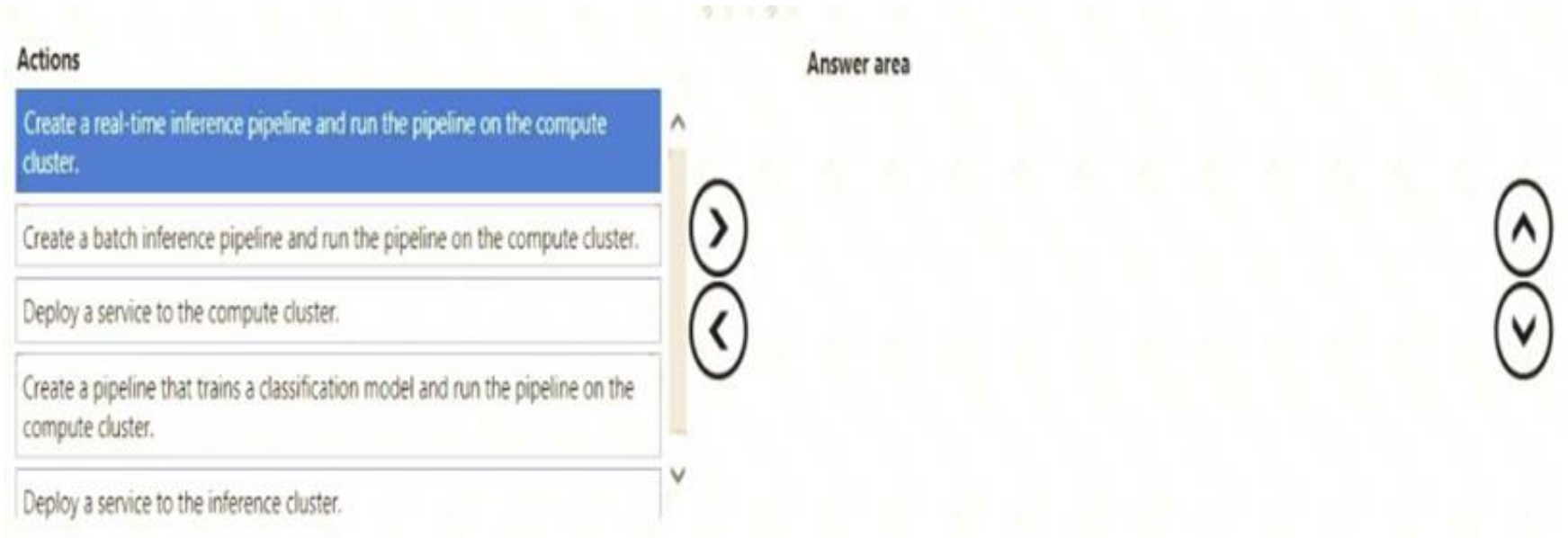
Microsoft

Exam Questions DP-100

Designing and Implementing a Data Science Solution on Azure

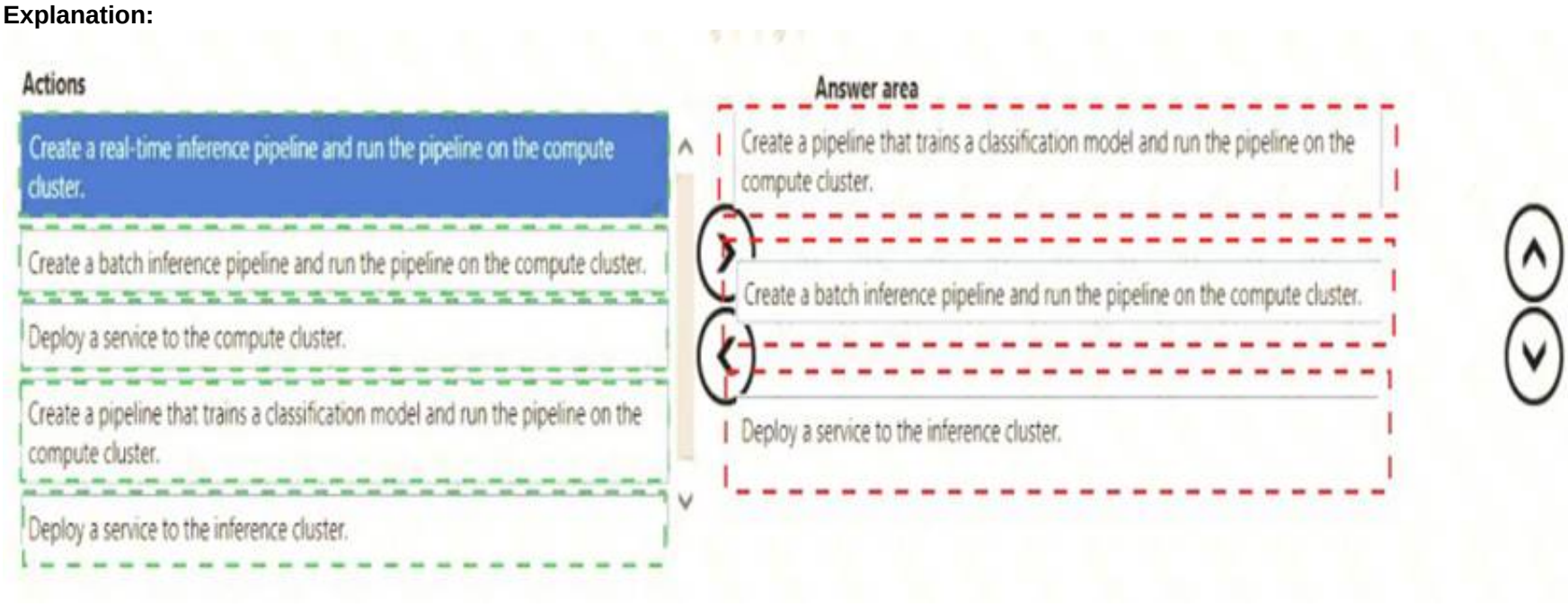
NEW QUESTION 1
- (Exam Topic 3)
You have an Azure Machine Learning workspace that contains a training cluster and an inference cluster. You plan to create a classification model by using the Azure Machine Learning designer.
You need to ensure that client applications can submit data as HTTP requests 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.



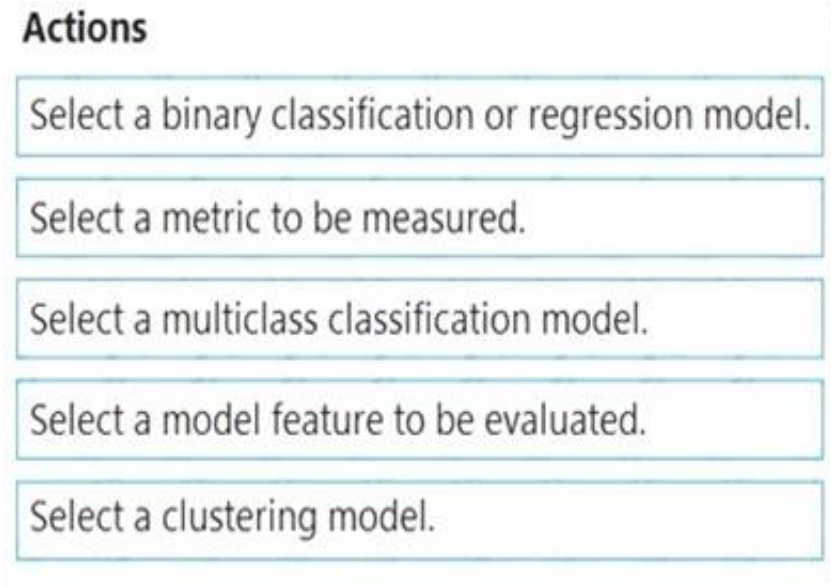
A. Mastered
B. Not Mastered

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.
You create a model to forecast weather conditions based on historical data.
You need to create a pipeline that runs a processing script to load data from a datastore and pass the processed data to a machine learning model training script.
Solution: Run the following code:

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

Does the solution meet the goal?

A. Yes
B. No

Answer: B

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

NEW QUESTION 4
- (Exam Topic 3)
You are using C-Support Vector classification to do a multi-class classification with an unbalanced training dataset. The C-Support Vector classification using Python code shown below:

```
from sklearn.svm import svc
import numpy as np
svc = SVC(kernel= 'linear', class_weight= 'balanced', C-1.0, random_state-0)
model1 = svc.fit(X_train, y)
```

You need to evaluate the C-Support Vector classification code.
Which evaluation statement should you use? To answer, select the appropriate options in the answer area. NOTE: Each correct selection is worth one point.

Code Segment	Evaluation Statement
class_weight=balanced	Automatically select the performance metrics for the classification Automatically adjust weights directly proportional to class frequencies in the input data Automatically adjust weights inversely proportional to class frequences in the input data
C parameter	Penalty parameter Degreee of polynomial kernel function Size of the kernel cache

A. Mastered

B. Not Mastered

Answer: A

Explanation:
Box 1: Automatically adjust weights inversely proportional to class frequencies in the input data
The “balanced” mode uses the values of y to automatically adjust weights inversely proportional to class frequencies in the input data as n_samples / (n_classes * np.bincount(y)).
Box 2: Penalty parameter
Parameter: C : float, optional (default=1.0) Penalty parameter C of the error term. References:
https://scikit-learn.org/stable/modules/generated/sklearn.svm.SVC.html

NEW QUESTION 5
- (Exam Topic 3)
HOTSPOT
You register the following versions of a model.

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

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

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

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

	Yes	No
The code will cause a previous version of the saved model to be overwritten.	○	○
The version number will now be 4.	○	○
The latest version of the stored model will have a property of value: 87.43.	○	○

A. Mastered
B. Not Mastered

Answer: A

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

NEW QUESTION 6
- (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 7
- (Exam Topic 3)
You train a machine learning model.
You must deploy the model as a real-time inference service for testing. The service requires low CPU utilization and less than 48 MB of RAM. The compute target for the deployed service must initialize automatically while minimizing cost and administrative overhead.
Which compute target should you use?

A. Azure Kubernetes Service (AKS) inference cluster
B. Azure Machine Learning compute cluster
C. Azure Container Instance (ACI)
D. attached Azure Databricks cluster

Answer: C

Explanation:
Azure Container Instances (ACI) are suitable only for small models less than 1 GB in size. Use it for low-scale CPU-based workloads that require less than 48 GB of RAM.
Note: Microsoft recommends using single-node Azure Kubernetes Service (AKS) clusters for dev-test of larger models.
Reference:
https://docs.microsoft.com/id-id/azure/machine-learning/how-to-deploy-and-where

NEW QUESTION 8
- (Exam Topic 3)
Note: This question is part of a series of questions that present the same scenario. Each question in the series contains a unique solution that might meet the stated goals. Some question sets might have more than one correct solution, while others might not have a correct solution.
After you answer a question in this section, you will NOT be able to return to it. As a result, these questions will not appear in the review screen.
You 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: Calculate the column median value and use the median value as the replacement for any missing value in the column.
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 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 new experiment in Azure Machine Learning Studio.
One class has a much smaller number of observations than the other classes in the training set. You need to select an appropriate data sampling strategy to compensate for the class imbalance. Solution: You use the Stratified split for the sampling mode.
Does the solution meet the goal?

A. Yes
B. No

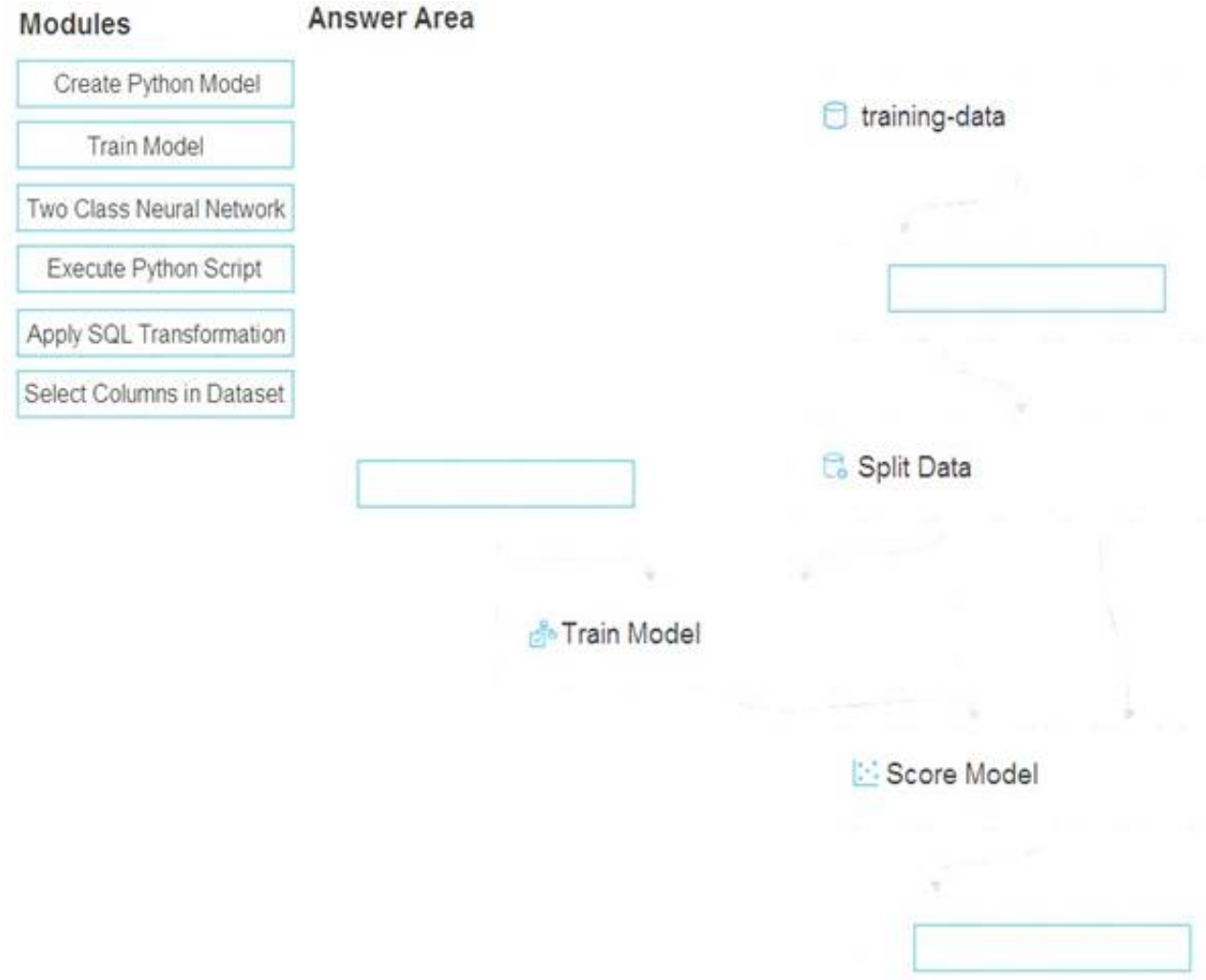
Answer: B

Explanation:
Instead use the Synthetic Minority Oversampling Technique (SMOTE) sampling mode.
Note: SMOTE is used to increase the number of 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 10

- (Exam Topic 3)
You create a training pipeline using the Azure Machine Learning designer. You upload a CSV file that contains the data from which you want to train your model.
You need to use the designer to create a pipeline that includes steps to perform the following tasks:

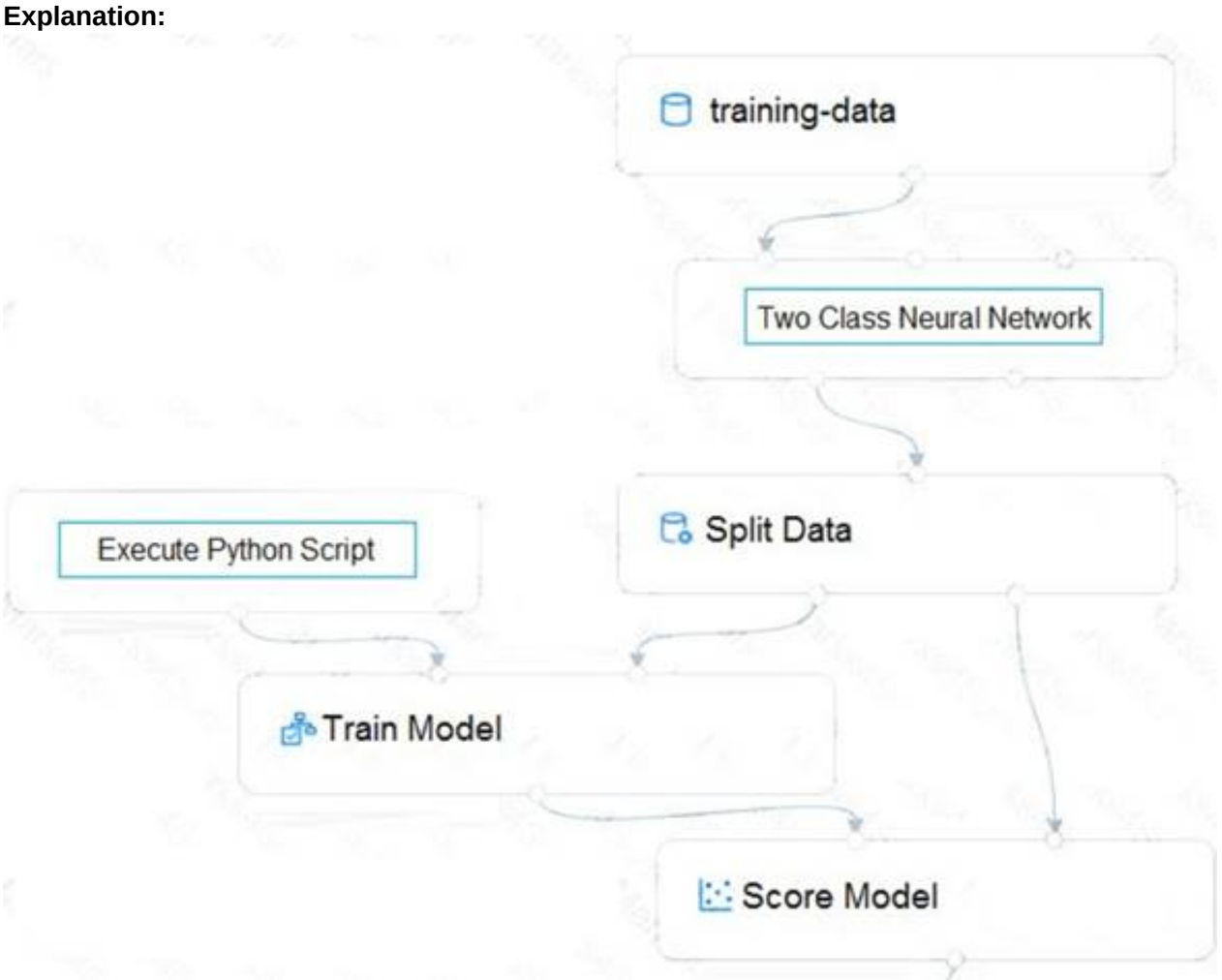
- Select the training features using the pandas filter method.
- Train a model based on the naive_bayes.GaussianNB algorithm.
- Return only the Scored Labels column by using the query SELECT [Scored Labels] FROM t1; Which modules should you use? To answer, drag the appropriate modules to the appropriate locations. Each
module name may be used once, more than once, or not at all. You may need to drag the split bar between panes or scroll to view content.
NOTE: Each correct selection is worth one point.



