DP-100受験対策解説集、DP-100対策学習

BONUS！！！ ShikenPASS DP-100ダンプの一部を無料でダウンロード：https://drive.google.com/open?id=1iFTcvlUxkPdEX3MQLGnJCXxf0mzD3E3i

当社Microsoftの専門家は、DP-100トレーニング資料を毎日更新し、最新の更新をタイムリーに提供します。当社の製品および購入手順に関する疑問または質問がある場合は、いつでも当社のオンライン顧客サービス担当者にご連絡ください。古いクライアントに割引を提供します。購入前にDP-100テスト問題を無料でダウンロードして試用できます。したがって、当社の製品には多くのメリットがあります。 DP-100試験問題を購入する前に、無料デモでDP-100模擬テストの特性と機能を知ることができます。

Microsoft DP-100 Exam Syllabus Topics:

Section	Objectives
Topic 1: Train machine learning models	- Tune hyperparameters and evaluate models - Train models using Azure Machine Learning
Topic 2: Optimize and manage models	- Track experiments and manage model lifecycle - Improve model performance
Topic 3: Deploy and consume models	- Deploy models to endpoints - Monitor deployed models and endpoints

Topic 4: Design and prepare a machine learning solution	- Plan and configure Azure Machine Learning workspace - Manage compute and data assets - Select appropriate Azure services for machine learning workloads
Topic 5: Explore and analyze data	- Perform exploratory data analysis - Ingest and prepare data for modeling

