labels] FROM t1; Which modules should you use? To answer, drag the appropriate modules to the appropriate locations. Each module name may be used once, more than once, or not at all. You may need to drag the split bar between panes or scroll to view content. NOTE: Each correct selection is worth one point.



- A. Mastered
- B. Not Mastered

Answer: A

Explanation:



NEW QUESTION 9

- (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 10

- (Exam Topic 3)

You plan to use automated machine learning to train a regression model. You have data that has features which have missing values, and categorical features with few distinct values.

You need to configure automated machine learning to automatically impute missing values and encode categorical features as part of the training task.

Which parameter and value pair should you use in the AutoMLConfig class?

- A. featurization = 'auto'
- B. enable_voting_ensemble = True
- C. task = 'classification'
- D. exclude_nan_labels = True
- E. enable_tf = True

Answer: A

Explanation:

Featurization str or FeaturizationConfig Values: 'auto' / 'off' / FeaturizationConfig

Indicator for whether featurization step should be done automatically or not, or whether customized featurization should be used.

Column type is automatically detected. Based on the detected column type preprocessing/featurization is done as follows:

Categorical: Target encoding, one hot encoding, drop high cardinality categories, impute missing values. Numeric: Impute missing values, cluster distance, weight of evidence.

DateTime: Several features such as day, seconds, minutes, hours etc. Text: Bag of words, pre-trained Word embedding, text target encoding. Reference:

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

NEW QUESTION 10

- (Exam Topic 3)

You are developing a deep learning model by using TensorFlow. You plan to run the model training workload on an Azure Machine Learning Compute Instance.

You must use CUDA-based model training. You need to provision the Compute Instance.

Which two virtual machines sizes can you use? To answer, select the appropriate virtual machine sizes in the answer area.

NOTE: Each correct selection is worth one point.

Virtual machine size

Name ↑	vCPUs	GPUs	RAM	Resource disk
BASIC_A0	1		0.75 GB	20 GB
STANDARD_D3_V2	4		14 GB	200 GB
STANDARD_E64_V3	64		432 GB	1,600 GB
STANDARD_M64LS	64		512 GB	2,000 GB
STANDARD_NC12	12	2	112 GB	680 GB
STANDARD_NC24	24	4	224 GB	1,440 GB

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

CUDA is a parallel computing platform and programming model developed by Nvidia for general computing on its own GPUs (graphics processing units). CUDA enables developers to speed up compute-intensive applications by harnessing the power of GPUs for the parallelizable part of the computation.

Reference:

<https://www.infoworld.com/article/3299703/what-is-cuda-parallel-programming-for-gpus.html>

NEW QUESTION 15

- (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 18

- (Exam Topic 3)

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

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

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

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

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

Does the solution meet the goal?

- A. Yes
- B. No

Answer: B

Explanation:

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

References:

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

NEW QUESTION 23

- (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	<div><div></div><div><div>the Microsoft Surface device</div><div>the ds-workstation notebook VM</div><div>the cpu-cluster compute target</div><div>the gpu-cluster compute target</div></div></div>
Run the training script	<div><div></div><div><div>the ds-workstation notebook VM</div><div>the cpu-compute target</div><div>the gpu-compute target</div><div>the Microsoft Surface device</div></div></div>

NEW QUESTION 27

- (Exam Topic 3)

You have a dataset that contains over 150 features. You use the dataset to train a Support Vector Machine (SVM) binary classifier. You need to use the Permutation Feature Importance module in Azure Machine Learning Studio to compute a set of feature importance scores for the dataset. In which order should you perform the actions? To answer, move all actions from the list of actions to the answer area and arrange them in the correct order.

Actions

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

Set the Metric for measuring performance property to **Classification - Accuracy** and then run the experiment.

Add a Permutation Feature Importance module and connect the trained model and test dataset.

Add a dataset to the experiment.

Add a Split Data module to create training and test datasets.

Answer Area

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

Step 1: Add a Two-Class Support Vector Machine module to initialize the SVM classifier.
Step 2: Add a dataset to the experiment
Step 3: Add a Split Data module to create training and test dataset.
To generate a set of feature scores requires that you have an already trained model, as well as a test dataset. Step 4: Add a Permutation Feature Importance module and connect to the trained model and test dataset. Step 5: Set the Metric for measuring performance property to Classification - Accuracy and then run the experiment.
Reference:
<https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/two-class-support-vector-mac> <https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/permutation-feature-importan>

NEW QUESTION 31

- (Exam Topic 3)

You use the designer to create a training pipeline for a classification model. The pipeline uses a dataset that includes the features and labels required for model training.

You create a real-time inference pipeline from the training pipeline. You observe that the schema for the generated web service input is based on the dataset and includes the label column that the model predicts. Client applications that use the service must not be required to submit this value. You need to modify the inference pipeline to meet the requirement. What should you do?

- A. Add a Select Columns in Dataset module to the inference pipeline after the dataset and use it to select all columns other than the label.
- B. Delete the dataset from the training pipeline and recreate the real-time inference pipeline.
- C. Delete the Web Service Input module from the inference pipeline.
- D. Replace the dataset in the inference pipeline with an Enter Data Manually module that includes data for the feature columns but not the label column.

Answer: A

Explanation:

By default, the Web Service Input will expect the same data schema as the module output data which connects to the same downstream port as it. You can remove the target variable column in the inference pipeline using Select Columns in Dataset module. Make sure that the output of Select Columns in Dataset removing target variable column is connected to the same port as the output of the Web Service Input module.

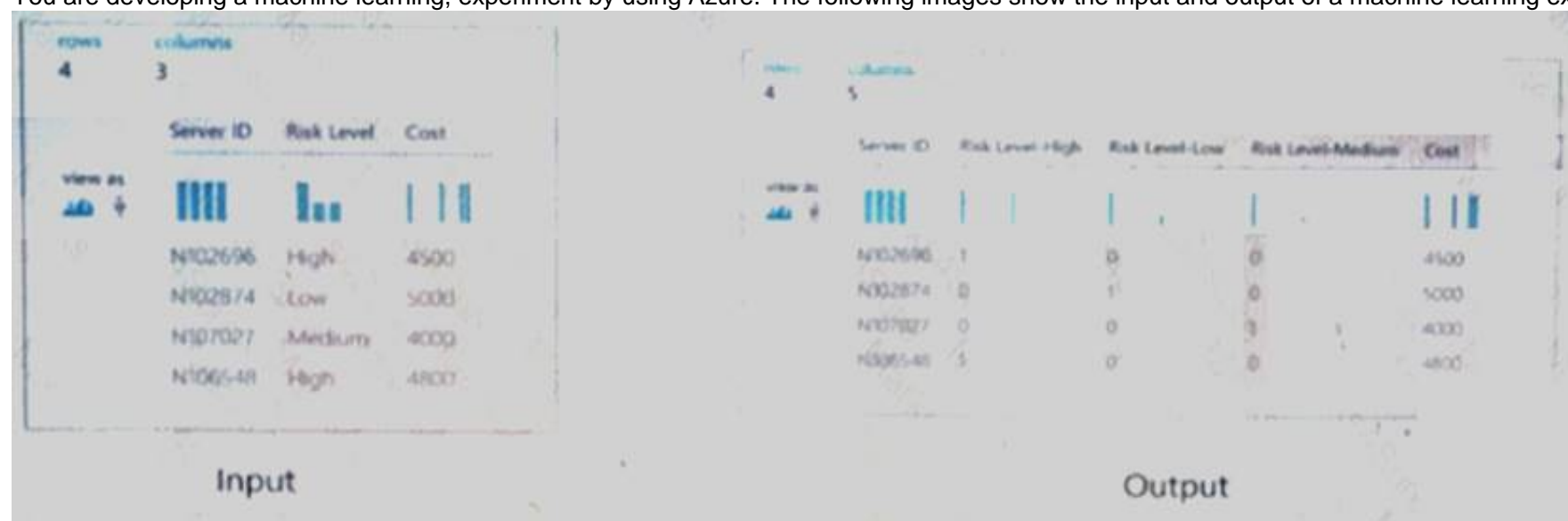
Reference:

https://docs.microsoft.com/en-us/azure/machine-learning/tutorial-designer-automobile-price-deploy

NEW QUESTION 32

- (Exam Topic 3)

You are developing a machine learning, experiment by using Azure. The following images show the input and output of a machine learning experiment:



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.

You need to perform the data transformation applied to the Risk Level column. Which module should you use?

