Microsoft

Exam Questions DP-100

Designing and Implementing a Data Science Solution on Azure

NEW QUESTION 1
- (Exam Topic 3)
You have an Azure Machine Learning workspace that contains a training cluster and an inference cluster. You plan to create a classification model by using the Azure Machine Learning designer.
You need to ensure that client applications can submit data as HTTP requests and receive predictions as responses.
Which three actions should you perform in sequence? To answer, move the appropriate actions from the list of actions to the answer area and arrange them in the correct order.

A. Mastered
B. Not Mastered

Answer: A

Explanation:

NEW QUESTION 2
- (Exam Topic 3)
You have several machine learning models registered in an Azure Machine Learning workspace. You must use the Fairlearn dashboard to assess fairness in a selected model.
Which three actions should you perform in sequence? To answer, move the appropriate actions from the list of actions to the answer area and arrange them in the correct order.

A. Mastered

B. Not Mastered

Answer: A

Explanation:
Graphical user interface, text, application Description automatically generated
Step 1: Select a model feature to be evaluated.
Step 2: Select a binary classification or regression model.
Register your models within Azure Machine Learning. For convenience, store the results in a dictionary,
which maps the id of the registered model (a string in name:version format) to the predictor itself. Example:
model_dict = {}
lr_reg_id = register_model("fairness_logistic_regression", lr_predictor) model_dict[lr_reg_id] = lr_predictor
svm_reg_id = register_model("fairness_svm", svm_predictor) model_dict[svm_reg_id] = svm_predictor
Step 3: Select a metric to be measured Precompute fairness metrics.
Create a dashboard dictionary using Fairlearn's metrics package. Reference:
https://docs.microsoft.com/en-us/azure/machine-learning/how-to-machine-learning-fairness-aml

NEW QUESTION 3
- (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	• 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: • Operating system: Ubuntu Linux • Software installed: Python 3.6 and Jupyter Notebooks • Size: NC6 (6 vCPUs, 1 vGPU, 56 Gb RAM)

The IT department creates an Azure Kubernetes Service (AKS)-based inference compute target named
aks-cluster in the Azure Machine Learning workspace. You have a Microsoft Surface Book computer with a GPU. Python 3.6 and Visual Studio Code are installed.
You need to run a script that trains a deep neural network (DNN) model and logs the loss and accuracy metrics.
Solution: Install the Azure ML SDK on the Surface Book. Run Python code to connect to the workspace. Run the training script as an experiment on the aks-cluster compute target.
Does the solution meet the goal?

A. Yes
B. No

Answer: B

Explanation:
Need to attach the mlvm virtual machine as a compute target in the Azure Machine Learning workspace.
Reference:
https://docs.microsoft.com/en-us/azure/machine-learning/concept-compute-target

NEW QUESTION 4
- (Exam Topic 3)
You create a Python script named train.py and save it in a folder named scripts. The script uses the scikit-learn framework to train a machine learning model.
You must run the script as an Azure Machine Learning experiment on your local workstation. You need to write Python code to initiate an experiment that runs the train.py script.
How should you complete the code segment? To answer, select the appropriate options in the answer area.
NOTE: Each correct selection is worth one point.

Answer Area

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

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

                                                     ▼  = 'train.py',
                                  script
                                  arguments
                                  environment
                                  compute_target

                                                     ▼  =py_sk)
                                  arguments
                                  resume_from
                                  environment
                                  compute_target

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

A. Mastered
B. Not Mastered

Answer: A

Explanation:
Graphical user interface, text, application, table, Word Description automatically generated
Box 1: source_directory
source_directory: A local directory containing code files needed for a run. Box 2: script
Script: The file path relative to the source_directory of the script to be run. Box 3: environment
Reference:
https://docs.microsoft.com/en-us/python/api/azureml-core/azureml.core.scriptrunconfig

NEW QUESTION 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

Create Python Model

Train Model

Two Class Neural Network

Execute Python Script

Apply SQL Transformation

Select Columns in Dataset

Answer Area

training-data

Split Data

Train Model

Score Model

A. Mastered
B. Not Mastered

Answer: A

Explanation:

training-data

Two Class Neural Network

Execute Python Script

Split Data

Train Model

Score Model
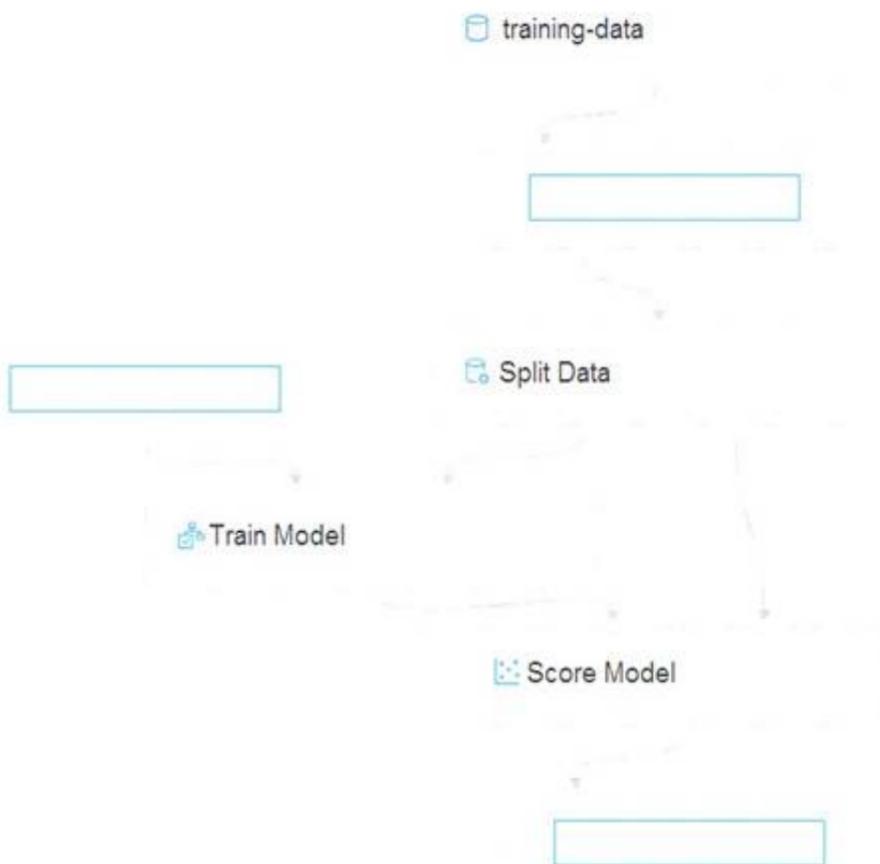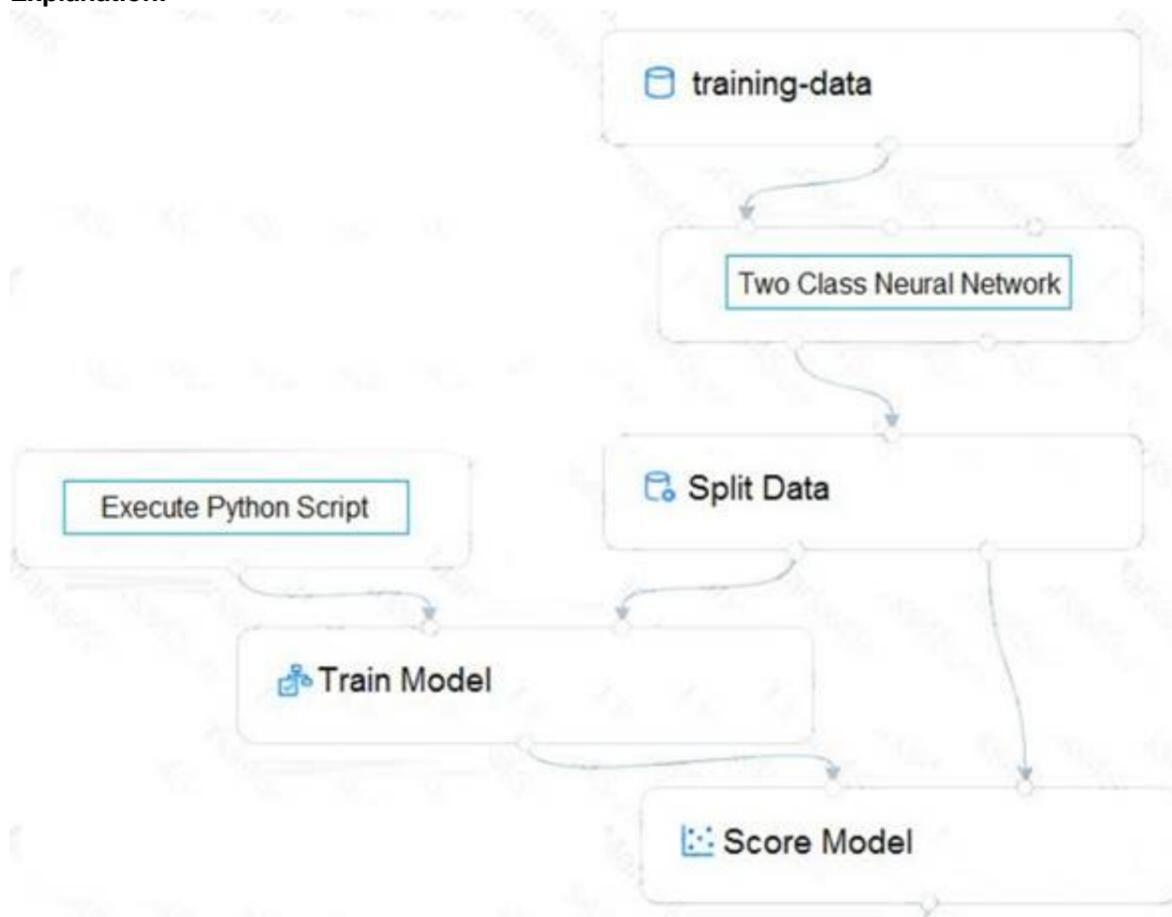
NEW QUESTION 6
- (Exam Topic 3)
You are developing a deep learning model by using TensorFlow. You plan to run the model training workload on an Azure Machine Learning Compute Instance.
You must use CUDA-based model training. You need to provision the Compute Instance.
Which two virtual machines sizes can you use? To answer, select the appropriate virtual machine sizes in the answer area.
NOTE: Each correct selection is worth one point.

Virtual machine size

Search by name...

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

A. Mastered
B. Not Mastered

Answer: A

Explanation:
CUDA is a parallel computing platform and programming model developed by Nvidia for general computing on its own GPUs (graphics processing units). CUDA enables developers to speed up compute-intensive applications by harnessing the power of GPUs for the parallelizable part of the computation.
Reference:
https://www.infoworld.com/article/3299703/what-is-cuda-parallel-programming-for-gpus.html

NEW QUESTION 7
- (Exam Topic 3)
Your Azure Machine Learning workspace has a dataset named real_estate_data. A sample of the data in the dataset follows.

postal_code	num_bedrooms	sq_feet	garage	price
12345	3	1300	0	23,9000
54321	1	950	0	11,0000
12346	2	1200	1	15,0000

You want to use automated machine learning to find the best regression model for predicting the price column. You need to configure an automated machine learning experiment using the Azure Machine Learning SDK. How should you complete the code? To answer, select the appropriate options in the answer area.
NOTE: Each correct selection is worth one point.