A. Mastered
B. Not Mastered

Answer: A

Explanation:



NEW QUESTION 10
- (Exam Topic 3)
You are creating a classification model for a banking company to identify possible instances of credit card fraud. You plan to create the model in Azure Machine

Learning by using automated machine learning.
The training dataset that you are using is highly unbalanced. You need to evaluate the classification model.
Which primary metric should you use?

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

Answer: C

Explanation:
AUC_weighted is a Classification metric.
Note: AUC is the Area under the Receiver Operating Characteristic Curve. Weighted is the arithmetic mean of the score for each class, weighted by the number of true instances in each class.
Reference:
https://docs.microsoft.com/en-us/azure/machine-learning/how-to-understand-automated-ml

NEW QUESTION 12
- (Exam Topic 3)
You plan to use automated machine learning to train a regression model. You have data that has features which have missing values, and categorical features with few distinct values.
You need to configure automated machine learning to automatically impute missing values and encode categorical features as part of the training task.
Which parameter and value pair should you use in the AutoMLConfig class?

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

Answer: A

Explanation:
Featurization str or FeaturizationConfig Values: 'auto' / 'off' / FeaturizationConfig
Indicator for whether featurization step should be done automatically or not, or whether customized featurization should be used.
Column type is automatically detected. Based on the detected column type preprocessing/featurization is done as follows:
Categorical: Target encoding, one hot encoding, drop high cardinality categories, impute missing values. Numeric: Impute missing values, cluster distance, weight of evidence.
DateTime: Several features such as day, seconds, minutes, hours etc. Text: Bag of words, pre-trained Word embedding, text target encoding. Reference:
https://docs.microsoft.com/en-us/python/api/azureml-train-automl-client/azureml.train.automl.automlconfig.auto

NEW QUESTION 17
- (Exam Topic 3)
You create an Azure Machine Learning workspace.
You need to detect data drift between a baseline dataset and a subsequent target dataset by using the DataDriftDetector class.
How should you complete the code segment? To answer, select the appropriate options in the answer area.
NOTE: Each correct selection is worth one point.

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

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

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

backfill
create_from_datasets
create_from_model

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

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

backfill
list
update

A. Mastered
B. Not Mastered

Answer: A

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

NEW QUESTION 20
- (Exam Topic 3)
Note: This question is part of a series of questions that present the same scenario. Each question in the series contains a unique solution that might meet the stated goals. Some question sets might have more than one correct solution, while others might not have a correct solution.
After you answer a question in this section, you will NOT be able to return to it. As a result, these questions will not appear in the review screen.
You have a Python script named train.py in a local folder named scripts. The script trains a regression model by using scikit-learn. The script includes code to load a training data file which is also located in the scripts folder.
You must run the script as an Azure ML experiment on a compute cluster named aml-compute.
You need to configure the run to ensure that the environment includes the required packages for model training. You have instantiated a variable named aml-compute that references the target compute cluster.
Solution: Run the following code:

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

Does the solution meet the goal?

A. Yes
B. No

Answer: B

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

NEW QUESTION 24
- (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 28
- (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 32
- (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 36
- (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 39
- (Exam Topic 3)
You are with a time series dataset in Azure Machine Learning Studio.
You need to split your dataset into training and testing subsets by using the Split Data module. Which splitting mode should you use?

A. Regular Expression Split
B. Split Rows with the Randomized split parameter set to true
C. Relative Expression Split
D. Recommender Split

Answer: B

Explanation:
Split Rows: Use this 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.
References:
https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/split-data

NEW QUESTION 40
- (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 42
- (Exam Topic 3) You are solving a classification task. The dataset is imbalanced.
You need to select an Azure Machine Learning Studio module to improve the classification accuracy. Which module should you use?

A. Fisher Linear Discriminant Analysis.
B. Filter Based Feature Selection
C. Synthetic Minority Oversampling Technique (SMOTE)
D. Permutation Feature Importance

Answer: C

Explanation:
Use the SMOTE module in Azure Machine Learning Studio (classic) to increase the number of underepresented cases in a dataset used for machine learning. SMOTE is a better way of increasing the number of rare cases than simply duplicating existing cases.
You connect the SMOTE module to a dataset that is imbalanced. There are many reasons why a dataset might be imbalanced: the category you are targeting might be very rare in the population, or the data might simply be difficult to collect. Typically, you use SMOTE when the class you want to analyze is under-represented.
Reference:
https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/smote

NEW QUESTION 44
- (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 47
- (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 52
- (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 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 55
- (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 57
- (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 62
- (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 65
- (Exam Topic 3)
You are preparing to use the Azure ML SDK to run an experiment and need to create compute. You run the following code:

```
from azureml.core.compute import ComputeTarget, AmlCompute
from azureml.core.compute_target import ComputeTargetException
ws = Workspace.from_config()
cluster_name = 'aml-cluster'
try:
  training_compute = ComputeTarget(workspace=ws, name=cluster_name)
except ComputeTargetException:
  compute_config = AmlCompute.provisioning_configuration(vm_size='STANDARD_D2_V2', vm_priority='lowpriority',
max_nodes=4)
  training_compute = ComputeTarget.create(ws, cluster_name, compute_config)
  training_compute.wait_for_completion(show_output=True)
```

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
If a training cluster named aml-cluster already exists in the workspace, it will be deleted and replaced.	O	O
The `wait_for_completion()` method will not return until the aml-cluster compute has four active nodes.	O	O
If the code creates a new aml-cluster compute target, it may be preempted due to capacity constraints.	O	O
The aml-cluster compute target is deleted from the workspace after the training experiment completes.	O	O

A. Mastered
B. Not Mastered

Answer: A

Explanation:
Box 1: No
If a training cluster already exists it will be used. Box 2: Yes
The wait_for_completion method waits for the current provisioning operation to finish on the cluster. Box 3: Yes
Low Priority VMs use Azure's excess capacity and are thus cheaper but risk your run being pre-empted.

Box 4: No
Need to use training_compute.delete() to deprovision and delete the AmlCompute target. Reference:
https://notebooks.azure.com/azureml/projects/azureml-getting-started/html/how-to-use-azureml/training/train-on https://docs.microsoft.com/en-us/python/api/azureml-core/azureml.core.compute.computetarget

NEW QUESTION 67
- (Exam Topic 3)
You create a multi-class image classification deep learning model that uses a set of labeled images. You create a script file named train.py that uses the PyTorch 1.3 framework to train the model.
You must run the script by using an estimator. The code must not require any additional Python libraries to be installed in the environment for the estimator. The time required for model training must be minimized.
You need to define the estimator that will be used to run the script. Which estimator type should you use?

A. TensorFlow
B. PyTorch
C. SKLearn
D. Estimator

Answer: B

Explanation:
For PyTorch, TensorFlow and Chainer tasks, Azure Machine Learning provides respective PyTorch, TensorFlow, and Chainer estimators to simplify using these frameworks.
Reference:
https://docs.microsoft.com/en-us/azure/machine-learning/how-to-train-ml-models

NEW QUESTION 70
- (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

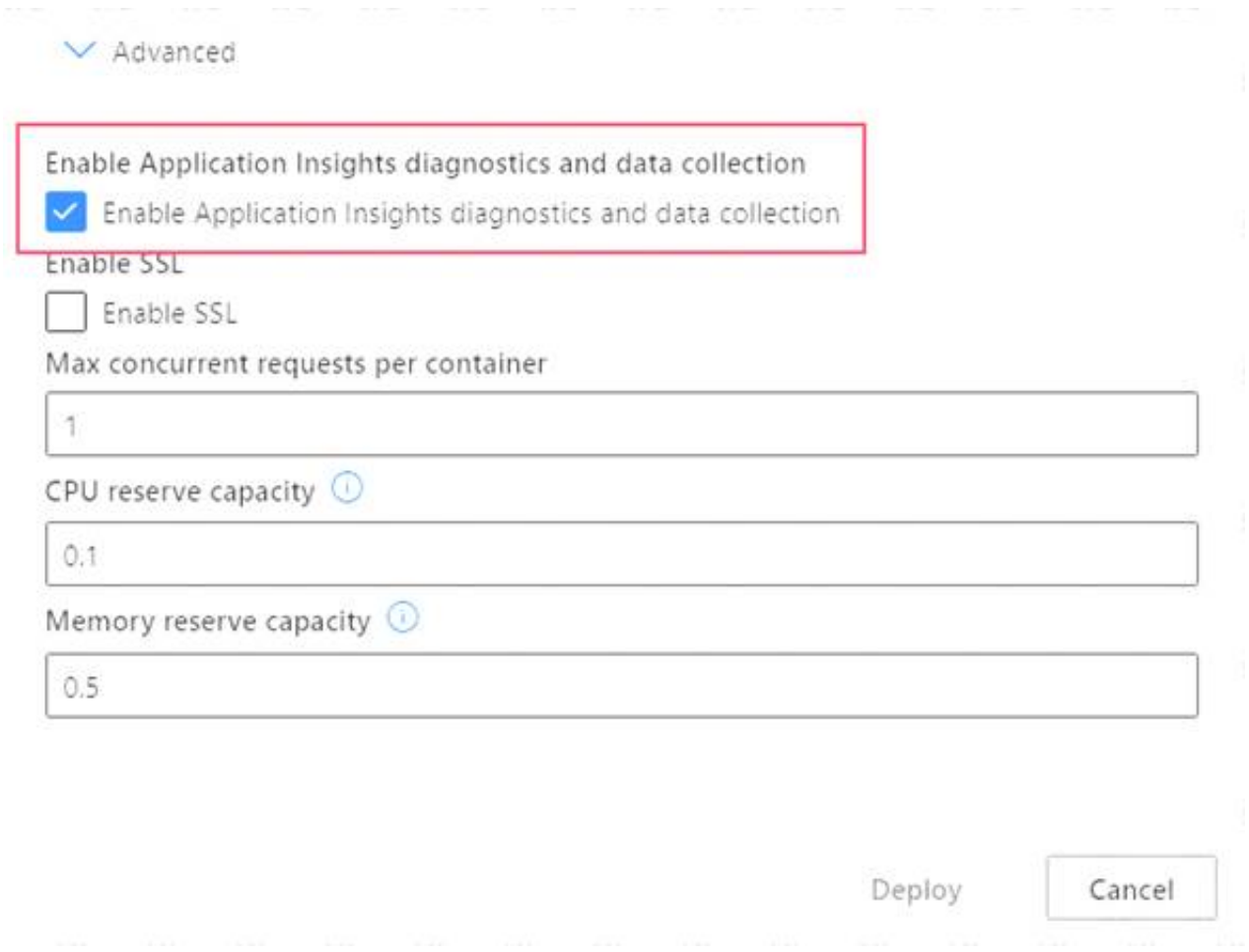
Explanation:
A run represents a single trial of an experiment. Runs are used to monitor the asynchronous execution of a trial, log metrics and store output of the trial, and to analyze results and access artifacts generated by the trial.
The run Class get_all_logs method downloads all logs for the run to a directory. Reference:
https://docs.microsoft.com/en-us/python/api/azureml-core/azureml.core.run(class)

NEW QUESTION 71
- (Exam Topic 3)
You deploy a real-time inference service for a trained model.
The deployed model supports a business-critical application, and it is important to be able to monitor the data submitted to the web service and the predictions the data generates.
You need to implement a monitoring solution for the deployed model using minimal administrative effort. What should you do?

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

Answer: B

Explanation:
Configure logging with Azure Machine Learning studio
You can also enable Azure Application Insights from Azure Machine Learning studio. When you're ready to deploy your model as a web service, use the following steps to enable Application Insights:
* 1. Sign in to the studio at https://ml.azure.com.
* 2. Go to Models and select the model you want to deploy.
* 3. Select +Deploy.
* 4. Populate the Deploy model form.
* 5. Expand the Advanced menu.
* 6. Select Enable Application Insights diagnostics and data collection.



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

NEW QUESTION 72
- (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.

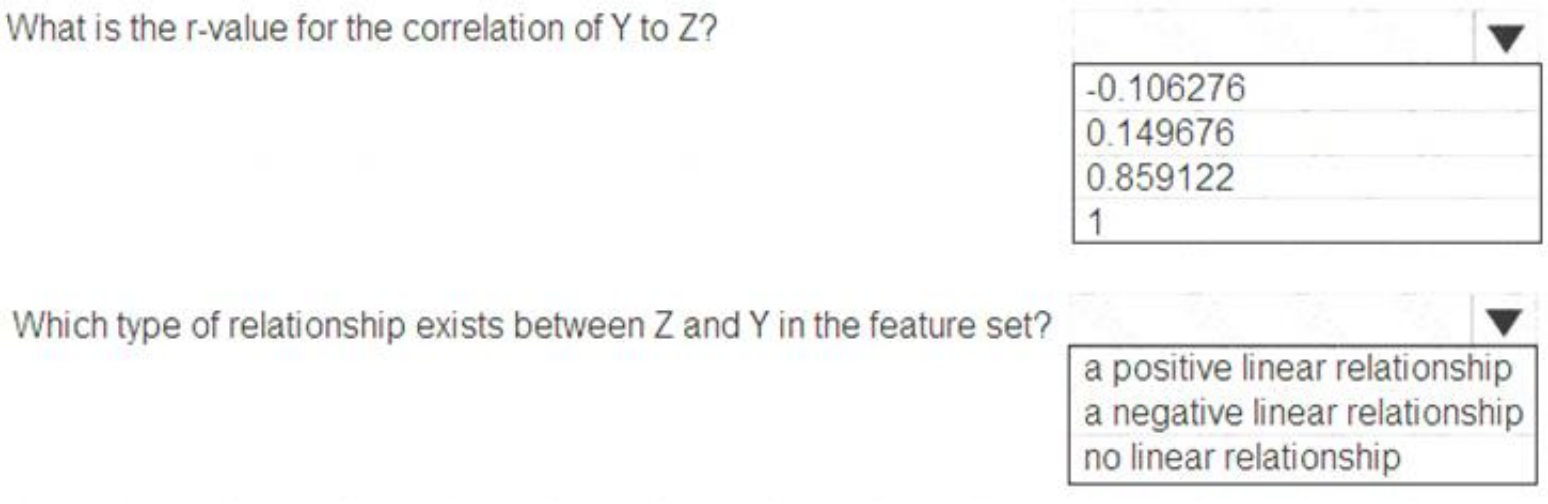
Answer: ADE

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 76
- (Exam Topic 3)
You have a feature set containing the following numerical features: X, Y, and Z.
The Poisson correlation coefficient (r-value) of X, Y, and Z features is shown in the following image:

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

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



A. Mastered
B. Not Mastered

Answer: A

Explanation:
Box 1: 0.859122
Box 2: a positively linear relationship
+1 indicates a strong positive linear relationship
-1 indicates a strong negative linear correlation
0 denotes no linear relationship between the two variables. References:
https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/compute-linear-correlation

NEW QUESTION 80
- (Exam Topic 3)
You previously deployed a model that was trained using a tabular dataset named training-dataset, which is based on a folder of CSV files.
Over time, you have collected the features and predicted labels generated by the model in a folder containing a CSV file for each month. You have created two tabular datasets based on the folder containing the inference data: one named predictions-dataset with a schema that matches the training data exactly, including the predicted label; and another named features-dataset with a schema containing all of the feature columns and a timestamp column based on the filename, which includes the day, month, and year.
You need to create a data drift monitor to identify any changing trends in the feature data since the model was trained. To accomplish this, you must define the required datasets for the data drift monitor.
Which datasets should you use to configure the data drift monitor? To answer, drag the appropriate datasets to the correct data drift monitor options. Each source may be used once, more than once, or not at all. You may need to drag the split bar between panes or scroll to view content.
NOTE: Each correct selection is worth one point.

A. Mastered
B. Not Mastered

Answer: A

Explanation:
Text Description automatically generated with medium confidence
Box 1: training-dataset
Baseline dataset - usually the training dataset for a model. Box 2: predictions-dataset
Target dataset - usually model input data - is compared over time to your baseline dataset. This comparison means that your target dataset must have a timestamp column specified.
The monitor will compare the baseline and target datasets. Reference:
https://docs.microsoft.com/en-us/azure/machine-learning/how-to-monitor-datasets

NEW QUESTION 84
- (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 87
- (Exam Topic 3)
You are developing a data science workspace that uses an Azure Machine Learning service. You need to select a compute target to deploy the workspace.
What should you use?

