

## Exam Questions DP-100

Designing and Implementing a Data Science Solution on Azure

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



**NEW QUESTION 1**

- (Exam Topic 3)

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.

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

- A. Mastered
- B. Not Mastered

**Answer: A**

**Explanation:**

Box 1: Yes

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

Box 2: Yes

script\_params is a dictionary of command-line arguments to pass to the training script specified in entry\_script.

Box 3: No

Box 4: Yes

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

**NEW QUESTION 2**

- (Exam Topic 3)

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

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

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

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

Does the solution meet the goal?

- A. Yes
- B. No

**Answer: B**

**Explanation:**

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

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

References:

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

**NEW QUESTION 3**

- (Exam Topic 3)

You are developing a deep learning model by using TensorFlow. You plan to run the model training workload on an Azure Machine Learning Compute Instance. You must use CUDA-based model training. You need to provision the Compute Instance.

Which two virtual machines sizes can you use? To answer, select the appropriate virtual machine sizes in the answer area.

NOTE: Each correct selection is worth one point.

Virtual machine size

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

- A. Mastered
- B. Not Mastered

**Answer: A**

**Explanation:**

CUDA is a parallel computing platform and programming model developed by Nvidia for general computing on its own GPUs (graphics processing units). CUDA enables developers to speed up compute-intensive applications by harnessing the power of GPUs for the parallelizable part of the computation.

Reference:

<https://www.infoworld.com/article/3299703/what-is-cuda-parallel-programming-for-gpus.html>

**NEW QUESTION 4**

- (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 5**

- (Exam Topic 3)

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

- a Microsoft Surface device on which Microsoft Office has been installed. Corporate IT policies prevent the installation of additional software
- a Compute Instance named ds-workstation in the workspace with 2 CPUs and 8 GB of memory
- an Azure Machine Learning compute target named cpu-cluster with eight CPU-based nodes
- an Azure Machine Learning compute target named gpu-cluster with four CPU and GPU-based nodes

You need to specify the compute resources to be used for running the code to submit the experiment, and for running the script in order to minimize model training time.

Which resources should the data scientist use? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.

**Resource type**

**Option**

Run code to submit the experiment

|                                |   |
|--------------------------------|---|
|                                | ▼ |
| the Microsoft Surface device   |   |
| the ds-workstation notebook VM |   |
| the cpu-cluster compute target |   |
| the gpu-cluster compute target |   |

Run the training script

|                                |   |
|--------------------------------|---|
|                                | ▼ |
| the ds-workstation notebook VM |   |
| the cpu-compute target         |   |
| the gpu-compute target         |   |
| the Microsoft Surface device   |   |

- A. Mastered
- B. Not Mastered

**Answer:** A

**Explanation:**

**Resource type**

**Option**

Run code to submit the experiment

|                                |   |
|--------------------------------|---|
|                                | ▼ |
| the Microsoft Surface device   |   |
| the ds-workstation notebook VM |   |
| the cpu-cluster compute target |   |
| the gpu-cluster compute target |   |

Run the training script

|                                |   |
|--------------------------------|---|
|                                | ▼ |
| the ds-workstation notebook VM |   |
| the cpu-compute target         |   |
| the gpu-compute target         |   |
| the Microsoft Surface device   |   |

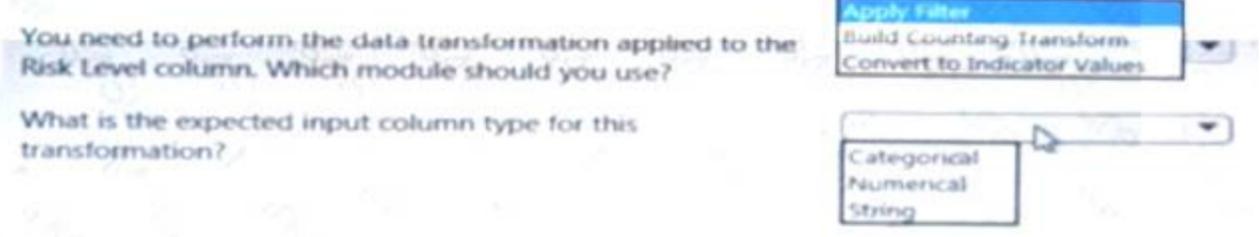
**NEW QUESTION 6**

- (Exam Topic 3)

You are developing a machine learning, experiment by using Azure. The following images show the input and output of a machine learning experiment:



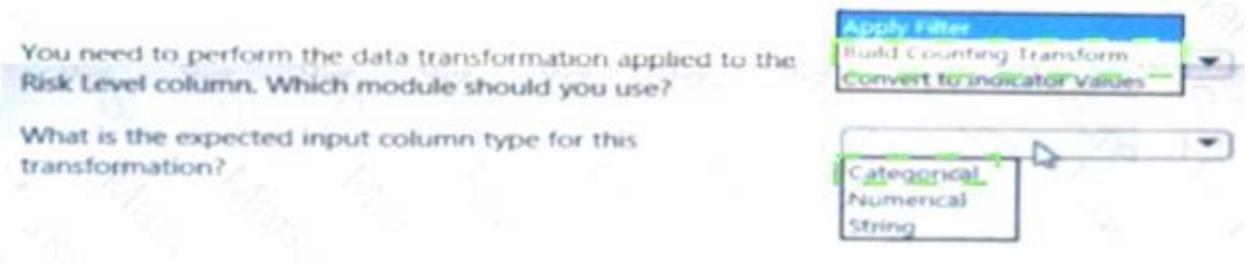
Use the drop-down menus to select the answer choice that answers each question based on the information presented in the graphic.  
 NOTE: Each correct selection is worth one point.



- A. Mastered
- B. Not Mastered

Answer: A

Explanation:



**NEW QUESTION 7**

- (Exam Topic 3)

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

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

NOTE: Each correct selection is worth one point.

| Model type  | Explainer   |
|---|---|
| A random forest model for predicting the alcohol content in wine given a set of covariates            | <input type="checkbox"/> Tabular<br><input type="checkbox"/> HAN<br><input type="checkbox"/> Text<br><input type="checkbox"/> Image |
| A natural language processing model for analyzing field reports                                       | <input type="checkbox"/> Tree<br><input type="checkbox"/> HAN<br><input type="checkbox"/> Text<br><input type="checkbox"/> Image    |
| An image classifier that determines the quality of the grape based upon its physical characteristics. | <input type="checkbox"/> Kernel<br><input type="checkbox"/> HAN<br><input type="checkbox"/> Text<br><input type="checkbox"/> Image  |

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

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

**NEW QUESTION 8**

- (Exam Topic 3)

You are solving a classification task.

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

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

- A. k=0.5

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

**Answer: C**

**Explanation:**

Leave One Out (LOO) cross-validation

Setting K = n (the number of observations) yields n-fold and is called leave-one out cross-validation (LOO), a special case of the K-fold approach.

LOO CV is sometimes useful but typically doesn't shake up the data enough. The estimates from each fold are highly correlated and hence their average can have high variance.

