

DP-100 Dumps

Designing and Implementing a Data Science Solution on Azure

<https://www.certleader.com/DP-100-dumps.html>



NEW QUESTION 1

- (Exam Topic 3)

HOTSPOT

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

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

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

NOTE: Each correct selection is worth one point.

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

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

Box 1: Yes

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

Box 2: Yes

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

Box 3: No

Box 4: Yes

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

NEW QUESTION 2

- (Exam Topic 3)

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 = PipelineData("processed_data", datastore=datastore)
process_step = PythonScriptStep(script_name="process.py",
    arguments=["--data_for_train", data_output],
    outputs=[data_output], compute_target=aml_compute,
    source_directory=process_directory)
pipeline = Pipeline(workspace=ws, steps=[process_step])
```

Does the solution meet the goal?

- A. Yes

B. No

Answer: B

Explanation:

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

NEW QUESTION 3

- (Exam Topic 3)

You are using C-Support Vector classification to do a multi-class classification with an unbalanced training dataset. The C-Support Vector classification using Python code shown below:

```
from sklearn.svm import svc
import numpy as np
svc = SVC(kernel= 'linear', class_weight= 'balanced', C=1.0, random_state=0)
modell = svc.fit(X_train, y)
```

You need to evaluate the C-Support Vector classification code.

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

Code Segment	Evaluation Statement
class_weight=balanced	<ul style="list-style-type: none"> Automatically select the performance metrics for the classification. Automatically adjust weights directly proportional to class frequencies in the input data. Automatically adjust weights inversely proportional to class frequencies in the input data.
C parameter	<ul style="list-style-type: none"> Penalty parameter Degree of polynomial kernel function Size of the kernel cache

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

Box 1: Automatically adjust weights inversely proportional to class frequencies in the input data
The "balanced" mode uses the values of y to automatically adjust weights inversely proportional to class frequencies in the input data as $n_{\text{samples}} / (n_{\text{classes}} * \text{np.bincount}(y))$.
Box 2: Penalty parameter
Parameter: C : float, optional (default=1.0) Penalty parameter C of the error term. References:
<https://scikit-learn.org/stable/modules/generated/sklearn.svm.SVC.html>

NEW QUESTION 4

- (Exam Topic 3)

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

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

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

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

Does the solution meet the goal?

- A. Yes
- B. No

Answer: B

Explanation:

Instead use the Synthetic Minority Oversampling Technique (SMOTE) sampling mode.
Note: SMOTE is used to increase the number of underrepresented cases in a dataset used for machine learning. SMOTE is a better way of increasing the number of rare cases than simply duplicating existing cases.
References:
<https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/smote>

NEW QUESTION 5

- (Exam Topic 3)

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.

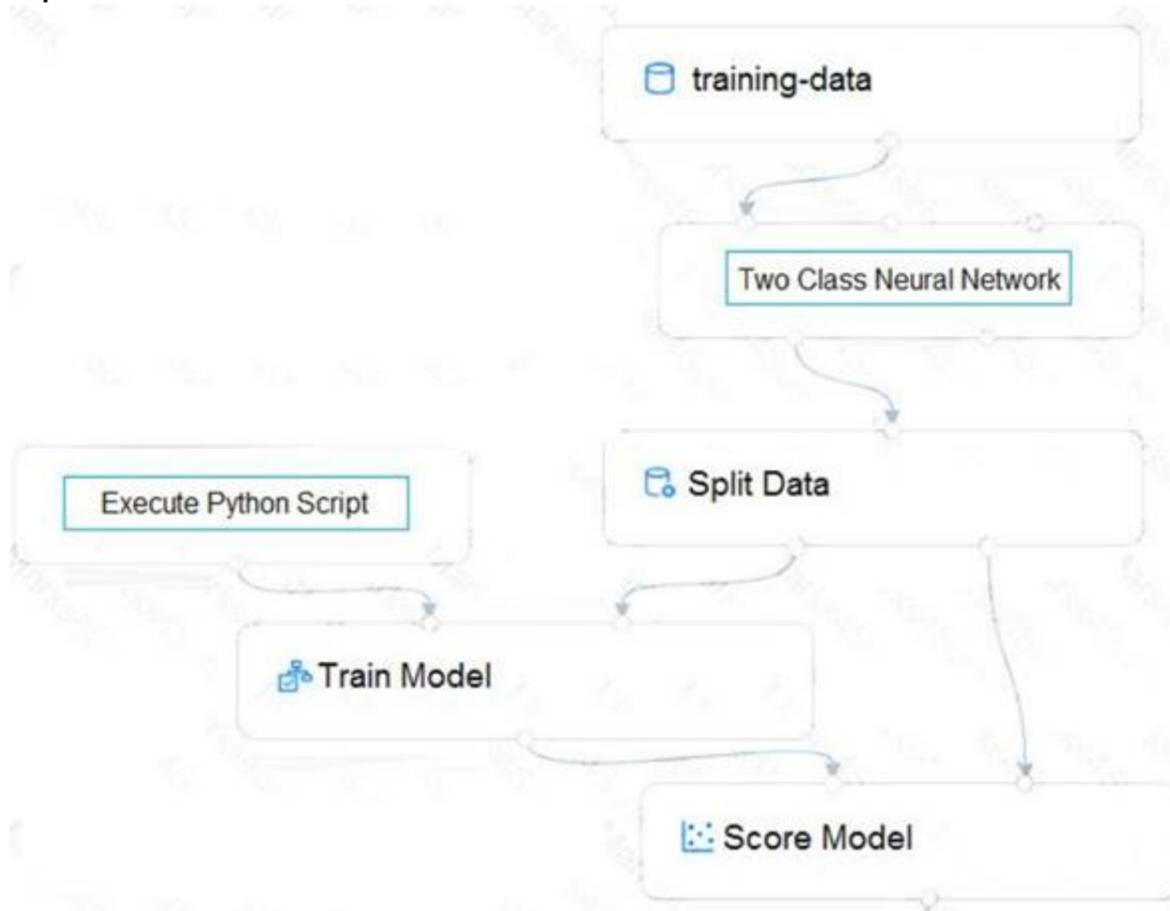
- > Train a model based on the naive_bayes.GaussianNB algorithm.
- > Return only the Scored Labels column by using the query SELECT [Scored Labels] FROM t1; Which modules should you use? To answer, drag the appropriate modules to the appropriate locations. Each module name may be used once, more than 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.

Modules	Answer Area
Create Python Model	
Train Model	
Two Class Neural Network	
Execute Python Script	
Apply SQL Transformation	
Select Columns in Dataset	

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:



NEW QUESTION 6

- (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 7

- (Exam Topic 3)

You create an Azure Machine Learning workspace.
You need to detect data drift between a baseline dataset and a subsequent target dataset by using the DataDriftDetector class.
How should you complete the code segment? To answer, select the appropriate options in the answer area.
NOTE: Each correct selection is worth one point.

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

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

monitor = DataDriftDetector. (ws, 'drift-monitor', baseline,
                             backfill
                             create_from_datasets
                             create_from_model
                             target, compute_target='cpu-cluster', frequency='Week', feature_list=None,
                             drift_threshold=.6, latency=24)

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

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

Graphical user interface, text, application, Word Description automatically generated
Box 1: create_from_datasets
The create_from_datasets method creates a new DataDriftDetector object from a baseline tabular dataset and a target time series dataset.
Box 2: backfill
The backfill method runs a backfill job over a given specified start and end date.
Syntax: backfill(start_date, end_date, compute_target=None, create_compute_target=False) Reference:
[https://docs.microsoft.com/en-us/python/api/azureml-datadrift/azureml.datadrift.datadriftdetector\(class\)](https://docs.microsoft.com/en-us/python/api/azureml-datadrift/azureml.datadrift.datadriftdetector(class))

NEW QUESTION 8

- (Exam Topic 3)

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

Actions

Answer Area