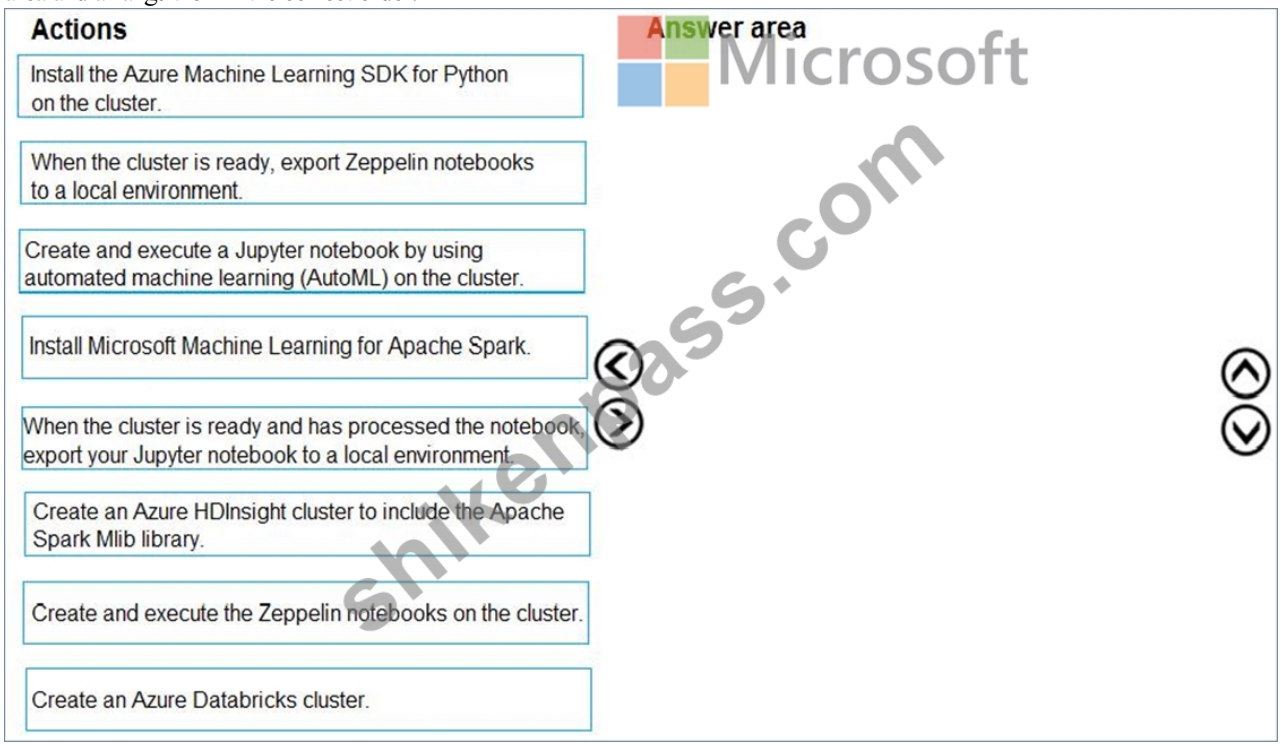
>> DP-100受験対策解説集 <<

DP-100対策学習 & DP-100試験問題集

私たちの専門家は、あなたがDP-100テストのわずかな変更に追いつくことができるように、日々献身的な最新情報を提供するよう努めています。したがって、お客様は生産性が高く効率的なユーザーエクスペリエンスを楽しむことができます。この状況では、お客様の提案と需要が合理的である限り、1年間の更新システムを無料でお楽しみいただけることを保証する義務があります。 DP-100テスト準備を購入した後、DP-100試験問題を購入してから1年間、無料アップデートをお楽しみいただけます。

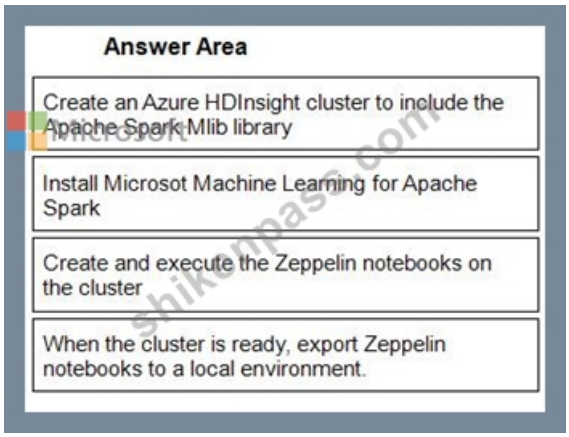
Microsoft Designing and Implementing a Data Science Solution on Azure 認定 DP-100 試験問題 (Q448-Q453):

質問 # 448
You are building an intelligent solution using machine learning models.
The environment must support the following requirements:
Data scientists must build notebooks in a cloud environment
Data scientists must use automatic feature engineering and model building in machine learning pipelines.
Notebooks must be deployed to retrain using Spark instances with dynamic worker allocation.
Notebooks must be exportable to be version controlled locally.
You need to create the environment.
Which four actions should you perform in sequence? To answer, move the appropriate actions from the list of actions to the answer area and arrange them in the correct order.



