

**TEST 1 - QUESTION: 1/50**

This question is included in a number of questions that depicts the identical set-up. However, every question has a distinctive result. Establish if the recommendation satisfies the requirements. You have been tasked with evaluating your model on a partial data sample via k-fold cross-validation. You have already configured a k parameter as the number of splits. You now have to configure the k parameter for the cross-validation with the usual value choice. Recommendation: You configure the use of the value  $k = 3$ . Will the requirements be satisfied?

☐ A

Yes

☐ B

No

**CORRECT ANSWER: B****EXPLANATION:**

As you can see at the moment our simulator just shows to you the right answer. We are trying our best to add also a reasonable explanation for the provided answer. We are sorry for the inconvenience

## TEST 1 - QUESTION: 2/50

You are preparing to train a regression model via automated machine learning. The data available to you has features with missing values, as well as categorical features with little discrete values. You want to make sure that automated machine learning is configured as follows: missing values must be automatically imputed. categorical features must be encoded as part of the training task. Which of the following actions should you take?

☐ A

You should make use of the featurization parameter with the 'auto' value pair.

☐ B

You should make use of the featurization parameter with the 'FeaturizationConfig' value pair.

☐ C

You should make use of the featurization parameter with the 'on' value pair.

☐ D

You should make use of the featurization parameter with the 'off' value pair.

## CORRECT ANSWER: A

### EXPLANATION:

Explanation: Featurization str or FeaturizationConfig Values: 'auto' / 'off' / FeaturizationConfig Indicator for whether featurization step should be done automatically or not, or whether customized featurization should be used. Column type is automatically detected. Based on the detected column type preprocessing/featurization is done as follows: Categorical: Target encoding, one hot encoding, drop high cardinality categories, impute missing values. Numeric: Impute missing values, cluster distance, weight of evidence. DateTime: Several features such as day, seconds, minutes, hours etc. Text: Bag of words, pre-trained Word embedding, text target encoding. Reference: <https://docs.microsoft.com/en-us/python/api/azureml-train-automl->

# Constructor

Python

 Copy

```
AutoMLConfig(task: str, path: typing.Union[str, NoneType] = None, iterations:
typing.Union[int, NoneType] = None, primary_metric: typing.Union[str, NoneType] =
None, positive_label: typing.Union[typing.Any, NoneType] = None, compute_target:
typing.Union[typing.Any, NoneType] = None, spark_context:
typing.Union[typing.Any, NoneType] = None, X: typing.Union[typing.Any, NoneType]
= None, y: typing.Union[typing.Any, NoneType] = None, sample_weight:
typing.Union[typing.Any, NoneType] = None, X_valid: typing.Union[typing.Any,
NoneType] = None, y_valid: typing.Union[typing.Any, NoneType] = None,
sample_weight_valid: typing.Union[typing.Any, NoneType] = None,
cv_splits_indices: typing.Union[typing.List[typing.List[typing.Any]], NoneType] =
None, validation_size: typing.Union[float, NoneType] = None, n_cross_validations:
typing.Union[int, NoneType] = None, y_min: typing.Union[float, NoneType] = None,
y_max: typing.Union[float, NoneType] = None, num_classes: typing.Union[int,
NoneType] = None, featurization: typing.Union[str,
azureml.automl.core.featurization.featurizationconfig.FeaturizationConfig] =
'auto', max_cores_per_iteration: int = 1, max_concurrent_iterations: int = 1,
iteration_timeout_minutes: typing.Union[int, NoneType] = None, mem_in_mb:
typing.Union[int, NoneType] = None, enforce_time_on_windows: bool = True
```

praw343167

## TEST 1 - QUESTION: 3/50

You need to implement a Data Science Virtual Machine (DSVM) that supports the Caffe2 deep learning framework. Which of the following DSVM should you create?

☐

A

Ubuntu 16.04 DSVM

☐

B

CentOS 7.4 DSVM

☐

C

Windows Server 2016 DSVM

☐

D

Windows Server 2012 DSVM

### CORRECT ANSWER: A

#### EXPLANATION:

Explanation: Caffe2 is supported by Data Science Virtual Machine for Linux. Microsoft offers Linux editions of the DSVM on Ubuntu 16.04 LTS and CentOS 7.4. However, only the DSVM on Ubuntu is preconfigured for Caffe2. Reference: <https://docs.microsoft.com/en-us/azure/machine-learning/data-science-virtual-machine/overview>

## Comparison with Azure Machine Learning

The DSVM is a customized VM image for Data Science but [Azure Machine Learning](#) (AzureML) is an end-to-end platform that encompasses:

- Fully Managed Compute
  - Compute Instances
  - Compute Clusters for distributed ML tasks
  - Inference Clusters for real-time scoring
- Datastores (for example Blob, ADLS Gen2, SQL DB)
- Experiment tracking
- Model management
- Notebooks
- Environments (manage conda and R dependencies)
- Labeling
- Pipelines (automate End-to-End Data science workflows)

praw343167

**TEST 1 - QUESTION: 4/50**

This question is included in a number of questions that depicts the identical set-up. However, every question has a distinctive result. Establish if the recommendation satisfies the requirements. You are in the process of carrying out feature engineering on a dataset. You want to add a feature to the dataset and fill the column value. Recommendation: You must make use of the Group Categorical Values Azure Machine Learning Studio module. Will the requirements be satisfied?

