DP-100 Q&As

Designing and Implementing a Data Science Solution on Azure

Pass Microsoft DP-100 Exam with 100% Guarantee

Free Download Real Questions & Answers **PDF** and **VCE** file from:

https://www.lead4pass.com/dp-100.html

100% Passing Guarantee
100% Money Back Assurance

Following Questions and Answers are all new published by Microsoft Official Exam Center

QUESTION 1

HOTSPOT

You write code to retrieve an experiment that is run from your Azure Machine Learning workspace.

The run used the model interpretation support in Azure Machine Learning to generate and upload a model explanation.

Business managers in your organization want to see the importance of the features in the model.

You need to print out the model features and their relative importance in an output that looks similar to the following.

Feature	Importance
0	1.5627435610083558
2	0.6077689312583112
4	0.5574002432900718
3	0.42858759955671777
1	0.3501361539771977

How should you complete the code? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.

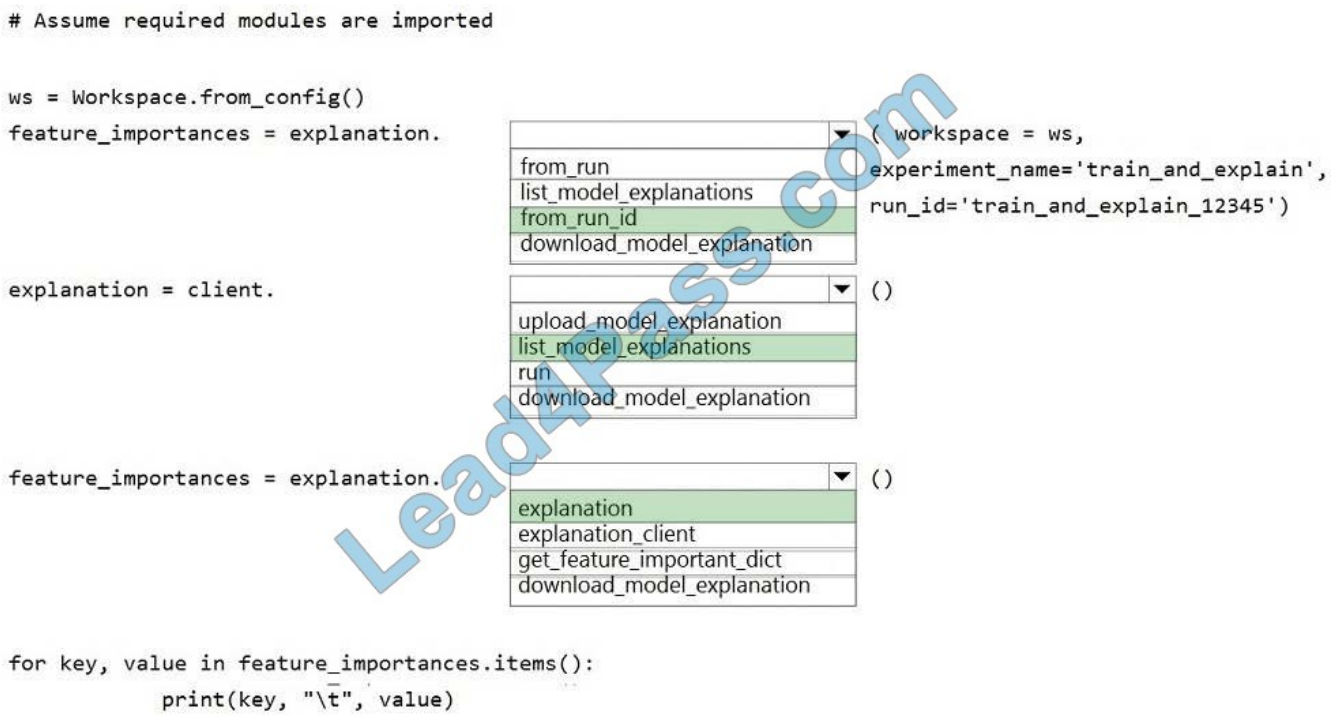
Hot Area:

Answer Area

```
# Assume required modules are imported

ws = Workspace.from_config()
feature_importances = explanation.[ from_run | list_model_explanations | from_run_id | download_model_explanation ]( workspace = ws,
                                                                                                                    experiment_name='train_and_explain',
                                                                                                                    run_id='train_and_explain_12345')

explanation = client.[ upload_model_explanation | list_model_explanations | run | download_model_explanation ]()

feature_importances = explanation.[ explanation | explanation_client | get_feature_important_dict | download_model_explanation ]()

for key, value in feature_importances.items():
            print(key, "\t", value)
```

Correct Answer:

Answer Area



Box 1: from_run_id

from_run_id(workspace, experiment_name, run_id)

Create the client with factory method given a run ID.