正解：

解説：



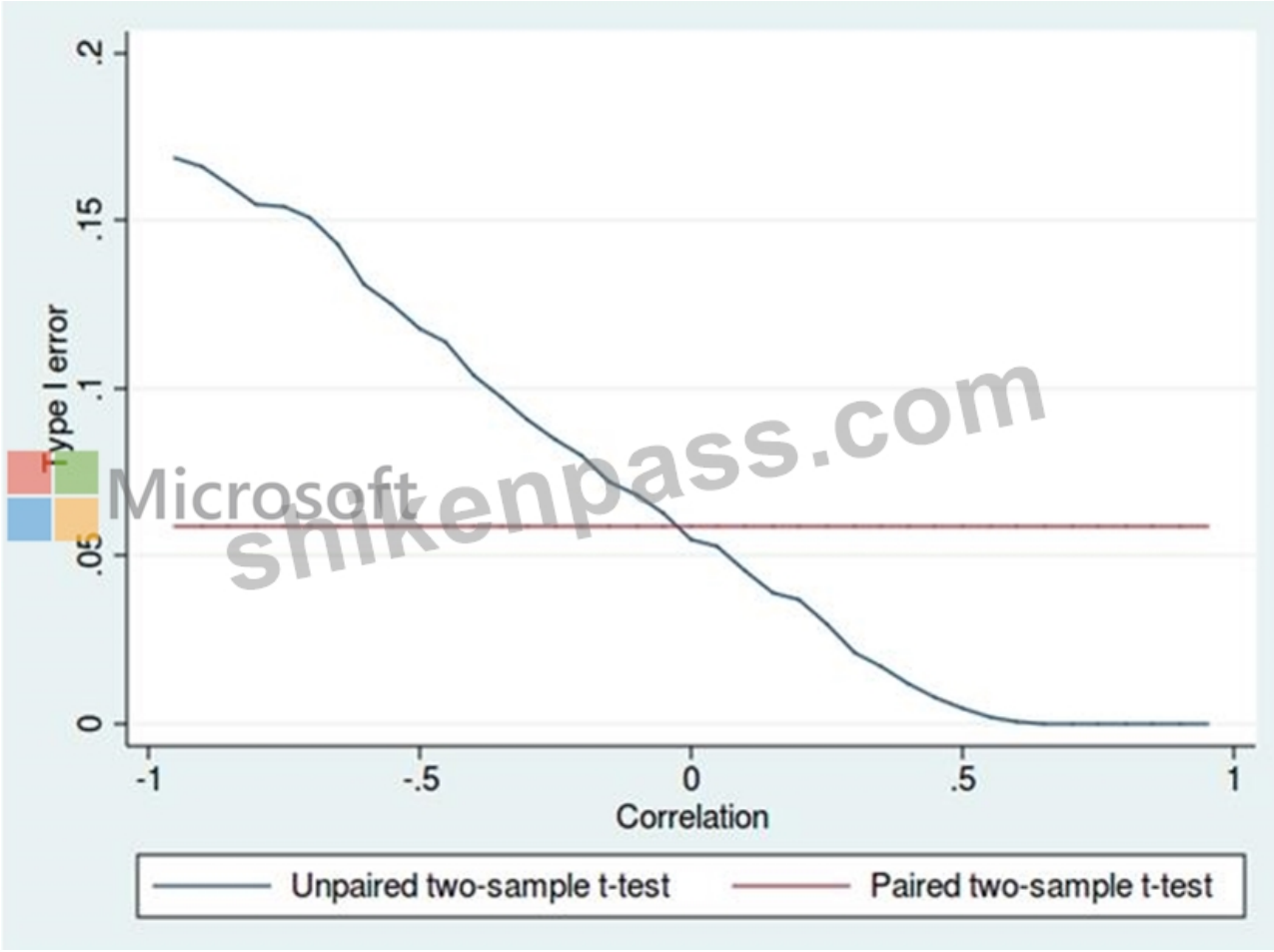
1 - Create an Azure HDInsight cluster to include the Apache Spark Mlib library
2 - Install Microsot Machine Learning for Apache Spark
3 - Create and execute the Zeppelin notebooks on the cluster
4 - When the cluster is ready, export Zeppelin notebooks to a local environment.
References:
https://docs.microsoft.com/en-us/azure/hdinsight/spark/apache-spark-zeppelin-notebook
https://azuremlbuild.blob.core.windows.net/pysparkapi/intro.html

質問 # 449
You are determining if two sets of data are significantly different from one another by using Azure Machine Learning Studio.
Estimated values in one set of data may be more than or less than reference values in the other set of data.
You must produce a distribution that has a constant Type I error as a function of the correlation.
You need to produce the distribution.
Which type of distribution should you produce?

- A. Paired t-test with a one-tail option
- B. Paired t-test with a two-tail option
- C. Unpaired t-test with a one-tail option
- D. Unpaired t-test with a two tail option

正解：B

解説：
Choose a one-tail or two-tail test. The default is a two-tailed test. This is the most common type of test, in which the expected distribution is symmetric around zero.
Example: Type I error of unpaired and paired two-sample t-tests as a function of the correlation. The simulated random numbers originate from a bivariate normal distribution with a variance of 1.

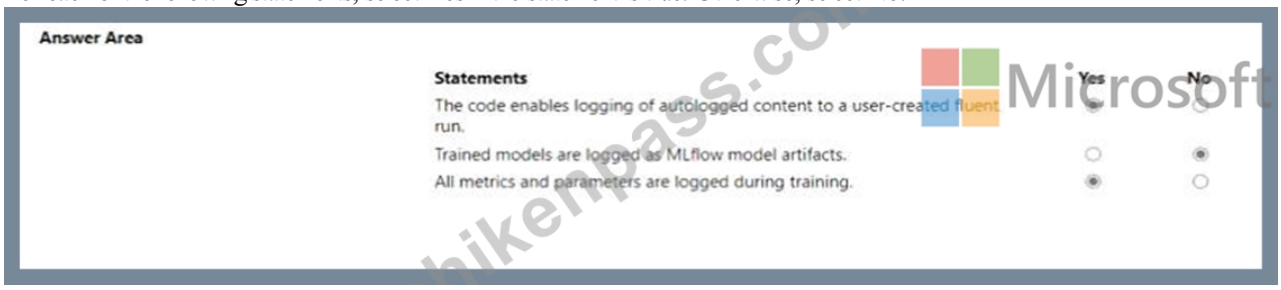


Reference:
https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/test-hypothesis-using-t-test
https://en.wikipedia.org/wiki/Student%27s_t-test

質問 # 450
You manage an Azure Machine Learning workspace. You submit a training job with the Azure Machine Learning Python SDK v2.
You must use MLflow to log metrics, model parameters, and mode! artifacts automatically when training a model.
You start by writing the following code segment:

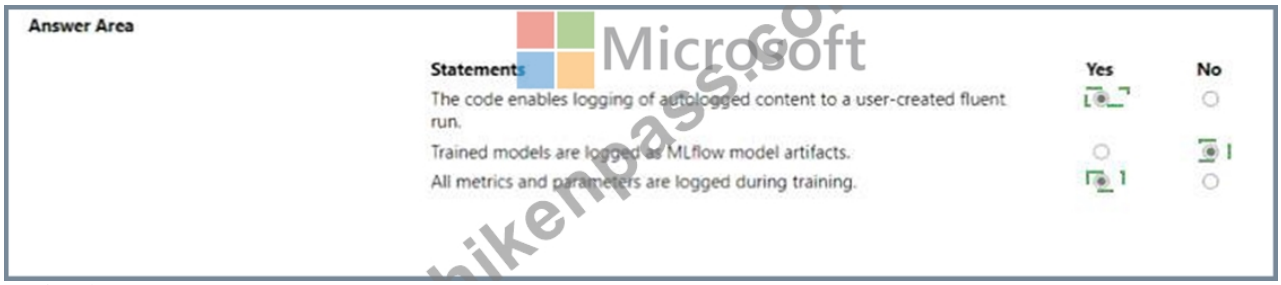
```
import mlflow
mlflow.autolog(log_models=False, exclusive=True)
```