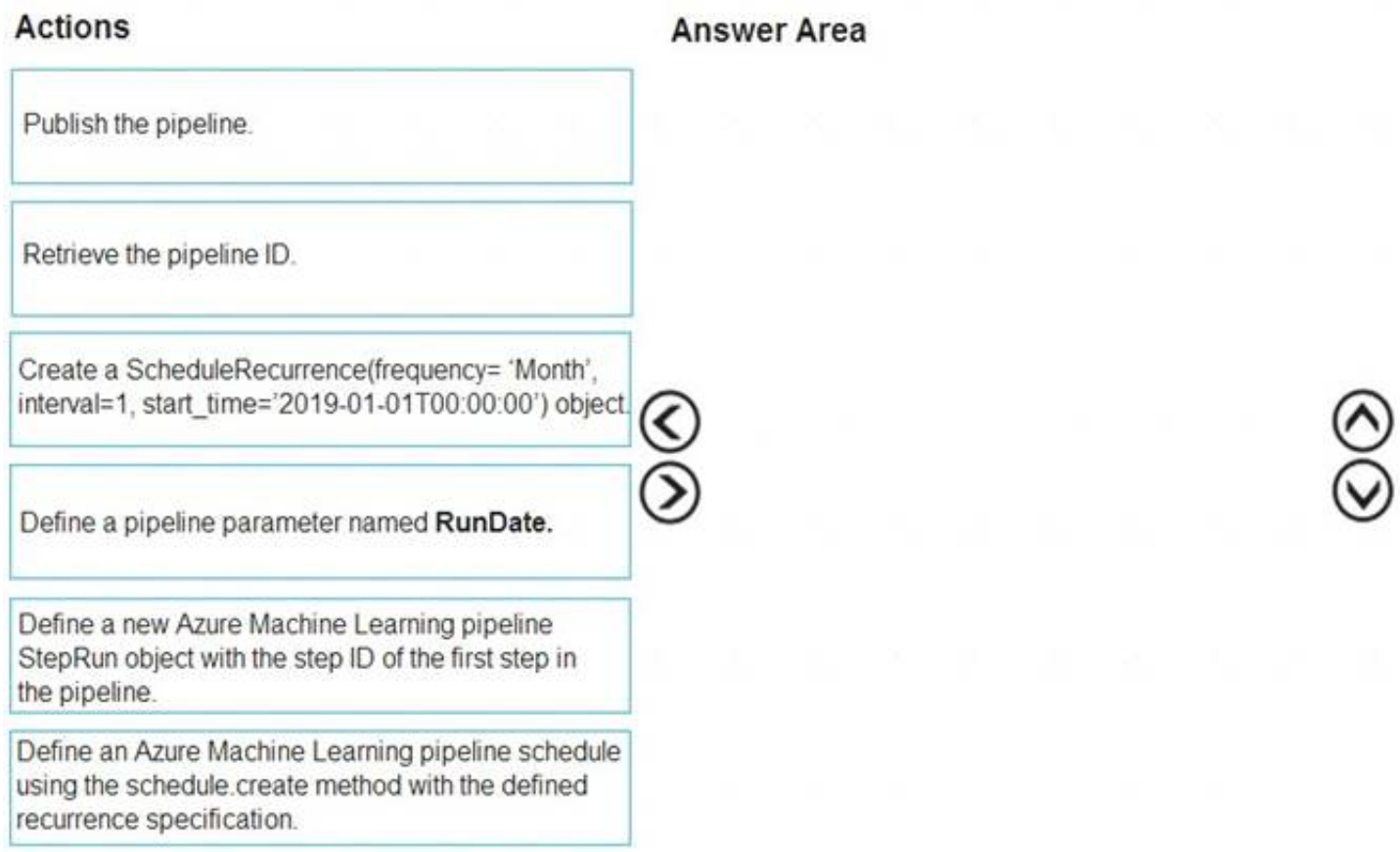
A. Azure Data Lake Analytics

B. Azure Databrick .
C. Apache Spark for HDInsight.
D. Azure Container Service

Answer: D

Explanation:
Azure Container Instances can be used as compute target for testing or development. Use for low-scale CPU-based workloads that require less than 48 GB of RAM.
Reference:
https://docs.microsoft.com/en-us/azure/machine-learning/service/how-to-deploy-and-where

NEW QUESTION 89
- (Exam Topic 3)
You create a multi-class image classification deep learning model.
The model must be retrained monthly with the new image data fetched from a public web portal. You create an Azure Machine Learning pipeline to fetch new data, standardize the size of images, and retrain the model.
You need to use the Azure Machine Learning SDK to configure the schedule for the pipeline.
Which four actions should you perform in sequence? To answer, move the appropriate actions from the list of actions to the answer area and arrange them in the correct order.



A. Mastered
B. Not Mastered

Answer: A

Explanation:
Step 1: Publish the pipeline.
To schedule a pipeline, you'll need a reference to your workspace, the identifier of your published pipeline, and the name of the experiment in which you wish to create the schedule.
Step 2: Retrieve the pipeline ID. Needed for the schedule.
Step 3: Create a ScheduleRecurrence..
To run a pipeline on a recurring basis, you'll create a schedule. A Schedule associates a pipeline, an experiment, and a trigger.
First create a schedule. Example: Create a Schedule that begins a run every 15 minutes: recurrence = ScheduleRecurrence(frequency="Minute", interval=15)
Step 4: Define an Azure Machine Learning pipeline schedule.. Example, continued:
recurring_schedule = Schedule.create(ws, name="MyRecurringSchedule", description="Based on time",
pipeline_id=pipeline_id, experiment_name=experiment_name, recurrence=recurrence)
Reference:
https://docs.microsoft.com/en-us/azure/machine-learning/how-to-schedule-pipelines

NEW QUESTION 91
- (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 94
- (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= [ pca | model | sklearn.decomposition ].fit_transform(X_train)
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

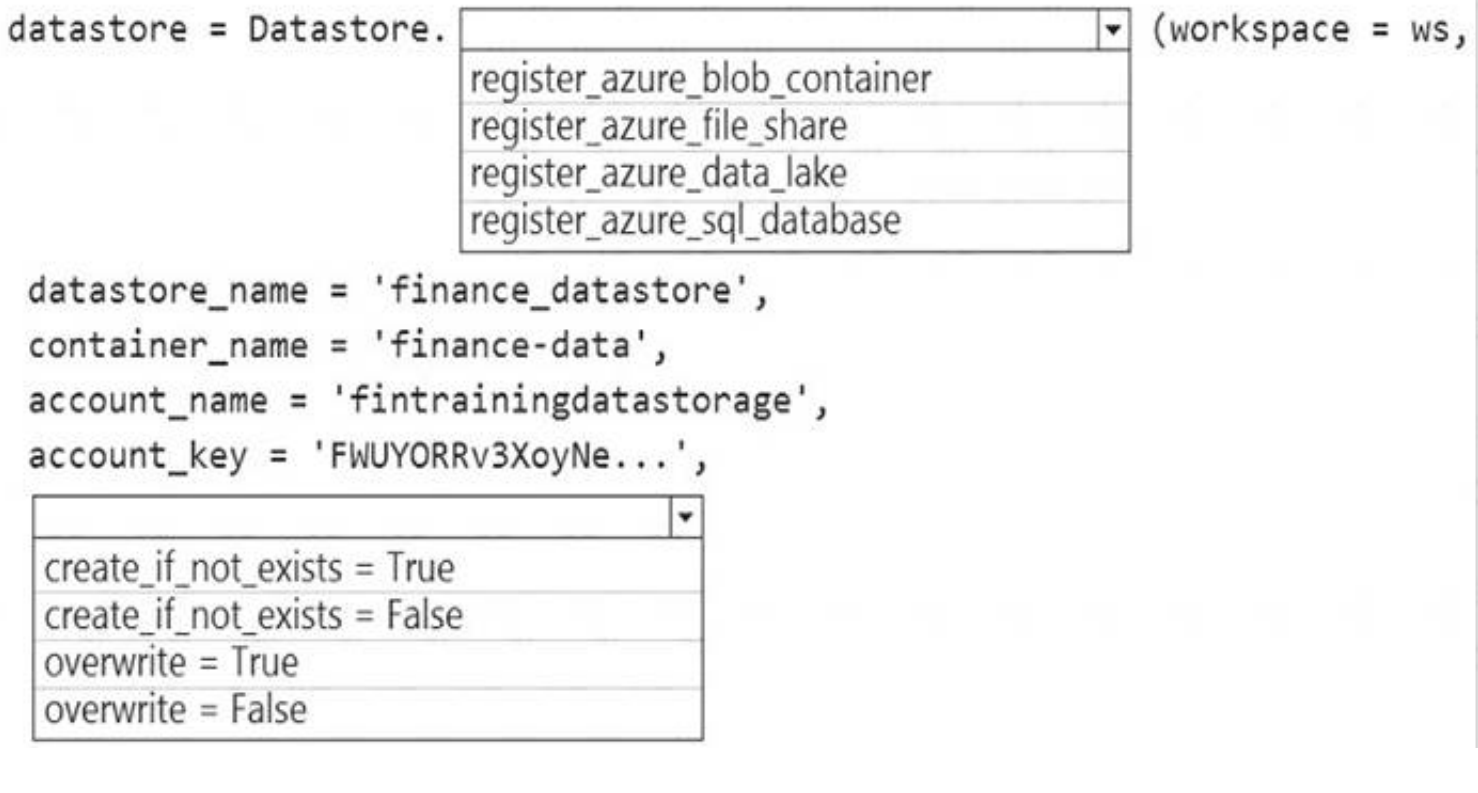
NEW QUESTION 96
- (Exam Topic 3)
You are implementing a machine learning model to predict stock prices. The model uses a PostgreSQL database and requires GPU processing.
You need to create a virtual machine that is pre-configured with the required tools. What should you do?

A. Create a Data Science Virtual Machine (DSVM) Windows edition.
B. Create a Geo Al 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 100
- (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 101
- (Exam Topic 3)
Note: This question is part of a series of questions that present the same scenario. Each question in the series contains a unique solution that might meet the stated goals. Some question sets might have more than one correct solution, while others might not have a correct solution.
After you answer a question in this section, you will NOT be able to return to it. As a result, these questions will not appear in the review screen.
You create a model to forecast weather conditions based on historical data.
You need to create a pipeline that runs a processing script to load data from a datastore and pass the processed data to a machine learning model training script.
Solution: Run the following code:

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

Does the solution meet the goal?

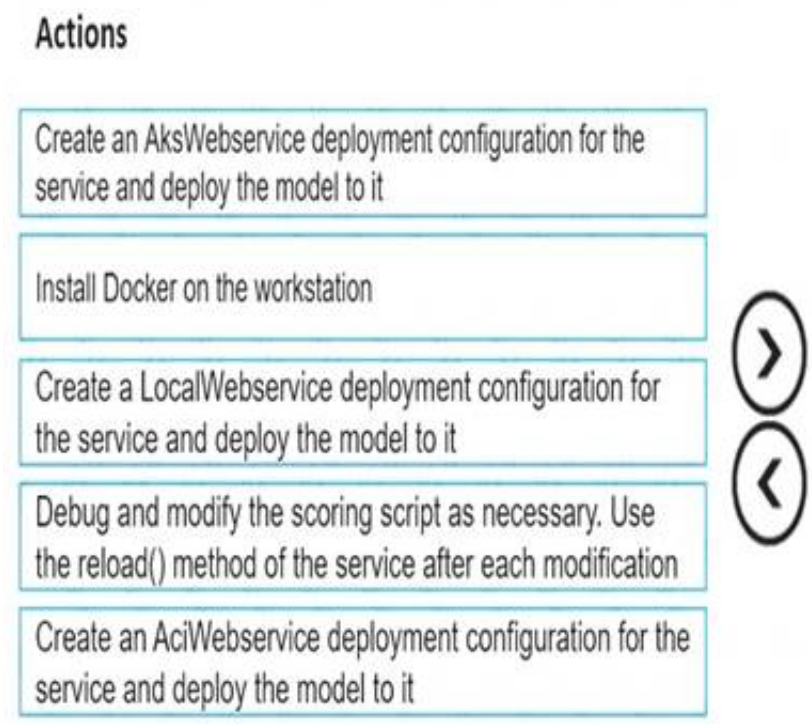
A. Yes
B. No

Answer: B

Explanation:
Note: Data used in pipeline can be produced by one step and consumed in another step by providing a PipelineData object as an output of one step and an input of one or more subsequent steps.
Compare with this example, the pipeline train step depends on the process_step_output output of the pipeline process step:
from azureml.pipeline.core import Pipeline, PipelineData from azureml.pipeline.steps import PythonScriptStep
datastore = ws.get_default_datastore()
process_step_output = PipelineData("processed_data", datastore=datastore) process_step = PythonScriptStep(script_name="process.py", arguments=["--data_for_train", process_step_output], outputs=[process_step_output], compute_target=aml_compute, source_directory=process_directory)
train_step = PythonScriptStep(script_name="train.py", arguments=["--data_for_train", process_step_output], inputs=[process_step_output],

compute_target=aml_compute, source_directory=train_directory)
pipeline = Pipeline(workspace=ws, steps=[process_step, train_step]) Reference:
https://docs.microsoft.com/en-us/python/api/azureml-pipeline-core/azureml.pipeline.core.pipelinedata?view=azu

NEW QUESTION 105
- (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.



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

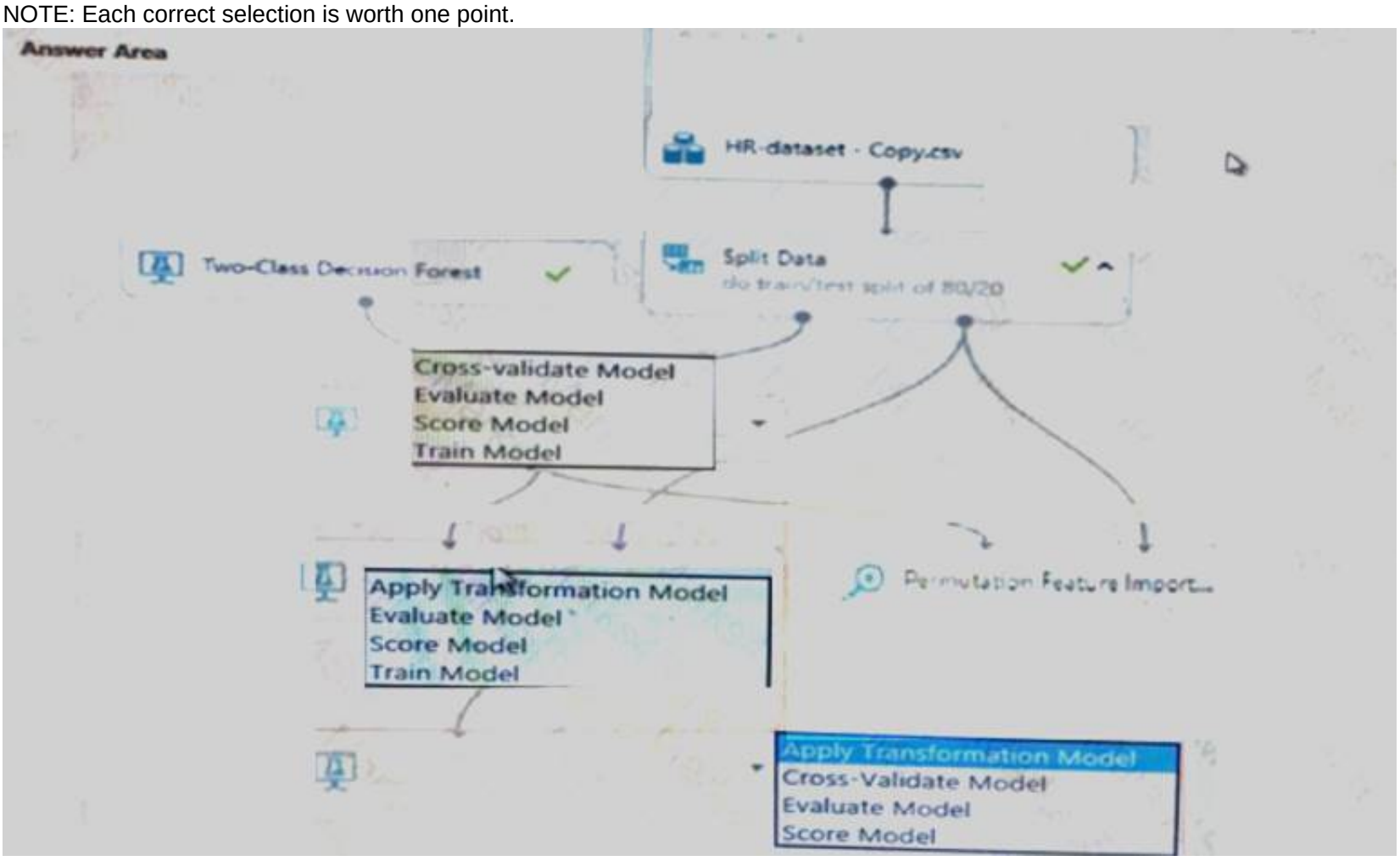
NEW QUESTION 108
- (Exam Topic 3)
You are building a recurrent neural network to perform a binary classification. You review the training loss, validation loss, training accuracy, and validation accuracy for each training epoch.
You need to analyze model performance.
Which observation indicates that the classification model is over fitted?

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

Answer: B

NEW QUESTION 111
- (Exam Topic 3)

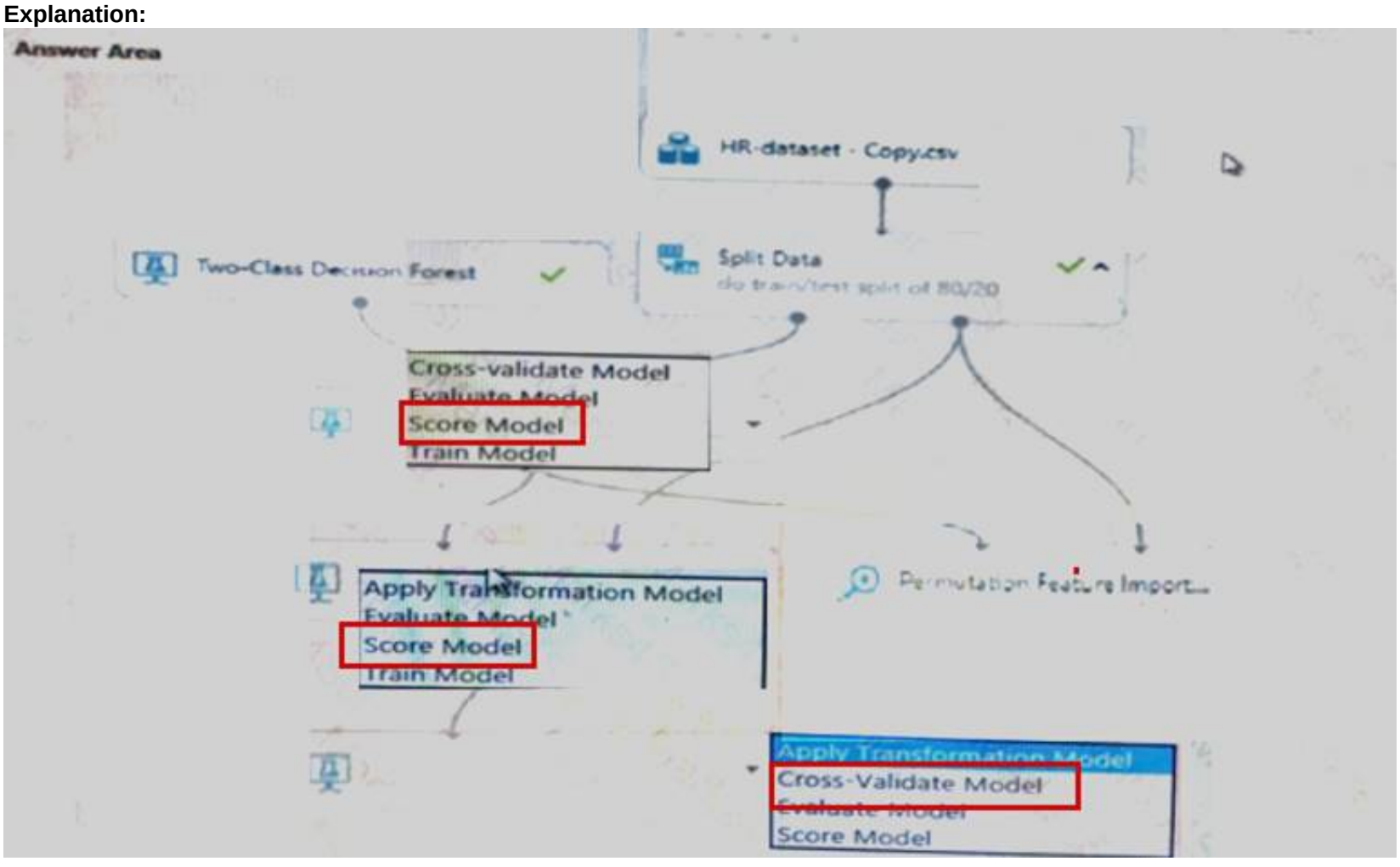
You create a binary classification model using Azure Machine Learning Studio.
You must use a Receiver Operating Characteristic (RO C) curve and an F1 score to evaluate the model. You need to create the required business metrics.
How should you complete the experiment? To answer, select the appropriate 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 114
- (Exam Topic 3)
HOTSPOT
You have an Azure blob container that contains a set of TSV files. The Azure blob container is registered as a datastore for an Azure Machine Learning service workspace. Each TSV file uses the same data schema.
You plan to aggregate data for all of the TSV files together and then register the aggregated data as a dataset in an Azure Machine Learning workspace by using the Azure Machine Learning SDK for Python.
You run the following code.