Answer Area

```
from azureml.core import Workspace
from azureml.core.compute import ComputeTarget
from azureml.core.runconfig import RunConfiguration
from azureml.train.automl import AutoMLConfig

ws = Workspace.from_config()
training_cluster = ComputeTarget(workspace=ws, name= 'aml-cluster1')
real_estate_ds = ws.datasets.get('real_estate_data')
split1_ds, split2_ds = real_estate_ds.random_split(percentage=0.7, seed=123)
automl_run_config = RunConfiguration(framework= "python")
automl_config = AutoMLConfig(         task= 'regression',
                        compute_target= training_cluster,
                        run_configuration=automl_run_config,
                        primary_metric='r2_score',
                                    ▼ =split1_ds,
```

```
X
Y
X_valid
Y_valid
training_data
```

```
                                    ▼ =split2_ds
```

```
X
Y
X_valid
Y_valid
validation_data
training_data
```

```
                                    ▼ ='price')
```

```
y
y_valid
y_max
label_column_name
exclude_nan_labels
```

A. Mastered
B. Not Mastered

Answer: A

Explanation:
Box 1: training_data
The training data to be used within the experiment. It should contain both training features and a label column (optionally a sample weights column). If training_data is specified, then the label_column_name parameter must also be specified.
Box 2: validation_data
Provide validation data: In this case, you can either start with a single data file and split it into training and validation sets or you can provide a separate data file for the validation set. Either way, the validation_data parameter in your AutoMLConfig object assigns which data to use as your validation set.
Example, the following code example explicitly defines which portion of the provided data in dataset to use for training and validation.
dataset = Dataset.Tabular.from_delimited_files(data)
training_data, validation_data = dataset.random_split(percentage=0.8, seed=1) automl_config = AutoMLConfig(compute_target = aml_remote_compute, task = 'classification',
primary_metric = 'AUC_weighted', training_data = training_data,
validation_data = validation_data, label_column_name = 'Class'
)
Box 3: label_column_name label_column_name:
The name of the label column. If the input data is from a pandas.DataFrame which doesn't have column names, column indices can be used instead, expressed as integers.
This parameter is applicable to training_data and validation_data parameters. Reference:
https://docs.microsoft.com/en-us/python/api/azureml-train-automl-client/azureml.train.automl.automlconfig.auto

NEW QUESTION 8
- (Exam Topic 3)
You train a model and register it in your Azure Machine Learning workspace. You are ready to deploy the model as a real-time web service.
You deploy the model to an Azure Kubernetes Service (AKS) inference cluster, but the deployment fails because an error occurs when the service runs the entry script that is associated with the model deployment.
You need to debug the error by iteratively modifying the code and reloading the service, without requiring a re-deployment of the service for each code update.
What should you do?

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

Answer: C

Explanation:
How to work around or solve common Docker deployment errors with Azure Container Instances (ACI) and Azure Kubernetes Service (AKS) using Azure Machine Learning.
The recommended and the most up to date approach for model deployment is via the Model.deploy() API using an Environment object as an input parameter. In this case our service will create a base docker image for you during deployment stage and mount the required models all in one call. The basic deployment tasks are:
* 1. Register the model in the workspace model registry.
* 2. Define Inference Configuration:
* a. Create an Environment object based on the dependencies you specify in the environment yaml file or use one of our procured environments.
* b. Create an inference configuration (InferenceConfig object) based on the environment and the scoring script.
* 3. Deploy the model to Azure Container Instance (ACI) service or to Azure Kubernetes Service (AKS).

NEW QUESTION 9
- (Exam Topic 3)
Note: This question is part of a series of questions that present the same scenario. Each question in the series contains a unique solution that might meet the stated goals. Some question sets might have more than one correct solution, while others might not have a correct solution.
After you answer a question in this section, you will NOT be able to return to it. As a result, these questions will not appear in the review screen.
You are creating a model to predict the price of a student's artwork depending on the following variables: the student's length of education, degree type, and art form.
You start by creating a linear regression model. You need to evaluate the linear regression model.
Solution: Use the following metrics: Relative Squared Error, Coefficient of Determination, Accuracy, Precision, Recall, F1 score, and AUC.
Does the solution meet the goal?

A. Yes
B. No

Answer: B

Explanation:
Relative Squared Error, Coefficient of Determination are good metrics to evaluate the linear regression model, but the others are metrics for classification models.
References:
https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/evaluate-model

NEW QUESTION 10
- (Exam Topic 3)
Note: This question is part of a series of questions that present the same scenario. Each question in the series contains a unique solution that might meet the stated goals. Some question sets might have more than one correct solution, while others might not have a correct solution.
After you answer a question in this section, you will NOT be able to return to it. As a result, these questions will not appear in the review screen.
You plan to use a Python script to run an Azure Machine Learning experiment. The script creates a reference to the experiment run context, loads data from a file,

identifies the set of unique values for the label column, and completes the experiment run:

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

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

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

A. Yes
B. No

Answer: A

Explanation:
run.log_list log a list of values to the run with the given name using log_list. Example: run.log_list("accuracies", [0.6, 0.7, 0.87])
Note:
Data= pd.read_csv('data.csv')
Data is read into a pandas.DataFrame, which is a two-dimensional, size-mutable, potentially heterogeneous tabular data.
label_vals =data['label'].unique
label_vals contains a list of unique label values. Reference:
https://www.element61.be/en/resource/azure-machine-learning-services-complete-toolbox-ai https://docs.microsoft.com/en-us/python/api/azureml-core/azureml.core.run(class) https://pandas.pydata.org/docs/reference/api/pandas.DataFrame.html

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 in 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 mont
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 registere
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 mont
F. Register the dataset with the name sales_dataset_MM-YYYY each month with appropriate MM and YYYY values for the month and yea
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' fil
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 registere
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 use the Azure Machine Learning designer to create and run a training pipeline.
The pipeline must be run every night to inference predictions from a large volume of files. The folder where the files will be stored is defined as a dataset.
You need to publish the pipeline as a REST service that can be used for the nightly inferencing run. What should you do?

A. Create a batch inference pipeline
B. Set the compute target for the pipeline to an inference cluster
C. Create a real-time inference pipeline
D. Clone the pipeline

Answer: A

Explanation:
Azure Machine Learning Batch Inference targets large inference jobs that are not time-sensitive. Batch Inference provides cost-effective inference compute scaling, with unparalleled throughput for asynchronous applications. It is optimized for high-throughput, fire-and-forget inference over large collections of data.
You can submit a batch inference job by pipeline_run, or through REST calls with a published pipeline. Reference:
https://github.com/Azure/MachineLearningNotebooks/blob/master/how-to-use-azureml/machine-learning-pipeli

NEW QUESTION 13
- (Exam Topic 3)
You are training machine learning models in Azure Machine Learning. You use Hyperdrive to tune the hyperparameters. In previous model training and tuning runs, many models showed similar performance. You need to select an early termination policy that meets the following requirements:
• accounts for the performance of all previous runs when evaluating the current run
• avoids comparing the current run with only the best performing run to date
Which two early termination policies should you use? Each correct answer presents part of the solution. NOTE: Each correct selection is worth one point.

A. Bandit
B. Median stopping
C. Default
D. Truncation selection

Answer: BC

Explanation:
The Median Stopping policy computes running averages across all runs and cancels runs whose best performance is worse than the median of the running averages.
If no policy is specified, the hyperparameter tuning service will let all training runs execute to completion. Reference:
https://docs.microsoft.com/en-us/python/api/azureml-train-core/azureml.train.hyperdrive.medianstoppingpolicy https://docs.microsoft.com/en-us/python/api/azureml-train-core/azureml.train.hyperdrive.truncationselectionpoli https://docs.microsoft.com/en-us/python/api/azureml-train-core/azureml.train.hyperdrive.banditpolicy

NEW QUESTION 16
- (Exam Topic 3)
You are running a training experiment on remote compute in Azure Machine Learning.
The experiment is configured to use a conda environment that includes the mlflow and azureml-contrib-run packages.
You must use MLflow as the logging package for tracking metrics generated in the experiment. You need to complete the script for the experiment.
How should you complete the code? To answer, select the appropriate options in the answer area.
NOTE: Each correct selection is worth one point.

```
import numpy as np
# Import library to log metrics
```

▼

```
from azureml.core import Run
import mlflow
import logging
```

```
# Start logging for this run
```

▼

```
run = Run.get_context()
mlflow.start_run()
logger = logging.getLogger('Run')
```

```
reg_rate = 0.01
# Log the reg_rate metric
```

▼

```
run.log('reg_rate', np.float(reg_rate))
mlflow.log_metric('reg_rate', np.float(reg_rate))
logger.info(np.float(reg_rate))
```

```
# Stop logging for this run
```

▼

```
run.complete()
mlflow.end_run()
logger.setLevel(logging.INFO)
```

A. Mastered
B. Not Mastered

Answer: A

Explanation:
Box 1: import mlflow
Import the mlflow and Workspace classes to access MLflow's tracking URI and configure your workspace. Box 2: mlflow.start_run()
Set the MLflow experiment name with set_experiment() and start your training run with start_run(). Box 3: mlflow.log_metric(' ..')
Use log_metric() to activate the MLflow logging API and begin logging your training run metrics. Box 4: mlflow.end_run()
Close the run: run.endRun() Reference:
https://docs.microsoft.com/en-us/azure/machine-learning/how-to-use-mlflow

NEW QUESTION 21
- (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 22
- (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 overcloking
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 27
- (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 32
- (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 33
- (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":[],
"AssignaleScopes":[
    "/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 36
- (Exam Topic 3)
You are moving a large dataset from Azure Machine Learning Studio to a Weka environment. You need to format the data for the Weka environment.
Which module should you use?

A. Convert to CSV
B. Convert to Dataset
C. Convert to ARFF
D. Convert to SVMLight

Answer: C

Explanation:
Use the Convert to ARFF module in Azure Machine Learning Studio, to convert datasets and results in Azure Machine Learning to the attribute-relation file format used by the Weka toolset. This format is known as ARFF.
The ARFF data specification for Weka supports multiple machine learning tasks, including data preprocessing, classification, and feature selection. In this format, data is organized by entites and their attributes, and is contained in a single text file.
References:
https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/convert-to-arff

NEW QUESTION 39
- (Exam Topic 3)
You deploy a model in Azure Container Instance.
You must use the Azure Machine Learning SDK to call the model API.
You need to invoke the deployed model using native SDK classes and methods.
How should you complete the command? To answer, select the appropriate options in the answer areas.
NOTE: Each correct selection is worth one point.

```
from azureml.core import Workspace

┌─────────────────────────────────────────────────────┬──┐
│                                                       │▼ │
├───────────────────────────────────────────────────────┤
│ from azureml.core.webservice import requests          │
│ from azureml.core.webservice import Webservice         │
│ from azureml.core.webservice import LocalWebservice    │
└───────────────────────────────────────────────────────┘

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

┌─────────────────────────────────────────────────────┬──┐
│                                                       │▼ │
├───────────────────────────────────────────────────────┤
│ predictions = service.run(input_json)                  │
│ predictions = requests.post(service.scoring_uri, input_json) │
│ predictions = service.deserialize(ws, input_json)      │
└───────────────────────────────────────────────────────┘
```

A. Mastered
B. Not Mastered

Answer: A

Explanation:
Box 1: from azureml.core.webservice import Webservice
The following code shows how to use the SDK to update the model, environment, and entry script for a web service to Azure Container Instances:
from azureml.core import Environment
from azureml.core.webservice import Webservice
from azureml.core.model import Model, InferenceConfig Box 2: predictions = service.run(input_json)
Example: The following code demonstrates sending data to the service: import json
test_sample = json.dumps({'data': [[1, 2, 3, 4, 5, 6, 7, 8, 9, 10],
[10, 9, 8, 7, 6, 5, 4, 3, 2, 1]
]})
test_sample = bytes(test_sample, encoding='utf8') prediction = service.run(input_data=test_sample)
print(prediction) Reference:
https://docs.microsoft.com/bs-latn-ba/azure/machine-learning/how-to-deploy-azure-container-instance https://docs.microsoft.com/en-us/azure/machine-learning/how-to-troubleshoot-deployment

NEW QUESTION 44
- (Exam Topic 3)

You are using the Azure Machine Learning Service to automate hyperparameter exploration of your neural network classification model.