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

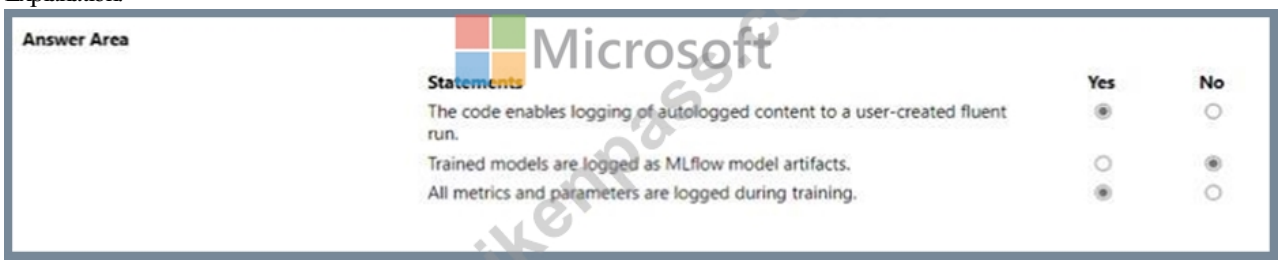


正解：

解説：



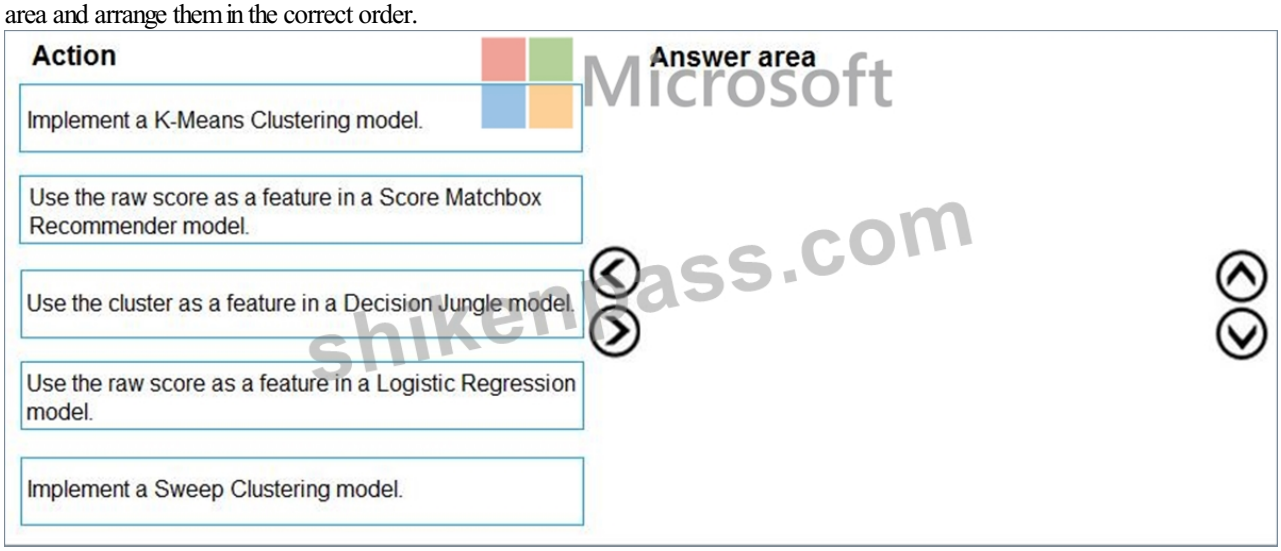
Explanation:



質問 # 451

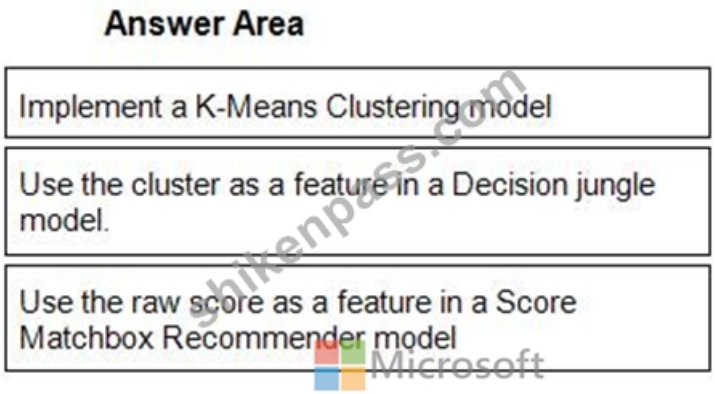
You need to define a modeling strategy for ad response.

Which three actions should you perform in sequence? To answer, move the appropriate actions from the list of actions to the answer area and arrange them in the correct order.



