Exam Questions DP-100

Designing and Implementing a Data Science Solution on Azure

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

NEW QUESTION 1
- (Exam Topic 3)
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 2
- (Exam Topic 3)
You create a pipeline in designer to train a model that predicts automobile prices.
Because of non-linear relationships in the data, the pipeline calculates the natural log (Ln) of the prices in the training data, trains a model to predict this natural log of price value, and then calculates the exponential of the scored label to get the predicted price.
The training pipeline is shown in the exhibit. (Click the Training pipeline tab.)
Training pipeline
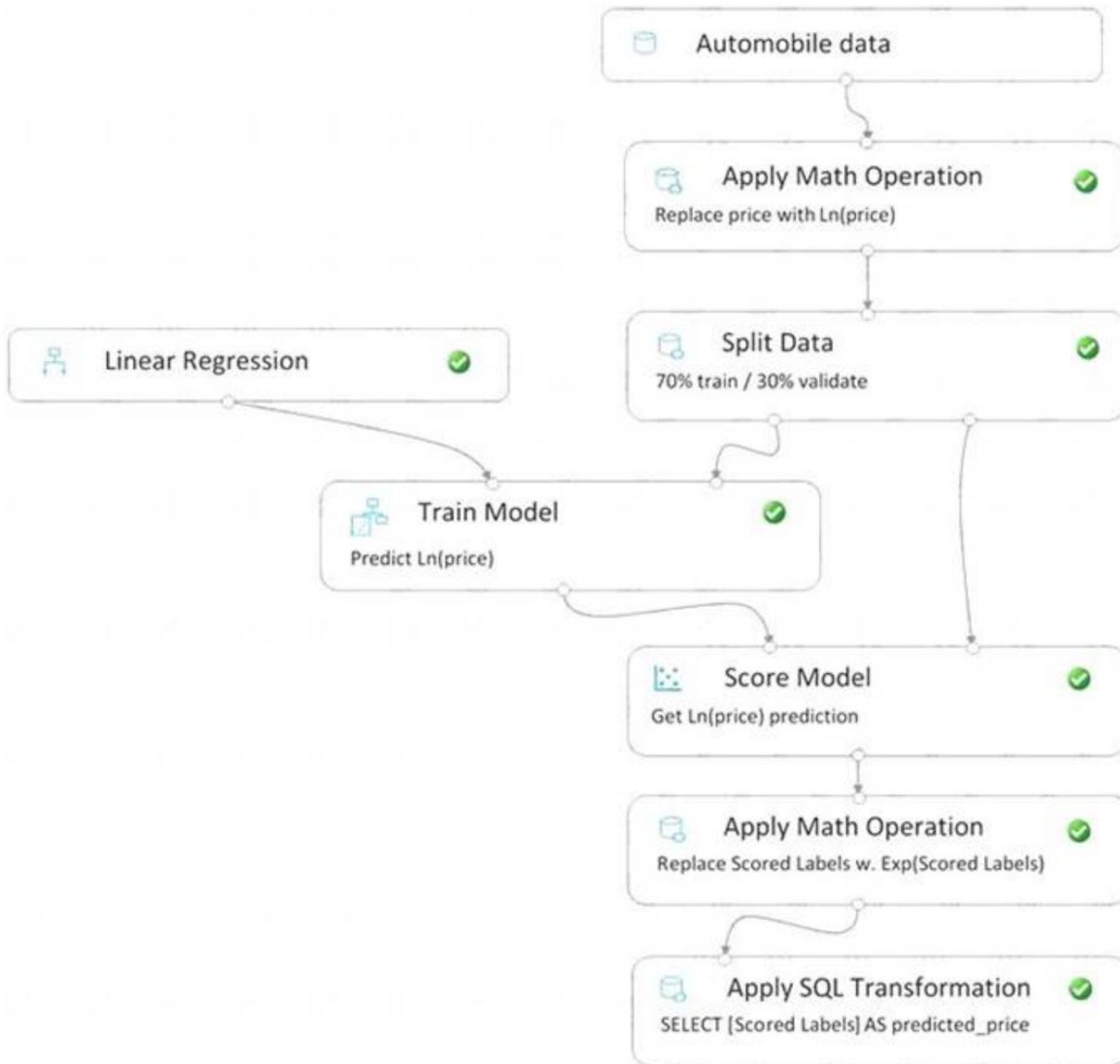
You create a real-time inference pipeline from the training pipeline, as shown in the exhibit. (Click the Real-time pipeline tab.)
Real-time pipeline
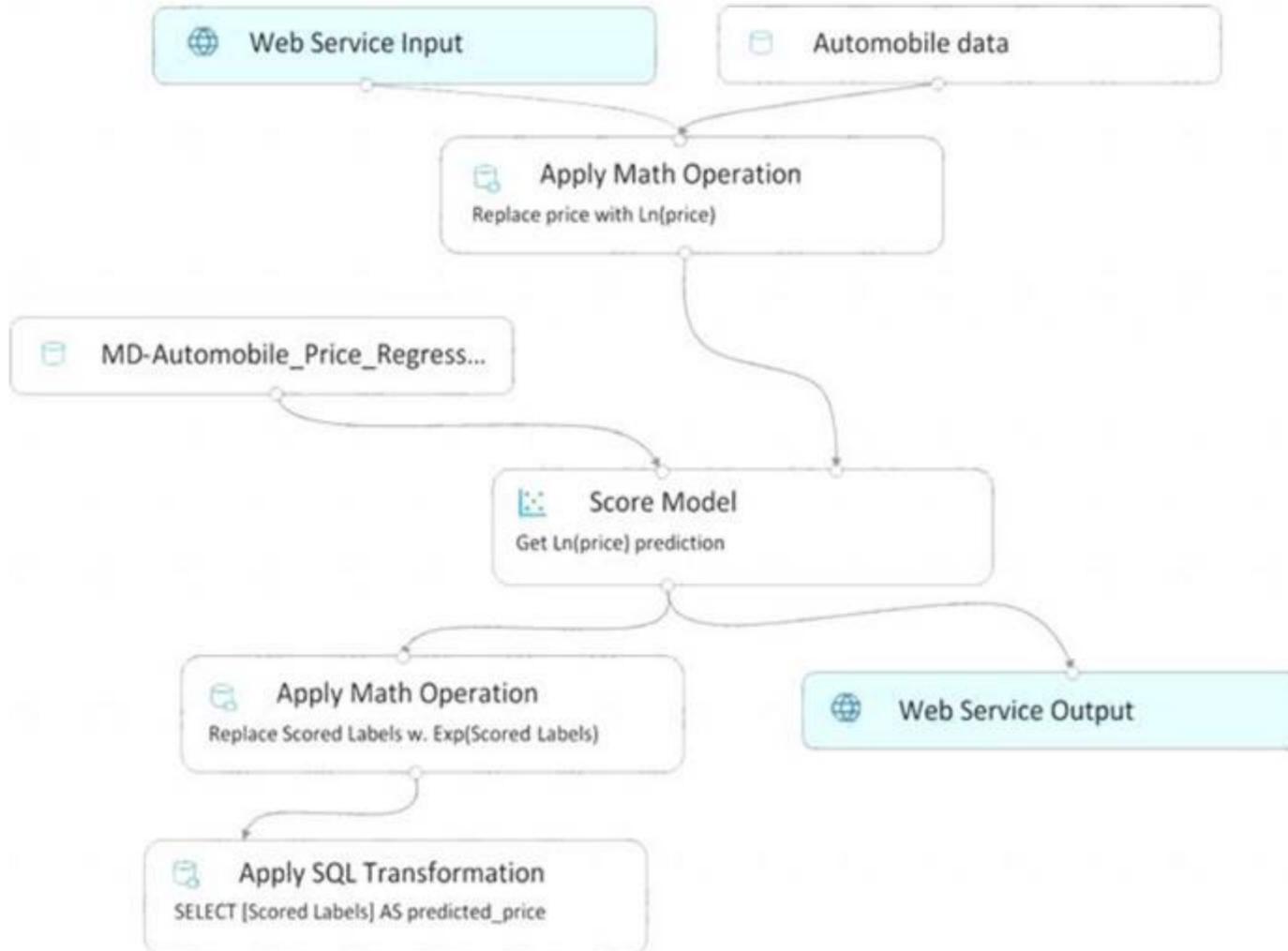
You need to modify the inference pipeline to ensure that the web service returns the exponential of the scored label as the predicted automobile price and that client applications are not required to include a price value in the input values.
Which three modifications must you make to the inference pipeline? Each correct answer presents part of the solution.
NOTE: Each correct selection is worth one point.

A. Connect the output of the Apply SQL Transformation to the Web Service Output module.
B. Replace the Web Service Input module with a data input that does not include the price column.
C. Add a Select Columns module before the Score Model module to select all columns other than price.
D. Replace the training dataset module with a data input that does not include the price column.
E. Remove the Apply Math Operation module that replaces price with its natural log from the data flow.

F. Remove the Apply SQL Transformation module from the data flow.

Answer: ACE

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 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 4
- (Exam Topic 3)
You are preparing to build a deep learning convolutional neural network model for image classification. You create a script to train the model using CUDA devices.
You must submit an experiment that runs this script in the Azure Machine Learning workspace. The following compute resources are available:

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

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

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

➢ an Azure Machine Learning compute target named gpu-cluster with four CPU and GPU-based nodes
You need to specify the compute resources to be used for running the code to submit the experiment, and for running the script in order to minimize model training time.
Which resources should the data scientist use? To answer, select the appropriate options in the answer area.
NOTE: Each correct selection is worth one point.

Resource type	Option
Run code to submit the experiment	▼
	the Microsoft Surface device
	the ds-workstation notebook VM
	the cpu-cluster compute target
	the gpu-cluster compute target
Run the training script	▼
	the ds-workstation notebook VM
	the cpu-compute target
	the gpu-compute target
	the Microsoft Surface device

A. Mastered
B. Not Mastered

Answer: A

Explanation:

Resource type	Option
Run code to submit the experiment	the Microsoft Surface device the ds-workstation notebook VM the cpu-cluster compute target the gpu-cluster compute target
Run the training script	the ds-workstation notebook VM the cpu-compute target the gpu-compute target the Microsoft Surface device

NEW QUESTION 5
- (Exam Topic 3)
You use the designer to create a training pipeline for a classification model. The pipeline uses a dataset that includes the features and labels required for model training.
You create a real-time inference pipeline from the training pipeline. You observe that the schema for the generated web service input is based on the dataset and includes the label column that the model predicts. Client applications that use the service must not be required to submit this value.
You need to modify the inference pipeline to meet the requirement. What should you do?

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

Answer: A

Explanation:
By default, the Web Service Input will expect the same data schema as the module output data which connects to the same downstream port as it. You can remove the target variable column in the inference pipeline using Select Columns in Dataset module. Make sure that the output of Select Columns in Dataset removing target variable column is connected to the same port as the output of the Web Service Intput module.
Reference:
https://docs.microsoft.com/en-us/azure/machine-learning/tutorial-designer-automobile-price-deploy

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.
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 7
- (Exam Topic 3)
You are developing a machine learning, experiment by using Azure. The following images show the input and output of a machine learning experiment:

Use the drop-down menus to select the answer choice that answers each question based on the information presented in the graphic.
NOTE: Each correct selection is worth one point.
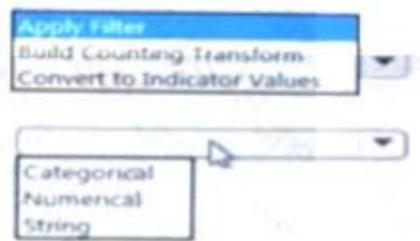
A. Mastered
B. Not Mastered

Answer: A

Explanation:
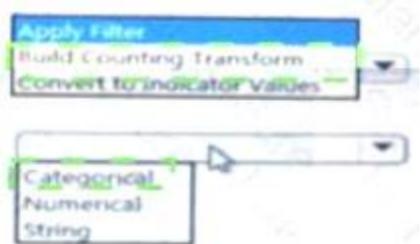
NEW QUESTION 8
- (Exam Topic 3)
A set of CSV files contains sales records. All the CSV files have the same data schema.
Each CSV file contains the sales record for a particular month and has the filename sales.csv. Each file in stored in a folder that indicates the month and year when the data was recorded. 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 9
- (Exam Topic 3)
You retrain an existing model.
You need to register the new version of a model while keeping the current version of the model in the registry.
What should you do?