Returns an instance of the ExplanationClient.

Parameters

Workspace Workspace An object that represents a workspace.

experiment_name str The name of an experiment.

run_id str A GUID that represents a run.

Box 2: list_model_explanations

list_model_explanations returns a dictionary of metadata for all model explanations available.

Returns

A dictionary of metadata such as id, data type, method, model type, and upload time, sorted by upload time

Box 3:

Reference:

https://docs.microsoft.com/en-us/python/api/azureml-contrib-interpret/azureml.contrib.interpret.explanation.explanation_client.explanationclient?view=azure-ml-py

QUESTION 2

HOTSPOT

You have a dataset that contains 2,000 rows. You are building a machine learning classification model by using Azure Learning Studio. You add a Partition and Sample module to the experiment.

You need to configure the module. You must meet the following requirements:

1.

Divide the data into subsets

2.

Assign the rows into folds using a round-robin method

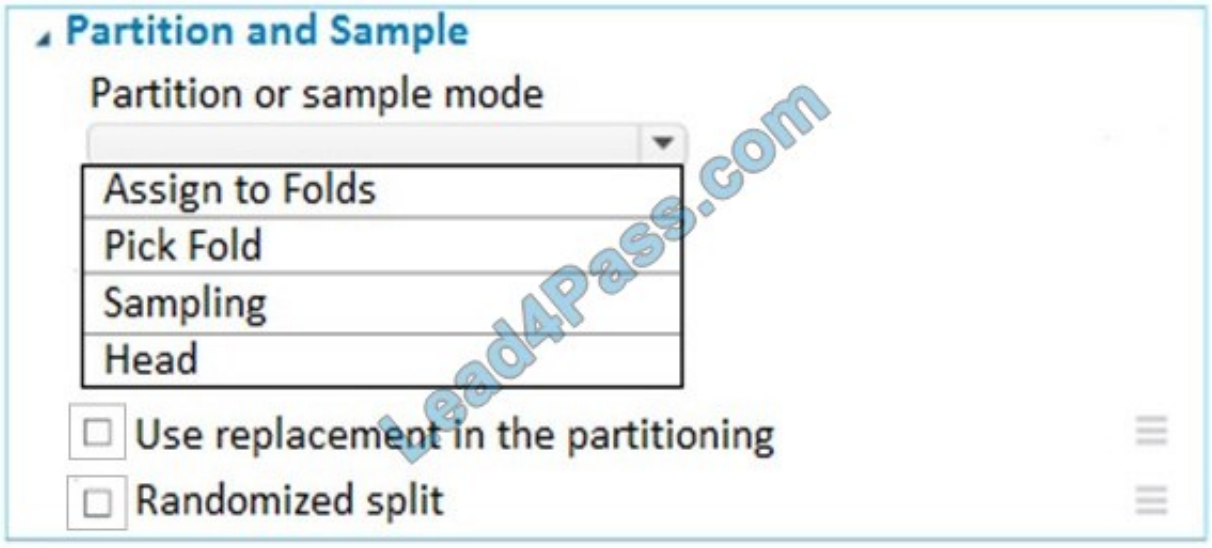
3.