What is the expected input column type for this transformation?

Apply Filter

Build Counting Transform

Convert to Indicator Values

Categorical

Numerical

String

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

You need to perform the data transformation applied to the Risk Level column. Which module should you use?

What is the expected input column type for this transformation?

Apply Filter

Build Counting Transform

Convert to Indicator Values

Categorical

Numerical

String

NEW QUESTION 37

- (Exam Topic 3)

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

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

You 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 38

- (Exam Topic 3)

A set of CSV files contains sales records. All the CSV files have the same data schema.

Each CSV file contains the sales record for a particular month and has the filename `sales.csv`. Each file is stored in a folder that indicates the month and year when the data was recorded. The folders are in an Azure blob container for which a datastore has been defined in an Azure Machine Learning workspace. The folders are organized in a parent folder named `sales` to create the following hierarchical structure:

```
/sales
  /01-2019
    /sales.csv
  /02-2019
    /sales.csv
  /03-2019
    /sales.csv
  ...
```

At the end of each month, a new folder with that month's sales file is added to the `sales` folder.

You plan to use the sales data to train a machine learning model based on the following requirements:

- You must define a dataset that loads all of the sales data to date into a structure that can be easily converted to a dataframe.
- You must be able to create experiments that use only data that was created before a specific previous month, ignoring any data that was added after that month.
- You must register the minimum number of datasets possible.

You need to register the sales data as a dataset in Azure Machine Learning service workspace. What should you do?

- A. Create a tabular dataset that references the datastore and explicitly specifies each `'sales/mm-yyyy/ sales.csv'` file every month
- B. Register the dataset with the name `sales_dataset` each month, replacing the existing dataset and specifying a tag named `month` indicating the month and year it was registered
- C. Use this dataset for all experiments.
- D. Create a tabular dataset that references the datastore and specifies the path `'sales/*/sales.csv'`, register the dataset with the name `sales_dataset` and a tag named `month` indicating the month and year it was registered, and use this dataset for all experiments.
- E. Create a new tabular dataset that references the datastore and explicitly specifies each `'sales/mm-yyyy/ sales.csv'` file every month
- F. Register the dataset with the name `sales_dataset_MM-YYYY` each month with appropriate `MM` and `YYYY` values for the month and year
- G. Use the appropriate month-specific dataset for experiments.
- H. Create a tabular dataset that references the datastore and explicitly specifies each `'sales/mm-yyyy/ sales.csv'` file
- I. Register the dataset with the name `sales_dataset` each month as a new version and with a tag named `month` indicating the month and year it was registered
- J. Use this dataset for all experiments, identifying the version to be used based on the month tag as necessary.

Answer: B

Explanation:

Specify the path. Example:

The following code gets the workspace existing workspace and the desired datastore by name. And then passes the datastore and file locations to the `path` parameter to create a new `TabularDataset`, `weather_ds`.

```
from azureml.core import Workspace, Datastore, Dataset
datastore_name = 'your datastore name'
```



```
# get existing workspace
workspace = Workspace.from_config()
# retrieve an existing datastore in the workspace by name
datastore = Datastore.get(workspace, datastore_name)
# create a TabularDataset from 3 file paths in datastore
datastore_paths = [(datastore, 'weather/2018/11.csv'), (datastore, 'weather/2018/12.csv'),
(datastore, 'weather/2019/*.csv')]
weather_ds = Dataset.Tabular.from_delimited_files(path=datastore_paths)
```

NEW QUESTION 43

- (Exam Topic 3)

You retrain an existing model.

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

What should you do?

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

Answer: B

Explanation:

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

Reference:

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

NEW QUESTION 48

- (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 this 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 51

- (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 56

- (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 58

- (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.MachineLearnin"
]
```

Reference:

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

NEW QUESTION 61

- (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 65

- (Exam Topic 3)

You deploy a model in Azure Container Instance.

You must use the Azure Machine Learning SDK to call the model API.

You need to invoke the deployed model using native SDK classes and methods.

How should you complete the command? To answer, select the appropriate options in the answer areas.

NOTE: Each correct selection is worth one point.

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

```
from azureml.core.webservice import requests
from azureml.core.webservice import Webservice
from azureml.core.webservice import LocalWebservice
```

```
import json
ws = Workspace.from_config()
service_name = "mlmodel1-service"
service = Webservice(name=service_name, workspace=ws)
x_new = [[2,101.5,1,24,21], [1,89.7,4,41,21]]
input_json = json.dumps({"data": x_new})
```

```
predictions = service.run(input_json)
predictions = requests.post(service.scoring_uri, input_json)
predictions = service.deserialize(ws, input_json)
```

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

Box 1: from azureml.core.webservice import Webservice

The following code shows how to use the SDK to update the model, environment, and entry script for a web service to Azure Container Instances:

from azureml.core import Environment

from azureml.core.webservice import Webservice

from azureml.core.model import Model, InferenceConfig Box 2: predictions = service.run(input_json)

Example: The following code demonstrates sending data to the service: import json

```
test_sample = json.dumps({'data': [ [1, 2, 3, 4, 5, 6, 7, 8, 9, 10],
[10, 9, 8, 7, 6, 5, 4, 3, 2, 1]
]})
```

```
test_sample = bytes(test_sample, encoding='utf8') prediction = service.run(input_data=test_sample)
```

```
print(prediction) Reference:
```

<https://docs.microsoft.com/bs-latn-ba/azure/machine-learning/how-to-deploy-azure-container-instance> <https://docs.microsoft.com/en-us/azure/machine-learning/how-to-troubleshoot-deployment>

NEW QUESTION 66

- (Exam Topic 3)

You are conducting feature engineering to prepuce data for further analysis. The data includes seasonal patterns on inventory requirements.

You need to select the appropriate method to conduct feature engineering on the data. Which method should you use?

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

Answer: D

NEW QUESTION 67

- (Exam Topic 3)

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

You must tune hyperparameters by performing a parameter sweep of the model. The parameter sweep must meet the following requirements:

- iterate all possible combinations of hyperparameters
- minimize computing resources required to perform the sweep
- You need to perform a parameter sweep of the model.

Which parameter sweep mode should you use?

- A. Random sweep
- B. Sweep clustering
- C. Entire grid
- D. Random grid
- E. Random seed

Answer: D

Explanation:

Maximum number of runs on random grid: This option also controls the number of iterations over a random sampling of parameter values, but the values are not generated randomly from the specified range; instead, a matrix is created of all possible combinations of parameter values and a random sampling is taken over the matrix. This method is more efficient and less prone to regional oversampling or undersampling.

If you are training a model that supports an integrated parameter sweep, you can also set a range of seed values to use and iterate over the random seeds as well. This is optional, but can be useful for avoiding bias introduced by seed selection.

NEW QUESTION 68

- (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 71

- (Exam Topic 3)

You are building a binary classification model by using a supplied training set. The training set is imbalanced between two classes.

You need to resolve the data imbalance.

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. Penalize the classification
- B. Resample the data set using under sampling or oversampling

- C. Generate synthetic samples in the minority class.
- D. Use accuracy as the evaluation metric of the model.
- E. Normalize the training feature set.

Answer: ABD

Explanation:

References:

<https://machinelearningmastery.com/tactics-to-combat-imbalanced-classes-in-your-machine-learning-dataset/>

NEW QUESTION 74

- (Exam Topic 3)

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

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

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

- A. run.upload_file('row_count', './data.csv')
- B. run.log('row_count', rows)
- C. run.tag('row_count', rows)
- D. run.log_table('row_count', rows)
- E. run.log_row('row_count', rows)

Answer: B

Explanation:

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

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

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

NEW QUESTION 77

- (Exam Topic 3)

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

You must deploy the model into production as a real-time web service to an inference cluster named service-compute that the IT department has created in the Azure Machine Learning workspace.

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

You need to write a script that uses the Azure Machine Learning SDK to deploy the model. The necessary modules have been imported.

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

NOTE: Each correct selection is worth one point.

```
# Assume the necessary modules have been imported
deploy_target = ▼ (ws, "service-compute")
AksCompute
AmlCompute
RemoteCompute
BatchCompute

deployment_config = ▼ .deploy_configuration(cpu_cores=1, memory_gb=1,
AksWebservice
AciWebservice
LocalWebService

▼ )
token_auth_enabled=True
token_auth_enabled=False
auth_enabled=True
auth_enabled=False

service = Model.deploy(ws, "ml-service",
    [model], inference_config, deployment_config, deploy_target)
service.wait_for_deployment(show_output = True)
```

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

Box 1: AksCompute Example:

```
aks_target = AksCompute(ws,"myaks")
```

If deploying to a cluster configured for dev/test, ensure that it was created with enough
 # cores and memory to handle this deployment configuration. Note that memory is also used by
 # things such as dependencies and AML components.