A. Register a model with a different name from the existing model and a custom property named versionwith the value 2.

B. Register the model with the same name as the existing model.

C. Save the new model in the default datastore with the same name as the existing mode

D. Do not register the new model.

E. Delete the existing model and register the new one with the same name.

Answer: B

Explanation:
Model version: A version of a registered model. When a new model is added to the Model Registry, it is added as Version 1. Each model registered to the same model name increments the version number.
Reference:
https://docs.microsoft.com/en-us/azure/databricks/applications/mlflow/model-registry

NEW QUESTION 10
- (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 10
- (Exam Topic 3)
You run an experiment that uses an AutoMLConfig class to define an automated machine learning task with a maximum of ten model training iterations. The task will attempt to find the best performing model based on a metric named accuracy.
You submit the experiment with the following code:
You need to create Python code that returns the best model that is generated by the automated machine learning task. Which code segment should you use?
A)

```
best_model = automl_run.get_details()
```

B)

```
best_model = automl_run.get_output()[1]
```

C)

```
best_model = automl_run.get_file_names()[1]
```

D)

```
best_model = automl_run.get_metrics()
```

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

Answer: B

Explanation:
The get_output method returns the best run and the fitted model. Reference:
https://notebooks.azure.com/azureml/projects/azureml-getting-started/html/how-to-use-azureml/automated-mach

NEW QUESTION 15
- (Exam Topic 3)
You are solving a classification task.
You must evaluate your model on a limited data sample by using k-fold cross validation. You start by configuring a k parameter as the number of splits.
You need to configure the k parameter for the cross-validation. Which value should you use?

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

Answer: C

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

NEW QUESTION 16
- (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 17
- (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 22
- (Exam Topic 3)
You create a batch inference pipeline by using the Azure ML SDK. You run the pipeline by using the following code:
from azureml.pipeline.core import Pipeline
from azureml.core.experiment import Experiment
pipeline = Pipeline(workspace=ws, steps=[parallelrun_step]) pipeline_run = Experiment(ws, 'batch_pipeline').submit(pipeline) You need to monitor the progress of the pipeline execution.
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. Run the following code in a notebook:

```
from azureml.contrib.interpret.explanation.explanation_client import ExplanationClient
client = ExplanationClient.from_run(pipeline_run)
explanation = client.download_model_explanation()
explanation = client.download_model_explanation(top_k=4)
global_importance_values = explanation.get_ranked_global_values()
global_importance_names = explanation.get_ranked_global_names()
print('global importance values: {}'.format(global_importance_values))
print('global importance names: {}'.format(global_importance_names))
```

B. Use the Inference Clusters tab in Machine Learning Studio.

C. Use the Activity log in the Azure portal for the Machine Learning workspace.

D. Run the following code in a notebook:

```
from azureml.widgets import RunDetails
RunDetails(pipeline_run).show()
```

E. Run the following code and monitor the console output from the PipelineRun object:

```
pipeline_run.wait_for_completion(show_output=True)
```

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

Answer: DE

Explanation:
A batch inference job can take a long time to finish. This example monitors progress by using a Jupyter widget. You can also manage the job's progress by using:

≫ Azure Machine Learning Studio.

≫ Console output from the PipelineRun object. from azureml.widgets import RunDetails RunDetails(pipeline_run).show()
pipeline_run.wait_for_completion(show_output=True) Reference:
https://docs.microsoft.com/en-us/azure/machine-learning/how-to-use-parallel-run-step#monitor-the-parallel-run

NEW QUESTION 27
- (Exam Topic 3)
You use Azure Machine Learning to train and register a model.
You must deploy the model into production as a real-time web service to an inference cluster named service-compute that the IT department has created in the Azure Machine Learning workspace.
Client applications consuming the deployed web service must be authenticated based on their Azure Active Directory service principal.
You need to write a script that uses the Azure Machine Learning SDK to deploy the model. The necessary modules have been imported.
How should you complete the code? To answer, select the appropriate options in the answer area.
NOTE: Each correct selection is worth one point.

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

deployment_config =                ▼ .deploy_configuration(cpu_cores=1, memory_gb=1,
                    ┌──────────────────┐
                    │ AksWebservice    │
                    │ AciWebservice    │
                    │ LocalWebService  │
                    └──────────────────┘

          ▼ )
┌──────────────────────────┐
│ token_auth_enabled=True   │
│ token_auth_enabled=False  │
│ auth_enabled=True         │
│ auth_enabled=False        │
└──────────────────────────┘

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

A. Mastered
B. Not Mastered

Answer: A

Explanation:
Box 1: AksCompute Example:
aks_target = AksCompute(ws,"myaks")

# If deploying to a cluster configured for dev/test, ensure that it was created with enough
# cores and memory to handle this deployment configuration. Note that memory is also used by
# things such as dependencies and AML components.
deployment_config = AksWebservice.deploy_configuration(cpu_cores = 1, memory_gb = 1)
service = Model.deploy(ws, "myservice", [model], inference_config, deployment_config, aks_target) Box 2: AksWebservice
Box 3: token_auth_enabled=Yes
Whether or not token auth is enabled for the Webservice.
Note: A Service principal defined in Azure Active Directory (Azure AD) can act as a principal on which authentication and authorization policies can be enforced in Azure Databricks.
The Azure Active Directory Authentication Library (ADAL) can be used to programmatically get an Azure AD access token for a user.
Reference:
https://docs.microsoft.com/en-us/azure/machine-learning/how-to-deploy-azure-kubernetes-service https://docs.microsoft.com/en-us/azure/databricks/dev-tools/api/latest/aad/service-prin-aad-token

NEW QUESTION 32
- (Exam Topic 3)
You are building a machine learning model for translating English language textual content into French language textual content.
You need to build and train the machine learning model to learn the sequence of the textual content. Which type of neural network should you use?

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

Answer: C

Explanation:
To translate a corpus of English text to French, we need to build a recurrent neural network (RNN).
Note: RNNs are designed to take sequences of text as inputs or return sequences of text as outputs, or both. They're called recurrent because the network's hidden layers have a loop in which the output and cell state from each time step become inputs at the next time step. This recurrence serves as a form of memory. It allows contextual information to flow through the network so that relevant outputs from previous time steps can be applied to network operations at the current time step.
References:
https://towardsdatascience.com/language-translation-with-rnns-d84d43b40571

NEW QUESTION 37
- (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 39
- (Exam Topic 3)
You create an Azure Machine Learning workspace named ML-workspace. You also create an Azure Databricks workspace named DB-workspace. DB-workspace contains a cluster named DB-cluster.
You must use DB-cluster to run experiments from notebooks that you import into DB-workspace.
You need to use ML-workspace to track MLflow metrics and artifacts generated by experiments running on DB-cluster. The solution must minimize the need for custom code.
What should you do?

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

Answer: B

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

NEW QUESTION 43
- (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 47
- (Exam Topic 3)
You plan to create a speech recognition deep learning model. The model must support the latest version of Python.
You need to recommend a deep learning framework for speech recognition to include in the Data Science Virtual Machine (DSVM).
What should you recommend?

A. Apache Drill
B. Tensorflow

C. Rattle
D. Weka

Answer: B

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

NEW QUESTION 52
- (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 57
- (Exam Topic 3)
You create a machine learning model by using the Azure Machine Learning designer. You publish the model as a real-time service on an Azure Kubernetes Service (AKS) inference compute cluster. You make no changes to the deployed endpoint configuration.
You need to provide application developers with the information they need to consume the endpoint.
Which two values should you provide to application developers? Each correct answer presents part of the solution.
NOTE: Each correct selection is worth one point.

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

Answer: CE

Explanation:
Deploying an Azure Machine Learning model as a web service creates a REST API endpoint. You can send data to this endpoint and receive the prediction returned by the model.
You create a web service when you deploy a model to your local environment, Azure Container Instances, Azure Kubernetes Service, or field-programmable gate arrays (FPGA). You retrieve the URI used to access the web service by using the Azure Machine Learning SDK. If authentication is enabled, you can also use the SDK to get the authentication keys or tokens.
Example:
# URL for the web service
scoring_uri = '<your web service URI>'
# If the service is authenticated, set the key or token key = '<your key or token>'
Reference:
https://docs.microsoft.com/en-us/azure/machine-learning/how-to-consume-web-service

NEW QUESTION 59
- (Exam Topic 3)
You are determining if two sets of data are significantly different from one another by using Azure Machine Learning Studio.
Estimated values in one set of data may be more than or less than reference values in the other set of data. You must produce a distribution that has a constant Type I error as a function of the correlation.
You need to produce the distribution.
Which type of distribution should you produce?

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

Answer: A

Explanation:
Choose a one-tail or two-tail test. The default is a two-tailed test. This is the most common type of test, in which the expected distribution is symmetric around zero.
Example: Type I error of unpaired and paired two-sample t-tests as a function of the correlation. The simulated random numbers originate from a bivariate normal distribution with a variance of 1.
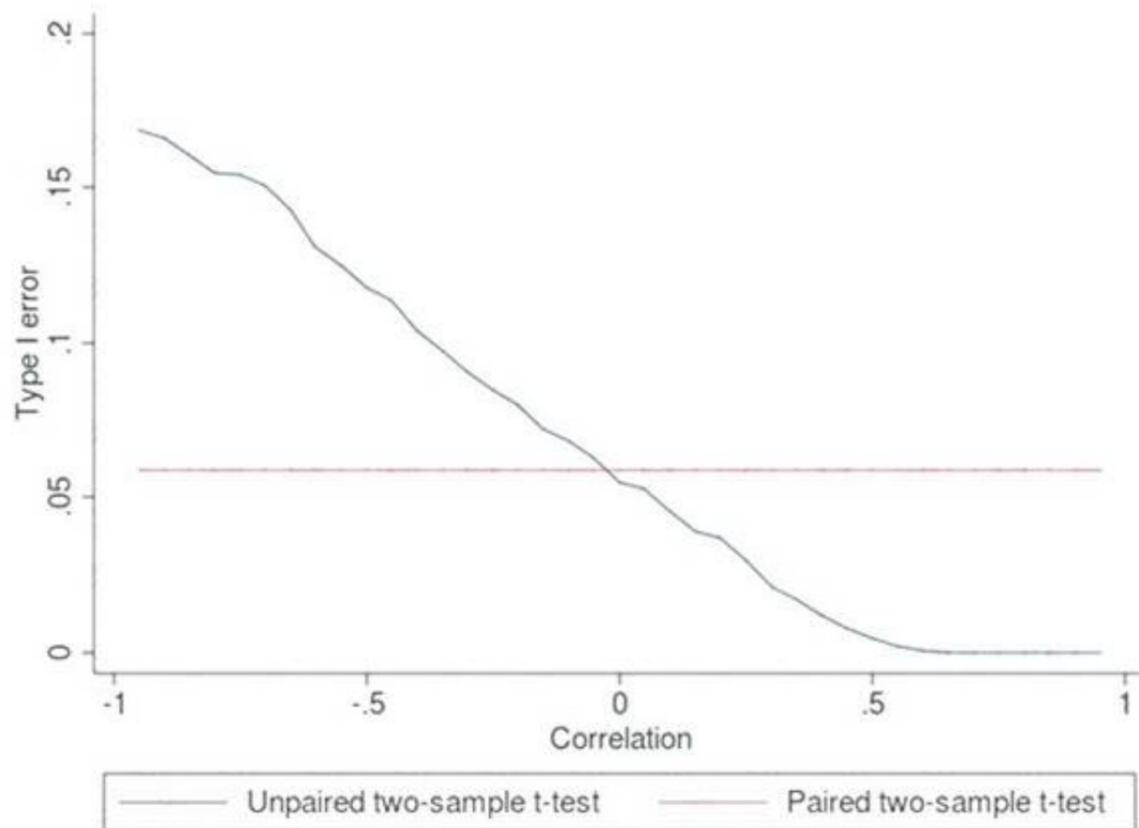
Reference:
https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/test-hypothesis-using-t-test https://en.wikipedia.org/wiki/Student%27s_t-test

NEW QUESTION 62
- (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 66
- (Exam Topic 3)
You plan to explore demographic data for home ownership in various cities. The data is in a CSV file with the following format:
age,city,income,home_owner 21,Chicago,50000,0 35,Seattle,120000,1 23,Seattle,65000,0 45,Seattle,130000,1 18,Chicago,48000,0

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

≫ the number of observations in the dataset

≫ a box plot of income by home_owner

≫ a dictionary containing the city names and the average income for each city

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

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

NOTE: Each correct selection is worth one point.

