

DP-100 Dumps

Designing and Implementing a Data Science Solution on Azure

<https://www.certleader.com/DP-100-dumps.html>



NEW QUESTION 1

- (Exam Topic 3)

You have several machine learning models registered in an Azure Machine Learning workspace. You must use the Fairlearn dashboard to assess fairness in a selected model.

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

Actions

Select a binary classification or regression model.

Select a metric to be measured.

Select a multiclass classification model.

Select a model feature to be evaluated.

Select a clustering model.

Answer Area

- A. Mastered
B. Not Mastered

Answer: A

Explanation:

Graphical user interface, text, application Description automatically generated

Step 1: Select a model feature to be evaluated.

Step 2: Select a binary classification or regression model.

Register your models within Azure Machine Learning. For convenience, store the results in a dictionary, which maps the id of the registered model (a string in name:version format) to the predictor itself. Example:

```
model_dict = {}
```

```
lr_reg_id = register_model("fairness_logistic_regression", lr_predictor) model_dict[lr_reg_id] = lr_predictor
```

```
svm_reg_id = register_model("fairness_svm", svm_predictor) model_dict[svm_reg_id] = svm_predictor
```

Step 3: Select a metric to be measured Precompute fairness metrics.

Create a dashboard dictionary using Fairlearn's metrics package. Reference:

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

NEW QUESTION 2

- (Exam Topic 3)

HOTSPOT

You create a script for training a machine learning model in Azure Machine Learning service. You create an estimator by running the following code:

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

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

NOTE: Each correct selection is worth one point.

	Yes	No
The estimator will look for the files it needs to run an experiment in the training-experiment directory of the local compute environment.	<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 3

- (Exam Topic 3)

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

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

You create a model to forecast weather conditions based on historical data.

You need to create a pipeline that runs a processing script to load data from a datastore and pass the processed data to a machine learning model training script.

Solution: Run the following code:

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

Does the solution meet the goal?

- A. Yes
B. No

Answer: B

Explanation:

train_step is missing. Reference:

<https://docs.microsoft.com/en-us/python/api/azureml-pipeline-core/azureml.pipeline.core.pipelinedata?view=azu>

NEW QUESTION 4

- (Exam Topic 3)

You are using C-Support Vector classification to do a multi-class classification with an unbalanced training dataset. The C-Support Vector classification using Python code shown below:

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

You need to evaluate the C-Support Vector classification code.

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

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

- A. Mastered
B. Not Mastered

Answer: A

Explanation:

Box 1: Automatically adjust weights inversely proportional to class frequencies in the input data

The “balanced” mode uses the values of y to automatically adjust weights inversely proportional to class frequencies in the input data as $n_{\text{samples}} / (n_{\text{classes}} * \text{np.bincount}(y))$.

Box 2: Penalty parameter

Parameter: C : float, optional (default=1.0) Penalty parameter C of the error term. References:

<https://scikit-learn.org/stable/modules/generated/sklearn.svm.SVC.html>

NEW QUESTION 5

- (Exam Topic 3)

You have a dataset that contains over 150 features. You use the dataset to train a Support Vector Machine (SVM) binary classifier.

You need to use the Permutation Feature Importance module in Azure Machine Learning Studio to compute a set of feature importance scores for the dataset.

In which order should you perform the actions? To answer, move all actions from the list of actions to the answer area and arrange them in the correct order.

Actions	Answer Area
Add a Two-Class Support Vector Machine module to initialize the SVM classifier.	
Set the Metric for measuring performance property to Classification - Accuracy and then run the experiment.	
Add a Permutation Feature Importance module and connect the trained model and test dataset.	<div>⬅</div> <div>➡</div>
Add a dataset to the experiment.	<div>⬆</div> <div>⬇</div>
Add a Split Data module to create training and test datasets.	

- A. Mastered
B. Not Mastered

Answer: A

Explanation:

Step 1: Add a Two-Class Support Vector Machine module to initialize the SVM classifier.

Step 2: Add a dataset to the experiment

Step 3: Add a Split Data module to create training and test dataset.

To generate a set of feature scores requires that you have an already trained model, as well as a test dataset. Step 4: Add a Permutation Feature Importance module and connect to the trained model and test dataset. Step 5: Set the Metric for measuring performance property to Classification - Accuracy and then run the experiment.

Reference:

<https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/two-class-support-vector-mac> <https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/permutation-feature-importan>

NEW QUESTION 6

- (Exam Topic 3)

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

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

You plan to use a Python script to run an Azure Machine Learning experiment. The script creates a reference to the experiment run context, loads data from a file, identifies the set of unique values for the label column, and completes the experiment run:

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

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

The experiment must record the unique labels in the data as metrics for the run that can be reviewed later. You must add code to the script to record the unique label values as run metrics at the point indicated by the comment.

Solution: Replace the comment with the following code:

`run.log_list('Label Values', label_vals)` Does the solution meet the goal?

- A. Yes
- B. No

Answer: A

Explanation:

`run.log_list` log a list of values to the run with the given name using `log_list`. Example: `run.log_list("accuracies", [0.6, 0.7, 0.87])`

Note:

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

Data is read into a `pandas.DataFrame`, which is a two-dimensional, size-mutable, potentially heterogeneous tabular data.

`label_vals = data['label'].unique`

`label_vals` contains a list of unique label values. Reference:

<https://www.element61.be/en/resource/azure-machine-learning-services-complete-toolbox-ai> [https://docs.microsoft.com/en-us/python/api/azureml-core/azureml.core.run\(class\)](https://docs.microsoft.com/en-us/python/api/azureml-core/azureml.core.run(class)) <https://pandas.pydata.org/docs/reference/api/pandas.DataFrame.html>

NEW QUESTION 7

- (Exam Topic 3)

You use the Two-Class Neural Network module in Azure Machine Learning Studio to build a binary classification model. You use the Tune Model Hyperparameters module to tune accuracy for the model.

You need to select the hyperparameters that should be tuned using the Tune Model Hyperparameters module. Which two hyperparameters should you use? Each correct answer presents part of the solution.

NOTE: Each correct selection is worth one point.

- A. Number of hidden nodes
- B. Learning Rate
- C. The type of the normalizer
- D. Number of learning iterations
- E. Hidden layer specification

Answer: DE

Explanation:

D: For Number of learning iterations, specify the maximum number of times the algorithm should process the training cases.

E: For Hidden layer specification, select the type of network architecture to create.

Between the input and output layers you can insert multiple hidden layers. Most predictive tasks can be accomplished easily with only one or a few hidden layers.

References:

<https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/two-class-neural-network>

NEW QUESTION 8

- (Exam Topic 3)

A set of CSV files contains sales records. All the CSV files have the same data schema.

Each CSV file contains the sales record for a particular month and has the filename `sales.csv`. Each file is stored in a folder that indicates the month and year when the data was recorded. The folders are in an Azure blob container for which a datastore has been defined in an Azure Machine Learning workspace. The folders are organized in a parent folder named `sales` to create the following hierarchical structure:

```
/sales
  /01-2019
    /sales.csv
  /02-2019
    /sales.csv
  /03-2019
    /sales.csv
  ...
```

At the end of each month, a new folder with that month's sales file is added to the sales folder.

You plan to use the sales data to train a machine learning model based on the following requirements:

- You must define a dataset that loads all of the sales data to date into a structure that can be easily converted to a dataframe.

➤ You must be able to create experiments that use only data that was created before a specific previous month, ignoring any data that was added after that month.

➤ You must register the minimum number of datasets possible.

You need to register the sales data as a dataset in Azure Machine Learning service workspace. What should you do?

- A. Create a tabular dataset that references the datastore and explicitly specifies each 'sales/mm-yyyy/ sales.csv' file every mont
- B. Register the dataset with the name sales_dataset each month, replacing the existing dataset and specifying a tag named month indicating the month and year it was registere
- C. Use this dataset for all experiments.
- D. Create a tabular dataset that references the datastore and specifies the path 'sales/*/sales.csv', register the dataset with the name sales_dataset and a tag named month indicating the month and year it was registered, and use this dataset for all experiments.
- E. Create a new tabular dataset that references the datastore and explicitly specifies each 'sales/mm-yyyy/ sales.csv' file every mont
- F. Register the dataset with the name sales_dataset_MM-YYYY each month with appropriate MM and YYYY values for the month and yea
- G. Use the appropriate month-specific dataset for experiments.
- H. Create a tabular dataset that references the datastore and explicitly specifies each 'sales/mm-yyyy/ sales.csv' fil
- I. Register the dataset with the name sales_dataset each month as a new version and with a tag named month indicating the month and year it was registere
- J. Use this dataset for all experiments, identifying the version to be used based on the month tag as necessary.

