

New DP-100 Test Prep - Valid Exam DP-100 Preparation



BONUS!!! Download part of Rea4exams DP-100 dumps for free: <https://drive.google.com/open?id=1kQDBrryHv5eM7FDI0T70yzq1C0m4XeFc>

With Rea4exams's Microsoft DP-100 Exam Training materials you can pass the Microsoft DP-100 exam easily. The training tools which designed by our website can help you pass the exam the first time. You only need to download the Rea4exams Microsoft DP-100 exam training materials, namely questions and answers, the exam will become very easy. Rea4exams guarantee that you will be able to pass the exam. If you are still hesitant, download our sample of material, then you can know the effect. Do not hesitate, add the exam material to your shopping cart quickly. If you miss it you will regret for a lifetime.

Do not waste further time and money, get real Microsoft DP-100 pdf questions and practice test software, and start DP-100 test preparation today. Rea4exams will also provide you with up to 365 days of free exam questions updates. Free demo of DP-100 Dumps PDF allowing you to try before you buy and one-year free update will be allowed after purchased.

>> New DP-100 Test Prep <<

Valid Exam Microsoft DP-100 Preparation | DP-100 Latest Test Cost

Before clients buy our DP-100 questions torrent they can download them and try out them freely. The pages of our product provide the demo and the aim is to let the client know part of our titles before their purchase and what form our DP-100 guide torrent is. You can visit our website and read the pages of our product. The pages introduce the quantity of our questions and answers of our DP-100 Guide Torrent, the time of update, the versions for you to choose and the price of our product. After you try out the free demo you could decide whether our DP-100 exam torrent is worthy to buy or not. So you needn't worry that you will waste your money or our DP-100 exam torrent is useless and boosts no values.

To prepare for the DP-100 Exam, candidates should have a solid understanding of machine learning concepts, programming languages such as Python and R, and Azure services such as Azure Machine Learning and Azure Databricks. Microsoft offers a range of training resources, including online courses and practice exams, to help candidates prepare for the exam.

Microsoft Designing and Implementing a Data Science Solution on Azure Sample Questions (Q300-Q305):

NEW QUESTION # 300

You need to implement a feature engineering strategy for the crowd sentiment local models. What should you do?

- A. Apply a Pearson correlation coefficient.
- **B. Apply a linear discriminant analysis.**
- C. Apply a Spearman correlation coefficient.
- D. Apply an analysis of variance (ANOVA).

Answer: B

Explanation:

Explanation/Reference:

Explanation:

The linear discriminant analysis method works only on continuous variables, not categorical or ordinal variables.

Linear discriminant analysis is similar to analysis of variance (ANOVA) in that it works by comparing the means of the variables.

Scenario:

Data scientists must build notebooks in a local environment using automatic feature engineering and model building in machine learning pipelines.

Experiments for local crowd sentiment models must combine local penalty detection data.

All shared features for local models are continuous variables.

Incorrect Answers:

B: The Pearson correlation coefficient, sometimes called Pearson's R test, is a statistical value that measures the linear relationship between two variables. By examining the coefficient values, you can infer something about the strength of the relationship between the two variables, and whether they are positively correlated or negatively correlated.

C: Spearman's correlation coefficient is designed for use with non-parametric and non-normally distributed data. Spearman's coefficient is a nonparametric measure of statistical dependence between two variables, and is sometimes denoted by the Greek letter rho. The Spearman's coefficient expresses the degree to which two variables are monotonically related. It is also called Spearman rank correlation, because it can be used with ordinal variables.

References:

<https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/fisher-linear-discriminant-analysis>

<https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/compute-linear-correlation> Testlet 2 Case study

Overview You are a data scientist for Fabrikam Residences, a company specializing in quality private and commercial property in the United States. Fabrikam Residences is considering expanding into Europe and has asked you to investigate prices for private residences in major European cities. You use Azure Machine Learning Studio to measure the median value of properties. You produce a regression model to predict property prices by using the Linear Regression and Bayesian Linear Regression modules.