You must define the hyperparameter space to automatically tune hyperparameters using random sampling according to following requirements:

» The learning rate must be selected from a normal distribution with a mean value of 10 and a standard deviation of 3.

» Batch size must be 16, 32 and 64.

» Keep probability must be a value selected from a uniform distribution between the range of 0.05 and 0.1.

You need to use the param_sampling method of the Python API for the Azure Machine Learning Service. 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.train.hyperdrive import RandomParameterSampling
param_sampling = RandomParameterSampling( {

"learning_rate" :        [▼] ,
                  uniform(10,3)
                  normal(10,3)
                  choice(10,3)
                  Loguniform(10,3)

"batch_size":            [▼] ,
                  choice(16,32,64)
                  choice(range(16,64))
                  normal(16,32,64)
                  normal(range(16,64))

"keep_probability" :     [▼]
                  choice(range(0.05, 0.1))
                  uniform(0.05, 0.1)
                  normal(0.05, 0.1)
                  lognormal(0.05, 0.1)

  }
)
```

A. Mastered
B. Not Mastered

Answer: A

Explanation:
In random sampling, hyperparameter values are randomly selected from the defined search space. Random sampling allows the search space to include both discrete and continuous hyperparameters.
Example:
from azureml.train.hyperdrive import RandomParameterSampling param_sampling = RandomParameterSampling({ "learning_rate": normal(10, 3),
"keep_probability": uniform(0.05, 0.1),
"batch_size": choice(16, 32, 64)
}
Reference:
https://docs.microsoft.com/en-us/azure/machine-learning/service/how-to-tune-hyperparameters

NEW QUESTION 48
- (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 49
- (Exam Topic 3)
Note: This question is part of a series of questions that present the same scenario. Each question in the series contains a unique solution that might meet the stated goals. Some question sets might have more than one correct solution, while others might not have a correct solution.
After you answer a question in this section, you will NOT be able to return to it. As a result, these questions will not appear in the review screen.

You are using Azure Machine Learning to run an experiment that trains a classification model.
You want to use Hyperdrive to find parameters that optimize the AUC metric for the model. You configure a HyperDriveConfig for the experiment by running the following code:

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

You plan to use this configuration to run a script that trains a random forest model and then tests it with validation data. The label values for the validation data are stored in a variable named y_test variable, and the predicted probabilities from the model are stored in a variable named y_predicted.
You need to add logging to the script to allow Hyperdrive to optimize hyperparameters for the AUC metric. Solution: Run the following code:

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

Does the solution meet the goal?

A. Yes
B. No

Answer: B

Explanation:
Use a solution with logging.info(message) instead. Note: Python printing/logging example: logging.info(message)
Destination: Driver logs, Azure Machine Learning designer Reference:
https://docs.microsoft.com/en-us/azure/machine-learning/how-to-debug-pipelines

NEW QUESTION 54
- (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 58
- (Exam Topic 3)
You are conducting feature engineering to prepuce data for further analysis. The data includes seasonal patterns on inventory requirements.
You need to select the appropriate method to conduct feature engineering on the data. Which method should you use?

A. Exponential Smoothing (ETS) function.
B. One Class Support Vector Machine module
C. Time Series Anomaly Detection module

D. Finite Impulse Response (FIR) Filter module.

Answer: D

NEW QUESTION 61
- (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 65
- (Exam Topic 3)
You have a dataset that includes home sales data for a city. The dataset includes the following columns.

Name	Description
Price	The sales price for the house.
Bedrooms	The number of bedrooms in the house.
Size	The size of the house in square feet.
HasGarage	A binary value indicating whether or not the house has a garage.
HomeType	The category of home, for example, apartment, townhouse, single-family home.

Each row in the dataset corresponds to an individual home sales transaction.
You need to use automated machine learning to generate the best model for predicting the sales price based on the features of the house.
Which values should you use? To answer, select the appropriate options in the answer area. NOTE: Each correct selection is worth one point.

Setting   Value

Prediction task
Classification
Forecasting
Regression
Outlier

Target column
Price
Bedrooms
Size
HasGarage
HomeType

A. Mastered
B. Not Mastered

Answer: A

Explanation:
Box 1: Regression
Regression is a supervised machine learning technique used to predict numeric values. Box 2: Price
Reference:
https://docs.microsoft.com/en-us/learn/modules/create-regression-model-azure-machine-learning-designer

NEW QUESTION 67
- (Exam Topic 3)
You create a binary classification model by using Azure Machine Learning Studio.
You must tune hyperparameters by performing a parameter sweep of the model. The parameter sweep must meet the following requirements:

› iterate all possible combinations of hyperparameters

› minimize computing resources required to perform the sweep

› You need to perform a parameter sweep of the model.
Which parameter sweep mode should you use?

A. Random sweep
B. Sweep clustering

C. Entire grid
D. Random grid
E. Random seed

Answer: D

Explanation:
Maximum number of runs on random grid: This option also controls the number of iterations over a random sampling of parameter values, but the values are not generated randomly from the specified range; instead, a matrix is created of all possible combinations of parameter values and a random sampling is taken over the matrix. This method is more efficient and less prone to regional oversampling or undersampling.
If you are training a model that supports an integrated parameter sweep, you can also set a range of seed values to use and iterate over the random seeds as well. This is optional, but can be useful for avoiding bias introduced by seed selection.

NEW QUESTION 69
- (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 70
- (Exam Topic 3)
You are building a binary classification model by using a supplied training set. The training set is imbalanced between two classes.
You need to resolve the data imbalance.
What are three possible ways to achieve this goal? Each correct answer presents a complete solution NOTE: Each correct selection is worth one point.

A. Penalize the classification
B. Resample the data set using under sampling or oversampling
C. Generate synthetic samples in the minority class.
D. Use accuracy as the evaluation metric of the model.
E. Normalize the training feature set.

Answer: ABD

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

NEW QUESTION 75
- (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 79
- (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. ScripcRunConfig
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)

NEW QUESTION 83
- (Exam Topic 3)
You create a datastore named training_data that references a blob container in an Azure Storage account. The blob container contains a folder named csv_files in which multiple comma-separated values (CSV) files are stored.
You have a script named train.py in a local folder named ./script that you plan to run as an experiment using an estimator. The script includes the following code to read data from the csv_files folder:

```
import os
import argparse
import pandas as pd

from sklearn.model_selection import train_test_split
from sklearn.linear_model import LogisticRegression
from azureml.core import Run

run = Run.get_context()
parser = argparse.ArgumentParser()
parser.add_argument('--data-folder', type=str, dest='data_folder', help='data reference')
args = parser.parse_args()

data_folder = args.data_folder
csv_files = os.listdir(data_folder)
training_data = pd.concat((pd.read_csv(os.path.join(data_folder,csv_file)) for csv_file in csv_files))

# Code goes on to split the training data and train a logistic regression model
```

You have the following script.

```
from azureml.core import Workspace, Datastore, Experiment
from azureml.train.sklearn import SKLearn

ws = Workspace.from_config()
exp = Experiment(workspace=ws, name='csv_training')
ds = Datastore.get(ws, datastore_name='training_data')
data_ref = ds.path('csv_files')

# Code to define estimator goes here

run = exp.submit(config=estimator)
run.wait_for_completion(show_output=True)
```

You need to configure the estimator for the experiment so that the script can read the data from a data reference named data_ref that references the csv_files folder in the training_data datastore.

Which code should you use to configure the estimator?

```
A. estimator = SKLearn(source_directory='./script',
       inputs=[data_ref.as_named_input('data-folder').to_pandas_dataframe()],
       compute_target='local',
       entry_script='train.py')

B. script_params = {
        '--data-folder': data_ref.as_mount()
   }
   estimator = SKLearn(source_directory='./script',
    script_params=script_params,
    compute_target='local',
    entry_script='train.py'

C. estimator = SKLearn(source_directory='./script',
       inputs=[data_ref.as_named_input('data-folder').as_mount()],
       compute_target='local',
       entry_script='train.py')

D. script_params = {
        '--data-folder': data_ref.as_download(path_on_compute='csv_files')
   }
   estimator = SKLearn(source_directory='./script',
    script_params=script_params,
    compute_target='local',
    entry_script='train.py'