Code segments

```
log
log_list
log_row
log_table
log_image
```

Answer Area

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

A. Mastered
B. Not Mastered

Answer: A

Explanation:
Box 1: log
The number of observations in the dataset. run.log(name, value, description='')
Scalar values: Log a numerical or string value to the run with the given name. Logging a metric to a run causes that metric to be stored in the run record in the experiment. You can log the same metric multiple times within a run, the result being considered a vector of that metric.
Example: run.log("accuracy", 0.95)
Box 2: log_image
A box plot of income by home_owner.
log_image Log an image to the run record. Use log_image to log a .PNG image file or a matplotlib plot to the run. These images will be visible and comparable in the run record.
Example: run.log_image("ROC", plot=plt) Box 3: log_table
A dictionary containing the city names and the average income for each city. log_table: Log a dictionary object to the run with the given name.

NEW QUESTION 68
- (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 71
- (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 73
- (Exam Topic 3)
You have a Python data frame named salesData in the following format:

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

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

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

You need to use the pandas.melt() function in Python to perform the transformation.
How should you complete the code segment? To answer, select the appropriate options in the answer area. NOTE: Each correct selection is worth one point.

Answer Area

```
import pandas as pd
salesData = pd.melt(          ▼  ,id_vars='       ▼ ',value_vars=     ▼  )
               dataFrame                shop                'shop'
               pandas                   year                'year'
               salesData                value                ['year']
               year                     Shop X, Shop Y, Shop  ['2017', '2018']
                                        Z
```

A. Mastered
B. Not Mastered

Answer: A

Explanation:
Box 1: dataFrame
Syntax: pandas.melt(frame, id_vars=None, value_vars=None, var_name=None, value_name='value', col_level=None)[source]
Where frame is a DataFrame Box 2: shop
Paramter id_vars id_vars : tuple, list, or ndarray, optional Column(s) to use as identifier variables.
Box 3: ['2017','2018']
value_vars : tuple, list, or ndarray, optional
Column(s) to unpivot. If not specified, uses all columns that are not set as id_vars. Example:
df = pd.DataFrame({'A': {0: 'a', 1: 'b', 2: 'c'},
'B': {0: 1, 1: 3, 2: 5},
'C': {0: 2, 1: 4, 2: 6}})
pd.melt(df, id_vars=['A'], value_vars=['B', 'C']) A variable value
0 a B 1
1 b B 3
2 c B 5
3 a C 2
4 b C 4
5 c C 6
References:
https://pandas.pydata.org/pandas-docs/stable/reference/api/pandas.melt.html

NEW QUESTION 78
- (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 83

- (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 88
- (Exam Topic 3)
You create an experiment in Azure Machine Learning Studio. You add a training dataset that contains 10,000 rows. The first 9,000 rows represent class 0 (90 percent).
The remaining 1,000 rows represent class 1 (10 percent).
The training set is imbalances between two classes. You must increase the number of training examples for class 1 to 4,000 by using 5 data rows. You add the Synthetic Minority Oversampling Technique (SMOTE) module to the experiment.
You need to configure the module.
Which values should you use? To answer, select the appropriate options in the dialog box in the answer area. NOTE: Each correct selection is worth one point.


A. Mastered
B. Not Mastered

Answer: A

Explanation:
Box 1: 300
You type 300 (%), the module triples the percentage of minority cases (3000) compared to the original dataset (1000).
Box 2: 5
We should use 5 data rows.
Use the Number of nearest neighbors option to determine the size of the feature space that the SMOTE algorithm uses when in building new cases. A nearest neighbor is a row of data (a case) that is very similar to some target case. The distance between any two cases is measured by combining the weighted vectors of all features.
By increasing the number of nearest neighbors, you get features from more cases.
By keeping the number of nearest neighbors low, you use features that are more like those in the original sample.
References:
https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/smote

NEW QUESTION 89
- (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 90
- (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.
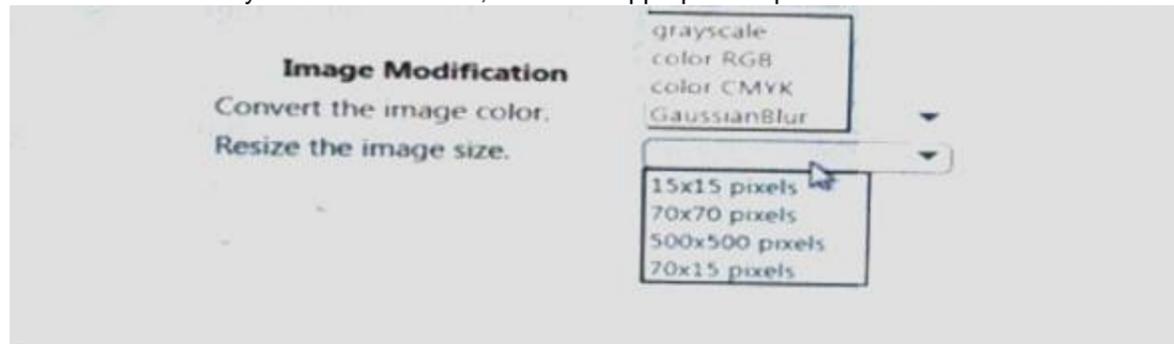
A. Mastered
B. Not Mastered

Answer: A

Explanation:
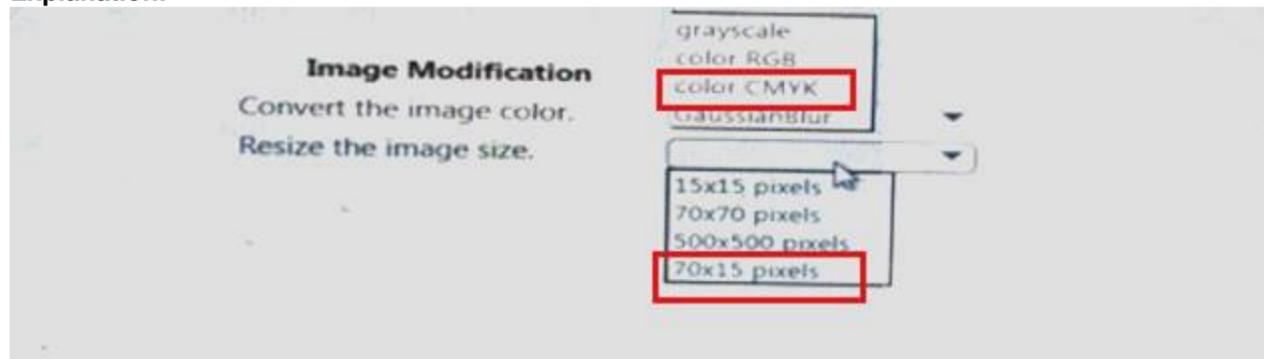
NEW QUESTION 95
- (Exam Topic 3)

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 and then run the training script as an experiment on local compute.

A. Yes
B. No

Answer: B

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

NEW QUESTION 96
- (Exam Topic 3)
You define a datastore named ml-data for an Azure Storage blob container. In the container, you have a folder named train that contains a file named data.csv.
You plan to use the file to train a model by using the Azure Machine Learning SDK.
You plan to train the model by using the Azure Machine Learning SDK to run an experiment on local compute.
You define a DataReference object by running the following code:

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

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

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

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

B. import os
   import argparse
   import pandas as pd

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

C. import pandas as pd

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

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

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

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

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

Answer: E

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

NEW QUESTION 98
- (Exam Topic 3)
You must store data in Azure Blob Storage to support Azure Machine Learning. You need to transfer the data into Azure Blob Storage.
What are three possible ways to achieve the goal? Each correct answer presents a complete solution.
NOTE: Each correct selection is worth one point.

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

Answer: BCD

Explanation:
You can move data to and from Azure Blob storage using different technologies: Azure Storage-Explorer
AzCopy Python SSIS
References:
https://docs.microsoft.com/en-us/azure/machine-learning/team-data-science-process/move-azure-blob

NEW QUESTION 99
- (Exam Topic 3)
Note: This question is part of a series of questions that present the same scenario. Each question in the series contains a unique solution that might meet the stated goals. Some question sets might have more than one correct solution, while others might not have a correct solution.
After you answer a question in this section, you will NOT be able to return to it. As a result, these questions will not appear in the review screen.
You are creating a new experiment in Azure Machine Learning Studio.
One class has a much smaller number of observations than the other classes in the training set. You need to select an appropriate data sampling strategy to compensate for the class imbalance. Solution: You use the Scale and Reduce sampling mode.
Does the solution meet the goal?

A. Yes
B. No

Answer: B

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

NEW QUESTION 102
- (Exam Topic 3)
You are analyzing the asymmetry in a statistical distribution.

The following image contains two density curves that show the probability distribution of two datasets.

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

NEW QUESTION 107
- (Exam Topic 3)
You are evaluating a Python NumPy array that contains six data points defined as follows: data = [10, 20, 30, 40, 50, 60]
You must generate the following output by using the k-fold algorithm implantation in the Python Scikit-learn machine learning library:
train: [10 40 50 60], test: [20 30]
train: [20 30 40 60], test: [10 50]
train: [10 20 30 50], test: [40 60]
You need to implement a cross-validation to generate the output.
How should you complete the code segment? To answer, select the appropriate code segment in the dialog box in the answer area.
NOTE: Each correct selection is worth one point.

A. Mastered
B. Not Mastered

Answer: A

Explanation:
Box 1: k-fold
Box 2: 3
K-F olds cross-validator provides train/test indices to split data in train/test sets. Split dataset into k consecutive folds (without shuffling by default).
The parameter n_splits (int, default=3) is the number of folds. Must be at least 2. Box 3: data
Example: Example:
>>>
>>> from sklearn.model_selection import KFold
>>> X = np.array([[1, 2], [3, 4], [1, 2], [3, 4]])
>>> y = np.array([1, 2, 3, 4])
>>> kf = KFold(n_splits=2)
>>> kf.get_n_splits(X) 2
>>> print(kf)
KFold(n_splits=2, random_state=None, shuffle=False)
>>> for train_index, test_index in kf.split(X): print("TRAIN:", train_index, "TEST:", test_index) X_train, X_test = X[train_index], X[test_index] y_train, y_test = y[train_index], y[test_index] TRAIN: [2 3] TEST: [0 1]
TRAIN: [0 1] TEST: [2 3]
References:
https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.KFold.html

NEW QUESTION 110
- (Exam Topic 3)
You are performing sentiment analysis using a CSV file that includes 12,000 customer reviews written in a short sentence format. You add the CSV file to Azure Machine Learning Studio and configure it as the starting point dataset of an experiment. You add the Extract N-Gram Features from Text module to the experiment to extract key phrases from the customer review column in the dataset.
You must create a new n-gram dictionary from the customer review text and set the maximum n-gram size to trigrams.
What should you select? To answer, select the appropriate options in the answer area.
NOTE: Each correct selection is worth one point.

Properties   Project

Extract N-Gram Features from Text

Text column

Selected columns
Column type: String Feature

Launch column selector

Vocabulary mode

Create
ReadOnly
Update
Merge

N-Grams size

3
4
4,000
12,000

0

Weighting function

Minimum word length

3

Maximum word length

25

Minimum n-gram document absolu...

5

Maximum n-gram document ratio

1

A. Mastered
B. Not Mastered

Answer: A

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

NEW QUESTION 113
- (Exam Topic 3)
You are building recurrent neural network to perform a binary classification.
The training loss, validation loss, training accuracy, and validation accuracy of each training epoch has been provided. You need to identify whether the classification model is over fitted.
Which of the following is correct?

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

Answer: B

Explanation:
An overfit model is one where performance on the train set is good and continues to improve, whereas performance on the validation set improves to a point and then begins to degrade.
References:
https://machinelearningmastery.com/diagnose-overfitting-underfitting-lstm-models/

NEW QUESTION 115
- (Exam Topic 3)
Your team is building a data engineering and data science development environment. The environment must support the following requirements:

≫ support Python and Scala

≫ compose data storage, movement, and processing services into automated data pipelines

≫ the same tool should be used for the orchestration of both data engineering and data science

≫ support workload isolation and interactive workloads

≫ enable scaling across a cluster of machines You need to create the environment.
What should you do?

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

Answer: B

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

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

≫ High-concurrency
Azure Databricks is fully integrated with Azure Data Factory.

NEW QUESTION 118
- (Exam Topic 3)
You plan to use the Hyperdrive feature of Azure Machine Learning to determine the optimal hyperparameter values when training a model.
You must use Hyperdrive to try combinations of the following hyperparameter values:
• learning_rate: any value between 0.001 and 0.1
• batch_size: 16, 32, or 64
You need to configure the search space for the Hyperdrive experiment.
Which two parameter expressions should you use? Each correct answer presents part of the solution. NOTE: Each correct selection is worth one point.