☐

A

No

☐

B

Yes

**CORRECT ANSWER: A****EXPLANATION:**

As you can see at the moment our simulator just shows to you the right answer. We are trying our best to add also a reasonable explanation for the provided answer. We are sorry for the inconvenience.

### TEST 1 - QUESTION: 5/50

You make use of Azure Machine Learning Studio to develop a linear regression model. You perform an experiment to assess various algorithms. Which of the following is an algorithm that reduces the variances between actual and predicted values?

☐

A

Fast Forest Quantile Regression

☐

B

Poisson Regression

☐

C

Linear Regression

☐

D

Boosted Decision Tree Regression

### CORRECT ANSWER: D

#### EXPLANATION:

Explanation: Mean absolute error (MAE) measures how close the predictions are to the actual outcomes; thus, a lower score is better. Reference: <https://docs.microsoft.com/en-us/azure/machine-learning/algorithm-module-reference/boosted-decision-tree-regression> <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>

**TEST 1 - QUESTION: 6/50**

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

Scikit-learn

☐ B

Rattle

☐ C

TensorFlow

☐ D

Weka

**CORRECT ANSWER: C****EXPLANATION:**

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

**TEST 1 - QUESTION: 7/50**

This question is included in a number of questions that depicts the identical set-up. However, every question has a distinctive result. Establish if the recommendation satisfies the requirements. You have been tasked with employing a machine learning model, which makes use of a PostgreSQL database and needs GPU processing, to forecast prices. You are preparing to create a virtual machine that has the necessary tools built into it. You need to make use of the correct virtual machine type. Recommendation: You make use of a Data Science Virtual Machine (DSVM) Windows edition. Will the requirements be satisfied?

☐ A

No

☐ B

Yes

**CORRECT ANSWER: B****EXPLANATION:**

Explanation: In the DSVM, your training models can use deep learning algorithms on hardware that's based on graphics processing units (GPUs). PostgreSQL is available for the following operating systems: Linux (all recent distributions), 64-bit installers available for macOS (OS X) version 10.6 and newer – Windows (with installers available for 64-bit version; tested on latest versions and back to Windows 2012 R2. Reference: <https://docs.microsoft.com/en-us/azure/machine-learning/data-science-virtual-machine/overview>



**TEST 1 - QUESTION: 8/50**

This question is included in a number of questions that depicts the identical set-up. However, every question has a distinctive result. Establish if the recommendation satisfies the requirements. You are in the process of creating a machine learning model. Your dataset includes rows with null and missing values. You plan to make use of the Clean Missing Data module in Azure Machine Learning Studio to detect and fix the null and missing values in the dataset. Recommendation: You make use of the Replace with median option. Will the requirements be satisfied?

☐ A

No

☐ B

Yes

**CORRECT ANSWER: A****EXPLANATION:**

Reference: <https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/clean-missing-data>

**TEST 1 - QUESTION: 9/50**

This question is included in a number of questions that depicts the identical set-up. However, every question has a distinctive result. Establish if the recommendation satisfies the requirements. You have been tasked with evaluating your model on a partial data sample via k-fold cross-validation. You have already configured a k parameter as the number of splits. You now have to configure the k parameter for the cross-validation with the usual value choice. Recommendation: You configure the use of the value  $k = 1$ . Will the requirements be satisfied?

☐ A

Yes

☐ B

No

**CORRECT ANSWER: B****EXPLANATION:**

As you can see at the moment our simulator just shows to you the right answer. We are trying our best to add also a reasonable explanation for the provided answer. We are sorry for the inconvenience.

**TEST 1 - QUESTION: 10/50**

This question is included in a number of questions that depicts the identical set-up. However, every question has a distinctive result. Establish if the recommendation satisfies the requirements. You are planning to make use of Azure Machine Learning designer to train models. You need choose a suitable compute type. Recommendation: You choose Inference cluster. Will the requirements be satisfied?

☐ A

No

☐ B

Yes

**CORRECT ANSWER: A****EXPLANATION:**

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

**TEST 1 - QUESTION: 11/50**

You are making use of the Azure Machine Learning designer to construct an experiment. After dividing a dataset into training and testing sets, you configure the algorithm to be Two-Class Boosted Decision Tree. You are preparing to ascertain the Area Under the Curve (AUC). Which of the following is a sequential combination of the models required to achieve your goal?

☐ A

Evaluate, Export Data, Train.

☐ B

Train, Score, Export Data.

☐ C

Train, Score, Evaluate.

☐ D

Score, Evaluate, Train.

**CORRECT ANSWER: C****EXPLANATION:**

As you can see at the moment our simulator just shows to you the right answer. We are trying our best to add also a reasonable explanation for the provided answer. We are sorry for the inconvenience.

**TEST 1 - QUESTION: 12/50**

This question is included in a number of questions that depicts the identical set-up. However, every question has a distinctive result. Establish if the recommendation satisfies the requirements. You are in the process of creating a machine learning model. Your dataset includes rows with null and missing values. You plan to make use of the Clean Missing Data module in Azure Machine Learning Studio to detect and fix the null and missing values in the dataset. Recommendation: You make use of the Remove entire row option. Will the requirements be satisfied?