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

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

	Yes	No
The myDataset_1 dataset can be converted into a pandas dataframe by using the following method: `using myDataset_1.to_pandas_dataframe()`	○	○
The myDataset_1.to_path() method returns an array of file paths for all of the TSV files in the dataset.	○	○
The myDataset_2 dataset can be converted into a pandas dataframe by using the following method: `myDataset_2.to_pandas_dataframe()`	○	○

A. Mastered
B. Not Mastered

Answer: A

Explanation:
Box 1: No
FileDataset references single or multiple files in datastores or from public URLs. The TSV files need to be parsed.
Box 2: Yes
to_path() gets a list of file paths for each file stream defined by the dataset. Box 3: Yes
TabularDataset.to_pandas_dataframe loads all records from the dataset into a pandas DataFrame. TabularDataset represents data in a tabular format created by parsing the provided file or list of files.
Note: TSV is a file extension for a tab-delimited file used with spreadsheet software. TSV stands for Tab Separated Values. TSV files are used for raw data and can be imported into and exported from spreadsheet software. TSV files are essentially text files, and the raw data can be viewed by text editors, though they are often used when moving raw data between spreadsheets.
Reference:
https://docs.microsoft.com/en-us/python/api/azureml-core/azureml.data.tabulardataset

NEW QUESTION 115
- (Exam Topic 3)
You are developing a hands-on workshop to introduce Docker for Windows to attendees. You need to ensure that workshop attendees can install Docker on their devices.
Which two prerequisite components should attendees install on the devices? Each correct answer presents part of the solution.
NOTE: Each correct selection is worth one point.

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

Answer: CE

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

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

NEW QUESTION 116
- (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 119
- (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 124
- (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 129
- (Exam Topic 3)
You are planning to register a trained model in an Azure Machine Learning workspace.
You must store additional metadata about the model in a key-value format. You must be able to add new metadata and modify or delete metadata after creation.
You need to register the model. Which parameter should you use?

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

Answer: D

Explanation:
azureml.core.Model.properties:
Dictionary of key value properties for the Model. These properties cannot be changed after registration, however new key value pairs can be added.
Reference:
https://docs.microsoft.com/en-us/python/api/azureml-core/azureml.core.model.model

NEW QUESTION 131
- (Exam Topic 3)
Note: This question is part of a series of questions that present the same scenario. Each question in the series contains a unique solution that might meet the stated goals. Some question sets might have more than one correct solution, while others might not have a correct solution.
After you answer a question in this section, you will NOT be able to return to it. As a result, these questions will not appear in the review screen.
You are using Azure Machine Learning to run an experiment that trains a classification model.
You want to use Hyperdrive to find parameters that optimize the AUC metric for the model. You configure a HyperDriveConfig for the experiment by running the following code:

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

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

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

Does the solution meet the goal?

A. Yes
B. No

Answer: A

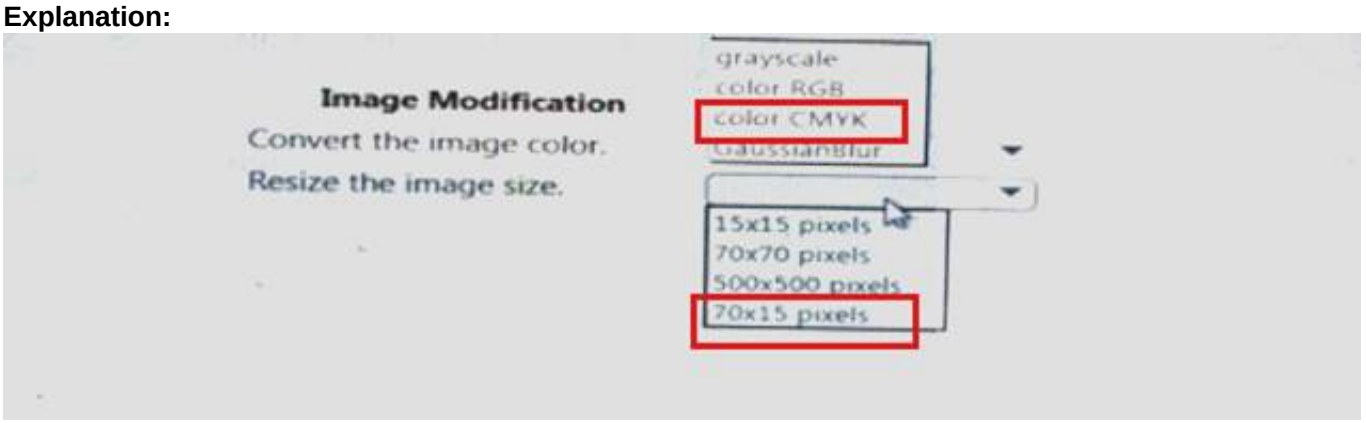
Explanation:
Python printing/logging example: logging.info(message)
Destination: Driver logs, Azure Machine Learning designer
Reference:
https://docs.microsoft.com/en-us/azure/machine-learning/how-to-debug-pipelines

NEW QUESTION 134
- (Exam Topic 3)
You arc I mating a deep learning model to identify cats and dogs. You have 25,000 color images. You must meet the following requirements:
• Reduce the number of training epochs.
• Reduce the size of the neural network.
• Reduce over-fitting of the neural network.
You need to select the image modification values.
Which value should you use? To answer, select the appropriate Options in the answer area. NOTE: Each correct selection is worth one point.

A. Mastered
B. Not Mastered

Answer: A

Explanation:



NEW QUESTION 137
- (Exam Topic 3)
You are analyzing a dataset containing historical data from a local taxi company. You arc developing a regression a regression model.
You must predict the fare of a taxi trip.
You need to select performance metrics to correctly evaluate the- regression model. Which two metrics can you use? Each correct answer presents a complete solution. NOTE: Each correct selection is worth one point.

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

Answer: BE

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

NEW QUESTION 141
- (Exam Topic 3)
You use the Azure Machine Learning SDK to run a training experiment that trains a classification model and calculates its accuracy metric.
The model will be retrained each month as new data is available. You must register the model for use in a batch inference pipeline.
You need to register the model and ensure that the models created by subsequent retraining experiments are registered only if their accuracy is higher than the currently registered model.
What are two possible ways to achieve this goal? Each correct answer presents a complete solution. NOTE: Each correct selection is worth one point.

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

Answer: CE

Explanation:
E: Using tags, you can track useful information such as the name and version of the machine learning library

used to train the model. Note that tags must be alphanumeric.
Reference:
https://notebooks.azure.com/xavierheriat/projects/azureml-getting-started/html/how-to-use-azureml/deployment/

NEW QUESTION 142
- (Exam Topic 3)
You arc creating a new experiment in Azure Machine Learning Studio. You have a small dataset that has missing values in many columns. The data does not require the application of predictors for each column. You plan to use the Clean Missing Data module to handle the missing data.
You need to select a data cleaning method. Which method should you use?

A. Synthetic Minority
B. Replace using Probabilistic PAC
C. Replace using MICE
D. Normalization

Answer: B

NEW QUESTION 146
- (Exam Topic 3)
You are performing sentiment analysis using a CSV file that includes 12,000 customer reviews written in a short sentence format. You add the CSV file to Azure Machine Learning Studio and configure it as the starting point dataset of an experiment. You add the Extract N-Gram Features from Text module to the experiment to extract key phrases from the customer review column in the dataset.
You must create a new n-gram dictionary from the customer review text and set the maximum n-gram size to trigrams.
What should you select? To answer, select the appropriate options in the answer area.
NOTE: Each correct selection is worth one point.

A. Mastered

B. Not Mastered

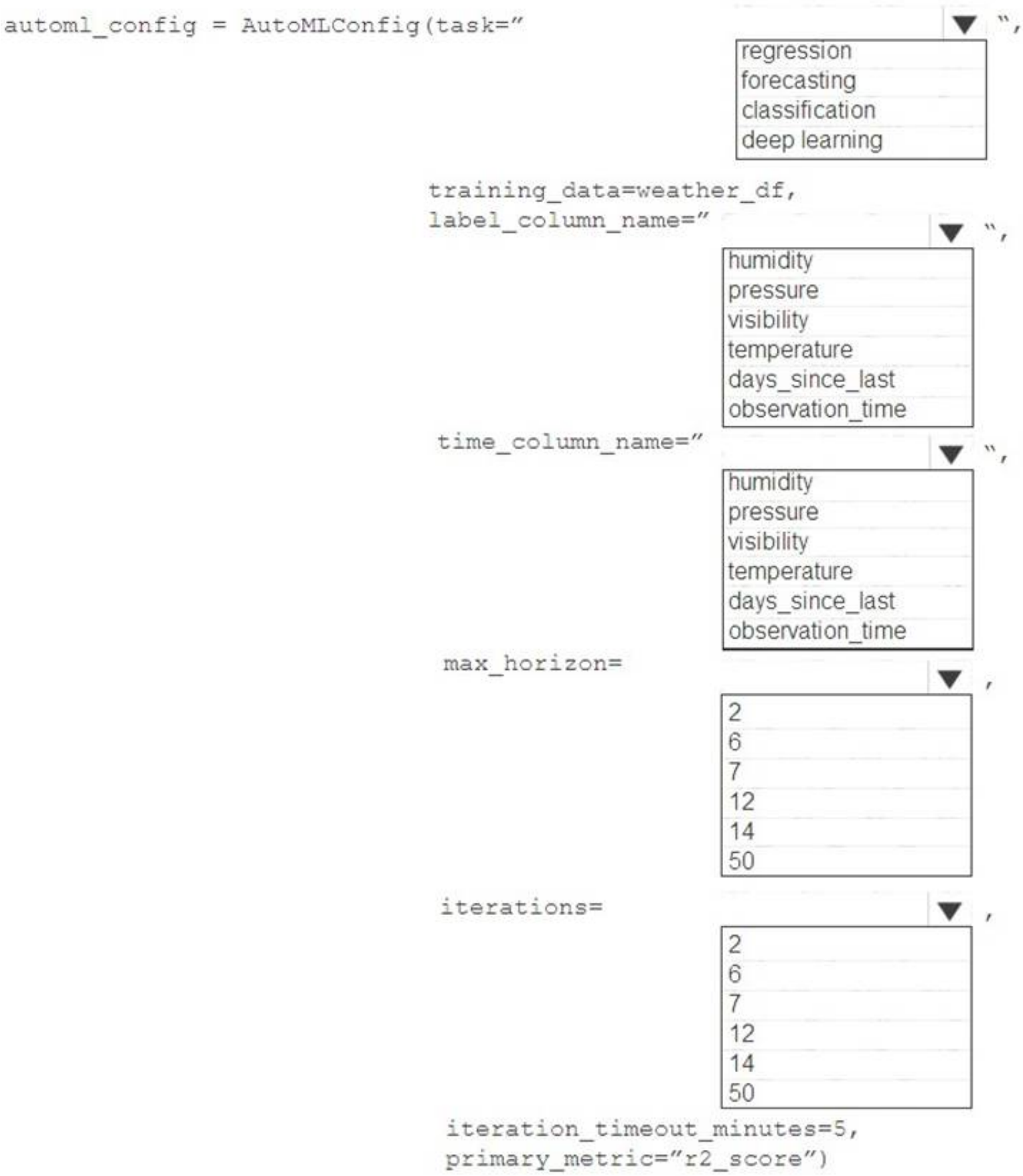
Answer: A

Explanation:
Vocabulary mode: Create
For Vocabulary mode, select Create to indicate that you are creating a new list of n-gram features. N-Grams size: 3
For N-Grams size, type a number that indicates the maximum size of the n-grams to extract and store. For example, if you type 3, unigrams, bigrams, and trigrams will be created.
Weighting function: Leave blank
The option, Weighting function, is required only if you merge or update vocabularies. It specifies how terms in the two vocabularies and their scores should be weighted against each other.
References:
https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/extract-n-gram-features-from

NEW QUESTION 148
- (Exam Topic 3)
HOTSPOT
You collect data from a nearby weather station. You have a pandas dataframe named weather_df that includes the following data:

Temperature	Observation_time	Humidity	Pressure	Visibility	Days_since_last observation
74	2019/10/2 00:00	0.62	29.87	3	0.5
89	2019/10/2 12:00	0.70	28.88	10	0.5
72	2019/10/3 00:00	0.64	30.00	8	0.5
80	2019/10/3 12:00	0.66	29.75	7	0.5

The data is collected every 12 hours: noon and midnight.
You plan to use automated machine learning to create a time-series model that predicts temperature over the next seven days. For the initial round of training, you want to train a maximum of 50 different models.
You must use the Azure Machine Learning SDK to run an automated machine learning experiment to train these models.
You need to configure the automated machine learning run.
How should you complete the AutoMLConfig definition? 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: forcasting
Task: The type of task to run. Values can be 'classification', 'regression', or 'forecasting' depending on the type of automated ML problem to solve.
Box 2: temperature
The training data to be used within the experiment. It should contain both training features and a label column (optionally a sample weights column).
Box 3: observation_time
time_column_name: The name of the time column. This parameter is required when forecasting to specify the datetime column in the input data used for building the time series and inferring its frequency. This setting is being deprecated. Please use forecasting_parameters instead.
Box 4: 7
"predicts temperature over the next seven days"
max_horizon: The desired maximum forecast horizon in units of time-series frequency. The default value is 1. Units are based on the time interval of your training data, e.g., monthly, weekly that the forecaster should
predict out. When task type is forecasting, this parameter is required.
Box 5: 50
"For the initial round of training, you want to train a maximum of 50 different models."
Iterations: The total number of different algorithm and parameter combinations to test during an automated
ML experiment. Reference:
https://docs.microsoft.com/en-us/python/api/azureml-train-automl-client/azureml.train.automl.automlconfig.auto

NEW QUESTION 150
- (Exam Topic 3)
You are analyzing a raw dataset that requires cleaning.

You must perform transformations and manipulations by using Azure Machine Learning Studio. You need to identify the correct modules to perform the transformations.
Which modules should you choose? To answer, drag the appropriate modules to the correct scenarios. Each module may be used once, more than once, or not at all.
You may need to drag the split bar between panes or scroll to view content. NOTE: Each correct selection is worth one point.

Answer Area

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

A. Mastered
B. Not Mastered

Answer: A

Explanation:
Box 1: Clean Missing Data Box 2: SMOTE
Use the SMOTE module in Azure Machine Learning Studio to increase the number of 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.
Box 3: Convert to Indicator Values
Use the Convert to Indicator Values module in Azure Machine Learning Studio. The purpose of this module is to convert columns that contain categorical values into a series of binary indicator columns that can more easily be used as features in a machine learning model.
Box 4: Remove Duplicate Rows References:
https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/smote https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/convert-to-indicator-values

NEW QUESTION 154
- (Exam Topic 3)
Note: This question is part of a series of questions that present the same scenario. Each question in the series contains a unique solution that might meet the stated goals. Some question sets might have more than one correct solution, while others might not have a correct solution.
After you answer a question in this section, you will NOT be able to return to it. As a result, these questions will not appear in the review screen.
You plan to use a Python script to run an Azure Machine Learning experiment. The script creates a reference to the experiment run context, loads data from a file, identifies the set of unique values for the label column, and completes the experiment run:
from azureml.core import Run import pandas as pd
run = Run.get_context()
data = pd.read_csv('data.csv') label_vals = data['label'].unique()
# Add code to record metrics here run.complete()
The experiment must record the unique labels in the data as metrics for the run that can be reviewed later.
You must add code to the script to record the unique label values as run metrics at the point indicated by the comment.
Solution: Replace the comment with the following code:
run.upload_file('outputs/labels.csv', './data.csv') Does the solution meet the goal?

A. Yes
B. No

Answer: B

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

NEW QUESTION 158
- (Exam Topic 3)
You plan to use the Hyperdrive feature of Azure Machine Learning to determine the optimal hyperparameter values when training a model.
You must use Hyperdrive to try combinations of the following hyperparameter values. You must not apply an early termination policy.
learning_rate: any value between 0.001 and 0.1
• batch_size: 16, 32, or 64
You need to configure the sampling method for the Hyperdrive experiment
Which two sampling methods can you use? Each correct answer is a complete solution. NOTE: Each correct selection is worth one point.

A. Grid sampling
B. No sampling
C. Bayesian sampling
D. Random sampling

Answer: CD

Explanation:
C: Bayesian sampling is based on the Bayesian optimization algorithm and makes intelligent choices on the hyperparameter values to sample next. It picks the sample based on how the previous samples performed, such that the new sample improves the reported primary metric.
Bayesian sampling does not support any early termination policy Example:
from azureml.train.hyperdrive import BayesianParameterSampling from azureml.train.hyperdrive import uniform, choice param_sampling = BayesianParameterSampling({ "learning_rate": uniform(0.05, 0.1),
"batch_size": choice(16, 32, 64, 128)
}
)
D: In random sampling, hyperparameter values are randomly selected from the defined search space. Random sampling allows the search space to include both discrete and continuous hyperparameters.
Reference:
https://docs.microsoft.com/en-us/azure/machine-learning/how-to-tune-hyperparameters

NEW QUESTION 161
- (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 165
- (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: Accuracy, Precision, Recall, F1 score and AUC. Does the solution meet the goal?