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

Answer: BD

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

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

D: Discrete hyperparameters are specified as a choice among discrete values. choice can be:

≫ one or more comma-separated values

≫ a range object

≫ any arbitrary list object Reference:
https://docs.microsoft.com/en-us/azure/machine-learning/how-to-tune-hyperparameters

NEW QUESTION 123
- (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 125
- (Exam Topic 3)
You are building an experiment using the Azure Machine Learning designer.
You split a dataset into training and testing sets. You select the Two-Class Boosted Decision Tree as the algorithm.
You need to determine the Area Under the Curve (AUC) of the model.
Which three modules should you use in sequence? To answer, move the appropriate modules from the list of modules to the answer area and arrange them in the correct order.

A. Mastered
B. Not Mastered

Answer: A

Explanation:
Step 1: Train Model
Two-Class Boosted Decision Tree
First, set up the boosted decision tree model.
* 1. Find the Two-Class Boosted Decision Tree module in the module palette and drag it onto the canvas.
* 2. Find the Train Model module, drag it onto the canvas, and then connect the output of the Two-Class Boosted Decision Tree module to the left input port of the Train Model module.
The Two-Class Boosted Decision Tree module initializes the generic model, and Train Model uses training data to train the model.
* 3. Connect the left output of the left Execute R Script module to the right input port of the Train Model
module (in this tutorial you used the data coming from the left side of the Split Data module for training). This portion of the experiment now looks something like this:

Step 2: Score Model
Score and evaluate the models
You use the testing data that was separated out by the Split Data module to score our trained models. You can then compare the results of the two models to see which generated better results.
Add the Score Model modules
* 1. Find the Score Model module and drag it onto the canvas.
* 2. Connect the Train Model module that's connected to the Two-Class Boosted Decision Tree module to the left input port of the Score Model module.
* 3. Connect the right Execute R Script module (our testing data) to the right input port of the Score Model module.
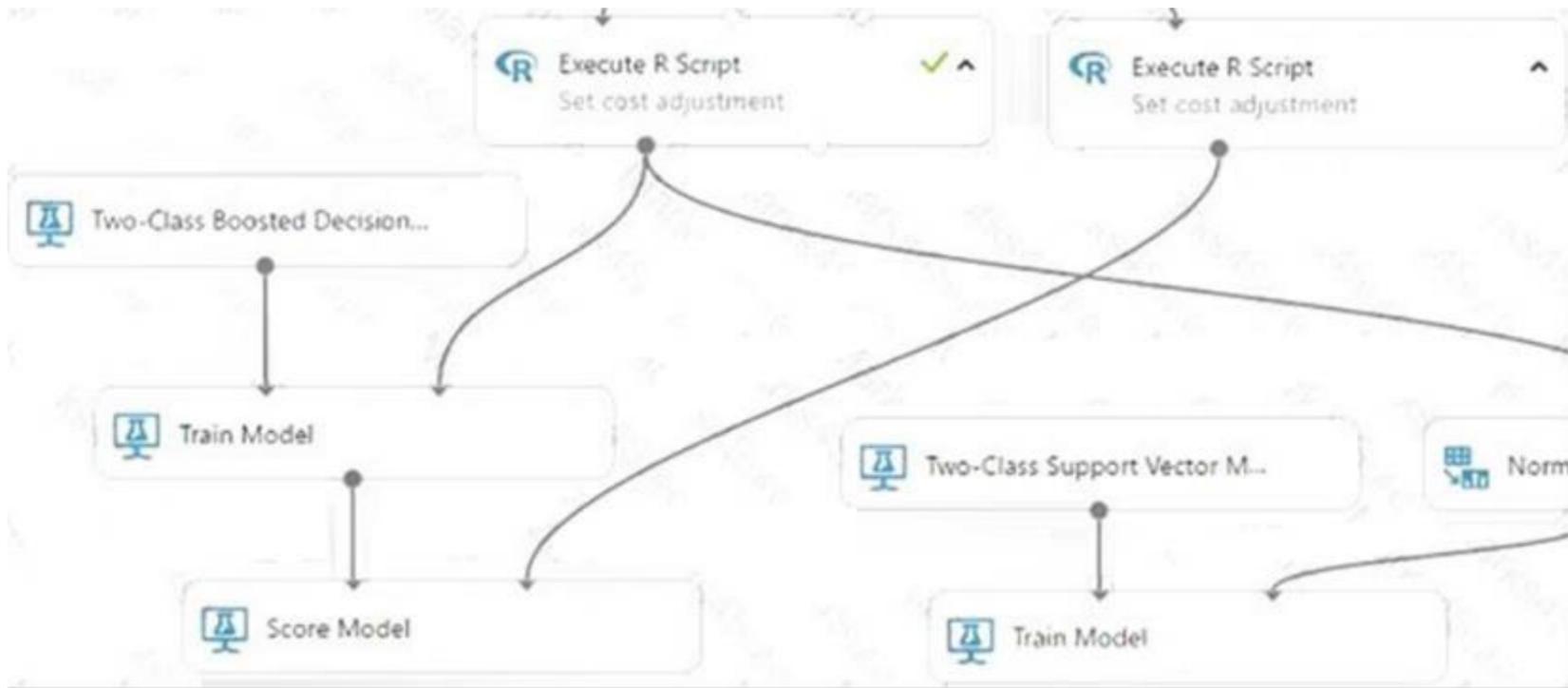
Step 3: Evaluate Model
To evaluate the two scoring results and compare them, you use an Evaluate Model module.
* 1. Find the Evaluate Model module and drag it onto the canvas.
* 2. Connect the output port of the Score Model module associated with the boosted decision tree model to the left input port of the Evaluate Model module.
* 3. Connect the other Score Model module to the right input port.
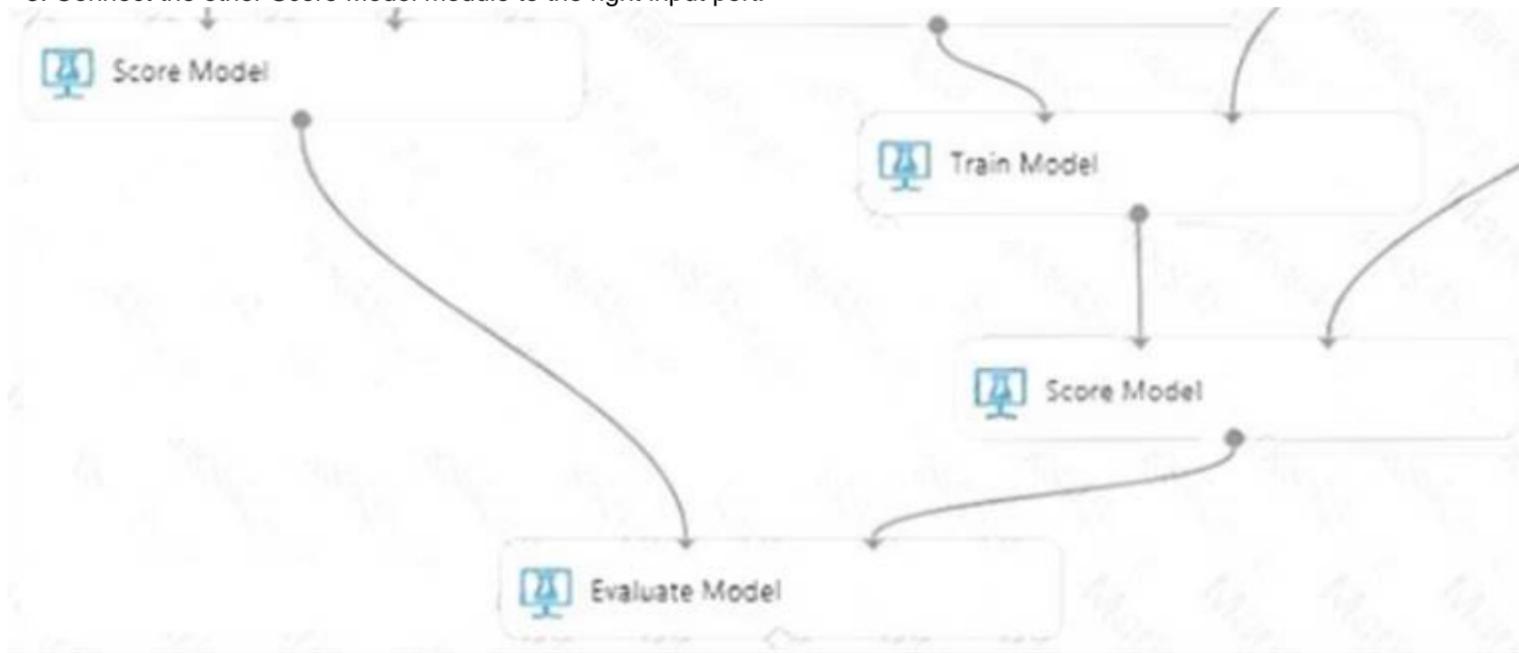
NEW QUESTION 130
- (Exam Topic 3)
You publish a batch inferencing pipeline that will be used by a business application.
The application developers need to know which information should be submitted to and returned by the REST interface for the published pipeline.
You need to identify the information required in the REST request and returned as a response from the published pipeline.
Which values should you use in the REST request and to expect in the response? 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: JSON containing an OAuth bearer token Specify your authentication header in the request.
To run the pipeline from the REST endpoint, you need an OAuth2 Bearer-type authentication header. Box 2: JSON containing the experiment name
Add a JSON payload object that has the experiment name. Example:
rest_endpoint = published_pipeline.endpoint response = requests.post(rest_endpoint, headers=auth_header, json={"ExperimentName": "batch_scoring",
"ParameterAssignments": {"process_count_per_node": 6}}) run_id = response.json()["Id"]
Box 3: JSON containing the run ID
Make the request to trigger the run. Include code to access the Id key from the response dictionary to get the value of the run ID.
Reference:
https://docs.microsoft.com/en-us/azure/machine-learning/tutorial-pipeline-batch-scoring-classification

NEW QUESTION 131
- (Exam Topic 3)

```
source_directory=scripts_folder,
entry_script="batch_pipeline.py",
mini_batch_size="5",
error_threshold=10,
output_action="append_row",
environment=batch_env,
compute_target=compute_target,
logging_level="DEBUG",
node_count=4)
```

You need to obtain the output from the pipeline execution. Where will you find the output?

A. the Activity Log in the Azure portal for the Machine Learning workspace
B. a file named parallel_run_step.txt located in the output folder
C. the digitjdentification.py script
D. the Inference Clusters tab in Machine Learning studio
E. the debug log

Answer: B

Explanation:
output_action (str): How the output is to be organized. Currently supported values are 'append_row' and 'summary_only'.
➢ 'append_row' – All values output by run() method invocations will be aggregated into one unique file named parallel_run_step.txt that is created in the output location.
➢ 'summary_only' Reference:
https://docs.microsoft.com/en-us/python/api/azureml-contrib-pipeline-steps/azureml.contrib.pipeline.steps.parall

NEW QUESTION 136
- (Exam Topic 3)
You plan to use a Data Science Virtual Machine (DSVM) with the open source deep learning frameworks Caffe2 and Theano. You need to select a pre configured DSVM to support the framework.
What should you create?

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

Answer: E

NEW QUESTION 139
- (Exam Topic 3)
Note: This question is part of a series of questions that present the same scenario. Each question in the series contains a unique solution that might meet the stated goals. Some question sets might have more
than one correct solution, while others might not have a correct solution.
After you answer a question in this section, you will NOT be able to return to it. As a result, these questions will not appear in the review screen.
You train a classification model by using a logistic regression algorithm.
You must be able to explain the model's predictions by calculating the importance of each feature, both as an overall global relative importance value and as a measure of local importance for a specific set of predictions.
You need to create an explainer that you can use to retrieve the required global and local feature importance values.
Solution: Create a TabularExplainer. Does the solution meet the goal?