```
deployment_config = AksWebService.deploy_configuration(cpu_cores = 1, memory_gb = 1)
```

```
service = Model.deploy(ws, "myservice", [model], inference_config, deployment_config, aks_target)
```

Box 3: token_auth_enabled=Yes

Whether or not token auth is enabled for the Webservice.

Note: A Service principal defined in Azure Active Directory (Azure AD) can act as a principal on which authentication and authorization policies can be enforced in Azure Databricks.

The Azure Active Directory Authentication Library (ADAL) can be used to programmatically get an Azure AD access token for a user.

Reference:

<https://docs.microsoft.com/en-us/azure/machine-learning/how-to-deploy-azure-kubernetes-service> <https://docs.microsoft.com/en-us/azure/databricks/dev-tools/api/latest/aad/service-prin-aad-token>

NEW QUESTION 82

- (Exam Topic 3)

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

```
Datastore.register_azure_blob_container(workspace=ws,
    datastore_name='demo_datastore',
    container_name='demo_datacontainer',
    account_name='demo_account',
    account_key='0A0A0A-0A0A00A-0A00A0A0A0A0A',
    create_if_not_exists=True)
```

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

Answer Area

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

datastore = Workspace.get(ws, 'demo_datastore')
```

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

Box 1: DataStore

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

Get a named datastore from the current workspace

```
datastore = Datastore.get(ws, datastore_name='your datastore name')
```

Box 2: ws

Box 3: demo_datastore Reference:

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

NEW QUESTION 87

- (Exam Topic 3)

You are building a machine learning model for translating English language textual content into French language textual content.

You need to build and train the machine learning model to learn the sequence of the textual content. Which type of neural network should you use?

- A. Multilayer Perceptions (MLPs)
- B. Convolutional Neural Networks (CNNs)
- C. Recurrent Neural Networks (RNNs)
- D. Generative Adversarial Networks (GANs)

Answer: C

Explanation:

To translate a corpus of English text to French, we need to build a recurrent neural network (RNN).

Note: RNNs are designed to take sequences of text as inputs or return sequences of text as outputs, or both. They're called recurrent because the network's hidden layers have a loop in which the output and cell state from each time step become inputs at the next time step. This recurrence serves as a form of memory. It allows contextual information to flow through the network so that relevant outputs from previous time steps can be applied to network operations at the current time step.

References:

<https://towardsdatascience.com/language-translation-with-rnns-d84d43b40571>

NEW QUESTION 88

- (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 90

- (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 91

- (Exam Topic 3)

You create an Azure Machine Learning workspace named ML-workspace. You also create an Azure Databricks workspace named DB-workspace. DB-workspace contains a cluster named DB-cluster.

You must use DB-cluster to run experiments from notebooks that you import into DB-workspace.

You need to use ML-workspace to track MLflow metrics and artifacts generated by experiments running on DB-cluster. The solution must minimize the need for custom code.

What should you do?

- A. From DB-cluster, configure the Advanced Logging option.
- B. From DB-workspace
- C. configure the Link Azure ML workspace option.
- D. From ML-workspace
- E. create an attached compute.
- F. From ML-workspace
- G. create a compute cluster.

Answer: B

Explanation:

Reference:

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

NEW QUESTION 92

- (Exam Topic 3)

You are using a Git repository to track work in an Azure Machine Learning workspace. You need to authenticate a Git account by using SSH.

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

Actions	Answer Area
Generate a public/private key pair	
Add the private key to the Git account	
Clone the Git repository by using an SSH repository URL	
Add the public key to the Git account	
Create a new Azure Key Vault resource	

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

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

Authenticate your Git Account with SSH: Step 1: Generating a public/private key pair Generate a new SSH key

* 1. Open the terminal window in the Azure Machine Learning Notebook Tab.

* 2. Paste the text below, substituting in your email address. `ssh-keygen -t rsa -b 4096 -C "your_email@example.com"`

This creates a new ssh key, using the provided email as a label.

> Generating public/private rsa key pair.

Step 2: Add the public key to the Git Account

In your terminal window, copy the contents of your public key file. Step 3: Clone the Git repository by using an SSH repository URL

* 1. Copy the SSH Git clone URL from the Git repo.

* 2. Paste the url into the git clone command below, to use your SSH Git repo URL. This will look something like:

`git clone git@example.com:GitUser/azureml-example.git` Cloning into 'azureml-example'.

Reference:

<https://docs.microsoft.com/en-us/azure/machine-learning/concept-train-model-git-integration>

NEW QUESTION 97

- (Exam Topic 3)

You plan to run a Python script as an Azure Machine Learning experiment.

The script must read files from a hierarchy of folders. The files will be passed to the script as a dataset argument.

You must specify an appropriate mode for the dataset argument.

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

NOTE: Each correct selection is worth one point.

- A. to_pandas_dataframe ()
- B. as_download()
- C. as_upload()
- D. as mount ()

Answer: B

Explanation:

Reference:

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

NEW QUESTION 101

- (Exam Topic 3)

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

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

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

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

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

Solution: Run the following code:

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

Does the solution meet the goal?

- A. Yes
- B. No

Answer: B

Explanation:

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

Example:

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

Reference:

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

NEW QUESTION 106

- (Exam Topic 3)

You register a file dataset named csvfolder that references a folder. The folder includes multiple com

ma-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)

```
inputs=[file_dataset.as_named_input('training_files').to_pandas_dataframe()],
```

B)

```
inputs=[file_dataset.as_named_input('training_files').as_mount()],
```

C)

```
script_params={'--training_files': file_dataset},
```

D)

```
inputs=[file_dataset.as_named_input('training_files')],
```

- 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 108

- (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 109

- (Exam Topic 3)

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

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

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

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

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

- A. Yes
- B. No

Answer: A

Explanation:

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

References:

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

NEW QUESTION 113

- (Exam Topic 3)

You create a multi-class image classification deep learning model.

The model must be retrained monthly with the new image data fetched from a public web portal. You create an Azure Machine Learning pipeline to fetch new data, standardize the size of images, and retrain the model.

You need to use the Azure Machine Learning SDK to configure the schedule for the pipeline.

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

correct order.

Actions	Answer Area
Publish the pipeline.	
Retrieve the pipeline ID.	
Create a ScheduleRecurrence(frequency= 'Month', interval=1, start_time='2019-01-01T00:00:00') object.	⬅
Define a pipeline parameter named RunDate .	➡
Define a new Azure Machine Learning pipeline StepRun object with the step ID of the first step in the pipeline.	⬆
Define an Azure Machine Learning pipeline schedule using the schedule.create method with the defined recurrence specification.	⬇

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

Step 1: Publish the pipeline.

To schedule a pipeline, you'll need a reference to your workspace, the identifier of your published pipeline, and the name of the experiment in which you wish to create the schedule.

Step 2: Retrieve the pipeline ID. Needed for the schedule.

Step 3: Create a ScheduleRecurrence..

To run a pipeline on a recurring basis, you'll create a schedule. A Schedule associates a pipeline, an experiment, and a trigger.

First create a schedule. Example: Create a Schedule that begins a run every 15 minutes: recurrence = ScheduleRecurrence(frequency="Minute", interval=15)

Step 4: Define an Azure Machine Learning pipeline schedule.. Example, continued:

```
recurring_schedule = Schedule.create(ws, name="MyRecurringSchedule", description="Based on time", pipeline_id=pipeline_id, experiment_name=experiment_name, recurrence=recurrence)
```

Reference:

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

NEW QUESTION 114

- (Exam Topic 3)

You need to select a pre built development environment for a series of data science experiments. You must use the R language for the experiments. Which three environments can you use? Each correct answer presents a complete solution. NOTE: Each correct selection is worth one point.