A. Yes
B. No

Answer: B

Explanation:
Those are metrics for evaluating classification models, instead use: Mean Absolute Error, Root Mean Absolute Error, Relative Absolute Error, Relative Squared Error, and the Coefficient of Determination.
References:
https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/evaluate-model

NEW QUESTION 170
- (Exam Topic 3)

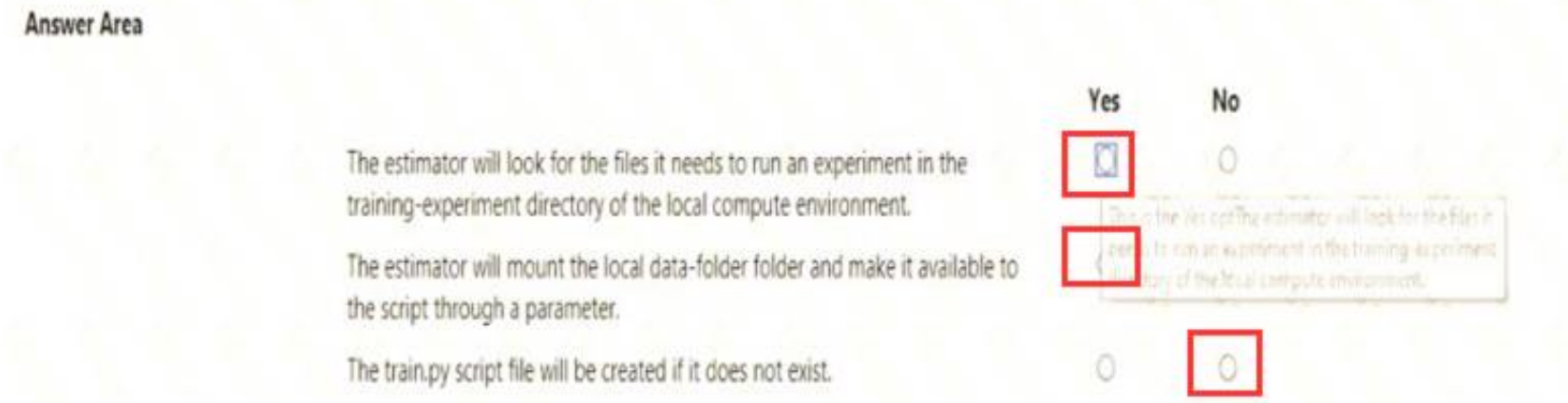
```
train_cluster = ComputeTarget(workspace=work_space, name='train-cluster')
estimator = Estimator(source_directory =
 'training-experiment',
 script_params = {'--data-folder': data_source.as_mount(), '--regularization': 0.8},
 compute_target = train_cluster,
 entry_script = 'train.py',
 conda_packages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

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

A. Mastered
B. Not Mastered

Answer: A

Explanation:



NEW QUESTION 171
- (Exam Topic 3)
You are a data scientist building a deep convolutional neural network (CNN) for image classification. The CNN model you built shows signs of overfitting.
You need to reduce overfitting and converge the model to an optimal fit.
Which two actions should you perform? Each correct answer presents a complete solution.
NOTE: Each correct selection is worth one point.

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

Answer: AC

Explanation:
References:
https://machinelearningmastery.com/how-to-reduce-overfitting-in-deep-learning-with-weight-regularization/
https://en.wikipedia.org/wiki/Convolutional_neural_network

NEW QUESTION 173
- (Exam Topic 3)
You create a classification model with a dataset that contains 100 samples with Class A and 10,000 samples with Class B
The variation of Class B is very high. You need to resolve imbalances. Which method should you use?

A. Partition and Sample
B. Cluster Centroids
C. Tomek links
D. Synthetic Minority Oversampling Technique (SMOTE)

Answer: D

NEW QUESTION 175
- (Exam Topic 3)
Note: This question is part of a series of questions that present the same scenario. Each question in the series contains a unique solution that might meet the stated goals. Some question sets might have more than one correct solution, while others might not have a correct solution.
After you answer a question in this section, you will NOT be able to return to it. As a result, these questions will not appear in the review screen.
You 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

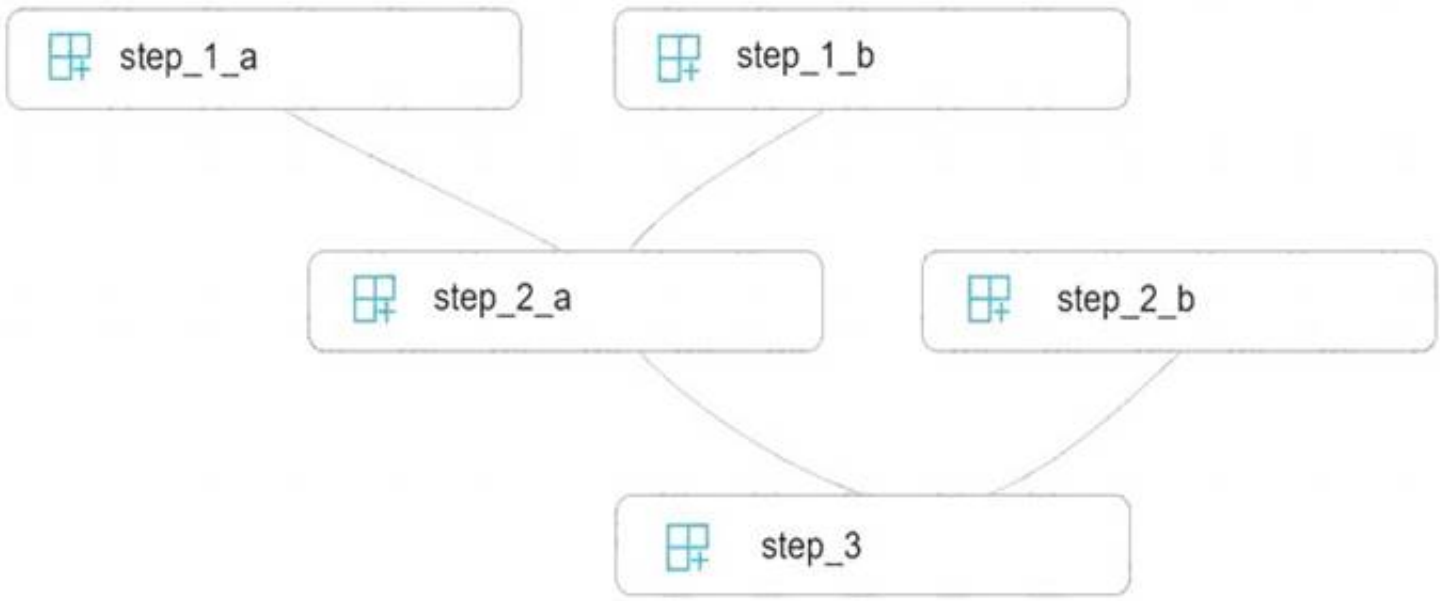
NEW QUESTION 180
- (Exam Topic 3)
Note: This question is part of a series of questions that present the same scenario. Each question in the series contains a unique solution that might meet the stated goals. Some question sets might have more than one correct solution, while others might not have a correct solution.
After you answer a question in this section, you will NOT be able to return to it. As a result, these questions will not appear in the review screen.
You are a data scientist using Azure Machine Learning Studio.
You need to normalize values to produce an output column into bins to predict a target column. Solution: Apply a Quantiles normalization with a QuantileIndex normalization.
Does the solution meet the GOAL?

A. Yes
B. No

Answer: B

Explanation:
Use the Entropy MDL binning mode which has a target column. References:
https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/group-data-into-bins

NEW QUESTION 182
- (Exam Topic 3)
You write five Python scripts that must be processed in the order specified in Exhibit A – which allows the same modules to run in parallel, but will wait for modules with dependencies.
You must create an Azure Machine Learning pipeline using the Python SDK, because you want to script to create the pipeline to be tracked in your version control system. You have created five PythonScriptSteps and have named the variables to match the module names.



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

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

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

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

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

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

Answer: A

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

NEW QUESTION 184
- (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')
```

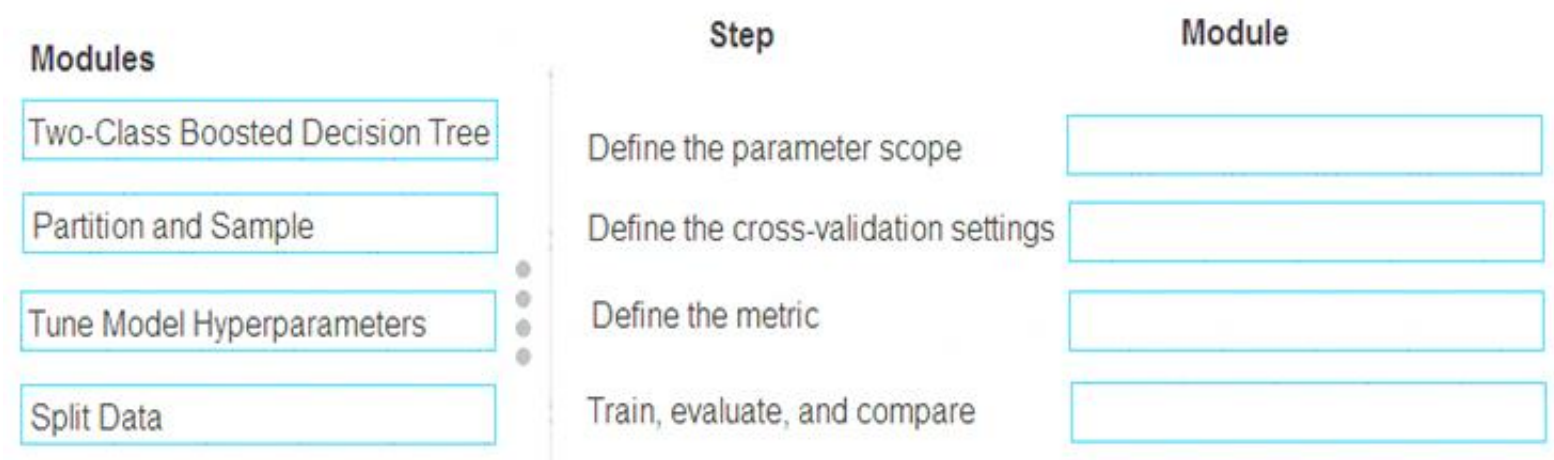
The experiment must be run on local computer using the default environment. You need to add code to start the experiment and run the script.
Which code segment should you use?

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 188
- (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.



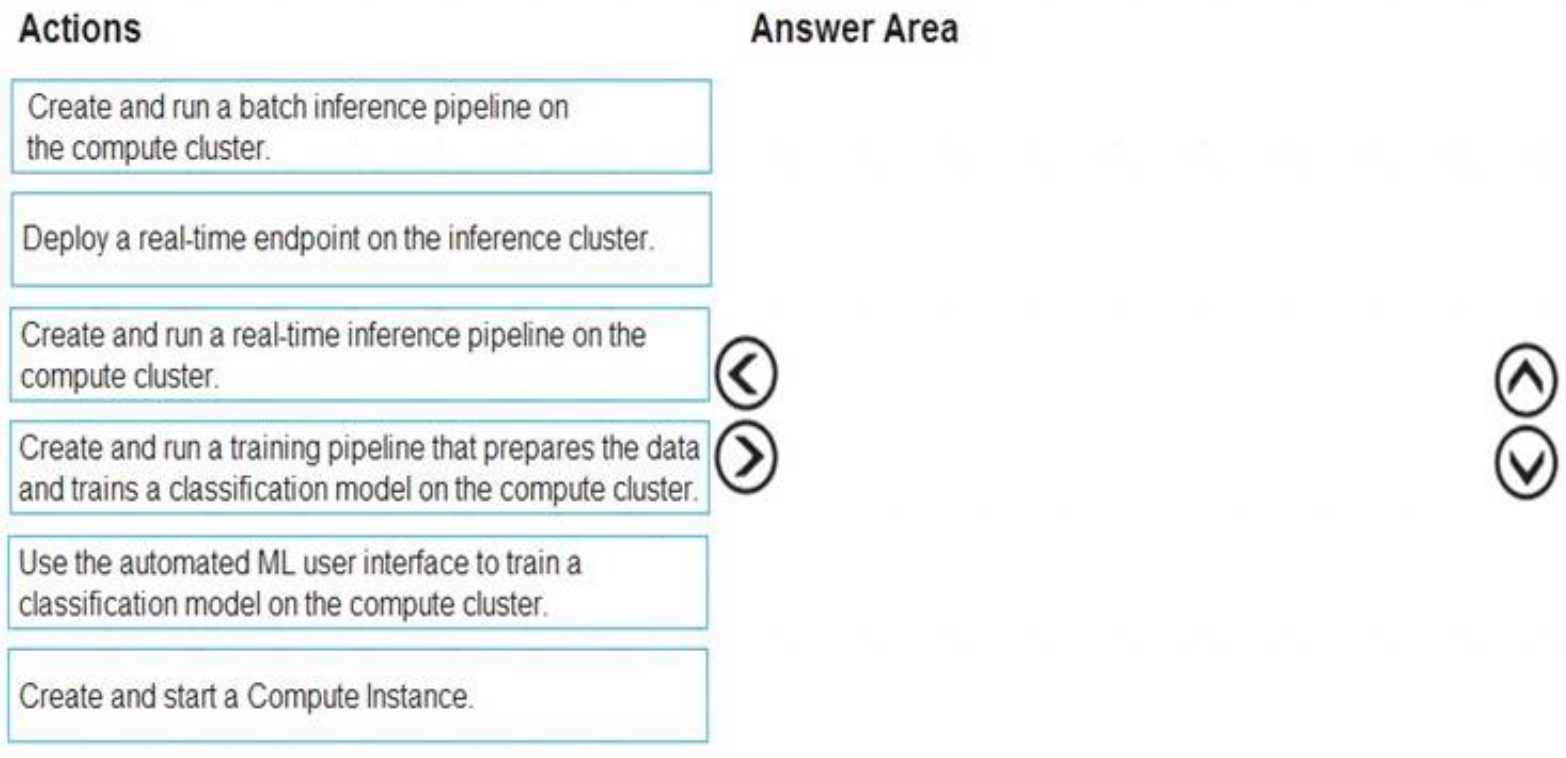
A. Mastered
B. Not Mastered

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 193
- (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

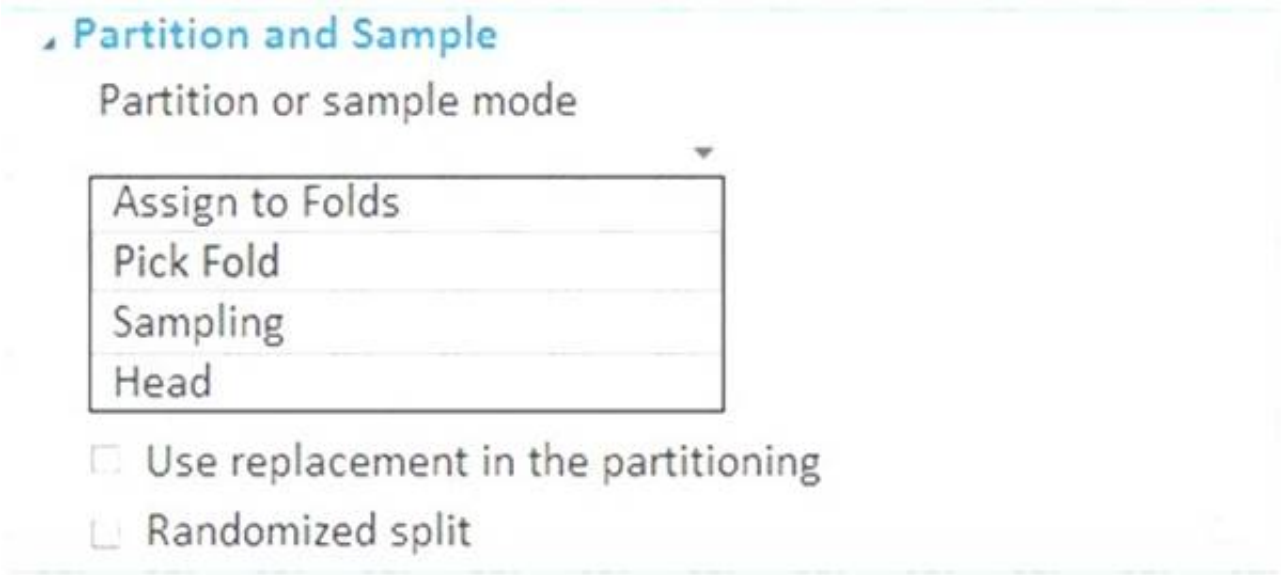
Answer: A

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 197
- (Exam Topic 3)
You have a dataset that contains 2,000 rows. You are building a machine learning classification model by using Azure Learning Studio. You add a Partition and Sample module to the experiment.
You need to configure the module. You must meet the following requirements:

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

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



A. Mastered
B. Not Mastered

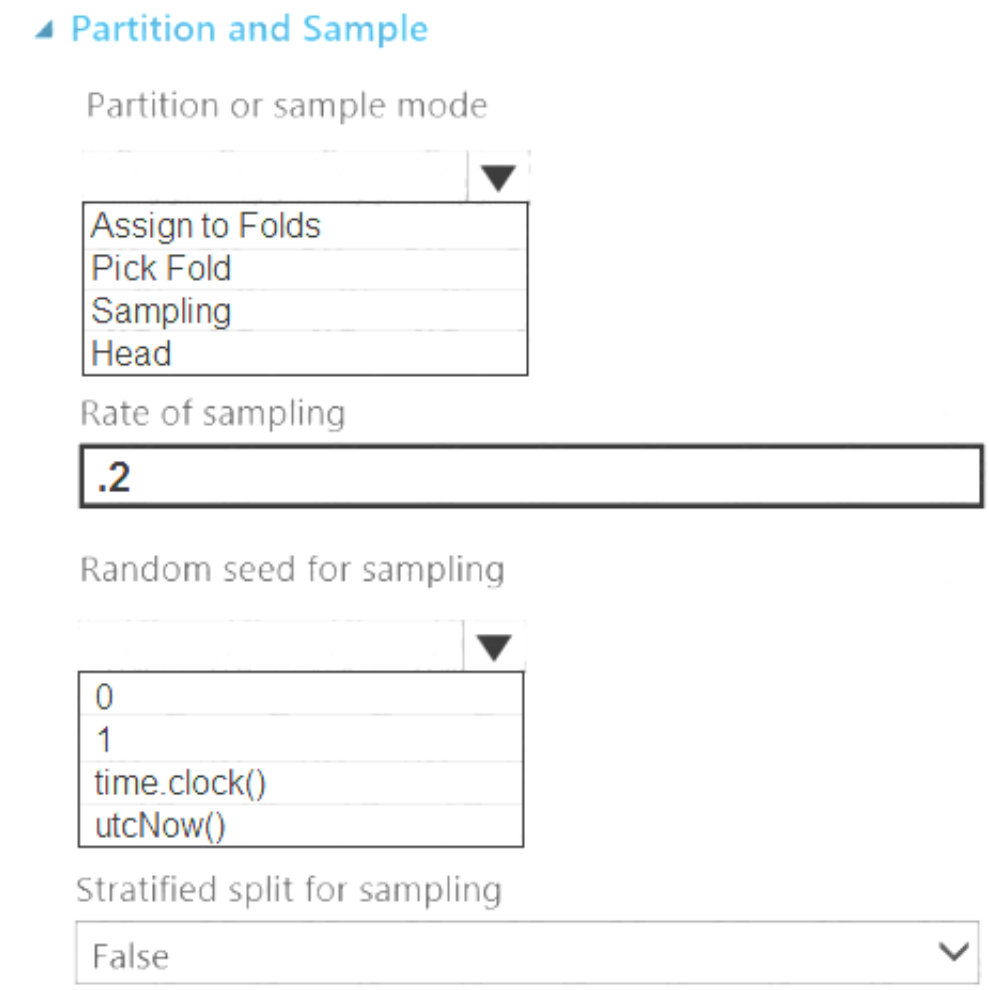
Answer: A

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

- For Partition or sample mode, select Assign to Folds.
- Use replacement in the partitioning: Select this option if you want the sampled row to be put back into the pool of rows for potential reuse. As a result, the same row might be assigned to several folds.
- If you do not use replacement (the default option), the sampled row is not put back into the pool of rows for potential reuse. As a result, each row can be assigned to only one fold.
- Randomized split: Select this option if you want rows to be randomly assigned to folds. If you do not select this option, rows are assigned to folds using the round-robin method. References:

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

NEW QUESTION 199
- (Exam Topic 3)
You are retrieving data from a large datastore by using Azure Machine Learning Studio.
You must create a subset of the data for testing purposes using a random sampling seed based on the system clock.
You add the Partition and Sample module to your experiment. You need to select the properties for the module.
Which values should you select? To answer, select the appropriate options in the answer area.
NOTE: Each correct selection is worth one point.



A. Mastered
B. Not Mastered

Answer: A

Explanation:
Box 1: Sampling Create a sample of data
This option supports simple random sampling or stratified random sampling. This is useful if you want to create a smaller representative sample dataset for testing.
* 1. Add the Partition and Sample module to your experiment in Studio, and connect the dataset.
* 2. Partition or sample mode: Set this to Sampling.
* 3. Rate of sampling.
See box 2 below.
Box 2: 0
* 3. Rate of sampling. Random seed for sampling: Optionally, type an integer to use as a seed value.
This option is important if you want the rows to be divided the same way every time. The default value is 0, meaning that a starting seed is generated based on the system clock. This can lead to slightly different results each time you run the experiment.
References:
https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/partition-and-sample

NEW QUESTION 204
- (Exam Topic 3)
You create an Azure Machine Learning workspace named workspace1. You assign a custom role to a user of workspace1.
The custom role has the following JSON definition:

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

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

Statements	Yes	No
The user can perform all actions in the workspace	○	○
The user can delete a compute resource in the workspace	○	○
The user can write metrics to the workspace	○	○

A.

Answer:

Explanation:
Graphical user interface, text, application, email Description automatically generated
Box 1: No
The actions listed in NotActions are prohibited.
If the roles include Actions that have a wildcard (*), the effective permissions are computed by subtracting the NotActions from the allowed Actions.
Box 2: No
Deleting compute resources in the workspace is in the NotActions list. Box 3: Yes
Writing metrics is not listed in NotActions. Reference:
https://docs.microsoft.com/en-us/azure/role-based-access-control/overview#how-azure-rbac-determines-if-a-use

NEW QUESTION 209
- (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 214
- (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 215
- (Exam Topic 3)
You have a comma-separated values (CSV) file containing data from which you want to train a classification model.
You are using the Automated Machine Learning interface in Azure Machine Learning studio to train the classification model. You set the task type to Classification.
You need to ensure that the Automated Machine Learning process evaluates only linear models. What should you do?

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

Answer: C

Explanation:
Automatic featurization can fit non-linear models. Reference: https://econml.azurewebsites.net/spec/estimation/dml.html
https://docs.microsoft.com/en-us/azure/machine-learning/how-to-use-automated-ml-for-ml-models

NEW QUESTION 216
- (Exam Topic 3)
You are creating a new Azure Machine Learning pipeline using the designer.
The pipeline must train a model using data in a comma-separated values (CSV) file that is published on a website. You have not created a dataset for this file.
You need to ingest the data from the CSV file into the designer pipeline using the minimal administrative effort.
Which module should you add to the pipeline in Designer?

A. Convert to CSV
B. Enter Data Manually D
C. Import Data
D. Dataset

Answer: D

Explanation:
The preferred way to provide data to a pipeline is a Dataset object. The Dataset object points to data that lives in or is accessible from a datastore or at a Web URL. The Dataset class is abstract, so you will create an instance of either a FileDataset (referring to one or more files) or a TabularDataset that's created by from one or more files with delimited columns of data.
Example:
from azureml.core import Dataset
iris_tabular_dataset = Dataset.Tabular.from_delimited_files([(def_blob_store, 'train-dataset/iris.csv')]) Reference:
https://docs.microsoft.com/en-us/azure/machine-learning/how-to-create-your-first-pipeline

NEW QUESTION 218
- (Exam Topic 3)
You run an automated machine learning experiment in an Azure Machine Learning workspace. Information about the run is listed in the table below:

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

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

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

B)

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

C)

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

D)

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

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

Answer: A

Explanation:
The get_output method on automl_classifier returns the best run and the fitted model for the last invocation. Overloads on get_output allow you to retrieve the best run and fitted model for any logged metric or for a particular iteration.
In [ ]:
best_run, fitted_model = local_run.get_output() Reference:
https://notebooks.azure.com/azureml/projects/azureml-getting-started/html/how-to-use-azureml/automated-mach

NEW QUESTION 221
- (Exam Topic 3)
You create an Azure Machine Learning compute resource to train models. The compute resource is configured as follows:
- Minimum nodes: 2
- Maximum nodes: 4

You must decrease the minimum number of nodes and increase the maximum number of nodes to the following values:
- Minimum nodes: 0
- Maximum nodes: 8

You need to reconfigure the compute resource.
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. Use the Azure Machine Learning studio.
B. Run the update method of the AmlCompute class in the Python SDK.
C. Use the Azure portal.
D. Use the Azure Machine Learning designer.
E. Run the refresh_state() method of the BatchCompute class in the Python SDK

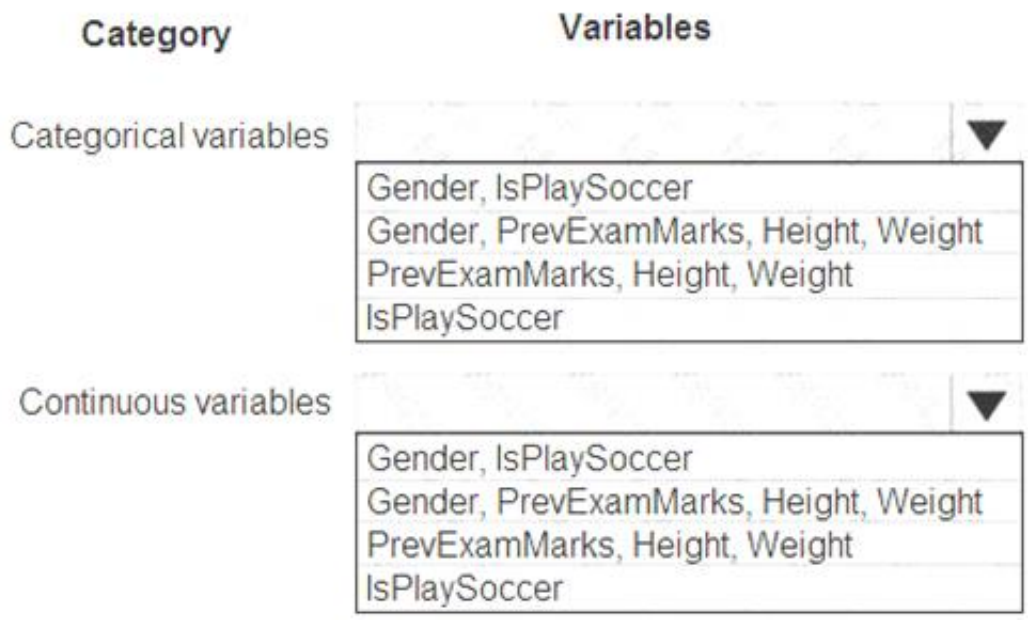
Answer: ABC

Explanation:
Reference:
https://docs.microsoft.com/en-us/python/api/azureml-core/azureml.core.compute.amlcompute(class)

NEW QUESTION 222
- (Exam Topic 3)
You are working on a classification task. You have a dataset indicating whether a student would like to play soccer and associated attributes. The dataset includes the following columns:

Name	Description
IsPlaySoccer	Values can be 1 and 0.
Gender	Values can be M or F.
PrevExamMarks	Stores values from 0 to 100
Height	Stores values in centimeters
Weight	Stores values in kilograms

You need to classify variables by type.
Which variable should you add to each category? To answer, select the appropriate options in the answer area. NOTE: Each correct selection is worth one point.

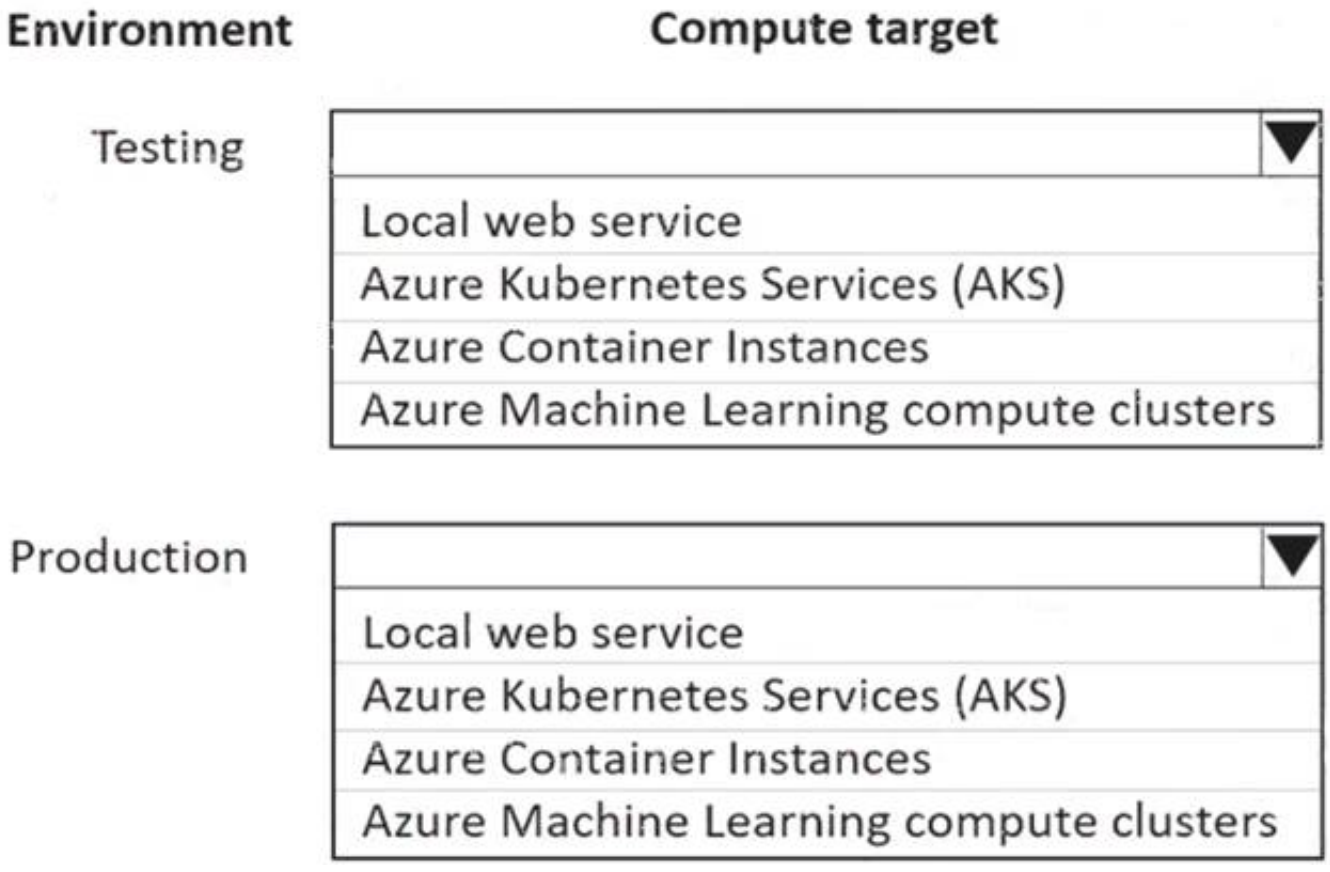


A. Mastered
B. Not Mastered

Answer: A

Explanation:
References: https://www.edureka.co/blog/classification-algorithms/

NEW QUESTION 223
- (Exam Topic 3)
You are using an Azure Machine Learning workspace. You set up an environment for model testing and an environment for production.
The compute target for testing must minimize cost and deployment efforts. The compute target for production must provide fast response time, autoscaling of the deployed service, and support real-time inferencing.
You need to configure compute targets for model testing and production.
Which compute targets should you use? To answer, select the appropriate options in the answer area.
NOTE: Each correct selection is worth one point.



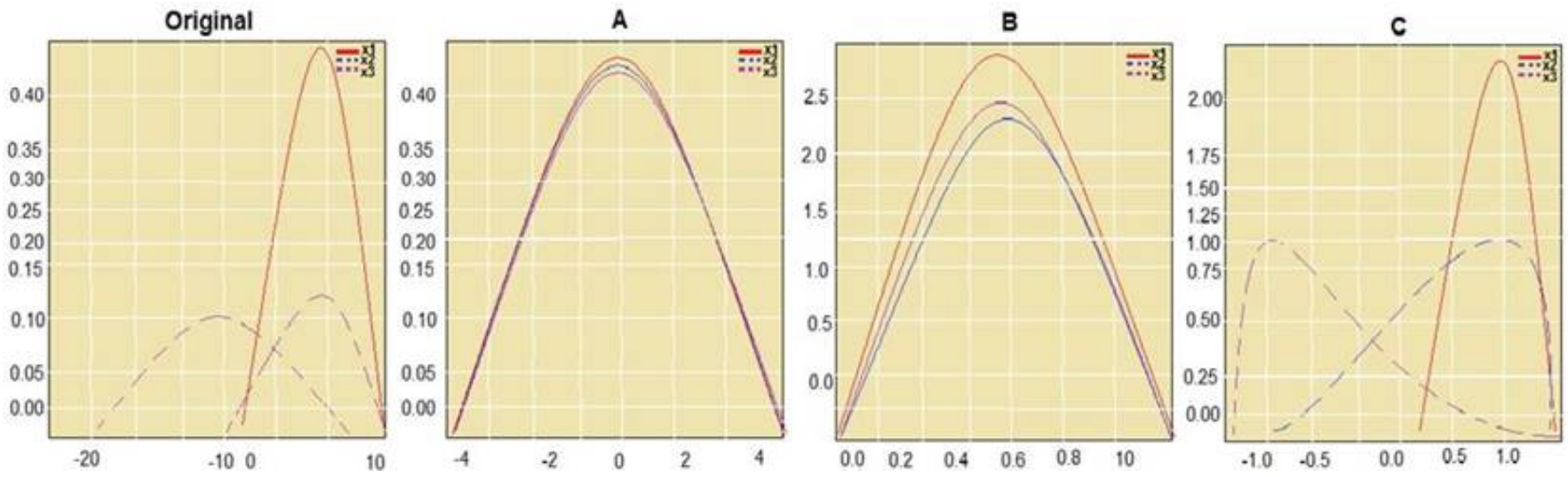
A. Mastered
B. Not Mastered

Answer: A

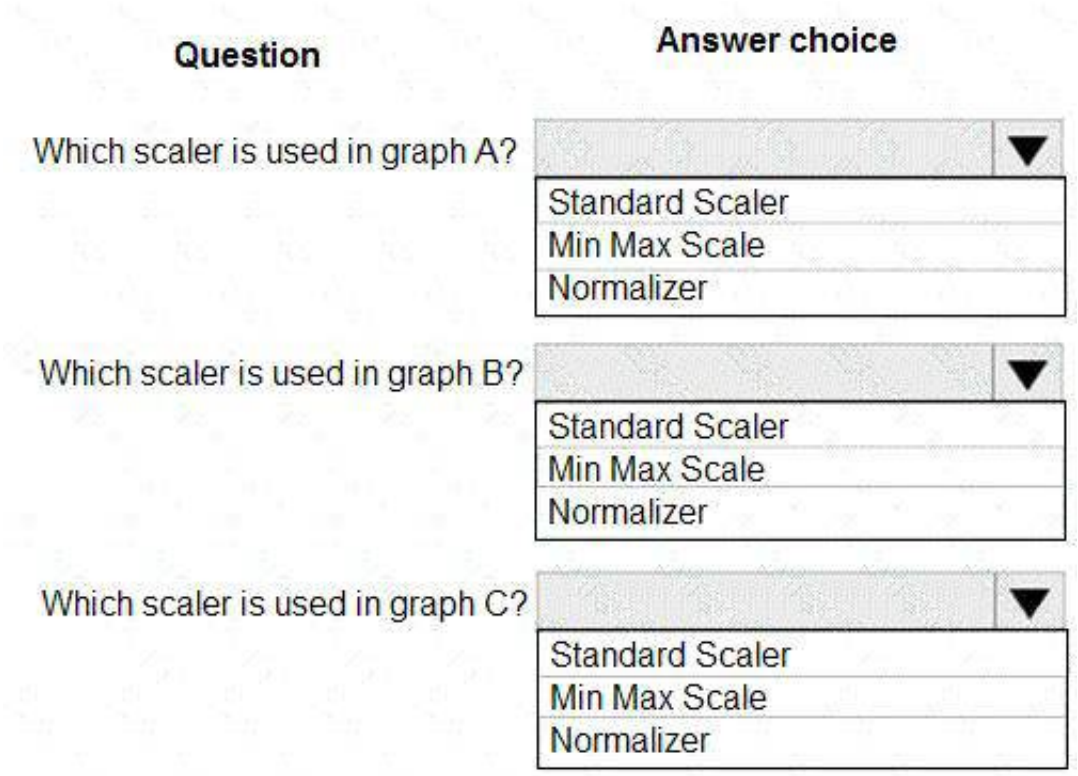
Explanation:
Text, application Description automatically generated
Box 1: Local web service
The Local web service compute target is used for testing/debugging. Use it for limited testing and troubleshooting. Hardware acceleration depends on use of libraries in the local system.
Box 2: Azure Kubernetes Service (AKS)
Azure Kubernetes Service (AKS) is used for Real-time inference. Recommended for production workloads.
Use it for high-scale production deployments. Provides fast response time and autoscaling of the deployed service
Reference:
https://docs.microsoft.com/en-us/azure/machine-learning/concept-compute-target

NEW QUESTION 224

- (Exam Topic 3)
You are performing feature scaling by using the scikit-learn Python library for x.1 x2, and x3 features. Original and scaled data is shown in the following image.