A. Yes
B. No

Answer: B

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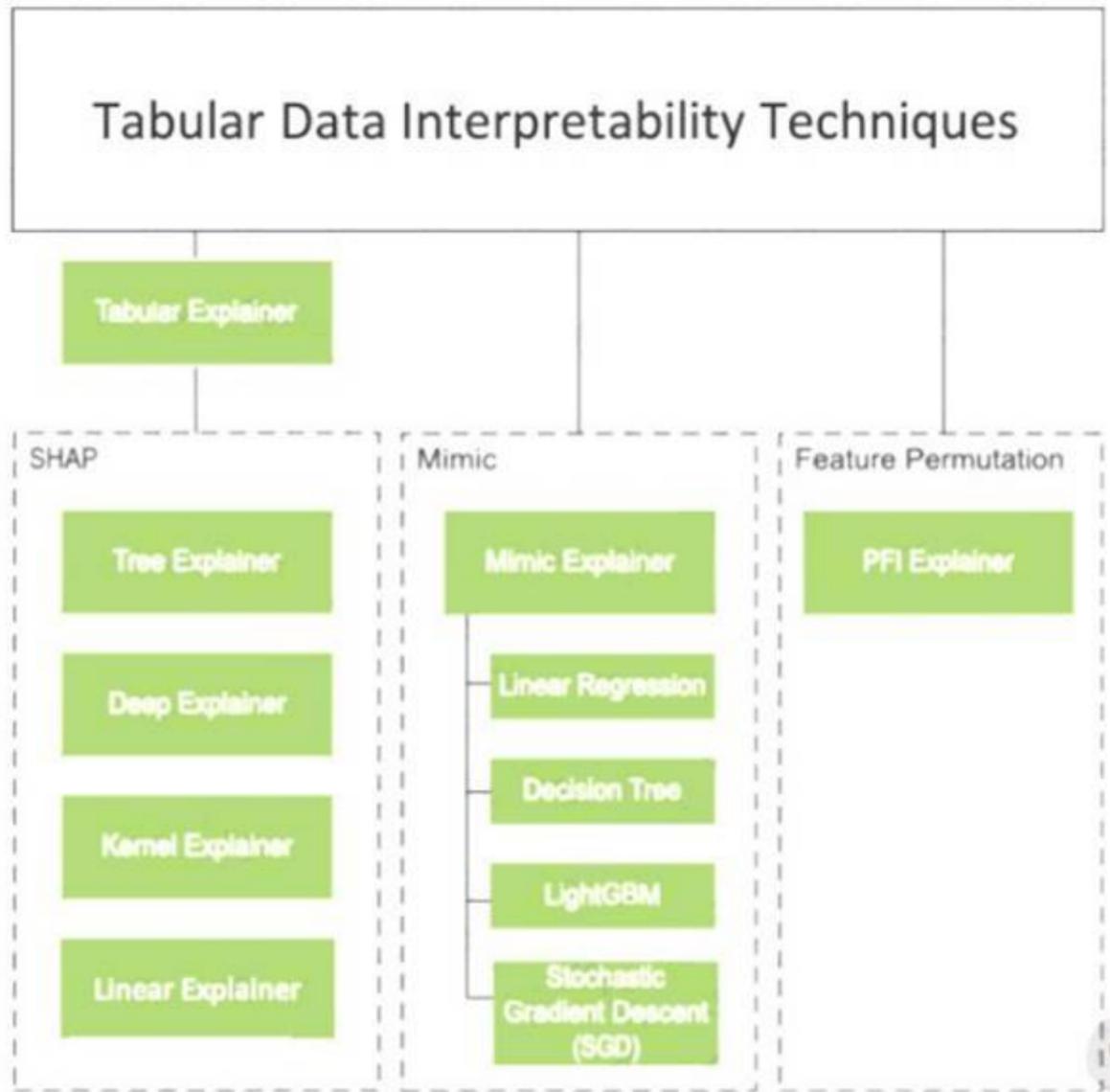
Note 2: Permutation Feature Importance Explainer (PFI): Permutation Feature Importance is a technique used to explain classification and regression models. At a high level, the way it works is by randomly shuffling data one feature at a time for the entire dataset and calculating how much the performance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

NEW QUESTION 140
- (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.
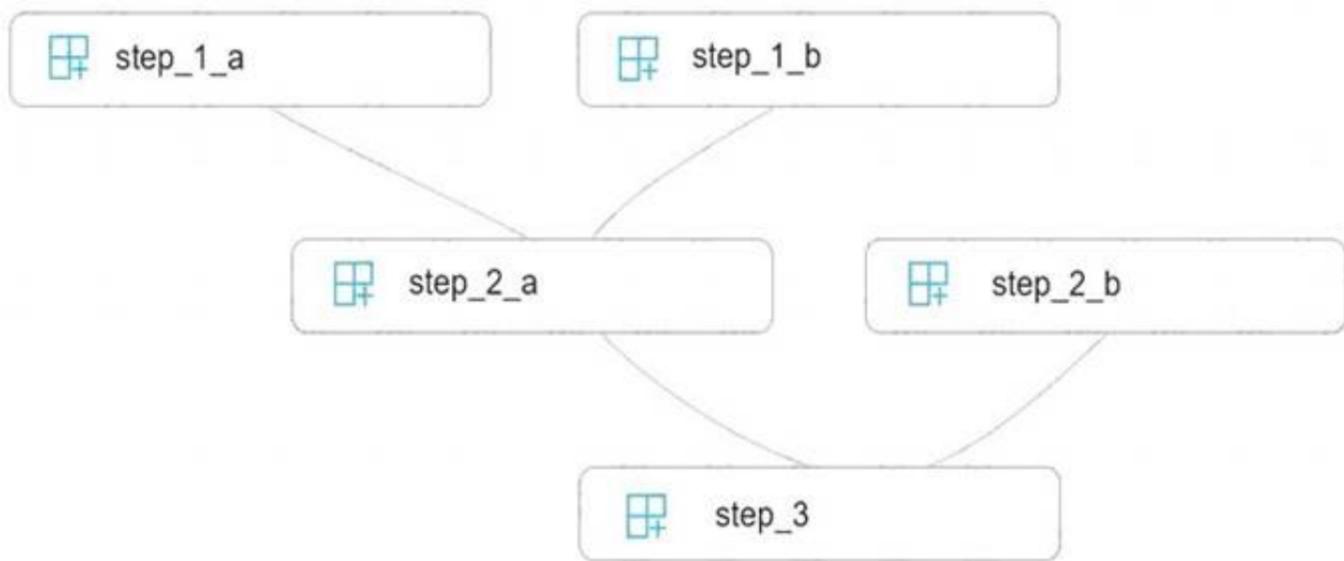
You need to create the pipeline shown. Assume all relevant imports have been done. Which Python code segment should you use?

```
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 142
- (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 147
- (Exam Topic 3)
Note: This question is part of a series of questions that present the same scenario. Each question in the series contains a unique solution that might meet the stated goals. Some question sets might have more than one correct solution, while others might not have a correct solution.
After you answer a question in this section, you will NOT be able to return to it. As a result, these questions will not appear in the review screen.
You train and register a machine learning model.
You plan to deploy the model as a real-time web service. Applications must use key-based authentication to use the model.
You need to deploy the web service. Solution:
Create an AciWebservice instance.
Set the value of the ssl_enabled property to True.
Deploy the model to the service. Does the solution meet the goal?

A. Yes
B. No

Answer: B

Explanation:
Instead use only auth_enabled = TRUE Note: Key-based authentication.
Web services deployed on AKS have key-based auth enabled by default. ACI-deployed services have
key-based auth disabled by default, but you can enable it by setting auth_enabled = TRUE when creating the ACI web service. The following is an example of creating an ACI deployment configuration with key-based auth enabled.
deployment_config <- aci_webservice_deployment_config(cpu_cores = 1,
memory_gb = 1, auth_enabled = TRUE) Reference:
https://azure.github.io/azureml-sdk-for-r/articles/deploying-models.html

NEW QUESTION 151
- (Exam Topic 3)
You are performing feature engineering on a dataset.
You must add a feature named CityName and populate the column value with the text London.
You need to add the new feature to the dataset.
Which Azure Machine Learning Studio module should you use?

A. Edit Metadata
B. Preprocess Text
C. Execute Python Script
D. Latent Dirichlet Allocation

Answer: A

Explanation:
Typical metadata changes might include marking columns as features. References:
https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/edit-metadata

NEW QUESTION 155
- (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 157
- (Exam Topic 3)
You use Azure Machine Learning Studio to build a machine learning experiment. You need to divide data into two distinct datasets.
Which module should you use?

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

Answer: D

Explanation:
The Group Data into Bins module supports multiple options for binning data. You can customize how the bin edges are set and how values are apportioned into the bins.
References:
https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/group-data-into-bins

NEW QUESTION 160
- (Exam Topic 3)
You create an Azure Machine Learning workspace and set up a development environment. You plan to train a deep neural network (DNN) by using the Tensorflow framework and by using estimators to submit training scripts.
You must optimize computation speed for training runs.
You need to choose the appropriate estimator to use as well as the appropriate training compute target configuration.
Which values should you use? To answer, select the appropriate options in the answer area. NOTE: Each correct selection is worth one point.

Answer Area

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

A. Mastered
B. Not Mastered

Answer: A

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

NEW QUESTION 165
- (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 166
- (Exam Topic 3)
You use Data Science Virtual Machines (DSVMs) for Windows and Linux in Azure. You need to access the DSVMs.
Which utilities should you use? To answer, select the appropriate options in the answer area. NOTE: Each correct selection is worth one point.
