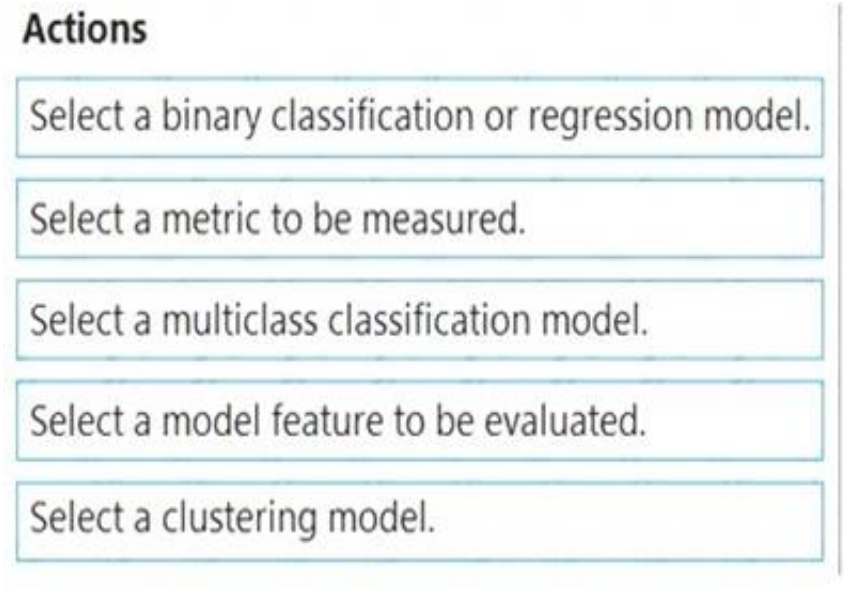
# Microsoft

## Exam Questions DP-100

Designing and Implementing a Data Science Solution on Azure

**NEW QUESTION 1**
- (Exam Topic 3)
You have several machine learning models registered in an Azure Machine Learning workspace. You must use the Fairlearn dashboard to assess fairness in a selected model.
Which three actions should you perform in sequence? To answer, move the appropriate actions from the list of actions to the answer area and arrange them in the correct order.




A. Mastered
B. Not Mastered

**Answer:** A

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

**NEW QUESTION 2**
- (Exam Topic 3)
HOTSPOT
You create a script for training a machine learning model in Azure Machine Learning service. You create an estimator by running the following code:

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

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

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

A. Mastered
B. Not Mastered

**Answer:** A

**Explanation:**
Box 1: Yes
Parameter source_directory is a local directory containing experiment configuration and code files needed for a training job.
Box 2: Yes
script_params is a dictionary of command-line arguments to pass to the training script specified in entry_script.
Box 3: No
Box 4: Yes
The conda_packages parameter is a list of strings representing conda packages to be added to the Python environment for the experiment.

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

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

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

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

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

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

A. Mastered
B. Not Mastered

**Answer:** A

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

**NEW QUESTION 4**
- (Exam Topic 3)
Note: This question is part of a series of questions that present the same scenario. Each question in the series contains a unique solution that might meet the stated goals. Some question sets might have more than one correct solution, while others might not have a correct solution.
After you answer a question in this section, you will NOT be able to return to it. As a result, these questions will not appear in the review screen.
An IT department creates the following Azure resource groups and resources:

| Resource group | Resources |
|---|---|
| ml_resources | • an Azure Machine Learning workspace named amlworkspace<br>• an Azure Storage account named amlworkspace12345<br>• an Application Insights instance named amlworkspace54321<br>• an Azure Key Vault named amlworkspace67890<br>• an Azure Container Registry named amlworkspace09876 |
| general_compute | A virtual machine named mlvm with the following configuration:<br>• Operating system: Ubuntu Linux<br>• Software installed: Python 3.6 and Jupyter Notebooks<br>• 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 5**
- (Exam Topic 3)
You train a machine learning model.
You must deploy the model as a real-time inference service for testing. The service requires low CPU utilization and less than 48 MB of RAM. The compute target for the deployed service must initialize automatically while minimizing cost and administrative overhead.
Which compute target should you use?

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

**Answer:** C

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

**NEW QUESTION 6**
- (Exam Topic 3)
You use an Azure Machine Learning workspace. You create the following Python code:

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

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

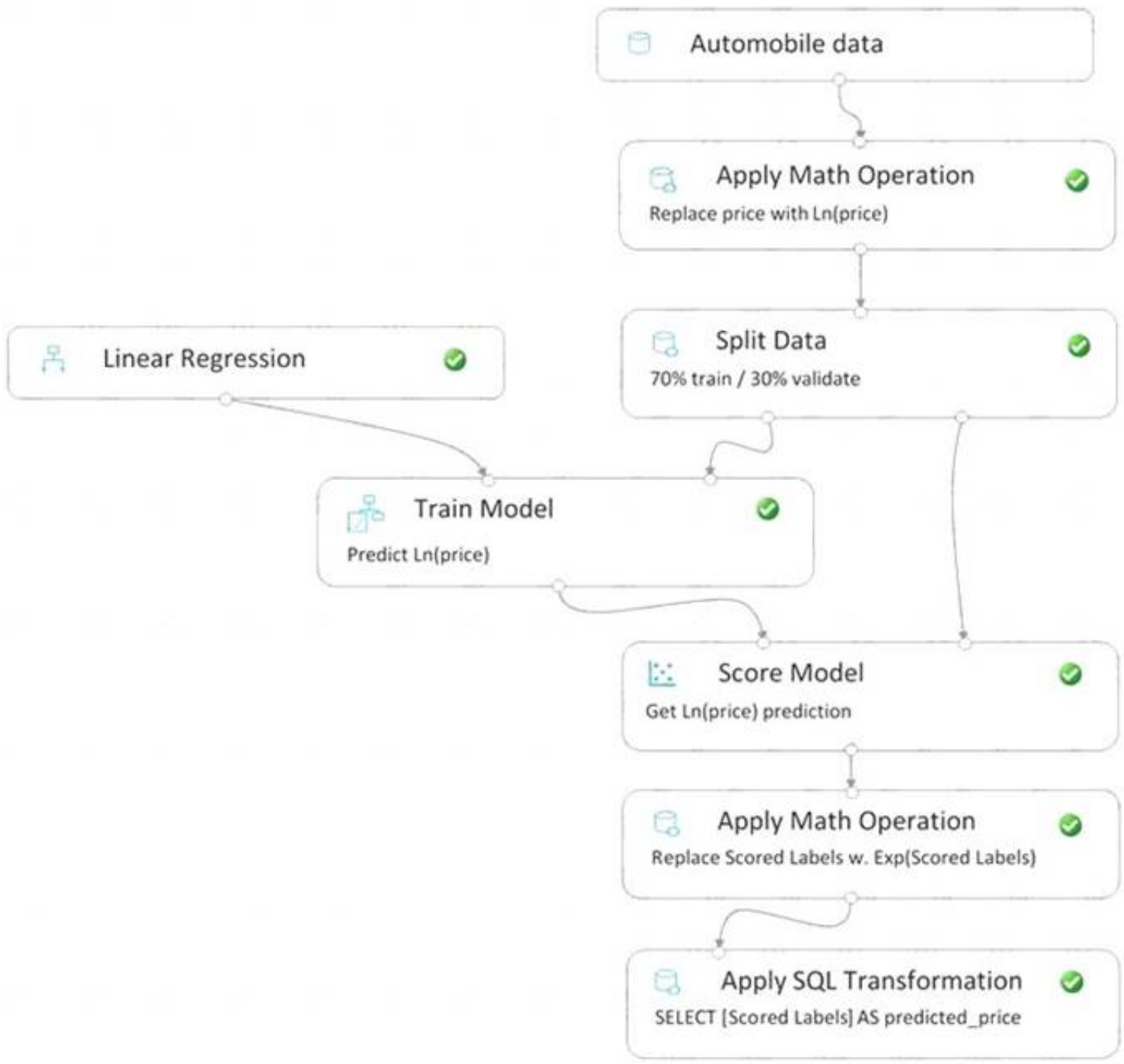
| Statements | Yes | No |
|---|---|---|
| The default environment will be created | ○ | ○ |
| The training script will run on local compute | ○ | ○ |
| A script run configuration runs a training script named `train.py` located in a directory defined by the `project_folder` variable | ○ | ○ |

A. Mastered
B. Not Mastered

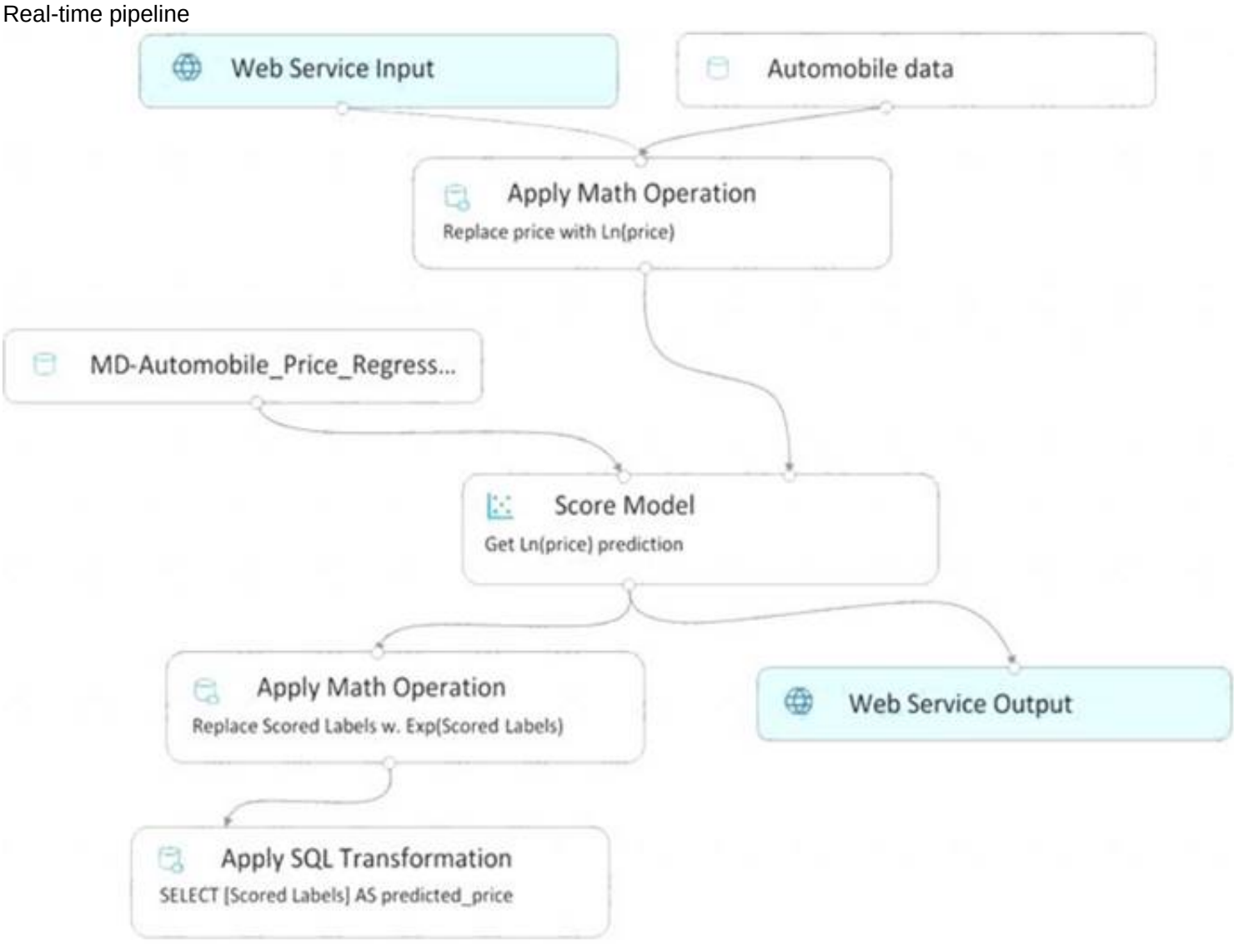
**Answer:** A

**Explanation:**
Graphical user interface, text, application Description automatically generated
Box 1: No
Environment is a required parameter. The environment to use for the run. If no environment is specified, azureml.core.runconfig.DEFAULT_CPU_IMAGE will be used as the Docker image for the run.
The following example shows how to instantiate a new environment. from azureml.core import Environment
myenv = Environment(name="myenv") Box 2: Yes
Parameter compute_target: The compute target where training will happen. This can either be a ComputeTarget object, the name of an existing ComputeTarget, or the string "local". If no compute target is specified, your local machine will be used.
Box 3: Yes
Parameter source_directory. A local directory containing code files needed for a run. Parameter script. The file path relative to the source_directory of the script to be run. Reference:
https://docs.microsoft.com/en-us/python/api/azureml-core/azureml.core.scriptrunconfig https://docs.microsoft.com/en-us/python/api/azureml-core/azureml.core.environment.environment

**NEW QUESTION 7**
- (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.

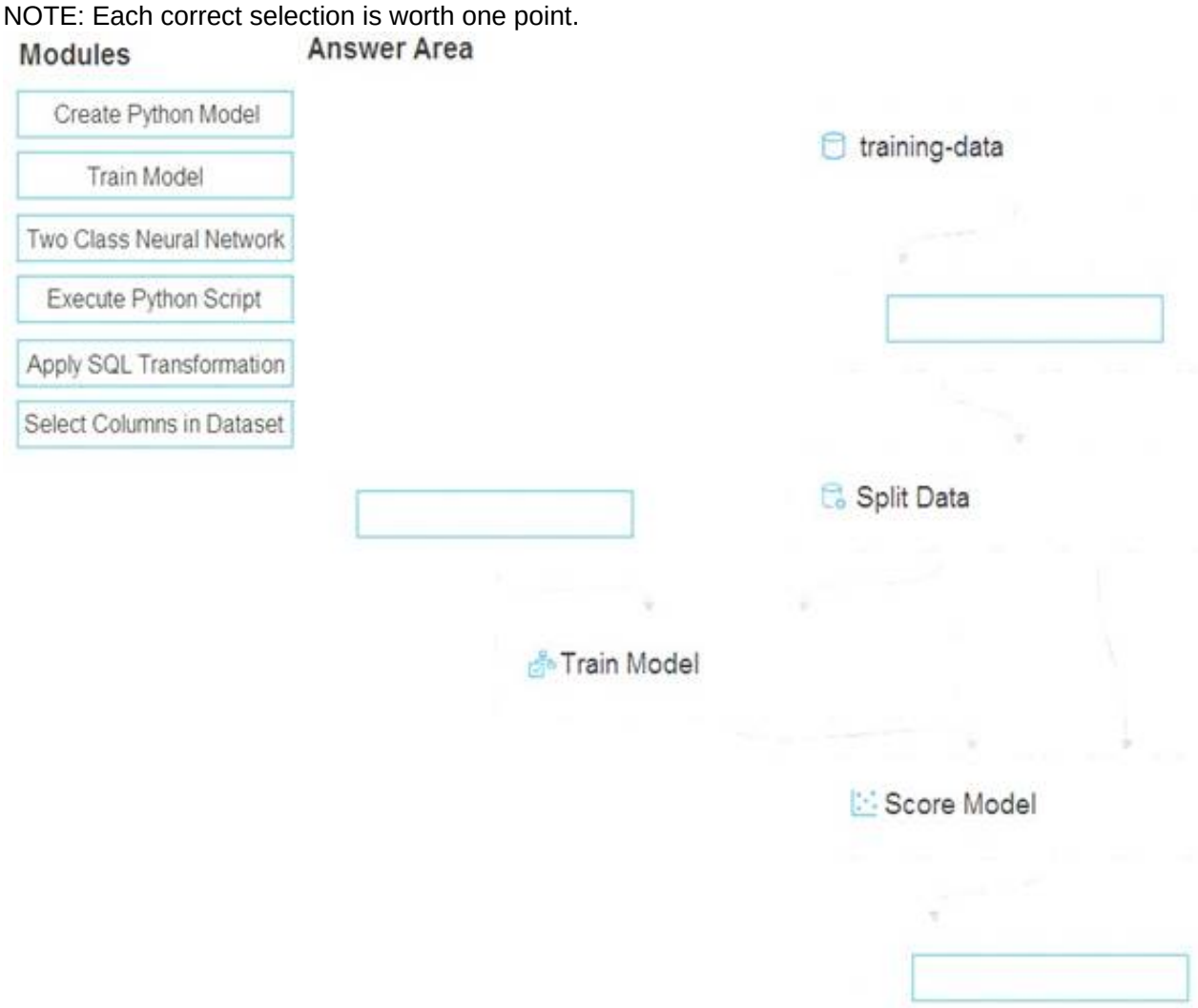
F. Remove the Apply SQL Transformation module from the data flow.

**Answer:** ACE

**NEW QUESTION 8**
- (Exam Topic 3)
You create a training pipeline using the Azure Machine Learning designer. You upload a CSV file that contains the data from which you want to train your model. You need to use the designer to create a pipeline that includes steps to perform the following tasks:

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

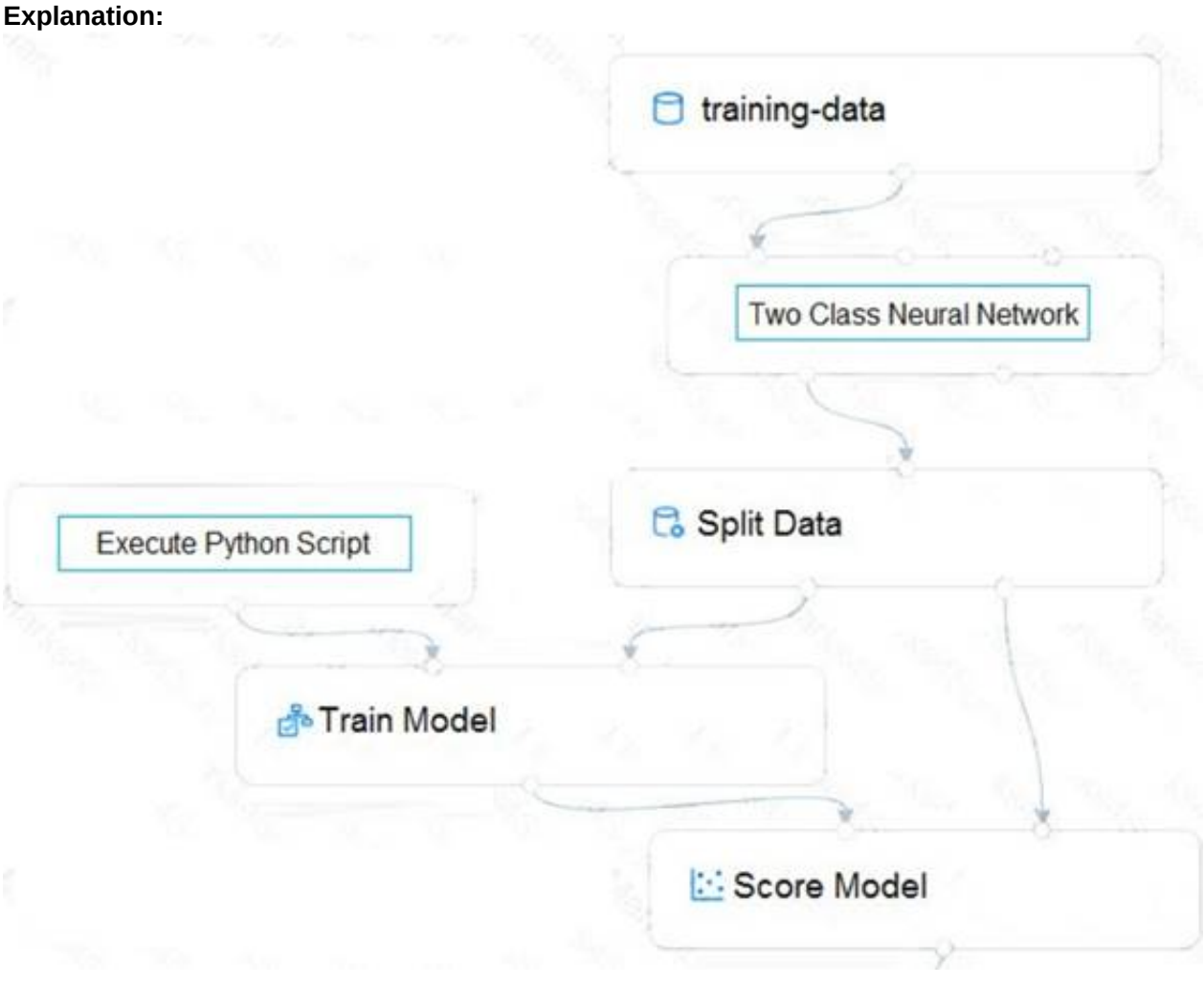
NOTE: Each correct selection is worth one point.




A. Mastered
B. Not Mastered

**Answer:** A

**Explanation:**



**NEW QUESTION 9**
- (Exam Topic 3)
Your Azure Machine Learning workspace has a dataset named real_estate_data. A sample of the data in the dataset follows.

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

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

**Answer Area**

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

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

▼ =split1_ds,
- X
- Y
- X_valid
- Y_valid
- training_data

▼ =split2_ds
- X
- Y
- X_valid
- Y_valid
- validation_data
- training_data

▼ ='price')
- y
- y_valid
- y_max
- label_column_name
- exclude_nan_labels

A. Mastered
B. Not Mastered

**Answer:** A

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

**NEW QUESTION 10**
- (Exam Topic 3)
You train a model and register it in your Azure Machine Learning workspace. You are ready to deploy the model as a real-time web service.
You deploy the model to an Azure Kubernetes Service (AKS) inference cluster, but the deployment fails because an error occurs when the service runs the entry

script that is associated with the model deployment.
You need to debug the error by iteratively modifying the code and reloading the service, without requiring a re-deployment of the service for each code update.
What should you do?

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

**Answer:** C

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

**NEW QUESTION 10**
- (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 12**
- (Exam Topic 3)
You are developing deep learning models to analyze semi-structured, unstructured, and structured data types. You have the following data available for model building:
- Video recordings of sporting events
- Transcripts of radio commentary about events
- Logs from related social media feeds captured during sporting events You need to select an environment for creating the model.
Which environment should you use?

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

**Answer:** A

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

**NEW QUESTION 16**
- (Exam Topic 3)
Note: This question is part of a series of questions that present the same scenario. Each question in the series contains a unique solution that might meet the stated goals. Some question sets might have more than one correct solution, while others might not have a correct solution.
After you answer a question in this section, you will NOT be able to return to it. As a result, these questions will not appear in the review screen.
You plan to use a Python script to run an Azure Machine Learning experiment. The script creates a reference to the experiment run context, loads data from a file, identifies the set of unique values for the label column, and completes the experiment run:

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

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

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

A. Yes
B. No

**Answer:** A

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

**NEW QUESTION 21**
- (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
```
▼ (1 of 3)
- from azureml.core import Run
- import mlflow
- import logging

```
# Start logging for this run
```
▼ (2 of 3)
- run = Run.get_context()
- mlflow.start_run()
- logger = logging.getLogger('Run')

```
reg_rate = 0.01
# Log the reg_rate metric
```
▼ (3 of 3)
- 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
```
▼ (4 of 3)
- 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 23**
- (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 26**
- (Exam Topic 3)
Note: This question is part of a series of questions that present the same scenario. Each question in the series contains a unique solution that might meet the stated goals. Some question sets might have more than one correct solution, while others might not have a correct solution.
After you answer a question in this section, you will NOT be able to return to it. As a result, these questions will not appear in the review screen.
You are analyzing a numerical dataset which contains missing values in several columns.
You must clean the missing values using an appropriate operation without affecting the dimensionality of the feature set.
You need to analyze a full dataset to include all values.
Solution: Remove the entire column that contains the missing data point. Does the solution meet the goal?

A. Yes
B. No

**Answer:** B

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

**NEW QUESTION 31**
- (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 33**
- (Exam Topic 3)
You are creating a binary classification by using a two-class logistic regression model. You need to evaluate the model results for imbalance.
Which evaluation metric should you use?

A. Relative Absolute Error
B. AUC Curve
C. Mean Absolute Error
D. Relative Squared Error

**Answer:** B

**Explanation:**
One can inspect the true positive rate vs. the false positive rate in the Receiver Operating Characteristic (ROC) curve and the corresponding Area Under the Curve (AUC) value. The closer this curve is to the upper left corner, the better the classifier's performance is (that is maximizing the true positive rate while minimizing the false positive rate). Curves that are close to the diagonal of the plot, result from classifiers that tend to make predictions that are close to random guessing.
References:
https://docs.microsoft.com/en-us/azure/machine-learning/studio/evaluate-model-performance#evaluating-a-bina

**NEW QUESTION 37**
- (Exam Topic 3) You are solving a classification task. The dataset is imbalanced.
You need to select an Azure Machine Learning Studio module to improve the classification accuracy. Which module should you use?

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

**Answer:** C

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

**NEW QUESTION 42**
- (Exam Topic 3)
You are moving a large dataset from Azure Machine Learning Studio to a Weka environment. You need to format the data for the Weka environment.
Which module should you use?

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

**Answer:** C

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

**NEW QUESTION 43**
- (Exam Topic 3)
You are using the Azure Machine Learning Service to automate hyperparameter exploration of your neural network classification model.
You must define the hyperparameter space to automatically tune hyperparameters using random sampling according to following requirements:

- The learning rate must be selected from a normal distribution with a mean value of 10 and a standard deviation of 3.
- Batch size must be 16, 32 and 64.
- Keep probability must be a value selected from a uniform distribution between the range of 0.05 and 0.1.

You need to use the param_sampling method of the Python API for the Azure Machine Learning Service. How should you complete the code segment? To answer, select the appropriate options in the answer area. NOTE: Each correct selection is worth one point.

```
from azureml.train.hyperdrive import RandomParameterSampling
param_sampling = RandomParameterSampling( {
"learning_rate" :  ▼ ,
                   uniform(10,3)
                   normal(10,3)
                   choice(10,3)
                   Loguniform(10,3)
"batch_size":      ▼ ,
                   choice(16,32,64)
                   choice(range(16,64))
                   normal(16,32,64)
                   normal(range(16,64))
"keep_probability" :  ▼
                   choice(range(0.05, 0.1))
                   uniform(0.05, 0.1)
                   normal(0.05, 0.1)
                   lognormal(0.05, 0.1)
    }
)
```

A. Mastered
B. Not Mastered

**Answer:** A

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

**NEW QUESTION 46**
- (Exam Topic 3)
You are using the Hyperdrive feature in Azure Machine Learning to train a model. You configure the Hyperdrive experiment by running the following code:

```
from azureml.train.hyperdrive import RandomParameterSampling
param_sampling = RandomParameterSampling( {
    "learning_rate": normal(10, 3),
    "keep_probability": uniform(0.05, 0.1),
    "batch_size": choice(16, 32, 64, 128)
    "number_of_hidden_layers": choice(range(3,5))
  }
)
```

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 |
|---|---|---|
| By defining sampling in this manner, every possible combination of the parameters will be tested. | ○ | ○ |
| Random values of the learning_rate parameter will be selected from a normal distribution with a mean of 10 and a standard deviation of 3. | ○ | ○ |
| The keep_probability parameter value will always be either **0.05** or **0.1**. | ○ | ○ |
| Random values for the number_of_hidden_layers parameter will be selected from a normal distribution with a mean of 3 and a standard deviation of 5. | ○ | ○ |

A. Mastered
B. Not Mastered

**Answer:** A

**Explanation:**
Box 1: Yes
In random sampling, hyperparameter values are randomly selected from the defined search space. Random sampling allows the search space to include both discrete and continuous hyperparameters.
Box 2: Yes
learning_rate has a normal distribution with mean value 10 and a standard deviation of 3.
Box 3: No
keep_probability has a uniform distribution with a minimum value of 0.05 and a maximum value of 0.1.
Box 4: No
number_of_hidden_layers takes on one of the values [3, 4, 5].
Reference:
https://docs.microsoft.com/en-us/azure/machine-learning/how-to-tune-hyperparameters

**NEW QUESTION 47**
- (Exam Topic 3)
You have a multi-class image classification deep learning model that uses a set of labeled photographs. You create the following code to select hyperparameter values when training the model.

```
from azureml.train.hyperdrive import BayesianParameterSampling
param_sampling = BayesianParametersSampling ({
 "learning_rate": uniform(0.01, 0.1),
 "batch_size": choice(16, 32, 64, 128)}
)
```

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

| | Yes | No |
|---|---|---|
| Hyperparameter combinations for the runs are selected based on how previous samples performed in the previous experiment run. | ○ | ○ |
| The learning rate value 0.09 might be used during model training. | ○ | ○ |
| You can define an early termination policy for this hyperparameter tuning run. | ○ | ○ |

A. Mastered
B. Not Mastered

**Answer:** A

**Explanation:**
Box 1: Yes
Hyperparameters are adjustable parameters you choose to train a model that govern the training process itself. Azure Machine Learning allows you to automate hyperparameter exploration in an efficient manner, saving you significant time and resources. You specify the range of hyperparameter values and a maximum number of training runs. The system then automatically launches multiple simultaneous runs with different parameter configurations and finds the configuration that results in the best performance, measured by the metric you choose. Poorly performing training runs are automatically early terminated, reducing wastage of compute resources. These resources are instead used to explore other hyperparameter configurations.
Box 2: Yes
uniform(low, high) - Returns a value uniformly distributed between low and high Box 3: No
Bayesian sampling does not currently support any early termination policy. Reference:
https://docs.microsoft.com/en-us/azure/machine-learning/how-to-tune-hyperparameters

**NEW QUESTION 49**
- (Exam Topic 3)
You create a multi-class image classification deep learning model that uses a set of labeled images. You create a script file named train.py that uses the PyTorch 1.3 framework to train the model.
You must run the script by using an estimator. The code must not require any additional Python libraries to be installed in the environment for the estimator. The time required for model training must be minimized.
You need to define the estimator that will be used to run the script. Which estimator type should you use?

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

**Answer:** B

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

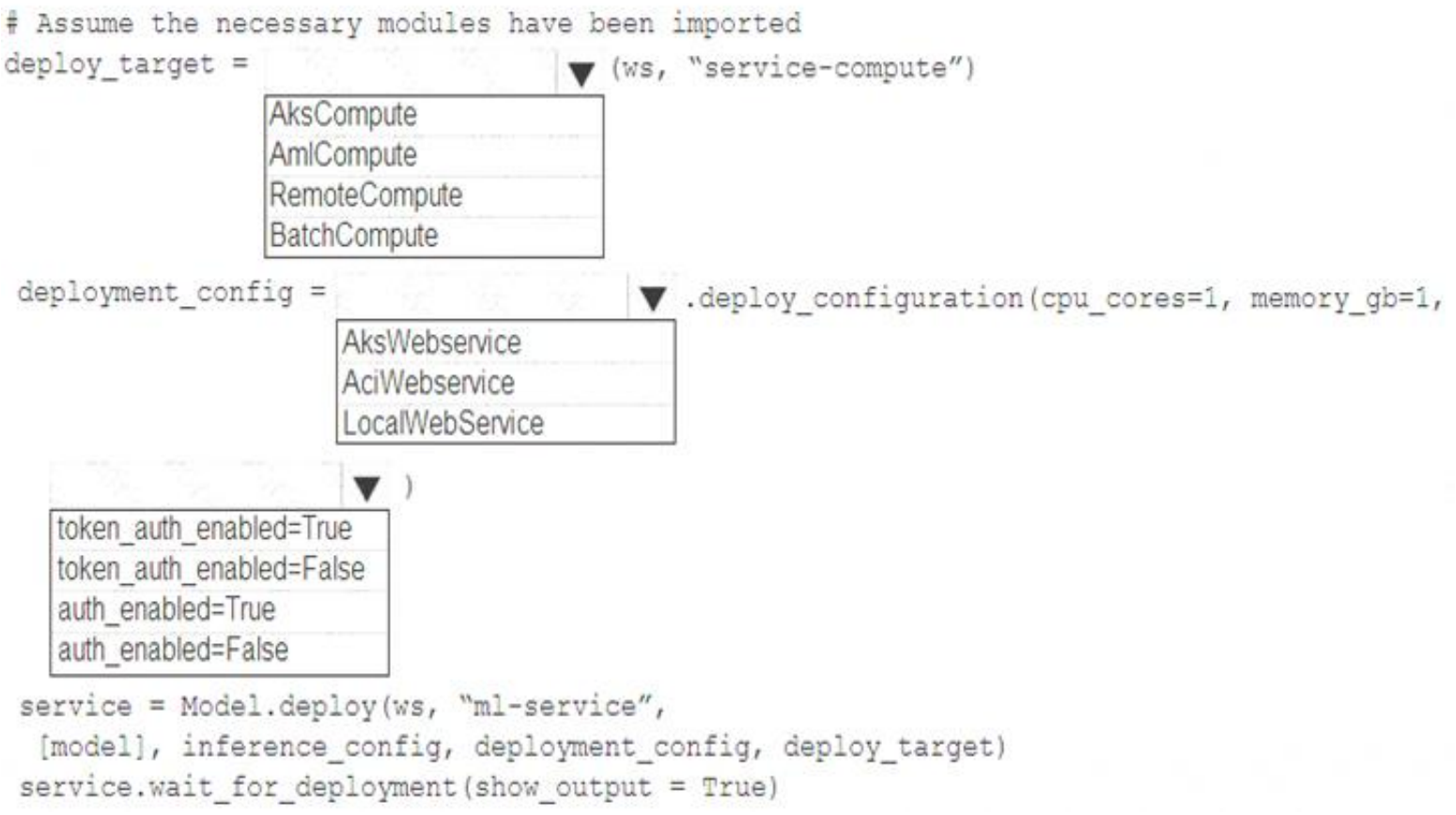
**NEW QUESTION 52**
- (Exam Topic 3)
You are building a binary classification model by using a supplied training set. The training set is imbalanced between two classes.
You need to resolve the data imbalance.
What are three possible ways to achieve this goal? Each correct answer presents a complete solution NOTE: Each correct selection is worth one point.

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

**Answer:** ABD

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

**NEW QUESTION 53**
- (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.




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 56**
- (Exam Topic 3)
A coworker registers a datastore in a Machine Learning services workspace by using the following code:

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

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

```
import azureml.core
from azureml.core import Workspace, Datastore
ws = Workspace.from_config()
datastore = [Workspace / Datastore / Experiment / Run] .get( [ws / run / experiment / log] , ' [demo_datastore / demo_datacontainer / demo_account / Datastore]
```

A. Mastered
B. Not Mastered

**Answer:** A

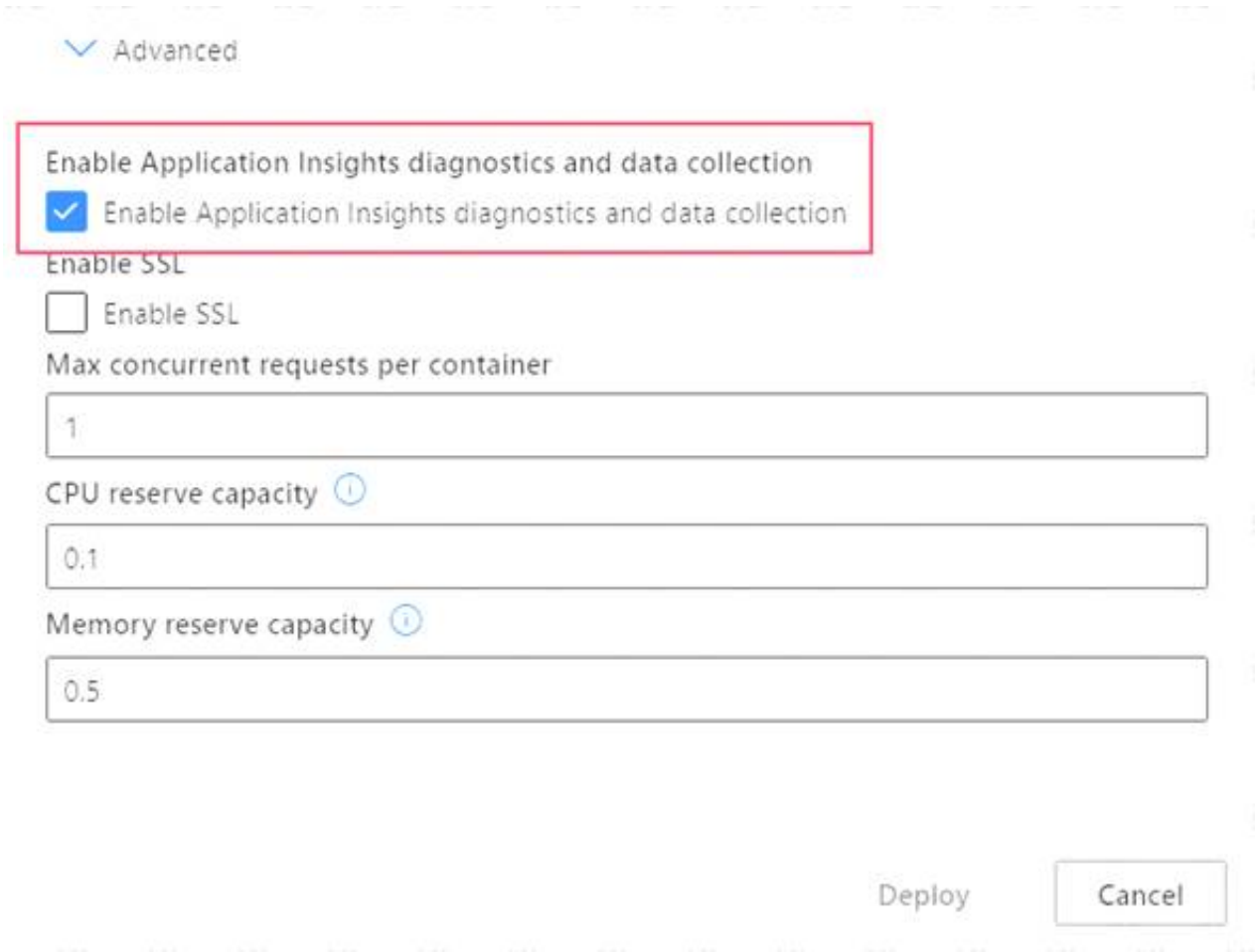
**Explanation:**
Box 1: DataStore
To get a specific datastore registered in the current workspace, use the get() static method on the Datastore class:
# Get a named datastore from the current workspace
datastore = Datastore.get(ws, datastore_name='your datastore name') Box 2: ws
Box 3: demo_datastore Reference:
https://docs.microsoft.com/en-us/azure/machine-learning/how-to-access-data

**NEW QUESTION 60**
- (Exam Topic 3)
You deploy a real-time inference service for a trained model.
The deployed model supports a business-critical application, and it is important to be able to monitor the data submitted to the web service and the predictions the data generates.
You need to implement a monitoring solution for the deployed model using minimal administrative effort. What should you do?

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

**Answer:** B

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




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

**NEW QUESTION 64**
- (Exam Topic 3)
You use the following code to run a script as an experiment in Azure Machine Learning:

```
from azureml.core import Workspace, Experiment, Run
from azureml.core import RunConfig, ScriptRunConfig
ws = Workspace.from_config()
run_config = RunConfiguration()
run_config.target='local'
script_config = ScriptRunConfig(source_directory='./script', script='experiment.py', run_config=run_config)
experiment = Experiment(workspace=ws, name='script experiment')
run = experiment.submit(config=script_config)
run.wait_for_completion()
```

You must identify the output files that are generated by the experiment run. You need to add code to retrieve the output file names.
Which code segment should you add to the script?

A. files = run.get_properties()
B. files= run.get_file_names()
C. files = run.get_details_with_logs()
D. files = run.get_metrics()
E. files = run.get_details()

**Answer:** B

**Explanation:**
You can list all of the files that are associated with this run record by called run.get_file_names() Reference:
https://docs.microsoft.com/en-us/azure/machine-learning/how-to-track-experiments

**NEW QUESTION 65**
- (Exam Topic 3)
You are the owner of an Azure Machine Learning workspace.
You must prevent the creation or deletion of compute resources by using a custom role. You must allow all other operations inside the workspace.
You need to configure the custom role.
How should you complete the configuration? To answer, select the appropriate options in the answer area.
NOTE: Each correct selection is worth one point.

**Answer Area**

```
{
 "Name": "Data Scientist Custom",
 "IsCustom": true
 "Description": "Description"
 "Actions": [
        [ Microsoft.MachineLearningServices/workspaces/*/read
          Microsoft.MachineLearningServices/workspaces/computes/*/write
          Microsoft.MachineLearningServices/workspaces/delete ]
        [ Microsoft.MachineLearningServices/workspaces/*/write
          Microsoft.MachineLearningServices/workspaces/computes/*/write
          Microsoft.MachineLearningServices/workspaces/delete ]
 ],
 "NotActions": [
        [ Microsoft.MachineLearningServices/workspaces/*/read
          Microsoft.MachineLearningServices/workspaces/*/write
          Microsoft.MachineLearningServices/workspaces/computes/*/delete ]
        [ Microsoft.MachineLearningServices/workspaces/*/read
          Microsoft.MachineLearningServices/workspaces/*/write
          Microsoft.MachineLearningServices/workspaces/computes/*/write ]
 ],
 "AssignableScopes": [
   "/subscriptions/<subscription_id>"
 ]
}
```

A. Mastered
B. Not Mastered

**Answer:** A

**Explanation:**
Graphical user interface, application Description automatically generated
Graphical user interface, application Description automatically generated
Box 1: Microsoft.MachineLearningServices/workspaces/*/read
Reader role: Read-only actions in the workspace. Readers can list and view assets, including datastore credentials, in a workspace. Readers can't create or update these assets.
Box 2: Microsoft.MachineLearningServices/workspaces/*/write
If the roles include Actions that have a wildcard (*), the effective permissions are computed by subtracting the NotActions from the allowed Actions.
Box 3: Box 2: Microsoft.MachineLearningServices/workspaces/computes/*/delete
Box 4: Microsoft.MachineLearningServices/workspaces/computes/*/write Reference:
https://docs.microsoft.com/en-us/azure/role-based-access-control/overview#how-azure-rbac-determines-if-a-use

**NEW QUESTION 68**
- (Exam Topic 3)
You previously deployed a model that was trained using a tabular dataset named training-dataset, which is based on a folder of CSV files.
Over time, you have collected the features and predicted labels generated by the model in a folder containing a CSV file for each month. You have created two tabular datasets based on the folder containing the inference data: one named predictions-dataset with a schema that matches the training data exactly, including the predicted label; and another named features-dataset with a schema containing all of the feature columns and a timestamp column based on the filename, which includes the day, month, and year.
You need to create a data drift monitor to identify any changing trends in the feature data since the model was trained. To accomplish this, you must define the required datasets for the data drift monitor.
Which datasets should you use to configure the data drift monitor? To answer, drag the appropriate datasets to the correct data drift monitor options. Each source may be used once, more than once, or not at all. You may need to drag the split bar between panes or scroll to view content.
NOTE: Each correct selection is worth one point.


A. Mastered

B. Not Mastered

**Answer:** A

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

**NEW QUESTION 73**
- (Exam Topic 3)
Note: This question is part of a series of questions that present the same scenario. Each question in the series contains a unique solution that might meet the stated goals. Some question sets might have more than one correct solution, while others might not have a correct solution.
After you answer a question in this section, you will NOT be able to return to it. As a result, these questions will not appear in the review screen.
You have a Python script named train.py in a local folder named scripts. The script trains a regression model by using scikit-learn. The script includes code to load a training data file which is also located in the scripts folder.
You must run the script as an Azure ML experiment on a compute cluster named aml-compute.
You need to configure the run to ensure that the environment includes the required packages for model training. You have instantiated a variable named aml-compute that references the target compute cluster.
Solution: Run the following code:

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

Does the solution meet the goal?

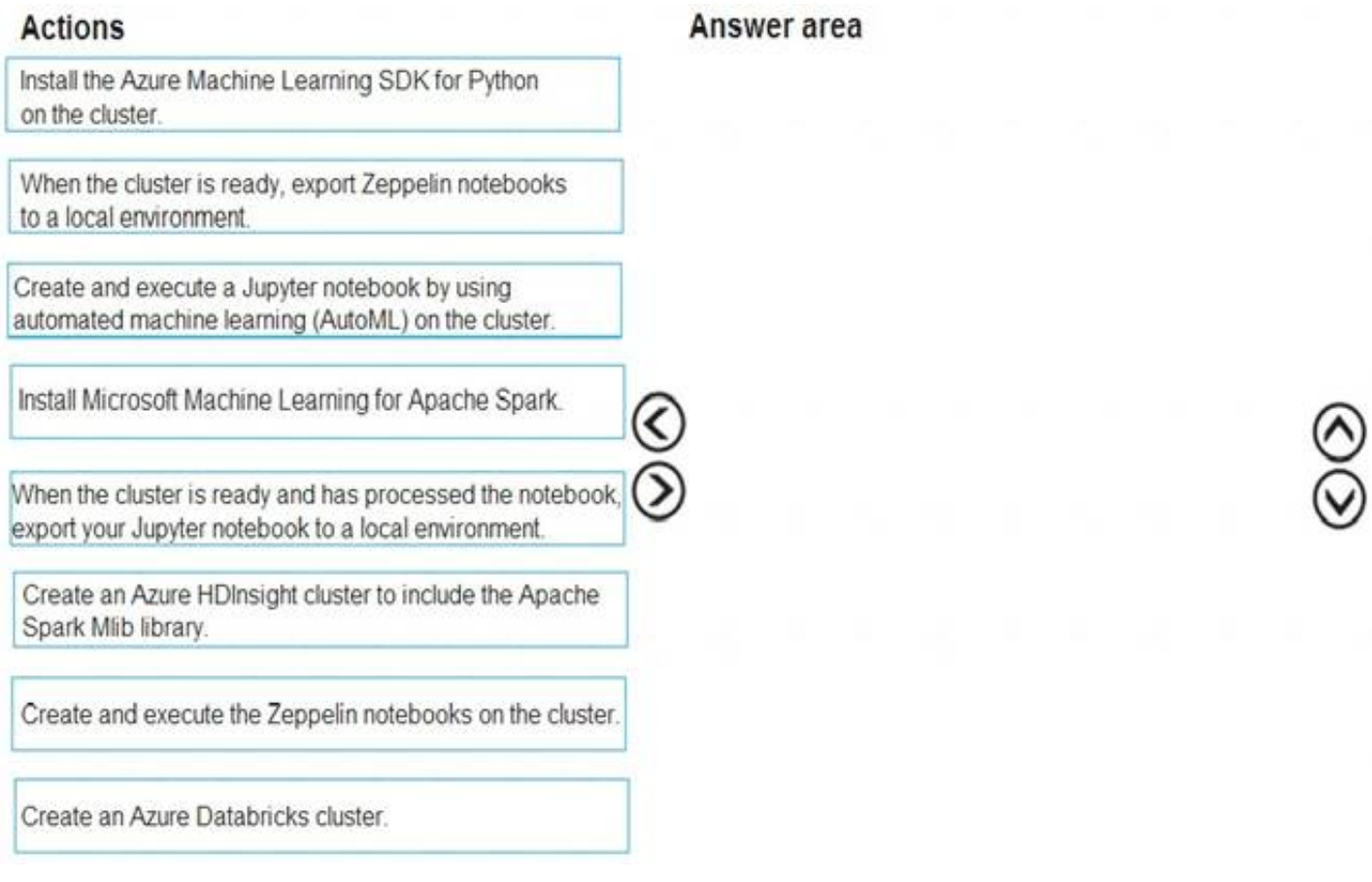
A. Yes
B. No

**Answer:** B

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

**NEW QUESTION 75**
- (Exam Topic 3)
You are building an intelligent solution using machine learning models. The environment must support the following requirements:
- Data scientists must build notebooks in a cloud environment
- Data scientists must use automatic feature engineering and model building in machine learning pipelines.
- Notebooks must be deployed to retrain using Spark instances with dynamic worker allocation.
- Notebooks must be exportable to be version controlled locally.

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




A. Mastered
B. Not Mastered

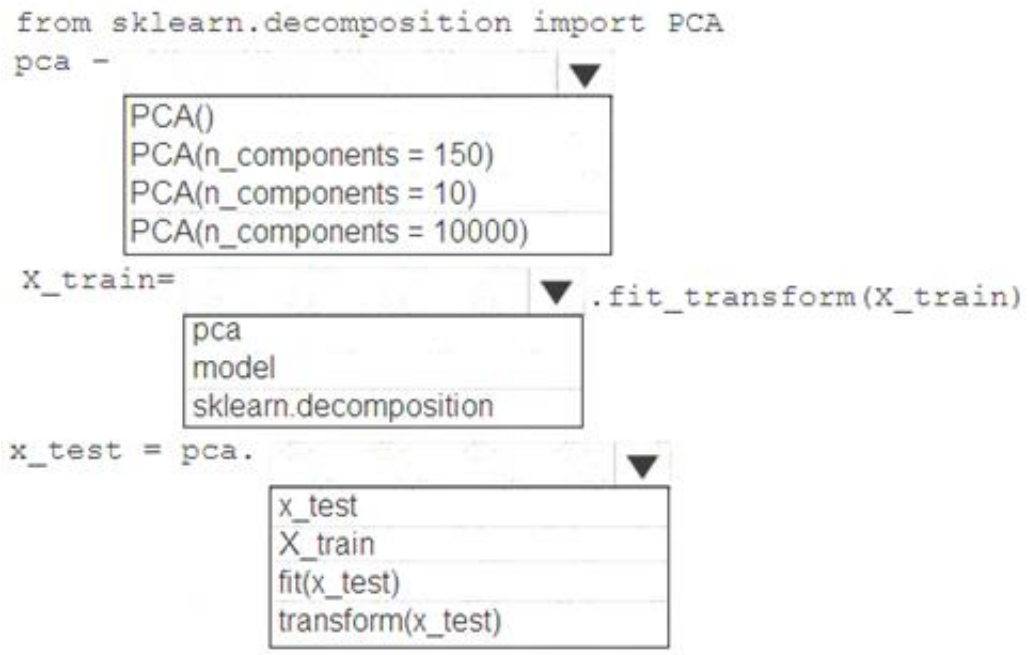
**Answer:** A

**Explanation:**
Step 1: Create an Azure HDInsight cluster to include the Apache Spark Mlib library
Step 2: Install Microsot Machine Learning for Apache Spark You install AzureML on your Azure HDInsight cluster.
Microsoft Machine Learning for Apache Spark (MMLSpark) provides a number of deep learning and data science tools for Apache Spark, including seamless integration of Spark Machine Learning pipelines with Microsoft Cognitive Toolkit (CNTK) and OpenCV, enabling you to quickly create powerful, highly-scalable predictive and analytical models for large image and text datasets.
Step 3: Create and execute the Zeppelin notebooks on the cluster
Step 4: When the cluster is ready, export Zeppelin notebooks to a local environment. Notebooks must be exportable to be version controlled locally.
References:
https://docs.microsoft.com/en-us/azure/hdinsight/spark/apache-spark-zeppelin-notebook https://azuremlbuild.blob.core.windows.net/pysparkapi/intro.html

**NEW QUESTION 76**
- (Exam Topic 3)
You have a dataset created for multiclass classification tasks that contains a normalized numerical feature set with 10,000 data points and 150 features.
You use 75 percent of the data points for training and 25 percent for testing. You are using the scikit-learn machine learning library in Python. You use X to denote the feature set and Y to denote class labels.
You create the following Python data frames:

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

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




A. Mastered
B. Not Mastered

**Answer:** A

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

**NEW QUESTION 79**
- (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 81**
- (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, Relative Squared Error, and the Coefficient of Determination.
Does the solution meet the goal?

A. Yes
B. No

**Answer:** A

**Explanation:**
The following metrics are reported for evaluating regression models. When you compare models, they are ranked by the metric you select for evaluation.
Mean absolute error (MAE) measures how close the predictions are to the actual outcomes; thus, a lower score is better.
Root mean squared error (RMSE) creates a single value that summarizes the error in the model. By squaring the difference, the metric disregards the difference between over-prediction and under-prediction.
Relative absolute error (RAE) is the relative absolute difference between expected and actual values; relative because the mean difference is divided by the arithmetic mean.
Relative squared error (RSE) similarly normalizes the total squared error of the predicted values by dividing by the total squared error of the actual values.
Mean Zero One Error (MZOE) indicates whether the prediction was correct or not. In other words: ZeroOneLoss(x,y) = 1 when x!=y; otherwise 0.
Coefficient of determination, often referred to as R2, represents the predictive power of the model as a value between 0 and 1. Zero means the model is random (explains nothing); 1 means there is a perfect fit. However, caution should be used in interpreting R2 values, as low values can be entirely normal and high values can be suspect.
AUC.
References:
https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/evaluate-model

**NEW QUESTION 84**
- (Exam Topic 3)
You use the Azure Machine Learning SDK in a notebook to run an experiment using a script file in an experiment folder.
The experiment fails.
You need to troubleshoot the failed experiment.
What are two possible ways to achieve this goal? Each correct answer presents a complete solution.

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

**Answer:** BD

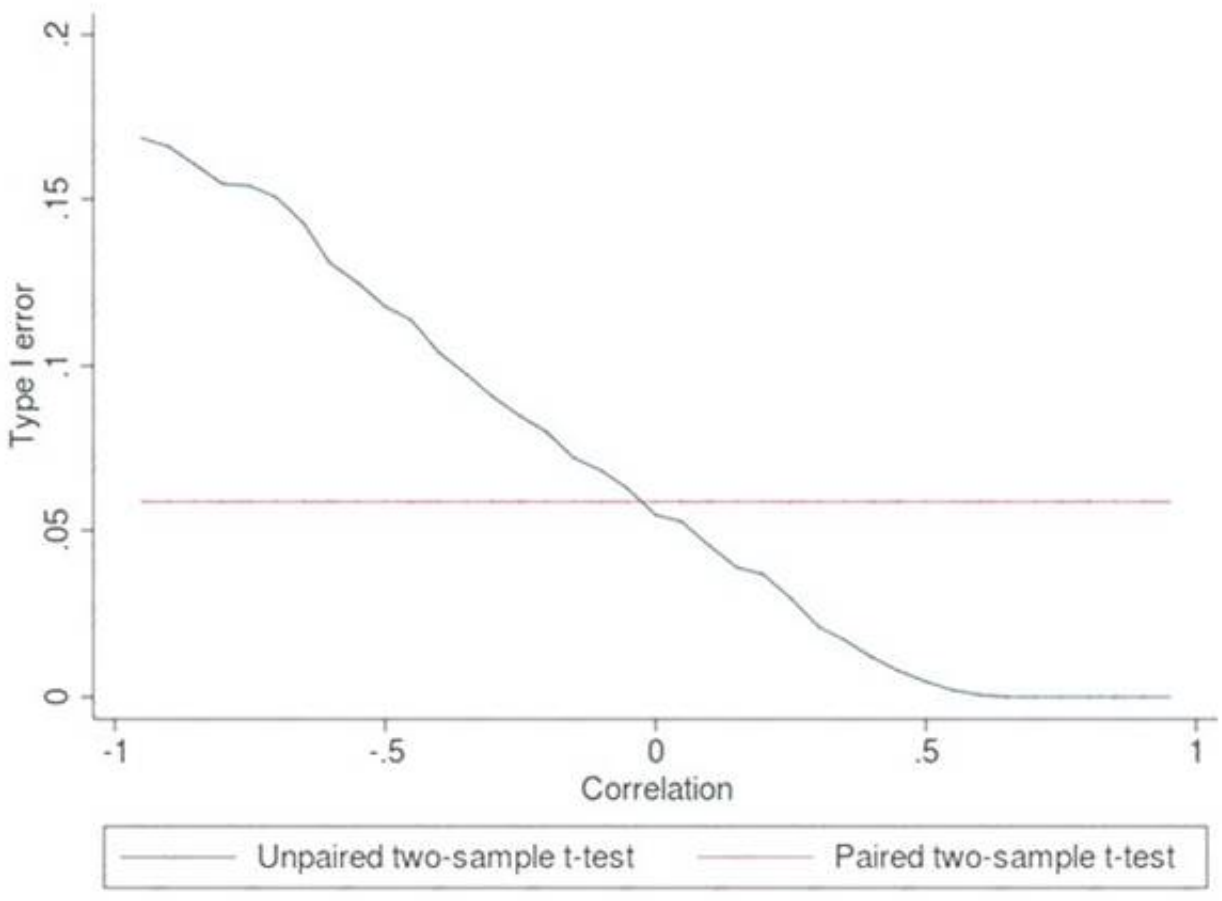
**Explanation:**
Use get_details_with_logs() to fetch the run details and logs created by the run.
You can monitor Azure Machine Learning runs and view their logs with the Azure Machine Learning studio. Reference:
https://docs.microsoft.com/en-us/python/api/azureml-pipeline-core/azureml.pipeline.core.steprun https://docs.microsoft.com/en-us/azure/machine-learning/how-to-monitor-view-training-logs

**NEW QUESTION 87**
- (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 89**
- (Exam Topic 3)
You create an Azure Databricks workspace and a linked Azure Machine Learning workspace. You have the following Python code segment in the Azure Machine Learning workspace: import mlflow
import mlflow.azureml import azureml.mlflow import azureml.core
from azureml.core import Workspace subscription_id = 'subscription_id' resourse_group = 'resource_group_name' workspace_name = 'workspace_name'
ws = Workspace.get(name=workspace_name, subscription_id=subscription_id, resource_group=resource_group)
experimentName = "/Users/{user_name}/{experiment_folder}/{experiment_name}" mlflow.set_experiment(experimentName)
uri = ws.get_mlflow_tracking_uri() mlflow.set_tracking_uri(uri)
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.

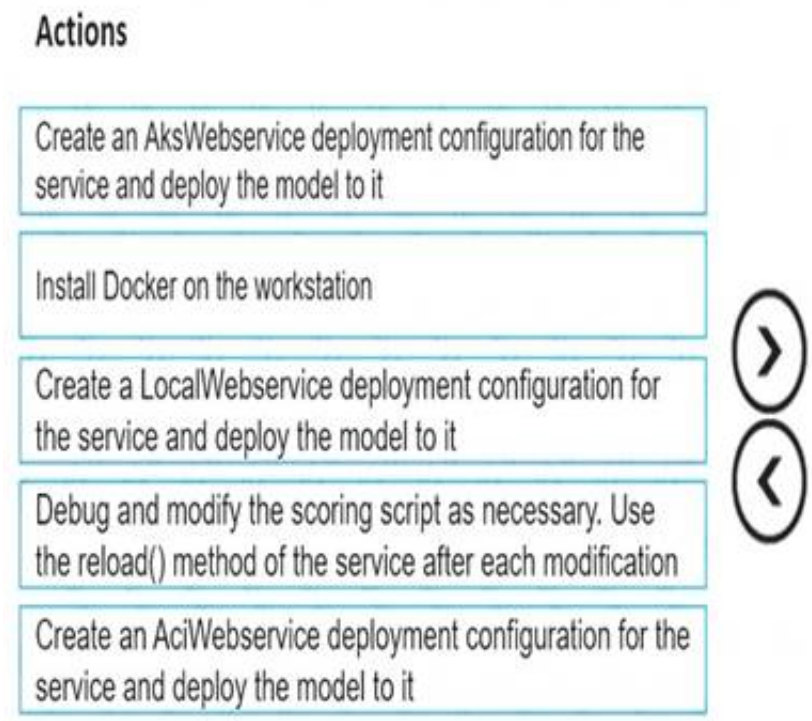
| | Yes | No |
|---|---|---|
| A resource group and Azure Machine Learning workspace will be created. | ○ | ○ |
| An Azure Databricks experiment will be tracked only in the Azure Machine Learning workspace. | ○ | ○ |
| The epoch loss metric is set to be tracked. | ○ | ○ |

A. Mastered
B. Not Mastered

**Answer:** A

**Explanation:**
A screenshot of a computer Description automatically generated with medium confidence
Box 1: No
The Workspace.get method loads an existing workspace without using configuration files. ws = Workspace.get(name="myworkspace", subscription_id='<azure-subscription-id>', resource_group='myresourcegroup')
Box 2: Yes
MLflow Tracking with Azure Machine Learning lets you store the logged metrics and artifacts from your local runs into your Azure Machine Learning workspace. The get_mlflow_tracking_uri() method assigns a unique tracking URI address to the workspace, ws, and set_tracking_uri() points the MLflow tracking URI to that address.
Box 3: Yes
Note: In Deep Learning, epoch means the total dataset is passed forward and backward in a neural network once.
Reference:
https://docs.microsoft.com/en-us/python/api/azureml-core/azureml.core.workspace.workspace https://docs.microsoft.com/en-us/azure/machine-learning/how-to-use-mlflow

**NEW QUESTION 91**
- (Exam Topic 3)
You train and register a model by using the Azure Machine Learning SDK on a local workstation. Python 3.6 and Visual Studio Code are installed on the workstation.
When you try to deploy the model into production as an Azure Kubernetes Service (AKS)-based web service, you experience an error in the scoring script that causes deployment to fail.
You need to debug the service on the local workstation before deploying the service to production.
Which four actions should you perform in sequence? To answer, move the appropriate actions from the list of actions to the answer area and arrange them in the correct order.



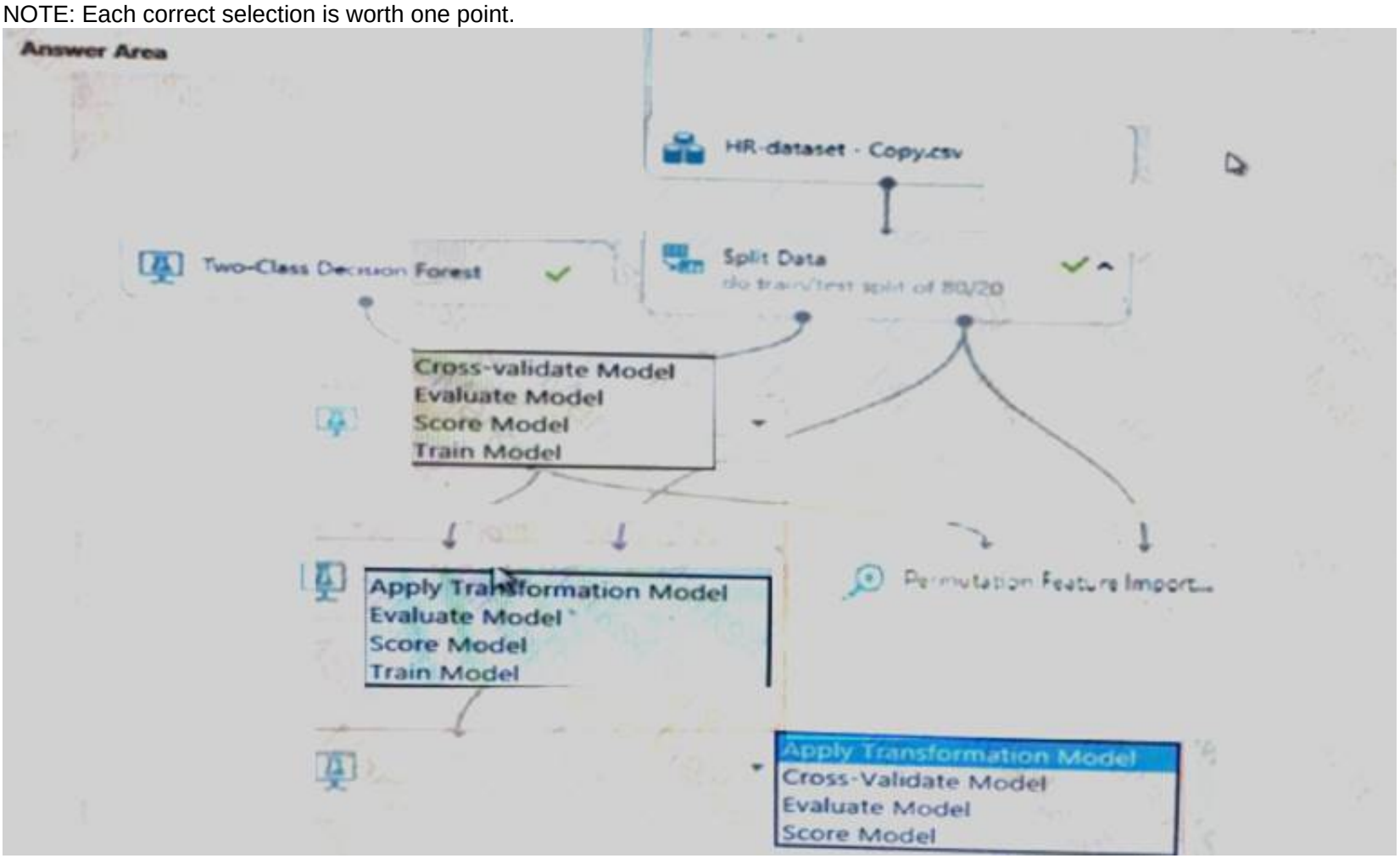

A. Mastered
B. Not Mastered

**Answer:** A

**Explanation:**
Graphical user interface, text, application, email Description automatically generated
Step 1: Install Docker on the workstation
Prerequisites include having a working Docker installation on your local system. Build or download the dockerfile to the compute node.
Step 2: Create an AksWebservice deployment configuration and deploy the model to it
To deploy a model to Azure Kubernetes Service, create a deployment configuration that describes the compute resources needed.
# If deploying to a cluster configured for dev/test, ensure that it was created with enough
# cores and memory to handle this deployment configuration. Note that memory is also used by
# things such as dependencies and AML components.

deployment_config = AksWebservice.deploy_configuration(cpu_cores = 1, memory_gb = 1)
service = Model.deploy(ws, "myservice", [model], inference_config, deployment_config, aks_target) service.wait_for_deployment(show_output = True)
print(service.state) print(service.get_logs())
Step 3: Create a LocalWebservice deployment configuration for the service and deploy the model to it
To deploy locally, modify your code to use LocalWebservice.deploy_configuration() to create a deployment configuration. Then use Model.deploy() to deploy the service.
Step 4: Debug and modify the scoring script as necessary. Use the reload() method of the service after each modification.
During local testing, you may need to update the score.py file to add logging or attempt to resolve any problems that you've discovered. To reload changes to the score.py file, use reload(). For example, the following code reloads the script for the service, and then sends data to it.
Reference:
https://docs.microsoft.com/en-us/azure/machine-learning/how-to-deploy-azure-kubernetes-service https://docs.microsoft.com/en-us/azure/machine-learning/how-to-troubleshoot-deployment-local

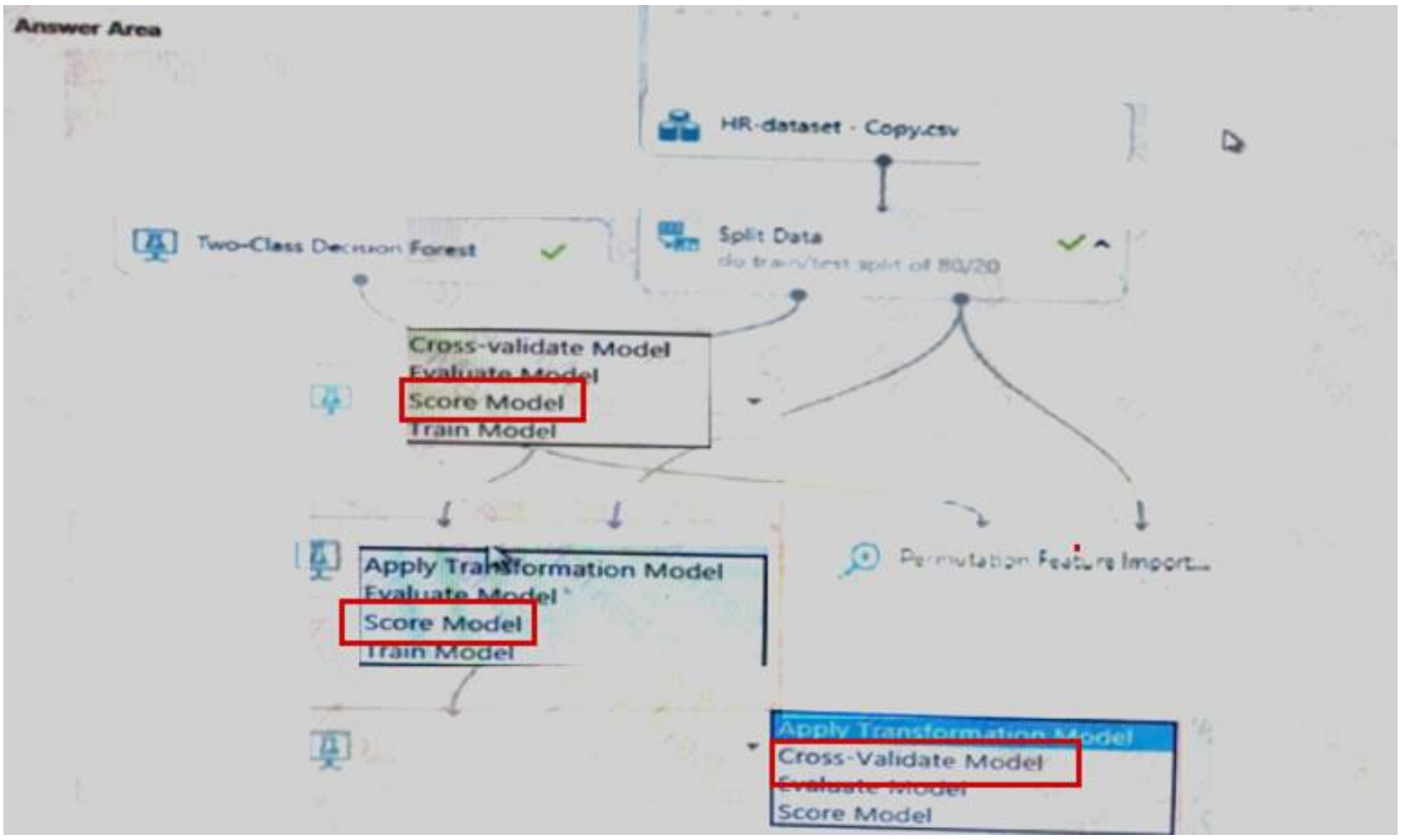
**NEW QUESTION 93**
- (Exam Topic 3)
You create a binary classification model using Azure Machine Learning Studio.
You must use a Receiver Operating Characteristic (RO C) curve and an F1 score to evaluate the model. You need to create the required business metrics.
How should you complete the experiment? To answer, select the appropriate options in the dialog box in the answer area.
NOTE: Each correct selection is worth one point.




A. Mastered
B. Not Mastered

**Answer:** A

**Explanation:**




**NEW QUESTION 95**
- (Exam Topic 3)
You deploy a model as an Azure Machine Learning real-time web service using the following code.

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

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

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

**Answer:** A

**Explanation:**
You can print out detailed Docker engine log messages from the service object. You can view the log for ACI, AKS, and Local deployments. The following example demonstrates how to print the logs.
# if you already have the service object handy print(service.get_logs())
# if you only know the name of the service (note there might be multiple services with the same name but different version number)
print(ws.webservices['mysvc'].get_logs()) Reference:
https://docs.microsoft.com/en-us/azure/machine-learning/how-to-troubleshoot-deployment

**NEW QUESTION 99**
- (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 104**
- (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 105**
- (Exam Topic 3)
Note: This question is part of a series of questions that present the same scenario. Each question in the series contains a unique solution that might meet the stated goals. Some question sets might have more than one correct solution, while others might not have a correct solution.
After you answer a question in this section, you will NOT be able to return to it. As a result, these questions will not appear in the review screen.
You create a model to forecast weather conditions based on historical data.
You need to create a pipeline that runs a processing script to load data from a datastore and pass the processed data to a machine learning model training script.
Solution: Run the following code:

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

Does the solution meet the goal?

A. Yes
B. No

**Answer:** A

**Explanation:**
The two steps are present: process_step and train_step Note:
Data used in pipeline can be produced by one step and consumed in another step by providing a PipelineData object as an output of one step and an input of one or more subsequent steps.
PipelineData objects are also used when constructing Pipelines to describe step dependencies. To specify that a step requires the output of another step as input, use a PipelineData object in the constructor of both steps.
For example, the pipeline train step depends on the process_step_output output of the pipeline process step: from azureml.pipeline.core import Pipeline, PipelineData
from azureml.pipeline.steps import PythonScriptStep datastore = ws.get_default_datastore()
process_step_output = PipelineData("processed_data", datastore=datastore) process_step = PythonScriptStep(script_name="process.py", arguments=["--data_for_train", process_step_output], outputs=[process_step_output], compute_target=aml_compute, source_directory=process_directory)
train_step = PythonScriptStep(script_name="train.py", arguments=["--data_for_train", process_step_output], inputs=[process_step_output], compute_target=aml_compute, source_directory=train_directory)
pipeline = Pipeline(workspace=ws, steps=[process_step, train_step]) Reference:
https://docs.microsoft.com/en-us/python/api/azureml-pipeline-core/azureml.pipeline.core.pipelinedata?view=azu

**NEW QUESTION 109**
- (Exam Topic 3)
Note: This question is part of a series of questions that present the same scenario. Each question in the series contains a unique solution that might meet the stated goals. Some question sets might have more than one correct solution, while others might not have a correct solution.
After you answer a question in this section, you will NOT be able to return to it. As a result, these questions will not appear in the review screen.
An IT department creates the following Azure resource groups and resources:

| Resource group | Resources |
|---|---|
| ml_resources | • an Azure Machine Learning workspace named amlworkspace<br>• an Azure Storage account named amlworkspace12345<br>• an Application Insights instance named amlworkspace54321<br>• an Azure Key Vault named amlworkspace67890<br>• an Azure Container Registry named amlworkspace09876 |
| general_compute | A virtual machine named mlvm with the following configuration:<br>• Operating system: Ubuntu Linux<br>• Software installed: Python 3.6 and Jupyter Notebooks<br>• Size: NC6 (6 vCPUs, 1 vGPU, 56 Gb RAM) |

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

A. Yes
B. No

**Answer:** A

**Explanation:**
Use the VM as a compute target.
Note: A compute target is a designated compute resource/environment where you run your training script or host your service deployment. This location may be your local machine or a cloud-based compute resource.
Reference:
https://docs.microsoft.com/en-us/azure/machine-learning/concept-compute-target

**NEW QUESTION 111**
- (Exam Topic 3)
Note: This question is part of a series of questions that present the same scenario. Each question in the series contains a unique solution that might meet the stated goals. Some question sets might have more than one correct solution, while others might not have a correct solution.
After you answer a question in this section, you will NOT be able to return to it. As a result, these questions will not appear in the review screen.
You 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 116**
- (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 120**
- (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 124**
- (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 128**
- (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 131**
- (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.

**Answer Area**

| REST Request | Value |
|---|---|
| Request Header | JSON containing the run ID<br>JSON containing the pipeline ID<br>JSON containing the experiment name<br>JSON containing an OAuth bearer token |
| Response | JSON containing the run ID<br>JSON containing the pipeline ID<br>JSON containing the experiment name<br>JSON containing an OAuth bearer token |
| Response | JSON containing the run ID<br>JSON containing a list of predictions<br>JSON containing the experiment name<br>JSON containing a path to the parallel_run_step.txt output file |

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 133**
- (Exam Topic 3)
Note: This question is part of a series of questions that present the same scenario. Each question in the series contains a unique solution that might meet the stated goals. Some question sets might have more than one correct solution, while others might not have a correct solution.
After you answer a question in this section, you will NOT be able to return to it. As a result, these questions will not appear in the review screen.
You train and register a machine learning model.
You plan to deploy the model as a real-time web service. Applications must use key-based authentication to use the model.
You need to deploy the web service.
Solution:
Create an AksWebservice instance.
Set the value of the auth_enabled property to False.
Set the value of the token_auth_enabled property to True.
Deploy the model to the service. Does the solution meet the goal?

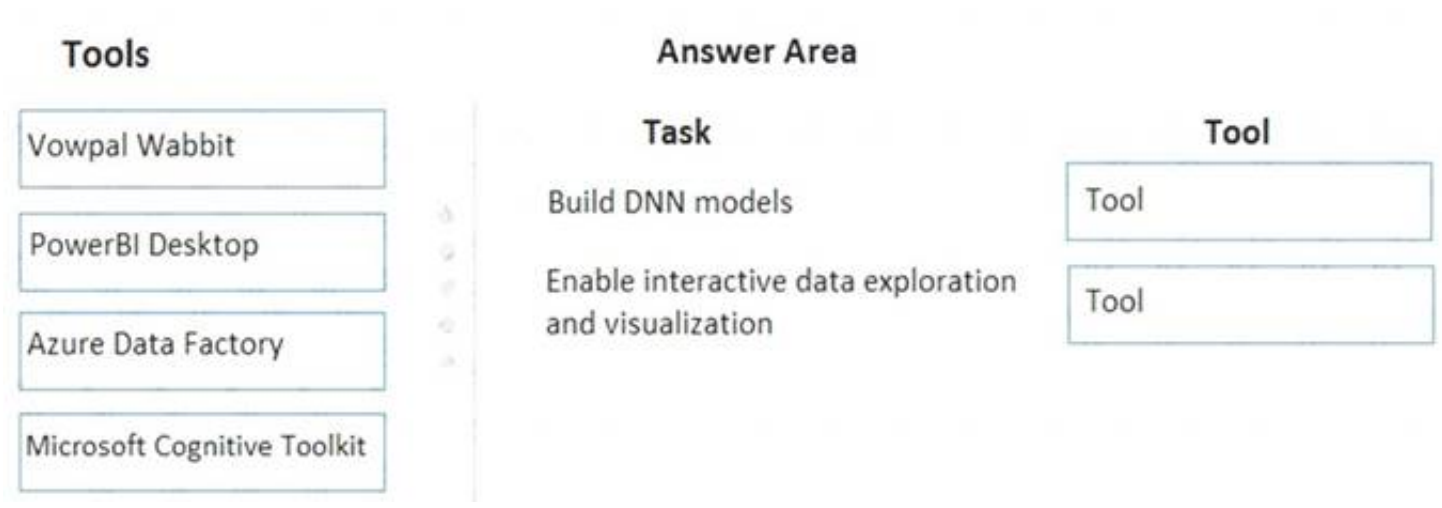
A. Yes
B. No

**Answer:** B

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

**NEW QUESTION 136**
- (Exam Topic 3)
You configure a Deep Learning Virtual Machine for Windows.
You need to recommend tools and frameworks to perform the following:
- Build deep neural network (DNN) models
- Perform interactive data exploration and visualization

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




A. Mastered
B. Not Mastered

**Answer:** A

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

**NEW QUESTION 138**
- (Exam Topic 3)
You are creating a machine learning model in Python. The provided dataset contains several numerical columns and one text column. The text column represents a product's category. The product category will always be one of the following:

- Bikes
- Cars
- Vans
- Boats

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


A. Mastered
B. Not Mastered

**Answer:** A

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

**NEW QUESTION 140**
- (Exam Topic 3)
You are a data scientist building a deep convolutional neural network (CNN) for image classification. The CNN model you built shows signs of overfitting.
You need to reduce overfitting and converge the model to an optimal fit.
Which two actions should you perform? Each correct answer presents a complete solution.
NOTE: Each correct selection is worth one point.

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

**Answer:** AC

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

**NEW QUESTION 145**
- (Exam Topic 3)
Note: This question is part of a series of questions that present the same scenario. Each question in the series contains a unique solution that might meet the stated goals. Some question sets might have more than one correct solution, while others might not have a correct solution.
After you answer a question in this section, you will NOT be able to return to it. As a result, these questions will not appear in the review screen.
You are a data scientist using Azure Machine Learning Studio.
You need to normalize values to produce an output column into bins to predict a target column. Solution: Apply an Equal Width with Custom Start and Stop binning mode.
Does the solution meet the goal?

A. Yes
B. No

**Answer:** B

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

**NEW QUESTION 146**
- (Exam Topic 3)
You are a data scientist working for a bank and have used Azure ML to train and register a machine learning model that predicts whether a customer is likely to repay a loan.
You want to understand how your model is making selections and must be sure that the model does not violate government regulations such as denying loans based on where an applicant lives.
You need to determine the extent to which each feature in the customer data is influencing predictions. What should you do?

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

**Answer:** D

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

**NEW QUESTION 147**
- (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

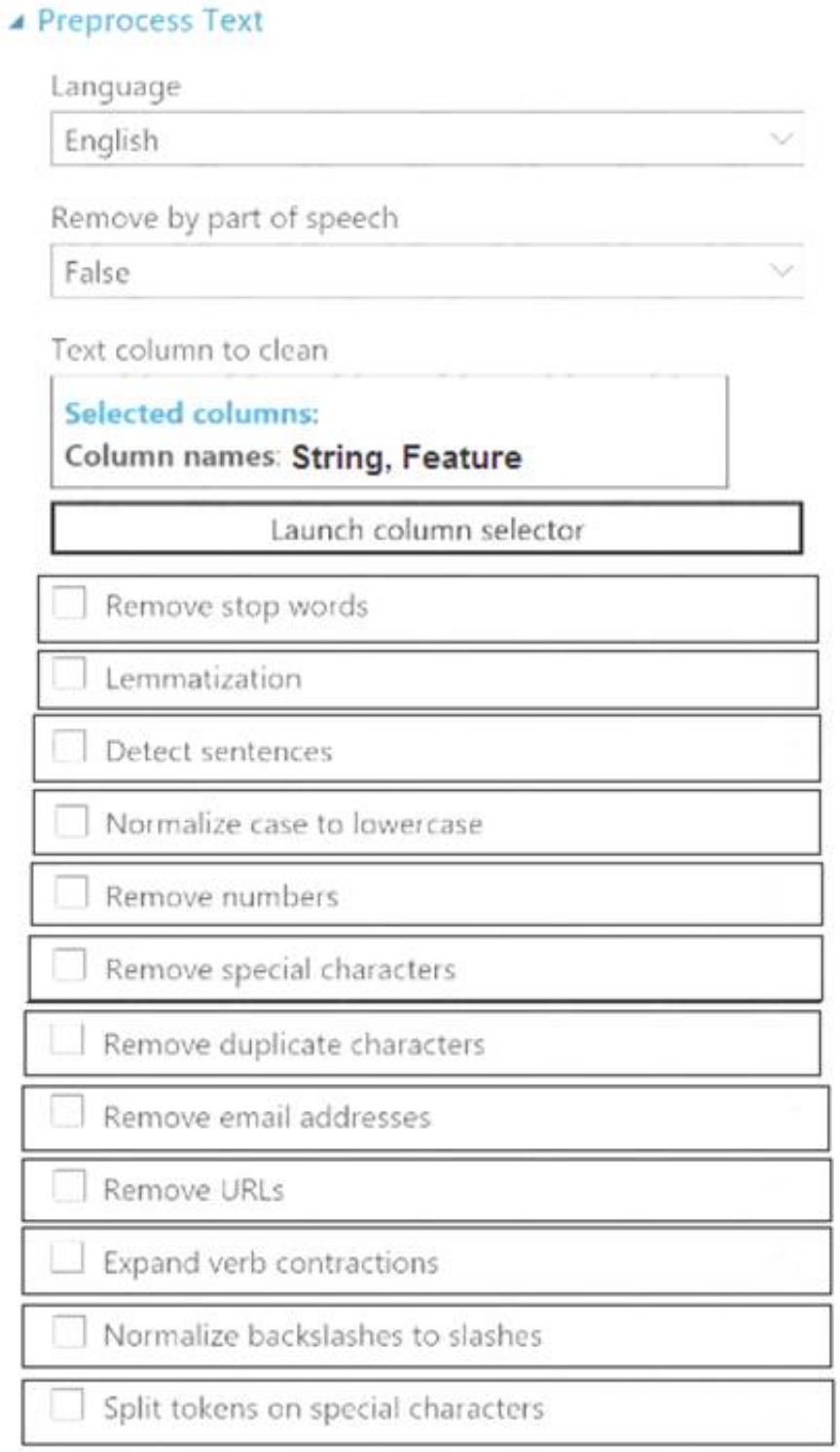
**Answer:** A

**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 148**
- (Exam Topic 3)
You plan to preprocess text from CSV files. You load the Azure Machine Learning Studio default stop words list.
You need to configure the Preprocess Text module to meet the following requirements:

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

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




A. Mastered
B. Not Mastered

**Answer:** A

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

**NEW QUESTION 149**
- (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?

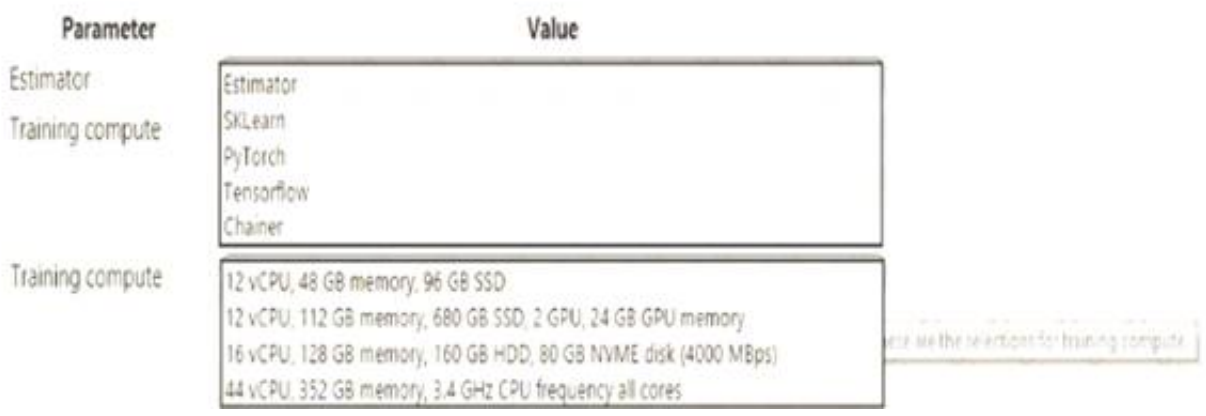
A. Split Data
B. Load Trained Model
C. Assign Data to Clusters
D. Group Data into Bins

**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 152**
- (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




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 155**
- (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.

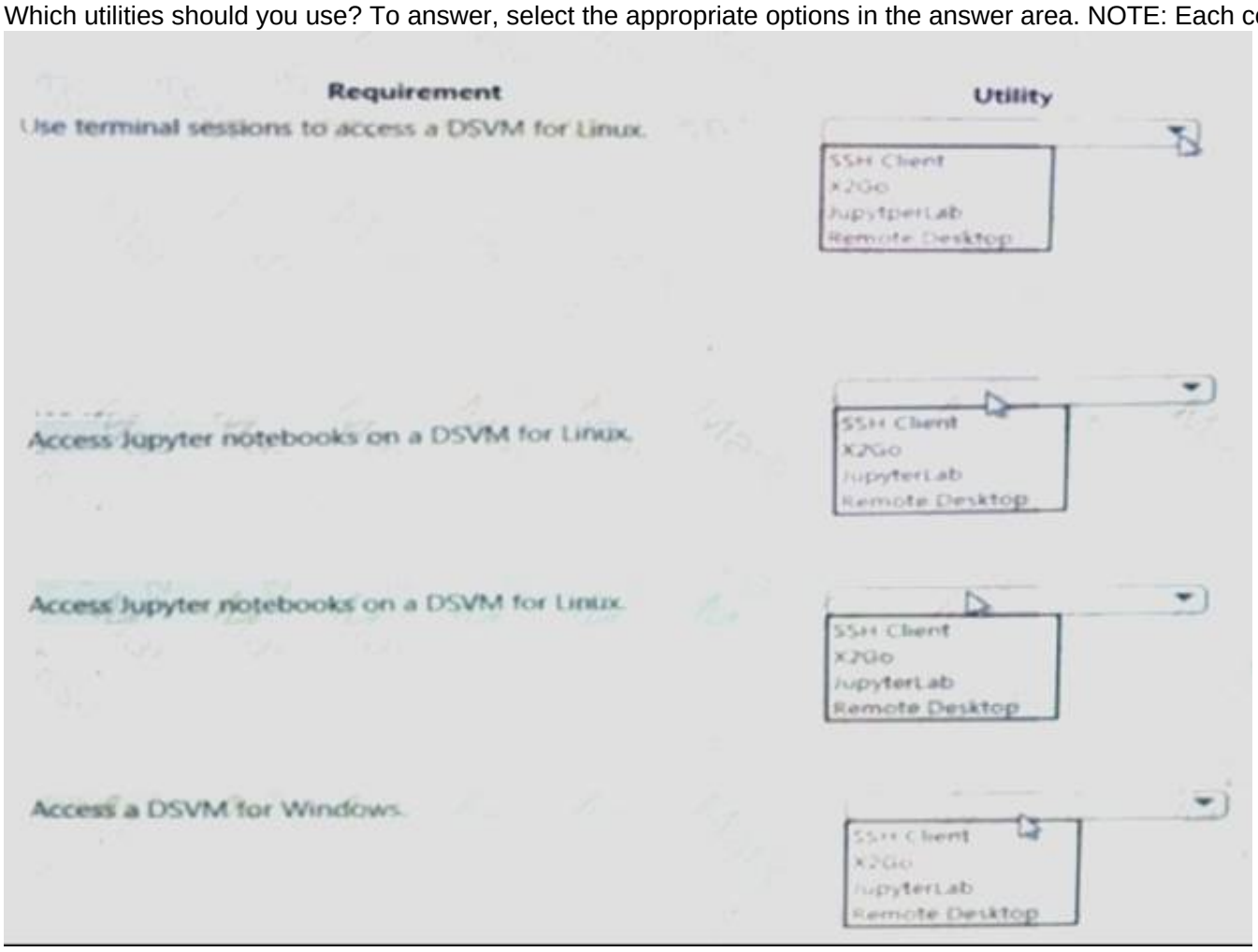
| Statements | Yes | No |
|---|---|---|
| The user can perform all actions in the workspace | ○ | ○ |
| The user can delete a compute resource in the workspace | ○ | ○ |
| The user can write metrics to the workspace | ○ | ○ |

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 159**
- (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 163**
- (Exam Topic 3)
Note: This question is part of a series of questions that present the same scenario. Each question in the series contains a unique solution that might meet the stated goals. Some question sets might have more than one correct solution, while others might not have a correct solution.
After you answer a question in this section, you will NOT be able to return to it. As a result, these questions will not appear in the review screen.
You are using Azure Machine Learning Studio to perform feature engineering on a dataset. You need to normalize values to produce a feature column grouped into bins.
Solution: Apply an Entropy Minimum Description Length (MDL) binning mode. Does the solution meet the goal?

A. Yes
B. No

**Answer:** A

**Explanation:**
Entropy MDL binning mode: This method requires that you select the column you want to predict and the column or columns that you want to group into bins. It then makes a pass over the data and attempts to determine the number of bins that minimizes the entropy. In other words, it chooses a number of bins that allows the data column to best predict the target column. It then returns the bin number associated with each row of your data in a column named <colname>quantized.
References:
https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/group-data-into-bins

**NEW QUESTION 165**
- (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 169**
- (Exam Topic 3)
You are performing a filter based feature selection for a dataset 10 build a multi class classifies by using Azure Machine Learning Studio.
The dataset contains categorical features that are highly correlated to the output label column.
You need to select the appropriate feature scoring statistical method to identify the key predictors. Which method should you use?

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

**Answer:** D

**Explanation:**
Pearson's correlation statistic, or Pearson's correlation coefficient, is also known in statistical models as the r value. For any two variables, it returns a value that indicates the strength of the correlation
Pearson's correlation coefficient is the test statistics that measures the statistical relationship, or association, between two continuous variables. It is known as the best method of measuring the association between variables of interest because it is based on the method of covariance. It gives information about the magnitude of the association, or correlation, as well as the direction of the relationship.
Reference:
https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/filter-based-feature-selection https://www.statisticssolutions.com/pearsons-correlation-coefficient/

**NEW QUESTION 174**
- (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 175**
- (Exam Topic 3)
Note: This question is part of a series of questions that present the same scenario. Each question in the series contains a unique solution that might meet the stated goals. Some question sets might have more than one correct solution, while others might not have a correct solution.
After you answer a question in this section, you will NOT be able to return to it. As a result, these questions will not appear in the review screen.
You are analyzing a numerical dataset which contain missing values in several columns.
You must clean the missing values using an appropriate operation without affecting the dimensionality of the feature set.
You need to analyze a full dataset to include all values.
Solution: Use the last Observation Carried Forward (IOCF) method to impute the missing data points. Does the solution meet the goal?

A. Yes
B. No

**Answer:** B

**Explanation:**
Instead use the Multiple Imputation by Chained Equations (MICE) method.
Replace using MICE: For each missing value, this option assigns a new value, which is calculated by using a method described in the statistical literature as "Multivariate Imputation using Chained Equations" or "Multiple Imputation by Chained Equations". With a multiple imputation method, each variable with missing data is modeled conditionally using the other variables in the data before filling in the missing values.
Note: Last observation carried forward (LOCF) is a method of imputing missing data in longitudinal studies. If a person drops out of a study before it ends, then his or her last observed score on the dependent variable is used for all subsequent (i.e., missing) observation points. LOCF is used to maintain the sample size and to reduce the bias caused by the attrition of participants in a study.
References:
https://methods.sagepub.com/reference/encyc-of-research-design/n211.xml https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3074241/

**NEW QUESTION 178**
- (Exam Topic 3)
You run an automated machine learning experiment in an Azure Machine Learning workspace. Information about the run is listed in the table below:

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

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

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

B)

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

C)

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

D)

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

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

**Answer:** A

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

**NEW QUESTION 180**
- (Exam Topic 3)
You create an Azure Machine Learning compute resource to train models. The compute resource is configured as follows:

- Minimum nodes: 2
- Maximum nodes: 4

You must decrease the minimum number of nodes and increase the maximum number of nodes to the following values:

- Minimum nodes: 0
- Maximum nodes: 8

You need to reconfigure the compute resource.
What are three possible ways to achieve this goal? Each correct answer presents a complete solution.
NOTE: Each correct selection is worth one point.

A. Use the Azure Machine Learning studio.
B. Run the update method of the AmlCompute class in the Python SDK.
C. Use the Azure portal.
D. Use the Azure Machine Learning designer.
E. Run the refresh_state() method of the BatchCompute class in the Python SDK

**Answer:** ABC

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

**NEW QUESTION 183**
- (Exam Topic 3)
You are a data scientist creating a linear regression model.
You need to determine how closely the data fits the regression line. Which metric should you review?

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

**Answer:** A

**Explanation:**
Coefficient of determination, often referred to as R2, represents the predictive power of the model as a value between 0 and 1. Zero means the model is random (explains nothing); 1 means there is a perfect fit. However, caution should be used in interpreting R2 values, as low values can be entirely normal and high values can be suspect.
References:
https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/evaluate-model

**NEW QUESTION 186**
- (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

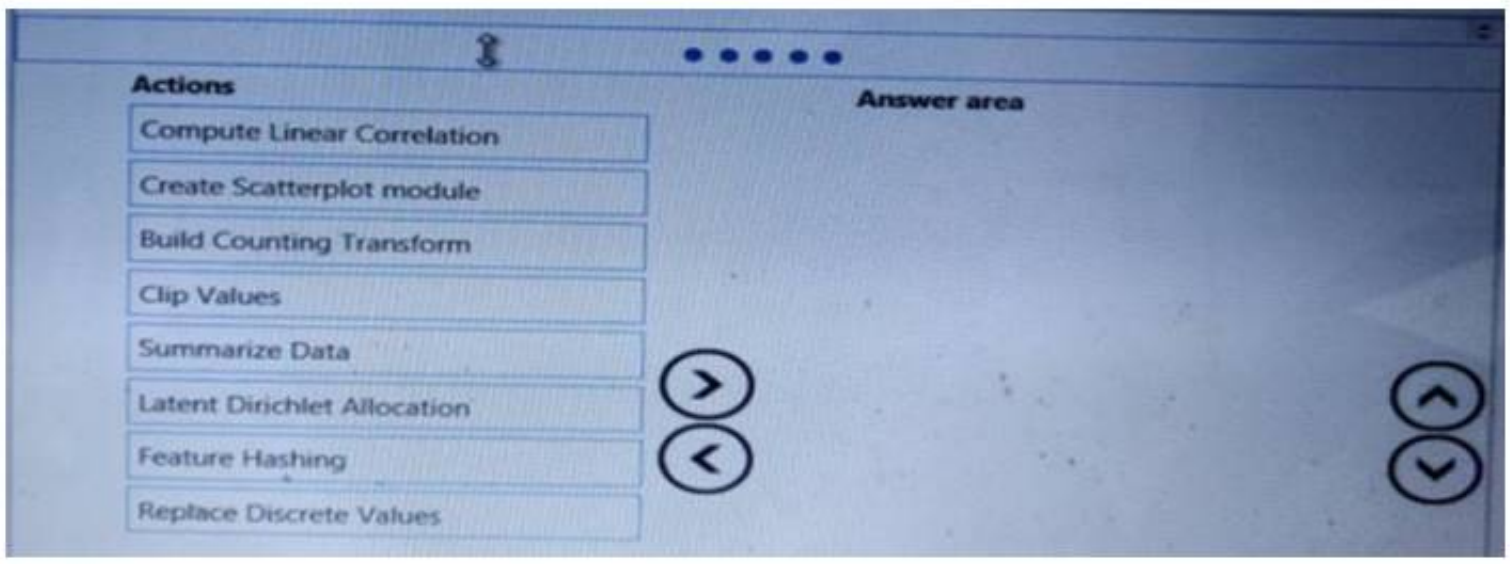
**NEW QUESTION 189**
- (Exam Topic 3)
You register a model that you plan to use in a batch inference pipeline.
The batch inference pipeline must use a ParallelRunStep step to process files in a file dataset. The script has the ParallelRunStep step runs must process six input files each time the inferencing function is called.
You need to configure the pipeline.
Which configuration setting should you specify in the ParallelRunConfig object for the PrallelRunStep step?

A. process_count_per_node= "6"
B. node_count= "6"
C. mini_batch_size= "6"
D. error_threshold= "6"

**Answer:** B

**Explanation:**
node_count is the number of nodes in the compute target used for running the ParallelRunStep. Reference:
https://docs.microsoft.com/en-us/python/api/azureml-contrib-pipeline-steps/azureml.contrib.pipeline.steps.parall

**NEW QUESTION 194**
- (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.

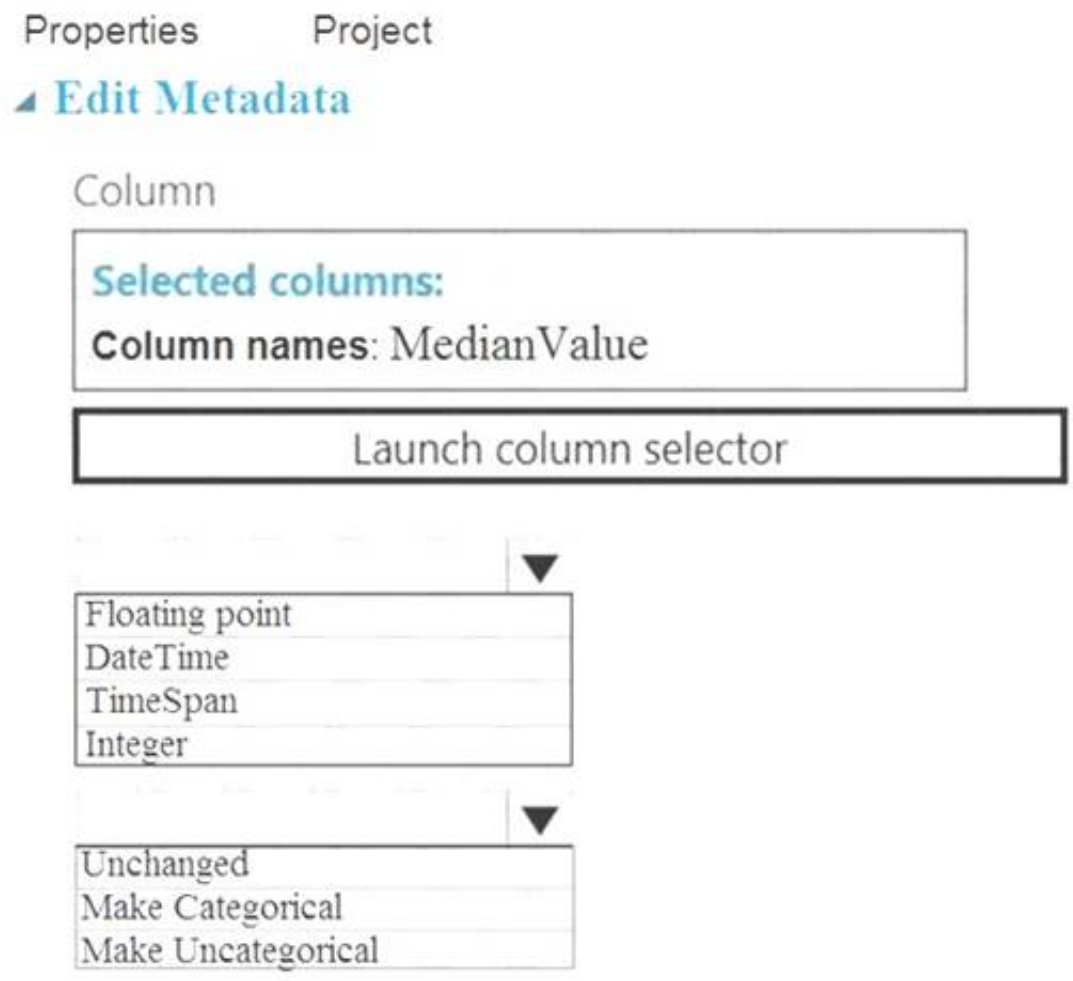



A. Mastered
B. Not Mastered

**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 195**
- (Exam Topic 2)
You need to configure the Edit Metadata module so that the structure of the datasets match.
Which configuration options should you select? To answer, select the appropriate options in the answer area.
NOTE: Each correct selection is worth one point.




A. Mastered
B. Not Mastered

**Answer:** A

**Explanation:**
Box 1: Floating point
Need floating point for Median values.
Scenario: An initial investigation shows that the datasets are identical in structure apart from the MedianValue column. The smaller Paris dataset contains the MedianValue in text format, whereas the larger London dataset contains the MedianValue in numerical format.
Box 2: Unchanged
Note: Select the Categorical option to specify that the values in the selected columns should be treated as categories.
For example, you might have a column that contains the numbers 0,1 and 2, but know that the numbers actually mean "Smoker", "Non smoker" and "Unknown". In that case, by flagging the column as categorical you can ensure that the values are not used in numeric calculations, only to group data.

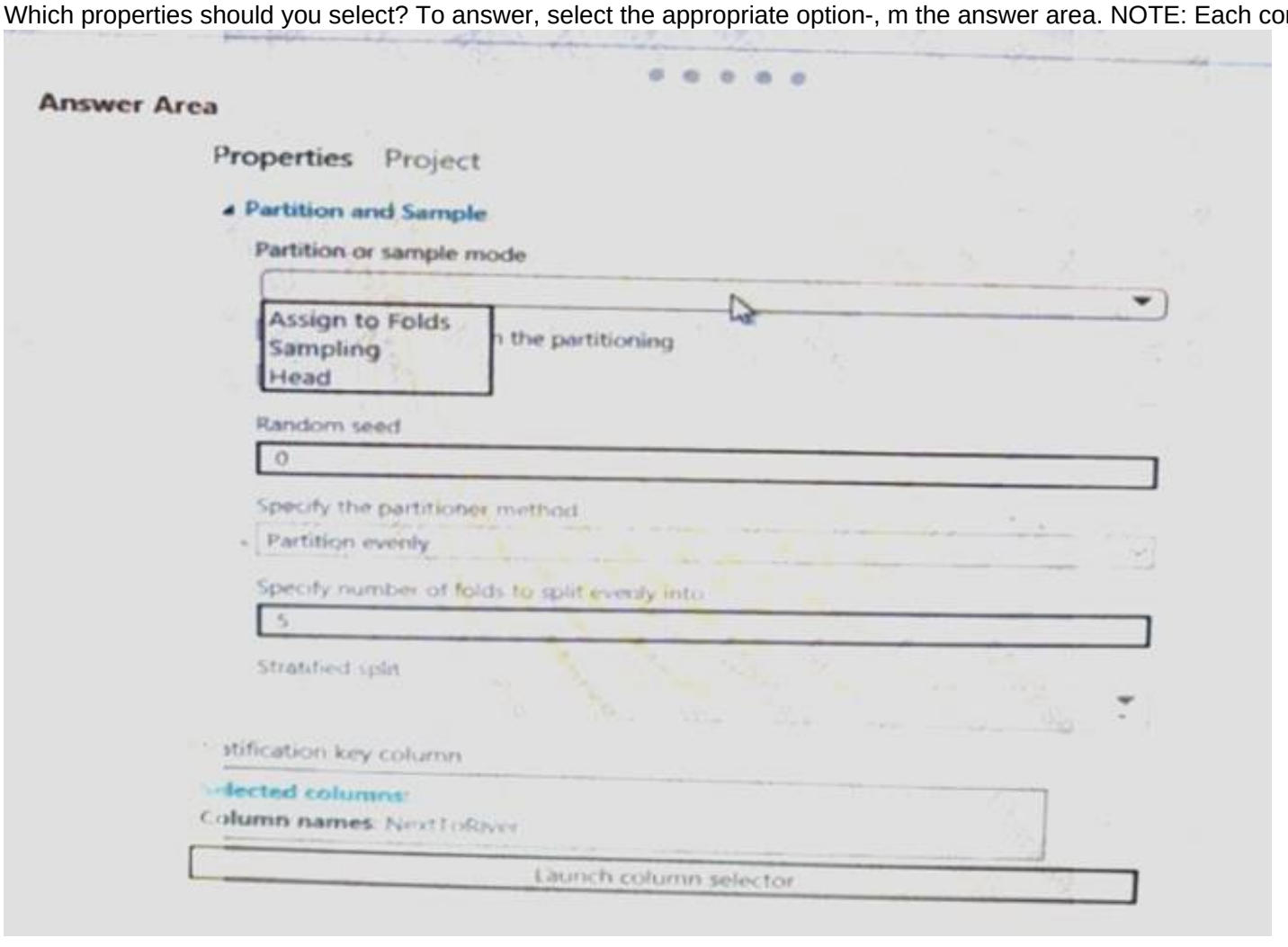
**NEW QUESTION 196**
- (Exam Topic 2)
You need to select a feature extraction method. Which method should you use?

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

**Answer:** C

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

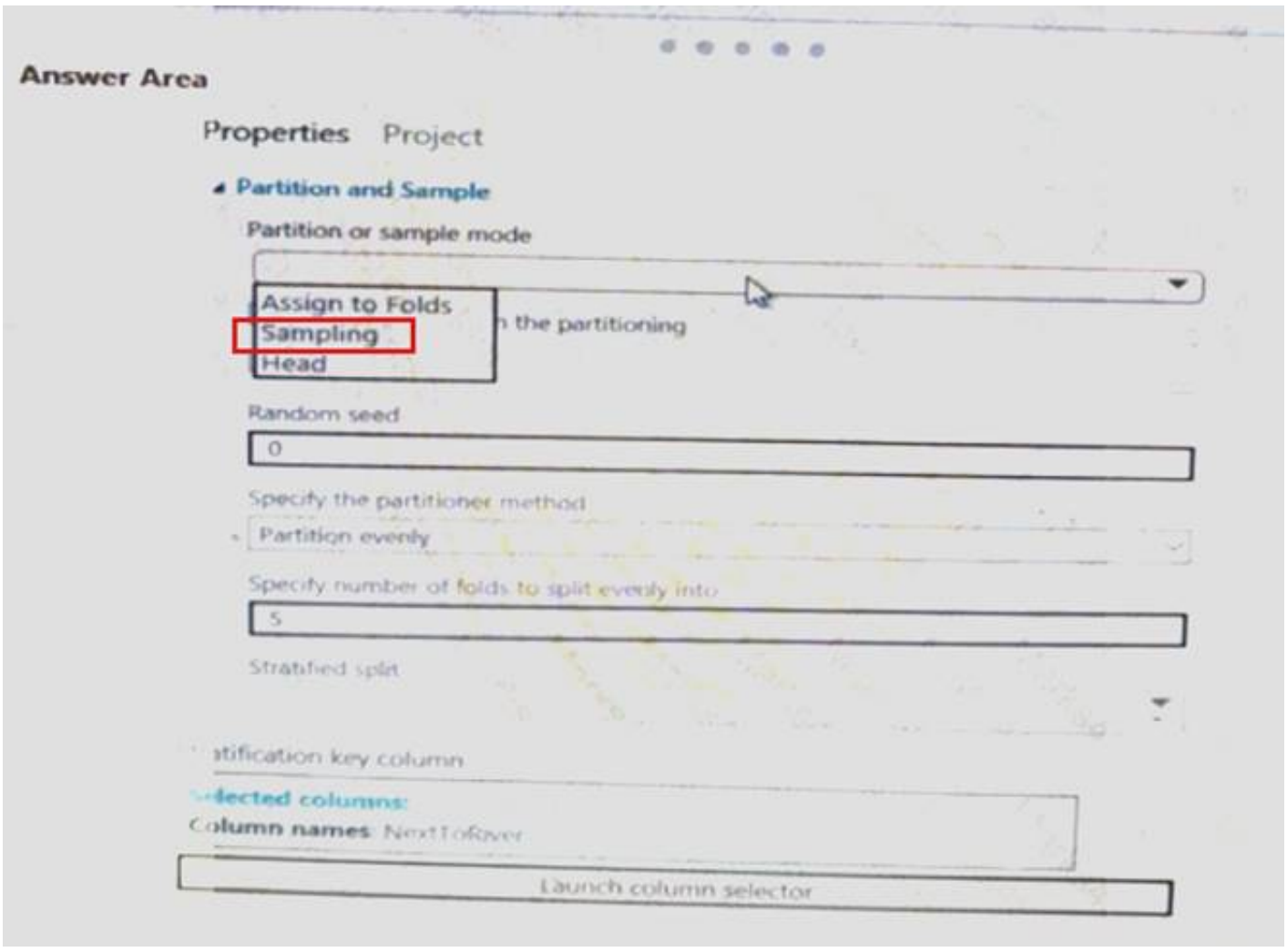
**NEW QUESTION 201**
- (Exam Topic 2)
You need to identify the methods for dividing the data according, to the testing requirements.
Which properties should you select? To answer, select the appropriate option-, m the answer area. NOTE: Each correct selection is worth one point.




A. Mastered
B. Not Mastered

**Answer:** A

**Explanation:**




**NEW QUESTION 204**
- (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.

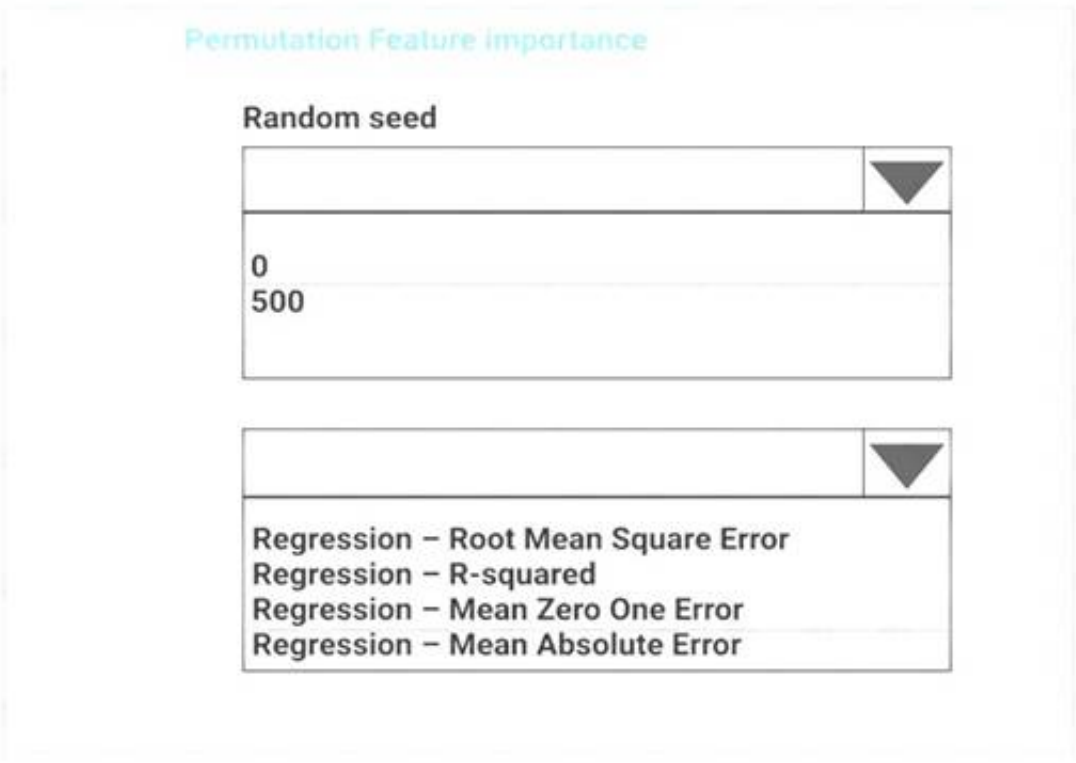

A. Mastered
B. Not Mastered

**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 209**
- (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




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

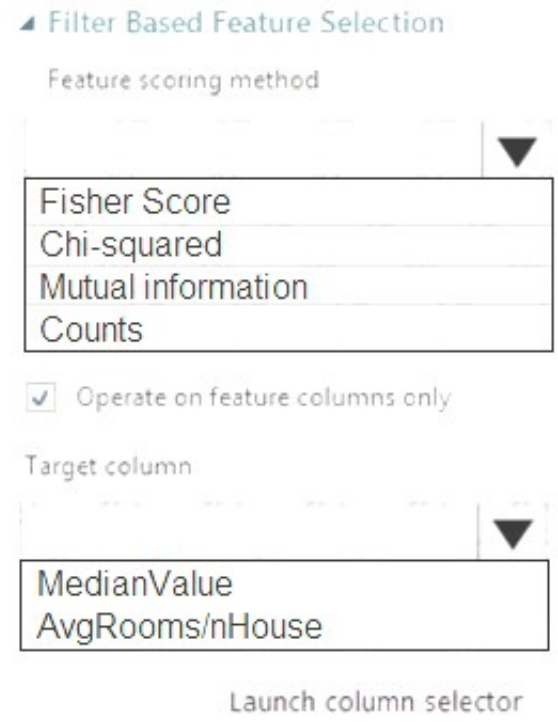
**NEW QUESTION 210**
- (Exam Topic 1)
You need to implement a feature engineering strategy for the crowd sentiment local models. What should you do?

A. Apply an analysis of variance (ANOVA).
B. Apply a Pearson correlation coefficient.
C. Apply a Spearman correlation coefficient.
D. Apply a linear discriminant analysis.

**Answer:** D

**Explanation:**
The linear discriminant analysis method works only on continuous variables, not categorical or ordinal variables.
Linear discriminant analysis is similar to analysis of variance (ANOVA) in that it works by comparing the means of the variables.
Scenario:
Data scientists must build notebooks in a local environment using automatic feature engineering and model building in machine learning pipelines.
Experiments for local crowd sentiment models must combine local penalty detection data. All shared features for local models are continuous variables.

**NEW QUESTION 213**
- (Exam Topic 2)
You need to configure the Feature Based Feature Selection module based on the experiment requirements and datasets.
How should you configure the module properties? To answer, select the appropriate options in the dialog box in the answer area.
NOTE: Each correct selection is worth one point.




A. Mastered
B. Not Mastered

**Answer:** A

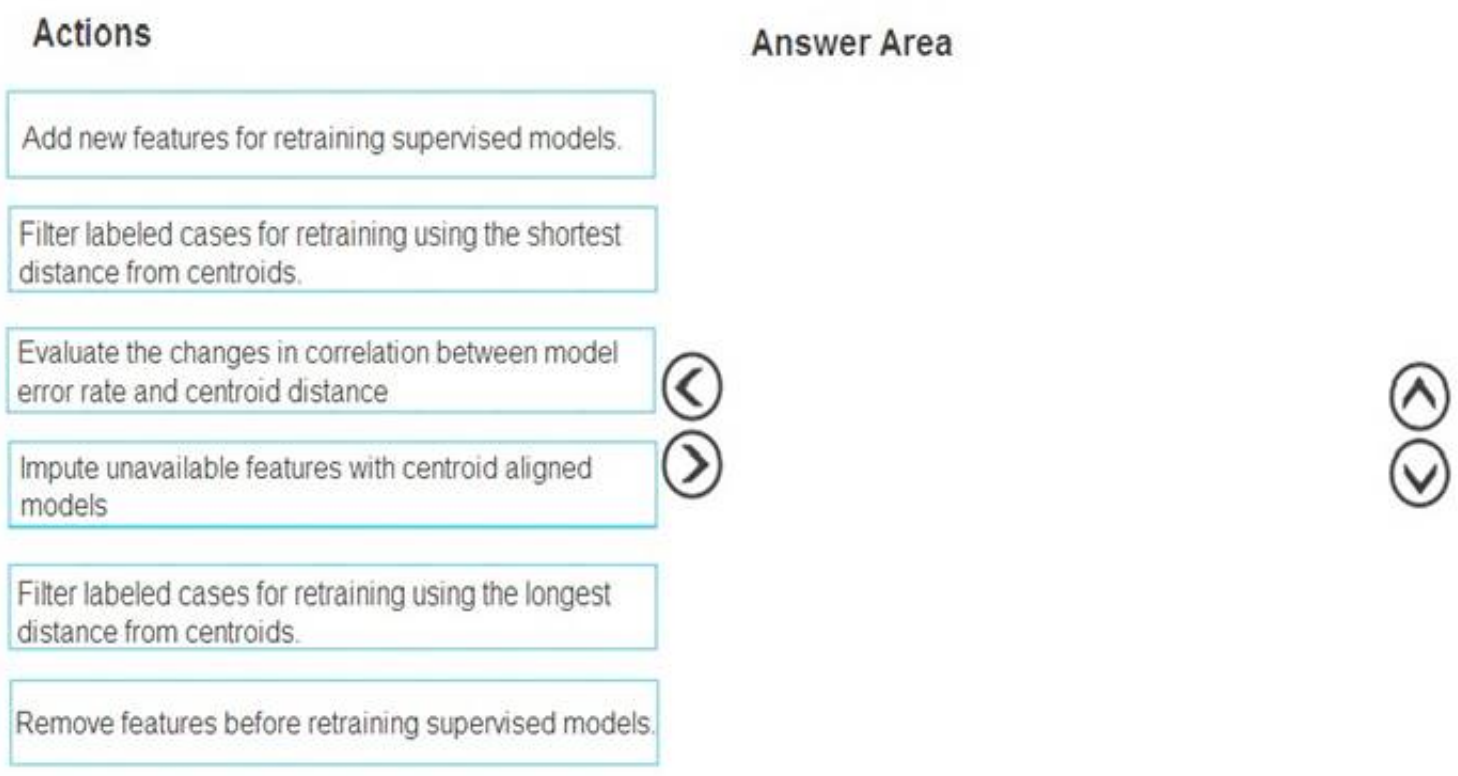
**Explanation:**
Box 1: Mutual Information.
The mutual information score is particularly useful in feature selection because it maximizes the mutual information between the joint distribution and target variables in datasets with many dimensions.
Box 2: MedianValue
MedianValue is the feature column, , it is the predictor of the dataset.
Scenario: The MedianValue and AvgRoomsinHouse columns both hold data in numeric format. You need to select a feature selection algorithm to analyze the relationship between the two columns in more detail.
References:
https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/filter-based-feature-selection

**NEW QUESTION 218**
- (Exam Topic 1)
You need to select an environment that will meet the business and data requirements. Which environment should you use?

A. Azure HDInsight with Spark MLlib
B. Azure Cognitive Services
C. Azure Machine Learning Studio
D. Microsoft Machine Learning Server

**Answer:** D

**NEW QUESTION 223**
- (Exam Topic 1)
You need to define an evaluation strategy for the crowd sentiment models.
Which three actions should you perform in sequence? To answer, move the appropriate actions from the list of actions to the answer area and arrange them in the correct order.




A. Mastered
B. Not Mastered

**Answer:** A

**Explanation:**
Scenario:
Experiments for local crowd sentiment models must combine local penalty detection data.
Crowd sentiment models must identify known sounds such as cheers and known catch phrases. Individual crowd sentiment models will detect similar sounds.
Note: Evaluate the changed in correlation between model error rate and centroid distance
In machine learning, a nearest centroid classifier or nearest prototype classifier is a classification model that assigns to observations the label of the class of training samples whose mean (centroid) is closest to the observation.
References: https://en.wikipedia.org/wiki/Nearest_centroid_classifier
https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/sweep-clustering

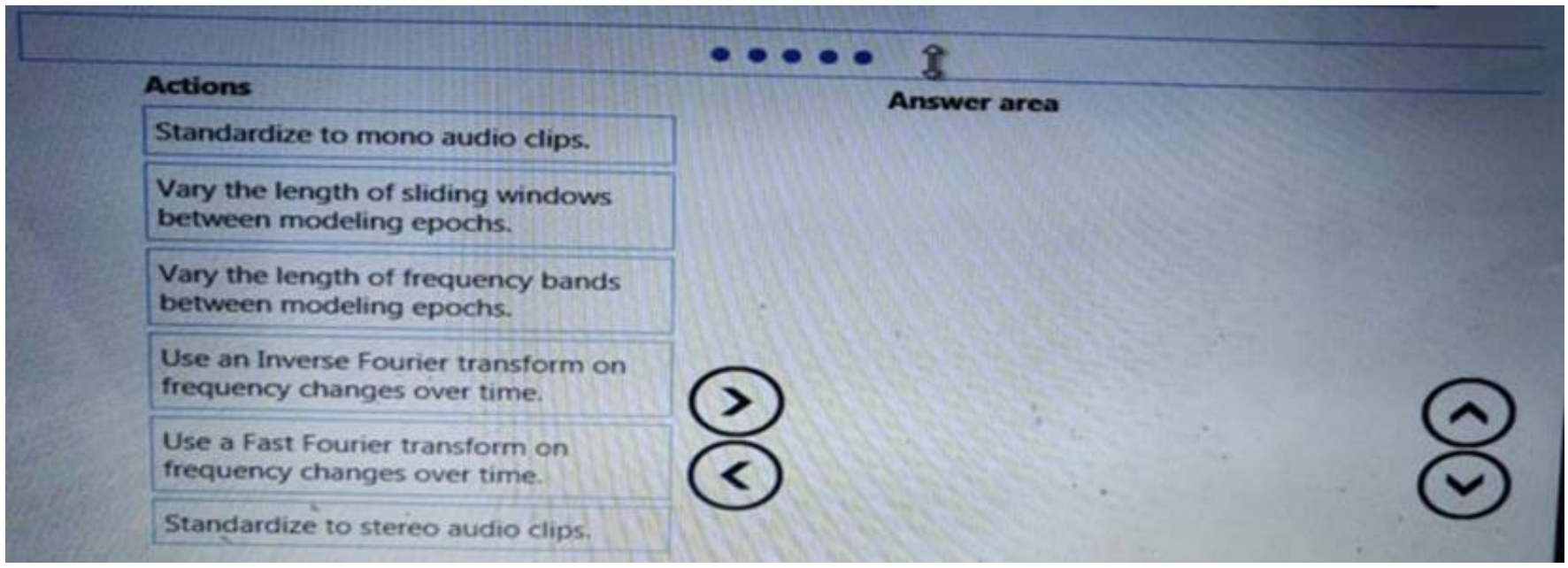
**NEW QUESTION 224**
- (Exam Topic 1)
You need to implement a scaling strategy for the local penalty detection data.
Which normalization type should you use?

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

**Answer:** C

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

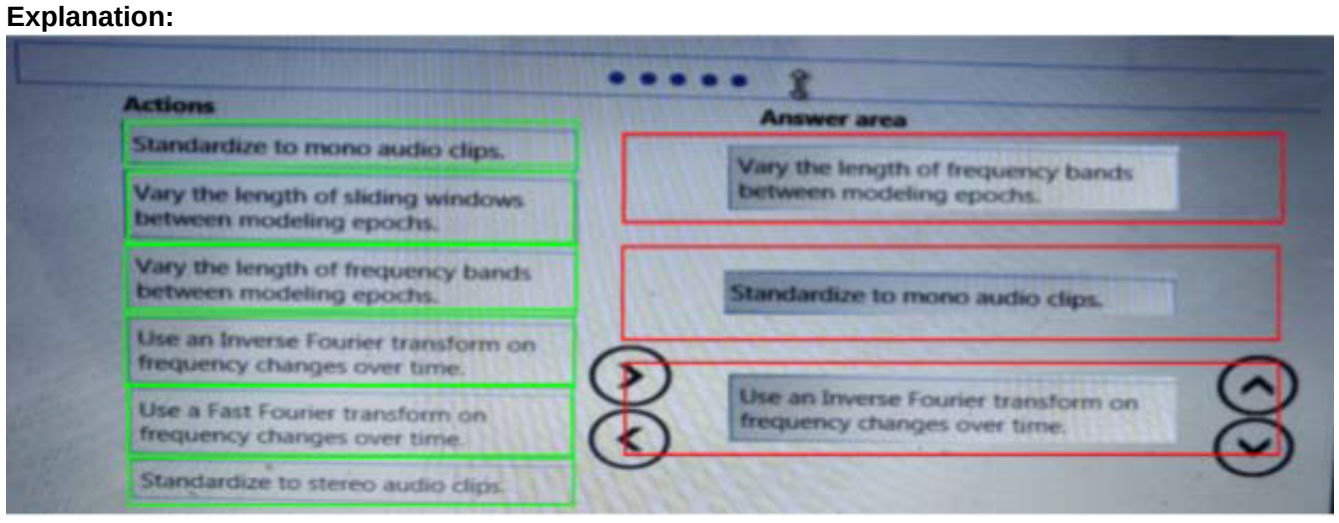
**NEW QUESTION 229**
- (Exam Topic 1)
You need to define a process for penalty event detection.
Which three actions should you perform in sequence? To answer, move the appropriate actions from the list of actions to the answer area and arrange them in the correct order.



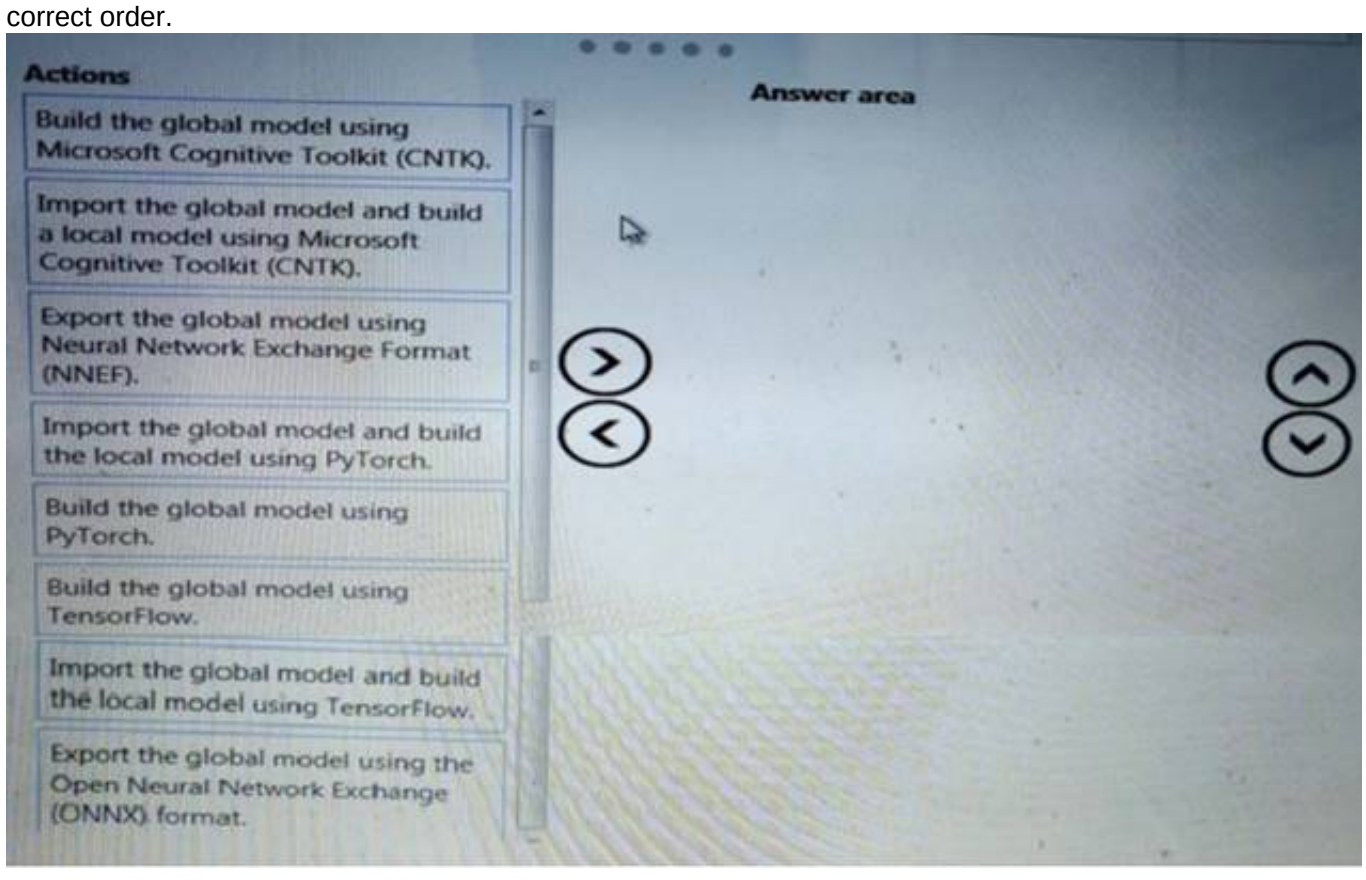

A. Mastered
B. Not Mastered

**Answer:** A

**Explanation:**



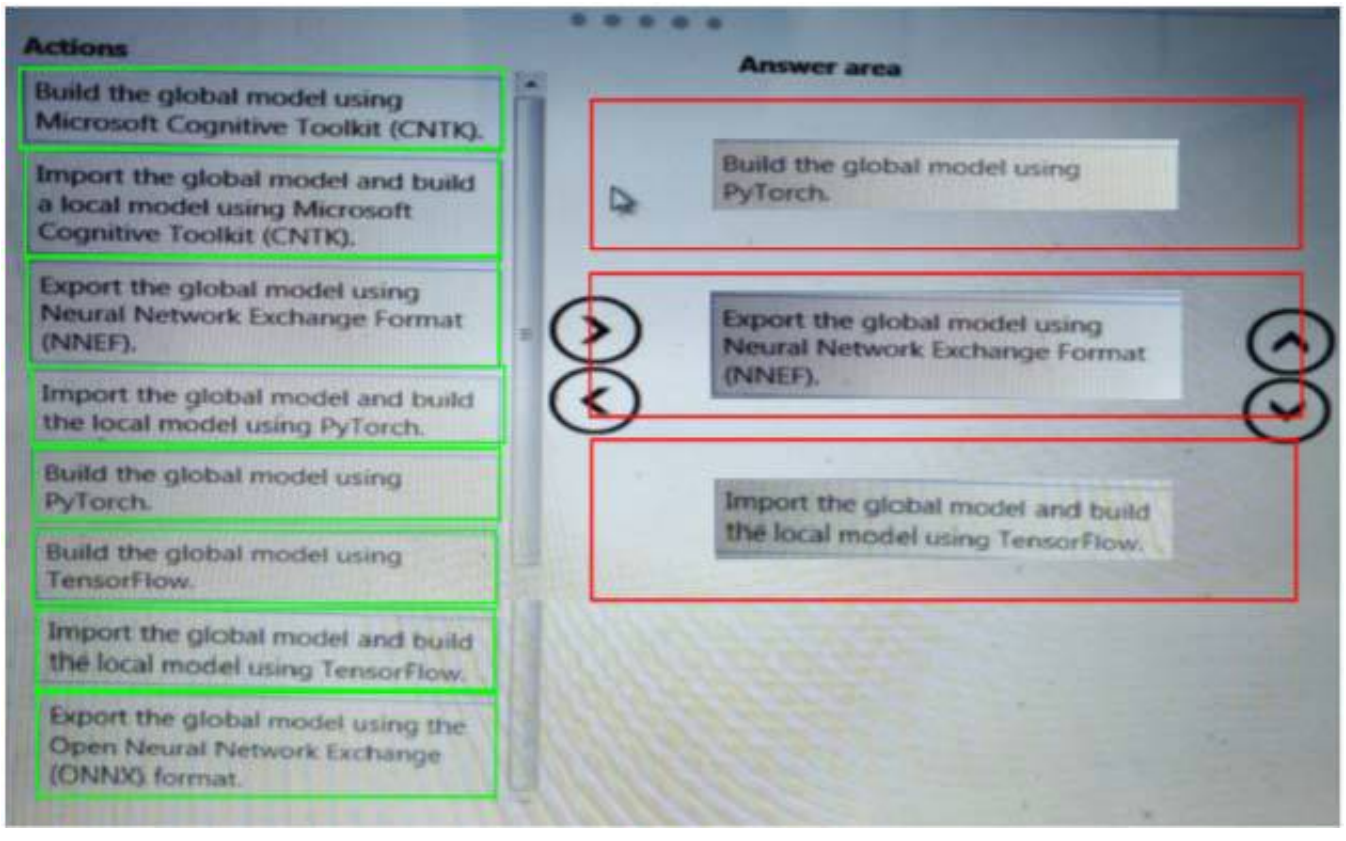

**NEW QUESTION 231**
- (Exam Topic 1)
You need to define a process for penalty event detection.
Which three actions should you perform in sequence? To answer, move the appropriate actions from the list of actions to the answer area and arrange them in the correct order.




A. Mastered
B. Not Mastered

**Answer:** A

**Explanation:**




**NEW QUESTION 234**

.....

# Thank You for Trying Our Product

## We offer two products:

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

2nd - Questons and Answers in PDF Format

## DP-100 Practice Exam Features:

* DP-100 Questions and Answers Updated Frequently

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

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

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