正解：

解説：



1 - Implement a K-Means Clustering model

2 - Use the cluster as a feature in a Decision jungle model.

3 - Use the raw score as a feature in a Score Matchbox Recommender model Reference:

https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/multiclass-decision-jungle

https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/score-matchbox-recommender

Topic 2, Overview

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.

質問 # 452

You collect data from a nearby weather station. You have a pandas dataframe named weather_df that includes the following data:

Temperature	Observation_time	Humidity	Pressure	Visibility	Days_since_last observation
74	2019/10/2 00:00	0.62	29.87	3	0.5
89	2019/10/2 12:00	0.70	28.88	10	0.5
72	2019/10/3 00:00	0.64	30.00	8	0.5
80	2019/10/3 12:00	0.66	29.75	7	0.5

The data is collected every 12 hours: noon and midnight.

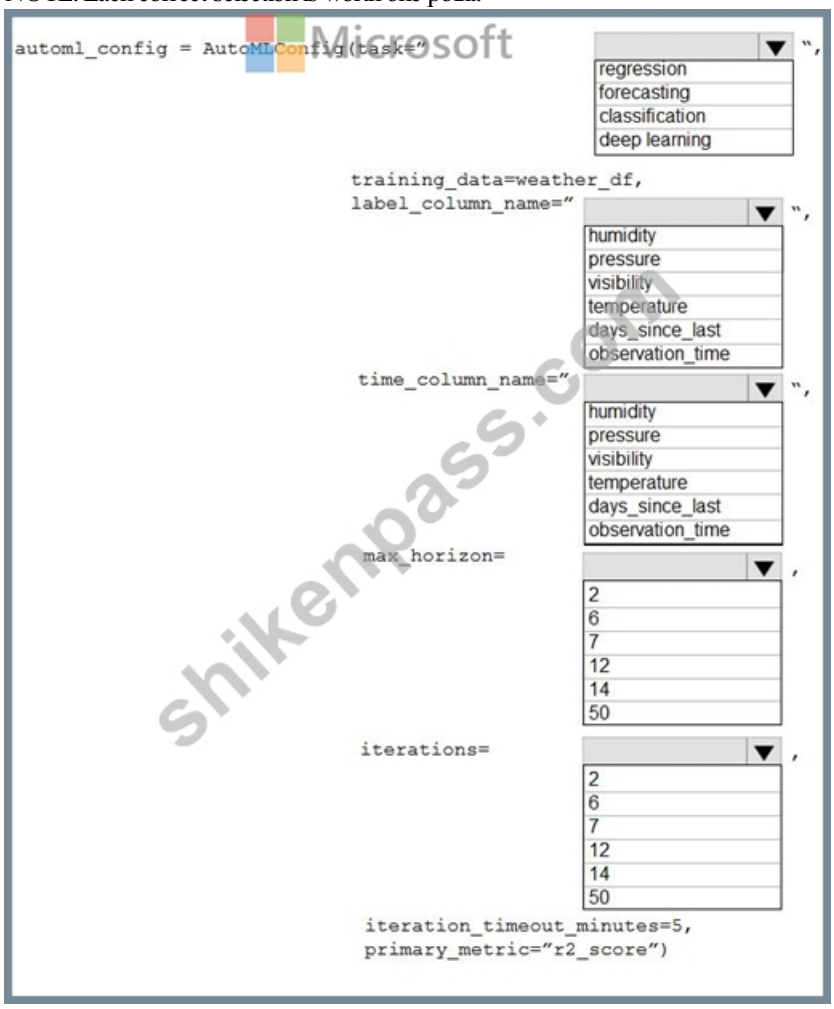
You plan to use automated machine learning to create a time-series model that predicts temperature over the next seven days. For the initial round of training, you want to train a maximum of 50 different models.