Datasets

There are two datasets in CSV format that contain property details for two cities, London and Paris, with the following columns:

Column heading	Description
CapitaCrimeRate	per capita crime rate by town
Zoned	proportion of residential land zoned for lots over 25,000 square feet
NonRetailAcres	proportion of retail business acres per town
NextToRiver	proximity of the property to the river
NitrogenOxideConcentration	nitric oxides concentration (parts per 10 million)
AvgRoomsPerHouse	average number of rooms per dwelling
Age	proportion of owner-occupied units built prior to 1940
DistanceToEmploymentCenter	weighted distances to employment centers
AccessibilityToHighway	index of accessibility to radial highways to a value of two decimal places
Tax	full value property tax rate per \$10,000
PupilTeacherRatio	pupil to teacher ratio by town
ProfessionalClass	professional class percentage
LowerStatus	percentage lower status of the population
MedianValue	median value of owner-occupied homes in \$1000s

The two datasets have been added to Azure Machine Learning Studio as separate datasets and included as the starting point of the experiment.

Dataset issues

The AccessibilityToHighway column in both datasets contains missing values. The missing data must be replaced with new data so that it is modeled conditionally using the other variables in the data before filling in the missing values.

Columns in each dataset contain missing and null values. The dataset also contains many outliers. The Age column has a high proportion of outliers. You need to remove the rows that have outliers in the Age column. The MedianValue and AvgRoomsInHouse columns both hold data in numeric format. You need to select a feature selection algorithm to analyze the relationship between the two columns in more detail.

Model fit

The model shows signs of overfitting. You need to produce a more refined regression model that reduces the overfitting.

Experiment requirements

You must set up the experiment to cross-validate the Linear Regression and Bayesian Linear Regression modules to evaluate performance.

In each case, the predictor of the dataset is the column named MedianValue. An initial investigation showed that the datasets are identical in structure apart from the MedianValue column. The smaller Paris dataset contains the MedianValue in text format, whereas the larger London dataset contains the MedianValue in numerical format. You must ensure that the datatype of the MedianValue column of the Paris dataset matches the structure of the London dataset.

You must prioritize the columns of data for predicting the outcome. You must use non-parameters statistics to measure the relationships.

You must use a feature selection algorithm to analyze the relationship between the MedianValue and AvgRoomsInHouse columns.

Model training

Given a trained model and a test dataset, you need to compute the permutation feature importance scores of feature variables. You need to set up the Permutation Feature Importance module to select the correct metric to investigate the model's accuracy and replicate the findings.

You want to configure hyperparameters in the model learning process to speed the learning phase by using hyperparameters. In addition, this configuration should cancel the lowest performing runs at each evaluation interval, thereby directing effort and resources towards models that are more likely to be successful.

You are concerned that the model might not efficiently use compute resources in hyperparameter tuning.

You also are concerned that the model might prevent an increase in the overall tuning time. Therefore, you need to implement an early stopping criterion on models that provides savings without terminating promising jobs.

Testing

You must produce multiple partitions of a dataset based on sampling using the Partition and Sample module in Azure Machine Learning Studio. You must create three equal partitions for cross-validation. You must also configure the cross-validation process so that the rows in the test and training datasets are divided evenly by properties that are near each city's main river. The data that identifies that a property is near a river is held in the column named NextToRiver. You want to complete this task before the data goes through the sampling process.

When you train a Linear Regression module using a property dataset that shows data for property prices for a large city, you need to determine the best features to use in a model. You can choose standard metrics provided to measure performance before and after the feature importance process completes. You must ensure that the distribution of the features across multiple training models is consistent.

Data visualization

You need to provide the test results to the Fabrikam Residences team. You create data visualizations to aid in presenting the results.

You must produce a Receiver Operating Characteristic (ROC) curve to conduct a diagnostic test evaluation of the model. You need to select appropriate methods for producing the ROC curve in Azure Machine Learning Studio to compare the Two-Class Decision Forest and the Two-Class Decision Jungle modules with one another.