☐

A

No

☐

B

Yes

**CORRECT ANSWER: B****EXPLANATION:**

Explanation: Remove entire row: Completely removes any row in the dataset that has one or more missing values. This is useful if the missing value can be considered randomly missing. Reference: <https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/clean-missing-data>

**TEST 1 - QUESTION: 13/50**

This question is included in a number of questions that depicts the identical set-up. However, every question has a distinctive result. Establish if the recommendation satisfies the requirements. You are planning to make use of Azure Machine Learning designer to train models. You need choose a suitable compute type. Recommendation: You choose Attached compute. Will the requirements be satisfied?

☐ A

No

☐ B

Yes

### TEST 1 - QUESTION: 14/50

You are in the process of constructing a deep convolutional neural network (CNN). The CNN will be used for image classification. You notice that the CNN model you constructed displays hints of overfitting. You want to make sure that overfitting is minimized, and that the model is converged to an optimal fit. Which of the following is TRUE with regards to achieving your goal?

☐ A

You have to add an additional dense layer with 512 input units, and reduce the amount of training data.

☐ B

You have to add an additional dense layer with 512 input units, and add L1/L2 regularization.

☐ C

You have to add L1/L2 regularization, and make use of training data augmentation.

☐ D

You have to add L1/L2 regularization, and reduce the amount of training data.

☐ E

You have to reduce the amount of training data and make use of training data augmentation.

#### EXPLANATION:

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

**CORRECT ANSWER: A**

**CORRECT ANSWER: D**

KEEP OPEN

#### EXPLANATION:

Explanation: B: Weight regularization provides an approach to reduce the overfitting of a deep learning neural network model on the training data and improve the performance of the model on new data, such as the holdout test set. Keras provides a weight regularization API that allows you to add a penalty for weight size to the loss function. Three different regularizer instances are provided; they are: L1: Sum of the absolute weights. L2: Sum of the squared weights. L1L2: Sum of the absolute and the squared weights. Because a fully connected layer occupies most of the parameters, it is prone to overfitting. One method to reduce overfitting is dropout. At each training stage, individual nodes are either "dropped out" of the net with probability  $1-p$  or kept with probability  $p$ , so that a reduced network is left; incoming and outgoing edges to a dropped-out node are also removed. By avoiding training all nodes on all training data,

dropout decreases overfitting. Reference:

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

[https://en.wikipedia.org/wiki/Convolutional\\_neural\\_network](https://en.wikipedia.org/wiki/Convolutional_neural_network)



**TEST 1 - QUESTION: 15/50**

This question is included in a number of questions that depicts the identical set-up. However, every question has a distinctive result. Establish if the recommendation satisfies the requirements. You have been tasked with evaluating your model on a partial data sample via k-fold cross-validation. You have already configured a k parameter as the number of splits. You now have to configure the k parameter for the cross-validation with the usual value choice. Recommendation: You configure the use of the value  $k = 10$ . Will the requirements be satisfied?

☐ A

Yes

☐ B

No

**CORRECT ANSWER: A****EXPLANATION:**

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.

### TEST 1 - QUESTION: 16/50

You construct a machine learning experiment via Azure Machine Learning Studio. You would like to split data into two separate datasets. Which of the following actions should you take?

☐ A

You should make use of the Group Categorical Values module.

☐ B

You should make use of the Clip Values module.

☐ C

You should make use of the Group Data into Bins module.

☐ D

You should make use of the Split Data module.

### CORRECT ANSWER: C

KEEP OPEN

#### EXPLANATION:

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

## Module overview

This article describes how to use the Group Data into Bins module in Machine Learning Studio (classic), to group numbers or change the distribution of continuous data.

The Group Data into Bins module supports multiple options for binning data. You can customize how the bin edges are set and how values are apportioned into the bins. For example, you can:

- Manually type a series of values to serve as the bin boundaries.
- Calculate entropy scores to determine an information values for each range, to optimize the bins in the predictive model. + Assign values to bins by using *quantiles*, or percentile ranks.
- Control the number of values in each bin can also be controlled.
- Force an even distribution of values into the bins.

**TEST 1 - QUESTION: 17/50**

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



CHECK BELOW THE RIGHT ANSWER

Properties

Project

### Extract N-Gram Features from Text

Text column

Selected columns:

Column type: String Feature

Launch column selector

Vocabulary mode

	▼
Create	
ReadOnly	
Update	
Merge	

N-Grams size

	▼
3	
4	
4,000	
12,000	

0

Weighting function

	▼
--	---

Minimum word length

3

Maximum word length

25

Minimum n-gram document absolu...

5

Maximum n-gram document ratio

1

**CORRECT ANSWER:**

KEEP OPEN

**EXPLANATION:**

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

Reference: <https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/extract-n-gram-features-from-text>

**TEST 1 - QUESTION: 18/50**

This question is included in a number of questions that depicts the identical set-up. However, every question has a distinctive result. Establish if the recommendation satisfies the requirements. You have been tasked with constructing a machine learning model that translates language text into a different language text. The machine learning model must be constructed and trained to learn the sequence of the. Recommendation: You make use of Recurrent Neural Networks (RNNs). Will the requirements be satisfied?