Allow rows in the dataset to be reused

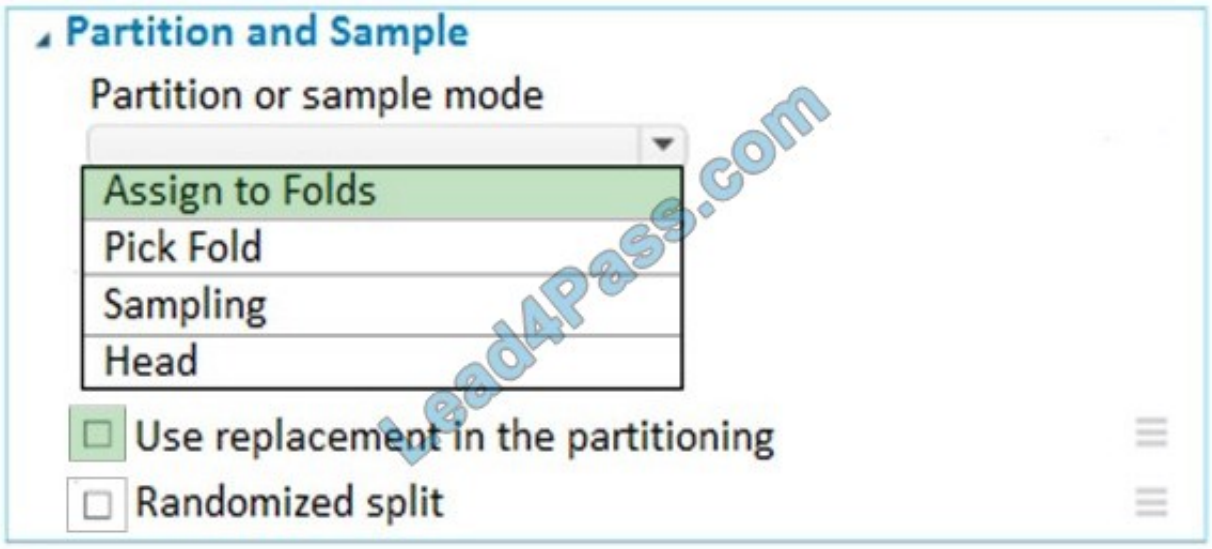
How should you configure the module? To answer, select the appropriate options in the dialog box in the answer area.

NOTE: Each correct selection is worth one point.

Hot Area:



Correct Answer:



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

Add the Partition and Sample module to your experiment in Studio (classic), and connect the dataset.

For Partition or sample mode, select Assign to Folds.

Use replacement in the partitioning: Select this option if you want the sampled row to be put back into the pool of rows for potential reuse. As a result, the same row might be assigned to several folds.

If you do not use replacement (the default option), the sampled row is not put back into the pool of rows for potential reuse. As a result, each row can be assigned to only one fold.

Randomized split: Select this option if you want rows to be randomly assigned to folds.

If you do not select this option, rows are assigned to folds using the round-robin method.

References:

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

QUESTION 3

You are retrieving data from a large datastore by using Azure Machine Learning Studio.

You must create a subset of the data for testing purposes using a random sampling seed based on the system clock.

You add the Partition and Sample module to your experiment.

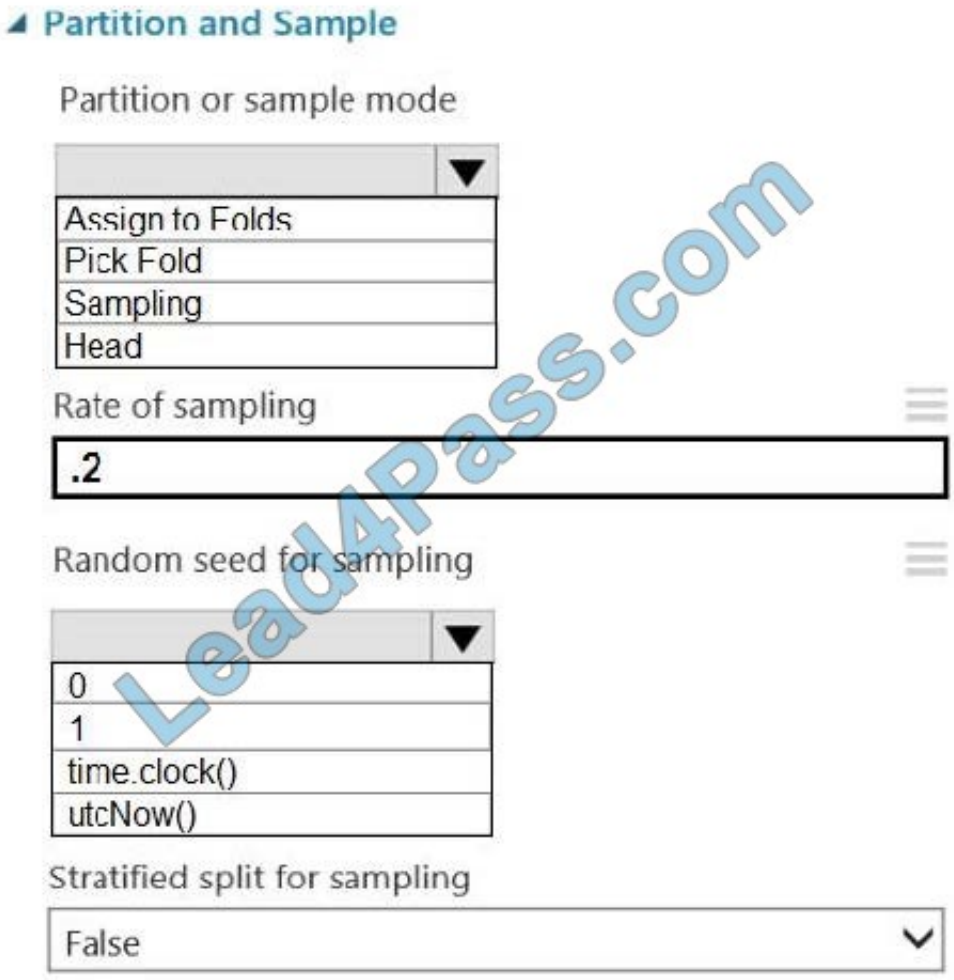
You need to select the properties for the module.

