DP-100 Q&As

Designing and Implementing a Data Science Solution on Azure

Pass Microsoft DP-100 Exam with 100% Guarantee

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

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

100% Passing Guarantee
100% Money Back Assurance

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

- **Instant Download** After Purchase
- **100% Money Back** Guarantee
- **365 Days** Free Update
- **800,000+** Satisfied Customers

QUESTION 1

You create a binary classification model by using Azure Machine Learning Studio.

You must tune hyperparameters by performing a parameter sweep of the model. The parameter sweep must meet the following requirements:

1.

iterate all possible combinations of hyperparameters

2.

minimize computing resources required to perform the sweep

You need to perform a parameter sweep of the model.

Which parameter sweep mode should you use?

A. Random sweep

B. Sweep clustering

C. Entire grid

D. Random grid

Correct Answer: D

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 Answers:

B: If you are building a clustering model, use Sweep Clustering to automatically determine the optimum number of clusters and other parameters.

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

QUESTION 2

You have a Jupyter Notebook that contains Python code that is used to train a model.

You must create a Python script for the production deployment. The solution must minimize code maintenance.

Which two actions should you perform? Each correct answer presents part of the solution.

NOTE: Each correct selection is worth one point.

A. Refactor the Jupyter Notebook code into functions

B. Save each function to a separate Python file

C. Define a main() function in the Python script

D. Remove all comments and functions from the Python script

Correct Answer: AC

C: Python main function is a starting point of any program. When the program is run, the python interpreter runs the code sequentially. Main function is executed only when it is run as a Python program.

A: Refactoring, code style and testing

The first step is to modularise the notebook into a reasonable folder structure, this effectively means to convert files from .ipynb format to .py format, ensure each script has a clear distinct purpose and organise these files in a coherent way.

Once the project is nicely structured we can tidy up or refactor the code.

Reference: https://www.guru99.com/learn-python-main-function-with-examples-understand-main.html
https://towardsdatascience.com/from-jupyter-notebook-to-deployment-a-straightforward-example-1838c203a437

QUESTION 3

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 \\'off\\' 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 \\'FeaturizationConfig\\' value pair.

Correct Answer: A

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-client/azureml.train.automl.automlconfig.automlconfig

QUESTION 4

DRAG 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

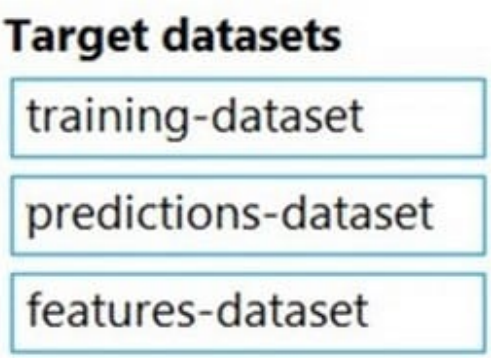
predictions-dataset with a schema that matches the training data exactly, including the predicted label; and another named features-dataset with a schema containing all of the feature columns and a timestamp column based on the filename, which includes the day, month, and year.

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.

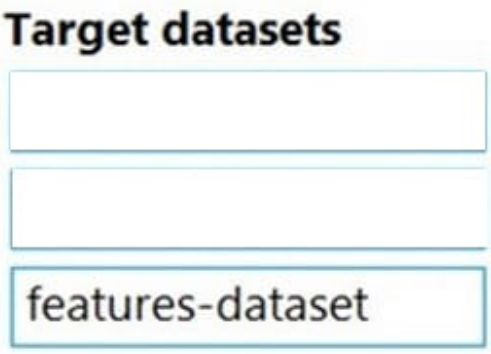
Which datasets should you use to configure the data drift monitor? To answer, drag the appropriate datasets to the correct data drift monitor options. Each source may be used once, more than once, or not at all. You may need to drag the

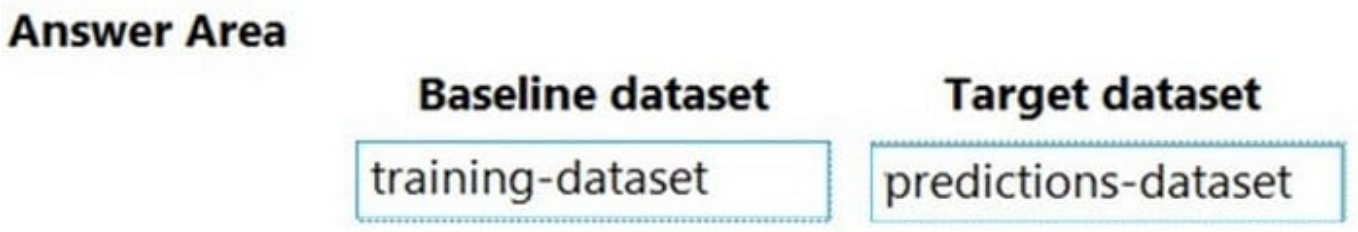
split bar between panes or scroll to view content.

NOTE: Each correct selection is worth one point. Select and Place:



Correct Answer:





Box 1: training-dataset

Baseline dataset - usually the training dataset for a model.

Box 2: predictions-dataset

Target dataset - usually model input data - is compared over time to your baseline dataset. This comparison means that your target dataset must have a timestamp column specified. The monitor will compare the baseline and target datasets.

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

QUESTION 5

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

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

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

Which evaluation statement should you use? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.

Hot Area:

Answer Area

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

Correct Answer:

Answer Area

Code Segment	Evaluation Statement
class_weight=balanced	Automatically adjust weights inversely proportional to class frequences in the input data.
C parameter	Penalty parameter

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

The “balanced” mode uses the values of y to automatically adjust weights inversely proportional to class frequencies in the input data as n_samples / (n_classes * np.bincount(y)).

Box 2: Penalty parameter

Parameter: C : float, optional (default=1.0)

Penalty parameter C of the error term.

References:

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

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

Algorithm	Mean Absolute Error	Root Mean Squared Error	Relative Absolute Error	Relative Squared Error
Bayesian Liner	3.276025	4.655442	0.511436	0.282138
Neural Network	2.676538	3.621476	0.417847	0.17073
Boosted Decision Tree	2.168847	2.878077	0.338589	0.107831
Linear	6.350005	8.720718	0.99133	0.99002
Decision Forest	2.390206	3.315 164	0.373146	0.14307