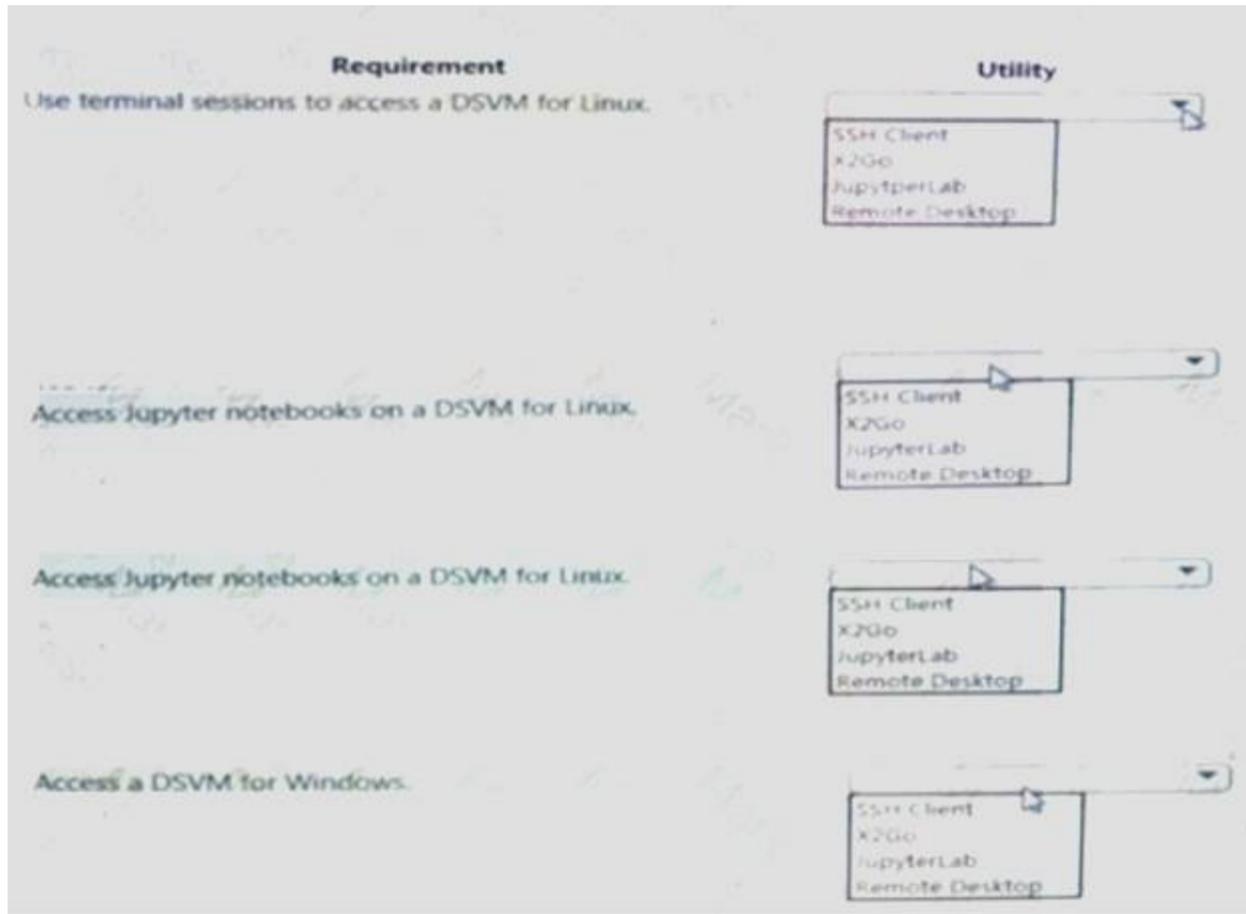
A. Mastered
B. Not Mastered

Answer: A

Explanation:
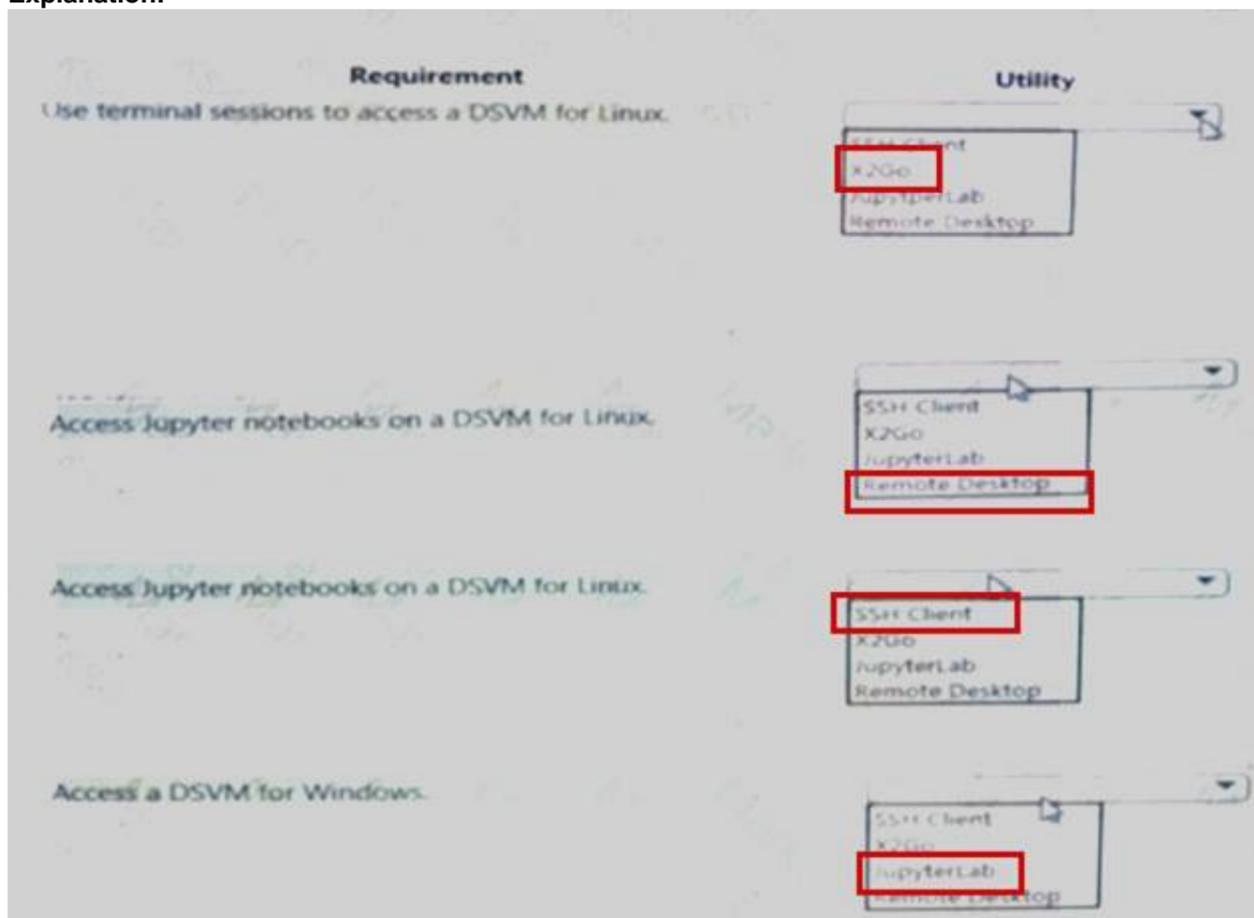
NEW QUESTION 171
- (Exam Topic 3)
You have a comma-separated values (CSV) file containing data from which you want to train a classification model.
You are using the Automated Machine Learning interface in Azure Machine Learning studio to train the classification model. You set the task type to Classification.
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 176
- (Exam Topic 3)
You are building a regression model tot estimating the number of calls during an event.
You need to determine whether the feature values achieve the conditions to build a Poisson regression model. Which two conditions must the feature set contain?
I ach correct answer presents part of the solution. NOTE:
Each correct selection is worth one point.

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

Answer: CE

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

≫ The response variable has a Poisson distribution.

≫ Counts cannot be negative. The method will fail outright if you attempt to use it with negative labels.

≫ A Poisson distribution is a discrete distribution; therefore, it is not meaningful to use this method with non-whole numbers.
References:
https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/poisson-regression

NEW QUESTION 181
- (Exam Topic 3)
Note: This question is part of a series of questions that present the same scenario. Each question in the series contains a unique solution that might meet the stated goals. Some question sets might have more than one correct solution, while others might not have a correct solution.
After you answer a question in this section, you will NOT be able to return to it. As a result, these questions will not appear in the review screen.
You train a classification model by using a logistic regression algorithm.
You must be able to explain the model's predictions by calculating the importance of each feature, both as an overall global relative importance value and as a measure of local importance for a specific set of predictions.
You need to create an explainer that you can use to retrieve the required global and local feature importance values.
Solution: Create a PFIExplainer. Does the solution meet the goal?

A. Yes
B. No

Answer: A

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

NEW QUESTION 182
- (Exam Topic 3)
Note: This question is part of a series of questions that present the same scenario. Each question in the series contains a unique solution that might meet the stated goals. Some question sets might have more than one correct solution, while others might not have a correct solution.
After you answer a question in this section, you will NOT be able to return to it. As a result, these questions will not appear in the review screen.
You create an Azure Machine Learning service datastore in a workspace. The datastore contains the following files:
• /data/2018/Q1 .csv
• /data/2018/Q2.csv
• /data/2018/Q3.csv
• /data/2018/Q4.csv
• /data/2019/Q1.csv
All files store data in the following format:
• id,f1,f2,l
• 1,1,2,0
• 2,1,1,1
• 3.2.1.0
You run the following code:

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

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

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

Solution: Run the following code:

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

Does the solution meet the goal?

A. Yes
B. No

Answer: A

Explanation:
Use two file paths.
Use Dataset.Tabular_from_delimeted as the data isn't cleansed. Note:
A TabularDataset represents data in a tabular format by parsing the provided file or list of files. This provides you with the ability to materialize the data into a pandas or Spark DataFrame so you can work with familiar data preparation and training libraries without having to leave your notebook. You can create a TabularDataset object from .csv, .tsv, .parquet, .jsonl files, and from SQL query results.
Reference:
https://docs.microsoft.com/en-us/azure/machine-learning/how-to-create-register-datasets

NEW QUESTION 183
- (Exam Topic 3)
You create an Azure Machine Learning compute target named ComputeOne by using the STANDARD_D1 virtual machine image.
You define a Python variable named was that references the Azure Machine Learning workspace. You run the following Python code:

```
from azureml.core.compute import ComputeTarget, AmlCompute
from azureml.core.compute_target import ComputeTargetException
the_cluster_name = "ComputeOne"
try:
    the_cluster = ComputeTarget(workspace=ws, name=the_cluster_name)
    print('Step1')
except ComputeTargetException:
    config = AmlCompute.provisioning_configuration(vm_size='STANDARD_DS12_v2', max_nodes=4)
    the_cluster = ComputeTarget.create(ws, the_cluster_name, config)
    print('Step2')
```

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

	Yes	No
A new machine learning compute resource is created with a virtual machine size of STANDARD_DS12_v2 and a maximum of four nodes.	O	O
Any experiments configured to use `the_cluster` will run on ComputeOne.	O	O
The text **Step1** will be printed to the screen.	O	O

A. Mastered
B. Not Mastered

Answer: A

Explanation:
Box 1:Yes
ComputeTargetException class: An exception related to failures when creating, interacting with, or configuring a compute target. This exception is commonly raised for failures attaching a compute target,
missing headers, and unsupported configuration values. Create(workspace, name, provisioning_configuration)
Provision a Compute object by specifying a compute type and related configuration. This method creates a new compute target rather than attaching an existing one. Box 2: Yes
Box 3: No
The line before print('Step1') will fail. Reference:
https://docs.microsoft.com/en-us/python/api/azureml-core/azureml.core.compute.computetarget

NEW QUESTION 184
- (Exam Topic 3)
Note: This question is part of a series of questions that present the same scenario. Each question in the series contains a unique solution that might meet the stated goals. Some question sets might have more than one correct solution, while others might not have a correct solution.
After you answer a question in this section, you will NOT be able to return to it. As a result, these
questions will not appear in the review screen.
You are creating a model to predict the price of a student's artwork depending on the following variables: the student's length of education, degree type, and art form.
You start by creating a linear regression model. You need to evaluate the linear regression model.
Solution: Use the following metrics: Mean Absolute Error, Root Mean Absolute Error, Relative Absolute Error, Accuracy, Precision, Recall, F1 score, and AUC.
Does the solution meet the goal?

A. Yes
B. No

Answer: B

Explanation:
Accuracy, Precision, Recall, F1 score, and AUC are metrics for evaluating classification models. Note: Mean Absolute Error, Root Mean Absolute Error, Relative Absolute Error are OK for the linear
regression model.
References:
https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/evaluate-model

NEW QUESTION 185
- (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.

Category	Variables
Categorical variables	▼
	Gender, IsPlaySoccer
	Gender, PrevExamMarks, Height, Weight
	PrevExamMarks, Height, Weight
	IsPlaySoccer
Continuous variables	▼
	Gender, IsPlaySoccer
	Gender, PrevExamMarks, Height, Weight
	PrevExamMarks, Height, Weight
	IsPlaySoccer

A. Mastered
B. Not Mastered

Answer: A

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

NEW QUESTION 189
- (Exam Topic 3)
You create a deep learning model for image recognition on Azure Machine Learning service using GPU-based training.
You must deploy the model to a context that allows for real-time GPU-based inferencing. You need to configure compute resources for model inferencing.
Which compute type should you use?

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

Answer: B

Explanation:
You can use Azure Machine Learning to deploy a GPU-enabled model as a web service. Deploying a model on Azure Kubernetes Service (AKS) is one option.
The AKS cluster provides a GPU resource that is used by the model for inference.
Inference, or model scoring, is the phase where the deployed model is used to make predictions. Using GPUs instead of CPUs offers performance advantages on highly parallelizable computation.
Reference:
https://docs.microsoft.com/en-us/azure/machine-learning/how-to-deploy-inferencing-gpus

NEW QUESTION 191
- (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.

Environment	Compute target

Testing
▼
Local web service
Azure Kubernetes Services (AKS)
Azure Container Instances
Azure Machine Learning compute clusters

Production
▼
Local web service
Azure Kubernetes Services (AKS)
Azure Container Instances
Azure Machine Learning compute clusters

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 194
- (Exam Topic 3)
You create a multi-class image classification deep learning model. You train the model by using PyTorch version 1.2.
You need to ensure that the correct version of PyTorch can be identified for the inferencing environment when the model is deployed.
What should you do?

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

Answer: D

Explanation:
framework_version: The PyTorch version to be used for executing training code. Reference:
https://docs.microsoft.com/en-us/python/api/azureml-train-core/azureml.train.dnn.pytorch?view=azure-ml-py

NEW QUESTION 199
- (Exam Topic 3)
You create a binary classification model. The model is registered in an Azure Machine Learning workspace. You use the Azure Machine Learning Fairness SDK to assess the model fairness.
You develop a training script for the model on a local machine.
You need to load the model fairness metrics into Azure Machine Learning studio. What should you do?

A. Implement the download_dashboard_by_upload_id function
B. Implement the creace_group_metric_sec function
C. Implement the upload_dashboard_dictionary function
D. Upload the training script

Answer: C

Explanation:
import azureml.contrib.fairness package to perform the upload:
from azureml.contrib.fairness import upload_dashboard_dictionary, download_dashboard_by_upload_id Reference:
https://docs.microsoft.com/en-us/azure/machine-learning/how-to-machine-learning-fairness-aml

NEW QUESTION 204
- (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 207
- (Exam Topic 2)
You need to visually identify whether outliers exist in the Age column and quantify the outliers before the
outliers are removed.
Which three Azure Machine Learning Studio modules should you use in sequence? To answer, move the appropriate modules from the list of modules to the answer area and arrange them in the correct order.
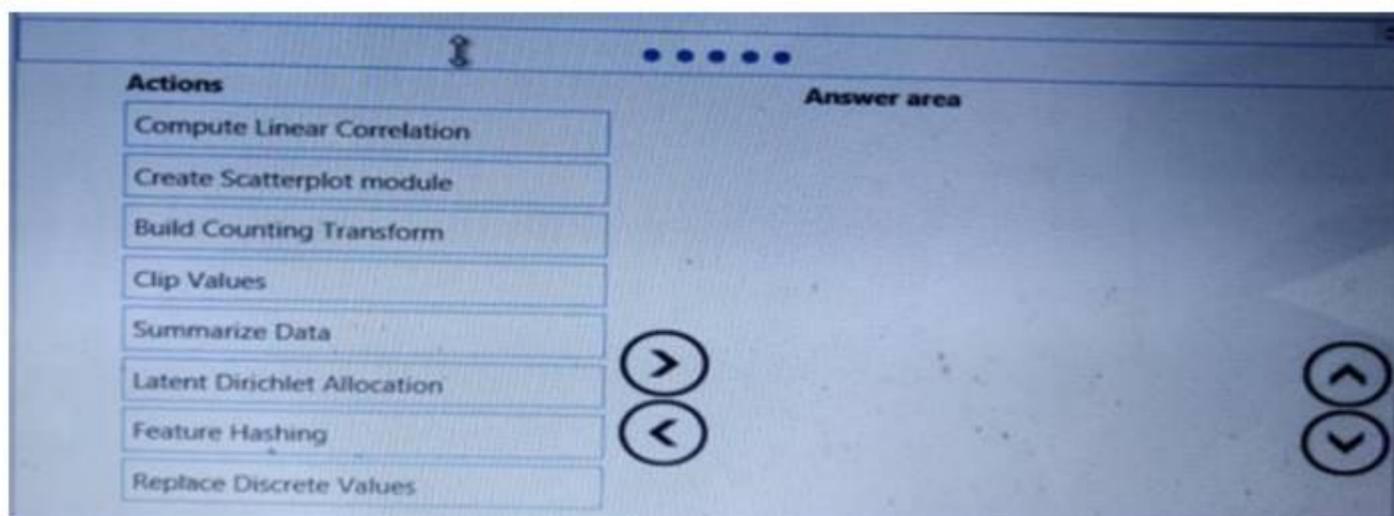
A. Mastered
B. Not Mastered

Answer: A

Explanation:

Create Scatterplot Summarize Data Clip Values
You can use the Clip Values module in Azure Machine Learning Studio, to identify and optionally replace data values that are above or below a specified threshold. This is useful when you want to remove outliers or replace them with a mean, a constant, or other substitute value.
References:
https://blogs.msdn.microsoft.com/azuredev/2017/05/27/data-cleansing-tools-in-azure-machine-learning/ https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/clip-values

NEW QUESTION 209
- (Exam Topic 2)
You need to correct the model fit issue.
Which three actions should you perform in sequence? To answer, move the appropriate actions from the list of actions to the answer area and arrange them in the correct order.

Actions   Answer Area

Add the Ordinal Regression module.

Add the Two-Class Averaged Perception module.

Augment the data.

Add the Bayesian Linear Regression module.

Decrease the memory size for L-BFGS.

Add the Multiclass Decision Jungle module.

Configure the regularization weight.

A. Mastered
B. Not Mastered

Answer: A

Explanation:
Step 1: Augment the data
Scenario: Columns in each dataset contain missing and null values. The datasets also contain many outliers.
Step 2: Add the Bayesian Linear Regression module.
Scenario: You produce a regression model to predict property prices by using the Linear Regression and Bayesian Linear Regression modules.
Step 3: Configure the regularization weight.
Regularization typically is used to avoid overfitting. For example, in L2 regularization weight, type the value to use as the weight for L2 regularization. We recommend that you use a non-zero value to avoid overfitting.
Scenario:
Model fit: The model shows signs of overfitting. You need to produce a more refined regression model that reduces the overfitting.

NEW QUESTION 214
- (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 217
- (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.
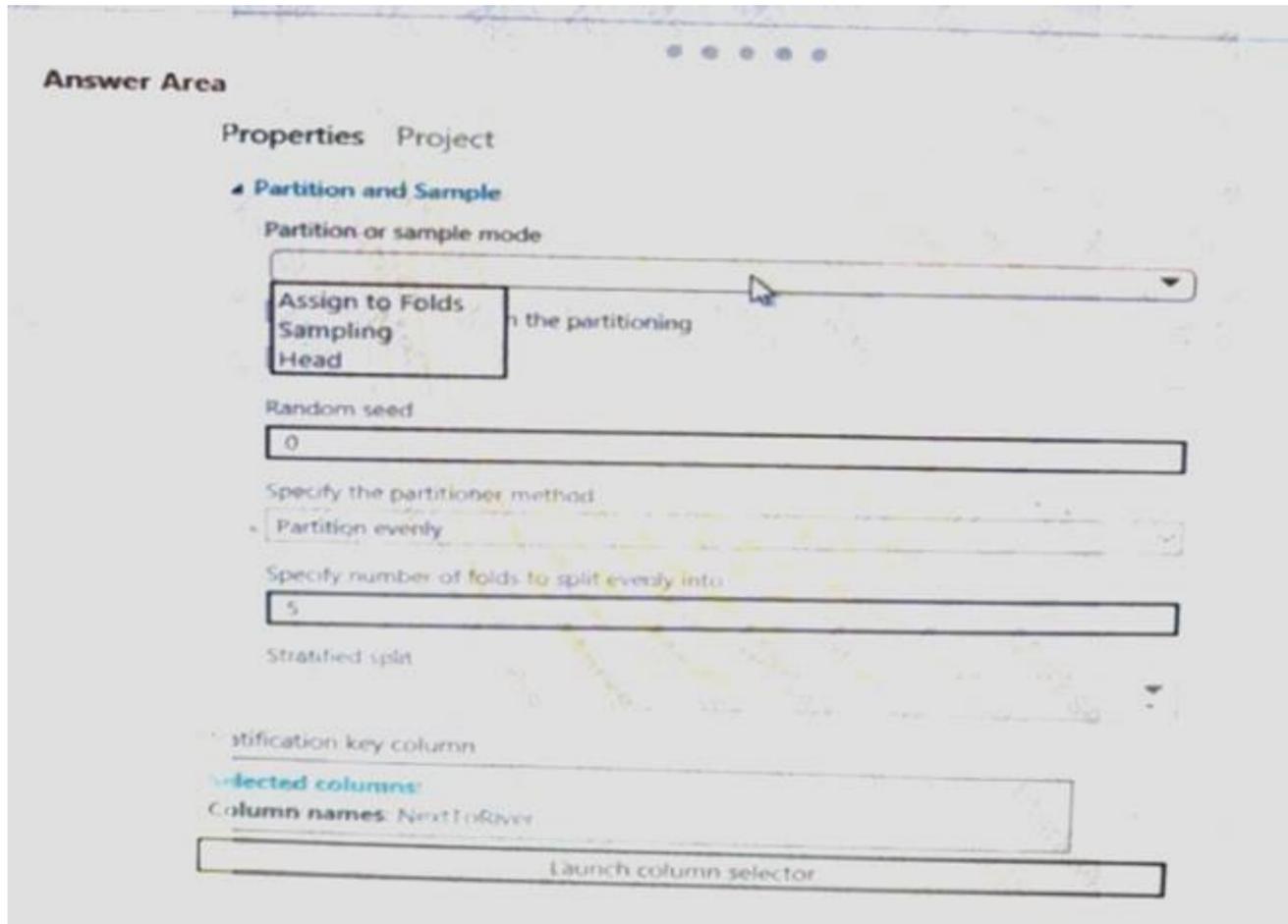
A. Mastered
B. Not Mastered

Answer: A

Explanation:
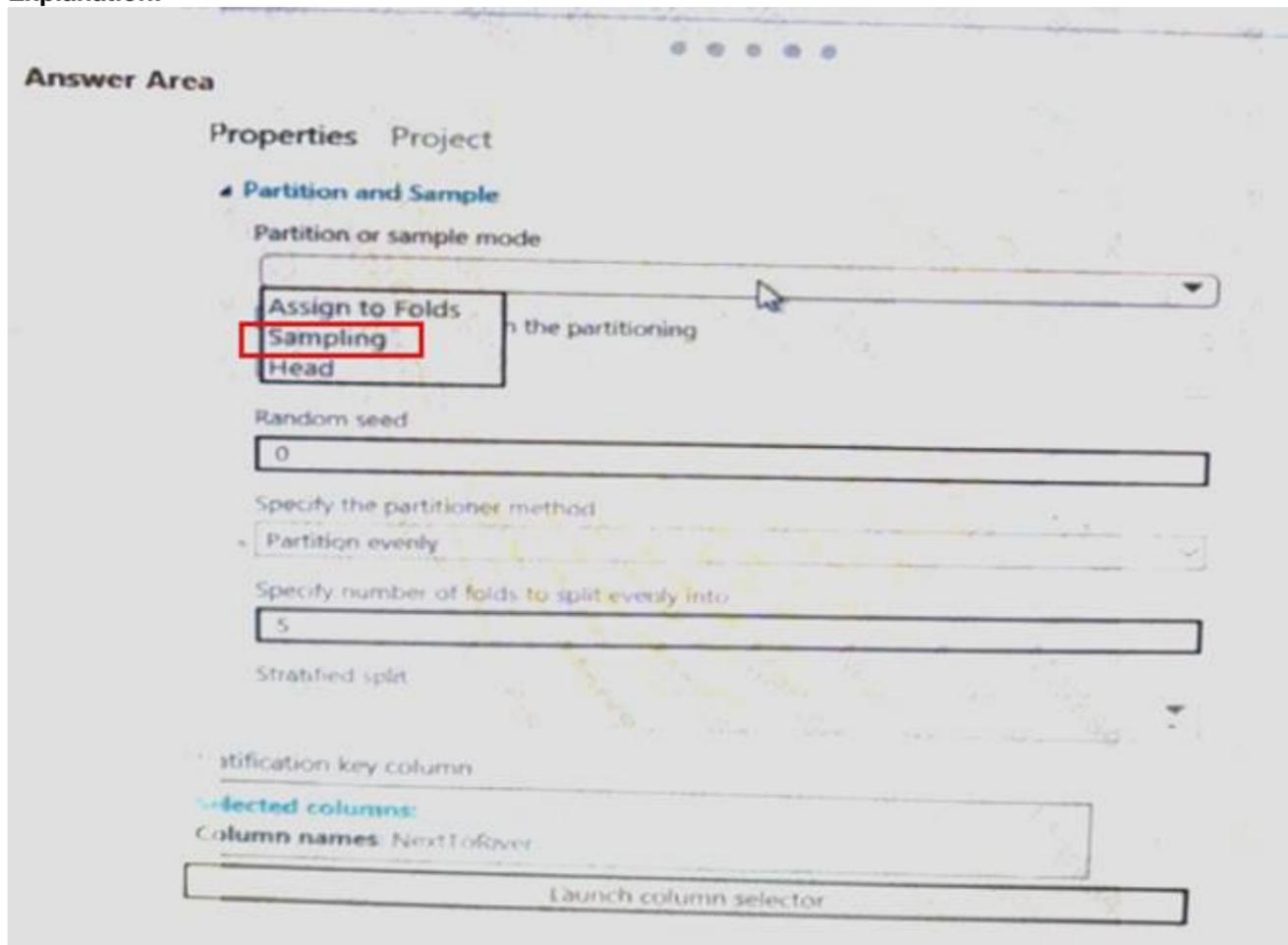
NEW QUESTION 219
- (Exam Topic 2)
You need to implement early stopping criteria as suited in the model training requirements.
Which three code segments should you use to develop the solution? To answer, move the appropriate code segments from the list of code segments to the answer area and arrange them in the correct order.
NOTE: More than one order of answer choices is correct. You will receive credit for any of the correct orders you select.

Code segments   Answer Area