NEW QUESTION # 301

You are tuning a hyperparameter for an algorithm. The following table shows a data set with different hyperparameter, training error, and validation errors.

Hyperparameter (H)	Training error (TE)	Validation error (VE)
1	105	95
2	200	85
3	250	100
4	105	100
5	400	50

Use the drop-down menus to select the answer choice that answers each question based on the information presented in the graphic.

Question

Answer Choice

Which H value should you select based on the data?

What H value displays the poorest training result?



Answer:

Explanation:

Question

Answer Choice

Which H value should you select based on the data?

	▼
1	
2	
3	
4	
5	

What H value displays the poorest training result?

	▼
1	
2	
3	
4	
5	



Explanation:

Question

Answer Choice

Which H value should you select based on the data?

	▼
1	
2	
3	
4	
5	

What H value displays the poorest training result?

	▼
1	
2	
3	
4	
5	

Box 1: 4

Choose the one which has lower training and validation error and also the closest match.

Minimize variance (difference between validation error and train error).

Box 2: 5

Minimize variance (difference between validation error and train error).

Reference:

<https://medium.com/comet-ml/organizing-machine-learning-projects-project-management-guidelines-2d2b85651>

NEW QUESTION # 302

You have an Azure blob container that contains a set of TSV files. The Azure blob container is registered as a datastore for an Azure Machine Learning service workspace. Each TSV file uses the same data schema.

You plan to aggregate data for all of the TSV files together and then register the aggregated data as a dataset in an Azure Machine Learning workspace by using the Azure Machine Learning SDK for Python.

You run the following code.

```

from azureml.core.workspace import Workspace
from azureml.core.datastore import Datastore
from azureml.core.dataset import Dataset
import pandas as pd
datastore_paths = (datastore, './data/*.tsv')
myDataset_1 = Dataset.File.from_files(path=datastore_paths)
myDataset_2 = Dataset.Tabular.from_delimited_files(path=datastore_paths, separator='\t')

```

For each of the following statements, select Yes if the statement is true. Otherwise, select No.

NOTE: Each correct selection is worth one point.

	Yes	No
The myDataset_1 dataset can be converted into a pandas dataframe by using the following method: using myDataset_1.to_pandas_dataframe()	<input type="radio"/>	<input type="radio"/>
The myDataset_1.to_path() method returns an array of file paths for all of the TSV files in the dataset.	<input type="radio"/>	<input type="radio"/>
The myDataset_2 dataset can be converted into a pandas dataframe by using the following method: myDataset_2.to_pandas_dataframe()	<input type="radio"/>	<input type="radio"/>

Answer:

Explanation:

	Yes	No
The myDataset_1 dataset can be converted into a pandas dataframe by using the following method: using myDataset_1.to_pandas_dataframe()	<input type="radio"/>	<input checked="" type="radio"/>
The myDataset_1.to_path() method returns an array of file paths for all of the TSV files in the dataset.	<input checked="" type="radio"/>	<input type="radio"/>
The myDataset_2 dataset can be converted into a pandas dataframe by using the following method: myDataset_2.to_pandas_dataframe()	<input checked="" type="radio"/>	<input type="radio"/>

Reference:

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

NEW QUESTION # 303

You are evaluating a completed binary classification machine learning model.

You need to use the precision as the evaluation metric.

Which visualization should you use?

- A. Scatter plot
- B. violin plot
- C. Gradient descent
- D. Receiver Operating Characteristic (ROC) curve

Answer: D

Explanation:

Receiver operating characteristic (or ROC) is a plot of the correctly classified labels vs. the incorrectly classified labels for a particular model.

Incorrect Answers:

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

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

C: A scatter plot graphs the actual values in your data against the values predicted by the model.

The scatter plot displays the actual values along the X-axis, and displays the predicted values along the Y-axis.