E. estimator = SKLearn(source_directory='./script',
       inputs=[data_ref.as_named_input('data-folder').as_download(path_on_compute='csv_files')],
       compute_target='local',
       entry_script='train.py')
```

A. Option A
B. Option B
C. Option C
D. Option D
E. Option E

Answer: B

Explanation:
Besides passing the dataset through the inputs parameter in the estimator, you can also pass the dataset through script_params and get the data path (mounting point) in your training script via arguments. This way, you can keep your training script independent of azureml-sdk. In other words, you will be able use the same training script for local debugging and remote training on any cloud platform.
Example:
from azureml.train.sklearn import SKLearn script_params = {
# mount the dataset on the remote compute and pass the mounted path as an argument to the training script '--data-folder':
mnist_ds.as_named_input('mnist').as_mount(),
'--regularization': 0.5
}
est = SKLearn(source_directory=script_folder, script_params=script_params, compute_target=compute_target, environment_definition=env, entry_script='train_mnist.py')
# Run the experiment
run = experiment.submit(est) run.wait_for_completion(show_output=True) Reference:
https://docs.microsoft.com/es-es/azure/machine-learning/how-to-train-with-datasets

NEW QUESTION 84
- (Exam Topic 3)
You use Azure Machine Learning to train and register a model.
You must deploy the model into production as a real-time web service to an inference cluster named service-compute that the IT department has created in the Azure Machine Learning workspace.
Client applications consuming the deployed web service must be authenticated based on their Azure Active Directory service principal.
You need to write a script that uses the Azure Machine Learning SDK to deploy the model. The necessary modules have been imported.

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

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

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

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

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

A. Mastered
B. Not Mastered

Answer: A

Explanation:
Box 1: AksCompute Example:
aks_target = AksCompute(ws,"myaks")
# If deploying to a cluster configured for dev/test, ensure that it was created with enough
# cores and memory to handle this deployment configuration. Note that memory is also used by
# things such as dependencies and AML components.
deployment_config = AksWebservice.deploy_configuration(cpu_cores = 1, memory_gb = 1)
service = Model.deploy(ws, "myservice", [model], inference_config, deployment_config, aks_target) Box 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 87
- (Exam Topic 3)
A coworker registers a datastore in a Machine Learning services workspace by using the following code:

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

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

Answer Area

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

datastore = | Workspace   | .get( | ws          | , ' | demo_datastore    |
            | Datastore   |       | run         |     | demo_datacontainer |
            | Experiment  |       | experiment  |     | demo_account      |
            | Run         |       | log         |     | 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 90

- (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 91

- (Exam Topic 3)
You are performing clustering by using the K-means algorithm. You need to define the possible termination conditions.
Which three conditions can you use? Each correct answer presents a complete solution. NOTE: Each correct selection is worth one point.

A. A fixed number of iterations is executed.
B. The residual sum of squares (RSS) rises above a threshold.
C. The sum of distances between centroids reaches a maximum.
D. The residual sum of squares (RSS) falls below a threshold.
E. Centroids do not change between iterations.

Answer: ADE

Explanation:
 References:
https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/k-means-clustering https://nlp.stanford.edu/IR-book/html/htmledition/k-means-1.html

NEW QUESTION 93

- (Exam Topic 3)
You plan to create a speech recognition deep learning model. The model must support the latest version of Python.
You need to recommend a deep learning framework for speech recognition to include in the Data Science Virtual Machine (DSVM).
What should you recommend?

A. Apache Drill
B. Tensorflow
C. Rattle
D. Weka

Answer: B

Explanation:
TensorFlow is an open source library for numerical computation and large-scale machine learning. It uses Python to provide a convenient front-end API for building applications with the framework
TensorFlow can train and run deep neural networks for handwritten digit classification, image recognition, word embeddings, recurrent neural networks, sequence-to-sequence models for machine translation, natural language processing, and PDE (partial differential equation) based simulations.
References:
https://www.infoworld.com/article/3278008/what-is-tensorflow-the-machine-learning-library-explained.html

NEW QUESTION 95

- (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 98
- (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.

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 103
- (Exam Topic 3)
Note: This question is part of a series of questions that present the same scenario. Each question in the series contains a unique solution that might meet the stated goals. Some question sets might have more than one correct solution, while others might not have a correct solution.
After you answer a question in this section, you will NOT be able to return to it. As a result, these questions will not appear in the review screen.
You have a Python script named train.py in a local folder named scripts. The script trains a regression model by using scikit-learn. The script includes code to load a training data file which is also located in the scripts folder.
You must run the script as an Azure ML experiment on a compute cluster named aml-compute.
You need to configure the run to ensure that the environment includes the required packages for model training. You have instantiated a variable named aml-compute that references the target compute cluster.
Solution: Run the following code:

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

Does the solution meet the goal?

A. Yes
B. No

Answer: B

Explanation:

The scikit-learn estimator provides a simple way of launching a scikit-learn training job on a compute target. It is implemented through the SKLearn class, which can be used to support single-node CPU training.
Example:
from azureml.train.sklearn import SKLearn
}
estimator = SKLearn(source_directory=project_folder, compute_target=compute_target, entry_script='train_iris.py'
)
Reference:
https://docs.microsoft.com/en-us/azure/machine-learning/how-to-train-scikit-learn

NEW QUESTION 107
- (Exam Topic 3)
You register a file dataset named csvjolder that references a folder. The folder includes multiple com
ma-separated values (CSV) files in an Azure storage blob container. You plan to use the following code to run a script that loads data from the file dataset. You create and instantiate the following variables:

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

You have the following code:

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

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

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

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

B)

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

C)

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

D)

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

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

Answer: D

Explanation:
 Example:
from azureml.train.estimator import Estimator script_params = {
# to mount files referenced by mnist dataset
'--data-folder': mnist_file_dataset.as_named_input('mnist_opendataset').as_mount(),
'--regularization': 0.5
}
est = Estimator(source_directory=script_folder, script_params=script_params, compute_target=compute_target, environment_definition=env,
entry_script='train.py')
Reference:
https://docs.microsoft.com/en-us/azure/machine-learning/tutorial-train-models-with-aml

NEW QUESTION 108
- (Exam Topic 3)
Note: This question is part of a series of questions that present the same scenario. Each question in the series contains a unique solution that might meet the stated goals. Some question sets might have more than one correct solution, while others might not have a correct solution.
After you answer a question in this section, you will NOT be able to return to it. As a result, these questions will not appear in the review screen.
You are creating a new experiment in Azure Learning learning Studio.
One class has a much smaller number of observations than the other classes in the training
You need to select an appropriate data sampling strategy to compensate for the class imbalance. Solution: You use the Synthetic Minority Oversampling Technique (SMOTE) sampling mode. Does the solution meet the goal?

A. Yes
B. No

Answer: A

Explanation:
SMOTE is used to increase the number of underrepresented cases in a dataset used for machine learning. SMOTE is a better way of increasing the number of rare cases than simply duplicating existing cases.
References:
https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/smote

NEW QUESTION 112
- (Exam Topic 3)
You plan to use Hyperdrive to optimize the hyperparameters selected when training a model. You create the following code to define options for the hyperparameter experiment

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

param_sampling = GridParameterSampling({
    "max_depth" : pe.choice(6, 7, 8, 9),
    "learning_rate" : pe.choice(0.05, 0.1, 0.15)
    })
hyperdrive_run_config = HyperDriveConfig(
    estimator = estimator,
    hyperparameter_sampling = param_sampling,
    policy = None,
    primary_metric_name = "auc",
    primary_metric_goal = PrimaryMetricGoal.MAXIMIZE,

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

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

Answer Area

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

A. Mastered
B. Not Mastered

Answer: A

Explanation:
Box 1: No
max_total_runs (50 here)
The maximum total number of runs to create. This is the upper bound; there may be fewer runs when the sample space is smaller than this value.
Box 2: Yes
Policy EarlyTerminationPolicy
The early termination policy to use. If None - the default, no early termination policy will be used. Box 3: No
Discrete hyperparameters are specified as a choice among discrete values. choice can be: one or more comma-separated values
> a range object
> any arbitrary list object Reference:
https://docs.microsoft.com/en-us/python/api/azureml-train-core/azureml.train.hyperdrive.hyperdriveconfig https://docs.microsoft.com/en-us/azure/machine-learning/how-to-tune-hyperparameters

NEW QUESTION 116
- (Exam Topic 3)
You create a multi-class image classification deep learning model that uses the PyTorch deep learning framework.
You must configure Azure Machine Learning Hyperdrive to optimize the hyperparameters for the classification model.
You need to define a primary metric to determine the hyperparameter values that result in the model with the best accuracy score.
Which three actions must you perform? Each correct answer presents part of the solution.

NOTE: Each correct selection is worth one point.

A. Set the primary_metric_goal of the estimator used to run the bird_classifier_train.py script to maximize.
B. Add code to the bird_classifier_train.py script to calculate the validation loss of the model and log it as a float value with the key loss.
C. Set the primary_metric_goal of the estimator used to run the bird_classifier_train.py script to minimize.
D. Set the primary_metric_name of the estimator used to run the bird_classifier_train.py script to accuracy.
E. Set the primary_metric_name of the estimator used to run the bird_classifier_train.py script to loss.
F. Add code to the bird_classifier_train.py script to calculate the validation accuracy of the model and log it as a float value with the key accuracy.

Answer: ADF

Explanation:
AD:
primary_metric_name="accuracy", primary_metric_goal=PrimaryMetricGoal.MAXIMIZE
Optimize the runs to maximize "accuracy". Make sure to log this value in your training script. Note:
primary_metric_name: The name of the primary metric to optimize. The name of the primary metric needs to exactly match the name of the metric logged by the training script.
primary_metric_goal: It can be either PrimaryMetricGoal.MAXIMIZE or PrimaryMetricGoal.MINIMIZE and determines whether the primary metric will be maximized or minimized when evaluating the runs.
F: The training script calculates the val_accuracy and logs it as "accuracy", which is used as the primary metric.

NEW QUESTION 120
- (Exam Topic 3)
You train and register a machine learning model. You create a batch inference pipeline that uses the model to generate predictions from multiple data files.
You must publish the batch inference pipeline as a service that can be scheduled to run every night. You need to select an appropriate compute target for the inference service.
Which compute target should you use?

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

Answer: B

Explanation:
Azure Machine Learning compute clusters is used for Batch inference. Run batch scoring on serverless compute. Supports normal and low-priority VMs. No support for real-time inference.
Reference:
https://docs.microsoft.com/en-us/azure/machine-learning/concept-compute-target

NEW QUESTION 125
- (Exam Topic 3)
You have a dataset created for multiclass classification tasks that contains a normalized numerical feature set with 10,000 data points and 150 features.
You use 75 percent of the data points for training and 25 percent for testing. You are using the scikit-learn machine learning library in Python. You use X to denote the feature set and Y to denote class labels.
You create the following Python data frames:

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

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

```
from sklearn.decomposition import PCA
pca = ▼
        ┌──────────────────────────────┐
        │ PCA()                        │
        │ PCA(n_components = 150)       │
        │ PCA(n_components = 10)        │
        │ PCA(n_components = 10000)     │
        └──────────────────────────────┘

X_train=              ▼    .fit_transform(X_train)
        ┌──────────────────────────────┐
        │ pca                          │
        │ model                        │
        │ sklearn.decomposition        │
        └──────────────────────────────┘