This is why the usual choice is K=5 or 10. It provides a good compromise for the bias-variance tradeoff.

**NEW QUESTION 9**

- (Exam Topic 3)

You create an Azure Machine Learning workspace.

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

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

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

Which JSON code segment should you use?

A)

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

B)

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

C)

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

D)

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

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

**Answer:** A

**Explanation:**

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

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

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

data\_scientist\_custom\_role.json :

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

Reference:

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

**NEW QUESTION 10**

- (Exam Topic 3)

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

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

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

- A. box plot
- B. scatter
- C. random forest diagram
- D. Venn diagram
- E. ROC curve

**Answer:** AB

**Explanation:**

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

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

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

**NEW QUESTION 10**

- (Exam Topic 3)

You have a dataset that includes home sales data for a city. The dataset includes the following columns.

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

Each row in the dataset corresponds to an individual home sales transaction.

You need to use automated machine learning to generate the best model for predicting the sales price based on the features of the house.

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

| Setting         | Value  |
|-----------------|--|
| Prediction task | <div style="border: 1px solid black; padding: 5px;">                     Classification<br/>                     Forecasting<br/>                     Regression<br/>                     Outlier                 </div>                   |
| Target column   | <div style="border: 1px solid black; padding: 5px;">                     Price<br/>                     Bedrooms<br/>                     Size<br/>                     HasGarage<br/>                     HomeType                 </div> |

- A. Mastered
- B. Not Mastered

**Answer:** A

**Explanation:**

Box 1: Regression

Regression is a supervised machine learning technique used to predict numeric values. Box 2: Price

Reference:

<https://docs.microsoft.com/en-us/learn/modules/create-regression-model-azure-machine-learning-designer>

**NEW QUESTION 12**

- (Exam Topic 3)

You create a batch inference pipeline by using the Azure ML SDK. You run the pipeline by using the following code:

```
from azureml.pipeline.core import Pipeline
```

```
from azureml.core.experiment import Experiment
```

```
pipeline = Pipeline(workspace=ws, steps=[parallelrun_step]) pipeline_run = Experiment(ws, 'batch_pipeline').submit(pipeline)
```

You need to monitor the progress of the pipeline execution.

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

NOTE: Each correct selection is worth one point.

A. Run the following code in a notebook:

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

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

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

D. Run the following code in a notebook:

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

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

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

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

**Answer:** DE

**Explanation:**

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

> Azure Machine Learning Studio.

> Console output from the PipelineRun object. `from azureml.widgets import RunDetails RunDetails(pipeline_run).show()`

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

<https://docs.microsoft.com/en-us/azure/machine-learning/how-to-use-parallel-run-step#monitor-the-parallel-run>

**NEW QUESTION 17**

- (Exam Topic 3)

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

run = Run.get_context()
data = pd.read_csv('./data.csv')
rows = (len(data))
# record row_count metric here
...
```

You need to record the row count as a metric named row\_count that can be returned using the get\_metrics method of the Run object after the experiment run completes. Which code should you use?

- A. run.upload\_file('row\_count', './data.csv')
- B. run.log('row\_count', rows)
- C. run.tag('row\_count', rows)
- D. run.log\_table('row\_count', rows)
- E. run.log\_row('row\_count', rows)

**Answer: B**

**Explanation:**

Log a numerical or string value to the run with the given name using log(name, value, description="). Logging a metric to a run causes that metric to be stored in the run record in the experiment. You can log the same metric multiple times within a run, the result being considered a vector of that metric.

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

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

**NEW QUESTION 20**

- (Exam Topic 3)

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

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

Which class object should you use?

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

**Answer: A**

**Explanation:**

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

The run Class get\_all\_logs method downloads all logs for the run to a directory. Reference:

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

**NEW QUESTION 22**

- (Exam Topic 3)

You create a datastore named training\_data that references a blob container in an Azure Storage account. The blob container contains a folder named csv\_files in which multiple comma-separated values (CSV) files are stored.

You have a script named train.py in a local folder named ./script that you plan to run as an experiment using an estimator. The script includes the following code to read data from the csv\_files folder:

```
import os
import argparse
import pandas as pd

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

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

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

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

You have the following script.

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

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

```
# Code to define estimator goes here
```

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

You need to configure the estimator for the experiment so that the script can read the data from a data reference named data\_ref that references the csv\_files folder in the training\_data datastore.

Which code should you use to configure the estimator?

- A. estimator = SKLearn(source\_directory='./script',  
inputs=[data\_ref.as\_named\_input('data-folder').to\_pandas\_dataframe()],  
compute\_target='local',  
entry\_script='train.py')
- B. script\_params = {  
    '--data-folder': data\_ref.as\_mount()  
}  
estimator = SKLearn(source\_directory='./script',  
script\_params=script\_params,  
compute\_target='local',  
entry\_script='train.py')
- C. estimator = SKLearn(source\_directory='./script',  
inputs=[data\_ref.as\_named\_input('data-folder').as\_mount()],  
compute\_target='local',  
entry\_script='train.py')
- D. script\_params = {  
    '--data-folder': data\_ref.as\_download(path\_on\_compute='csv\_files')  
}  
estimator = SKLearn(source\_directory='./script',  
script\_params=script\_params,  
compute\_target='local',  
entry\_script='train.py')
- E. estimator = SKLearn(source\_directory='./script',  
inputs=[data\_ref.as\_named\_input('data-folder').as\_download(path\_on\_compute='csv\_files')],  
compute\_target='local',  
entry\_script='train.py')

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

**Answer:** B

#### Explanation:

Besides passing the dataset through the inputs parameter in the estimator, you can also pass the dataset through script\_params and get the data path (mounting point) in your training script via arguments. This way, you can keep your training script independent of azureml-sdk. In other words, you will be able use the same training script for local debugging and remote training on any cloud platform.

Example:

```
from azureml.train.sklearn import SKLearn
script_params = {
# mount the dataset on the remote compute and pass the mounted path as an argument to the training script '--data-folder':
mnist_ds.as_named_input('mnist').as_mount(),
'--regularization': 0.5
}
est = SKLearn(source_directory=script_folder, script_params=script_params, compute_target=compute_target, environment_definition=env,
entry_script='train_mnist.py')
# Run the experiment
run = experiment.submit(est)
run.wait_for_completion(show_output=True)
Reference:
https://docs.microsoft.com/es-es/azure/machine-learning/how-to-train-with-datasets
```

#### NEW QUESTION 27

- (Exam Topic 3)

You use Azure Machine Learning to train and register a model.

You must deploy the model into production as a real-time web service to an inference cluster named service-compute that the IT department has created in the Azure Machine Learning workspace.

Client applications consuming the deployed web service must be authenticated based on their Azure Active Directory service principal.

You need to write a script that uses the Azure Machine Learning SDK to deploy the model. The necessary modules have been imported.

