

➤ **Vendor: Microsoft**

➤ **Exam Code: DP-100**

➤ **Exam Name: Designing and Implementing a Data Science Solution on Azure**

➤ **New Updated Questions from [Braindump2go](#) (Updated in [May/2020](#))**

Visit Braindump2go and Download Full Version DP-100 Exam Dumps

QUESTION 72

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

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

You are creating a new experiment in Azure Machine Learning Studio. One class has a much smaller number of observations than the other classes in the training set. You need to select an appropriate data sampling strategy to compensate for the class imbalance.

Solution: You use the Scale and Reduce sampling mode.

Does the solution meet the goal?

- A. Yes
- B. No

Answer: B

Explanation:

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

Note: SMOTE is used to increase the number of underrepresented cases in a dataset used for machine learning.

SMOTE is a better way of increasing the number of rare cases than simply duplicating existing cases.

References:

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

QUESTION 73

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

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

You are creating a new experiment in Azure Learning learning Studio. One class has a much smaller number of observations than the other classes in the training You need to select an appropriate data sampling strategy to compensate for the class imbalance.

Solution: You use the Synthetic Minority Oversampling Technique (SMOTE) sampling mode.

Does the solution meet the goal?

- A. Yes
- B. No

Answer: A

Explanation:

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

References:

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

QUESTION 74

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

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

You are creating a new experiment in Azure Machine Learning Studio. One class has a much smaller number of observations than the other classes in the training set. You need to select an appropriate data sampling strategy to compensate for the class imbalance.

Solution: You use the Principal Components Analysis (PCA) sampling mode.

Does the solution meet the goal?

- A. Yes
- B. No

Answer: B

QUESTION 75

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>

QUESTION 76

You are conducting feature engineering to prepare data for further analysis. The data includes seasonal patterns on inventory requirements.

You need to select the appropriate method to conduct feature engineering on the data.

Which method should you use?

- A. Exponential Smoothing (ETS) function.
- B. One Class Support Vector Machine module
- C. Time Series Anomaly Detection module
- D. Finite Impulse Response (FIR) Filter module.

Answer: D

QUESTION 77

You are solving a classification task.

The dataset is imbalanced.

[DP-100 Exam Dumps](#) [DP-100 Exam Questions](#) [DP-100 PDF Dumps](#) [DP-100 VCE Dumps](#)

<https://www.braindump2go.com/dp-100.html>

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>

QUESTION 78

You are performing a filter based feature selection for a dataset 10 build a multi class classifiers by using Azure Machine Learning Studio.

The dataset contains categorical features that are highly correlated to the output label column.

You need to select the appropriate feature scoring statistical method to identify the key predictors.

Which method should you use?

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

Answer: D

Explanation:

Pearson's correlation statistic, or Pearson's correlation coefficient, is also known in statistical models as the r value. For any two variables, it returns a value that indicates the strength of the correlation. Pearson's correlation coefficient is the test statistics that measures the statistical relationship, or association, between two continuous variables. It is known as the best method of measuring the association between variables of interest because it is based on the method of covariance. It gives information about the magnitude of the association, or correlation, as well as the direction of the relationship.

Reference:

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

<https://www.statisticssolutions.com/pearsons-correlation-coefficient/>

QUESTION 79

You plan to use a Data Science Virtual Machine (DSVM) with the open source deep learning frameworks Caffe2 and Theano.

You need to select a pre configured DSVM to support the framework.

What should you create?

- A. Data Science Virtual Machine for Linux (CentOS)
- B. Data Science Virtual Machine for Windows 2012
- C. Data Science Virtual Machine for Windows 2016
- D. Geo AI Data Science Virtual Machine with ArcGIS
- E. Data Science Virtual Machine for Linux (Ubuntu)

Answer: E

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

QUESTION 80

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 CUOA.
What should you implement?

- A. Intel Software Guard Extensions (Intel SGX) technology
- B. Solid State Drives (SSD)
- C. Graphic Processing Unit (GPU)
- D. Computer Processing Unit (CPU) speed increase by using overclocking
- E. High Random Access Memory (RAM) configuration

Answer: C

Explanation:

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

QUESTION 81

You need to select a pre built development environment for a series of data science experiments. You must use the R language for the experiments.
Which three environments can you use? Each correct answer presents a complete solution. NOTE:
Each correct selection is worth one point.

- A. ML.NET Library on a local environment
- B. Azure Machine Learning Studio
- C. Data Science Virtual Machine (DSVM)
- D. Azure Data bricks
- E. Azure Cognitive Services

Answer: ABD

QUESTION 82

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. Apache Drill
- B. Tensorflow
- C. Rattle
- D. Weka

Answer: B

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 library 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>