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

NOTE: Each correct selection is worth one point.

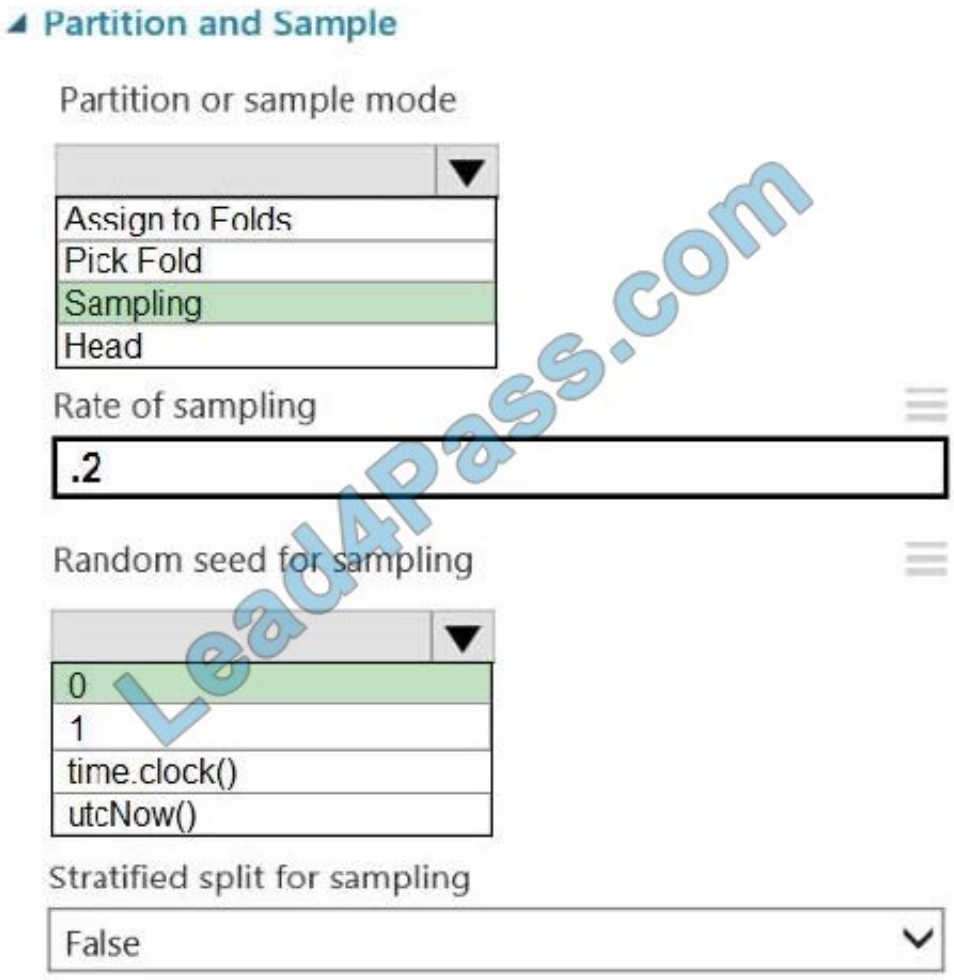
Hot Area:

Answer Area



Correct Answer:

Answer Area



Box 1: Sampling Create a sample of data This option supports simple random sampling or stratified random sampling. This is useful if you want to create a smaller representative sample dataset for testing.