Answer: B

Explanation:

Specify the path. Example:

The following code gets the workspace existing workspace and the desired datastore by name. And then passes the datastore and file locations to the path parameter to create a new TabularDataset, weather_ds.

```
from azureml.core import Workspace, Datastore, Dataset
datastore_name = 'your datastore name'
```

```
# get existing workspace
```

```
workspace = Workspace.from_config()
```

```
# retrieve an existing datastore in the workspace by name
datastore = Datastore.get(workspace, datastore_name)
```

```
# create a TabularDataset from 3 file paths in datastore
datastore_paths = [(datastore, 'weather/2018/11.csv'), (datastore, 'weather/2018/12.csv'), (datastore, 'weather/2019/*.csv')]
```

```
weather_ds = Dataset.Tabular.from_delimited_files(path=datastore_paths)
```

NEW QUESTION 9

- (Exam Topic 3)

You use the Azure Machine Learning designer to create and run a training pipeline.

The pipeline must be run every night to inference predictions from a large volume of files. The folder where the files will be stored is defined as a dataset.

You need to publish the pipeline as a REST service that can be used for the nightly inferencing run. What should you do?

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

Answer: A

Explanation:

Azure Machine Learning Batch Inference targets large inference jobs that are not time-sensitive. Batch Inference provides cost-effective inference compute scaling, with unparalleled throughput for asynchronous applications. It is optimized for high-throughput, fire-and-forget inference over large collections of data. You can submit a batch inference job by pipeline_run, or through REST calls with a published pipeline. Reference:

<https://github.com/Azure/MachineLearningNotebooks/blob/master/how-to-use-azureml/machine-learning-pipeli>

NEW QUESTION 10

- (Exam Topic 3)

You create a new Azure subscription. No resources are provisioned in the subscription. You need to create an Azure Machine Learning workspace.

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

- A. Run Python code that uses the Azure ML SDK library and calls the Workspace.create method with name, subscription_id, resource_group, and location parameters.
- B. Use an Azure Resource Management template that includes a Microsoft.MachineLearningServices/ workspaces resource and its dependencies.
- C. Use the Azure Command Line Interface (CLI) with the Azure Machine Learning extension to call the az group create function with --name and --location parameters, and then the az ml workspace create function, specifying -w and -g parameters for the workspace name and resource group.
- D. Navigate to Azure Machine Learning studio and create a workspace.
- E. Run Python code that uses the Azure ML SDK library and calls the Workspace.get method with name, subscription_id, and resource_group parameters.

Answer: BCD

Explanation:

B: You can use an Azure Resource Manager template to create a workspace for Azure Machine Learning. Example:

```
{"type": "Microsoft.MachineLearningServices/workspaces",
```

```
...
```

C: You can create a workspace for Azure Machine Learning with Azure CLI Install the machine learning extension.

Create a resource group: az group create --name <resource-group-name> --location <location>

To create a new workspace where the services are automatically created, use the following command: az ml workspace create -w <workspace-name> -g <resource-group-name>

D: You can create and manage Azure Machine Learning workspaces in the Azure portal.

➤ Sign in to the Azure portal by using the credentials for your Azure subscription.

➤ In the upper-left corner of Azure portal, select + Create a resource.

➤ Use the search bar to find Machine Learning.

➤ Select Machine Learning.

➤ In the Machine Learning pane, select Create to begin.

[Home](#) > [New](#) > [Machine Learning](#) >

Machine Learning

Create a machine learning workspace

[Basics](#) [Networking](#) [Advanced](#) [Tags](#) [Review + create](#)

Project details

Select the subscription to manage deployed resources and costs. Use resource groups like folders to organize and manage all your resources.

Subscription * ⓘ

Resource group * ⓘ [Create new](#)

Workspace details


Specify the name, region, and edition for the workspace.

Workspace name * ⓘ

Region * ⓘ

Workspace edition * ⓘ

- Basic
- Enterprise

 For your convenience, these resources are available: Application Insights, Azure Key Vault

Reference:

<https://docs.microsoft.com/en-us/azure/machine-learning/how-to-create-workspace-template> <https://docs.microsoft.com/en-us/azure/machine-learning/how-to-manage-workspace-cli> <https://docs.microsoft.com/en-us/azure/machine-learning/how-to-manage-workspace>

NEW QUESTION 10

- (Exam Topic 3)

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

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

You are analyzing a numerical dataset which contains missing values in several columns.

You must clean the missing values using an appropriate operation without affecting the dimensionality of the feature set.

You need to analyze a full dataset to include all values.

Solution: Remove the entire column that contains the missing data point. Does the solution meet the goal?

- A. Yes
- B. No

Answer: B

Explanation:

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

NEW QUESTION 12

- (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 14

- (Exam Topic 3) You are solving a classification task. The dataset is imbalanced.

You need to select an Azure Machine Learning Studio module to improve the classification accuracy. Which module should you use?

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

Answer: C

Explanation:

Use the SMOTE module in Azure Machine Learning Studio (classic) to increase the number of 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.

You connect the SMOTE module to a dataset that is imbalanced. There are many reasons why a dataset might be imbalanced: the category you are targeting might be very rare in the population, or the data might simply be difficult to collect. Typically, you use SMOTE when the class you want to analyze is under-represented.

Reference:

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

NEW QUESTION 18

- (Exam Topic 3)

You are moving a large dataset from Azure Machine Learning Studio to a Weka environment. You need to format the data for the Weka environment.

Which module should you use?

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

Answer: C

Explanation:

Use the Convert to ARFF module in Azure Machine Learning Studio, to convert datasets and results in Azure Machine Learning to the attribute-relation file format used by the Weka toolset. This format is known as ARFF.

The ARFF data specification for Weka supports multiple machine learning tasks, including data preprocessing, classification, and feature selection. In this format, data is organized by entites and their attributes, and is contained in a single text file.

References:

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

NEW QUESTION 22

- (Exam Topic 3)

You deploy a model in Azure Container Instance.

You must use the Azure Machine Learning SDK to call the model API.

You need to invoke the deployed model using native SDK classes and methods.

How should you complete the command? To answer, select the appropriate options in the answer areas.

NOTE: Each correct selection is worth one point.

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

```
from azureml.core.webservice import requests  
from azureml.core.webservice import Webservice  
from azureml.core.webservice import LocalWebservice
```

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

```
predictions = service.run(input_json)  
predictions = requests.post(service.scoring_uri, input_json)  
predictions = service.deserialize(ws, input_json)
```

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

Box 1: from azureml.core.webservice import Webservice

The following code shows how to use the SDK to update the model, environment, and entry script for a web service to Azure Container Instances:

from azureml.core import Environment

from azureml.core.webservice import Webservice

from azureml.core.model import Model, InferenceConfig Box 2: predictions = service.run(input_json)

Example: The following code demonstrates sending data to the service: import json

test_sample = json.dumps({'data': [[1, 2, 3, 4, 5, 6, 7, 8, 9, 10],

[10, 9, 8, 7, 6, 5, 4, 3, 2, 1]

]])

test_sample = bytes(test_sample, encoding='utf8') prediction = service.run(input_data=test_sample)

print(prediction) Reference:

<https://docs.microsoft.com/bs-latn-ba/azure/machine-learning/how-to-deploy-azure-container-instance> <https://docs.microsoft.com/en-us/azure/machine-learning/how-to-troubleshoot-deployment>

NEW QUESTION 23

- (Exam Topic 3)

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

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

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

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

```
from azureml.train.hyperdrive import RandomParameterSampling
```

```
param_sampling = RandomParameterSampling( {
```

```
"learning_rate" :
```

uniform(10,3)
normal(10,3)
choice(10,3)
Loguniform(10,3)

```
"batch_size":
```

choice(16,32,64)
choice(range(16,64))
normal(16,32,64)
normal(range(16,64))

```
"keep_probability" :
```

choice(range(0.05, 0.1))
uniform(0.05, 0.1)
normal(0.05, 0.1)
lognormal(0.05, 0.1)

```
    }
```

```
)
```

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

In random sampling, hyperparameter values are randomly selected from the defined search space. Random sampling allows the search space to include both discrete and continuous hyperparameters.

Example:

```
from azureml.train.hyperdrive import RandomParameterSampling param_sampling = RandomParameterSampling( { "learning_rate": normal(10, 3),
```

```
"keep_probability": uniform(0.05, 0.1),
```

```
"batch_size": choice(16, 32, 64)
```

```
}
```

Reference:

<https://docs.microsoft.com/en-us/azure/machine-learning/service/how-to-tune-hyperparameters>

NEW QUESTION 26

- (Exam Topic 3)

You use the Azure Machine Learning service to create a tabular dataset named training.data. You plan to use this dataset in a training script.