☐

A

No

☐

B

Yes

**CORRECT ANSWER: B****EXPLANATION:**

Explanation: Note: RNNs are designed to take sequences of text as inputs or return sequences of text as outputs, or both. They're called recurrent because the network's hidden layers have a loop in which the output and cell state from each time step become inputs at the next time step. This recurrence serves as a form of memory. It allows contextual information to flow through the network so that relevant outputs from previous time steps can be applied to network operations at the current time step. Reference:

<https://towardsdatascience.com/language-translation-with-rnns-d84d43b40571>

**TEST 1 - QUESTION: 19/50**

You plan to use a Data Science Virtual Machine (DSVM) with the open source deep learning frameworks Caffe2 and PyTorch. You need to select a pre-configured DSVM to support the frameworks. What should you create?

☐ A

Geo AI Data Science Virtual Machine with ArcGIS

☐ B

Data Science Virtual Machine for Linux (Ubuntu)

☐ C

Data Science Virtual Machine for Linux (CentOS)

☐ D

Data Science Virtual Machine for Windows 2012

☐ E

Data Science Virtual Machine for Windows 2016

**CORRECT ANSWER: B**

KEEP OPEN

**EXPLANATION:**

Explanation: Caffe2 and PyTorch is supported by Data Science Virtual Machine for Linux. Microsoft offers Linux editions of the DSVM on Ubuntu 16.04 LTS and CentOS 7.4. Only the DSVM on Ubuntu is preconfigured for Caffe2 and PyTorch. Incorrect Answers: D: Caffe2 and PyTorch are only supported in the Data Science Virtual Machine for Linux. Reference: <https://docs.microsoft.com/en-us/azure/machine-learning/data-science-virtual-machine/overview>

### TEST 1 - QUESTION: 20/50

This question is included in a number of questions that depicts the identical set-up. However, every question has a distinctive result. Establish if the recommendation satisfies the requirements. You are in the process of creating a machine learning model. Your dataset includes rows with null and missing values. You plan to make use of the Clean Missing Data module in Azure Machine Learning Studio to detect and fix the null and missing values in the dataset. Recommendation: You make use of the Custom substitution value option. Will the requirements be satisfied?

☐ A

No

☐ B

Yes

### CORRECT ANSWER: A

#### EXPLANATION:

Reference: <https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/clean-missing-data>

## How to use Clean Missing Data

This module lets you define a cleaning operation. You can also save the cleaning operation so that you can apply it later to new data. See the following links for a description of how to create and save a cleaning process:

- [To replace missing values](#)
- [To apply a cleaning transformation to new data](#)

#### 📘 Important

The cleaning method that you use for handling missing values can dramatically affect your results. We recommend that you experiment with different methods. Consider both the justification for use of a particular method, and the quality of the results.

praw343167



**TEST 1 - QUESTION: 21/50**

You make use of Azure Machine Learning Studio to create a binary classification model. You are preparing to carry out a parameter sweep of the model to tune hyperparameters. You have to make sure that the sweep allows for every possible combination of hyperparameters to be iterated. Also, the computing resources needed to carry out the sweep must be reduced. Which of the following actions should you take?

☐ A

You should consider making use of the Entire grid sweep mode.

☐ B

You should consider making use of the Random grid sweep mode.

☐ C

You should consider making use of the Selective grid sweep mode.

☐ D

You should consider making use of the Measured grid sweep mode.

**CORRECT ANSWER: B**

KEEP OPEN

**EXPLANATION:**

Explanation: Maximum number of runs on random grid: This option also controls the number of iterations over a random sampling of parameter values, but the values are not generated randomly from the specified range; instead, a matrix is created of all possible combinations of parameter values and a random sampling is taken over the matrix. This method is more efficient and less prone to regional oversampling or undersampling. If you are training a model that supports an integrated parameter sweep, you can also set a range of seed values to use and iterate over the random seeds as well. This is optional, but can be useful for avoiding bias introduced by seed selection. C: Entire grid: When you select this option, the module loops over a grid predefined by the system, to try different combinations and identify the best learner. This option is useful for cases where you don't know what the best parameter settings might be and want to try all possible combination of values. Reference: <https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/tune-model-hyperparameters>

**TEST 1 - QUESTION: 22/50**

This question is included in a number of questions that depicts the identical set-up. However, every question has a distinctive result. Establish if the recommendation satisfies the requirements. You have been tasked with constructing a machine learning model that translates language text into a different language text. The machine learning model must be constructed and trained to learn the sequence of the. Recommendation: You make use of Generative Adversarial Networks (GANs). Will the requirements be satisfied?

☐ A

No

☐ B

Yes

**CORRECT ANSWER: A**

KEEP OPEN

**EXPLANATION:**

As you can see at the moment our simulator just shows to you the right answer. We are trying our best to add also a reasonable explanation for the provided answer. We are sorry for the inconvenience.

### TEST 1 - QUESTION: 23/50

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 Deep Learning Virtual Machine (DLVM) Windows edition.

☐ C

Create a Geo AI Data Science Virtual Machine (Geo-DSVM) Windows edition.

☐ D

Create a Deep Learning Virtual Machine (DLVM) Linux edition.