x_test = pca.        ▼
        ┌──────────────────────────────┐
        │ x_test                       │
        │ X_train                      │
        │ fit(x_test)                  │
        │ transform(x_test)            │
        └──────────────────────────────┘
```

A. Mastered
B. Not Mastered

Answer: A

Explanation:
Box 1: PCA(n_components = 10)
Need to reduce the dimensionality of the feature set to 10 features in both training and testing sets. Example:
from sklearn.decomposition import PCA pca = PCA(n_components=2) ;2 dimensions principalComponents = pca.fit_transform(x)
Box 2: pca
fit_transform(X[, y])fits the model with X and apply the dimensionality reduction on X. Box 3: transform(x_test)
transform(X) applies dimensionality reduction to X. References:
https://scikit-learn.org/stable/modules/generated/sklearn.decomposition.PCA.html

NEW QUESTION 130
- (Exam Topic 3)
You are implementing a machine learning model to predict stock prices. The model uses a PostgreSQL database and requires GPU processing.
You need to create a virtual machine that is pre-configured with the required tools. What should you do?

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

Answer: E

NEW QUESTION 134
- (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.

A. Mastered
B. Not Mastered

Answer: A

Explanation:
Box 1: register_azure_blob_container
Register an Azure Blob Container to the datastore.
Box 2: create_if_not_exists = False
Create the file share if it does not exists, defaults to False. Reference:
https://docs.microsoft.com/en-us/python/api/azureml-core/azureml.core.datastore.datastore

NEW QUESTION 136
- (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 139
- (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 142
- (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.
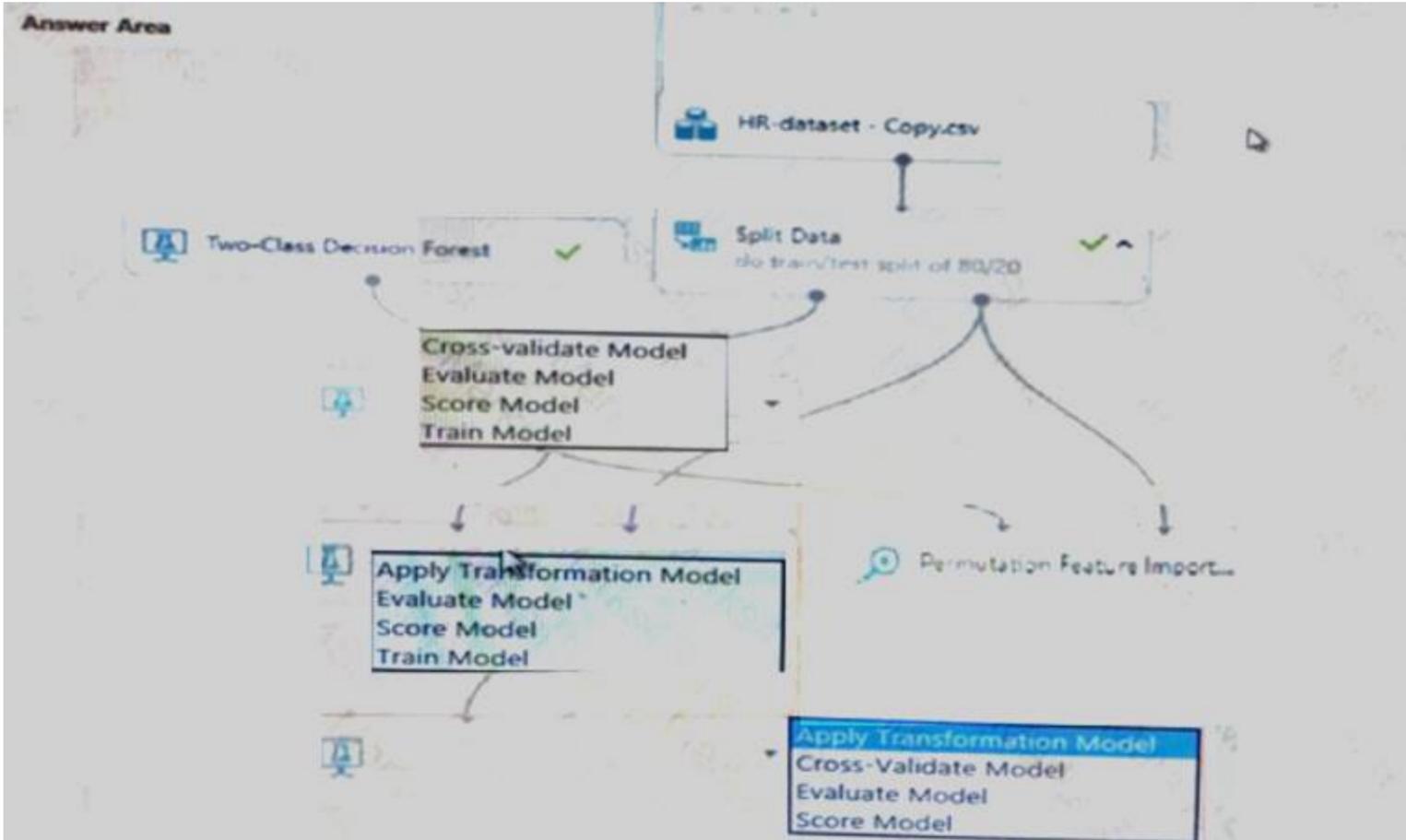
A. Mastered
B. Not Mastered

Answer: A

Explanation:
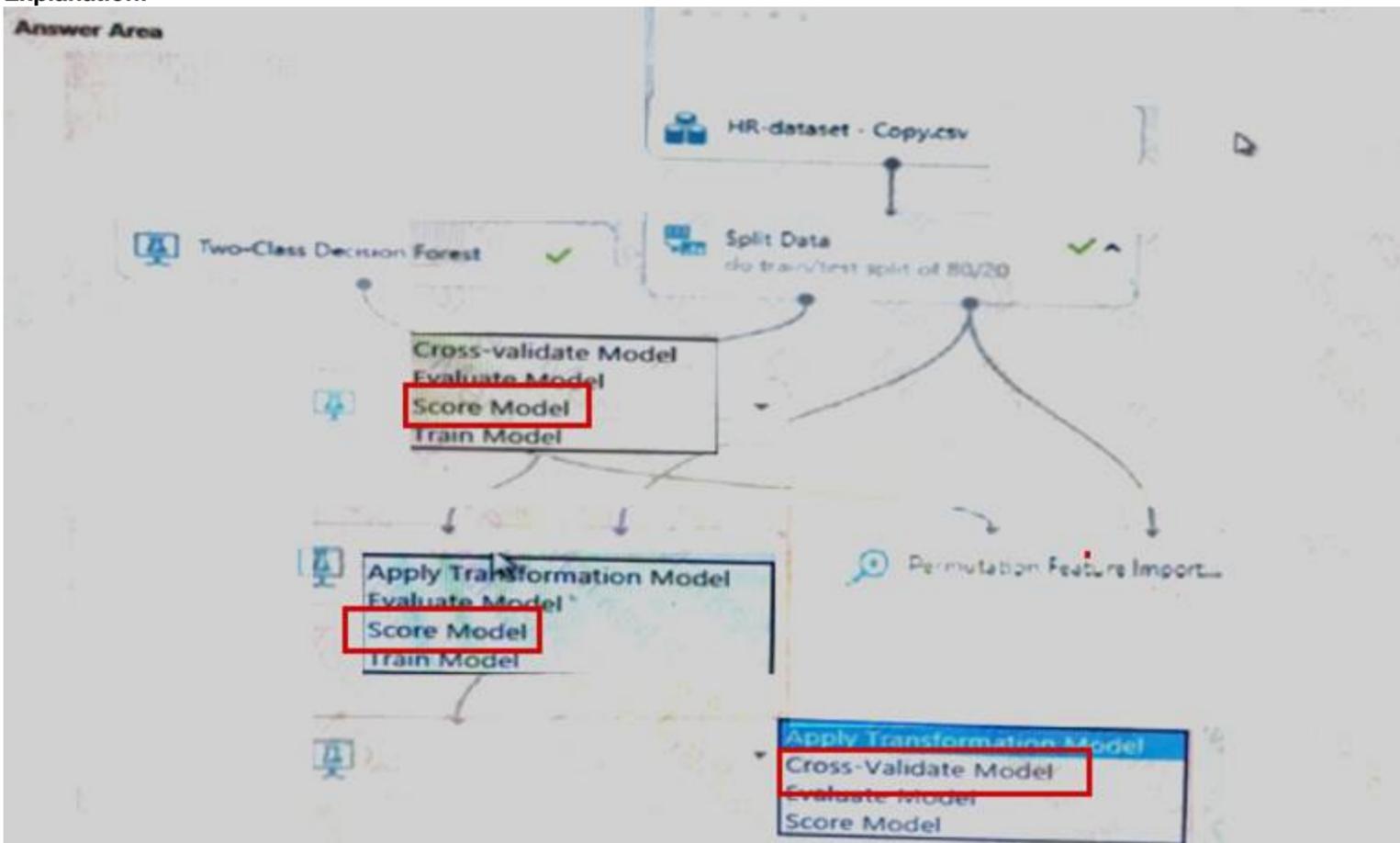
NEW QUESTION 144
- (Exam Topic 3)
You plan to explore demographic data for home ownership in various cities. The data is in a CSV file with the following format:
age,city,income,home_owner 21,Chicago,50000,0 35,Seattle,120000,1 23,Seattle,65000,0 45,Seattle,130000,1 18,Chicago,48000,0
You need to run 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

```python
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 148

- (Exam Topic 3)
You deploy a model as an Azure Machine Learning real-time web service using the following code.

```python
# ws, model, interference_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 149

- (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,
    ┌─────────────────────────────┐        experiment_name='train_and_explain',
    │ from_run                     │        run_id='train_and_explain_12345')
    │ list_model_explanations      │
    │ 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 150
- (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 155

- (Exam Topic 3)

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

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

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

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

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

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

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

```
import json, os
from sklearn.metrics import roc_auc_score
# code to train model omitted
auc = roc_auc_score(y_test, y_predicted)
os.makedirs("outputs", exist_ok = True)
with open("outputs/AUC.txt", "w") as file_cur:
    file_cur.write(auc)
```

Does the solution meet the goal?

A. Yes
B. No

Answer: B

Explanation:

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

Destination: Driver logs, Azure Machine Learning designer Reference:

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

NEW QUESTION 160

- (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 161

- (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 163

- (Exam Topic 3)
You use Azure Machine Learning to deploy a model as a real-time web service.
You need to create an entry script for the service that ensures that the model is loaded when the service starts and is used to score new data as it is received.
Which functions should you include in the script? To answer, drag the appropriate functions to the correct actions. Each function may be used once, more than once, or not at all. You may need to drag the split bar between panes or scroll to view content
NOTE: Each correct selection is worth one point.

Functions			Answer Area		
main()	score()	run()		Action	Function
init()	predict()			Load the model when the service starts.	
				Use the model to score new data.	
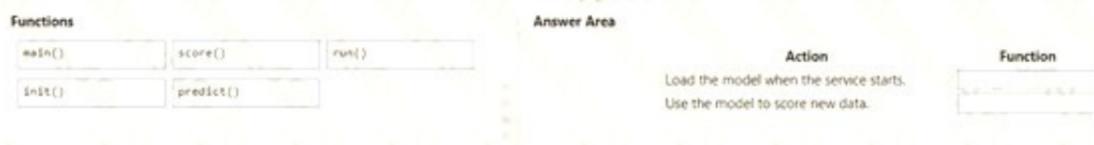
A. Mastered
B. Not Mastered

Answer: A

Explanation:
Box 1: init()
The entry script has only two required functions, init() and run(data). These functions are used to initialize the service at startup and run the model using request data passed in by a client. The rest of the script handles loading and running the model(s).
Box 2: run() Reference:
https://docs.microsoft.com/en-us/azure/machine-learning/how-to-deploy-existing-model

NEW QUESTION 167
- (Exam Topic 3)
You define a datastore named ml-data for an Azure Storage blob container. In the container, you have a folder named train that contains a file named data.csv.
You plan to use the file to train a model by using the Azure Machine Learning SDK.
You plan to train the model by using the Azure Machine Learning SDK to run an experiment on local compute.
You define a DataReference object by running the following code:

```
from azureml.core import Workspace, Datastore, Environment
from azureml.train.estimator import Estimator
ws = Workspace.from_config()
ml_data = = Datastore.get(ws, datastore_name='ml-data')
data_ref = ml_data.path('train').as_download(path_on_compute='train_data')
estimator = Estimator(source_directory='experiment_folder',
  script_params={'--data-folder': data_ref},
  compute_target = 'local',
  entry_script='training.py')
run = experiment.submit(config=estimator)
run.wait_for_completion(show_output=True)
```

You need to load the training data. Which code segment should you use?

```
A. import os
   import argparse
   import pandas as pd

   parser = argparse.ArgumentParser()
   parser.add_argument('--data-folder', type=str, dest='data_folder')
   data_folder = args.data_folder
   data = pd.read_csv(os.path.join(data_folder,'ml-data','train_data','data.csv'))

B. import os
   import argparse
   import pandas as pd

   parser = argparse.ArgumentParser()
   parser.add_argument('--data-folder', type=str, dest='data_folder')
   data_folder = args.data_folder
   data = pd.read_csv(os.path.join(data_folder,'train','data.csv'))

C. import pandas as pd

   data = pd.read_csv('./data.csv')

D. import os
   import argparse
   import pandas as pd

   parser = argparse.ArgumentParser()
   parser.add_argument('--data-folder', type=str, dest='data_folder')
   data_folder = args.data_folder
   data = pd.read_csv(os.path.join('ml_data', data_folder,'data.csv'))