You create a variable that references the dataset using the following code: training_ds = workspace.datasets.get("training_data")

You define an estimator to run the script.

You need to set the correct property of the estimator to ensure that your script can access the training.data dataset
Which property should you set?

A)

```
inputs = [training_ds.as_named_input('training_ds')]
```

B)

```
script_params = {"--training_ds":training_ds}
```

C)

```
environment_definition = {"training_data":training_ds}
```

D)

```
source_directory = training_ds
```

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

Answer: A

Explanation:

Example:

Get the training dataset

```
diabetes_ds = ws.datasets.get("Diabetes Dataset")
```

Create an estimator that uses the remote compute hyper_estimator = SKLearn(source_directory=experiment_folder,

inputs=[diabetes_ds.as_named_input('diabetes')], # Pass the dataset as an input compute_target = cpu_cluster, conda_packages=['pandas','ipykernel','matplotlib'],

pip_packages=['azureml-sdk','argparse','pyarrow'], entry_script='diabetes_training.py')

Reference:

<https://notebooks.azure.com/GraemeMalcolm/projects/azureml-primers/html/04%20-%20Optimizing%20Model>

NEW QUESTION 28

- (Exam Topic 3)

You are evaluating a completed binary classification machine learning model. You need to use the precision as the valuation metric.

Which visualization should you use?

- A. Binary classification confusion matrix
- B. box plot
- C. Gradient descent
- D. coefficient of determination

Answer: A

Explanation:

References:

<https://machinelearningknowledge.ai/confusion-matrix-and-performance-metrics-machine-learning/>

NEW QUESTION 30

- (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>Classification</div> <div>Forecasting</div> <div>Regression</div> <div>Outlier</div>
Target column	<div>Price</div> <div>Bedrooms</div> <div>Size</div> <div>HasGarage</div> <div>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 31

- (Exam Topic 3)

You have a multi-class image classification deep learning model that uses a set of labeled photographs. You create the following code to select hyperparameter values when training the model.

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

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

NOTE: Each correct selection is worth one point.

	Yes	No
Hyperparameter combinations for the runs are selected based on how previous samples performed in the previous experiment run.	<input type="radio"/>	<input type="radio"/>
The learning rate value 0.09 might be used during model training.	<input type="radio"/>	<input type="radio"/>
You can define an early termination policy for this hyperparameter tuning run.	<input type="radio"/>	<input type="radio"/>

- A. Mastered
B. Not Mastered

Answer: A

Explanation:

Box 1: Yes

Hyperparameters are adjustable parameters you choose to train a model that govern the training process itself. Azure Machine Learning allows you to automate hyperparameter exploration in an efficient manner, saving you significant time and resources. You specify the range of hyperparameter values and a maximum number of training runs. The system then automatically launches multiple simultaneous runs with different parameter configurations and finds the configuration that results in the best performance, measured by the metric you choose. Poorly performing training runs are automatically early terminated, reducing wastage of compute resources. These resources are instead used to explore other hyperparameter configurations.

Box 2: Yes

uniform(low, high) - Returns a value uniformly distributed between low and high Box 3: No

Bayesian sampling does not currently support any early termination policy. Reference:

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

NEW QUESTION 32

- (Exam Topic 3)

You use Azure Machine Learning designer to create a training pipeline for a regression model.

You need to prepare the pipeline for deployment as an endpoint that generates predictions asynchronously for a dataset of input data values.

What should you do?

- A. Clone the training pipeline.
- B. Create a batch inference pipeline from the training pipeline.
- C. Create a real-time inference pipeline from the training pipeline.
- D. Replace the dataset in the training pipeline with an Enter Data Manually module.

Answer: C

Explanation:

You must first convert the training pipeline into a real-time inference pipeline. This process removes training modules and adds web service inputs and outputs to handle requests.

Reference:

<https://docs.microsoft.com/en-us/azure/machine-learning/tutorial-designer-automobile-price-deploy> <https://docs.microsoft.com/en-us/azure/machine-learning/algorithm-module-reference/enter-data-manually>

NEW QUESTION 35

- (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 37

- (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 40

- (Exam Topic 3)

A coworker registers a datastore in a Machine Learning services workspace by using the following code:

```
Datastore.register_azure_blob_container(workspace=ws,
datastore_name='demo_datastore',
container_name='demo_datacontainer',
account_name='demo_account',
account_key='0A0A0A-0A0A00A-0A00A0A0A0A',
create_if_not_exists=True)
```

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

Answer Area

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

datastore =
```

Workspace
Datastore
Experiment
Run

```
.get(
```

ws
run
experiment
log

```
, '
```

demo_datastore
demo_datacontainer
demo_account
Datastore

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

Box 1: DataStore

To get a specific datastore registered in the current workspace, use the get() static method on the Datastore class:

Get a named datastore from the current workspace

datastore = Datastore.get(ws, datastore_name='your datastore name')

Box 2: ws

Box 3: demo_datastore Reference:

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

NEW QUESTION 44

- (Exam Topic 3)

You use the following code to run a script as an experiment in Azure Machine Learning:

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

You must identify the output files that are generated by the experiment run. You need to add code to retrieve the output file names.

Which code segment should you add to the script?

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

Answer: B

Explanation:

You can list all of the files that are associated with this run record by called run.get_file_names() Reference:

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

NEW QUESTION 47

- (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 49

- (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.

Target datasets	Answer Area
training-dataset	Baseline dataset
predictions-dataset	Target dataset
features-dataset	Target dataset

- 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 52

- (Exam Topic 3)

You are creating an experiment by using Azure Machine Learning Studio.

You must divide the data into four subsets for evaluation. There is a high degree of missing values in the data. You must prepare the data for analysis.

You need to select appropriate methods for producing the experiment.

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

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

Actions	Answer Area
Build Counting Transform	
Missing Values Scrubber	
Feature Hashing	
Clean Missing Data	⬅️ ⬆️
Replace Discrete Values	➡️ ⬇️
Import Data	
Latent Dirichlet Transformation	
Partition and Sample	

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

The Clean Missing Data module in Azure Machine Learning Studio, to remove, replace, or infer missing values.

NEW QUESTION 56

- (Exam Topic 3)

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

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

You train and register a machine learning model.

You plan to deploy the model as a real-time web service. Applications must use key-based authentication to use the model.

You need to deploy the web service. Solution:

Create an AksWebservice instance.

Set the value of the auth_enabled property to True.

Deploy the model to the service. Does the solution meet the goal?

A. Yes

B. No

Answer: A

Explanation:

Key-based authentication.

Web services deployed on AKS have key-based auth enabled by default. ACI-deployed services have key-based auth disabled by default, but you can enable it by setting auth_enabled = TRUE when creating the ACI web service. The following is an example of creating an ACI deployment configuration with key-based auth enabled.