**CORRECT ANSWER: A**

KEEP OPEN

#### EXPLANATION:

Explanation: In the DSVM, your training models can use deep learning algorithms on hardware that's based on graphics processing units (GPUs). PostgreSQL is available for the following operating systems: Linux (all recent distributions), 64-bit installers available for macOS (OS X) version 10.6 and newer – Windows (with installers available for 64-bit version; tested on latest versions and back to Windows 2012 R2. Incorrect Answers: B: The Azure Geo AI Data Science VM (Geo-DSVM) delivers geospatial analytics capabilities from Microsoft's Data Science VM. Specifically, this VM extends the AI and data science toolkits in the Data Science VM by adding ESRI's market-leading ArcGIS Pro Geographic Information System. C, D: DLVM is a template on top of DSVM image. In terms of the packages, GPU drivers etc are all there in the DSVM image. Mostly it is for convenience during creation where we only allow DLVM to be created on GPU VM instances on Azure. Reference: <https://docs.microsoft.com/en-us/azure/machine-learning/data-science-virtual-machine/overview>

**TEST 1 - QUESTION: 24/50**

This question is included in a number of questions that depicts the identical set-up. However, every question has a distinctive result. Establish if the recommendation satisfies the requirements. You have been tasked with employing a machine learning model, which makes use of a PostgreSQL database and needs GPU processing, to forecast prices. You are preparing to create a virtual machine that has the necessary tools built into it. You need to make use of the correct virtual machine type. Recommendation: You make use of a Geo AI Data Science Virtual Machine (Geo-DSVM) Windows edition. Will the requirements be satisfied?

☐ A

Yes

☐ B

No

**CORRECT ANSWER: B**

KEEP OPEN

**EXPLANATION:**

Explanation: The Azure Geo AI Data Science VM (Geo-DSVM) delivers geospatial analytics capabilities from Microsoft's Data Science VM. Specifically, this VM extends the AI and data science toolkits in the Data Science VM by adding ESRI's market-leading ArcGIS Pro Geographic Information System. Reference: <https://docs.microsoft.com/en-us/azure/machine-learning/data-science-virtual-machine/overview>

**TEST 1 - QUESTION: 25/50**

This question is included in a number of questions that depicts the identical set-up. However, every question has a distinctive result. Establish if the recommendation satisfies the requirements. You are in the process of carrying out feature engineering on a dataset. You want to add a feature to the dataset and fill the column value. Recommendation: You must make use of the Join Data Azure Machine Learning Studio module. Will the requirements be satisfied?

☐

A

Yes

☐

B

No

**CORRECT ANSWER: B**

KEEP OPEN

**EXPLANATION:**

As you can see at the moment our simulator just shows to you the right answer. We are trying our best to add also a reasonable explanation for the provided answer. We are sorry for the inconvenience.

**TEST 1 - QUESTION: 26/50**

You have been tasked with designing a deep learning model, which accommodates the most recent edition of Python, to recognize language. You have to include a suitable deep learning framework in the Data Science Virtual Machine (DSVM). Which of the following actions should you take?

☐ A

You should consider including Theano.

☐ B

You should consider including Rattle.

☐ C

You should consider including Chainer.

☐ D

You should consider including TensorFlow.

**CORRECT ANSWER: D**

KEEP OPEN

**EXPLANATION:**

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

Machine learning is a complex discipline. But implementing machine learning models is far less daunting and difficult than it used to be, thanks to machine learning frameworks—such as **Google's TensorFlow**—that ease the process of acquiring data, training models, serving predictions, and refining future results.

Created by the Google Brain team, TensorFlow is an open source library for numerical computation and large-scale machine learning. TensorFlow bundles together a slew of machine learning and deep learning (aka neural networking) models and algorithms and makes them useful by way of a common metaphor. It uses Python to provide a convenient front-end API for building applications with the framework, while executing those applications in high-performance C++.

draw343167

**TEST 1 - QUESTION: 27/50**

Your team is building a data engineering and data science development environment. The environment must support the following requirements: support Python and Scala compose data storage, movement, and processing services into automated data pipelines the same tool should be used for the orchestration of both data engineering and data science support workload isolation and interactive workloads enable scaling across a cluster of machines You need to create the environment. What should you do?

☐ A

Build the environment in Azure Databricks and use Azure Container Instances for orchestration.

☐ B

Build the environment in Azure Databricks and use Azure Data Factory for orchestration.

☐ C

Build the environment in Apache Hive for HDInsight and use Azure Data Factory for orchestration.

☐ D

Build the environment in Apache Spark for HDInsight and use Azure Container Instances for orchestration.

**CORRECT ANSWER: B**

KEEP OPEN

**EXPLANATION:**

Explanation: In Azure Databricks, we can create two different types of clusters. Standard, these are the default clusters and can be used with Python, R, Scala and SQL High-concurrency Azure Databricks is fully integrated with Azure Data Factory. Incorrect Answers: D: Azure Container Instances is good for development or testing. Not suitable for production workloads. Reference: <https://docs.microsoft.com/en-us/azure/architecture/data-guide/technology-choices/data-science-and-machine-learning>

**TEST 1 - QUESTION: 28/50**

You are developing a data science workspace that uses an Azure Machine Learning service. You need to select a compute target to deploy the workspace. What should you use?