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

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

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

)
token_auth_enabled=True
token_auth_enabled=False
auth_enabled=True
auth_enabled=False

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

- A. Mastered
- B. Not Mastered

Answer: A

**Explanation:**

Box 1: AksCompute Example:

```
aks_target = AksCompute(ws,"myaks")
```

# If deploying to a cluster configured for dev/test, ensure that it was created with enough  
 # cores and memory to handle this deployment configuration. Note that memory is also used by  
 # things such as dependencies and AML components.

```
deployment_config = AksWebservice.deploy_configuration(cpu_cores = 1, memory_gb = 1)
```

```
service = Model.deploy(ws, "myservice", [model], inference_config, deployment_config, aks_target)
```

Box 2: AksWebservice

Box 3: token\_auth\_enabled=Yes  
 Whether or not token auth is enabled for the Webservice.

Note: A Service principal defined in Azure Active Directory (Azure AD) can act as a principal on which authentication and authorization policies can be enforced in Azure Databricks.

The Azure Active Directory Authentication Library (ADAL) can be used to programmatically get an Azure AD access token for a user.

Reference:

<https://docs.microsoft.com/en-us/azure/machine-learning/how-to-deploy-azure-kubernetes-service> <https://docs.microsoft.com/en-us/azure/databricks/dev-tools/api/latest/aad/service-prin-aad-token>

**NEW QUESTION 31**

- (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>"
  ]
}
```

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

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

- A. Mastered
- B. Not Mastered

**Answer:** A

**Explanation:**

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

Box 1: Microsoft.MachineLearningServices/workspaces/\*/read

Reader role: Read-only actions in the workspace. Readers can list and view assets, including datastore credentials, in a workspace. Readers can't create or update these assets.

Box 2: Microsoft.MachineLearningServices/workspaces/\*/write

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

Box 3: Microsoft.MachineLearningServices/workspaces/computes/\*/delete

Box 4: Microsoft.MachineLearningServices/workspaces/computes/\*/write Reference:

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

**NEW QUESTION 36**

- (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: Replace each missing value using the Multiple Imputation by Chained Equations (MICE) method. Does the solution meet the goal?

- A. Yes
- B. NO

**Answer:** A

**Explanation:**

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: Multivariate imputation by chained equations (MICE), sometimes called "fully conditional specification" or "sequential regression multiple imputation" has emerged in the statistical literature as one principled method of addressing missing data. Creating multiple imputations, as opposed to single imputations, accounts for the statistical uncertainty in the imputations. In addition, the chained equations approach is very flexible and can handle variables of varying types (e.g., continuous or binary) as well as complexities such as bounds or survey skip patterns.

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 41**

- (Exam Topic 3)

You plan to run a Python script as an Azure Machine Learning experiment.

The script must read files from a hierarchy of folders. The files will be passed to the script as a dataset argument.

You must specify an appropriate mode for the dataset argument.

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

NOTE: Each correct selection is worth one point.

- A. to\_pandas\_dataframe ()
- B. as\_download()
- C. as\_upload()
- D. as mount ()

**Answer: B**

**Explanation:**

Reference:

<https://docs.microsoft.com/en-us/python/api/azureml-core/azureml.data.filedataset?view=azure-ml-py>

**NEW QUESTION 44**

- (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 46**

- (Exam Topic 3)

You use an Azure Machine Learning workspace.

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

What should you do?

- A. Deploy the model to Azure Kubernetes Service (AKS). During deployment, set the token\_auth\_enabled parameter of the target configuration object to true
- B. Deploy the model to Azure Container Instance
- C. During deployment, set the auch\_enabled parameter of the target configuration object to true
- D. Deploy the model to Azure Container Instance
- E. During deployment, set the coken\_auch\_enabled parameter of the target configuration object to true
- F. Deploy the model to Azure Kubernetes Service (AKS). During deployment, set the auc
- G. enabled parameter of the target configuration object to true

**Answer: A**

**Explanation:**

To control token authentication, use the token\_auth\_enabled parameter when you create or update a deployment

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

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

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

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

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

**NEW QUESTION 50**

- (Exam Topic 3)

You create a multi-class image classification deep learning model.

The model must be retrained monthly with the new image data fetched from a public web portal. You create an Azure Machine Learning pipeline to fetch new data, standardize the size of images, and retrain the model.

You need to use the Azure Machine Learning SDK to configure the schedule for the pipeline.

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

| Actions  | Answer Area |
|--|-------------|
| Publish the pipeline.  |             |
| Retrieve the pipeline ID.  |             |
| Create a ScheduleRecurrence(frequency= 'Month', interval=1, start_time='2019-01-01T00:00:00') object.                          |             |
| Define a pipeline parameter named <b>RunDate</b> .   |             |
| Define a new Azure Machine Learning pipeline StepRun object with the step ID of the first step in the pipeline.                |             |
| Define an Azure Machine Learning pipeline schedule using the schedule.create method with the defined recurrence specification. |             |

- A. Mastered
- B. Not Mastered

**Answer:** A

**Explanation:**

Step 1: Publish the pipeline.

To schedule a pipeline, you'll need a reference to your workspace, the identifier of your published pipeline, and the name of the experiment in which you wish to create the schedule.

Step 2: Retrieve the pipeline ID. Needed for the schedule.

Step 3: Create a ScheduleRecurrence..

To run a pipeline on a recurring basis, you'll create a schedule. A Schedule associates a pipeline, an experiment, and a trigger.

First create a schedule. Example: Create a Schedule that begins a run every 15 minutes: recurrence = ScheduleRecurrence(frequency="Minute", interval=15)

Step 4: Define an Azure Machine Learning pipeline schedule.. Example, continued:

```
recurring_schedule = Schedule.create(ws, name="MyRecurringSchedule", description="Based on time", pipeline_id=pipeline_id, experiment_name=experiment_name, recurrence=recurrence)
```

Reference:

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

**NEW QUESTION 55**

- (Exam Topic 3)

You have a dataset created for multiclass classification tasks that contains a normalized numerical feature set with 10,000 data points and 150 features.

You use 75 percent of the data points for training and 25 percent for testing. You are using the scikit-learn machine learning library in Python. You use X to denote the feature set and Y to denote class labels.

You create the following Python data frames:

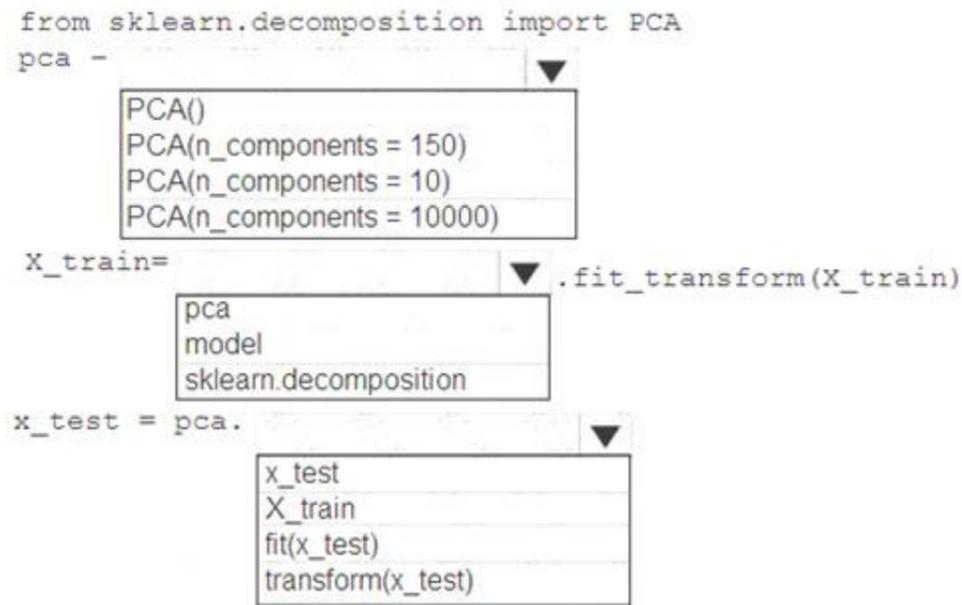
| Name    | Description           |
|---------|-----------------------|
| X_train | training feature set  |
| Y_train | training class labels |
| x_train | testing feature set   |
| y_train | testing class labels  |

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

```

from sklearn.decomposition import PCA
pca = PCA(n_components=10)
X_train = pca.fit_transform(X_train)
x_test = pca.transform(x_test)

```



- A. Mastered
- B. Not Mastered

Answer: A

**Explanation:**

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

**NEW QUESTION 57**

- (Exam Topic 3)

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

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

Answer: CE

**Explanation:**

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

**NEW QUESTION 60**

- (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' resource\_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 62**

- (Exam Topic 3)

You are building a recurrent neural network to perform a binary classification. You review the training loss, validation loss, training accuracy, and validation accuracy for each training epoch.

You need to analyze model performance.

Which observation indicates that the classification model is over fitted?

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

**Answer:** B

**NEW QUESTION 65**

- (Exam Topic 3)

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

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

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

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

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

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

NOTE: Each correct selection is worth one point.

## Code segments

log  
log\_list  
log\_row  
log\_table  
log\_image

## Answer Area

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

- A. Mastered  
B. Not Mastered

Answer: A

**Explanation:**

Box 1: log

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

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

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

Box 2: log\_image

A box plot of income by home\_owner.

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

Example: run.log\_image("ROC", plot=plt) Box 3: log\_table

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

**NEW QUESTION 69**

- (Exam Topic 3)

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

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

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

You need to configure the module.

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

SMOTE

Label column

Selected columns:  
**All labels**

Launch column selector

SMOTE percentage

0  
 300  
 3000  
 4000

Number of nearest neighbors

0  
 1  
 5  
 4000

Random seed

0

- A. Mastered
- B. Not Mastered

Answer: A

**Explanation:**

Box 1: 300

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

Box 2: 5

We should use 5 data rows.

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

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

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

References:

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

**NEW QUESTION 73**

- (Exam Topic 3)

A biomedical research company plans to enroll people in an experimental medical treatment trial.

You create and train a binary classification model to support selection and admission of patients to the trial. The model includes the following features: Age, Gender, and Ethnicity.

The model returns different performance metrics for people from different ethnic groups.

You need to use Fairlearn to mitigate and minimize disparities for each category in the Ethnicity feature. Which technique and constraint should you use? To answer, select the appropriate options in the answer area. NOTE: Each correct selection is worth one point.

**Option**

**Value**

Technique

Grid search  
 Outlier detection  
 Dimensionality reduction

Constraint

Demographic parity  
 False-positive rate parity

- A. Mastered
- B. Not Mastered

**Answer:** A

**Explanation:**

Graphical user interface, text, application, chat or text message Description automatically generated

Box 1: Grid Search

Fairlearn open-source package provides postprocessing and reduction unfairness mitigation algorithms: ExponentiatedGradient, GridSearch, and ThresholdOptimizer.

Note: The Fairlearn open-source package provides postprocessing and reduction unfairness mitigation algorithms types:

- Reduction: These algorithms take a standard black-box machine learning estimator (e.g., a LightGBM model) and generate a set of retrained models using a sequence of re-weighted training datasets.
- Post-processing: These algorithms take an existing classifier and the sensitive feature as input.

Box 2: Demographic parity

The Fairlearn open-source package supports the following types of parity constraints: Demographic parity, Equalized odds, Equal opportunity, and Bounded group loss.

Reference:

<https://docs.microsoft.com/en-us/azure/machine-learning/concept-fairness-ml>

**NEW QUESTION 77**

- (Exam Topic 3)

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

...

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

You need to add code to run the steps.

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

NOTE: Each correct selection is worth one point.

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

**Answer:** CD

**Explanation:**

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

# Build the pipeline. Example:

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

# Submit the pipeline to be run

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

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

**NEW QUESTION 79**

- (Exam Topic 3)

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

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

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

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

**Answer:** BE

**Explanation:**

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

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

What is the mean and standard deviation for each column?

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

References:

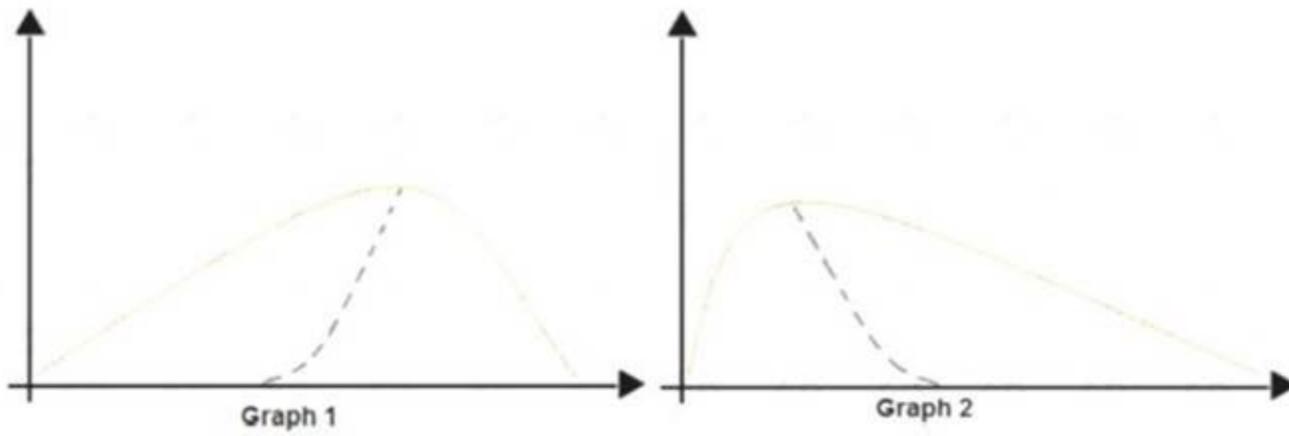
<https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/export-count-table> <https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/summarize-data>

**NEW QUESTION 82**

- (Exam Topic 3)

You are analyzing the asymmetry in a statistical distribution.

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



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

| Question  | Answer choice   |
|---|---|
| Which type of distribution is shown for the dataset density curve of Graph 1? | <ul style="list-style-type: none"> <li>Negative skew</li> <li>Positive skew</li> <li>Normal distribution</li> <li>Bimodal distribution</li> </ul> |
| Which type of distribution is shown for the dataset density curve of Graph 2? | <ul style="list-style-type: none"> <li>Negative skew</li> <li>Positive skew</li> <li>Normal distribution</li> <li>Bimodal distribution</li> </ul> |

- A. Mastered
- B. Not Mastered

**Answer:** A

**Explanation:**

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

**NEW QUESTION 84**

- (Exam Topic 3)

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

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

**Answer:** B

**NEW QUESTION 86**

- (Exam Topic 3)

You develop and train a machine learning model to predict fraudulent transactions for a hotel booking website. Traffic to the site varies considerably. The site experiences heavy traffic on Monday and Friday and much lower traffic on other days. Holidays are also high web traffic days. You need to deploy the model as an Azure Machine Learning real-time web service endpoint on compute that can dynamically scale up and down to support demand. Which deployment compute option should you use?

- A. attached Azure Databricks cluster
- B. Azure Container Instance (ACI)
- C. Azure Kubernetes Service (AKS) inference cluster
- D. Azure Machine Learning Compute Instance
- E. attached virtual machine in a different region

**Answer:** D

**Explanation:**

Azure Machine Learning compute cluster is a managed-compute infrastructure that allows you to easily create a single or multi-node compute. The compute is created within your workspace region as a resource that can be shared with other users in your workspace. The compute scales up automatically when a job is submitted, and can be put in an Azure Virtual Network.  
 Reference:

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

**NEW QUESTION 89**

- (Exam Topic 3)

You are analyzing a raw dataset that requires cleaning.

You must perform transformations and manipulations by using Azure Machine Learning Studio. You need to identify the correct modules to perform the transformations.

Which modules should you choose? To answer, drag the appropriate modules to the correct scenarios. Each module may be used once, more than once, or not at all.

You may need to drag the split bar between panes or scroll to view content. NOTE: Each correct selection is worth one point.

**Answer Area**

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

- A. Mastered
- B. Not Mastered

**Answer:** A

**Explanation:**

Box 1: Clean Missing Data Box 2: SMOTE

Use the SMOTE module in Azure Machine Learning Studio to increase the number of underrepresented cases in a dataset used for machine learning. SMOTE is a better way of increasing the number of rare cases than simply duplicating existing cases.

Box 3: Convert to Indicator Values

Use the Convert to Indicator Values module in Azure Machine Learning Studio. The purpose of this module is to convert columns that contain categorical values into a series of binary indicator columns that can more easily be used as features in a machine learning model.

Box 4: Remove Duplicate Rows References:

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

**NEW QUESTION 92**

- (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 93**

- (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 | <div style="border: 1px solid black; padding: 5px;">                     JSON containing the run ID<br/>                     JSON containing the pipeline ID<br/>                     JSON containing the experiment name<br/>                     JSON containing an OAuth bearer token                 </div>                                 |
| Response       | <div style="border: 1px solid black; padding: 5px;">                     JSON containing the run ID<br/>                     JSON containing the pipeline ID<br/>                     JSON containing the experiment name<br/>                     JSON containing an OAuth bearer token                 </div>                                 |
| Response       | <div style="border: 1px solid black; padding: 5px;">                     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                 </div> |

- 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 94

- (Exam Topic 3)

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

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

Answer Area

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

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

Answer Area

|   |  |     |    |                                  |                       |   |  |                       |                                  |
|---|--|-----|----|----------------------------------|-----------------------|---|--|-----------------------|----------------------------------|
| <p>The estimator will look for the files it needs to run an experiment in the training-experiment directory of the local compute environment.</p> <p>The estimator will mount the local data-folder folder and make it available to the script through a parameter.</p> <p>The train.py script file will be created if it does not exist.</p> | <table border="0"> <tr> <td style="padding-right: 20px;">Yes</td> <td>No</td> </tr> <tr> <td style="text-align: center;"><input checked="" type="radio"/></td> <td style="text-align: center;"><input type="radio"/></td> </tr> <tr> <td colspan="2" style="text-align: center; border: 1px solid gray; padding: 2px;">                 This is the way that the estimator will look for the files it needs to run an experiment in the training-experiment directory of the local compute environment.             </td> </tr> <tr> <td style="text-align: center;"><input type="radio"/></td> <td style="text-align: center;"><input checked="" type="radio"/></td> </tr> </table> | Yes | No | <input checked="" type="radio"/> | <input type="radio"/> | This is the way that the estimator will look for the files it needs to run an experiment in the training-experiment directory of the local compute environment. |  | <input type="radio"/> | <input checked="" type="radio"/> |
| Yes   | No   |     |    |                                  |                       |   |  |                       |                                  |
| <input checked="" type="radio"/>  | <input type="radio"/>  |     |    |                                  |                       |   |  |                       |                                  |
| This is the way that the estimator will look for the files it needs to run an experiment in the training-experiment directory of the local compute environment.   |  |     |    |                                  |                       |   |  |                       |                                  |
| <input type="radio"/>   | <input checked="" type="radio"/>   |     |    |                                  |                       |   |  |                       |                                  |

**NEW QUESTION 97**

- (Exam Topic 3)

You create a new Azure Databricks workspace.

You configure a new cluster for long-running tasks with mixed loads on the compute cluster as shown in the image below.

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

Code for each user runs as a separate process

|     |
|-----|
| ▼   |
| Yes |
| No  |

The number of workers is fixed for the entire duration of the job

|     |
|-----|
| ▼   |
| Yes |
| No  |

- A. Mastered
- B. Not Mastered

**Answer:** A

**Explanation:**

Box 1: No  
 Running user code in separate processes is not possible in Scala. Box 2: No  
 Autoscaling is enabled. Minimum 2 workers, Maximum 8 workers. Reference:  
<https://docs.databricks.com/clusters/configure.html>

**NEW QUESTION 100**

- (Exam Topic 3)

You create a classification model with a dataset that contains 100 samples with Class A and 10,000 samples with Class B. The variation of Class B is very high. You need to resolve imbalances. Which method should you use?

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

**Answer:** D

**NEW QUESTION 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 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:**

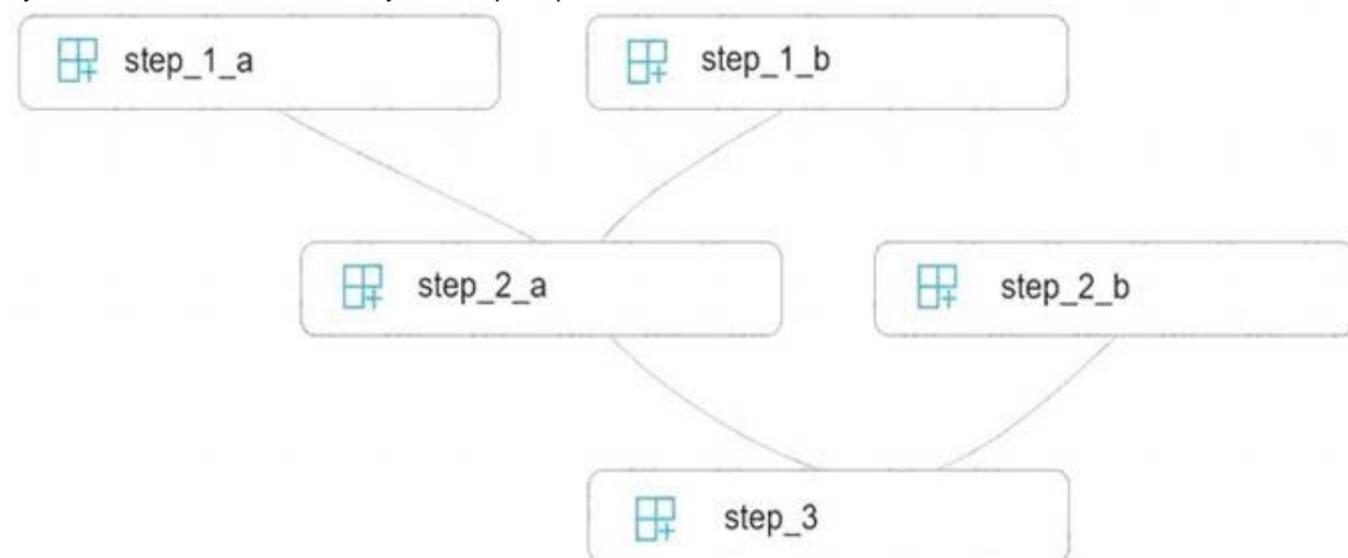
Use the Entropy MDL binning mode which has a target column. References:  
<https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/group-data-into-bins>

**NEW QUESTION 109**

- (Exam Topic 3)

You write five Python scripts that must be processed in the order specified in Exhibit A – which allows the same modules to run in parallel, but will wait for modules with dependencies.

You must create an Azure Machine Learning pipeline using the Python SDK, because you want to script to create the pipeline to be tracked in your version control system. You have created five PythonScriptSteps and have named the variables to match the module names.



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

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

```
pipeline_steps = {
  "Pipeline": {
    "run": step_3,
    "run_after": {[
      {"run": step_2_a,
        "run_after":
          [{"run": step_1_a},
           {"run": step_1_b}]
        },
      {"run": step_2_b}
    ]}
  }
}
```

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

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

**Answer:** A

**Explanation:**

The steps parameter is an array of steps. To build pipelines that have multiple steps, place the steps in order in this array.

Reference:

<https://docs.microsoft.com/en-us/azure/machine-learning/how-to-use-parallel-run-step>

**NEW QUESTION 110**

- (Exam Topic 3)

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

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

**Answer:** A

**Explanation:**

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

References:

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

**NEW QUESTION 112**

- (Exam Topic 3)

You plan to preprocess text from CSV files. You load the Azure Machine Learning Studio default stop words list.

You need to configure the Preprocess Text module to meet the following requirements:

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

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

Preprocess Text

Language

Remove by part of speech

Text column to clean  

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

Remove stop words

Lemmatization

Detect sentences

Normalize case to lowercase

Remove numbers

Remove special characters

Remove duplicate characters

Remove email addresses

Remove URLs

Expand verb contractions

Normalize backslashes to slashes

Split tokens on special characters

- A. Mastered
- B. Not Mastered

**Answer:** A

**Explanation:**

Box 1: Remove stop words

Remove words to optimize information retrieval.

Remove stop words: Select this option if you want to apply a predefined stopword list to the text column. Stop word removal is performed before any other processes.

Box 2: Lemmatization

Ensure that multiple related words from a single canonical form. Lemmatization converts multiple related words to a single canonical form

Box 3: Remove special characters  
 Remove special characters: Use this option to replace any non-alphanumeric special characters with the pipe | character.

References:

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

**NEW QUESTION 115**

- (Exam Topic 3)

You create an Azure Machine Learning workspace and set up a development environment. You plan to train a deep neural network (DNN) by using the Tensorflow framework and by using estimators to submit training scripts.

You must optimize computation speed for training runs.

You need to choose the appropriate estimator to use as well as the appropriate training compute target configuration.

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

Answer Area

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

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

Box 1: Tensorflow

TensorFlow represents an estimator for training in TensorFlow experiments. Box 2: 12 vCPU, 112 GB memory...,2 GPU,..

Use GPUs for the deep neural network. Reference:

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

NEW QUESTION 117

- (Exam Topic 3)

You have a Jupyter Notebook that contains Python code that is used to train a model.

You must create a Python script for the production deployment. The solution must minimize code maintenance.

Which two actions should you perform? Each correct answer presents part of the solution.

NOTE: Each correct selection is worth one point.

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

Answer: AC

Explanation:

Reference:

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

NEW QUESTION 120

- (Exam Topic 3)

You create an Azure Machine Learning compute target named ComputeOne by using the STANDARD\_D1 virtual machine image.

You define a Python variable named ws that references the Azure Machine Learning workspace. You run the following Python code:

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

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

NOTE: Each correct selection is worth one point.

|   | Yes                   | No                    |
|---|-----------------------|-----------------------|
| A new machine learning compute resource is created with a virtual machine size of STANDARD_DS12_v2 and a maximum of four nodes. | <input type="radio"/> | <input type="radio"/> |
| Any experiments configured to use the_cluster will run on ComputeOne.   | <input type="radio"/> | <input type="radio"/> |
| The text <b>Step1</b> will be printed to the screen.  | <input type="radio"/> | <input type="radio"/> |

- A. Mastered
- B. Not Mastered

Answer: A

**Explanation:**

Box 1: Yes

ComputeTargetException class: An exception related to failures when creating, interacting with, or configuring a compute target. This exception is commonly raised for failures attaching a compute target,

missing headers, and unsupported configuration values. Create(workspace, name, provisioning\_configuration)

Provision a Compute object by specifying a compute type and related configuration. This method creates a new compute target rather than attaching an existing one. Box 2: Yes

Box 3: No

The line before print('Step1') will fail. Reference:

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

**NEW QUESTION 124**

- (Exam Topic 3)

You are working on a classification task. You have a dataset indicating whether a student would like to play soccer and associated attributes. The dataset includes the following columns:

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

You need to classify variables by type.

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

| Category              | Variables  |
|-----------------------|--|
| Categorical variables | <ul style="list-style-type: none"> <li>Gender, IsPlaySoccer</li> <li>Gender, PrevExamMarks, Height, Weight</li> <li>PrevExamMarks, Height, Weight</li> <li>IsPlaySoccer</li> </ul> |
| Continuous variables  | <ul style="list-style-type: none"> <li>Gender, IsPlaySoccer</li> <li>Gender, PrevExamMarks, Height, Weight</li> <li>PrevExamMarks, Height, Weight</li> <li>IsPlaySoccer</li> </ul> |

- A. Mastered
- B. Not Mastered

Answer: A

**Explanation:**

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

**NEW QUESTION 125**

- (Exam Topic 3)

You create a deep learning model for image recognition on Azure Machine Learning service using GPU-based training.

You must deploy the model to a context that allows for real-time GPU-based inferencing. You need to configure compute resources for model inferencing.

Which compute type should you use?

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

Answer: B

**Explanation:**

You can use Azure Machine Learning to deploy a GPU-enabled model as a web service. Deploying a model on Azure Kubernetes Service (AKS) is one option. The AKS cluster provides a GPU resource that is used by the model for inference.

Inference, or model scoring, is the phase where the deployed model is used to make predictions. Using GPUs instead of CPUs offers performance advantages on highly parallelizable computation.

Reference:

<https://docs.microsoft.com/en-us/azure/machine-learning/how-to-deploy-inferencing-gpus>

**NEW QUESTION 126**

- (Exam Topic 3)

An organization creates and deploys a multi-class image classification deep learning model that uses a set of labeled photographs.

The software engineering team reports there is a heavy inferencing load for the prediction web services during the summer. The production web service for the

model fails to meet demand despite having a fully-utilized compute cluster where the web service is deployed. You need to improve performance of the image classification web service with minimal downtime and minimal administrative effort. What should you advise the IT Operations team to do?

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

**Answer: D**

**Explanation:**

The Azure Machine Learning SDK does not provide support scaling an AKS cluster. To scale the nodes in the cluster, use the UI for your AKS cluster in the Azure Machine Learning studio. You can only change the node count, not the VM size of the cluster.

Reference:

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

**NEW QUESTION 131**

- (Exam Topic 2)

You need to correct the model fit issue.

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

| Actions                                       |   | Answer Area |
|---|---|-------------|
| Add the Ordinal Regression module.            |   |             |
| Add the Two-Class Averaged Perception module. |   |             |
| Augment the data.                             | ➤ | ⤴           |
| Add the Bayesian Linear Regression module.    | ⬅ | ⤵           |
| Decrease the memory size for L-BFGS.          |   |             |
| Add the Multiclass Decision Jungle module.    |   |             |
| Configure the regularization weight.          |   |             |

- A. Mastered
- B. Not Mastered

**Answer: A**

**Explanation:**

Step 1: Augment the data

Scenario: Columns in each dataset contain missing and null values. The datasets also contain many outliers.

Step 2: Add the Bayesian Linear Regression module.

Scenario: You produce a regression model to predict property prices by using the Linear Regression and Bayesian Linear Regression modules.

Step 3: Configure the regularization weight.

Regularization typically is used to avoid overfitting. For example, in L2 regularization weight, type the value to use as the weight for L2 regularization. We recommend that you use a non-zero value to avoid overfitting.

Scenario:

Model fit: The model shows signs of overfitting. You need to produce a more refined regression model that reduces the overfitting.

**NEW QUESTION 135**

- (Exam Topic 2)

You need to implement early stopping criteria as suited in the model training requirements.

Which three code segments should you use to develop the solution? To answer, move the appropriate code segments from the list of code segments to the answer area and arrange them in the correct order.

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

Code segments

Answer Area

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

```
import TruncationSelectionPolicy
```

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

```
import BanditPolicy
```

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



- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

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

NEW QUESTION 136

- (Exam Topic 1)

You need to build a feature extraction strategy for the local models.

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

The code segment to be completed is:

```
with C.layers.default_options(init=C.glorot_uniform(), activation=C.relu):
h = features
```

The four options shown are:

- ```
h = C.layers.Convolution2D(num_filters=8...)(h)
h = C.layers.MaxPooling(filter_shape=(3,3...)(h)
h = C.layers.Convolution2D(num_filters=16...)(h)
h = C.layers.MaxPooling(filter_shape=(2,2...)(h)
```
- ```
h = C.layers.MaxPooling(filter_shape=(3,3...)(h)
h = C.layers.MaxPooling(filter_shape=(2,2...)(h)
h = C.layers.Convolution2D(num_filters=8...)(h)
h = C.layers.Convolution2D(num_filters=16...)(h)
```
- ```
h = C.layers.Convolution2D(num_filters=16...)(h)
h = C.layers.Convolution2D(num_filters=8...)(h)
h = C.layers.MaxPooling(filter_shape=(2,2...)(h)
h = C.layers.MaxPooling(filter_shape=(3,3...)(h)
```
- ```
h = C.layers.MaxPooling(filter_shape=(3,3...)(h)
h = C.layers.MaxPooling(filter_shape=(2,2...)(h)
h = C.layers.Convolution2D(num_filters=8...)(h)
h = C.layers.Convolution2D(num_filters=16...)(h)
```

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

```

Answer Area

with C.layers.default_options(init=C.glorot_uniform(), activation=C.relu):
h = features
h = C.layers.Convolution2D(num_filters=8...)(h)
h = C.layers.MaxPooling(filter_shape=(3,3)...)(h)
h = C.layers.Convolution2D(num_filters=16...)(h)
h = C.layers.MaxPooling(filter_shape=(2,2)...)(h)
r = C.layers.Dense...

h = C.layers.MaxPooling(filter_shape=(3,3)...)(h)
h = C.layers.MaxPooling(filter_shape=(2,2)...)(h)
h = C.layers.Convolution2D(num_filters=8...)(h)
h = C.layers.Convolution2D(num_filters=16...)(h)

h = C.layers.Convolution2D(num_filters=16...)(h)
h = C.layers.Convolution2D(num_filters=8...)(h)
h = C.layers.MaxPooling(filter_shape=(2,2)...)(h)
h = C.layers.MaxPooling(filter_shape=(3,3)...)(h)

h = C.layers.MaxPooling(filter_shape=(3,3)...)(h)
h = C.layers.MaxPooling(filter_shape=(2,2)...)(h)
h = C.layers.Convolution2D(num_filters=8...)(h)
h = C.layers.Convolution2D(num_filters=16...)(h)
    