```
deployment_config <- aci_webservice_deployment_config(cpu_cores = 1, memory_gb = 1,  
auth_enabled = TRUE)
```

 Reference:

<https://azure.github.io/azureml-sdk-for-r/articles/deploying-models.html>

NEW QUESTION 58

- (Exam Topic 3)

You are building an intelligent solution using machine learning models. The environment must support the following requirements:

- Data scientists must build notebooks in a cloud environment
- Data scientists must use automatic feature engineering and model building in machine learning pipelines.
- Notebooks must be deployed to retrain using Spark instances with dynamic worker allocation.
- Notebooks must be exportable to be version controlled locally.

You need to create the environment.

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

Actions

Install the Azure Machine Learning SDK for Python on the cluster.

When the cluster is ready, export Zeppelin notebooks to a local environment.

Create and execute a Jupyter notebook by using automated machine learning (AutoML) on the cluster.

Install Microsoft Machine Learning for Apache Spark.

When the cluster is ready and has processed the notebook, export your Jupyter notebook to a local environment.

Create an Azure HDInsight cluster to include the Apache Spark Mlib library.

Create and execute the Zeppelin notebooks on the cluster.

Create an Azure Databricks cluster.

Answer area



A. Mastered

B. Not Mastered

Answer: A

Explanation:

Step 1: Create an Azure HDInsight cluster to include the Apache Spark Mlib library

Step 2: Install Microsoft Machine Learning for Apache Spark You install AzureML on your Azure HDInsight cluster.

Microsoft Machine Learning for Apache Spark (MMLSpark) provides a number of deep learning and data science tools for Apache Spark, including seamless integration of Spark Machine Learning pipelines with Microsoft Cognitive Toolkit (CNTK) and OpenCV, enabling you to quickly create powerful, highly-scalable predictive and analytical models for large image and text datasets.

Step 3: Create and execute the Zeppelin notebooks on the cluster

Step 4: When the cluster is ready, export Zeppelin notebooks to a local environment. Notebooks must be exportable to be version controlled locally.

References:

<https://docs.microsoft.com/en-us/azure/hdinsight/spark/apache-spark-zeppelin-notebook> <https://azuremlbuild.blob.core.windows.net/pysparkapi/intro.html>

NEW QUESTION 60

- (Exam Topic 3)

You train and register a machine learning model. You create a batch inference pipeline that uses the model to generate predictions from multiple data files. You must publish the batch inference pipeline as a service that can be scheduled to run every night. You need to select an appropriate compute target for the inference service.

Which compute target should you use?

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

Answer: B

Explanation:

Azure Machine Learning compute clusters is used for Batch inference. Run batch scoring on serverless compute. Supports normal and low-priority VMs. No support for real-time inference.

Reference:

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

NEW QUESTION 64

- (Exam Topic 3)

You are implementing a machine learning model to predict stock prices. The model uses a PostgreSQL database and requires GPU processing.

You need to create a virtual machine that is pre-configured with the required tools. What should you do?

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

Answer: E

NEW QUESTION 67

- (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 70

- (Exam Topic 3)

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

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

You are creating a model to predict the price of a student's artwork depending on the following variables: the student's length of education, degree type, and art form.

You start by creating a linear regression model. You need to evaluate the linear regression model.

Solution: Use the following metrics: Mean Absolute Error, Root Mean Absolute Error, Relative Absolute Error, Relative Squared Error, and the Coefficient of Determination.

Does the solution meet the goal?

- A. Yes
- B. No

Answer: A

Explanation:

The following metrics are reported for evaluating regression models. When you compare models, they are ranked by the metric you select for evaluation.

Mean absolute error (MAE) measures how close the predictions are to the actual outcomes; thus, a lower score is better.

Root mean squared error (RMSE) creates a single value that summarizes the error in the model. By squaring the difference, the metric disregards the difference between over-prediction and under-prediction.

Relative absolute error (RAE) is the relative absolute difference between expected and actual values; relative because the mean difference is divided by the arithmetic mean.

Relative squared error (RSE) similarly normalizes the total squared error of the predicted values by dividing by the total squared error of the actual values.

Mean Zero One Error (MZOE) indicates whether the prediction was correct or not. In other words: $\text{ZeroOneLoss}(x,y) = 1$ when $x \neq y$; otherwise 0.

Coefficient of determination, often referred to as R^2 , represents the predictive power of the model as a value between 0 and 1. Zero means the model is random (explains nothing); 1 means there is a perfect fit. However, caution should be used in interpreting R^2 values, as low values can be entirely normal and high values can be suspect.

AUC.

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

NEW QUESTION 73

- (Exam Topic 3)

You train and register a model by using the Azure Machine Learning SDK on a local workstation. Python 3.6 and Visual Studio Code are installed on the workstation.

When you try to deploy the model into production as an Azure Kubernetes Service (AKS)-based web service, you experience an error in the scoring script that causes deployment to fail.

You need to debug the service on the local workstation before deploying the service to production.

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

Actions	Answer Area
Create an AksWebservice deployment configuration for the service and deploy the model to it	
Install Docker on the workstation	
Create a LocalWebservice deployment configuration for the service and deploy the model to it	
Debug and modify the scoring script as necessary. Use the reload() method of the service after each modification	
Create an AciWebservice deployment configuration for the service and deploy the model to it	

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

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

Step 1: Install Docker on the workstation

Prerequisites include having a working Docker installation on your local system. Build or download the dockerfile to the compute node.

Step 2: Create an AksWebservice deployment configuration and deploy the model to it

To deploy a model to Azure Kubernetes Service, create a deployment configuration that describes the compute resources needed.

If deploying to a cluster configured for dev/test, ensure that it was created with enough

cores and memory to handle this deployment configuration. Note that memory is also used by

things such as dependencies and AML components.

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

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

```
print(service.state) print(service.get_logs())
```

Step 3: Create a LocalWebservice deployment configuration for the service and deploy the model to it

To deploy locally, modify your code to use LocalWebservice.deploy_configuration() to create a deployment configuration. Then use Model.deploy() to deploy the service.

Step 4: Debug and modify the scoring script as necessary. Use the reload() method of the service after each modification.

During local testing, you may need to update the score.py file to add logging or attempt to resolve any problems that you've discovered. To reload changes to the score.py file, use reload(). For example, the following code reloads the script for the service, and then sends data to it.

Reference:

<https://docs.microsoft.com/en-us/azure/machine-learning/how-to-deploy-azure-kubernetes-service> <https://docs.microsoft.com/en-us/azure/machine-learning/how-to-troubleshoot-deployment-local>

NEW QUESTION 74

- (Exam Topic 3)

You have a Python script that executes a pipeline. The script includes the following code: from azureml.core import Experiment

pipeline_run = Experiment(ws, 'pipeline_test').submit(pipeline) You want to test the pipeline before deploying the script.

You need to display the pipeline run details written to the STDOUT output when the pipeline completes. Which code segment should you add to the test script?

- A. pipeline_run.get.metrics()
- B. pipeline_run.wait_for_completion(show_output=True)
- C. pipeline_param = PipelineParameter(name="stdout", default_value="console")
- D. pipeline_run.get_status()

Answer: B

Explanation:

wait_for_completion: Wait for the completion of this run. Returns the status object after the wait. Syntax: wait_for_completion(show_output=False, wait_post_processing=False, raise_on_error=True) Parameter: show_output
Indicates whether to show the run output on sys.stdout.

NEW QUESTION 75

- (Exam Topic 3)

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

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

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

You need to configure the module.

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

▲ 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 78

- (Exam Topic 3)

You are analyzing a dataset containing historical data from a local taxi company. You arc developing a regression a regression model.

You must predict the fare of a taxi trip.

You need to select performance metrics to correctly evaluate the- regression model. Which two metrics can you use? Each correct answer presents a complete solution. NOTE: Each correct selection is worth one point.

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

Answer: BE

Explanation:

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

NEW QUESTION 81

- (Exam Topic 3)

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

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

An IT department creates the following Azure resource groups and resources:

Resource group	Resources
ml_resources	<ul style="list-style-type: none"> an Azure Machine Learning workspace named amlworkspace an Azure Storage account named amlworkspace12345 an Application Insights instance named amlworkspace54321 an Azure Key Vault named amlworkspace67890 an Azure Container Registry named amlworkspace09876
general_compute	<p>A virtual machine named mlvm with the following configuration:</p> <ul style="list-style-type: none"> Operating system: Ubuntu Linux Software installed: Python 3.6 and Jupyter Notebooks Size: NC6 (6 vCPUs, 1 vGPU, 56 Gb RAM)

The IT department creates an Azure Kubernetes Service (AKS)-based inference compute target named aks-cluster in the Azure Machine Learning workspace. You have a Microsoft Surface Book computer with a GPU. Python 3.6 and Visual Studio Code are installed. You need to run a script that trains a deep neural network (DNN) model and logs the loss and accuracy metrics.

Solution: Install the Azure ML SDK on the Surface Book. Run Python code to connect to the workspace and then run the training script as an experiment on local compute.

- A. Yes
- B. No

Answer: B

Explanation:

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

NEW QUESTION 83

- (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 86

- (Exam Topic 3)

You must store data in Azure Blob Storage to support Azure Machine Learning. You need to transfer the data into Azure Blob Storage.

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

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

Answer: BCD

Explanation:

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

NEW QUESTION 91

- (Exam Topic 3)

You are evaluating a Python NumPy array that contains six data points defined as follows: data = [10, 20, 30, 40, 50, 60]

You must generate the following output by using the k-fold algorithm implantation in the Python Scikit-learn machine learning library:

train: [10 40 50 60], test: [20 30]

train: [20 30 40 60], test: [10 50]

train: [10 20 30 50], test: [40 60]

You need to implement a cross-validation to generate the output.

How should you complete the code segment? To answer, select the appropriate code segment in the dialog box in the answer area.

NOTE: Each correct selection is worth one point.