☐ A

Azure Databricks

☐ B

Azure Container Service

☐ C

Azure Data Lake Analytics

☐ D

Apache Spark for HDInsight

**CORRECT ANSWER: B**

KEEP OPEN

**EXPLANATION:**

Explanation: Azure Container Instances can be used as compute target for testing or development. Use for low-scale CPU-based workloads that require less than 48 GB of RAM. Reference: <https://docs.microsoft.com/en-us/azure/machine-learning/service/how-to-deploy-and-where>



**TEST 1 - QUESTION: 29/50**

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

**CORRECT ANSWER: C**

KEEP OPEN

**EXPLANATION:**

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>

**TEST 1 - QUESTION: 30/50**

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

☐

A

Azure Cognitive Services

☐

B

Azure Data Lake Analytics

☐

C

Azure Machine Learning Studio

☐

D

Azure HDInsight with Spark MLlib

**CORRECT ANSWER: A**

KEEP OPEN

**EXPLANATION:**

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

### TEST 1 - QUESTION: 31/50

You have been tasked with creating a new Azure pipeline via the Machine Learning designer. You have to make sure that the pipeline trains a model using data in a comma-separated values (CSV) file that is published on a website. A dataset for the file for this file does not exist. Data from the CSV file must be ingested into the designer pipeline with the least amount of administrative effort as possible. Which of the following actions should you take?

☐ A

You should add the Dataset object to the pipeline.

☐ B

You should make use of the Convert to TXT module.

☐ C

You should add the Import Data object to the pipeline.

☐ D

You should add the Copy Data object to the pipeline.

### CORRECT ANSWER: A

KEEP OPEN

### EXPLANATION:

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>

**TEST 1 - QUESTION: 32/50**

You are planning to host practical training to acquaint learners with data visualization creation using Python. Learner devices are able to connect to the internet. Learner devices are currently NOT configured for Python development. Also, learners are unable to install software on their devices as they lack administrator permissions. Furthermore, they are unable to access Azure subscriptions. It is imperative that learners are able to execute Python-based data visualization code. Which of the following actions should you take?

☐ A

You should consider configuring the use of Azure BatchAI.

☐ B

You should consider configuring the use of Azure Kubernetes Service.

☐ C

You should consider configuring the use of Azure Container Instance.

☐ D

You should consider configuring the use of Azure Notebooks.

**CORRECT ANSWER: D**

KEEP OPEN

**EXPLANATION:**

Reference: <https://notebooks.azure.com/>

## TEST 1 - QUESTION: 33/50

HOTSPOT Complete the sentence by selecting the correct option in the answer area.

☐ CHECK BELOW THE RIGHT ANSWER

### Answer Area

To move a large dataset from Azure Machine Learning Studio to a Weka environment, the data must be converted to the



CSV  
DOCX  
ARFF  
TXT

format.

### CORRECT ANSWER:

#### EXPLANATION:

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 entities and their attributes, and is contained in a single text file. Reference: <https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/convert-to-arff>

#### Module overview

This article describes how to use the Convert to ARFF module in Machine Learning Studio (classic), to convert datasets and results 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 entities and their attributes, and is contained in a single text file. You can find details of the Weka file format in the [Technical Notes](#) section.

In general, conversion to the Weka file format is required only if you want to use both Machine Learning and Weka, and intend to move your training data back and forth between them.

For more information about the Weka toolset, see this Wikipedia article: [Weka \(machine learning\)](#)

**TEST 1 - QUESTION: 34/50**

HOTSPOT Complete the sentence by selecting the correct option in the answer area.

☐ CHECK BELOW THE RIGHT ANSWER

**Answer Area**

	▼
SSD	
FPGA	
GPU	
Power BI	

is required for a Deep Learning Virtual Machine (DLVM) to support Compute Unified Device Architecture (CUDA) computations.

**CORRECT ANSWER:**

KEEP OPEN

**EXPLANATION:**

Explanation: A Deep Learning Virtual Machine is a pre-configured environment for deep learning using GPU instances.

**TEST 1 - QUESTION: 35/50**

This question is included in a number of questions that depicts the identical set-up. However, every question has a distinctive result. Establish if the recommendation satisfies the requirements. You have been tasked with employing a machine learning model, which makes use of a PostgreSQL database and needs GPU processing, to forecast prices. You are preparing to create a virtual machine that has the necessary tools built into it. You need to make use of the correct virtual machine type. Recommendation: You make use of a Deep Learning Virtual Machine (DLVM) Windows edition. Will the requirements be satisfied?

☐ A

No

☐ B

Yes

**CORRECT ANSWER: A**

KEEP OPEN

**EXPLANATION:**

Explanation: DLVM is a template on top of DSVM image. In terms of the packages, GPU drivers etc are all there in the DSVM image. Mostly it is for convenience during creation where we only allow DLVM to be created on GPU VM instances on Azure. Reference: <https://docs.microsoft.com/en-us/azure/machine-learning/data-science-virtual-machine/overview>

**TEST 1 - QUESTION: 36/50**

You plan to build a team data science environment. Data for training models in machine learning pipelines will be over 20 GB in size. You have the following requirements: Models must be built using Caffe2 or Chainer frameworks. Data scientists must be able to use a data science environment to build the machine learning pipelines and train models on their personal devices in both connected and disconnected network environments. Personal devices must support updating machine learning pipelines when connected to a network. You need to select a data science environment. Which environment should you use?

☐ A

Azure Databricks

☐ B

Azure Machine Learning Studio

☐ C

Azure Kubernetes Service (AKS)

☐ D

Azure Machine Learning Service

**CORRECT ANSWER: D**

KEEP OPEN

**EXPLANATION:**

Explanation: The Data Science Virtual Machine (DSVM) is a customized VM image on Microsoft's Azure cloud built specifically for doing data science. Caffe2 and Chainer are supported by DSVM. DSVM integrates with Azure Machine Learning. Incorrect Answers: B: Use Machine Learning Studio when you want to experiment with machine learning models quickly and easily, and the built-in machine learning algorithms are sufficient for your solutions.