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

Answer: ABD

NEW QUESTION 116

- (Exam Topic 3)

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


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

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

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

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

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

Answer Area

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

- A. Mastered
 B. Not Mastered

Answer: A

Explanation:

Box 1: No

max_total_runs (50 here)

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

Box 2: Yes

Policy EarlyTerminationPolicy

The early termination policy to use. If None - the default, no early termination policy will be used. Box 3: No

Discrete hyperparameters are specified as a choice among discrete values. choice can be: one or more comma-separated values

> a range object

> any arbitrary list object Reference:

<https://docs.microsoft.com/en-us/python/api/azureml-train-core/azureml.train.hyperdrive.hyperdriveconfig> <https://docs.microsoft.com/en-us/azure/machine-learning/how-to-tune-hyperparameters>

NEW QUESTION 119

- (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 122

- (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 124

- (Exam Topic 3)

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

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

NOTE: Each correct selection is worth one point.

datastore = Datastore.

register_azure_blob_container
register_azure_file_share
register_azure_data_lake
register_azure_sql_database

(workspace = ws,

datastore_name = 'finance_datastore',

container_name = 'finance-data',

account_name = 'fintrainingdatastorage',

account_key = 'FWUYORRv3XoyNe...',

create_if_not_exists = True
create_if_not_exists = False
overwrite = True
overwrite = False

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

Box 1: register_azure_blob_container

Register an Azure Blob Container to the datastore.

Box 2: create_if_not_exists = False

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

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

NEW QUESTION 129

- (Exam Topic 3)

You create a machine learning model by using the Azure Machine Learning designer. You publish the model as a real-time service on an Azure Kubernetes Service (AKS) inference compute cluster. You make no changes to the deployed endpoint configuration.

You need to provide application developers with the information they need to consume the endpoint.

Which two values should you provide to application developers? Each correct answer presents part of the solution.

NOTE: Each correct selection is worth one point.

- A. The name of the AKS cluster where the endpoint is hosted.
- B. The name of the inference pipeline for the endpoint.
- C. The URL of the endpoint.
- D. The run ID of the inference pipeline experiment for the endpoint.
- E. The key for the endpoint.

Answer: CE

Explanation:

Deploying an Azure Machine Learning model as a web service creates a REST API endpoint. You can send data to this endpoint and receive the prediction returned by the model.

You create a web service when you deploy a model to your local environment, Azure Container Instances, Azure Kubernetes Service, or field-programmable gate arrays (FPGA). You retrieve the URI used to access the web service by using the Azure Machine Learning SDK. If authentication is enabled, you can also use the SDK to get the authentication keys or tokens.

Example:

URL for the web service
 scoring_uri = '<your web service URI>
 # If the service is authenticated, set the key or token key = '<your key or token>
 Reference:
<https://docs.microsoft.com/en-us/azure/machine-learning/how-to-consume-web-service>

NEW QUESTION 130

- (Exam Topic 3)

You train and register a model by using the Azure Machine Learning SDK on a local workstation. Python 3.6 and Visual Studio Code are installed on the workstation.

When you try to deploy the model into production as an Azure Kubernetes Service (AKS)-based web service, you experience an error in the scoring script that causes deployment to fail.

You need to debug the service on the local workstation before deploying the service to production.

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

Actions	Answer Area
Create an AksWeb service deployment configuration for the service and deploy the model to it	
Install Docker on the workstation	
Create a LocalWeb service deployment configuration for the service and deploy the model to it	
Debug and modify the scoring script as necessary. Use the reload() method of the service after each modification	
Create an AciWeb service deployment configuration for the service and deploy the model to it	

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

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

Step 1: Install Docker on the workstation

Prerequisites include having a working Docker installation on your local system. Build or download the dockerfile to the compute node.

Step 2: Create an AksWeb service deployment configuration and deploy the model to it

To deploy a model to Azure Kubernetes Service, create a deployment configuration that describes the compute resources needed.

If deploying to a cluster configured for dev/test, ensure that it was created with enough

cores and memory to handle this deployment configuration. Note that memory is also used by

things such as dependencies and AML components.

deployment_config = AksWeb service.deploy_configuration(cpu_cores = 1, memory_gb = 1)

service = Model.deploy(ws, "myservice", [model], inference_config, deployment_config, aks_target) service.wait_for_deployment(show_output = True)

print(service.state) print(service.get_logs())

Step 3: Create a LocalWeb service deployment configuration for the service and deploy the model to it

To deploy locally, modify your code to use LocalWeb service.deploy_configuration() to create a deployment configuration. Then use Model.deploy() to deploy the service.

Step 4: Debug and modify the scoring script as necessary. Use the reload() method of the service after each modification.

During local testing, you may need to update the score.py file to add logging or attempt to resolve any problems that you've discovered. To reload changes to the score.py file, use reload(). For example, the following code reloads the script for the service, and then sends data to it.

Reference:

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

NEW QUESTION 135

- (Exam Topic 3)

You are building a recurrent neural network to perform a binary classification. You review the training loss, validation loss, training accuracy, and validation accuracy for each training epoch.

You need to analyze model performance.

Which observation indicates that the classification model is over fitted?

- A. 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.
- B. The training loss increases while the validation loss decreases when training the model.
- C. The training loss decreases while the validation loss increases when training the model.
- D. The training loss stays constant and the validation loss decreases when training the model.

Answer: B

NEW QUESTION 136

- (Exam Topic 3)

HOTSPOT

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

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

You run the following code.

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

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

NOTE: Each correct selection is worth one point.

	Yes	No
The myDataset_1 dataset can be converted into a pandas dataframe by using the following method: using myDataset_1.to_pandas_dataframe()	<input type="radio"/>	<input type="radio"/>
The myDataset_1.to_path() method returns an array of file paths for all of the TSV files in the dataset.	<input type="radio"/>	<input type="radio"/>
The myDataset_2 dataset can be converted into a pandas dataframe by using the following method: myDataset_2.to_pandas_dataframe()	<input type="radio"/>	<input type="radio"/>

A. Mastered

B. Not Mastered

Answer: A

Explanation:

Box 1: No

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

Box 2: Yes

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

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

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

Reference:

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

NEW QUESTION 141

- (Exam Topic 3)

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

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

NOTE: Each correct selection is worth one point. D

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

Answer: ABD

Explanation:

Reference:

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

NEW QUESTION 142

- (Exam Topic 3)

You are developing a hands-on workshop to introduce Docker for Windows to attendees. You need to ensure that workshop attendees can install Docker on their devices.

Which two prerequisite components should attendees install on the devices? Each correct answer presents part of the solution.
 NOTE: Each correct selection is worth one point.

- A. Microsoft Hardware-Assisted Virtualization Detection Tool
- B. Kitematic
- C. BIOS-enabled virtualization
- D. VirtualBox
- E. Windows 10 64-bit Professional

Answer: CE

Explanation:

C: Make sure your Windows system supports Hardware Virtualization Technology and that virtualization is enabled. Ensure that hardware virtualization support is turned on in the BIOS settings. For example:



E: To run Docker, your machine must have a 64-bit operating system running Windows 7 or higher. References:
https://docs.docker.com/toolbox/toolbox_install_windows/ <https://blogs.technet.microsoft.com/canitpro/2015/09/08/step-by-step-enabling-hyper-v-for-use-on-windows-10/>

NEW QUESTION 144

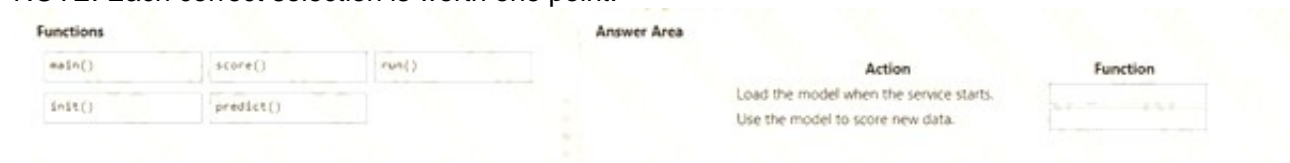
- (Exam Topic 3)

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

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

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

NOTE: Each correct selection is worth one point.



- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

Box 1: init()

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

Box 2: run() Reference:

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

NEW QUESTION 146

- (Exam Topic 3)

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

You must predict the fare of a taxi trip.

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

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

Answer: BE

Explanation:

References:

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

NEW QUESTION 147

- (Exam Topic 3)

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

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

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

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

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

Solution: Install the Azure ML SDK on the Surface Book. Run Python code to connect to the workspace and then run the training script as an experiment on local compute.

A. Yes

B. No

Answer: B

Explanation:

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

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

NEW QUESTION 149

- (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 151

- (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 155

- (Exam Topic 3)

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

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

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

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

Solution: Run the following code:

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

Does the solution meet the goal?