```
from numpy import array
from sklearn.model_selection import K-Means
k-fold
CrossValidation
ModelSelection

data = array([10, 20, 30, 40, 50, 60])
kfold = Kfold(n_splits=1
2
3
6, shuffle = True, random_state=1)

for train, test in kFold, split(data
k-fold
array
train, test):

print('train: %s, test: %s' % (data[train], data[test]))
```

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

Box 1: k-fold

Box 2: 3

K-F olds cross-validator provides train/test indices to split data in train/test sets. Split dataset into k consecutive folds (without shuffling by default).

The parameter n_splits (int, default=3) is the number of folds. Must be at least 2. Box 3: data

Example: Example:

>>>

>>> from sklearn.model_selection import KFold

>>> X = np.array([[1, 2], [3, 4], [1, 2], [3, 4]])

>>> y = np.array([1, 2, 3, 4])

>>> kf = KFold(n_splits=2)

>>> kf.get_n_splits(X) 2

>>> print(kf)

KFold(n_splits=2, random_state=None, shuffle=False)

>>> for train_index, test_index in kf.split(X): print("TRAIN:", train_index, "TEST:", test_index) X_train, X_test = X[train_index], X[test_index] y_train, y_test =

y[train_index], y[test_index] TRAIN: [2 3] TEST: [0 1]

TRAIN: [0 1] TEST: [2 3]

References:

https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.KFold.html

NEW QUESTION 94

- (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=1
- B. k=10
- C. k=0.5
- D. k=0.9

Answer: B

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 97

- (Exam Topic 3)

You are using Azure Machine Learning to train machine learning models. You need a compute target on which to remotely run the training script. You run the following Python code:

```
from azureml.core.compute import ComputeTarget, AmlCompute
from azureml.core.compute_target import ComputeTargetException
the_cluster_name = "NewCompute"
config = AmlCompute.provisioning_configuration(vm_size='STANDARD_D2', max_nodes=3)
the_cluster = ComputeTarget.create(ws, the_cluster_name, config)
```

Answer Area

	Yes	No
The compute is created in the same region as the Machine Learning service workspace.	<input checked="" type="checkbox"/>	<input type="checkbox"/>
The compute resource created by the code is displayed as a compute cluster in Azure Machine Learning studio.	<input type="checkbox"/>	<input type="checkbox"/>
The minimum number of nodes will be zero.	<input type="checkbox"/>	<input type="checkbox"/>

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

Box 1: Yes

The compute is created within your workspace region as a resource that can be shared with other users. Box 2: Yes

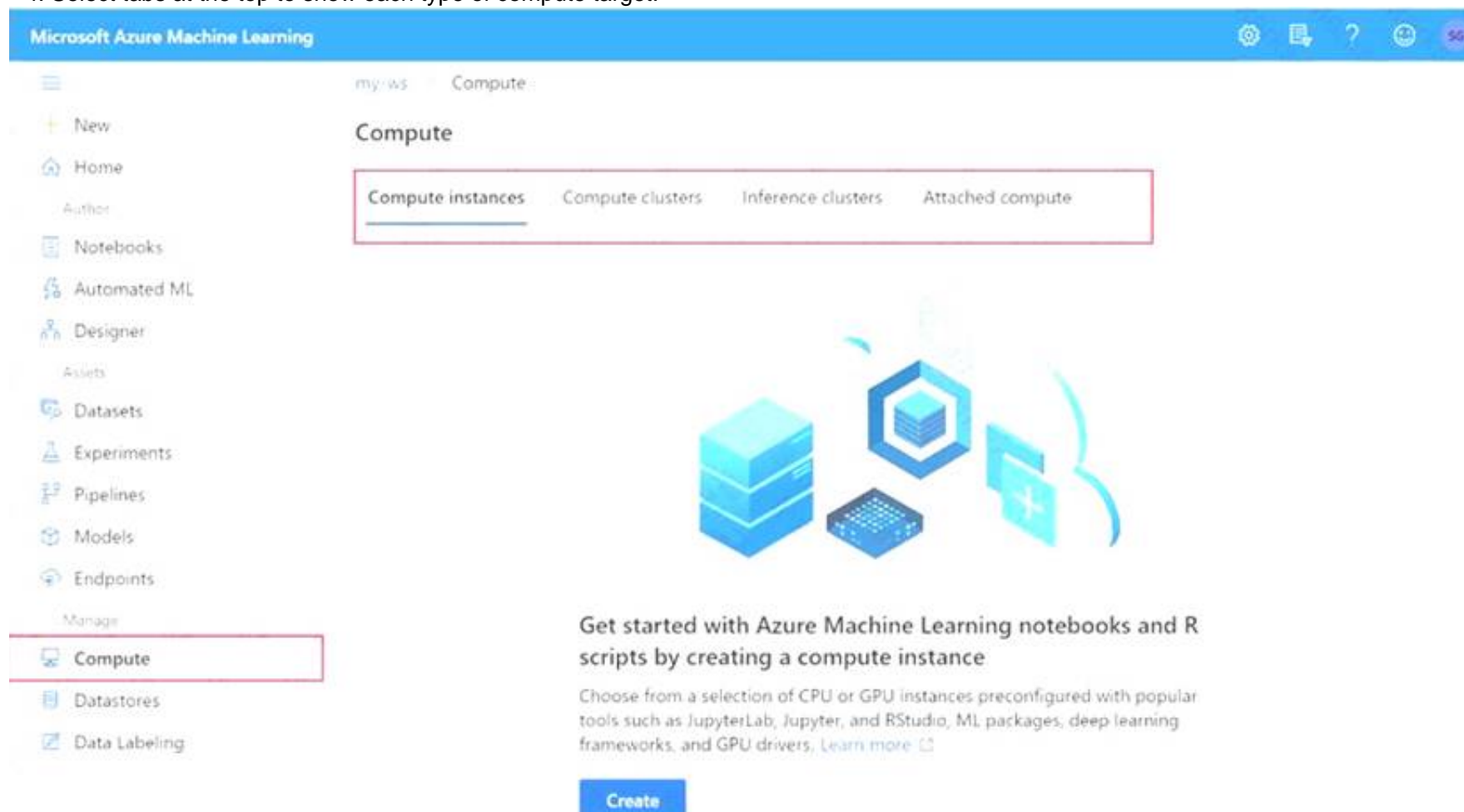
It is displayed as a compute cluster. View compute targets

* 1. To see all compute targets for your workspace, use the following steps:

* 2. Navigate to Azure Machine Learning studio.

* 3. Under Manage, select Compute.

* 4. Select tabs at the top to show each type of compute target.



Box 3: Yes

min_nodes is not specified, so it defaults to 0. Reference:

<https://docs.microsoft.com/en-us/python/api/azureml-core/azureml.core.compute.amlcompute.amlcomputeprovi> <https://docs.microsoft.com/en-us/azure/machine->

learning/how-to-create-attach-compute-studio

NEW QUESTION 100

- (Exam Topic 3)

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

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

You plan to use a Python script to run an Azure Machine Learning experiment. The script creates a reference to the experiment run context, loads data from a file, identifies the set of unique values for the label column, and completes the experiment run:

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

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

The experiment must record the unique labels in the data as metrics for the run that can be reviewed later.

You must add code to the script to record the unique label values as run metrics at the point indicated by the comment.

Solution: Replace the comment with the following code: for label_val in label_vals:

run.log('Label Values', label_val) Does the solution meet the goal?

- A. Yes
- B. No

Answer: A

Explanation:

The run_log function is used to log the contents in label_vals: for label_val in label_vals:

run.log('Label Values', label_val) Reference:

<https://www.element61.be/en/resource/azure-machine-learning-services-complete-toolbox-ai>

NEW QUESTION 102

- (Exam Topic 3)

You have an Azure Machine Learning workspace named workspace1 that is accessible from a public endpoint. The workspace contains an Azure Blob storage datastore named store1 that represents a blob container in an Azure storage account named account1. You configure workspace1 and account1 to be accessible by using private endpoints in the same virtual network.

You must be able to access the contents of store1 by using the Azure Machine Learning SDK for Python. You must be able to preview the contents of store1 by using Azure Machine Learning studio.

You need to configure store1.

What should you do? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.

Requirement	Action
Access the contents of store1 by using the Azure Machine Learning SDK for Python.	<div> <div>Set store1 as the default datastore.</div> <div>Disable data validation for store1.</div> <div>Update authentication for store1.</div> <div>Regenerate the keys of account1.</div> </div>
Preview the contents of store1 by using Azure Machine Learning studio.	<div> <div>Set store1 as the default datastore.</div> <div>Disable data validation for store1.</div> <div>Update authentication for store1.</div> <div>Regenerate the keys of account1.</div> </div>

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

Text, table Description automatically generated

Box 1: Regenerate the keys of account1.

Azure Blob Storage support authentication through Account key or SAS token.

To authenticate your access to the underlying storage service, you can provide either your account key, shared access signatures (SAS) tokens, or service principal

Box 2: Update the authentication for store1.

For Azure Machine Learning studio users, several features rely on the ability to read data from a dataset; such as dataset previews, profiles and automated machine learning. For these features to work with storage behind virtual networks, use a workspace managed identity in the studio to allow Azure Machine Learning to access the storage account from outside the virtual network.

Note: Some of the studio's features are disabled by default in a virtual network. To re-enable these features, you must enable managed identity for storage accounts you intend to use in the studio.

The following operations are disabled by default in a virtual network:

➤ Preview data in the studio.

Reference:

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

NEW QUESTION 104

- (Exam Topic 3)

You plan to run a Python script as an Azure Machine Learning experiment. The script contains the following code:

```
import os, argparse, glob
```

```
from azureml.core import Run parser = argparse.ArgumentParser() parser.add_argument('--input-data',
```

```
type=str, dest='data_folder') args = parser.parse_args() data_path = args.data_folder
```

```
file_paths = glob.glob(data_path + "/*.jpg")
```

You must specify a file dataset as an input to the script. The dataset consists of multiple large image files and must be streamed directly from its source.

You need to write code to define a ScriptRunConfig object for the experiment and pass the ds dataset as an argument.

Which code segment should you use?

A. arguments = ['--input-data', ds.to_pandas_dataframe()]

B. arguments = ['--input-data', ds.as_mount()]

C. arguments = ['--data-data', ds]

D. arguments = ['--input-data', ds.as_download()]

Answer: A

Explanation:

If you have structured data not yet registered as a dataset, create a TabularDataset and use it directly in your training script for your local or remote experiment.

To load the TabularDataset to pandas DataFrame df = dataset.to_pandas_dataframe()

Note: TabularDataset represents data in a tabular format created by parsing the provided file or list of files. Reference:

<https://docs.microsoft.com/en-us/azure/machine-learning/how-to-train-with-datasets>

NEW QUESTION 105

- (Exam Topic 3)

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

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

A. the Activity Log in the Azure portal for the Machine Learning workspace

B. a file named parallel_run_step.txt located in the output folder

C. the digitidentification.py script

D. the Inference Clusters tab in Machine Learning studio

E. the debug log

Answer: B

Explanation:

output_action (str): How the output is to be organized. Currently supported values are 'append_row' and 'summary_only'.

> 'append_row' – All values output by run() method invocations will be aggregated into one unique file named parallel_run_step.txt that is created in the output location.

> 'summary_only' Reference:

<https://docs.microsoft.com/en-us/python/api/azureml-contrib-pipeline-steps/azureml.contrib.pipeline.steps.parall>

NEW QUESTION 109

- (Exam Topic 3)

You are a data scientist building a deep convolutional neural network (CNN) for image classification. The CNN model you built shows signs of overfitting.

You need to reduce overfitting and converge the model to an optimal fit.

Which two actions should you perform? Each correct answer presents a complete solution.

NOTE: Each correct selection is worth one point.

A. Reduce the amount of training data.

B. Add an additional dense layer with 64 input units

C. Add L1/L2 regularization.

D. Use training data augmentation

E. Add an additional dense layer with 512 input units.

Answer: AC

Explanation:

References:

<https://machinelearningmastery.com/how-to-reduce-overfitting-in-deep-learning-with-weight-regularization/>

https://en.wikipedia.org/wiki/Convolutional_neural_network

NEW QUESTION 114

- (Exam Topic 3)

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

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

You are a data scientist using Azure Machine Learning Studio.

You need to normalize values to produce an output column into bins to predict a target column. Solution: Apply a Quantiles normalization with a QuantileIndex normalization.

Does the solution meet the GOAL?

- A. Yes
- B. No

Answer: B

Explanation:

Use the Entropy MDL binning mode which has a target column. References:

<https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/group-data-into-bins>

NEW QUESTION 115

- (Exam Topic 3)

You have the following code. The code prepares an experiment to run a script:

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

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

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

The experiment must be run on local computer using the default environment. You need to add code to start the experiment and run the script.

Which code segment should you use?

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

Answer: D

Explanation:

The experiment class submit method submits an experiment and return the active created run.

Syntax: submit(config, tags=None, **kwargs) Reference:

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

NEW QUESTION 118

- (Exam Topic 3)

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

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

You are using Azure Machine Learning Studio to perform feature engineering on a dataset. You need to normalize values to produce a feature column grouped into bins.

Solution: Apply an Entropy Minimum Description Length (MDL) binning mode. Does the solution meet the goal?

- A. Yes
- B. No

Answer: A

Explanation:

Entropy MDL binning mode: This method requires that you select the column you want to predict and the column or columns that you want to group into bins. It then makes a pass over the data and attempts to determine the number of bins that minimizes the entropy. In other words, it chooses a number of bins that allows the data column to best predict the target column. It then returns the bin number associated with each row of your data in a column named <colname>quantized.

References:

<https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/group-data-into-bins>

NEW QUESTION 122

- (Exam Topic 3)

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

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

Answer: C

Explanation:

Azure Kubernetes Service (AKS) can be used real-time inference. Reference:

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

NEW QUESTION 127

- (Exam Topic 3)

You are developing a linear regression model in Azure Machine Learning Studio. You run an experiment to compare different algorithms. The following image displays the results dataset output:

Algorithm	Mean Absolute Error	Root Mean Squared Error	Relative Absolute Error	Relative Squared Error
Bayesian Linear	3.276025	4.655442	0.511436	0.282138
Neural Network	2.676538	3.621476	0.417847	0.17073
Boosted Decision Tree	2.168847	2.878077	0.338589	0.107831
Linear	6.350005	8.720718	0.99133	0.99002
Decision Forest	2.390206	3.315164	0.373146	0.14307

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

NOTE: Each correct selection is worth one point.

Question	Answer choice
Which algorithm minimizes differences between actual and predicted values?	<div><div></div><div>Bayesian Linear Regression</div><div>Neural Network Regression</div><div>Boosted Decision Tree Regression</div><div>Linear Regression</div><div>Decision Forest Regression</div></div>
Which approach should you use to find the best parameters for a Linear Regression model for the Online Gradient Descent method?	<div><div></div><div>Set the Decrease learning rate option to True.</div><div>Set the Decrease learning rate option to True.</div><div>Set the Create trainer mode option to Parameter Range.</div><div>Increase the number of epochs.</div><div>Decrease the number of epochs.</div></div>

- A. Mastered
B. Not Mastered

Answer: A

Explanation:

Box 1: Boosted Decision Tree Regression

Mean absolute error (MAE) measures how close the predictions are to the actual outcomes; thus, a lower score is better.

Box 2:

Online Gradient Descent: If you want the algorithm to find the best parameters for you, set Create trainer mode option to Parameter Range. You can then specify multiple values for the algorithm to try.

References:

<https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/evaluate-model> <https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/linear-regression>

NEW QUESTION 132

- (Exam Topic 3)

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

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

You create an Azure Machine Learning service datastore in a workspace. The datastore contains the following files:

- /data/2018/Q1.csv
- /data/2018/Q2.csv
- /data/2018/Q3.csv
- /data/2018/Q4.csv
- /data/2019/Q1.csv

All files store data in the following format:

- id,f1,f2,l
- 1,1,2,0
- 2,1,1,1
- 3.2.1.0

You run the following code:

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

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

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

Solution: Run the following code:

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

Does the solution meet the goal?

- A. Yes
- B. No

Answer: A

Explanation:

Use two file paths.

Use Dataset.Tabular_from_delimited as the data isn't cleansed. Note:

A TabularDataset represents data in a tabular format by parsing the provided file or list of files. This provides you with the ability to materialize the data into a pandas or Spark DataFrame so you can work with familiar data preparation and training libraries without having to leave your notebook. You can create a TabularDataset object from .csv, .tsv, .parquet, .jsonl files, and from SQL query results.

Reference:

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

NEW QUESTION 137

- (Exam Topic 3)

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

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

You are creating a model to predict the price of a student's artwork depending on the following variables: the student's length of education, degree type, and art form.

You start by creating a linear regression model. You need to evaluate the linear regression model.

Solution: Use the following metrics: Mean Absolute Error, Root Mean Absolute Error, Relative Absolute Error, Accuracy, Precision, Recall, F1 score, and AUC.

Does the solution meet the goal?

- A. Yes
- B. No

Answer: B

Explanation:

Accuracy, Precision, Recall, F1 score, and AUC are metrics for evaluating classification models. Note: Mean Absolute Error, Root Mean Absolute Error, Relative Absolute Error are OK for the linear regression model.

References:

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

NEW QUESTION 141

- (Exam Topic 3)

You plan to run a script as an experiment using a Script Run Configuration. The script uses modules from the scipy library as well as several Python packages that are not typically installed in a default conda environment

You plan to run the experiment on your local workstation for small datasets and scale out the experiment by running it on more powerful remote compute clusters for larger datasets.

You need to ensure that the experiment runs successfully on local and remote compute with the least administrative effort.

What should you do?

- A. Create and register an Environment that includes the required package
- B. Use this Environment for all experiment runs.
- C. Always run the experiment with an Estimator by using the default packages.
- D. Do not specify an environment in the run configuration for the experimen
- E. Run the experiment by using the default environment.
- F. Create a config.yaml file defining the conda packages that are required and save the file in the experiment folder.
- G. Create a virtual machine (VM) with the required Python configuration and attach the VM as a compute target
- H. Use this compute target for all experiment runs.

Answer: A

Explanation:

If you have an existing Conda environment on your local computer, then you can use the service to create an environment object. By using this strategy, you can reuse your local interactive environment on remote runs.

Reference:

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

NEW QUESTION 144

- (Exam Topic 3)

You are creating a new Azure Machine Learning pipeline using the designer.

The pipeline must train a model using data in a comma-separated values (CSV) file that is published on a website. You have not created a dataset for this file.

You need to ingest the data from the CSV file into the designer pipeline using the minimal administrative effort.

Which module should you add to the pipeline in Designer?

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

Answer: D

Explanation:

The preferred way to provide data to a pipeline is a Dataset object. The Dataset object points to data that lives in or is accessible from a datastore or at a Web URL. The Dataset class is abstract, so you will create an instance of either a FileDataset (referring to one or more files) or a TabularDataset that's created by from one or more files with delimited columns of data.

Example:

```
from azureml.core import Dataset
```

```
iris_tabular_dataset = Dataset.Tabular.from_delimited_files([(def_blob_store, 'train-dataset/iris.csv')])
```

 Reference:

<https://docs.microsoft.com/en-us/azure/machine-learning/how-to-create-your-first-pipeline>

NEW QUESTION 148

- (Exam Topic 3)

You run an automated machine learning experiment in an Azure Machine Learning workspace. Information about the run is listed in the table below:

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

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

A)

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

B)

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

C)

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

D)

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

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

Answer: A

Explanation:

The get_output method on automl_classifier returns the best run and the fitted model for the last invocation. Overloads on get_output allow you to retrieve the best run and fitted model for any logged metric or for a particular iteration.

In []:

```
best_run, fitted_model = local_run.get_output()
```

 Reference:

<https://notebooks.azure.com/azureml/projects/azureml-getting-started/html/how-to-use-azureml/automated-mach>

NEW QUESTION 153

- (Exam Topic 3)

You are a data scientist creating a linear regression model.

You need to determine how closely the data fits the regression line. Which metric should you review?

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

Answer: A

Explanation:

Coefficient of determination, often referred to as R2, represents the predictive power of the model as a value between 0 and 1. Zero means the model is random (explains nothing); 1 means there is a perfect fit. However, caution should be used in interpreting R2 values, as low values can be entirely normal and high values can be suspect.

References:

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

NEW QUESTION 154

- (Exam Topic 3)

You create the following config.json file.

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

You must use the Azure Machine Learning SDK to interact with data and experiments in the workspace. You need to configure the config.json file to connect to the workspace from the Python environment. Which two additional parameters must you add to the config.json file in order to connect to the workspace?

Each correct answer presents part of the solution.

NOTE: Each correct selection is worth one point.

- A. subscription_id
- B. Key
- C. resource_group
- D. region
- E. Login

Answer: AC

Explanation:

To use the same workspace in multiple environments, create a JSON configuration file. The configuration file saves your subscription (subscription_id), resource (resource_group), and workspace name so that it can be easily loaded.

The following sample shows how to create a workspace. from azureml.core import Workspace