Reference: <https://docs.microsoft.com/en-us/azure/machine-learning/data-science-virtual-machine/overview>



**TEST 1 - QUESTION: 37/50**

You need to consider the underlined segment to establish whether it is accurate. To improve the amount of low incidence cases in a dataset, you should make use of the SMOTE module. Select “No adjustment required” if the underlined segment is accurate. If the underlined segment is inaccurate, select the accurate option.

☐

A

Edit Metadata

☐

B

No adjustment required.

☐

C

Remove Duplicate Rows

☐

D

Join Data

**CORRECT ANSWER: B**

KEEP OPEN

**EXPLANATION:**

Explanation: Use the SMOTE module in Azure Machine Learning Studio to increase the number of underrepresented cases in a dataset used for machine learning. SMOTE is a better way of increasing the number of rare cases than simply duplicating existing cases. Reference: <https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/smote>

**TEST 1 - QUESTION: 38/50**

This question is included in a number of questions that depicts the identical set-up. However, every question has a distinctive result. Establish if the recommendation satisfies the requirements. You have been tasked with constructing a machine learning model that translates language text into a different language text. The machine learning model must be constructed and trained to learn the sequence of the. Recommendation: You make use of Convolutional Neural Networks (CNNs). Will the requirements be satisfied?

☐ A

No

☐ B

Yes

**CORRECT ANSWER: A**

KEEP OPEN

**EXPLANATION:**

As you can see at the moment our simulator just shows to you the right answer. We are trying our best to add also a reasonable explanation for the provided answer. We are sorry for the inconvenience.

**TEST 1 - QUESTION: 39/50**

You have recently concluded the construction of a binary classification machine learning model. You are currently assessing the model. You want to make use of a visualization that allows for precision to be used as the measurement for the assessment. Which of the following actions should you take?

☐ A

You should consider using Receiver Operating Characteristic (ROC) curve visualization.

☐ B

You should consider using Venn diagram visualization.

☐ C

You should consider using Box plot visualization.

☐ D

You should consider using the Binary classification confusion matrix visualization.

**CORRECT ANSWER: D**

KEEP OPEN

**EXPLANATION:**

Reference: <https://docs.microsoft.com/en-us/azure/machine-learning/how-to-understand-automated-ml#confusion-matrix>

**TEST 1 - QUESTION: 40/50**

You plan to use a Deep Learning Virtual Machine (DLVM) to train deep learning models using Compute Unified Device Architecture (CUDA) computations. You need to configure the DLVM to support CUDA. What should you implement?

☐ A

Solid State Drives (SSD)

☐ B

Intel Software Guard Extensions (Intel SGX) technology

☐ C

Computer Processing Unit (CPU) speed increase by using overclocking

☐ D

High Random Access Memory (RAM) configuration

☐ E

Graphic Processing Unit (GPU)

**CORRECT ANSWER: E**

KEEP OPEN

**EXPLANATION:**

Explanation: A Deep Learning Virtual Machine is a pre-configured environment for deep learning using GPU instances. Reference: <https://azuremarketplace.microsoft.com/en-au/marketplace/apps/microsoft-ads.dsvm-deep-learning>

**TEST 1 - QUESTION: 41/50**

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

Azure Storage Explorer

☐ B

Bulk Insert SQL Query

☐ C

Python script

☐ D

AzCopy

☐ E

Bulk Copy Program (BCP)

**CORRECT ANSWERS: A,C,D**

KEEP OPEN

**EXPLANATION:**

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

**TEST 1 - QUESTION: 42/50**

This question is included in a number of questions that depicts the identical set-up. However, every question has a distinctive result. Establish if the recommendation satisfies the requirements. You are in the process of carrying out feature engineering on a dataset. You want to add a feature to the dataset and fill the column value. Recommendation: You must make use of the Edit Metadata Azure Machine Learning Studio module. Will the requirements be satisfied?

☐

A

No

☐

B

Yes

**CORRECT ANSWER: B**

KEEP OPEN

**EXPLANATION:**

Explanation: Typical metadata changes might include marking columns as features. Reference: <https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/edit-metadata>  
<https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/join-data> <https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/group-categorical-values>

## TEST 1 - QUESTION: 43/50

HOTSPOT Complete the sentence by selecting the correct option in the answer area.

☐ CHECK BELOW THE RIGHT ANSWER

### Answer Area

▼
Probabilistic PCA
<b>Median</b>
SMOTE
Custom substitution value

is a data cleaning option of the Clean Missing Data module that does not require predictors for each column.

### CORRECT ANSWER:

KEEP OPEN

### EXPLANATION:

Explanation: Replace using Probabilistic PCA: Compared to other options, such as Multiple Imputation using Chained Equations (MICE), this option has the advantage of not requiring the application of predictors for each column. Instead, it approximates the covariance for the full dataset. Therefore, it might offer better performance for datasets that have missing values in many columns. Reference: <https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/clean-missing-data>

### TEST 1 - QUESTION: 44/50

You are developing a hands-on workshop to introduce Docker for Windows to attendees. You need to ensure that workshop attendees can install Docker on their devices. Which two prerequisite components should attendees install on the devices? Each correct answer presents part of the solution. NOTE: Each correct selection is worth one point.