```
early_termination_policy =
TruncationSelectionPolicy(evaluation_interval=1,
truncation_percentage=20, delay_evaluation=5)
```

```
import TruncationSelectionPolicy
```

```
from azureml.train.hyperdrive
```

```
import BanditPolicy
```

```
early_termination_policy = BanditPolicy
(slack_factor = 0.1, evaluation_interval=1,
delay_evaluation=5)
```

A. Mastered
B. Not Mastered

Answer: A

Explanation:
You need to implement an early stopping criterion on models that provides savings without terminating promising jobs.
Truncation selection cancels a given percentage of lowest performing runs at each evaluation interval. Runs are compared based on their performance on the primary metric and the lowest X% are terminated. Example:
from azureml.train.hyperdrive import TruncationSelectionPolicy
early_termination_policy = TruncationSelectionPolicy(evaluation_interval=1, truncation_percentage=20, delay_evaluation=5)

NEW QUESTION 224
- (Exam Topic 2)
You need to configure the Permutation Feature Importance module for the model training requirements. What should you do? To answer, select the appropriate options in the dialog box in the answer area. NOTE: Each correct selection is worth one point.

Answer Area

Permutation Feature Importance

Random seed

```
0
500
```

```
Regression – Root Mean Square Error
Regression – R-squared
Regression – Mean Zero One Error
Regression – Mean Absolute Error
```

A. Mastered
B. Not Mastered

Answer: A

Explanation:
Box 1: 500
For Random seed, type a value to use as seed for randomization. If you specify 0 (the default), a number is generated based on the system clock.
A seed value is optional, but you should provide a value if you want reproducibility across runs of the same experiment.
Here we must replicate the findings. Box 2: Mean Absolute Error
Scenario: Given a trained model and a test dataset, you must compute the Permutation Feature Importance scores of feature variables. You need to set up the Permutation Feature Importance module to select the correct metric to investigate the model's accuracy and replicate the findings.
Regression. Choose one of the following: Precision, Recall, Mean Absolute Error , Root Mean Squared Error, Relative Absolute Error, Relative Squared Error,

Coefficient of Determination
References:
https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/permutation-feature-importan

NEW QUESTION 227
- (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.
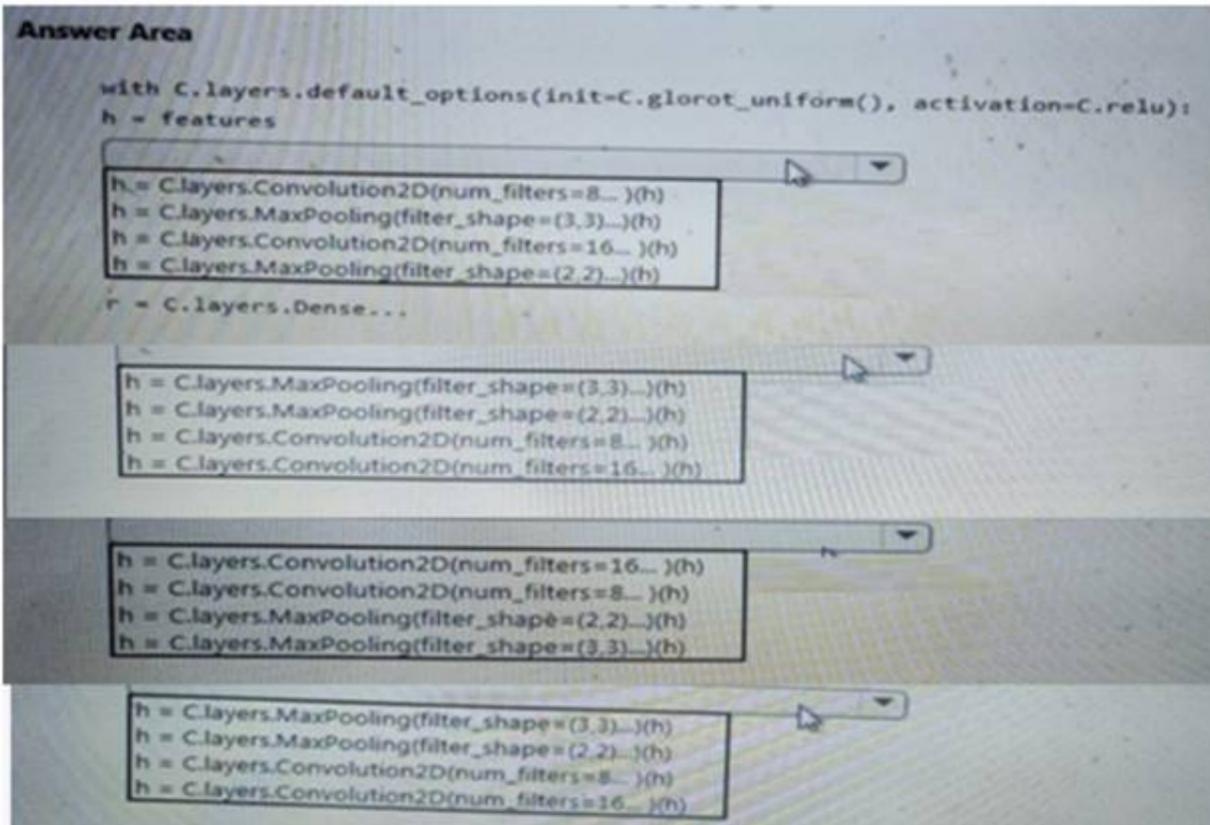

A. Mastered
B. Not Mastered

Answer: A

Explanation:
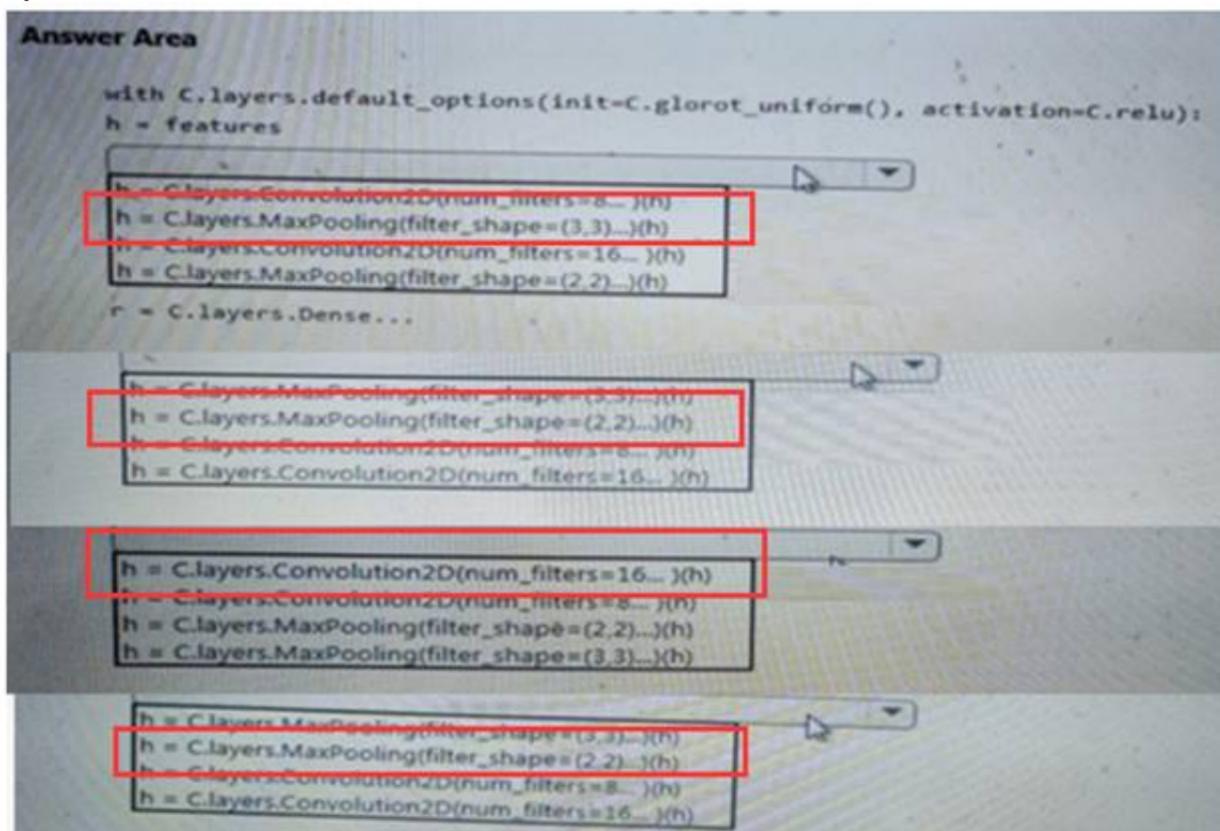
NEW QUESTION 231
- (Exam Topic 1)
You need to implement a new cost factor scenario for the ad response models as illustrated in the performance curve exhibit.
Which technique should you use?

A. Set the threshold to 0.5 and retrain if weighted Kappa deviates +/- 5% from 0.45.
B. Set the threshold to 0.05 and retrain if weighted Kappa deviates +/- 5% from 0.5.
C. Set the threshold to 0.2 and retrain if weighted Kappa deviates +/- 5% from 0.6.
D. Set the threshold to 0.75 and retrain if weighted Kappa deviates +/- 5% from 0.15.

Answer: A

Explanation:
Scenario:
Performance curves of current and proposed cost factor scenarios are shown in the following diagram:
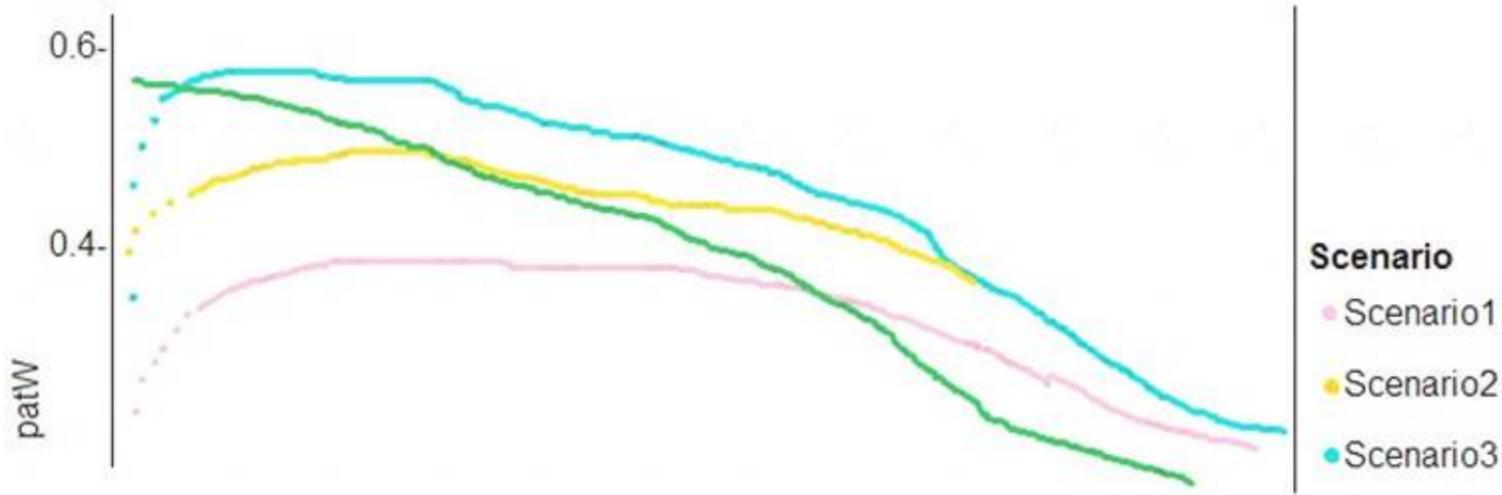
The ad propensity model uses a cut threshold is 0.45 and retrains occur if weighted Kappa deviated from 0.1 +/- 5%.

NEW QUESTION 236
- (Exam Topic 1)
You need to use the Python language to build a sampling strategy for the global penalty detection 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.

```
import pytorch as deeplearninglib
import tensorflow as deeplearninglib
import cntk as deeplearninglib
```

```
train_smapler = deeplearminglib.DistributedSampler.(penalty_video_dataset)
train_sampler = deeplearminglib.log_uniform_candidate_sampler.(penalty_video_dataset)
train_sampler = deeplearninglib.WeightedRandomSampler.(penalty_video_dataset)
train_sampler = deeplearninglib.all_candidate_sampler.(penalty_video_dataset)

...
train loader -

...
(train_smapler, penalty_video_dataset)
```

```
optimizer = deeplearninglib.optim.SGD(model.parameters().lr=0,01)
optimizer = deeplearninglib.train.GradientDescentOptimizer(learning_rate=0.10)
```

```
model = deeplearninglib.parallel.Distributed(DataParallel(model)
model = deeplearninglib.nn.parallel.DistributedDataParallelCPU(model)
model = deeplearninglib.keras.Model([
model = deeplearninglib.keras.Sequental([

...
train_sampler.set_epoch(epoch)
    for data, target in train_loader:
        data, target = data.to(device), target.to(device)
```

A. Mastered
B. Not Mastered

Answer: A

Explanation:
Box 1: import pytorch as deeplearninglib Box 2: ..DistributedSampler(Sampler).. DistributedSampler(Sampler):
Sampler that restricts data loading to a subset of the dataset.
It is especially useful in conjunction with class:`torch.nn.parallel.DistributedDataParallel`. In such case, each process can pass a DistributedSampler instance as a DataLoader sampler, and load a subset of the original dataset that is exclusive to it.
Scenario: Sampling must guarantee mutual and collective exclusively between local and global segmentation models that share the same features.
Box 3: optimizer = deeplearninglib.train. GradientDescentOptimizer(learning_rate=0.10)

NEW QUESTION 237
......

THANKS FOR TRYING THE DEMO OF OUR PRODUCT

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

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

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

Money Back Guarantee

DP-100 Practice Exam Features:

* DP-100 Questions and Answers Updated Frequently

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

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

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