

- **Vendor: Microsoft**
- **Exam Code: DP-100**
- **Exam Name: Designing and Implementing a Data Science Solution on Azure**
- **New Questions (Dec/2021)**

[Visit PassLeader and Download Full Version DP-100 Exam Dumps](#)

NEW QUESTION 271

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

Answer: B

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

NEW QUESTION 272

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

Answer: B

NEW QUESTION 273

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. Yes
- B. No

Answer: B

Explanation:

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

NEW QUESTION 274

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. Yes
- B. No

Answer: A

Explanation:

Typical metadata changes might include marking columns as features.

<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>

NEW QUESTION 275

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. Yes
- B. No

Answer: A

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.

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

NEW QUESTION 276

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. Yes
- B. No

Answer: B

Explanation:

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

NEW QUESTION 277

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. No adjustment required
- B. Convert to Indicator Values
- C. Apply SQL Transformation
- D. Group Categorical Values

Answer: B

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.

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

NEW QUESTION 278

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. No adjustment required
- B. Remove Duplicate Rows
- C. Join Data
- D. Edit Metadata

Answer: A

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.

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

NEW QUESTION 279

You create a Python script that runs a training experiment in Azure Machine Learning. The script uses the Azure Machine Learning SDK for Python. You must add a statement that retrieves the names of the logs and outputs generated by the script. You need to reference a Python class object from the SDK for the statement. Which class object should you use?

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

Answer: A

Explanation:

A run represents a single trial of an experiment. Runs are used to monitor the asynchronous execution of a trial, log metrics and store output of the trial, and to analyze results and access artifacts generated by the trial. The run Class get_all_logs method downloads all logs for the run to a directory.

Incorrect:

Not A: A run represents a single trial of an experiment. Runs are used to monitor the asynchronous execution of a trial, log metrics and store output of the trial, and to analyze results and access artifacts generated by the trial.

Not B: A ScriptRunConfig packages together the configuration information needed to submit a run in Azure ML, including the script, compute target, environment, and any distributed job-specific configs.

[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 280

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. Choose two.)

- 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.

Incorrect:

Not B: The run `get_file_names` list the files that are stored in association with the run.

[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 281

You create and register a model in an Azure Machine Learning workspace. You must use the Azure Machine Learning SDK to implement a batch inference pipeline that uses a `ParallelRunStep` to score input data using the model. You must specify a value for the `ParallelRunConfig` `compute_target` setting of the pipeline step. You need to create the compute target. Which class should you use?

- A. `BatchCompute`
- B. `AdlaCompute`
- C. `AmlCompute`
- D. `AksCompute`

Answer: C

Explanation:

Compute target to use for `ParallelRunStep`. This parameter may be specified as a compute target object or the string name of a compute target in the workspace. The `compute_target` target is of `AmlCompute` or string. Note: An Azure Machine Learning Compute (`AmlCompute`) is a managed-compute infrastructure that allows you to easily create a single or multi-node compute. The compute is created within your workspace region as a resource that can be shared with other users.

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

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

NEW QUESTION 282

You use the designer to create a training pipeline for a classification model. The pipeline uses a

dataset that includes the features and labels required for model training. You create a real-time inference pipeline from the training pipeline. You observe that the schema for the generated web service input is based on the dataset and includes the label column that the model predicts. Client applications that use the service must not be required to submit this value. You need to modify the inference pipeline to meet the requirement. What should you do?

- A. Add a Select Columns in Dataset module to the inference pipeline after the dataset and use it to select all columns other than the label.
- B. Delete the dataset from the training pipeline and recreate the real-time inference pipeline.
- C. Delete the Web Service Input module from the inference pipeline.
- D. Replace the dataset in the inference pipeline with an Enter Data Manually module that includes data for the feature columns but not the label column.

Answer: A

Explanation:

By default, the Web Service Input will expect the same data schema as the module output data which connects to the same downstream port as it. You can remove the target variable column in the inference pipeline using Select Columns in Dataset module. Make sure that the output of Select Columns in Dataset removing target variable column is connected to the same port as the output of the Web Service Input module.

<https://docs.microsoft.com/en-us/azure/machine-learning/tutorial-designer-automobile-price-deploy>

NEW QUESTION 283

You use the Azure Machine Learning designer to create and run a training pipeline. You then create a real-time inference pipeline. You must deploy the real-time inference pipeline as a web service. What must you do before you deploy the real-time inference pipeline?

- A. Run the real-time inference pipeline.
- B. Create a batch inference pipeline.
- C. Clone the training pipeline.
- D. Create an Azure Machine Learning compute cluster.

Answer: D

Explanation:

<https://docs.microsoft.com/en-us/azure/machine-learning/tutorial-designer-automobile-price-deploy>

NEW QUESTION 284

You create an Azure Machine Learning workspace named ML-workspace. You also create an Azure Databricks workspace named DB-workspace. DB-workspace contains a cluster named DB-cluster. You must use DB-cluster to run experiments from notebooks that you import into DB-workspace. You need to use ML-workspace to track MLflow metrics and artifacts generated by experiments running on DB-cluster. The solution must minimize the need for custom code. What should you do?

- A. From DB-cluster, configure the Advanced Logging option.
- B. From DB-workspace, configure the Link Azure ML workspace option.
- C. From ML-workspace, create an attached compute.
- D. From ML-workspace, create a compute cluster.

Answer: B

Explanation:

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

NEW QUESTION 285

You are planning to register a trained model in an Azure Machine Learning workspace. You must store additional metadata about the model in a key-value format. You must be able to add new metadata and modify or delete metadata after creation. You need to register the model. Which parameter should you use?

- A. description
- B. model_framework
- C. tags
- D. properties

Answer: D

Explanation:

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

NEW QUESTION 286

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. Windows Server 2012 DSVM
- B. Windows Server 2016 DSVM
- C. Ubuntu 16.04 DSVM
- D. CentOS 7.4 DSVM

Answer: C

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.

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

NEW QUESTION 287

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 Rattle.
- B. You should consider including TensorFlow.
- C. You should consider including Theano.
- D. You should consider including Chainer.

Answer: B

NEW QUESTION 288

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 Split Data module.
- B. You should make use of the Group Categorical Values module.
- C. You should make use of the Clip Values module.
- D. You should make use of the Group Data into Bins module.

Answer: D

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.

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

NEW QUESTION 289

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 make use of the Convert to TXT module.
- B. You should add the Copy Data object to the pipeline.
- C. You should add the Import Data object to the pipeline.
- D. You should add the Dataset object to the pipeline.

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.

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

NEW QUESTION 290

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 Venn diagram visualization.
- B. You should consider using Receiver Operating Characteristic (ROC) curve visualization.
- C. You should consider using Box plot visualization.
- D. You should consider using the Binary classification confusion matrix visualization.

Answer: D

Explanation:

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

NEW QUESTION 291

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 disable deep learning.
- B. You should enable automatic featurization.
- C. You should disable automatic featurization.
- D. You should set the task type to Forecasting.

Answer: C

Explanation:

<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

NEW QUESTION 292

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. Boosted Decision Tree Regression
- D. Linear Regression

Answer: C

Explanation:

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

<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>

NEW QUESTION 293

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 Selective grid sweep mode.
- B. You should consider making use of the Measured grid sweep mode.
- C. You should consider making use of the Entire grid sweep mode.
- D. You should consider making use of the Random grid sweep mode.

Answer: D

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.

Incorrect:

Not 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.

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

NEW QUESTION 294

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 L1/L2 regularization, and reduce the amount of training data.
- C. You have to reduce the amount of training data and make use of training data augmentation.
- D. You have to add L1/L2 regularization, and make use of training data augmentation.
- E. You have to add an additional dense layer with 512 input units, and add L1/L2 regularization.

Answer: B

Explanation:

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

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

NEW QUESTION 295

HotSpot

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.)

Answer Area

Requirement	Action
Access the contents of store1 by using the Azure Machine Learning SDK for Python.	<div>▼</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>
Preview the contents of store1 by using Azure Machine Learning studio.	<div>▼</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>

Answer:

Answer Area

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

Explanation:

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.

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

NEW QUESTION 296

Drag and Drop

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.)

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

Answer:

Target datasets

features-dataset

Answer Area

Baseline dataset

training-dataset

Target dataset

predictions-dataset

Explanation:

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.

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

NEW QUESTION 297

.....

[**Visit PassLeader and Download Full Version DP-100 Exam Dumps**](#)