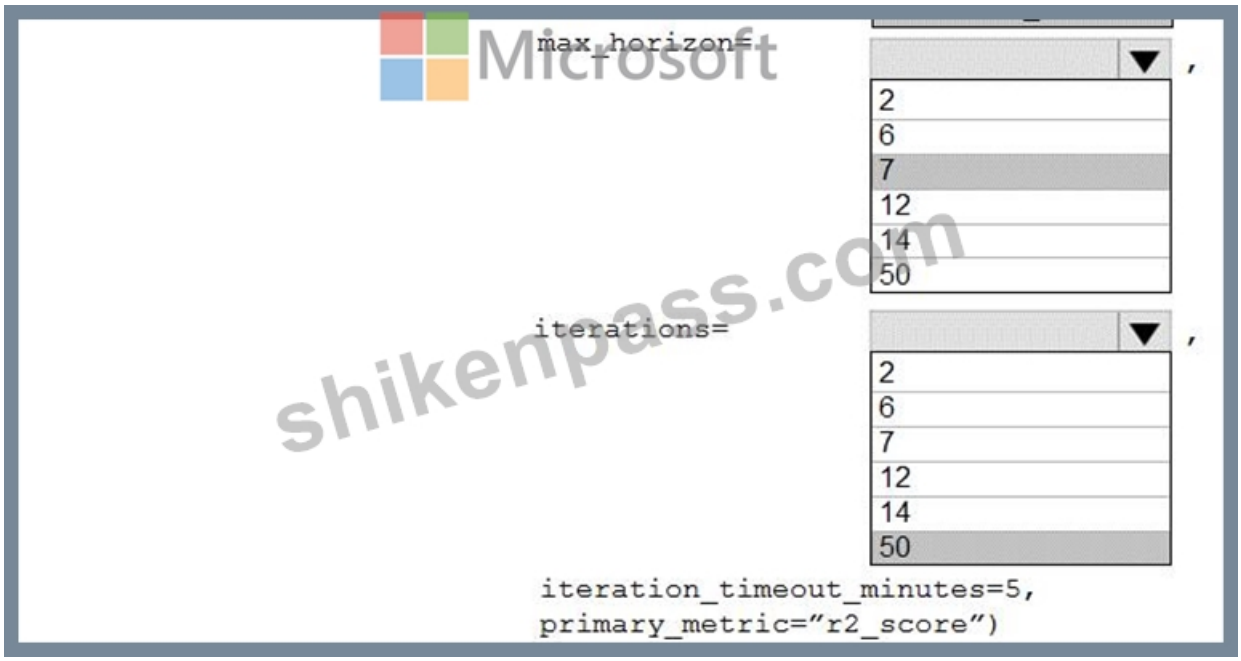
You must use the Azure Machine Learning SDK to run an automated machine learning experiment to train these models.

You need to configure the automated machine learning run.