- A. Yes
- B. No

Answer: A

Explanation:

The two steps are present: process_step and train_step Note:

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

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

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

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

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

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

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

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

NEW QUESTION 156

- (Exam Topic 3)

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

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

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

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

The IT department creates an Azure Kubernetes Service (AKS)-based inference compute target named aks-cluster in the Azure Machine Learning workspace.

You have a Microsoft Surface Book computer with a GPU. Python 3.6 and Visual Studio Code are installed. You need to run a script that trains a deep neural network (DNN) model and logs the loss and accuracy metrics.

Solution: Attach the mlvm virtual machine as a compute target in the Azure Machine Learning workspace. Install the Azure ML SDK on the Surface Book and run Python code to connect to the workspace. Run the training script as an experiment on the mlvm remote compute resource.

- A. Yes
- B. No

Answer: A

Explanation:

Use the VM as a compute target.

Note: A compute target is a designated compute resource/environment where you run your training script or host your service deployment. This location may be

your local machine or a cloud-based compute resource.

Reference:

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

NEW QUESTION 161

- (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 162

- (Exam Topic 3)

You are creating a new experiment in Azure Machine Learning Studio. You have a small dataset that has missing values in many columns. The data does not require the application of predictors for each column. You plan to use the Clean Missing Data module to handle the missing data.

You need to select a data cleaning method. Which method should you use?

A. Synthetic Minority

B. Replace using Probabilistic PAC

C. Replace using MICE

D. Normalization

Answer: B

NEW QUESTION 167

- (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 172

- (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 177

- (Exam Topic 3)

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

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

Answer Area

	Yes	No
The compute is created in the same region as the Machine Learning service workspace.	<input checked="" type="checkbox"/>	<input type="checkbox"/>
The compute resource created by the code is displayed as a compute cluster in Azure Machine Learning studio.	<input type="checkbox"/>	<input type="checkbox"/>
The minimum number of nodes will be zero.	<input type="checkbox"/>	<input type="checkbox"/>

A. Mastered

B. Not Mastered

Answer: A

Explanation:

Box 1: Yes

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

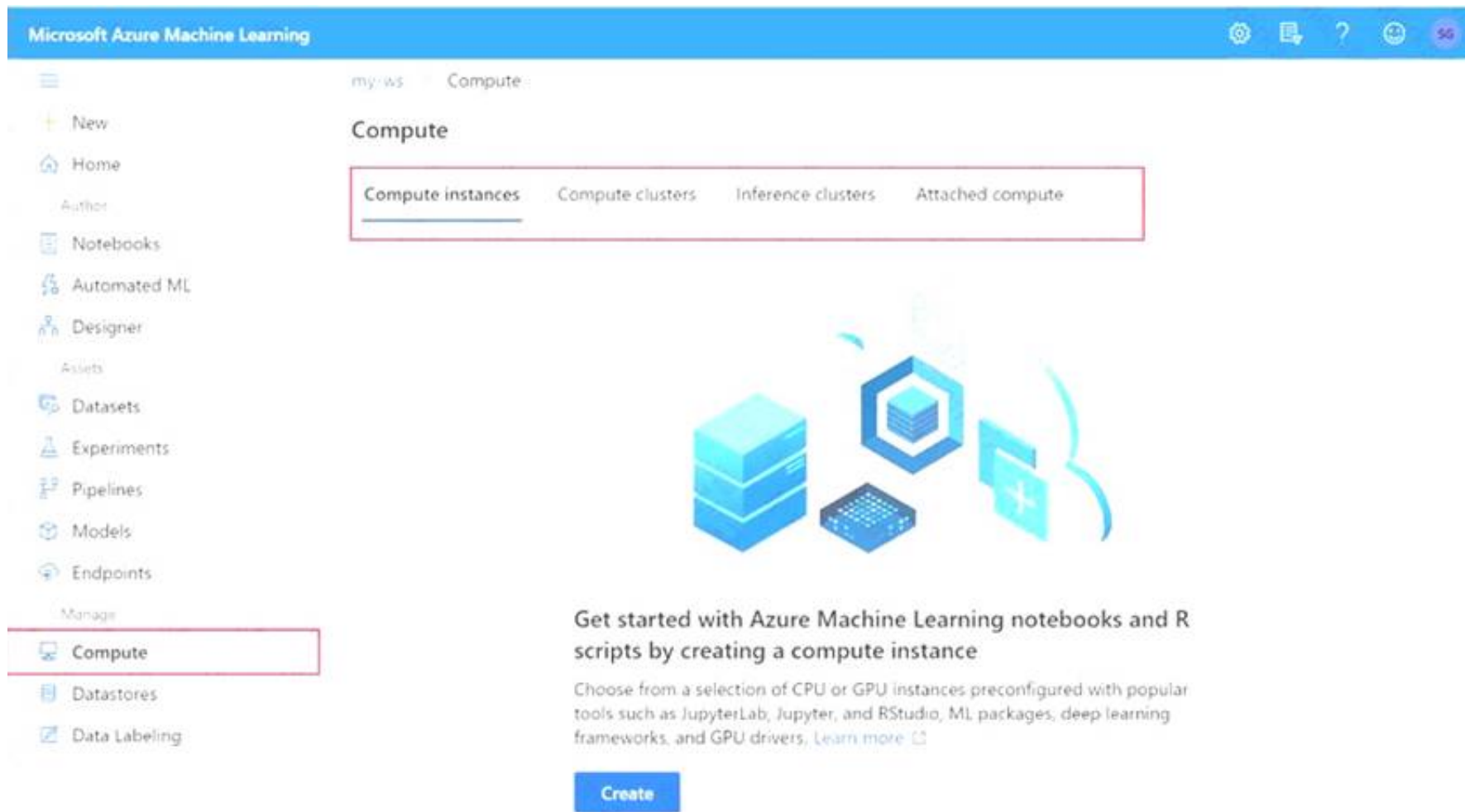
It is displayed as a compute cluster. View compute targets

* 1. To see all compute targets for your workspace, use the following steps:

* 2. Navigate to Azure Machine Learning studio.

* 3. Under Manage, select Compute.

* 4. Select tabs at the top to show each type of compute target.



Box 3: Yes
min_nodes is not specified, so it defaults to 0. Reference:
<https://docs.microsoft.com/en-us/python/api/azureml-core/azureml.core.compute.amlcompute.amlcomputeprovider> <https://docs.microsoft.com/en-us/azure/machine-learning/how-to-create-attach-compute-studio>

NEW QUESTION 178

- (Exam Topic 3)
You have an Azure Machine Learning workspace named workspace1 that is accessible from a public endpoint. The workspace contains an Azure Blob storage datastore named store1 that represents a blob container in an Azure storage account named account1. You configure workspace1 and account1 to be accessible by using private endpoints in the same virtual network.
You must be able to access the contents of store1 by using the Azure Machine Learning SDK for Python. You must be able to preview the contents of store1 by using Azure Machine Learning studio.
You need to configure store1.
What should you do? To answer, select the appropriate options in the answer area.
NOTE: Each correct selection is worth one point.

Requirement	Action
Access the contents of store1 by using the Azure Machine Learning SDK for Python.	<div><div>Set store1 as the default datastore.</div><div>Disable data validation for store1.</div><div>Update authentication for store1.</div><div>Regenerate the keys of account1.</div></div>
Preview the contents of store1 by using Azure Machine Learning studio.	<div><div>Set store1 as the default datastore.</div><div>Disable data validation for store1.</div><div>Update authentication for store1.</div><div>Regenerate the keys of account1.</div></div>

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

Text, table Description automatically generated
Box 1: Regenerate the keys of account1.
Azure Blob Storage support authentication through Account key or SAS token.
To authenticate your access to the underlying storage service, you can provide either your account key, shared access signatures (SAS) tokens, or service principal
Box 2: Update the authentication for store1.
For Azure Machine Learning studio users, several features rely on the ability to read data from a dataset; such as dataset previews, profiles and automated machine learning. For these features to work with storage behind virtual networks, use a workspace managed identity in the studio to allow Azure Machine Learning to access the storage account from outside the virtual network.
Note: Some of the studio's features are disabled by default in a virtual network. To re-enable these features, you must enable managed identity for storage accounts you intend to use in the studio.

The following operations are disabled by default in a virtual network:

➤ Preview data in the studio.

Reference:

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

NEW QUESTION 182

- (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 183

- (Exam Topic 3)

You are building an experiment using the Azure Machine Learning designer.

You split a dataset into training and testing sets. You select the Two-Class Boosted Decision Tree as the algorithm.

You need to determine the Area Under the Curve (AUC) of the model.

Which three 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.

Modules

- Export Data
- Tune Model Hyperparameters
- Cross Validate Model
- Evaluate Model
- Score Model
- Train Model

Answer Area

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

Step 1: Train Model

Two-Class Boosted Decision Tree

First, set up the boosted decision tree model.

* 1. Find the Two-Class Boosted Decision Tree module in the module palette and drag it onto the canvas.

* 2. Find the Train Model module, drag it onto the canvas, and then connect the output of the Two-Class Boosted Decision Tree module to the left input port of the Train Model module.

The Two-Class Boosted Decision Tree module initializes the generic model, and Train Model uses training data to train the model.

* 3. Connect the left output of the left Execute R Script module to the right input port of the Train Model module (in this tutorial you used the data coming from the left side of the Split Data module for training). This portion of the experiment now looks something like this:



Step 2: Score Model

Score and evaluate the models

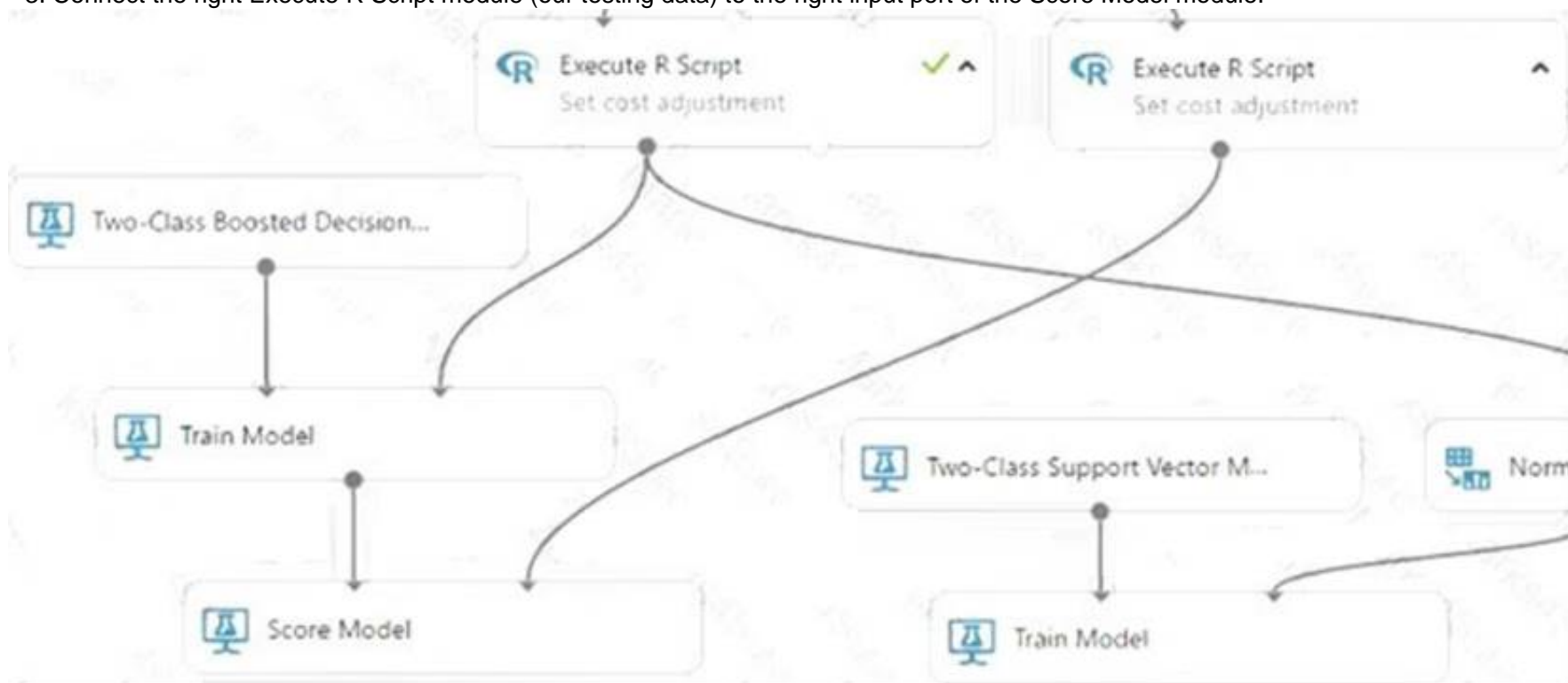
You use the testing data that was separated out by the Split Data module to score our trained models. You can then compare the results of the two models to see which generated better results.

Add the Score Model modules

* 1. Find the Score Model module and drag it onto the canvas.

* 2. Connect the Train Model module that's connected to the Two-Class Boosted Decision Tree module to the left input port of the Score Model module.

* 3. Connect the right Execute R Script module (our testing data) to the right input port of the Score Model module.



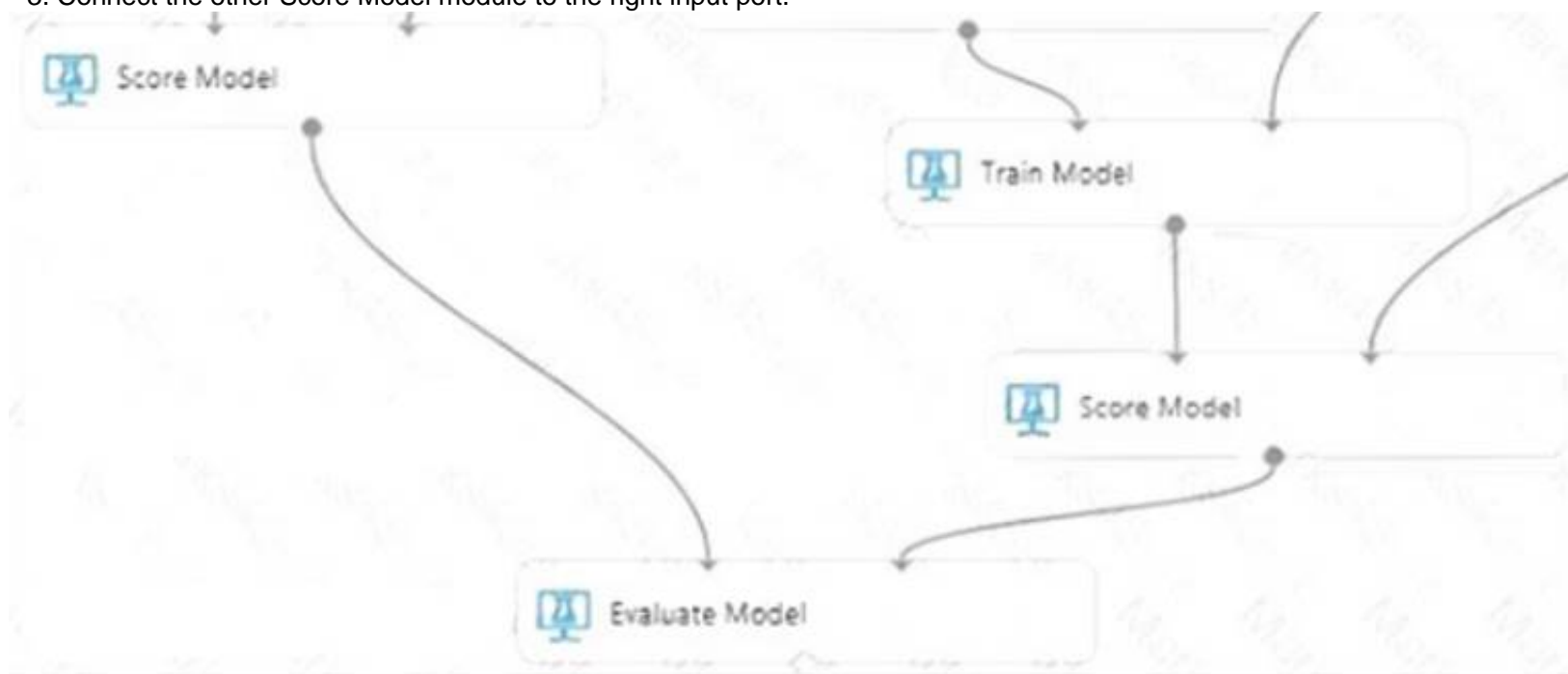
Step 3: Evaluate Model

To evaluate the two scoring results and compare them, you use an Evaluate Model module.

* 1. Find the Evaluate Model module and drag it onto the canvas.

* 2. Connect the output port of the Score Model module associated with the boosted decision tree model to the left input port of the Evaluate Model module.

* 3. Connect the other Score Model module to the right input port.



NEW QUESTION 188

- (Exam Topic 3)

You configure a Deep Learning Virtual Machine for Windows.

You need to recommend tools and frameworks to perform the following:

- Build deep neural network (DNN) models
- Perform interactive data exploration and visualization

Which tools and frameworks should you recommend? To answer, drag the appropriate tools to the correct tasks. Each tool may be used once, more than once, or not at all. You may need to drag the split bar between panes or scroll to view content.

NOTE: Each correct selection is worth one point.

Tools	Task	Tool
Vowpal Wabbit	Build DNN models	Tool
PowerBI Desktop	Enable interactive data exploration and visualization	Tool
Azure Data Factory		
Microsoft Cognitive Toolkit		

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

Box 1: Vowpal Wabbit

Use the Train Vowpal Wabbit Version 8 module in Azure Machine Learning Studio (classic), to create a machine learning model by using Vowpal Wabbit.

Box 2: PowerBI Desktop

Power BI Desktop is a powerful visual data exploration and interactive reporting tool

BI is a name given to a modern approach to business decision making in which users are empowered to find, explore, and share insights from data across the enterprise.

References:

<https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/train-vowpal-wabbit-version-> <https://docs.microsoft.com/en-us/azure/architecture/data-guide/scenarios/interactive-data-exploration>

NEW QUESTION 191

- (Exam Topic 3)

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

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

You are a data scientist using Azure Machine Learning Studio.

You need to normalize values to produce an output column into bins to predict a target column. Solution: Apply an Equal Width with Custom Start and Stop binning mode.

Does the solution meet the goal?

- A. Yes
- B. No

Answer: B

Explanation:

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

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

NEW QUESTION 195

- (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>

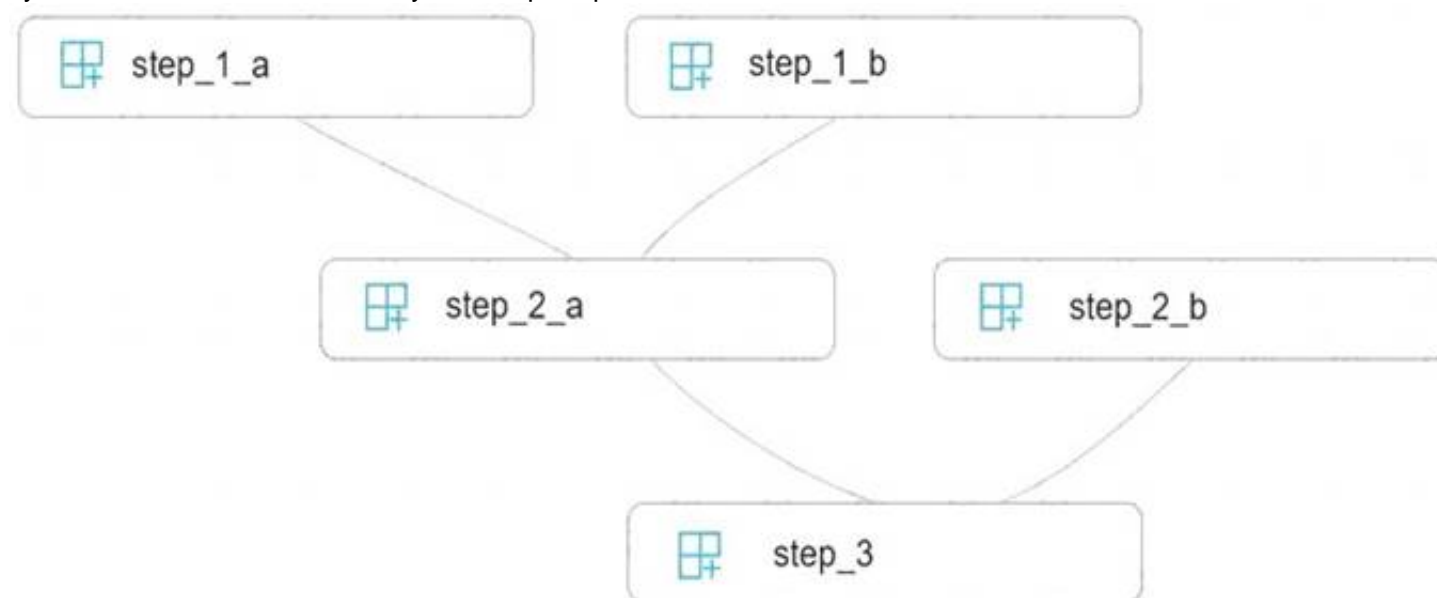
NEW QUESTION 196

- (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 198

- (Exam Topic 3)

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

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

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

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

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

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

NOTE: Each correct selection is worth one point.

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

- A. Mastered
B. Not Mastered

Answer: A

Explanation:

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

Box 3: No

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

Reference:

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

NEW QUESTION 200

- (Exam Topic 3)

You have a model with a large difference between the training and validation error values. You must create a new model and perform cross-validation.

You need to identify a parameter set for the new model using Azure Machine Learning Studio.

Which module you should use for each step? To answer, drag the appropriate modules to the correct steps. Each module may be used once or more than once, or not at all. You may need to drag the split bar between panes or scroll to view content.

NOTE: Each correct selection is worth one point.

Modules	Step	Module
Two-Class Boosted Decision Tree	Define the parameter scope	
Partition and Sample	Define the cross-validation settings	
Tune Model Hyperparameters	Define the metric	
Split Data	Train, evaluate, and compare	

- A. Mastered
B. Not Mastered

Answer: A

Explanation:

Box 1: Split data

Box 2: Partition and Sample

Box 3: Two-Class Boosted Decision Tree Box 4: Tune Model Hyperparameters

Integrated train and tune: You configure a set of parameters to use, and then let the module iterate over multiple combinations, measuring accuracy until it finds a "best" model. With most learner modules, you can choose which parameters should be changed during the training process, and which should remain fixed.

We recommend that you use Cross-Validate Model to establish the goodness of the model given the specified parameters. Use Tune Model Hyperparameters to identify the optimal parameters. References:

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

NEW QUESTION 203

- (Exam Topic 3)

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

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

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

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

Preprocess Text

Language
English

Remove by part of speech
False

Text column to clean

Selected columns:
Column names: **String, Feature**

Launch column selector

☐ Remove stop words

☐ Lemmatization

☐ Detect sentences

☐ Normalize case to lowercase

☐ Remove numbers

☐ Remove special characters

☐ Remove duplicate characters

☐ Remove email addresses

☐ Remove URLs

☐ Expand verb contractions

☐ Normalize backslashes to slashes

☐ Split tokens on special characters

- A. Mastered
B. Not Mastered

Answer: A

Explanation:

Box 1: Remove stop words

Remove words to optimize information retrieval.

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

Box 2: Lemmatization

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

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

References:

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

NEW QUESTION 205

- (Exam Topic 3)

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

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

Each correct answer presents part of the solution. NOTE:

Each correct selection is worth one point.

- A. The label data must be a negative value.
B. The label data can be positive or negative,
C. The label data must be a positive value
D. The label data must be non discrete.
E. The data must be whole numbers.

Answer: CE

Explanation:

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

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

References:

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

NEW QUESTION 209

- (Exam Topic 3)

You are performing a filter based feature selection for a dataset 10 build a multi class classifies by using Azure Machine Learning Studio.

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

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

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

Answer: D

Explanation:

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

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

Reference:

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

NEW QUESTION 211

- (Exam Topic 3)

You are developing a linear regression model in Azure Machine Learning Studio. You run an experiment to compare different algorithms.

The following image displays the results dataset output:

Algorithm	Mean Absolute Error	Root Mean Squared Error	Relative Absolute Error	Relative Squared Error
Bayesian Liner	3.276025	4.655442	0.511436	0.282138
Neural Network	2.676538	3.621476	0.417847	0.17073
Boosted Decision Tree	2.168847	2.878077	0.338589	0.107831
Linear	6.350005	8.720718	0.99133	0.99002
Decision Forest	2.390206	3.315 164	0.373146	0.14307

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

NOTE: Each correct selection is worth one point.

Question

Answer choice

Which algorithm minimizes differences between actual and predicted values?

- Bayesian Linear Regression
- Neural Network Regression
- Boosted Decision Tree Regression
- Linear Regression
- Decision Forest Regression

Which approach should you use to find the best parameters for a Linear Regression model for the Online Gradient Descent method?

- Set the Decrease learning rate option to True.
- Set the Decrease learning rate option to True.
- Set the Create trainer mode option to Parameter Range.
- Increase the number of epochs.
- Decrease the number of epochs.

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

Box 1: Boosted Decision Tree Regression

Mean absolute error (MAE) measures how close the predictions are to the actual outcomes; thus, a lower score is better.

Box 2:

Online Gradient Descent: If you want the algorithm to find the best parameters for you, set Create trainer mode option to Parameter Range. You can then specify multiple values for the algorithm to try.

References:

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

NEW QUESTION 212

- (Exam Topic 3)

You create an Azure Machine Learning compute target named ComputeOne by using the STANDARD_D1 virtual machine image.

You define a Python variable named ws that references the Azure Machine Learning workspace. You run the following Python code:

```
from azureml.core.compute import ComputeTarget, AmlCompute
from azureml.core.compute_target import ComputeTargetException
the_cluster_name = "ComputeOne"
try:
    the_cluster = ComputeTarget(workspace=ws, name=the_cluster_name)
    print('Step1')
except ComputeTargetException:
    config = AmlCompute.provisioning_configuration(vm_size='STANDARD_DS12_V2', max_nodes=4)
    the_cluster = ComputeTarget.create(ws, the_cluster_name, config)
    print('Step2')
```

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

NOTE: Each correct selection is worth one point.

	Yes	No
A new machine learning compute resource is created with a virtual machine size of STANDARD_DS12_V2 and a maximum of four nodes.	<input type="radio"/>	<input type="radio"/>
Any experiments configured to use the_cluster will run on ComputeOne.	<input type="radio"/>	<input type="radio"/>
The text Step1 will be printed to the screen.	<input type="radio"/>	<input type="radio"/>

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

Box 1: Yes

ComputeTargetException class: An exception related to failures when creating, interacting with, or configuring a compute target. This exception is commonly raised for failures attaching a compute target,

missing headers, and unsupported configuration values. Create(workspace, name, provisioning_configuration)

Provision a Compute object by specifying a compute type and related configuration. This method creates a new compute target rather than attaching an existing one. Box 2: Yes

Box 3: No

The line before print('Step1') will fail. Reference:

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

NEW QUESTION 216

- (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 contain missing values in several columns.

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

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

Solution: Use the last Observation Carried Forward (LOCF) method to impute the missing data points. Does the solution meet the goal?

- A. Yes
- B. No

Answer: B

Explanation:

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

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

Note: Last observation carried forward (LOCF) is a method of imputing missing data in longitudinal studies. If a person drops out of a study before it ends, then his

or her last observed score on the dependent variable is used for all subsequent (i.e., missing) observation points. LOCF is used to maintain the sample size and to reduce the bias caused by the attrition of participants in a study.

References:

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

NEW QUESTION 219

- (Exam Topic 3)

You are a data scientist creating a linear regression model.

You need to determine how closely the data fits the regression line. Which metric should you review?

- A. Coefficient of determination
- B. Recall
- C. Precision
- D. Mean absolute error
- E. Root Mean Square Error

Answer: A

Explanation:

Coefficient of determination, often referred to as R², represents the predictive power of the model as a value between 0 and 1. Zero means the model is random (explains nothing); 1 means there is a perfect fit. However, caution should be used in interpreting R² values, as low values can be entirely normal and high values can be suspect.

References:

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

NEW QUESTION 223

- (Exam Topic 3)

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

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

What should you do?

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

Answer: D

Explanation:

framework_version: The PyTorch version to be used for executing training code. Reference:

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

NEW QUESTION 224

- (Exam Topic 3)

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

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

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

Answer: C

Explanation:

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

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

NEW QUESTION 228

- (Exam Topic 3)

You create the following config.json file.

```
{
  "workspace_name" : "ml-workspace"
}
```

You must use the Azure Machine Learning SDK to interact with data and experiments in the workspace. You need to configure the config.json file to connect to the workspace from the Python environment. Which two additional parameters must you add to the config.json file in order to connect to the workspace?

Each correct answer presents part of the solution.

NOTE: Each correct selection is worth one point.

- A. subscription_id
- B. Key
- C. resource_group
- D. region
- E. Login

Answer: AC

Explanation:

To use the same workspace in multiple environments, create a JSON configuration file. The configuration file saves your subscription (subscription_id), resource (resource_group), and workspace name so that it can be easily loaded.

The following sample shows how to create a workspace. from azureml.core import Workspace

```
ws = Workspace.create(name='myworkspace', subscription_id='<azure-subscription-id>', resource_group='myresourcegroup', create_resource_group=True, location='eastus2')
```

Reference:

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

NEW QUESTION 233

- (Exam Topic 2)

You need to configure the Edit Metadata module so that the structure of the datasets match.

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

NOTE: Each correct selection is worth one point.

Properties Project

Edit Metadata

Column

Selected columns:

Column names: MedianValue

Launch column selector

Floating point
DateTime
TimeSpan
Integer

Unchanged
Make Categorical
Make Uncategorical

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

Box 1: Floating point

Need floating point for Median values.

Scenario: An initial investigation shows that the datasets are identical in structure apart from the MedianValue column. The smaller Paris dataset contains the MedianValue in text format, whereas the larger London dataset contains the MedianValue in numerical format.

Box 2: Unchanged

Note: Select the Categorical option to specify that the values in the selected columns should be treated as categories.

For example, you might have a column that contains the numbers 0,1 and 2, but know that the numbers actually mean "Smoker", "Non smoker" and "Unknown". In that case, by flagging the column as categorical you can ensure that the values are not used in numeric calculations, only to group data.

NEW QUESTION 234

- (Exam Topic 2)

You need to set up the Permutation Feature Importance module according to the model training requirements. Which properties should you select? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.

4 Tune Model Hyperparameters

Specify parameter sweeping mode

Random sweep

Maximum number of runs on random sweep

5

Random seed

0

Label column

Selected columns:
Column names: MedianValue

Launch column selector

Metric for measuring performance for classification

F-score
Precision
Recall
Accuracy

Metric for measuring performance for regression

Root of mean squared error
R-squared
Mean zero one error
Mean absolute error

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

Box 1: Accuracy

Scenario: You want to configure hyperparameters in the model learning process to speed the learning phase by using hyperparameters. In addition, this configuration should cancel the lowest performing runs at each evaluation interval, thereby directing effort and resources towards models that are more likely to be successful.

Box 2: R-Squared

NEW QUESTION 237

- (Exam Topic 2)

You need to select a feature extraction method. Which method should you use?

- A. Spearman correlation
- B. Mutual information
- C. Mann-Whitney test
- D. Pearson's correlation

Answer: A

Explanation:

Spearman's rank correlation coefficient assesses how well the relationship between two variables can be described using a monotonic function.

Note: Both Spearman's and Kendall's can be formulated as special cases of a more general correlation coefficient, and they are both appropriate in this scenario.

Scenario: The MedianValue and AvgRoomsInHouse columns both hold data in numeric format. You need to select a feature selection algorithm to analyze the relationship between the two columns in more detail. References:

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

NEW QUESTION 239

- (Exam Topic 2)

You need to identify the methods for dividing the data according, to the testing requirements.

Which properties should you select? To answer, select the appropriate option-, m the answer area. NOTE: Each correct selection is worth one point.

Answer Area

Properties Project

Partition and Sample

Partition or sample mode

Assign to Folds
Sampling
Head

Random seed

0

Specify the partitioner method

Partition evenly

Specify number of folds to split evenly into

5

Stratified split

Stratification key column

Selected columns:

Column names: NextToOver

Launch column selector

- A. Mastered
 B. Not Mastered

Answer: A

Explanation:

Answer Area

Properties Project

Partition and Sample

Partition or sample mode

Assign to Folds
Sampling
Head

Random seed

0

Specify the partitioner method

Partition evenly

Specify number of folds to split evenly into

5

Stratified split

Stratification key column

Selected columns:

Column names: NextToOver

Launch column selector

NEW QUESTION 244

.....

THANKS FOR TRYING THE DEMO OF OUR PRODUCT

Visit Our Site to Purchase the Full Set of Actual DP-100 Exam Questions With Answers.

We Also Provide Practice Exam Software That Simulates Real Exam Environment And Has Many Self-Assessment Features. Order the DP-100 Product From:

<https://www.2passeasy.com/dumps/DP-100/>

Money Back Guarantee

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