E. import os
   import argparse
   import pandas as pd

   parser = argparse.ArgumentParser()
   parser.add_argument('--data-folder', type=str, dest='data_folder')
   data_folder = args.data_folder
   data = pd.read_csv(os.path.join(data_folder,'data.csv'))
```

A. Option A
B. Option B
C. Option C
D. Option D
E. Option E

Answer: E

Explanation:
 Example:
data_folder = args.data_folder
# Load Train and Test data
train_data = pd.read_csv(os.path.join(data_folder, 'data.csv')) Reference:
https://www.element61.be/en/resource/azure-machine-learning-services-complete-toolbox-ai

NEW QUESTION 171
- (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	Grid search
	Outlier detection
	Dimensionality reduction
Constraint	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 176
- (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 179
- (Exam Topic 3)
You create an experiment in Azure Machine Learning Studio- You add a training dataset that contains 10.000 rows. The first 9.000 rows represent class 0 (90 percent). The first 1.000 rows represent class 1 (10 percent).
The training set is unbalanced between two Classes. You must increase the number of training examples for class 1 to 4,000 by using data rows. You add the Synthetic Minority Oversampling Technique (SMOTE) module to the experiment.
You need to configure the module.
Which values should you use? To answer, select the appropriate options in the dialog box in the answer area. NOTE: Each correct selection is worth one point.
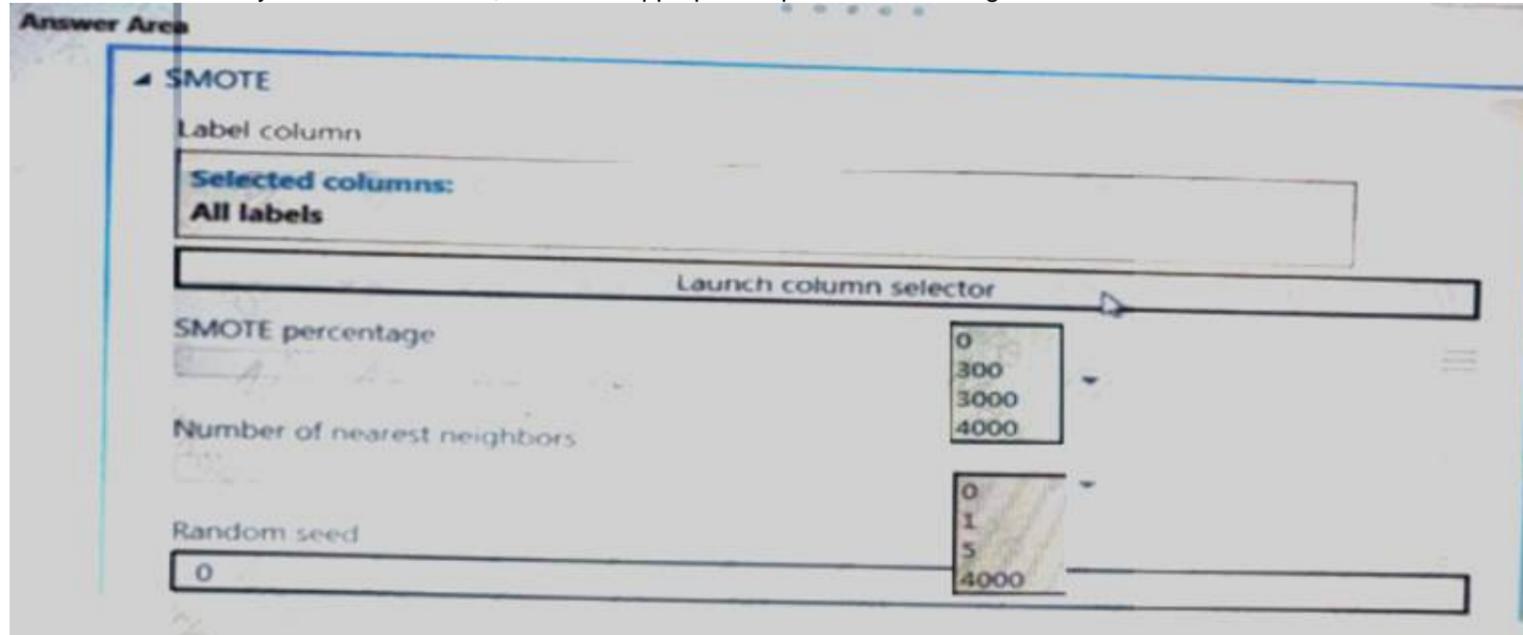
A. Mastered
B. Not Mastered

Answer: A

Explanation:
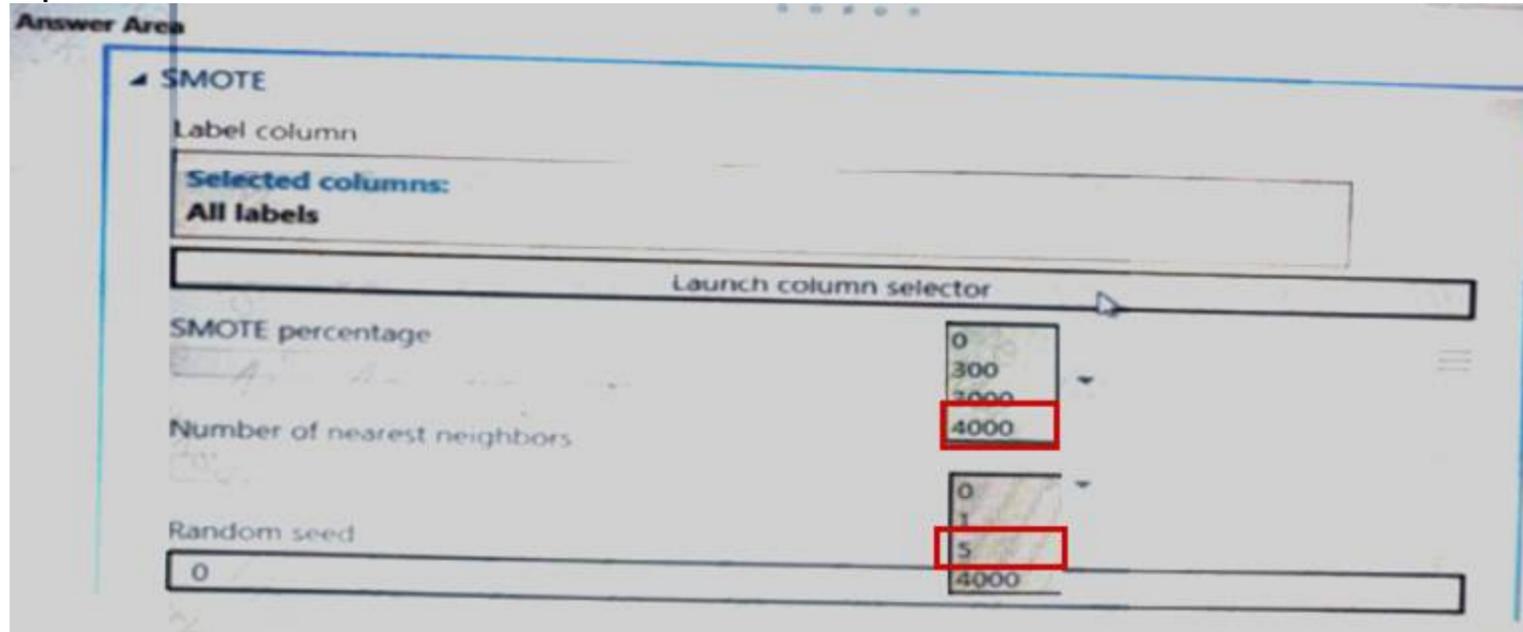
NEW QUESTION 181
- (Exam Topic 3)
Note: This question is part of a series of questions that present the same scenario. Each question in the series contains a unique solution that might meet the stated goals. Some question sets might have more than one correct solution, while others might not have a correct solution.
After you answer a question in this section, you will NOT be able to return to it. As a result, these questions will not appear in the review screen.
You are creating a new experiment in Azure Machine Learning Studio.
One class has a much smaller number of observations than the other classes in the training set. You need to select an appropriate data sampling strategy to compensate for the class imbalance. Solution: You use the Scale and Reduce sampling mode.
Does the solution meet the goal?

A. Yes
B. No

Answer: B

Explanation:
Instead use the Synthetic Minority Oversampling Technique (SMOTE) sampling mode.
Note: SMOTE is used to increase the number of underepresented cases in a dataset used for machine learning. SMOTE is a better way of increasing the number

of rare cases than simply duplicating existing cases.
References:
https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/smote

NEW QUESTION 184
- (Exam Topic 3)
You are analyzing the asymmetry in a statistical distribution.
The following image contains two density curves that show the probability distribution of two datasets.
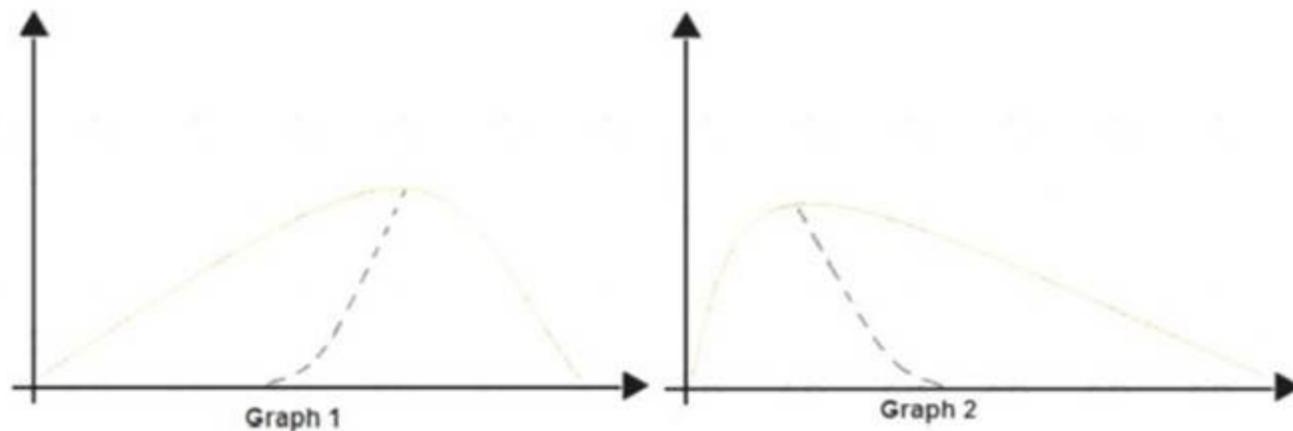
Graph 1   Graph 2

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.

Question	Answer choice
Which type of distribution is shown for the dataset density curve of Graph 1?	▼ Negative skew / Positive skew / Normal distribution / Bimodal distribution
Which type of distribution is shown for the dataset density curve of Graph 2?	▼ Negative skew / Positive skew / Normal distribution / Bimodal distribution

A. Mastered
B. Not Mastered

Answer: A

Explanation:
Box 1: Positive skew
Positive skew values means the distribution is skewed to the right. Box 2: Negative skew
Negative skewness values mean the distribution is skewed to the left. References:
https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/compute-elementary-statistic

NEW QUESTION 186
- (Exam Topic 3)
You are tuning a hyperparameter for an algorithm. The following table shows a data set with different hyperparameter, training error, and validation errors.

Hyperparameter (H)	Training error (TE)	Validation error (VE)
1	105	95
2	200	85
3	250	100
4	105	100
5	400	50

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

Question	Answer Choise
Which H value should you select based on the data?	1 2 3 4 5
What H value displays the poorest training result?	1 2 3 4 5

A. Mastered
B. Not Mastered

Answer: A

Explanation:
Box 1: 4
Choose the one which has lower training and validation error and also the closest match. Minimize variance (difference between validation error and train error).
Box 2: 5
Minimize variance (difference between validation error and train error). Reference:
https://medium.com/comet-ml/organizing-machine-learning-projects-project-management-guidelines-2d2b8565

NEW QUESTION 187
- (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 191
- (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.	☒	○
The compute resource created by the code is displayed as a compute cluster in Azure Machine Learning studio.	○	○
The minimum number of nodes will be zero.	○	○
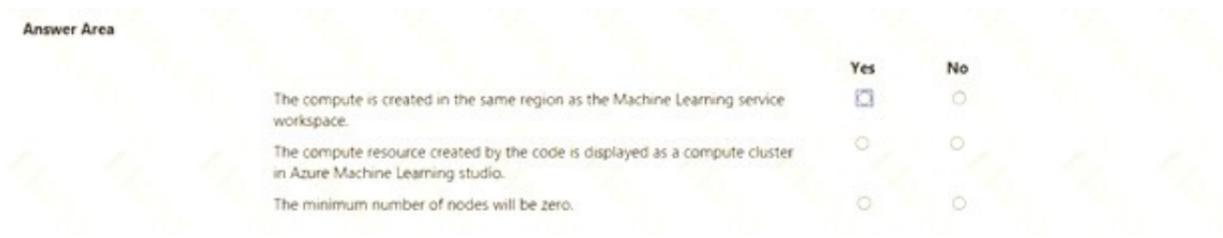
A. Mastered
B. Not Mastered

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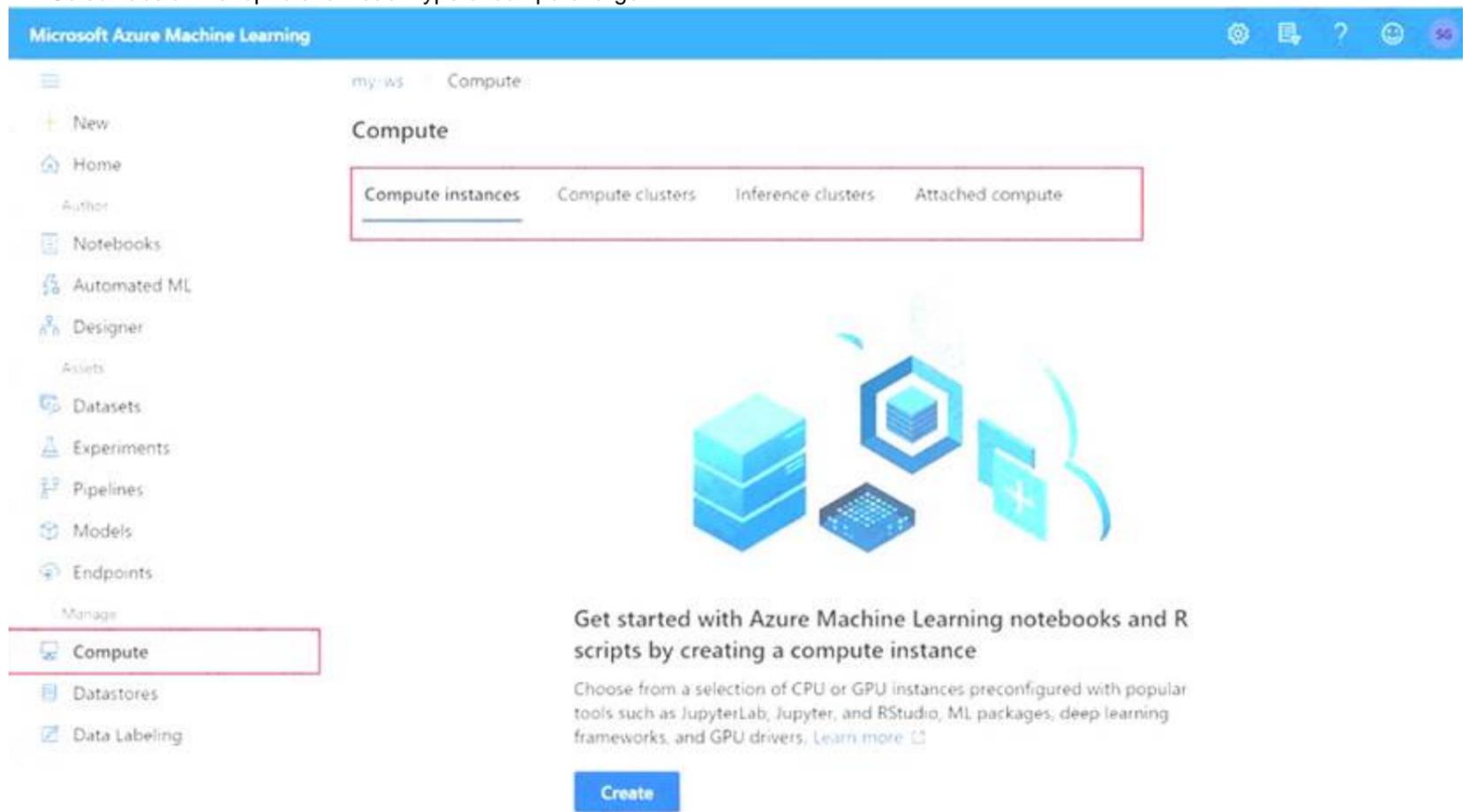
Box 3: Yes
min_nodes is not specified, so it defaults to 0. Reference:
https://docs.microsoft.com/en-us/python/api/azureml-core/azureml.core.compute.amlcompute.amlcomputeprovi https://docs.microsoft.com/en-us/azure/machine-learning/how-to-create-attach-compute-studio

NEW QUESTION 196
- (Exam Topic 3)
Note: This question is part of a series of questions that present the same scenario. Each question in the series contains a unique solution that might meet the stated goals. Some question sets might have more than one correct solution, while others might not have a correct solution.
After you answer a question in this section, you will NOT be able to return to it. As a result, these questions will not appear in the review screen.
You plan to use a Python script to run an Azure Machine Learning experiment. The script creates a reference to the experiment run context, loads data from a file, identifies the set of unique values for the label column, and completes the experiment run:
from azureml.core import Run import pandas as pd
run = Run.get_context()
data = pd.read_csv('data.csv') label_vals = data['label'].unique()
# Add code to record metrics here run.complete()
The experiment must record the unique labels in the data as metrics for the run that can be reviewed later.
You must add code to the script to record the unique label values as run metrics at the point indicated by the comment.
Solution: Replace the comment with the following code: for label_val in label_vals:
run.log('Label Values', label_val) Does the solution meet the goal?

A. Yes
B. No

Answer: A

Explanation:
The run_log function is used to log the contents in label_vals: for label_val in label_vals:
run.log('Label Values', label_val) Reference:
https://www.element61.be/en/resource/azure-machine-learning-services-complete-toolbox-ai

NEW QUESTION 201
- (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 206
- (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 CROC) 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 207
- (Exam Topic 3)
You configure a Deep Learning Virtual Machine for Windows.
You need to recommend tools and frameworks to perform the following:

≫ Build deep neural network (DNN) models

≫ Perform interactive data exploration and visualization
Which tools and frameworks should you recommend? To answer, drag the appropriate tools to the correct tasks. Each tool may be used once, more than once, or not at all. You may need to drag the split bar between panes or scroll to view content.
NOTE: Each correct selection is worth one point.

A. Mastered
B. Not Mastered

Answer: A

Explanation:
Box 1: Vowpal Wabbit
Use the Train Vowpal Wabbit Version 8 module in Azure Machine Learning Studio (classic), to create a machine learning model by using Vowpal Wabbit.
Box 2: PowerBI Desktop
Power BI Desktop is a powerful visual data exploration and interactive reporting tool
BI is a name given to a modern approach to business decision making in which users are empowered to find, explore, and share insights from data across the enterprise.
References:
https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/train-vowpal-wabbit-version- https://docs.microsoft.com/en-us/azure/architecture/data-guide/scenarios/interactive-data-exploration

NEW QUESTION 212

- (Exam Topic 3)
You are creating a machine learning model in Python. The provided dataset contains several numerical columns and one text column. The text column represents a product's category. The product category will always be one of the following:

≫ Bikes

≫ Cars

≫ Vans

≫ Boats

You are building a regression model using the scikit-learn Python package.
You need to transform the text data to be compatible with the scikit-learn Python package.
How should you complete the code segment? To answer, select the appropriate options in the answer area.
NOTE: Each correct selection is worth one point.