It also displays a line that illustrates the perfect prediction, where the predicted value exactly matches the actual value.

References:

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

NEW QUESTION # 304

Your Azure Machine Learning workspace has a dataset named `real_estate_data`. A sample of the data in the dataset follows.

postal_code	num_bedrooms	sq_feet	garage	price
12345	3	1300	0	23,9000
54321	1	950	0	11,0000
12346	2	1200	1	15,0000

You want to use automated machine learning to find the best regression model for predicting the price column.

You need to configure an automated machine learning experiment using the Azure Machine Learning SDK.

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

NOTE: Each correct selection is worth one point.

Answer Area

```
from azureml.core import Workspace
from azureml.core.compute import ComputeTarget
from azureml.core.runconfig import RunConfiguration
from azureml.train.automl import AutoMLConfig

ws = Workspace.from_config()
training_cluster = ComputeTarget(workspace=ws, name='aml-cluster1')
real_estate_ds = ws.datasets.get('real_estate_data')
split1_ds, split2_ds = real_estate_ds.random_split(percentage=0.7, seed=123)
automl_run_config = RunConfiguration(
    framework="python",
    task='regression',
    compute_target=training_cluster,
    run_configuration=automl_run_config,
    primary_metric='r2_score',
    =split1_ds,
```

X
Y
X_valid
Y_valid
training_data

=split2_ds

X
Y
X_valid
Y_valid
validation_data
training_data

='price')

y
y_valid
y_max
label_column_name
exclude_nan_labels

Answer:

Explanation:

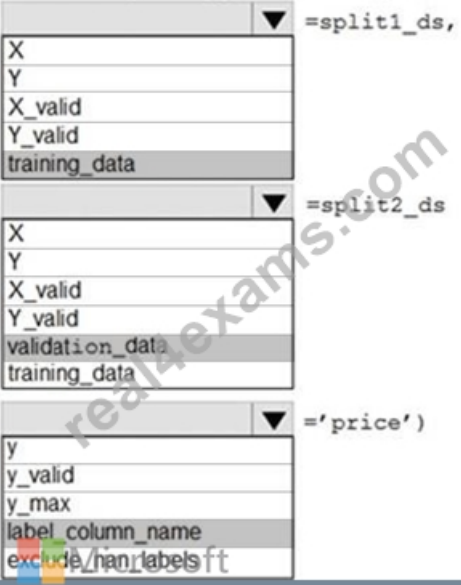
Answer Area

```
from azureml.core import Workspace
from azureml.core.compute import ComputeTarget
from azureml.core.runconfig import RunConfiguration
from azureml.train.automl import AutoMLConfig

ws = Workspace.from_config()
training_cluster = ComputeTarget(workspace=ws, name='aml-cluster1')
real_estate_ds = ws.datasets.get('real_estate_data')
split1_ds, split2_ds = real_estate_ds.random_split(percentage=0.7, seed=123)
automl_run_config = RunConfiguration(framework="python")
automl_config = AutoMLConfig(
    task='regression',
    compute_target=training_cluster,
    run_configuration=automl_run_config,
    primary_metric='r2_score',
    training_data=split1_ds,
    validation_data=split2_ds,
    label_column_name='price',
    exclude_nnan_labels=True)
```



Explanation:



Box 1: training_data

The training data to be used within the experiment. It should contain both training features and a label column (optionally a sample weights column). If training_data is specified, then the label_column_name parameter must also be specified.

Box 2: validation_data

Provide validation data: In this case, you can either start with a single data file and split it into training and validation sets or you can provide a separate data file for the validation set. Either way, the validation_data parameter in your AutoMLConfig object assigns which data to use as your validation set.

Example, the following code example explicitly defines which portion of the provided data in dataset to use for training and validation.

```
dataset = Dataset.Tabular.from_delimited_files(data)
```

```
training_data, validation_data = dataset.random_split(percentage=0.8, seed=1) automl_config = AutoMLConfig(compute_target =
```