- ☐ A  
VirtualBox
- ☐ B  
BIOS-enabled virtualization
- ☐ C  
Kitematic
- ☐ D  
Windows 10 64-bit Professional
- ☐ E  
Microsoft Hardware-Assisted Virtualization Detection Tool

### CORRECT ANSWERS: B,D

KEEP OPEN

### EXPLANATION:

Explanation: C: Make sure your Windows system supports Hardware Virtualization Technology and that virtualization is enabled. Ensure that hardware virtualization support is turned on in the BIOS settings. For example:





E: To run Docker, your machine must have a 64-bit operating system running Windows 7 or higher. Reference:

[https://docs.docker.com/toolbox/toolbox\\_install\\_windows/](https://docs.docker.com/toolbox/toolbox_install_windows/)

<https://blogs.technet.microsoft.com/canitpro/2015/09/08/step-by-step-enabling-hyper-v-for-use-on-windows-10/>

**TEST 1 - QUESTION: 45/50**

This question is included in a number of questions that depicts the identical set-up. However, every question has a distinctive result. Establish if the recommendation satisfies the requirements. You are planning to make use of Azure Machine Learning designer to train models. You need choose a suitable compute type. Recommendation: You choose Compute cluster. Will the requirements be satisfied?

☐ A

No

☐ B

Yes

**CORRECT ANSWER: B**

KEEP OPEN

**EXPLANATION:**

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

**TEST 1 - QUESTION: 46/50**

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 ARFF

☐ B

Convert to CSV

☐ C

Convert to Dataset

☐ D

Convert to SVMLight

**CORRECT ANSWER: A**

KEEP OPEN

**EXPLANATION:**

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 entities and their attributes, and is contained in a single text file. Reference: <https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/convert-to-arff>

**TEST 1 - QUESTION: 47/50**

You need to consider the underlined segment to establish whether it is accurate. To transform a categorical feature into a binary indicator, you should make use of the Clean Missing Data module. Select “No adjustment required” if the underlined segment is accurate. If the underlined segment is inaccurate, select the accurate option.

☐ A

Convert to Indicator Values

☐ B

Apply SQL Transformation

☐ C

Group Categorical Values

☐ D

No adjustment required.

**CORRECT ANSWER: A**

KEEP OPEN

**EXPLANATION:**

Explanation: Use the Convert to Indicator Values module in Azure Machine Learning Studio. The purpose of this module is to convert columns that contain categorical values into a series of binary indicator columns that can more easily be used as features in a machine learning model. Reference: <https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/convert-to-indicator-values>

**TEST 1 - QUESTION: 48/50**

You want to train a classification model using data located in a comma-separated values (CSV) file. The classification model will be trained via the Automated Machine Learning interface using the Classification task type. You have been informed that only linear models need to be assessed by the Automated Machine Learning. Which of the following actions should you take?

☐ A

You should enable automatic featurization.

☐ B

You should disable deep learning.

☐ C

You should disable automatic featurization.

☐ D

You should set the task type to Forecasting

**CORRECT ANSWER: C**

KEEP OPEN

**EXPLANATION:**

Reference: <https://econml.azurewebsites.net/spec/estimation/dml.html>  
<https://docs.microsoft.com/en-us/azure/machine-learning/how-to-use-automated-ml-for-ml-models>

### TEST 1 - QUESTION: 49/50

You have been tasked with ascertaining if two sets of data differ considerably. You will make use of Azure Machine Learning Studio to complete your task. You plan to perform a paired t-test. Which of the following are conditions that must apply to use a paired t-test? (Choose all that apply.)

☐ A

The sampling distribution of  $x_1 - x_2$  is normal.

☐ B

The sampling distribution of  $d$  is normal.

☐ C

You have a matched pairs of scores.

☐ D

All scores are independent from each other.

### CORRECT ANSWERS: B,C

KEEP OPEN

#### EXPLANATION:

Reference: <https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/test-hypothesis-using-t-test>

## How to configure Test Hypothesis Using t-Test

Use a single dataset as input. The columns that you are comparing must be in the same dataset.

If you need to compare columns from different datasets, you can isolate each column to compare by using [Select Columns in Dataset](#), and then merge them into one dataset by using [Add Columns](#).

1. Add the Test Hypothesis Using t-Test module to your experiment.

You can find this module in the [Statistical Functions](#) category in Studio (classic).

2. Add the dataset that contains the column or columns that you want to analyze.
3. Decide which kind of t-test is appropriate for your data. See [How to choose a t-test](#).
4. **Single sample:** If you are using a single sample, set these parameters:

prw343167

### TEST 1 - QUESTION: 50/50

HOTSPOT You need to consider the underlined segment to establish whether it is accurate.

☐ CHECK BELOW THE RIGHT ANSWER

### Answer Area

The

	▼
Venn diagram	
Box plot	
Gradient descent	
Violin plot	

visualization can be used to reveal outliers in your data.

### CORRECT ANSWER:

KEEP OPEN

### EXPLANATION:

Explanation: The box-plot algorithm can be used to display outliers. Reference: <https://medium.com/analytics-vidhya/what-is-an-outliers-how-to-detect-and-remove-them-which-algorithm-are-sensitive-towards-outliers-2d501993d59>