1.

Add the Partition and Sample module to your experiment in Studio, and connect the dataset.

2.

Partition or sample mode: Set this to Sampling.

3.

Rate of sampling. See box 2 below.

Box 2: 0

3. Rate of sampling. Random seed for sampling: Optionally, type an integer to use as a seed value.

This option is important if you want the rows to be divided the same way every time. The default value is 0, meaning that a starting seed is generated based on the system clock. This can lead to slightly different results each time you run the experiment.

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

QUESTION 4

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

Correct Answer: A

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 quantized.

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

QUESTION 5

You are evaluating a completed binary classification machine learning model.

You need to use the precision as the evaluation metric.

Which visualization should you use?

A. Violin plot

B. Gradient descent

C. Box plot

D. Binary classification confusion matrix

Correct Answer: D

Incorrect Answers:

A: A violin plot is a visual that traditionally combines a box plot and a kernel density plot.

B: Gradient descent is a first-order iterative optimization algorithm for finding the minimum of a function. To find a local minimum of a function using gradient descent, one takes steps proportional to the negative of the gradient (or approximate gradient) of the function at the current point.

C: A box plot lets you see basic distribution information about your data, such as median, mean, range and quartiles but doesn\\'t show you how your data looks throughout its range.

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

QUESTION 6

Note: This question is part of a series of questions that present the same scenario. Each question in the series contains a unique solution that might meet the stated goals. Some question sets might have more than one correct solution, while

others might not have a correct solution.

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

You have a Python script named train.py in a local folder named scripts. The script trains a regression model by using scikit-learn. The script includes code to load a training data file which is also located in the scripts folder.

You must run the script as an Azure ML experiment on a compute cluster named aml-compute.

You need to configure the run to ensure that the environment includes the required packages for model training. You have instantiated a variable named aml-compute that references the target compute cluster.

Solution: Run the following code:

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

Does the solution meet the goal?

A. Yes

B. No

Correct Answer: B

There is a missing line: conda_packages=[\\'scikit-learn\\'], which is needed.

Correct example:

sk_est = Estimator(source_directory=\\'./my-sklearn-proj\\', script_params=script_params,

compute_target=compute_target,

entry_script=\\'train.py\\',

conda_packages=[\\'scikit-learn\\'])

Note:

The Estimator class represents a generic estimator to train data using any supplied framework.

This class is designed for use with machine learning frameworks that do not already have an Azure Machine Learning pre-configured estimator. Pre-configured estimators exist for Chainer, PyTorch, TensorFlow, and SKLearn.

Example:

from azureml.train.estimator import Estimator

script_params = {

to mount files referenced by mnist dataset

\\'--data-folder\\': ds.as_named_input(\\'mnist\\').as_mount(),

\\'--regularization\\': 0.8

}

Reference:

https://docs.microsoft.com/en-us/python/api/azureml-train-core/azureml.train.estimator.estimator

QUESTION 7

HOTSPOT

You are performing sentiment analysis using a CSV file that includes 12.0O0 customer reviews written in a short sentence format.

You add the CSV file to Azure Machine Learning Studio and Configure it as the starting point dataset of an experiment.

You add the Extract N-Gram Features from Text module to the experiment to extract key phrases from the customer review column in the dataset.