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

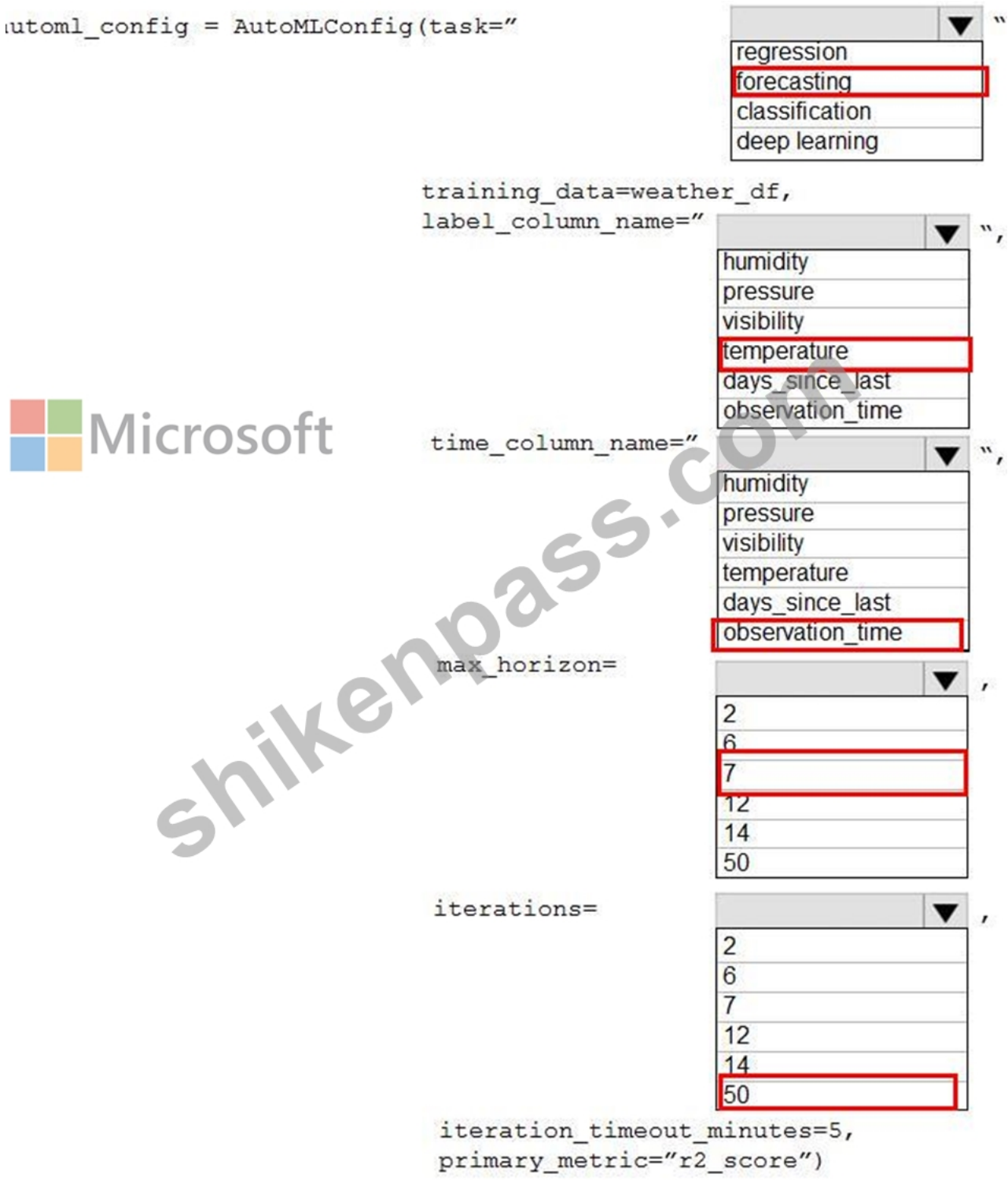
NOTE: Each correct selection is worth one point.





正解：

解説：



Explanation:
Box 1: forcasting
Task: The type of task to run. Values can be 'classification', 'regression', or 'forecasting' depending on the type of automated ML problem to solve.
Box 2: temperature
The training data to be used within the experiment. It should contain both training features and a label column (optionally a sample weights column).
Box 3: observation_time
time_column_name: The name of the time column. This parameter is required when forecasting to specify the datetime column in the input data used for building the time series and inferring its frequency. This setting is being deprecated. Please use forecasting_parameters instead.
Box 4: 7
"predicts temperature over the next seven days"

max_horizon: The desired maximum forecast horizon in units of time-series frequency. The default value is 1.
Units are based on the time interval of your training data, e.g., monthly, weekly that the forecaster should predict out. When task type is forecasting, this parameter is required.
Box 5: 50
"For the initial round of training, you want to train a maximum of 50 different models." Iterations: The total number of different algorithm and parameter combinations to test during an automated ML experiment.
Reference:
https://docs.microsoft.com/en-us/python/api/azureml-train-automl-client/azureml.train.automl.automlconfig.automlconfig

質問 # 453
.....