```

**NEW QUESTION 137**

- (Exam Topic 1)

You need to implement a new cost factor scenario for the ad response models as illustrated in the performance curve exhibit. Which technique should you use?

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

**Answer:** A

**Explanation:**

Scenario:

Performance curves of current and proposed cost factor scenarios are shown in the following diagram:



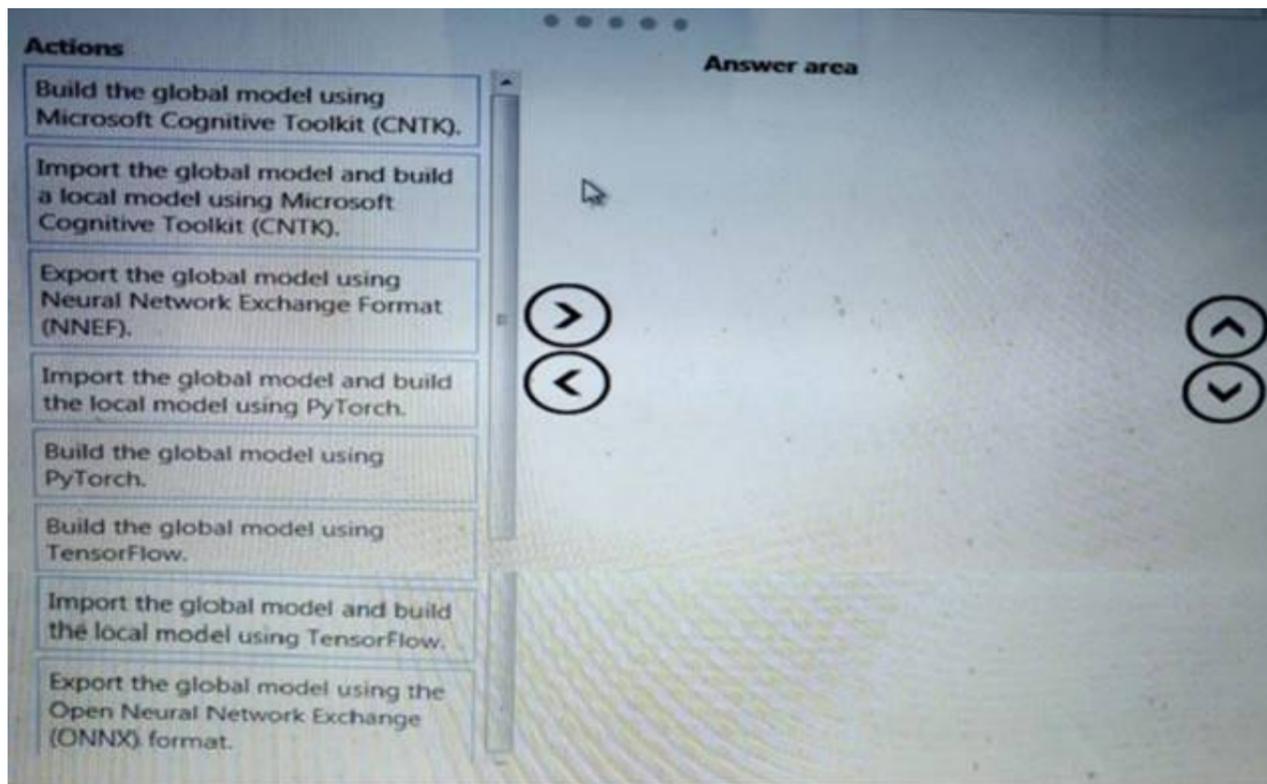
The ad propensity model uses a cut threshold is 0.45 and retrains occur if weighted Kappa deviated from 0.1 +/- 5%.

**NEW QUESTION 138**

- (Exam Topic 1)

You need to define a process for penalty event detection.

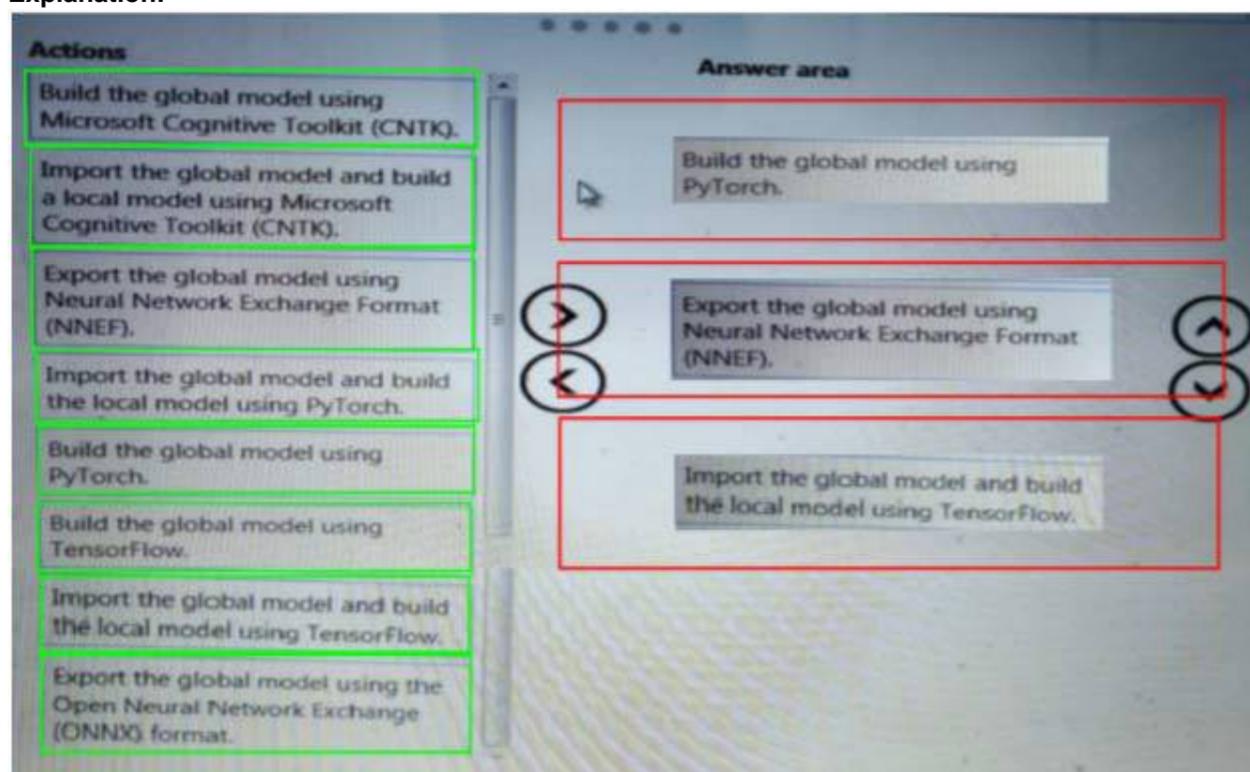
Which three actions should you perform in sequence? To answer, move the appropriate actions from the list of actions to the answer area and arrange them in the correct order.



- A. Mastered
- B. Not Mastered

Answer: A

Explanation:



NEW QUESTION 141

.....

## THANKS FOR TRYING THE DEMO OF OUR PRODUCT

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

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

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

### Money Back Guarantee

#### **DP-100 Practice Exam Features:**

- \* DP-100 Questions and Answers Updated Frequently
- \* DP-100 Practice Questions Verified by Expert Senior Certified Staff
- \* DP-100 Most Realistic Questions that Guarantee you a Pass on Your FirstTry
- \* DP-100 Practice Test Questions in Multiple Choice Formats and Updatesfor 1 Year