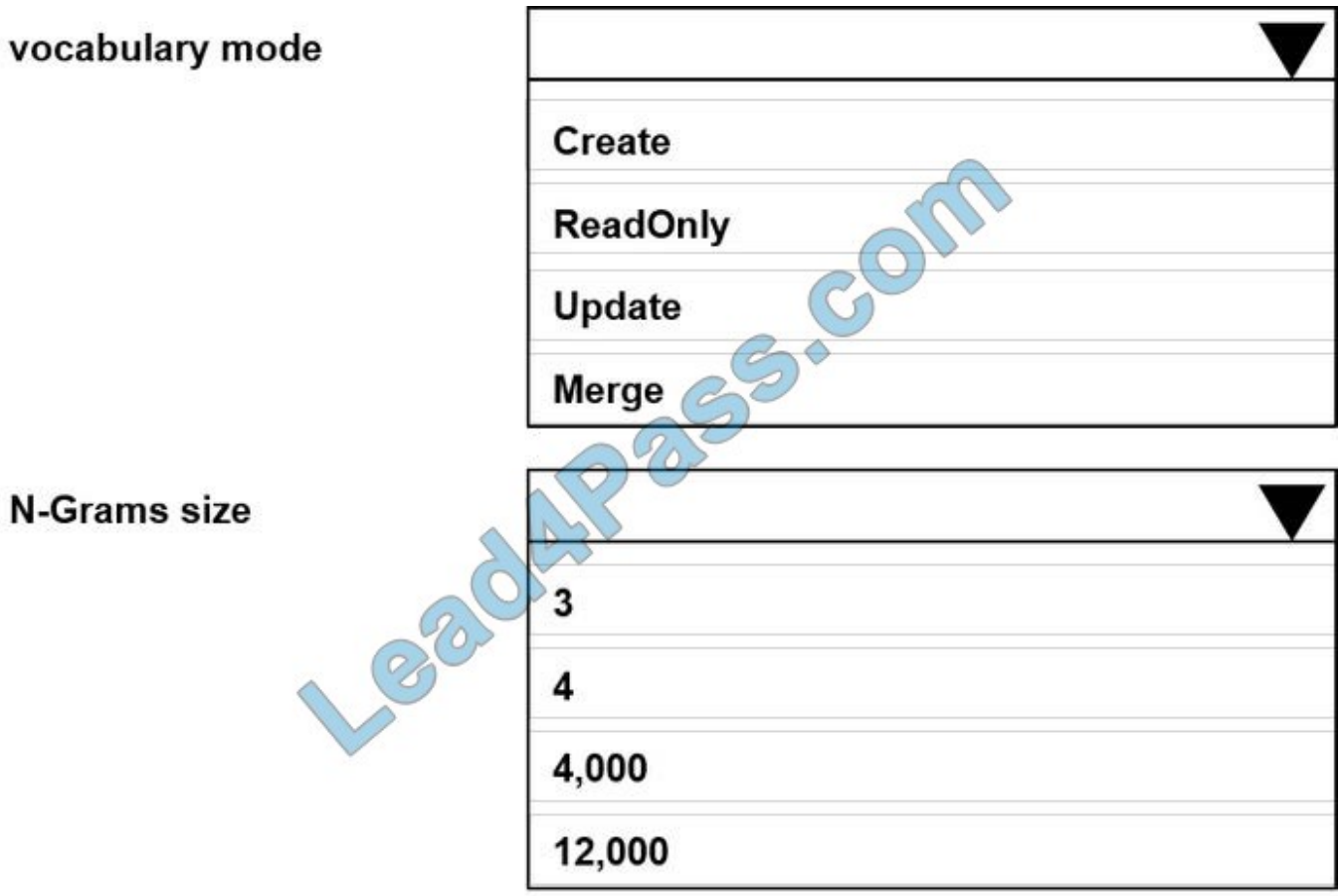
You must create a new n-gram text dictionary from the customer review text and set the maximum n-gram size to trigrams.

You need to configure the Extract N Gram features from Text module.