- Add a Two-Class Support Vector Machine module to initialize the SVM classifier.
- Set the Metric for measuring performance property to **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.



- 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 9

- (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

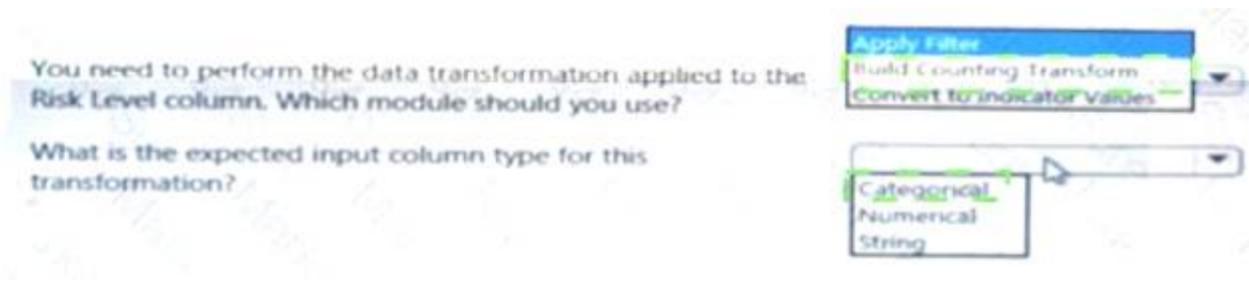
Numerical

String

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:



NEW QUESTION 10

- (Exam Topic 3)

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

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

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

Answer: A

Explanation:

Azure Cognitive Services expand on Microsoft’s evolving portfolio of machine learning APIs and enable developers to easily add cognitive features – such as emotion and video detection; facial, speech, and vision recognition; and speech and language understanding – into their applications. The goal of Azure Cognitive Services is to help developers create applications that can see, hear, speak, understand, and even begin to reason. The catalog of services within Azure Cognitive Services can be categorized into five main pillars - Vision, Speech, Language, Search, and Knowledge.

References:

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

NEW QUESTION 10

- (Exam Topic 3)

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 12

- (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: 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 17

- (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 20

- (Exam Topic 3)

You plan to use a Deep Learning Virtual Machine (DLVM) to train deep learning models using Compute Unified Device Architecture (CUDA) computations.

You need to configure the DLVM to support CUDA. What should you implement?

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

Answer: C

Explanation:

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

References:

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

NEW QUESTION 25

- (Exam Topic 3)

You are hired as a data scientist at a winery. The previous data scientist used Azure Machine Learning. You need to review the models and explain how each model makes decisions.

Which explainer modules should you use? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.

Model type	Explainer
A random forest model for predicting the alcohol content in wine given a set of covariates	<div style="border: 1px solid black; padding: 2px;"> <div style="text-align: right; margin-bottom: 5px;">▼</div> <div style="border: 1px solid black; padding: 2px;"> Tabular HAN Text Image </div> </div>
A natural language processing model for analyzing field reports	<div style="border: 1px solid black; padding: 2px;"> <div style="text-align: right; margin-bottom: 5px;">▼</div> <div style="border: 1px solid black; padding: 2px;"> Tree HAN Text Image </div> </div>
An image classifier that determines the quality of the grape based upon its physical characteristics.	<div style="border: 1px solid black; padding: 2px;"> <div style="text-align: right; margin-bottom: 5px;">▼</div> <div style="border: 1px solid black; padding: 2px;"> Kernel HAN Text Image </div> </div>

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

info based on the given model and data sets. The meta explainers leverage all the libraries (SHAP, LIME, Mimic, etc.) that we have integrated or developed. The following are the meta explainers available in the SDK:
 Tabular Explainer: Used with tabular datasets. Text Explainer: Used with text datasets. Image Explainer: Used with image datasets. Box 1: Tabular
 Box 2: Text
 Box 3: Image Reference:
<https://medium.com/microsoftazure/automated-and-interpretable-machine-learning-d07975741298>

NEW QUESTION 26

- (Exam Topic 3)

You are solving a classification task.

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

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

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

Answer: C

Explanation:

Leave One Out (LOO) cross-validation
 Setting $K = n$ (the number of observations) yields n-fold and is called leave-one out cross-validation (LOO), a special case of the K-fold approach.
 LOO CV is sometimes useful but typically doesn't shake up the data enough. The estimates from each fold are highly correlated and hence their average can have high variance.
 This is why the usual choice is $K=5$ or 10 . It provides a good compromise for the bias-variance tradeoff.

NEW QUESTION 30

- (Exam Topic 3)

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

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

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

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

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

Solution: Remove the entire column that contains the missing data point. Does the solution meet the goal?

- A. Yes
- B. No

Answer: B

Explanation:

Use the Multiple Imputation by Chained Equations (MICE) method. References: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3074241/>
<https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/clean-missing-data>

NEW QUESTION 32

- (Exam Topic 3)

You create an Azure Machine Learning workspace.

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

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

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

Which JSON code segment should you use?

A)

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

B)

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

C)

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

D)

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

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

Answer: A

Explanation:

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

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

> It can't delete the workspace.

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

data_scientist_custom_role.json :