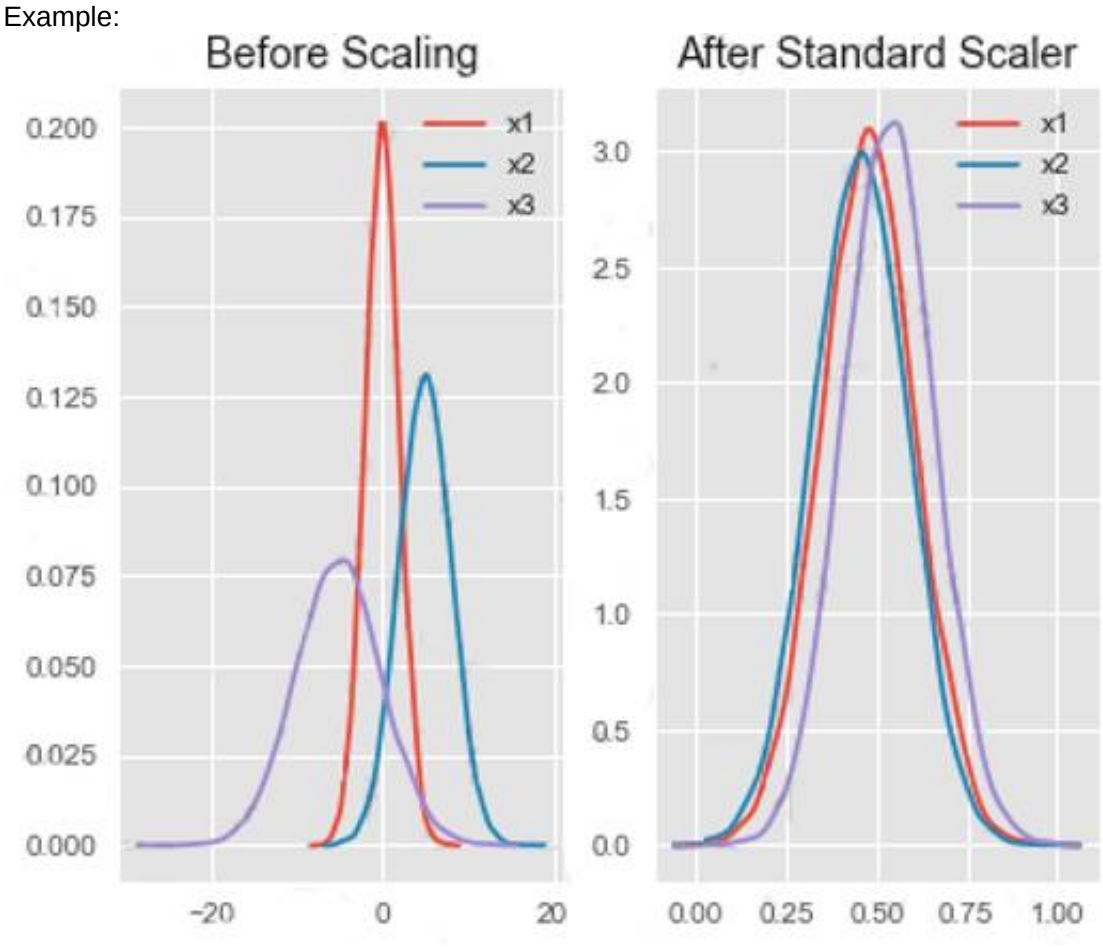
Use the drop-down menus to select the answer choice that answers each question based on the information presented in the graphic.
NOTE: Each correct selection is worth one point.



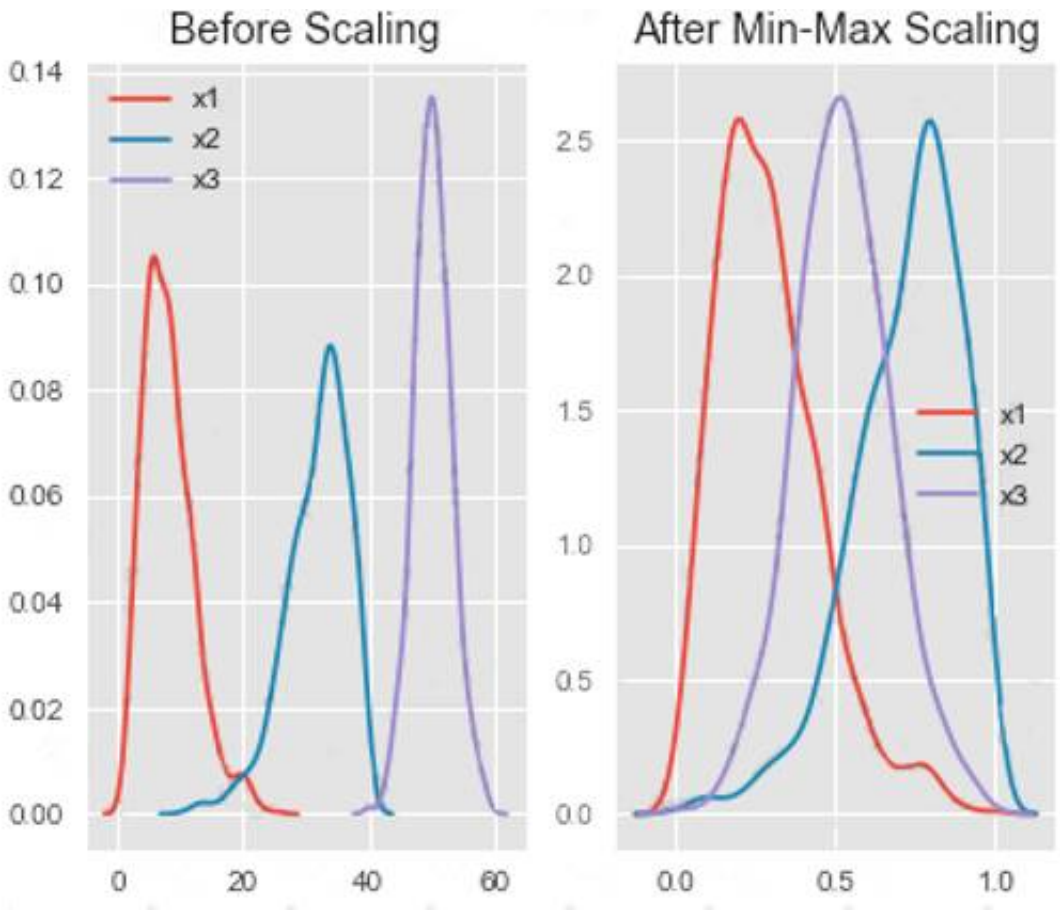
A. Mastered
B. Not Mastered

Answer: A

Explanation:
Box 1: StandardScaler
The StandardScaler assumes your data is normally distributed within each feature and will scale them such that the distribution is now centred around 0, with a standard deviation of 1.
Example:



All features are now on the same scale relative to one another. Box 2: Min Max Scaler



Notice that the skewness of the distribution is maintained but the 3 distributions are brought into the same scale so that they overlap.
Box 3: Normalizer
References:
http://benalexkeen.com/feature-scaling-with-scikit-learn/

NEW QUESTION 227
- (Exam Topic 3)
You are a lead data scientist for a project that tracks the health and migration of birds. You create a multi-class image classification deep learning model that uses a set of labeled bird photographs collected by experts.
You have 100,000 photographs of birds. All photographs use the JPG format and are stored in an Azure blob container in an Azure subscription.
You need to access the bird photograph files in the Azure blob container from the Azure Machine Learning service workspace that will be used for deep learning model training. You must minimize data movement.
What should you do?

A. Create an Azure Data Lake store and move the bird photographs to the store.
B. Create an Azure Cosmos DB database and attach the Azure Blob containing bird photographs storage to the database.
C. Create and register a dataset by using TabularDataset class that references the Azure blob storage containing bird photographs.
D. Register the Azure blob storage containing the bird photographs as a datastore in Azure Machine Learning service.
E. Copy the bird photographs to the blob datastore that was created with your Azure Machine Learning service workspace.

Answer: D

Explanation:
We recommend creating a datastore for an Azure Blob container. When you create a workspace, an Azure blob container and an Azure file share are automatically registered to the workspace.
Reference:
https://docs.microsoft.com/en-us/azure/machine-learning/how-to-access-data

NEW QUESTION 230
- (Exam Topic 3)
You use Azure Machine Learning to train a model based on a dataset named dataset1. You define a dataset monitor and create a dataset named dataset2 that contains new data.
You need to compare dataset1 and dataset2 by using the Azure Machine Learning SDK for Python. Which method of the DataDriftDetector class should you use?

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

Answer: C

Explanation:
A backfill run is used to see how data changes over time. Reference:
https://docs.microsoft.com/en-us/python/api/azureml-datadrift/azureml.datadrift.datadriftdetector.datadriftdetect

NEW QUESTION 233
- (Exam Topic 3)
You create the following config.json file.

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

You must use the Azure Machine Learning SDK to interact with data and experiments in the workspace. You need to configure the config.json file to connect to the workspace from the Python environment. Which two additional parameters must you add to the config.json file in order to connect to the workspace?
Each correct answer presents part of the solution.
NOTE: Each correct selection is worth one point.

A. subscription_Id
B. Key
C. resource_group
D. region
E. Login

Answer: AC

Explanation:
To use the same workspace in multiple environments, create a JSON configuration file. The configuration file saves your subscription (subscription_id), resource (resource_group), and workspace name so that it can be easily loaded.
The following sample shows how to create a workspace. from azureml.core import Workspace
ws = Workspace.create(name='myworkspace', subscription_id='<azure-subscription-id>', resource_group='myresourcegroup', create_resource_group=True, location='eastus2'
)
Reference:
https://docs.microsoft.com/en-us/python/api/azureml-core/azureml.core.workspace.workspace

NEW QUESTION 235
- (Exam Topic 3)
You run a script as an experiment in Azure Machine Learning.
You have a Run object named run that references the experiment run. You must review the log files that were generated during the experiment run.
You need to download the log files to a local folder for review.
Which two code segments can you run to achieve this goal? Each correct answer presents a complete solution.
NOTE: Each correct selection is worth one point.

A. run.get_details()
B. run.get_file_names()
C. run.get_metrics()
D. run.download_files(output_directory='./runfiles')
E. run.get_all_logs(destination='./runlogs')

Answer: AE

Explanation:
The run Class get_all_logs method downloads all logs for the run to a directory.
The run Class get_details gets the definition, status information, current log files, and other details of the run. Reference:
https://docs.microsoft.com/en-us/python/api/azureml-core/azureml.core.run(class)

NEW QUESTION 236
- (Exam Topic 3)
An organization creates and deploys a multi-class image classification deep learning model that uses a set of labeled photographs.
The software engineering team reports there is a heavy inferencing load for the prediction web services during the summer. The production web service for the model fails to meet demand despite having a fully-utilized compute cluster where the web service is deployed.
You need to improve performance of the image classification web service with minimal downtime and minimal administrative effort.
What should you advise the IT Operations team to do?

A. Increase the minimum node count of the compute cluster where the web service is deployed.
B. Create a new compute cluster by using larger VM sizes for the nodes, redeploy the web service to that cluster, and update the DNS registration for the service endpoint to point to the new cluster.
C. Increase the VM size of nodes in the compute cluster where the web service is deployed.
D. Increase the node count of the compute cluster where the web service is deployed.

Answer: D

Explanation:
The Azure Machine Learning SDK does not provide support scaling an AKS cluster. To scale the nodes in the cluster, use the UI for your AKS cluster in the Azure Machine Learning studio. You can only change the node count, not the VM size of the cluster.
Reference:
https://docs.microsoft.com/en-us/azure/machine-learning/how-to-create-attach-kubernetes

NEW QUESTION 240
- (Exam Topic 3)
Note: This question is part of a series of questions that present the same scenario. Each question in the series contains a unique solution that might meet the stated goals. Some question sets might have more than one correct solution, while others might not have a correct solution.
After you answer a question in this section, you will NOT be able to return to it. As a result, these questions will not appear in the review screen.
You are using Azure Machine Learning to run an experiment that trains a classification model.
You want to use Hyperdrive to find parameters that optimize the AUC metric for the model. You configure a HyperDriveConfig for the experiment by running the following code:

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

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

```
import numpy as np
from sklearn.metrics import roc_auc_score
from azureml.core.run import Run
run = Run.get_context()
# code to train model omitted
auc = roc_auc_score(y_test, y_predicted)
run.log("AUC", np.float(auc))
```

Does the solution meet the goal?

A. Yes
B. No

Answer: A

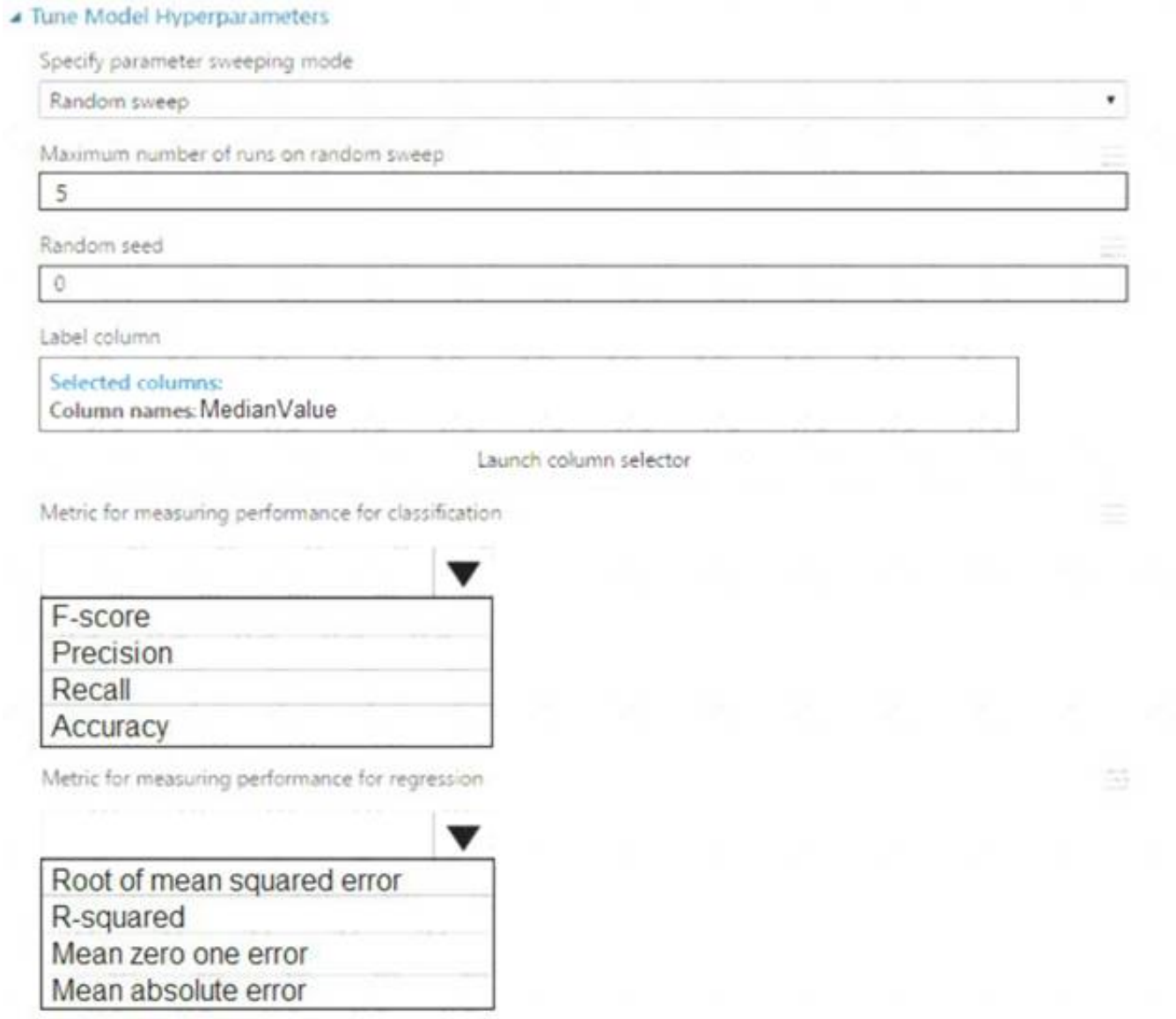
NEW QUESTION 244
- (Exam Topic 2)
You need to select a feature extraction method. Which method should you use?

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

Answer: C

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

NEW QUESTION 245
- (Exam Topic 2)
You need to set up the Permutation Feature Importance module according to the model training requirements. Which properties should you select? To answer, select the appropriate options in the answer area.
NOTE: Each correct selection is worth one point.