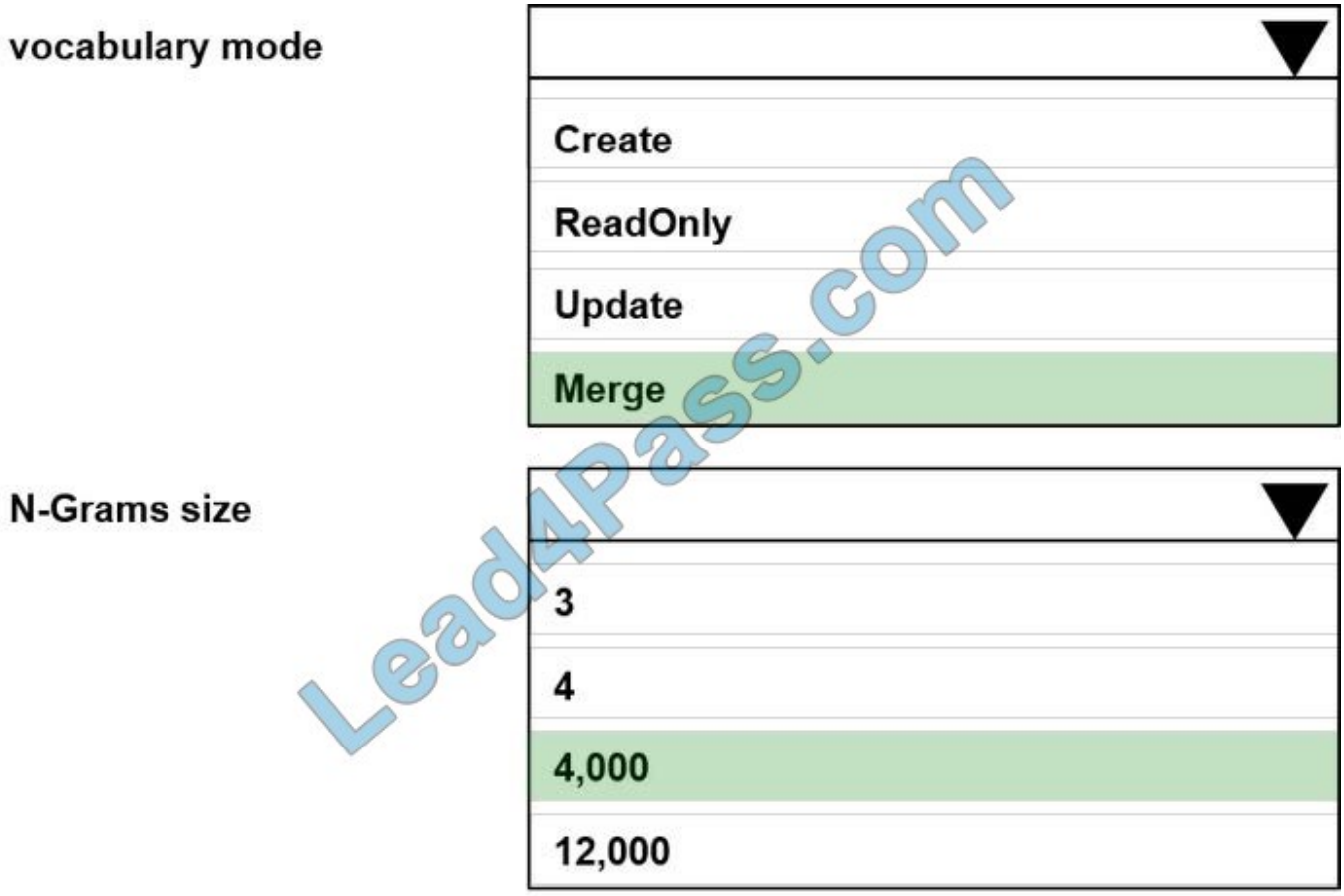
What should you select? To answer, select the appropriate options in the answer area;

NOTE: Each correct selection is worth one point.

Hot Area:



Correct Answer:



QUESTION 8

You plan to provision an Azure Machine Learning Basic edition workspace for a data science project.

You need to identify the tasks you will be able to perform in the workspace.

Which three tasks will you be able to perform? Each correct answer presents a complete solution.

NOTE: Each correct selection is worth one point.

A. Create a Compute Instance and use it to run code in Jupyter notebooks.

B. Create an Azure Kubernetes Service (AKS) inference cluster.

C. Use the designer to train a model by dragging and dropping pre-defined modules.

D. Create a tabular dataset that supports versioning.

E. Use the Automated Machine Learning user interface to train a model.

Correct Answer: ABD

Incorrect Answers:

C, E: The UI is included the Enterprise edition only.

Reference:

https://azure.microsoft.com/en-us/pricing/details/machine-learning/

QUESTION 9

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 count for each feature column.

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

NOTE: Each correct selection is worth one point.

A. Computer Linear Correlation

B. Export Count Table

C. Execute Python Script

D. Convert to Indicator Values

E. Summarize Data

Correct Answer: BE

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.

Incorrect Answers:

A: The Compute Linear Correlation module in Azure Machine Learning Studio is used to compute a set of Pearson correlation coefficients for each possible pair of variables in the input dataset.