```
{
  "Name": "Data Scientist Custom", "IsCustom": true,
  "Description": "Can run experiment but can't create or delete compute.", "Actions": ["*"],
  "NotActions": [
    "Microsoft.MachineLearningServices/workspaces/*/delete", "Microsoft.MachineLearningServices/workspaces/write",
    "Microsoft.MachineLearningServices/workspaces/computes/*/write", "Microsoft.MachineLearningServices/workspaces/computes/*/delete",
    "Microsoft.Authorization/*/write"
  ],
  "AssignableScopes": [ "/subscriptions/<subscription_id>/resourceGroups/<resource_group_name>/providers/Microsoft.MachineLearni"
]
```

Reference:

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

NEW QUESTION 33

- (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 38

- (Exam Topic 3)

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

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

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

- A. box plot
- B. scatter
- C. random forest diagram

- D. Venn diagram
- E. ROC curve

Answer: AB

Explanation:

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

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

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

NEW QUESTION 41

- (Exam Topic 3)

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

You create a variable that references the dataset using the following code: `training_ds = workspace.datasets.get("training_data")`

You define an estimator to run the script.

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

Which property should you set?

A)

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

B)

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

C)

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

D)

```
source_directory = training_ds
```

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

Answer: A

Explanation:

Example:

Get the training dataset

```
diabetes_ds = ws.datasets.get("Diabetes Dataset")
```

Create an estimator that uses the remote compute `hyper_estimator = SKLearn(source_directory=experiment_folder,`

```
inputs=[diabetes_ds.as_named_input('diabetes')], # Pass the dataset as an input compute_target = cpu_cluster, conda_packages=['pandas','ipykernel','matplotlib'],
```

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

```

Reference:

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

NEW QUESTION 43

- (Exam Topic 3)

You are evaluating a completed binary classification machine learning model. You need to use the precision as the valuation metric.

Which visualization should you use?

- A. Binary classification confusion matrix
- B. box plot
- C. Gradient descent
- D. coefficient of determination

Answer: A

Explanation:

References:

<https://machinelearningknowledge.ai/confusion-matrix-and-performance-metrics-machine-learning/>

NEW QUESTION 46

- (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.

The learning rate value 0.09 might be used during model training.

You can define an early termination policy for this hyperparameter tuning run.

- 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 49

- (Exam Topic 3)

You plan to build a team data science environment. Data for training models in machine learning pipelines will be over 20 GB in size.

You have the following requirements:

- Models must be built using Caffe2 or Chainer frameworks.
- Data scientists must be able to use a data science environment to build the machine learning pipelines and train models on their personal devices in both connected and disconnected network environments.
- Personal devices must support updating machine learning pipelines when connected to a network. You need to select a data science environment.

Which environment should you use?

- A. Azure Machine Learning Service
- B. Azure Machine Learning Studio
- C. Azure Databricks
- D. Azure Kubernetes Service (AKS)

Answer: A

Explanation:

The Data Science Virtual Machine (DSVM) is a customized VM image on Microsoft's Azure cloud built specifically for doing data science. Caffe2 and Chainer are supported by DSVM.

DSVM integrates with Azure Machine Learning.

NEW QUESTION 50

- (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 55

- (Exam Topic 3)

You create a Python script that runs a training experiment in Azure Machine Learning. The script uses the Azure Machine Learning SDK for Python.

You must add a statement that retrieves the names of the logs and outputs generated by the script. You need to reference a Python class object from the SDK for the statement.

Which class object should you use?

- A. Run
- B. ScriptRunConfig
- C. Workspace
- D. Experiment

Answer: A

Explanation:

A run represents a single trial of an experiment. Runs are used to monitor the asynchronous execution of a trial, log metrics and store output of the trial, and to analyze results and access artifacts generated by the trial.

The run Class `get_all_logs` method downloads all logs for the run to a directory. Reference:

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

NEW QUESTION 56

- (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")
deployment_config = .deploy_configuration(cpu_cores=1, memory_gb=1,
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 2: `AksWebservice`

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 59

- (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 =
```

- 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 60

- (Exam Topic 3)

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

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

You need to implement a monitoring solution for the deployed model using minimal administrative effort. What should you do?

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

Answer: B

Explanation:

Configure logging with Azure Machine Learning studio

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

- * 1. Sign in to the studio at <https://ml.azure.com>.
- * 2. Go to Models and select the model you want to deploy.
- * 3. Select +Deploy.
- * 4. Populate the Deploy model form.
- * 5. Expand the Advanced menu.
- * 6. Select Enable Application Insights diagnostics and data collection.

Advanced

Enable Application Insights diagnostics and data collection

Enable Application Insights diagnostics and data collection

Enable SSL

Enable SSL

Max concurrent requests per container

CPU reserve capacity ⓘ

Memory reserve capacity ⓘ

Reference:

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

NEW QUESTION 62

- (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 63

- (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: 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 66

- (Exam Topic 3)

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

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

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

One class has a much smaller number of observations than tin- other classes in the training set. You need to select an appropriate data sampling strategy to compensate for the class imbalance. Solution: You use the Principal Components Analysis (PCA) sampling mode.

Does the solution meet the goal?

- A. Yes
- B. No

Answer: B

Explanation:

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

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

NEW QUESTION 68

- (Exam Topic 3)

You previously deployed a model that was trained using a tabular dataset named training-dataset, which is based on a folder of CSV files.

Over time, you have collected the features and predicted labels generated by the model in a folder containing a CSV file for each month. You have created two

tabular datasets based on the folder containing the inference data: one named predictions-dataset with a schema that matches the training data exactly, including the predicted label; and another named features-dataset with a schema containing all of the feature columns and a timestamp column based on the filename, which includes the day, month, and year.

You need to create a data drift monitor to identify any changing trends in the feature data since the model was trained. To accomplish this, you must define the required datasets for the data drift monitor.

Which datasets should you use to configure the data drift monitor? To answer, drag the appropriate datasets to the correct data drift monitor options. Each source 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.

Target datasets	Answer Area
training-dataset	Baseline dataset
predictions-dataset	Target dataset
features-dataset	Target dataset

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

Text Description automatically generated with medium confidence

Box 1: training-dataset

Baseline dataset - usually the training dataset for a model. Box 2: predictions-dataset

Target dataset - usually model input data - is compared over time to your baseline dataset. This comparison means that your target dataset must have a timestamp column specified.

The monitor will compare the baseline and target datasets. Reference:

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

NEW QUESTION 70

- (Exam Topic 3)

You are using a decision tree algorithm. You have trained a model that generalizes well at a tree depth equal to 10.

You need to select the bias and variance properties of the model with varying tree depth values.

Which properties should you select for each tree depth? To answer, select the appropriate options in the answer area.

Answer Area

Tree Depth	Bias	Variance
5	<input type="text" value="High"/> <input type="text" value="Low"/> <input type="text" value="Identical"/>	<input type="text" value="High"/> <input type="text" value="Low"/> <input type="text" value="Identical"/>
15	<input type="text" value="High"/> <input type="text" value="Low"/> <input type="text" value="Identical"/>	<input type="text" value="High"/> <input type="text" value="Low"/> <input type="text" value="Identical"/>

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

In decision trees, the depth of the tree determines the variance. A complicated decision tree (e.g. deep) has low bias and high variance.

Note: In statistics and machine learning, the bias–variance tradeoff is the property of a set of predictive models whereby models with a lower bias in parameter estimation have a higher variance of the parameter estimates across samples, and vice versa. Increasing the bias will decrease the variance. Increasing the variance will decrease the bias.

References:

<https://machinelearningmastery.com/gentle-introduction-to-the-bias-variance-trade-off-in-machine-learning/>

NEW QUESTION 75

- (Exam Topic 3)

You use an Azure Machine Learning workspace.

You have a trained model that must be deployed as a web service. Users must authenticate by using Azure Active Directory.

What should you do?

- A. Deploy the model to Azure Kubernetes Service (AKS). During deployment, set the token_auth_enabled parameter of the target configuration object to true
- B. Deploy the model to Azure Container Instance

- C. During deployment, set the auch_enabled parameter of the target configuration object to true
- D. Deploy the model to Azure Container Instance
- E. During deployment, set the coken_auch_enabled parameter of the target configuration object to true
- F. Deploy the model to Azure Kubernetes Service (AKS). During deployment, set the auc
- G. enabled parameter of the target configuration object to true

Answer: A

Explanation:

To control token authentication, use the token_auth_enabled parameter when you create or update a deployment

Token authentication is disabled by default when you deploy to Azure Kubernetes Service.

Note: The model deployments created by Azure Machine Learning can be configured to use one of two authentication methods:

key-based: A static key is used to authenticate to the web service.

token-based: A temporary token must be obtained from the Azure Machine Learning workspace (using Azure Active Directory) and used to authenticate to the web service.

Reference:

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

NEW QUESTION 80

- (Exam Topic 3)

You are creating an experiment by using Azure Machine Learning Studio.

You must divide the data into four subsets for evaluation. There is a high degree of missing values in the data. You must prepare the data for analysis.

You need to select appropriate methods for producing the experiment.

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

NOTE: More than one order of answer choices is correct. You will receive credit for any of the correct orders you select.

Actions		Answer Area
Build Counting Transform		
Missing Values Scrubber		
Feature Hashing		
Clean Missing Data	⬅	⬆
Replace Discrete Values	➡	⬇
Import Data		
Latent Dirichlet Transformation		
Partition and Sample		

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

The Clean Missing Data module in Azure Machine Learning Studio, to remove, replace, or infer missing values.

NEW QUESTION 85

- (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 88

- (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 (OSVM)
- D. Azure Data bricks
- E. Azure Cognitive Services

Answer: ABD

NEW QUESTION 91

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

Does the solution meet the goal?

- A. Yes
- B. No

Answer: A

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 95

- (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.  (workspace = ws,
```

register_azure_blob_container
register_azure_file_share
register_azure_data_lake
register_azure_sql_database

```
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 exist, defaults to False. Reference:

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

NEW QUESTION 100

- (Exam Topic 3)

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

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

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

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

Answer: D

Explanation:

Reference:

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

NEW QUESTION 102

- (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 107

- (Exam Topic 3)

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

The experiment fails.

You need to troubleshoot the failed experiment.

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

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

Answer: BD

Explanation:

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

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

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

NEW QUESTION 111

- (Exam Topic 3)

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

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

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

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

Solution: Run the following code:

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

Does the solution meet the goal?

- A. Yes
- B. No

Answer: B

Explanation:

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

Compare with this example, the pipeline train step depends on the process_step_output output of the pipeline process step:

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

datastore = ws.get_default_datastore()

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

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

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

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

NEW QUESTION 112

- (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 AksWebservice deployment configuration for the service and deploy the model to it
- Install Docker on the workstation
- Create a LocalWebservice 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 AciWebservice 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 AksWebservice 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 = AksWebservice.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 LocalWebservice deployment configuration for the service and deploy the model to it

To deploy locally, modify your code to use LocalWebservice.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 113

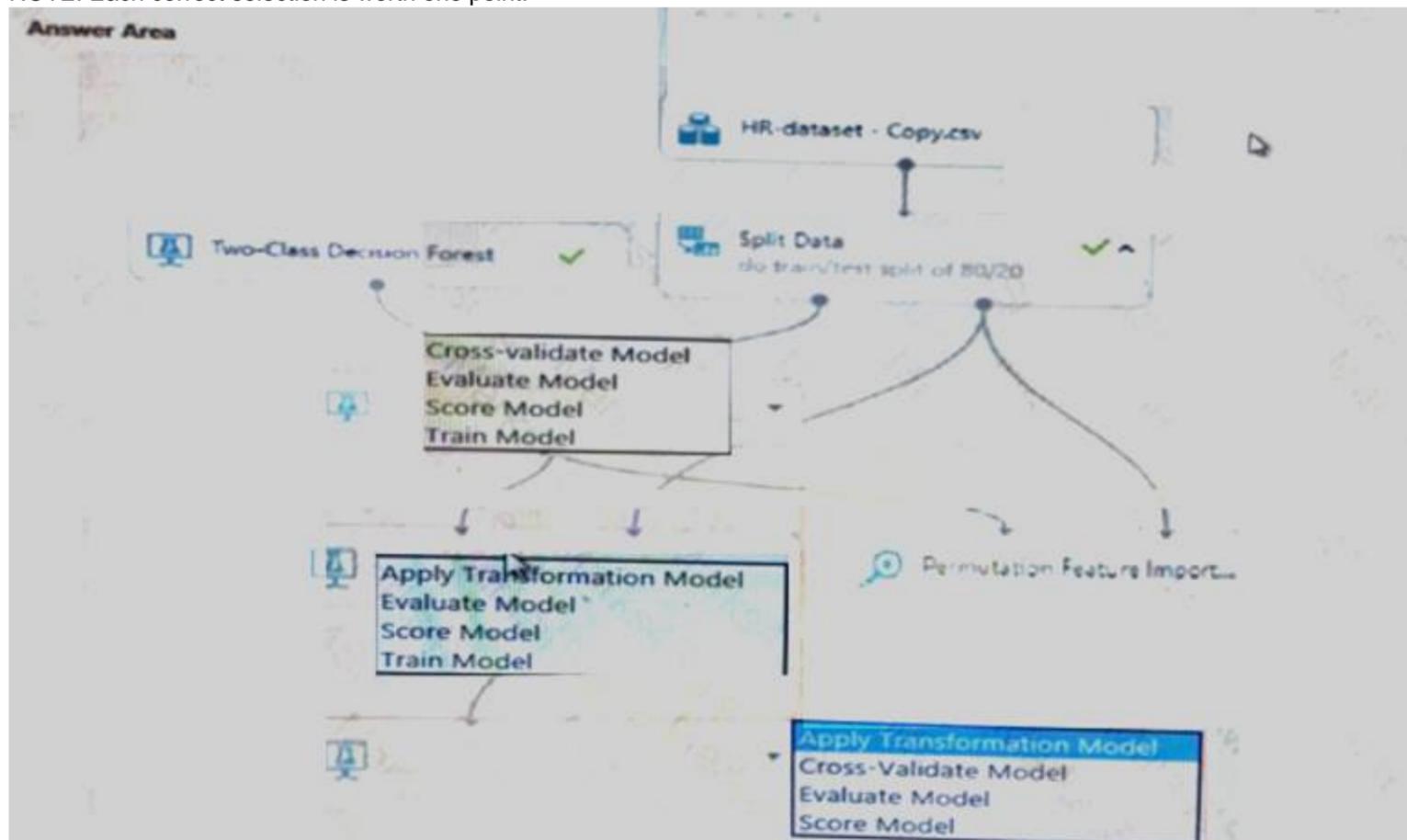
- (Exam Topic 3)

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

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

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

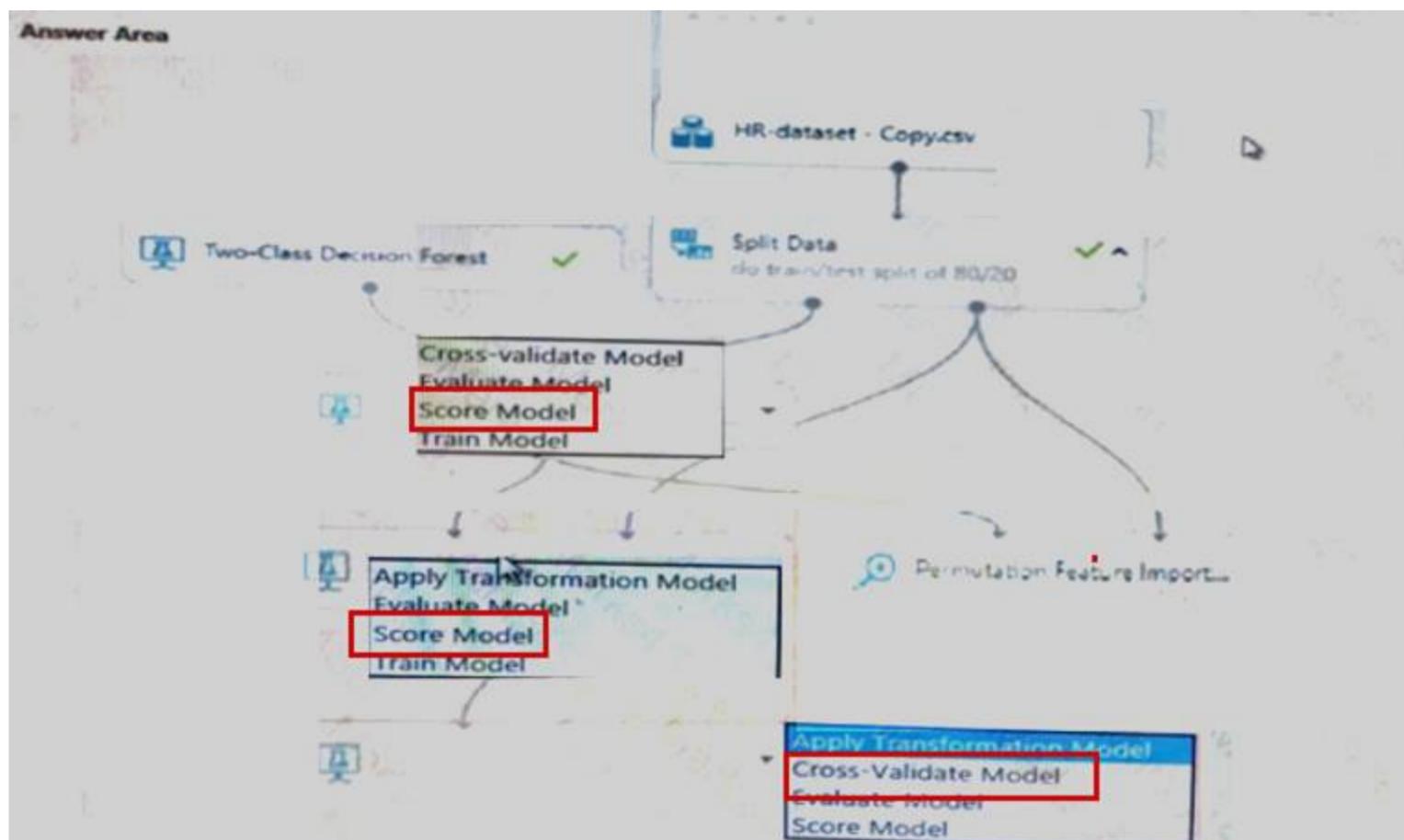
NOTE: Each correct selection is worth one point.



- A. Mastered
- B. Not Mastered

Answer: A

Explanation:



NEW QUESTION 115

- (Exam Topic 3)

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

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

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

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

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

How should you complete the code? To answer, drag the appropriate code segments to the correct locations. Each code segment may be used once, more than once, or not at all. You may need to drag the split bar between panes or scroll to view content.

NOTE: Each correct selection is worth one point.

Code segments

- log
- log_list
- log_row
- log_table
- log_image

Answer Area

```

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

```

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

Box 1: log

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

Scalar values: Log a numerical or string value to the run with the given name. Logging a metric to a run causes that metric to be stored in the run record in the experiment. You can log the same metric multiple times within a run, the result being considered a vector of that metric.

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

Box 2: log_image

A box plot of income by home_owner.

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

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

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

NEW QUESTION 118

- (Exam Topic 3)

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

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

The deployment fails.

You need to troubleshoot the deployment failure by determining the actions that were performed during deployment and identifying the specific action that failed. Which code segment should you run?

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

Answer: A

Explanation:

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

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

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

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

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

NEW QUESTION 122

- (Exam Topic 3)

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

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

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

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

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

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

NOTE: Each correct selection is worth one point.

```
# Assume required modules are imported
```

```
ws = Workspace.from_config()
```

```
feature_importances = explanation.
```

	(workspace = ws,
from_run	experiment_name='train_and_explain',
list_model_explanations	run_id='train_and_explain_12345')
from_run_id	
download_model_explanation	

```
explanation = client.
```

	()
upload_model_explanation	
list_model_explanations	
run	
download_model_explanation	

```
feature_importances = explanation.
```

	()
explanation	
explanation_client	
get_feature_important_dict	
download_model_explanation	

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

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

Box 1: from_run_id

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

Parameters

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

Box 2: list_model_explanations

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

Returns

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

Box 3: explanation:

Reference:

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

NEW QUESTION 126

- (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 127

- (Exam Topic 3)

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

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

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

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

Answer: B

Explanation:

wait_for_completion: Wait for the completion of this run. Returns the status object after the wait. Syntax: wait_for_completion(show_output=False, wait_post_processing=False, raise_on_error=True) Parameter: show_output Indicates whether to show the run output on sys.stdout.

NEW QUESTION 130

- (Exam Topic 3)

You are planning to register a trained model in an Azure Machine Learning workspace.

You must store additional metadata about the model in a key-value format. You must be able to add new metadata and modify or delete metadata after creation.

You need to register the model. Which parameter should you use?

- A. description
- B. model_framework
- C. cags
- D. properties

Answer: D

Explanation:

azureml.core.Model.properties:

Dictionary of key value properties for the Model. These properties cannot be changed after registration, however new key value pairs can be added.

Reference:

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

NEW QUESTION 131

- (Exam Topic 3)

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

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

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

You need to configure the module.

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

SMOTE

Label column

Selected columns:
All labels

Launch column selector

SMOTE percentage

0
300
3000
4000

Number of nearest neighbors

0
1
5
4000

Random seed

0

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

Box 1: 300

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

Box 2: 5

We should use 5 data rows.

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

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

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

References:

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

NEW QUESTION 135

- (Exam Topic 3)

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

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

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

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

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

variable named y_test variable, and the predicted probabilities from the model are stored in a variable named y_predicted. You need to add logging to the script to allow Hyperdrive to optimize hyperparameters for the AUC metric. Solution: Run the following code:

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

Does the solution meet the goal?

- A. Yes
- B. No

Answer: A

Explanation:

Python printing/logging example: logging.info(message)

Destination: Driver logs, Azure Machine Learning designer

Reference:

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

NEW QUESTION 138

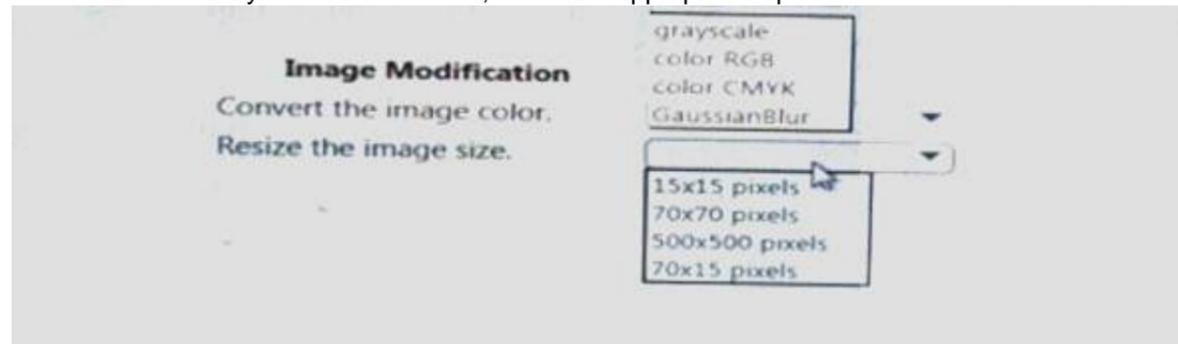
- (Exam Topic 3)

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

- Reduce the number of training epochs.
- Reduce the size of the neural network.
- Reduce over-fitting of the neural network.

You need to select the image modification values.

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



- A. Mastered
- B. Not Mastered

Answer: A

Explanation:



NEW QUESTION 140

- (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	<ul style="list-style-type: none"> Grid search Outlier detection Dimensionality reduction
Constraint	<ul style="list-style-type: none"> Demographic parity False-positive rate parity

- 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 144

- (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 145

- (Exam Topic 3)

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

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

You need to add code to run the steps.

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

NOTE: Each correct selection is worth one point.

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

Answer: CD

Explanation:

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

Build the pipeline. Example:

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

Submit the pipeline to be run

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

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

NEW QUESTION 150

- (Exam Topic 3)

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

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

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

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

Answer: BE

Explanation:

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

E: Summarize Data statistics are useful when you want to understand the characteristics of the complete dataset. For example, you might need to know: How many missing values are there in each column? How many unique values are there in a feature column? What is the mean and standard deviation for each column?

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

References:

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

NEW QUESTION 154

- (Exam Topic 3)

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 157

- (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 159

- (Exam Topic 3)

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

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

Which of the following is correct?

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

Answer: B

Explanation:

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

References:

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

NEW QUESTION 161

- (Exam Topic 3)

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

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

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

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

NOTE: Each correct selection is worth one point.

- A. run(mini_batch) D
- B. main()
- C. batch()
- D. init()
- E. score(mini_batch)

Answer: AD

Explanation:

Reference:

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

NEW QUESTION 166

- (Exam Topic 3)

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

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

What should you do?

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

Answer: B

Explanation:

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

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

Azure Databricks is fully integrated with Azure Data Factory.

NEW QUESTION 170

- (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:

- learning_rate: any value between 0.001 and 0.1
- batch_size: 16, 32, or 64

You need to configure the search space for the Hyperdrive experiment.

Which two parameter expressions should you use? Each correct answer presents part of the solution. NOTE: Each correct selection is worth one point.

- A. a choice expression for learning_rate
- B. a uniform expression for learning_rate
- C. a normal expression for batch_size
- D. a choice expression for batch_size
- E. a uniform expression for batch_size

Answer: BD

Explanation:

B: Continuous hyperparameters are specified as a distribution over a continuous range of values. Supported distributions include:

- uniform(low, high) - Returns a value uniformly distributed between low and high

D: 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/azure/machine-learning/how-to-tune-hyperparameters>

NEW QUESTION 173

- (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

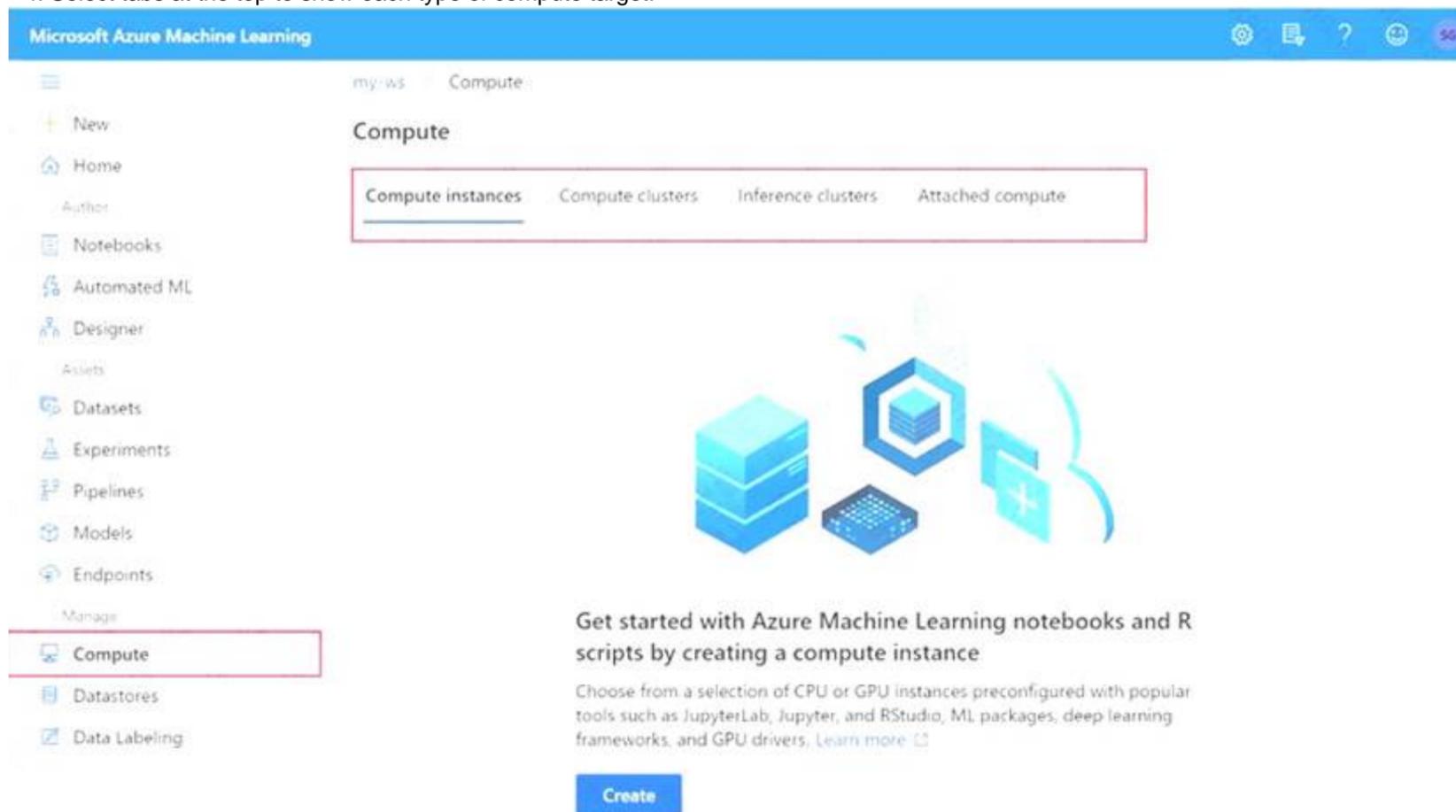
Answer: A

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

NEW QUESTION 175

- (Exam Topic 3)

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

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

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

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

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

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

Solution: Replace the comment with the following code:

```
run.upload_file('outputs/labels.csv', './data.csv') Does the solution meet the goal?
```

- A. Yes
- B. No

Answer: B

Explanation:

label_vals has the unique labels (from the statement label_vals = data['label'].unique()), and it has to be logged.

Note:

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

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

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

NEW QUESTION 177

- (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 Oversampling Technique (SMOTE)
- B. Replace using MICE
- C. Replace using; Probabilistic PCA
- D. Normalization

Answer: C

Explanation:

Replace using Probabilistic PCA: Compared to other options, such as Multiple Imputation using Chained Equations (MICE), this option has the advantage of not requiring the application of predictors for each column. Instead, it approximates the covariance for the full dataset. Therefore, it might offer better performance for datasets that have missing values in many columns.

References:

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

NEW QUESTION 181

- (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.	<ul style="list-style-type: none"> Set store1 as the default datastore. Disable data validation for store1. Update authentication for store1. Regenerate the keys of account1.
Preview the contents of store1 by using Azure Machine Learning studio.	<ul style="list-style-type: none"> Set store1 as the default datastore. Disable data validation for store1. Update authentication for store1. Regenerate the keys of account1.

- 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 184

- (Exam Topic 3)

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

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

You need to determine which compute environment to use for the following scenarios.

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

NOTE: Each correct selection is worth one point.

Environments	Answer Area						
<input type="text" value="nb_server"/> <input type="text" value="aks_cluster"/> <input type="text" value="mlc_cluster"/>	<table border="1"> <thead> <tr> <th>Scenario</th> <th>Environment</th> </tr> </thead> <tbody> <tr> <td>Run an Azure Machine Learning Designer training pipeline.</td> <td><input type="text" value="Environment"/></td> </tr> <tr> <td>Deploying a web service from the Azure Machine Learning designer.</td> <td><input type="text" value="Environment"/></td> </tr> </tbody> </table>	Scenario	Environment	Run an Azure Machine Learning Designer training pipeline.	<input type="text" value="Environment"/>	Deploying a web service from the Azure Machine Learning designer.	<input type="text" value="Environment"/>
Scenario	Environment						
Run an Azure Machine Learning Designer training pipeline.	<input type="text" value="Environment"/>						
Deploying a web service from the Azure Machine Learning designer.	<input type="text" value="Environment"/>						

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

Box 1: nb_server

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

Box 2: mlc_cluster

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

Reference:

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

NEW QUESTION 185

- (Exam Topic 3)

You are evaluating a completed binary classification machine. You need to use the precision as the evaluation metric. Which visualization should you use?

- A. scatter plot
- B. coefficient of determination
- C. Receiver Operating Characteristic (ROC) curve

D. Gradient descent

Answer: C

Explanation:

Receiver operating characteristic (or ROC) is a plot of the correctly classified labels vs. the incorrectly classified labels for a particular model.

References:

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

NEW QUESTION 188

- (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

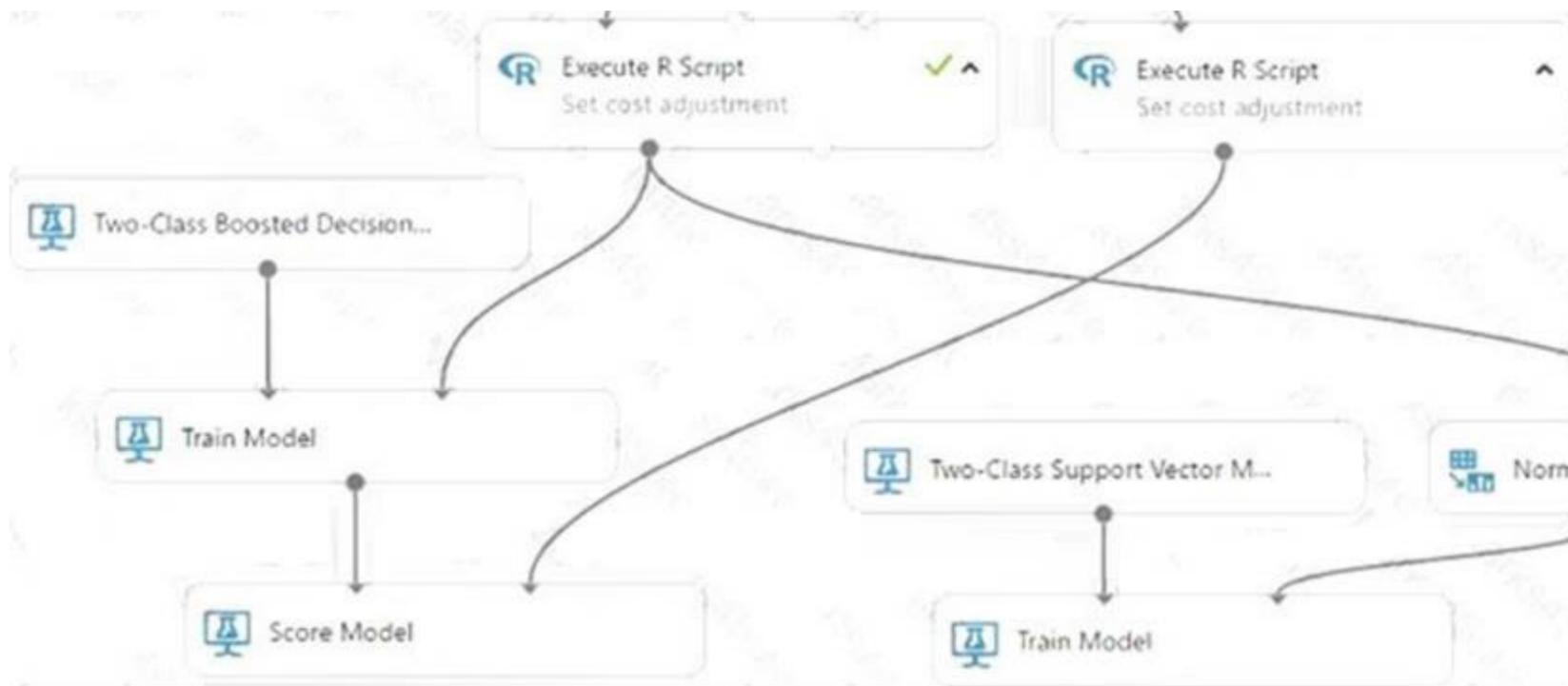
You use the testing data that was separated out by the Split Data module to score our trained models. You can then compare the results of the two models to see which generated better results.

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.

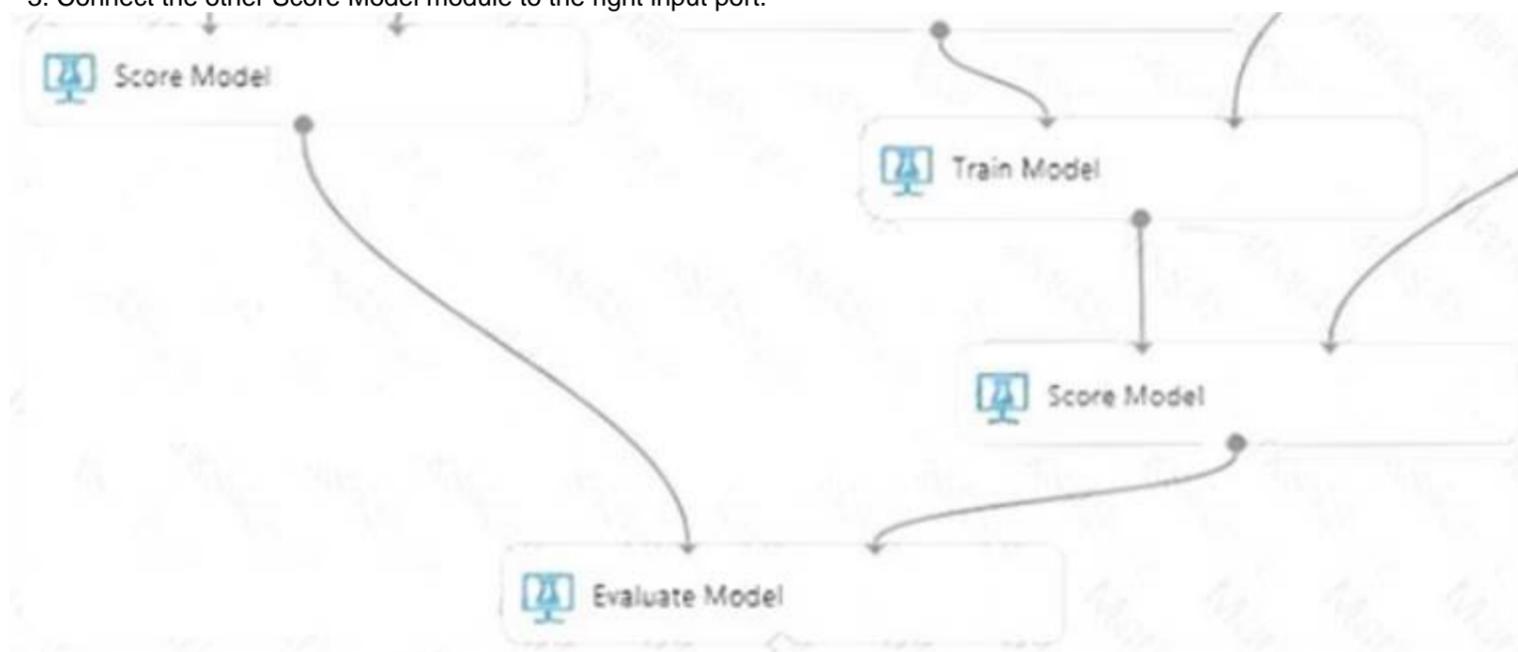
* 3. Connect the right Execute R Script module (our testing data) to the right input port of the Score Model module.



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 190

- (Exam Topic 3)

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

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

You train and register a machine learning model.

You plan to deploy the model as a real-time web service. Applications must use key-based authentication to use the model.

You need to deploy the web service.

Solution:

Create an AksWebservice instance.

Set the value of the auth_enabled property to False.

Set the value of the token_auth_enabled property to True.

Deploy the model to the service. Does the solution meet the goal?

- A. Yes
- B. No

Answer: B

Explanation:

Instead use only auth_enabled = TRUE Note: Key-based authentication.

Web services deployed on AKS have key-based auth enabled by default. ACI-deployed services have key-based auth disabled by default, but you can enable it by setting auth_enabled = TRUE when creating the ACI web service. The following is an example of creating an ACI deployment configuration with key-based auth enabled.

```
deployment_config <- aci_webservice_deployment_config(cpu_cores = 1, memory_gb = 1, auth_enabled = TRUE)
```

Reference:

<https://azure.github.io/azureml-sdk-for-r/articles/deploying-models.html>

NEW QUESTION 195

- (Exam Topic 3)

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

What should you create?

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

Answer: E

NEW QUESTION 196

- (Exam Topic 3)

You are a data scientist building a deep convolutional neural network (CNN) for image classification. The CNN model you built shows signs of overfitting. You need to reduce overfitting and converge the model to an optimal fit.

Which two actions should you perform? Each correct answer presents a complete solution.

NOTE: Each correct selection is worth one point.

- A. Reduce the amount of training data.
- B. Add an additional dense layer with 64 input units
- C. Add L1/L2 regularization.
- D. Use training data augmentation
- E. Add an additional dense layer with 512 input units.

Answer: AC

Explanation:

References:

<https://machinelearningmastery.com/how-to-reduce-overfitting-in-deep-learning-with-weight-regularization/>

https://en.wikipedia.org/wiki/Convolutional_neural_network

NEW QUESTION 197

- (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 201

- (Exam Topic 3)

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

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

You are a data scientist using Azure Machine Learning Studio.

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

Does the solution meet the GOAL?

- A. Yes
- B. No

Answer: B

Explanation:

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

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

NEW QUESTION 202

- (Exam Topic 3)

You use Azure Machine Learning Studio to build a machine learning experiment. You need to divide data into two distinct datasets.

Which module should you use?

- A. Partition and Sample
- B. Assign Data to Clusters
- C. Group Data into Bins
- D. Test Hypothesis Using t-Test

Answer: A

Explanation:

Partition and Sample with the Stratified split option outputs multiple datasets, partitioned using the rules you specified.

References:

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

NEW QUESTION 207

- (Exam Topic 3)

You use the Azure Machine Learning designer to create and run a training pipeline. You then create a real-time inference pipeline.

You must deploy the real-time inference pipeline as a web service.

What must you do before you deploy the real-time inference pipeline?

- A. Run the real-time inference pipeline.
- B. Create a batch inference pipeline.
- C. Clone the training pipeline.
- D. Create an Azure Machine Learning compute cluster.

Answer: D

Explanation:

You need to create an inferencing cluster. Deploy the real-time endpoint

After your AKS service has finished provisioning, return to the real-time inferencing pipeline to complete deployment.

- > Select Deploy above the canvas.
- > Select Deploy new real-time endpoint.
- > Select the AKS cluster you created.
- > Select Deploy. Reference:

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

NEW QUESTION 210

- (Exam Topic 3)

You are performing a classification task in Azure Machine Learning Studio.

You must prepare balanced testing and training samples based on a provided data set. You need to split the data with a 0.75:0.25 ratio.

Which value should you use for each parameter? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.

Parameter	Value
Splitting mode	<div style="border: 1px solid black; padding: 5px;"> Split rows Recommender Split Regular Expression Split Relative Expression Split </div>
Fraction of rows in the first output dataset	<div style="border: 1px solid black; padding: 5px;"> 0.75 0.25 0.5 1 </div>
Randomized split	<div style="border: 1px solid black; padding: 5px;"> True False </div>
Stratified split	<div style="border: 1px solid black; padding: 5px;"> True False </div>

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

Box 1: Split rows

Use the Split Rows option if you just want to divide the data into two parts. You can specify the percentage of data to put in each split, but by default, the data is divided 50-50.

You can also randomize the selection of rows in each group, and use stratified sampling. In stratified sampling, you must select a single column of data for which you want values to be apportioned equally among the two result datasets.

Box 2: 0.75

If you specify a number as a percentage, or if you use a string that contains the "%" character, the value is interpreted as a percentage. All percentage values must

be within the range (0, 100), not including the values 0 and 100.

Box 3: Yes

To ensure splits are balanced.

Box 4: No

If you use the option for a stratified split, the output datasets can be further divided by subgroups, by selecting a strata column.

Reference:

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

NEW QUESTION 212

- (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 213

- (Exam Topic 3)

You have a dataset that contains 2,000 rows. You are building a machine learning classification model by using Azure Learning Studio. You add a Partition and Sample module to the experiment.

You need to configure the module. You must meet the following requirements:

- > Divide the data into subsets
- > Assign the rows into folds using a round-robin method
- > Allow rows in the dataset to be reused

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

NOTE: Each correct selection is worth one point.

Partition and Sample

Partition or sample mode

Assign to Folds

Pick Fold

Sampling

Head

Use replacement in the partitioning

Randomized split

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

Use the Split data into partitions option when you want to divide the dataset into subsets of the data. This option is also useful when you want to create a custom number of folds for cross-validation, or to split rows into several groups.

- For Partition or sample mode, select Assign to Folds.
- Use replacement in the partitioning: Select this option if you want the sampled row to be put back into the pool of rows for potential reuse. As a result, the same row might be assigned to several folds.
- If you do not use replacement (the default option), the sampled row is not put back into the pool of rows for potential reuse. As a result, each row can be assigned to only one fold.
- Randomized split: Select this option if you want rows to be randomly assigned to folds. If you do not select this option, rows are assigned to folds using the round-robin method. References:
<https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/partition-and-sample>

NEW QUESTION 216

- (Exam Topic 3)

You are retrieving data from a large datastore by using Azure Machine Learning Studio. You must create a subset of the data for testing purposes using a random sampling seed based on the system clock. You add the Partition and Sample module to your experiment. You need to select the properties for the module. Which values should you select? To answer, select the appropriate options in the answer area.
NOTE: Each correct selection is worth one point.

▸ Partition and Sample

Partition or sample mode

▼

Assign to Folds
Pick Fold
Sampling
Head

Rate of sampling

.2

Random seed for sampling

▼

0
1
time.clock()
utcNow()

Stratified split for sampling

False ▼

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

Box 1: Sampling Create a sample of data

This option supports simple random sampling or stratified random sampling. This is useful if you want to create a smaller representative sample dataset for testing.

- * 1. Add the Partition and Sample module to your experiment in Studio, and connect the dataset.
- * 2. Partition or sample mode: Set this to Sampling.
- * 3. Rate of sampling.

See box 2 below.

Box 2: 0

- * 3. Rate of sampling. Random seed for sampling: Optionally, type an integer to use as a seed value.

This option is important if you want the rows to be divided the same way every time. The default value is 0, meaning that a starting seed is generated based on the system clock. This can lead to slightly different results each time you run the experiment.

References:

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

NEW QUESTION 217

- (Exam Topic 3)

You use Azure Machine Learning Studio to build a machine learning experiment. You need to divide data into two distinct datasets. Which module should you use?

- A. Split Data
- B. Load Trained Model
- C. Assign Data to Clusters
- D. Group Data into Bins

Answer: D

Explanation:

The Group Data into Bins module supports multiple options for binning data. You can customize how the bin edges are set and how values are apportioned into the bins.

References:

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

NEW QUESTION 220

- (Exam Topic 3)

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

The custom role has the following JSON definition:

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

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

NOTE: Each correct selection is worth one point.

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

A.

Answer:

Explanation:

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

Box 1: No

The actions listed in NotActions are prohibited.

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

Box 2: No

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

Writing metrics is not listed in NotActions. Reference:

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

NEW QUESTION 223

- (Exam Topic 3)

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

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

You are using Azure Machine Learning Studio to perform feature engineering on a dataset. You need to normalize values to produce a feature column grouped into bins.

Solution: Apply an Entropy Minimum Description Length (MDL) binning mode. Does the solution meet the goal?

A. Yes

B. No

Answer: A

Explanation:

Entropy MDL binning mode: This method requires that you select the column you want to predict and the column or columns that you want to group into bins. It then makes a pass over the data and attempts to determine the number of bins that minimizes the entropy. In other words, it chooses a number of bins that allows the data column to best predict the target column. It then returns the bin number associated with each row of your data in a column named <colname>quantized.

References:

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

NEW QUESTION 226

- (Exam Topic 3)

You have a comma-separated values (CSV) file containing data from which you want to train a classification model.

You are using the Automated Machine Learning interface in Azure Machine Learning studio to train the classification model. You set the task type to Classification.

You need to ensure that the Automated Machine Learning process evaluates only linear models. What should you do?

- A. Add all algorithms other than linear ones to the blocked algorithms list.
- B. Set the Exit criterion option to a metric score threshold.
- C. Clear the option to perform automatic featurization.
- D. Clear the option to enable deep learning.
- E. Set the task type to Regression.

Answer: C

Explanation:

Automatic featurization can fit non-linear models. Reference: <https://econml.azurewebsites.net/spec/estimation/dml.html>

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

NEW QUESTION 229

- (Exam Topic 3)

You are building a regression model to estimate 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 231

- (Exam Topic 3)

You are performing a filter based feature selection for a dataset that has 10 features. You build a multi class classifier 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. Pearson 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 233

- (Exam Topic 3)

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

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

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

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

All files store data in the following format:

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

You run the following code:

```
data_store = Datastore.register_azure_blob_container(workspace=ws,  
datastore_name='data_store',  
container_name='quarterly_data',  
account_name='companydata',  
account_key='NRPxk8duxBM3...'  
create_if_not_exists=False)
```

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

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

Solution: Run the following code:

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

Does the solution meet the goal?

- A. Yes
- B. No

Answer: A

Explanation:

Use two file paths.

Use Dataset.Tabular_from_delimited as the data isn't cleansed. Note:

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

Reference:

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

NEW QUESTION 237

.....

Thank You for Trying Our Product

* 100% Pass or Money Back

All our products come with a 90-day Money Back Guarantee.

* One year free update

You can enjoy free update one year. 24x7 online support.

* Trusted by Millions

We currently serve more than 30,000,000 customers.

* Shop Securely

All transactions are protected by VeriSign!

100% Pass Your DP-100 Exam with Our Prep Materials Via below:

<https://www.certleader.com/DP-100-dumps.html>