```
ws = Workspace.create(name='myworkspace', subscription_id='<azure-subscription-id>', resource_group='myresourcegroup', create_resource_group=True,  
location='eastus2'  
)
```

Reference:

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

NEW QUESTION 158

- (Exam Topic 3)

You run a script as an experiment in Azure Machine Learning.

You have a Run object named run that references the experiment run. You must review the log files that were generated during the experiment run.

You need to download the log files to a local folder for review.

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

NOTE: Each correct selection is worth one point.

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

Answer: AE

Explanation:

The run Class get_all_logs method downloads all logs for the run to a directory.

The run Class get_details gets the definition, status information, current log files, and other details of the run. Reference:

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

NEW QUESTION 162

- (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 165

- (Exam Topic 3)

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

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

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

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

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

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

Solution: Run the following code:

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

Does the solution meet the goal?

- A. Yes
- B. No

Answer: A

NEW QUESTION 168

- (Exam Topic 2)

You need to configure the Edit Metadata module so that the structure of the datasets match.

Which configuration options should you select? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.

Properties Project

▲ Edit Metadata

Column

Selected columns:

Column names: MedianValue

Launch column selector

▼

Floating point
DateTime
TimeSpan
Integer

▼

Unchanged
Make Categorical
Make Uncategorical

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

Box 1: Floating point

Need floating point for Median values.

Scenario: An initial investigation shows that the datasets are identical in structure apart from the MedianValue column. The smaller Paris dataset contains the MedianValue in text format, whereas the larger London dataset contains the MedianValue in numerical format.

Box 2: Unchanged

Note: Select the Categorical option to specify that the values in the selected columns should be treated as categories.

For example, you might have a column that contains the numbers 0,1 and 2, but know that the numbers actually mean "Smoker", "Non smoker" and "Unknown". In that case, by flagging the column as categorical you can ensure that the values are not used in numeric calculations, only to group data.

NEW QUESTION 173

- (Exam Topic 2)

You need to select a feature extraction method. Which method should you use?

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

Answer: A

Explanation:

Spearman's rank correlation coefficient assesses how well the relationship between two variables can be described using a monotonic function.

Note: Both Spearman's and Kendall's can be formulated as special cases of a more general correlation coefficient, and they are both appropriate in this scenario.

Scenario: The MedianValue and AvgRoomsInHouse columns both hold data in numeric format. You need to select a feature selection algorithm to analyze the relationship between the two columns in more detail. References:

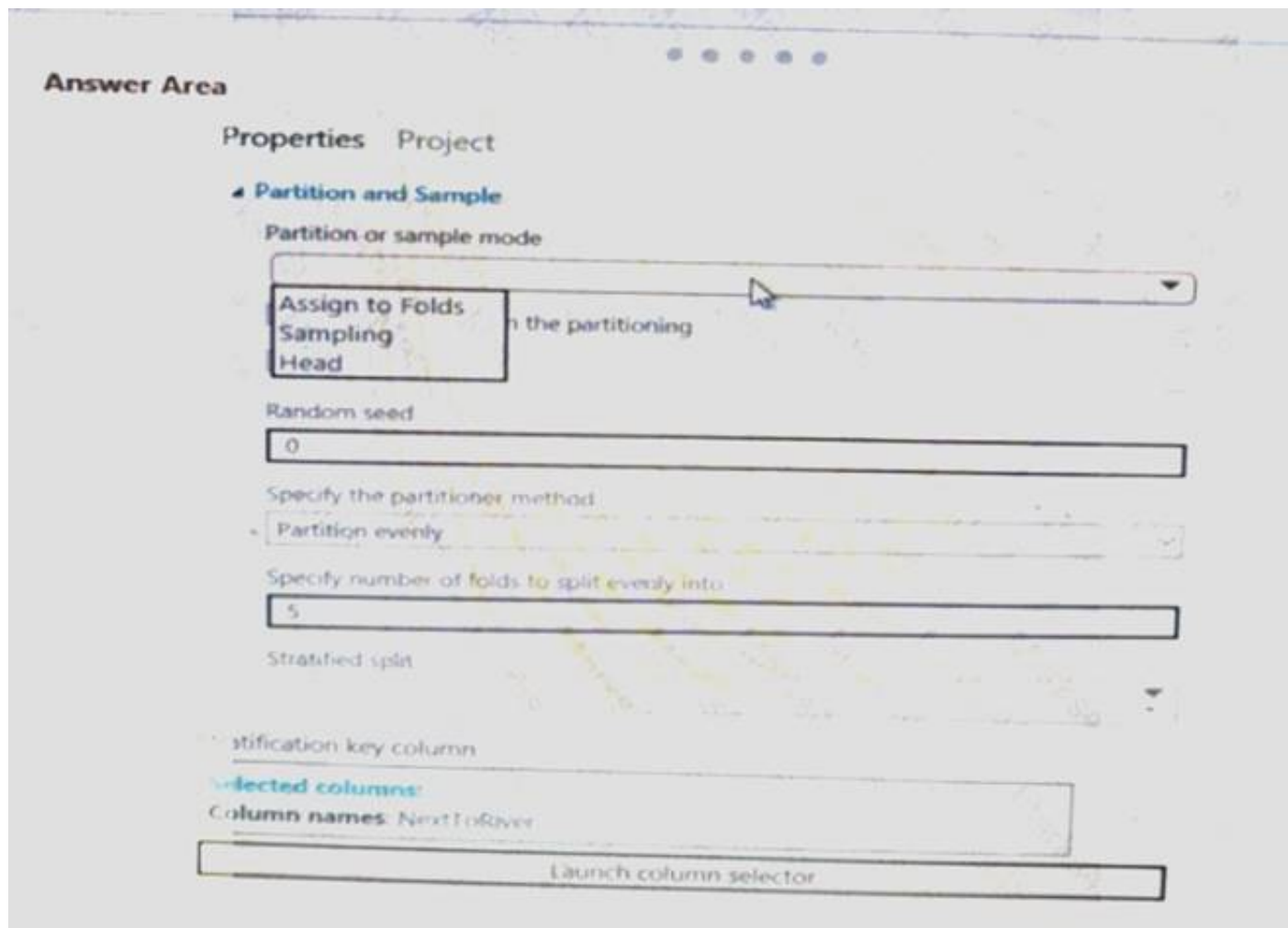
<https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/feature-selection-modules>

NEW QUESTION 175

- (Exam Topic 2)

You need to identify the methods for dividing the data according, to the testing requirements.

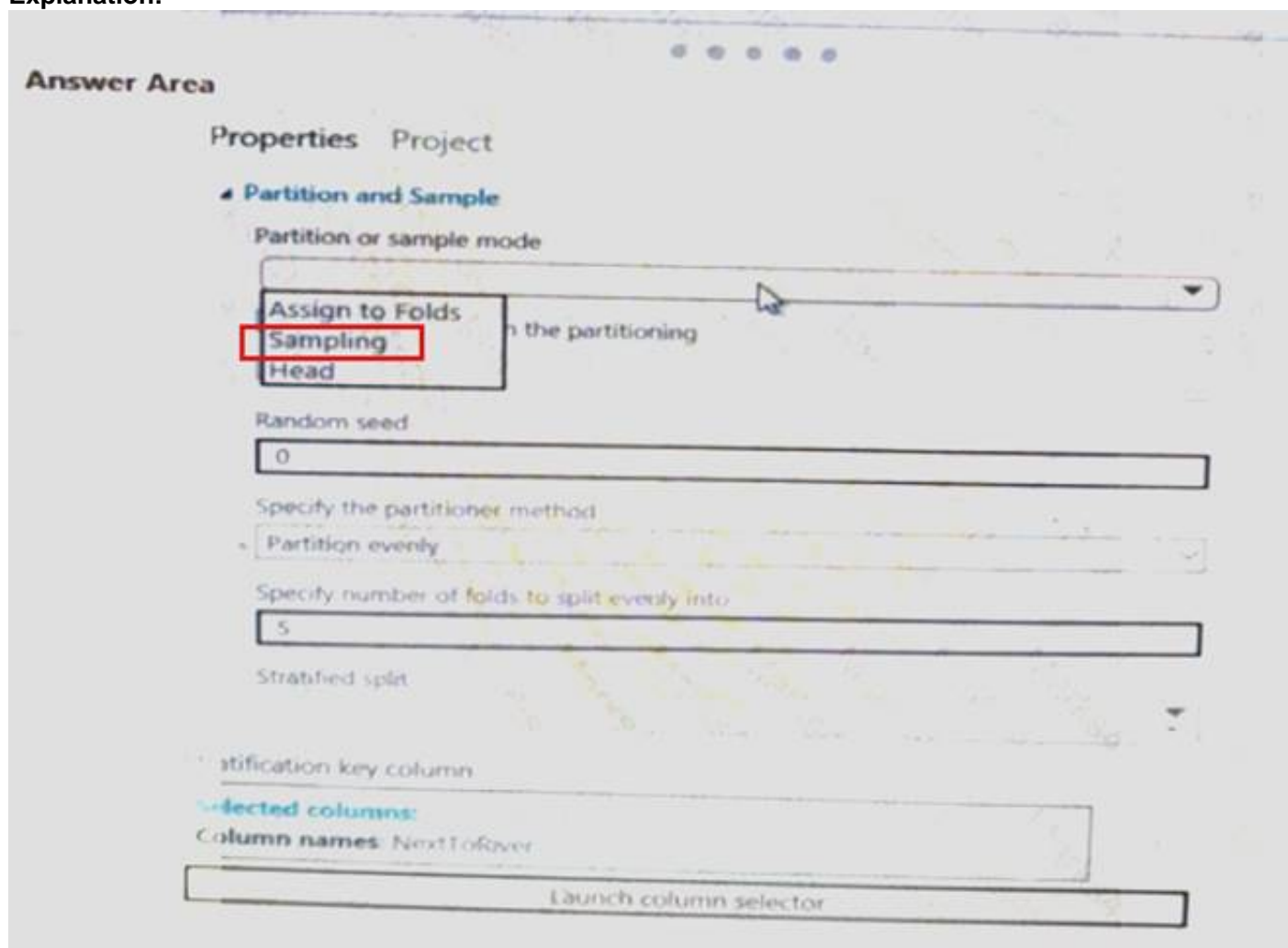
Which properties should you select? To answer, select the appropriate option-, m the answer area. NOTE: Each correct selection is worth one point.



- A. Mastered
- B. Not Mastered

Answer: A

Explanation:



NEW QUESTION 176

- (Exam Topic 1)

You need to implement a feature engineering strategy for the crowd sentiment local models. What should you do?

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

Answer: D

Explanation:

The linear discriminant analysis method works only on continuous variables, not categorical or ordinal variables.

Linear discriminant analysis is similar to analysis of variance (ANOVA) in that it works by comparing the means of the variables.

Scenario:

Data scientists must build notebooks in a local environment using automatic feature engineering and model building in machine learning pipelines.

Experiments for local crowd sentiment models must combine local penalty detection data. All shared features for local models are continuous variables.

NEW QUESTION 181

- (Exam Topic 1)

You need to implement a scaling strategy for the local penalty detection data.

Which normalization type should you use?

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

Answer: C

Explanation:

Post batch normalization statistics (PBN) is the Microsoft Cognitive Toolkit (CNTK) version of how to evaluate the population mean and variance of Batch Normalization which could be used in inference Original Paper.

In CNTK, custom networks are defined using the BrainScriptNetworkBuilder and described in the CNTK network description language "BrainScript."

Scenario:

Local penalty detection models must be written by using BrainScript. References:

<https://docs.microsoft.com/en-us/cognitive-toolkit/post-batch-normalization-statistics>

NEW QUESTION 183

- (Exam Topic 1)

You need to define an evaluation strategy for the crowd sentiment models.

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

Actions		Answer Area
Define a cross-entropy function activation.		
Add cost functions for each target state.		
Evaluate the classification error metric.	⬅	⬆
Evaluate the distance error metric.	➡	⬇
Add cost functions for each component metric.		
Define a sigmoid loss function activation.		

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

Step 1: Define a cross-entropy function activation

When using a neural network to perform classification and prediction, it is usually better to use cross-entropy error than classification error, and somewhat better to use cross-entropy error than mean squared error to evaluate the quality of the neural network.

Step 2: Add cost functions for each target state. Step 3: Evaluated the distance error metric. References:

<https://www.analyticsvidhya.com/blog/2018/04/fundamentals-deep-learning-regularization-techniques/>

NEW QUESTION 184

.....

Thank You for Trying Our Product

* 100% Pass or Money Back

All our products come with a 90-day Money Back Guarantee.

* One year free update

You can enjoy free update one year. 24x7 online support.

* Trusted by Millions

We currently serve more than 30,000,000 customers.

* Shop Securely

All transactions are protected by VeriSign!

100% Pass Your DP-100 Exam with Our Prep Materials Via below:

<https://www.certleader.com/DP-100-dumps.html>