```
from sklearn import linear_model
import              ▼
        ┌──────────────────────┐
        │ pandas as df         │
        │ numpy as df          │
        │ scipy as df          │
        └──────────────────────┘
dataset = df.read_csv("data\\ProductSales.csv")
ProductCategoryMapping = {"Bikes":1, "Cars":2, "Boats": 3,
"Vans": 4}
dataset['ProductCategoryMapping'] =
dataset['ProductCategory'].              ▼
                    ┌──────────────────────────────────┐
                    │ map[ProductCategoryMapping]       │
                    │ reduce[ProductCategoryMapping]    │
                    │ transpose[ProductCategoryMapping] │
                    └──────────────────────────────────┘
regr = linear_model.LinearRegression()
X_train = dataset[['ProductCategoryMapping', 'ProductSize',
'ProductCost']]
y_train = dataset[['Sales']]
regr.fit(X_train, y_train)
```

A. Mastered
B. Not Mastered

Answer: A

Explanation:
Box 1: pandas as df
Pandas takes data (like a CSV or TSV file, or a SQL database) and creates a Python object with rows and columns called data frame that looks very similar to table in a statistical software (think Excel or SPSS for example.
Box 2: transpose[ProductCategoryMapping] Reshape the data from the pandas Series to columns. Reference:
https://datascienceplus.com/linear-regression-in-python/

NEW QUESTION 215
- (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 220
- (Exam Topic 3)
Note: This question is part of a series of questions that present the same scenario. Each question in the series contains a unique solution that might meet the stated goals. Some question sets might have more
than one correct solution, while others might not have a correct solution.
After you answer a question in this section, you will NOT be able to return to it. As a result, these questions will not appear in the review screen.
You train a classification model by using a logistic regression algorithm.
You must be able to explain the model's predictions by calculating the importance of each feature, both as an overall global relative importance value and as a measure of local importance for a specific set of predictions.
You need to create an explainer that you can use to retrieve the required global and local feature importance values.
Solution: Create a TabularExplainer. Does the solution meet the goal?

A. Yes
B. No

Answer: B

Explanation:
Instead use Permutation Feature Importance Explainer (PFI). Note 1:
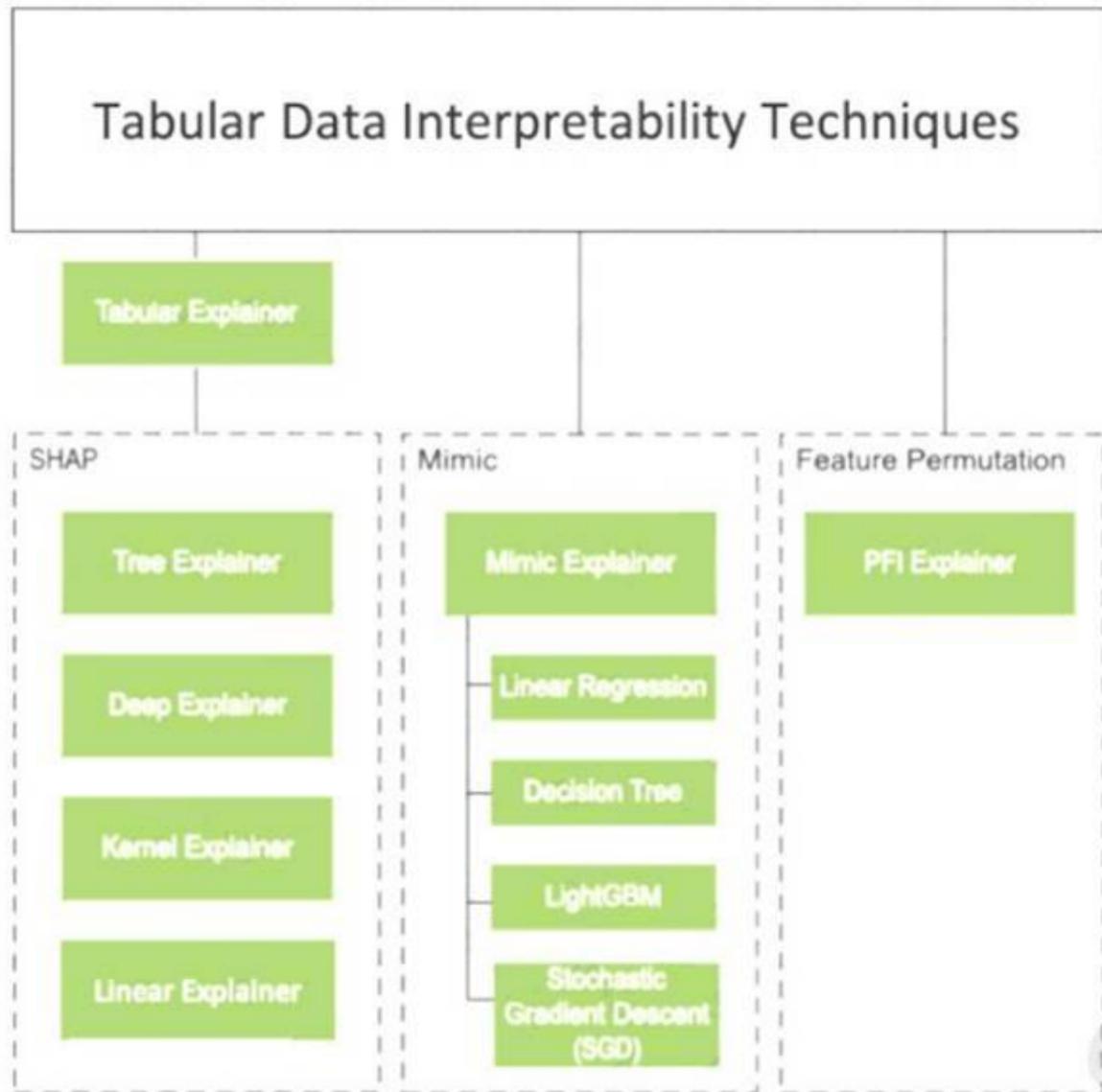
Note 2: Permutation Feature Importance Explainer (PFI): Permutation Feature Importance is a technique used to explain classification and regression models. At a high level, the way it works is by randomly shuffling data one feature at a time for the entire dataset and calculating how much the performance metric of interest changes. The larger the change, the more important that feature is. PFI can explain the overall behavior of any underlying model but does not explain individual predictions.
Reference:
https://docs.microsoft.com/en-us/azure/machine-learning/how-to-machine-learning-interpretability

NEW QUESTION 221
- (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 225
- (Exam Topic 3)
You train a classification model by using a decision tree algorithm.
You create an estimator by running the following Python code. The variable feature_names is a list of all feature names, and class_names is a list of all class names.
from interpret.ext.blackbox import TabularExplainer

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

You need to explain the predictions made by the model for all classes by determining the importance of all features.
For each of the following statements, select Yes if the statement is true. Otherwise, select No.
NOTE: Each correct selection is worth one point.

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

A. Mastered
B. Not Mastered

Answer: A

Explanation:
s and visualizations more informative, you can choose to pass in feature names and output class names if doing classification.
Box 3: No
TabularExplainer automatically selects the most appropriate one for your use case, but you can call each of its three underlying explainers underneath (TreeExplainer, DeepExplainer, or KernelExplainer) directly.
Reference:
https://docs.microsoft.com/en-us/azure/machine-learning/how-to-machine-learning-interpretability-aml

NEW QUESTION 227
- (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 230
- (Exam Topic 3)
You have the following code. The code prepares an experiment to run a script:

```
from azureml.core import Workspace, Experiment, Run, ScriptRunConfig

ws = Workspace.from_config()
script_config = ScriptRunConfig(source_directory='experiment_files',
                                script='experiment.py')

script_experiment = Experiment(workspace=ws, name='script-experiment')
```

The experiment must be run on local computer using the default environment. You need to add code to start the experiment and run the script.
Which code segment should you use?

A. run = script_experiment.start_logging()
B. run = Run(experiment=script_experiment)
C. ws.get_run(run_id=experiment.id)
D. run = script_experiment.submit(config=script_config)

Answer: D

Explanation:
The experiment class submit method submits an experiment and return the active created run.
Syntax: submit(config, tags=None, **kwargs) Reference:
https://docs.microsoft.com/en-us/python/api/azureml-core/azureml.core.experiment.experiment

NEW QUESTION 234
- (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	Split rows ▼
	Split rows
	Recommender Split
	Regular Expression Split
	Relative Expression Split
Fraction of rows in the first output dataset	▼
	0.75
	0.25
	0.5
	1
Randomized split	▼
	True
	False
Stratified split	▼
	True
	False

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 239
- (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 AciWebservice instance.
Set the value of the ssl_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 243
- (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.

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 248
- (Exam Topic 3)
You have an Azure Machine Learning workspace that contains a CPU-based compute cluster and an Azure Kubernetes Services (AKS) inference cluster. You create a tabular dataset containing data that you plan to use to create a classification model.
You need to use the Azure Machine Learning designer to create a web service through which client applications can consume the classification model by submitting new data and getting an immediate prediction as a response.
Which three actions should you perform in sequence? To answer, move the appropriate actions from the list of actions to the answer area and arrange them in the correct order.

A. Mastered
B. Not Mastered

Answer: A

Explanation:
Step 1: Create and start a Compute Instance
To train and deploy models using Azure Machine Learning designer, you need compute on which to run the training process, test the model, and host the model in a deployed service.
There are four kinds of compute resource you can create:
Compute Instances: Development workstations that data scientists can use to work with data and models. Compute Clusters: Scalable clusters of virtual machines for on-demand processing of experiment code. Inference Clusters: Deployment targets for predictive services that use your trained models.
Attached Compute: Links to existing Azure compute resources, such as Virtual Machines or Azure Databricks clusters.
Step 2: Create and run a training pipeline..
After you've used data transformations to prepare the data, you can use it to train a machine learning model. Create and run a training pipeline
Step 3: Create and run a real-time inference pipeline
After creating and running a pipeline to train the model, you need a second pipeline that performs the same data transformations for new data, and then uses the trained model to inference (in other words, predict) label values based on its features. This pipeline will form the basis for a predictive service that you can publish for applications to use.
Reference:
https://docs.microsoft.com/en-us/learn/modules/create-classification-model-azure-machine-learning-designer/

NEW QUESTION 250
- (Exam Topic 3)
You are a data scientist working for a bank and have used Azure ML to train and register a machine learning model that predicts whether a customer is likely to repay a loan.
You want to understand how your model is making selections and must be sure that the model does not violate government regulations such as denying loans based on where an applicant lives.
You need to determine the extent to which each feature in the customer data is influencing predictions. What should you do?

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

Answer: D

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

NEW QUESTION 255
- (Exam Topic 3)
You plan to preprocess text from CSV files. You load the Azure Machine Learning Studio default stop words list.
You need to configure the Preprocess Text module to meet the following requirements:

➢ Ensure that multiple related words from a single canonical form.

➢ Remove pipe characters from text.

➢ Remove words to optimize information retrieval.

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

◢ Preprocess Text

Language

English

Remove by part of speech

False

Text column to clean

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

Launch column selector

☐ Remove stop words

☐ Lemmatization

☐ Detect sentences

☐ Normalize case to lowercase

☐ Remove numbers

☐ Remove special characters

☐ Remove duplicate characters

☐ Remove email addresses

☐ Remove URLs

☐ Expand verb contractions

☐ Normalize backslashes to slashes

☐ Split tokens on special characters

A. Mastered
B. Not Mastered

Answer: A

Explanation:
Box 1: Remove stop words
Remove words to optimize information retrieval.
Remove stop words: Select this option if you want to apply a predefined stopword list to the text column. Stop word removal is performed before any other processes.
Box 2: Lemmatization
Ensure that multiple related words from a single canonical form. Lemmatization converts multiple related words to a single canonical form Box 3: Remove special characters
Remove special characters: Use this option to replace any non-alphanumeric special characters with the pipe | character.
References:
https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/preprocess-text

NEW QUESTION 257
- (Exam Topic 3)
You have a Jupyter Notebook that contains Python code that is used to train a model.
You must create a Python script for the production deployment. The solution must minimize code maintenance.
Which two actions should you perform? Each correct answer presents part of the solution.
NOTE: Each correct selection is worth one point.

A. Refactor the Jupyter Notebook code into functions
B. Save each function to a separate Python file
C. Define a main() function in the Python script
D. Remove all comments and functions from the Python script

Answer: AC

Explanation:
Reference:
https://www.guru99.com/learn-python-main-function-with-examples-understand-main.html https://towardsdatascience.com/from-jupyter-notebook-to-deployment-a-straightforward-example-1838c203a43

NEW QUESTION 260
- (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 262
- (Exam Topic 3)
You use Azure Machine Learning designer to create a real-time service endpoint. You have a single Azure Machine Learning service compute resource. You train the model and prepare the real-time pipeline for deployment You need to publish the inference pipeline as a web service. Which compute type should you use?

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

Answer: C

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

NEW QUESTION 267
- (Exam Topic 3)
You are creating a machine learning model. You have a dataset that contains null rows.
You need to use the Clean Missing Data module in Azure Machine Learning Studio to identify and resolve the null and missing data in the dataset.
Which parameter should you use?

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

Answer: B

Explanation:
Remove entire row: Completely removes any row in the dataset that has one or more missing values. This is useful if the missing value can be considered randomly missing.
References:
https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/clean-missing-data

NEW QUESTION 269
- (Exam Topic 3)
You create a binary classification model to predict whether a person has a disease. You need to detect possible classification errors.
Which error type should you choose for each description? To answer, select the appropriate options in the answer area.
NOTE: Each correct selection is worth one point.

Description	Error type
A person has a disease. The model classifies the case as having a disease.	▼ True Positives / True Negatives / False Positives / False Negatives
A person does not have a disease. The model classifies the case as having no disease.	▼ True Positives / True Negatives / False Positives / False Negatives
A person does not have a disease. The model classifies the case as having a disease.	▼ True Positives / True Negatives / False Positives / False Negatives
A person has a disease. The model classifies the case as having no disease.	▼ True Positives / True Negatives / False Positives / False Negatives

A. Mastered
B. Not Mastered

Answer: A

Explanation:
Box 1: True Positive
A true positive is an outcome where the model correctly predicts the positive class Box 2: True Negative
A true negative is an outcome where the model correctly predicts the negative class. Box 3: False Positive
A false positive is an outcome where the model incorrectly predicts the positive class. Box 4: False Negative
A false negative is an outcome where the model incorrectly predicts the negative class. Note: Let's make the following definitions:
"Wolf" is a positive class. "No wolf" is a negative class.
We can summarize our "wolf-prediction" model using a 2x2 confusion matrix that depicts all four possible outcomes:
Reference:
https://developers.google.com/machine-learning/crash-course/classification/true-false-positive-negative

NEW QUESTION 274
......

Thank You for Trying Our Product

We offer two products:

1st - We have Practice Tests Software with Actual Exam Questions

2nd - Questons and Answers in PDF Format

DP-100 Practice Exam Features:

* DP-100 Questions and Answers Updated Frequently

* DP-100 Practice Questions Verified by Expert Senior Certified Staff

* DP-100 Most Realistic Questions that Guarantee you a Pass on Your FirstTry

* DP-100 Practice Test Questions in Multiple Choice Formats and Updatesfor 1 Year

100% Actual & Verified — Instant Download, Please Click
[Order The DP-100 Practice Test Here](https://www.certshared.com/exam/DP-100/)