A. Mastered
B. Not Mastered

Answer: A

Explanation:
Box 1: Accuracy
Scenario: You want to configure hyperparameters in the model learning process to speed the learning phase by using hyperparameters. In addition, this configuration should cancel the lowest performing runs at each evaluation interval, thereby directing effort and resources towards models that are more likely to be successful.
Box 2: R-Squared

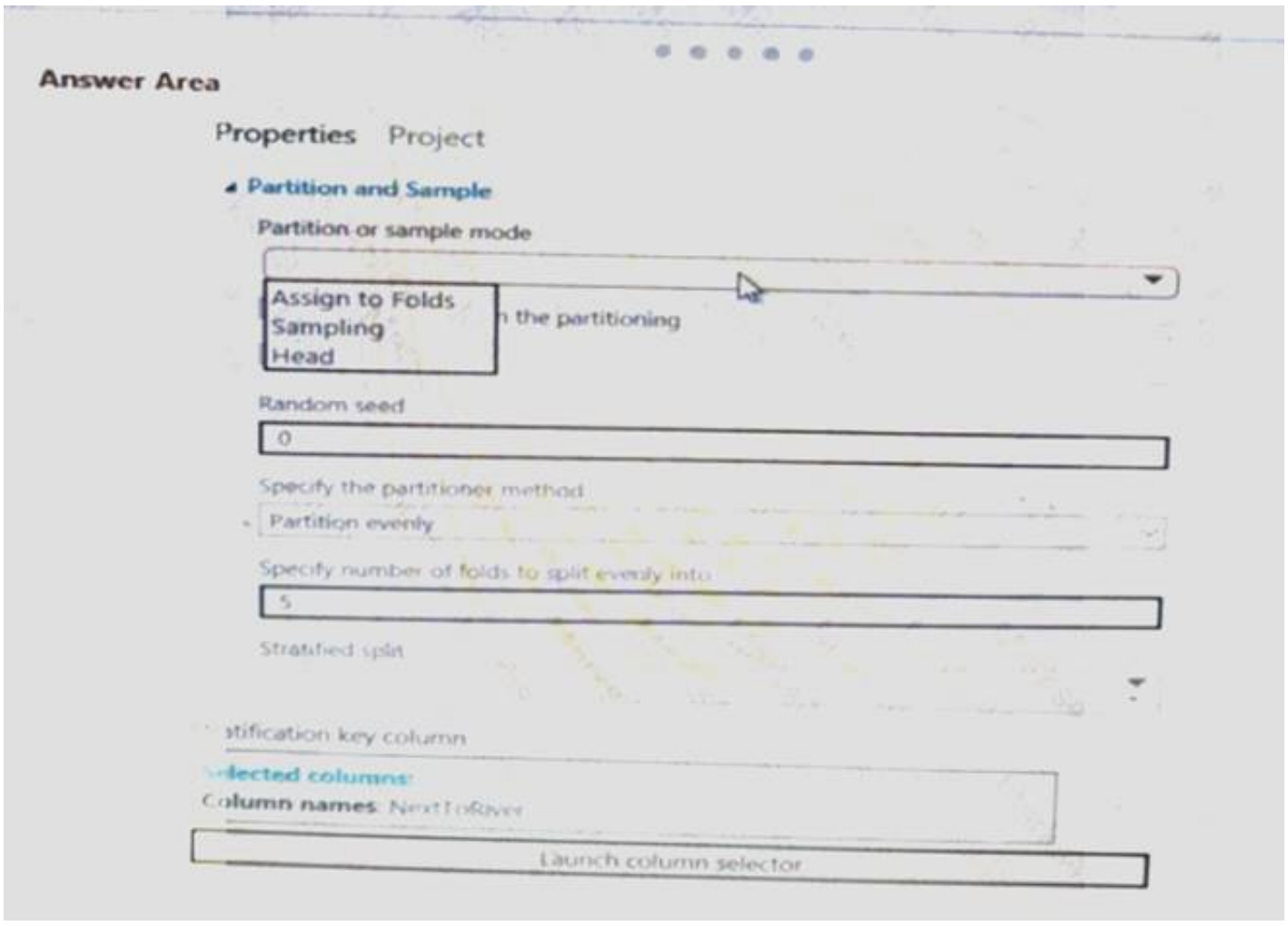
NEW QUESTION 247
- (Exam Topic 2)
You need to select a feature extraction method. Which method should you use?

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

Answer: A

Explanation:
Spearman's rank correlation coefficient assesses how well the relationship between two variables can be described using a monotonic function.
Note: Both Spearman's and Kendall's can be formulated as special cases of a more general correlation coefficient, and they are both appropriate in this scenario.
Scenario: The MedianValue and AvgRoomsInHouse columns both hold data in numeric format. You need to select a feature selection algorithm to analyze the relationship between the two columns in more detail. References:
https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/feature-selection-modules

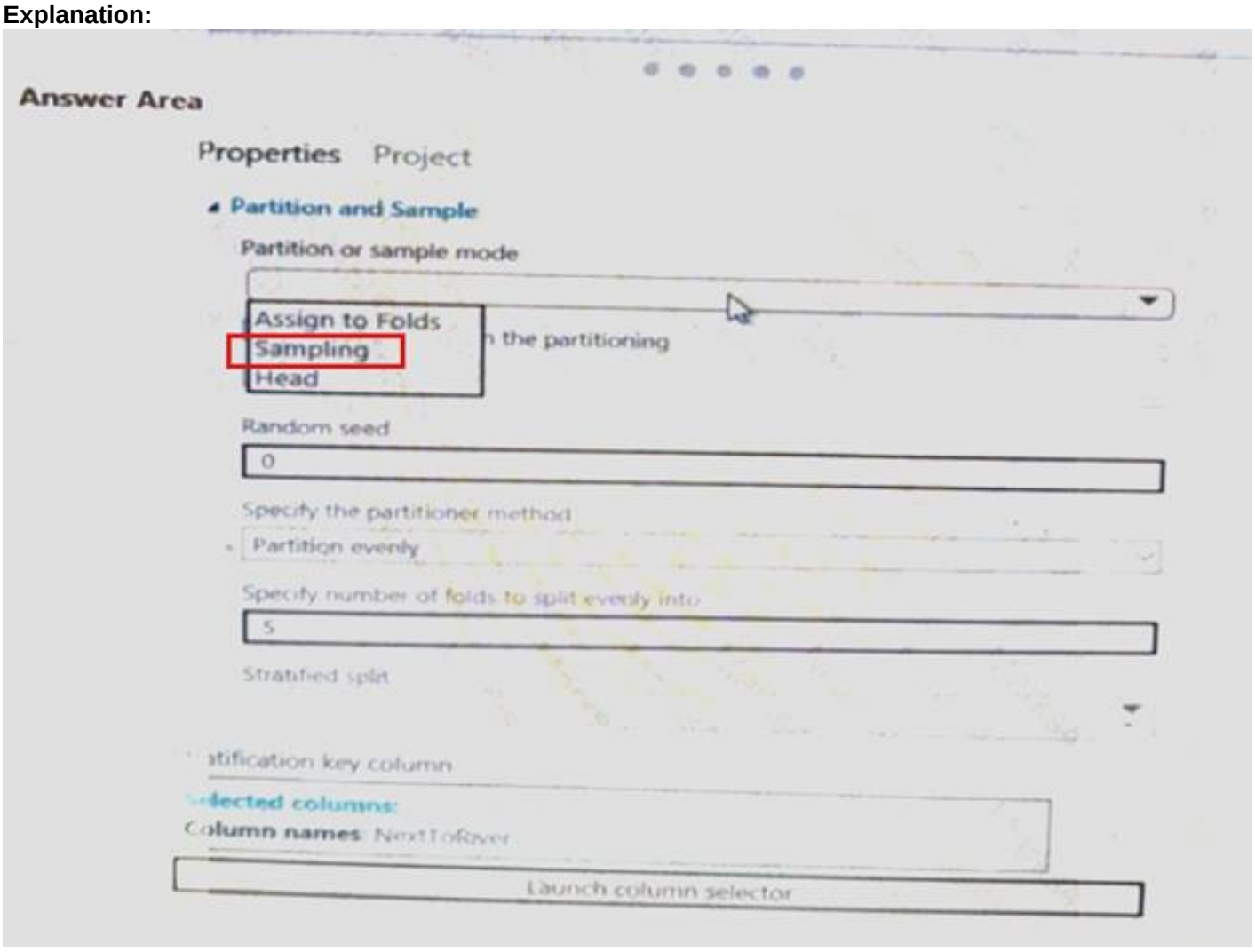
NEW QUESTION 250
- (Exam Topic 2)
You need to identify the methods for dividing the data according, to the testing requirements.
Which properties should you select? To answer, select the appropriate option-, m the answer area. NOTE: Each correct selection is worth one point.



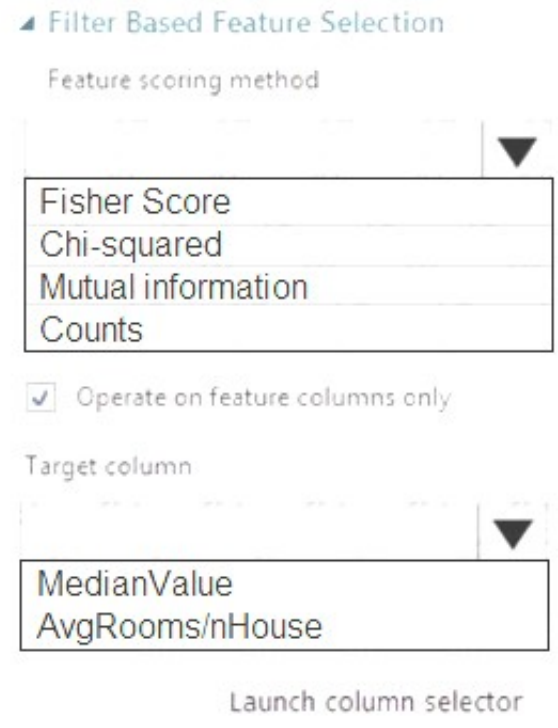
A. Mastered
B. Not Mastered

Answer: A

Explanation:



NEW QUESTION 251
- (Exam Topic 2)
You need to configure the Feature Based Feature Selection module based on the experiment requirements and datasets.
How should you configure the module properties? 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:
Box 1: Mutual Information.
The mutual information score is particularly useful in feature selection because it maximizes the mutual information between the joint distribution and target variables in datasets with many dimensions.
Box 2: MedianValue
MedianValue is the feature column, , it is the predictor of the dataset.
Scenario: The MedianValue and AvgRoomsinHouse columns both hold data in numeric format. You need to select a feature selection algorithm to analyze the relationship between the two columns in more detail.
References:
https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/filter-based-feature-selection

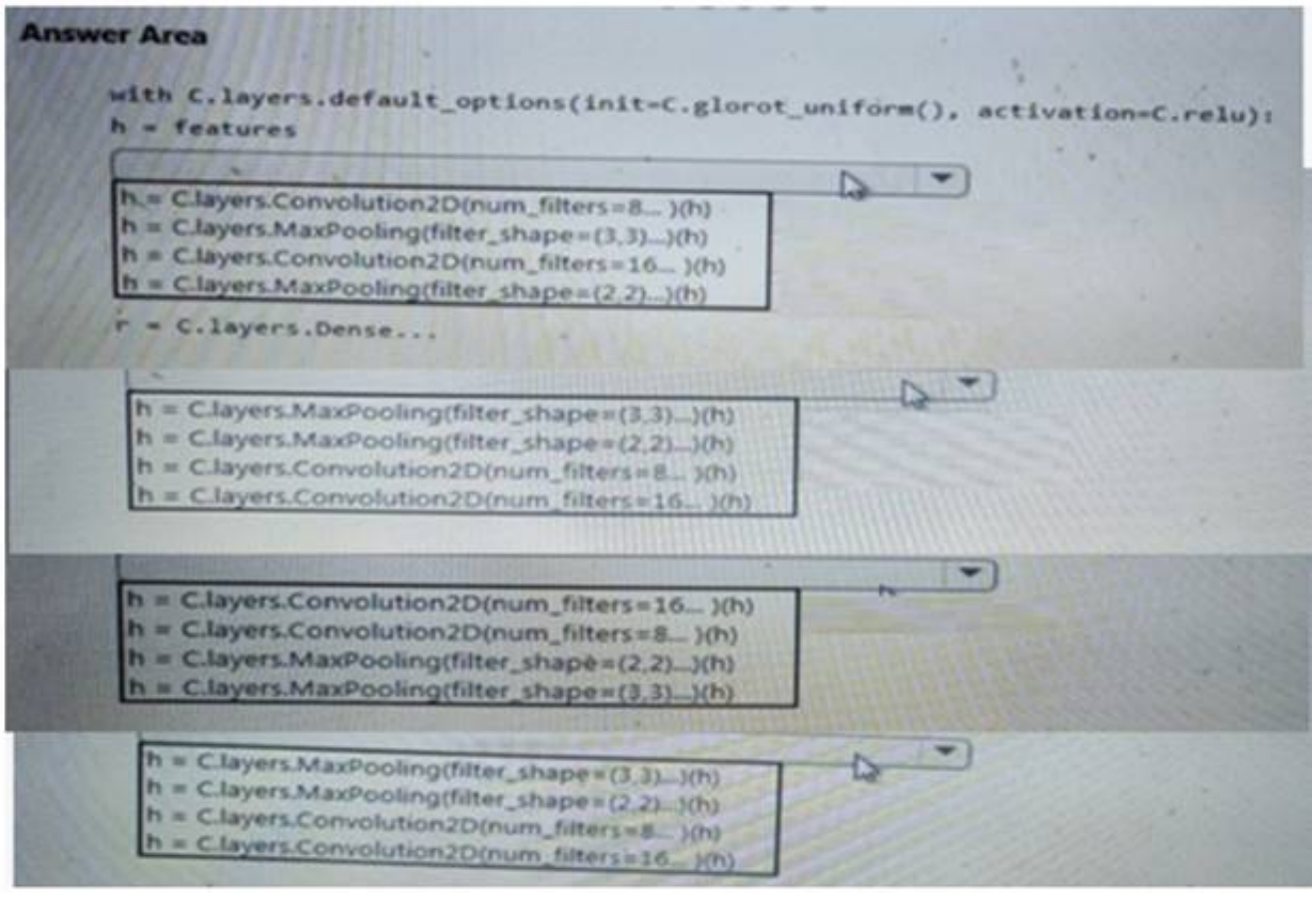
NEW QUESTION 253
- (Exam Topic 1)
You need to implement a model development strategy to determine a user's tendency to respond to an ad. Which technique should you use?

A. Use a Relative Expression Split module to partition the data based on centroid distance.
B. Use a Relative Expression Split module to partition the data based on distance travelled to the event.
C. Use a Split Rows module to partition the data based on distance travelled to the event.
D. Use a Split Rows module to partition the data based on centroid distance.

Answer: A

Explanation:
Split Data partitions the rows of a dataset into two distinct sets.
The Relative Expression Split option in the Split Data module of Azure Machine Learning Studio is helpful
when you need to divide a dataset into training and testing datasets using a numerical expression.
Relative Expression Split: Use this option whenever you want to apply a condition to a number column. The number could be a date/time field, a column containing age or dollar amounts, or even a percentage. For example, you might want to divide your data set depending on the cost of the items, group people by age ranges, or separate data by a calendar date.
Scenario:
Local market segmentation models will be applied before determining a user's propensity to respond to an advertisement.
The distribution of features across training and production data are not consistent References:
https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/split-data

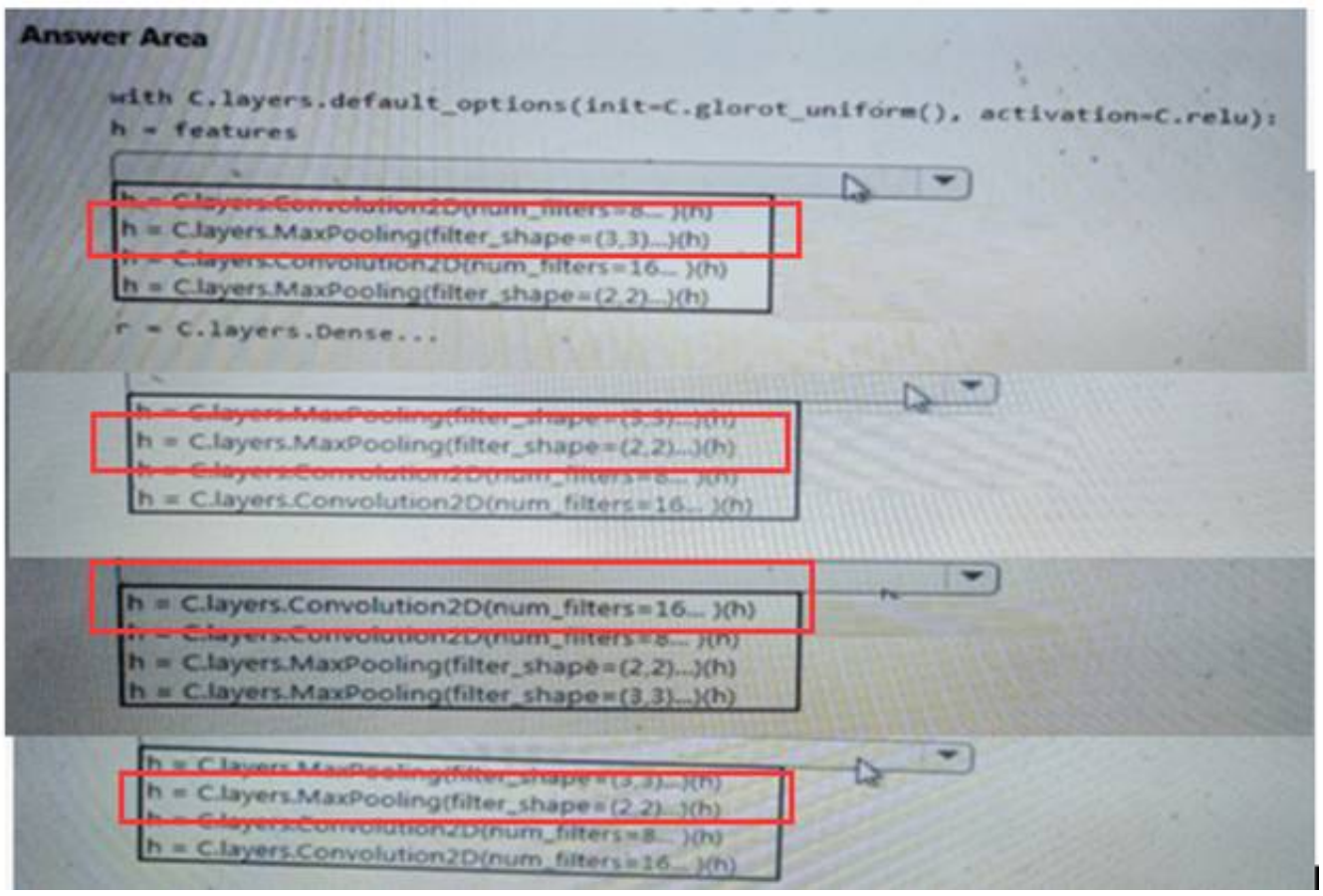
NEW QUESTION 256
- (Exam Topic 1)
You need to build a feature extraction strategy for the local models.
How should you complete the code segment? To answer, select the appropriate options in the answer area. NOTE: Each correct selection is worth one point.



A. Mastered
B. Not Mastered

Answer: A

Explanation:



NEW QUESTION 260
- (Exam Topic 1)
You need to implement a scaling strategy for the local penalty detection data.
Which normalization type should you use?

A. Streaming
B. Weight
C. Batch
D. Cosine

Answer: C

Explanation:
Post batch normalization statistics (PBN) is the Microsoft Cognitive Toolkit (CNTK) version of how to evaluate the population mean and variance of Batch Normalization which could be used in inference Original Paper.
In CNTK, custom networks are defined using the BrainScriptNetworkBuilder and described in the CNTK network description language "BrainScript."
Scenario:
Local penalty detection models must be written by using BrainScript. References:
https://docs.microsoft.com/en-us/cognitive-toolkit/post-batch-normalization-statistics

NEW QUESTION 263

.....

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