C: With Python, you can perform tasks that aren\\'t currently supported by existing Studio modules such as:

Visualizing data using matplotlib

Using Python libraries to enumerate datasets and models in your workspace Reading, loading, and manipulating data from sources not supported by the Import Data module

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

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

QUESTION 10

You need to implement a model development strategy to determine a user\\'s tendency to respond to an ad.

Which technique should you use?

A. Use a Relative Expression Split module to partition the data based on centroid distance.

B. Use a Relative Expression Split module to partition the data based on distance travelled to the event.

C. Use a Split Rows module to partition the data based on distance travelled to the event.

D. Use a Split Rows module to partition the data based on centroid distance.

Correct Answer: A

Split Data partitions the rows of a dataset into two distinct sets.

The Relative Expression Split option in the Split Data module of Azure Machine Learning Studio is helpful when you need to divide a dataset into training and testing datasets using a numerical expression.

Relative Expression Split: Use this option whenever you want to apply a condition to a number column. The number could be a date/time field, a column containing age or dollar amounts, or even a percentage. For example, you might want to

divide your data set depending on the cost of the items, group people by age ranges, or separate data by a calendar date.

Scenario:

Local market segmentation models will be applied before determining a user\\'s propensity to respond to an advertisement.

The distribution of features across training and production data are not consistent

References:

https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/split-data

QUESTION 11

Note: This question is part of a series of questions that present the same scenario. Each question in the series contains a unique solution that might meet the stated goals. Some question sets might have more than one correct solution, while

others might not have a correct solution.

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

appear in the review screen.

You are creating a model to predict the price of a student\\'s artwork depending on the following variables:

the student\\'s length of education, degree type, and art form.

You start by creating a linear regression model.

You need to evaluate the linear regression model.

Solution: Use the following metrics: Accuracy, Precision, Recall, F1 score, and AUC.

Does the solution meet the goal?

A. Yes

B. No

Correct Answer: B

Those are metrics for evaluating classification models, instead use: Mean Absolute Error, Root Mean Absolute Error, Relative Absolute Error, Relative Squared Error, and the Coefficient of Determination.

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

QUESTION 12

You plan to create a speech recognition deep learning model.

The model must support the latest version of Python.

You need to recommend a deep learning framework for speech recognition to include in the Data Science Virtual Machine (DSVM).

What should you recommend?

A. Rattle

B. TensorFlow

C. Weka

D. Scikit-learn

Correct Answer: B

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

Incorrect Answers:

A: Rattle is the R analytical tool that gets you started with data analytics and machine learning.

C: Weka is used for visual data mining and machine learning software in Java.

D: Scikit-learn is one of the most useful library for machine learning in Python. It is on NumPy, SciPy and matplotlib, this library contains a lot of effiecient tools for machine learning and statistical modeling including classification, regression, clustering and dimensionality reduction.

Reference: https://www.infoworld.com/article/3278008/what-is-tensorflow-the-machine-learning-library-explained.html

QUESTION 13

You define a datastore named ml-data for an Azure Storage blob container. In the container, you have a folder named train that contains a file named data.csv. You plan to use the file to train a model by using the Azure Machine Learning

SDK.

You plan to train the model by using the Azure Machine Learning SDK to run an experiment on local compute.

You define a DataReference object by running the following code:

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


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

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

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

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

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

C.
```
import pandas as pd

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

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

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

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

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

A. Option A

B. Option B

C. Option C

D. Option D

E. Option E

Correct Answer: E

Example:

data_folder = args.data_folder

Load Train and Test data

train_data = pd.read_csv(os.path.join(data_folder, \\'data.csv\\'))

Reference:

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

Latest DP-100 Dumps DP-100 VCE Dumps DP-100 Braindumps

To Read the Whole Q&As, please purchase the Complete Version from Our website.

Try our product !

100% Guaranteed Success
100% Money Back Guarantee
365 Days Free Update
Instant Download After Purchase
24x7 Customer Support
Average 99.9% Success Rate
More than 800,000 Satisfied Customers Worldwide
Multi-Platform capabilities - Windows, Mac, Android, iPhone, iPod, iPad, Kindle

We provide exam PDF and VCE of Cisco, Microsoft, IBM, CompTIA, Oracle and other IT Certifications.
You can view Vendor list of All Certification Exams offered:

https://www.lead4pass.com/allproducts

Need Help

Please provide as much detail as possible so we can best assist you.
To update a previously submitted ticket: