Microsoft

Exam Questions 70-774

Perform Cloud Data Science with Azure Machine Learning (beta)

NEW QUESTION 1
Note: This question is part of a series of questions that use the same or similar answer choices. An answer choice may be correct for more than one question in the series. Each question is independent of the other questions in this series. Information and details provided in a question apply only to that question.
You have a non-tabular file that is saved in Azure Blob storage.
You need to download the file locally, access the data in the file, and then format the data as a dataset. Which module should you use?

A. Execute Python Script
B. Tune Model Hyperparameters
C. Normalize Data
D. Select Columns in Dataset
E. Import Data
F. Edit Metadata
G. Clip Values
H. Clean Missing Data

Answer: E

Explanation: References:
https://msdn.microsoft.com/en-us/library/azure/mt674698.aspx

NEW QUESTION 2
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 sections, you will NOT be able to return to it. As a result, these questions will not appear in the review screen.
You are designing an Azure Machine Learning workflow.
You have a dataset that contains two million large digital photographs.
You plan to detect the presence of trees in the photographs. You need to ensure that your model supports the following:
Solution: You create an Azure notebook that supports the Microsoft Cognitive Toolkit. Does this meet the goal?

A. Yes
B. No

Answer: B

NEW QUESTION 3
Note: This question is part of a series of questions that use the same scenario. For your convenience, the scenario is repeated in each question. Each question presents a different goal and answer choices, but the text of the scenario is exactly the same in each question in this series.
Start of repeated scenario
You plan to use Azure platform tools to detect and analyze food items in smart refrigerators. To provide families with an integrated experience for grocery shopping and cooking, the refrigerators will connect to other smart appliances, such as stoves and microwave ovens, on a LAN.
You plan to build an object recognition model by using the Microsoft Cognitive Toolkit. The object recognition model will receive input from the connected devices and send results to applications.
The training data will be derived from more than 10 TB of images. You will convert the raw images to the sparse format.
End of repeated scenario.
You need to preprocess the training data by using a Principal Component Analysis (PCA) algorithm in the least amount of time possible. Which implementation method should you use?

A. Azure HDInsight using HiveML
B. Azure Machine Learning Studio and a custom Execute Python Script module
C. Azure HDInsight using Microsoft R Server
D. Azure Machine Learning Studio with a custom Execute R Script module

Answer: C

NEW QUESTION 4
You are analyzing taxi trips in New York City. You leverage the Azure Data Factory to create data pipelines and to orchestrate data movement.
You plan to develop a predictive model for 170 million rows (37 GB) of raw data in Apache Hive by using Microsoft R Server to identify which factors contribute to the passenger tipping behavior.
All of the platforms that are used for the analysis are the same. Each worker node has eight processor cores and 26 GB of memory.
Which type of Azure HDInsight cluster should you use to produce results as quickly as possible?

A. Hadoop
B. HBase
C. Interactive Hive
D. Spark

Answer: D

Explanation: References:
https://azure.microsoft.com/en-gb/blog/general-availability-of-hdinsight-interactive-query-blazing-fast-data-war

NEW QUESTION 5
Note: This question is part of a series of questions that use the same scenario. For your convenience, the scenario is repeated in each question. Each question presents a different goal and answer choices, but the text of the scenario is exactly the same in each question in this series.
You plan to create a predictive analytics solution for credit risk assessment and fraud prediction in Azure Machine Learning. The Machine Learning workspace for

the solution will be shared with other users in your organization. You will add assets to projects and conduct experiments in the workspace.

The experiments will be used for training models that will be published to provide scoring from web services. The experiment for fraud prediction will use Machine Learning modules and APIs to train the models and will predict probabilities in an Apache Hadoop ecosystem.

End of repeated scenario.

You need to alter the list of columns that will be used for predicting fraud for an input web service endpoint. The columns from the original data source must be retained while running the Machine Learning experiment.

Which module should you add after the web service input module and before the prediction module?

A. Edit Metadata
B. Import Data
C. SMOTE
D. Select Columns in Dataset

Answer: D

NEW QUESTION 6

Note: This question is part of a series of questions that use the same scenario. For your convenience, the scenario is repeated in each question. Each question presents a different goal and answer choices, but the text of the scenario is exactly the same in each question in this series.

Start of repeated scenario

You plan to use Azure platform tools to detect and analyze food items in smart refrigerators. To provide families with an integrated experience for grocery shopping and cooking, the refrigerators will connect to other smart appliances, such as stoves and microwave ovens, on a LAN.

You plan to build an object recognition model by using the Microsoft Cognitive Toolkit. The object recognition model will receive input from the connected devices and send results to applications.

The training data will be derived from more than 10 TB of images. You will convert the raw images to the sparse format.

End of repeated scenario.

You need to ensure that a web service endpoint can receive image data and use an object recognition model to return the expected object and the confidence level of the model. The solution must minimize the effort required to generate the client code to access the web service.

Which resource should you use?

A. the edX Data Science Learning Dashboard
B. Azure Machine Learning Studio
C. Cortana Intelligence Gallery
D. the Data Science Virtual Machine

Answer: B

NEW QUESTION 7

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 sections, you will NOT be able to return to it. As a result, these questions will not appear in the review screen.

You are designing an Azure Machine Learning workflow.

You have a dataset that contains two million large digital photographs. You plan to detect the presence of trees in the photographs.

You need to ensure that your model supports the following:

Solution: You create a Machine Learning experiment that implements the Multiclass Decision Jungle module. Does this meet the goal?

A. Yes
B. No

Answer: B

NEW QUESTION 8

The manager of a call center reports that staffing the center is difficult because the number of calls is unpredictable. You have historical data that contains information about the calls.

You need to build an Azure Machine Learning experiment to predict the number of total calls each hour. Which model should you use?

A. Multiclass Logistic Regression
B. Boosted Decision Tree Regression
C. Decision Forest Regression
D. Poisson Regression

Answer: D

NEW QUESTION 9

Note: This question is part of a series of questions that use the same scenario. For your convenience, the scenario is repeated in each question. Each question presents a different goal and answer choices, but the text of the scenario is exactly the same in each question in this series.

You plan to create a predictive analytics solution for credit risk assessment and fraud prediction in Azure Machine Learning. The Machine Learning workspace for the solution will be shared with other users in your organization. You will add assets to projects and conduct experiments in the workspace.

The experiments will be used for training models that will be published to provide scoring from web services. The experiment for fraud prediction will use Machine Learning modules and APIs to train the models and will predict probabilities in an Apache Hadoop ecosystem.

You plan to configure the resources for part of a workflow that will be used to preprocess data from files stored in Azure Blob storage. You plan to use Python to preprocess and store the data in Hadoop.

You need to get the data into Hadoop as quickly as possible.

Which three actions should you perform? Each correct answer presents part of the solution. NOTE: Each correct selection is worth one point.

A. Create an Azure virtual machine (VM), and then configure MapReduce on the VM.
B. Create an Azure HDInsight Hadoop cluster.
C. Create an Azure virtual machine (VM), and then install an IPython Notebook server.
D. Process the files by using Python to store the data to a Hadoop instance.
E. Create the Machine learning experiment, and then add an Execute Python Script module.

Answer: BDE

NEW QUESTION 10
You plan to use Azure Machine Learning to develop a predictive model. You plan to include an Execute Python Script module.
What capability does the module provide?

A. importing Python modules from a ZIP file for execution in a Machine Learning experiment
B. performing interactive debugging of a Python script
C. saving the results of a Python script run in a Machine Learning environment to a local file
D. returning multiple data frames

Answer: A

NEW QUESTION 10
Note: This question is part of a series of questions that use the same scenario. For your convenience, the scenario is repeated in each question. Each question presents a different goal and answer choices, but the text of the scenario is exactly the same in each question in this series.
A travel agency named Margie's Travel sells airline tickets to customers in the United States.
Margie's Travel wants you to provide insights and predictions on flight delays. The agency is considering implementing a system that will communicate to its customers as the flight departure nears about possible delays due to weather conditions. The flight data contains the following attributes:
The weather data contains the following attributes: AirportID, ReadingDate (YYYY/MM/DD HH), SkyConditionVisibility, WeatherType, WindSpeed, StationPressure, PressureChange, and HourlyPrecip.
You need to use historical data about on-time flight performance and the weather data to predict whether the departure of a scheduled flight will be delayed by more than 30 minutes.
Which method should you use?

A. clustering
B. linear regression
C. classification
D. anomaly detection

Answer: C

Explanation: References:
https://gallery.cortanaintelligence.com/Experiment/Binary-Classification-Flight-delay-prediction-3

NEW QUESTION 15
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 sections, you will NOT be able to return to it. As a result, these questions will not appear in the review screen.
You are working on an Azure Machine Learning experiment. You have the dataset configured as shown in the following table.

Model	Mean absolute error (MAE)
Boosted decision tree	.2
Relative absolute error (RAE)	.43

You need to ensure that you can compare the performance of the models and add annotations to the results. Solution: You connect the Score Model modules from each trained model as inputs for the Evaluate Model
module, and use the Execute R Script module.
Does this meet the goal?

A. Yes
B. No

Answer: A

Explanation: References:
https://msdn.microsoft.com/en-us/library/azure/dn905915.aspx

NEW QUESTION 18
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 sections, you will NOT be able to return to it. As a result, these questions will not appear in the review screen.
You are designing an Azure Machine Learning workflow.
You have a dataset that contains two million large digital photographs. You plan to detect the presence of trees in the photographs.
You need to ensure that your model supports the following: Solution: You create an endpoint to the Computer vision API. Does this meet the goal?

A. Yes
B. No

Answer: B

NEW QUESTION 22
Note: This question is part of a series of questions that use the same scenario. For your convenience, the scenario is repeated in each question. Each question presents a different goal and answer choices, but the text of the scenario is exactly the same in each question in this series.

Start of repeated scenario

You plan to use Azure platform tools to detect and analyze food items in smart refrigerators. To provide families with an integrated experience for grocery shopping and cooking, the refrigerators will connect to other smart appliances, such as stoves and microwave ovens, on a LAN.

You plan to build an object recognition model by using the Microsoft Cognitive Toolkit. The object recognition model will receive input from the connected devices and send results to applications.
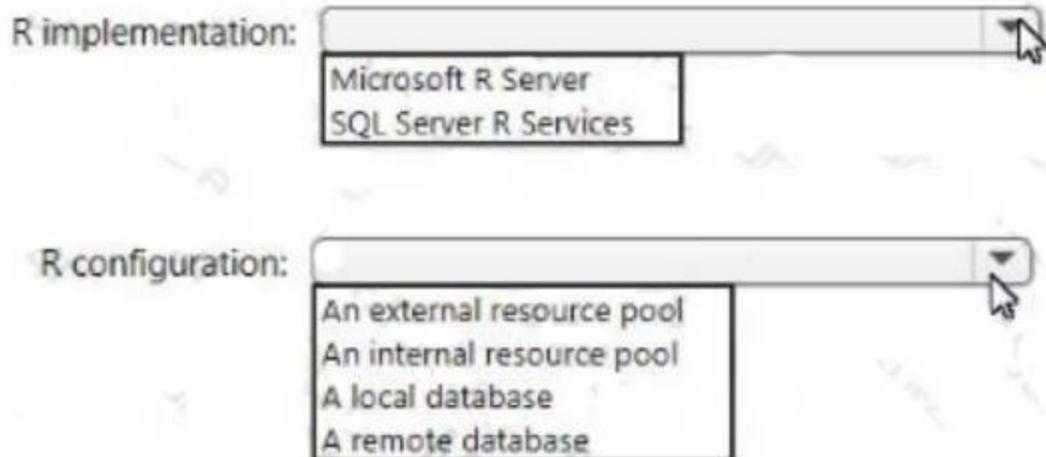
The training data will be derived from more than 10 TB of images. You will convert the raw images to the sparse format.

End of repeated scenario.

The image files to train the object recognition model are stored in a Microsoft SQL Server 2016 Standard edition database on an Azure virtual machine (VM).
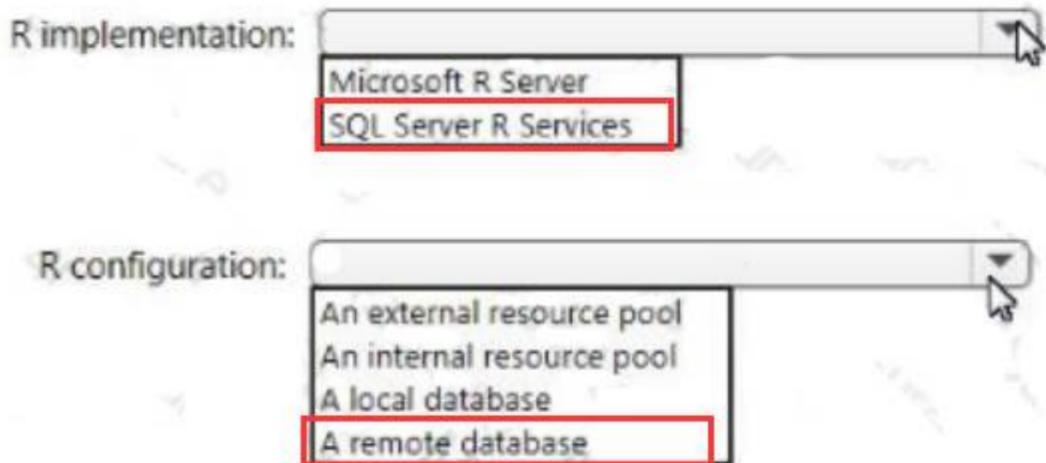
You need to support R packages that can use full parallel threading and processing for RevoScaleR.

How should you implement R? To answer, select the appropriate options in the answer area. NOTE: Each correct selection is worth one point.

R implementation:
- Microsoft R Server
- SQL Server R Services

R configuration:
- An external resource pool
- An internal resource pool
- A local database
- A remote database

Answer:

Explanation:

R implementation:
- Microsoft R Server
- **SQL Server R Services**

R configuration:
- An external resource pool
- An internal resource pool
- A local database
- **A remote database**

NEW QUESTION 25

You are building a classification experiment in Azure Machine Learning.

You need to ensure that you can use the Evaluate Model module the experiment.

Which two actions should you perform? Each correct answer presents part of the solution. NOTE: Each correct selection is worth one point.

A. Connect the input of the Score Model modules to the output of the Evaluate Model module.
B. Connect the input of the Score Model modules to the output of the Train Model modules and the output Split Data modules.
C. Connect the output of the Score Model modules to the input of the Evaluate Model module.
D. Connect the output of the Score Model modules to the input of the Train Model modules and the input of the Split Data modules.

Answer: AB

NEW QUESTION 28

Note: This question is part of a series of questions that use the same scenario. For your convenience, the scenario is repeated in each question. Each question presents a different goal and answer choices, but the text of the scenario is exactly the same in each question in this scries.

Start of repeated scenario

You plan to create a predictive analytics solution for credit risk assessment and fraud prediction in Azure Machine Learning. The Machine Learning workspace for the solution will be shared with other users in your organization. You will add assets to projects and conduct experiments in the workspace.

The experiments will be used for training models that will be published to provide scoring from web services. The experiment for fraud prediction will use Machine Learning modules and APIs to train the models and will predict probabilities in an Apache Hadoop ecosystem.

End of repeated scenario.

You plan to share the Machine Learning workspace with the other users.

You are evaluating whether to assign the User role or the Owner role to several of the users.

Which three actions can be performed by the users who are assigned the User role? Each correct answer presents a complete solution.

NOTE: Each correct selection is worth one point.

A. Create, open, modify, and delete datasets.
B. Create, open, modify, and delete experiments.
C. Invite users to the workspace.
D. Delete users from the workspace.
E. Create, open, modify, and delete web services.
F. Access notebooks.

Answer: CDF

NEW QUESTION 29
Note: This question is part of a series of questions that use the same or similar answer choices. An answer choice may be correct for more than one question in the series. Each question is independent of the other questions in this series. Information and details provided in a question apply only to that question.
You have a dataset that contains a column named Column1. Column1 is empty. You need to omit Column1 from the dataset. The solution must use a native module. Which module should you use?

A. Execute Python Script
B. Tune Model Hyperparameters
C. Normalize Data
D. Select Columns in Dataset
E. Import Data
F. Edit Metadata
G. Clip Values
H. Clean Missing Data

Answer: D

Explanation: References:
https://msdn.microsoft.com/en-us/library/azure/dn905883.aspx

NEW QUESTION 34
Note: This question is part of a series of questions that use the same scenario. For your convenience, the scenario is repeated in each question. Each question presents a different goal and answer choices, but the text of the scenario is exactly the same in each question in this series.
A travel agency named Margie's Travel sells airline tickets to customers in the United States.
Margie's Travel wants you to provide insights and predictions on flight delays. The agency is considering implementing a system that will communicate to its customers as the flight departure nears about possible delays due to weather conditions. The flight data contains the following attributes:
The weather data contains the following attributes: AirportID, ReadingDate (YYYY/MM/DD HH), SkyConditionVisibility, WeatherType, WindSpeed, StationPressure, PressureChange, and HourlyPrecip.
You have an untrained Azure Machine Learning model that you plan to train to predict flight delays.
You need to assess the variability of the dataset and the reliability of the predictions from the model. Which module should you use?

A. Cross-Validate Model
B. Evaluate Model
C. Tune Model Hyperparameters
D. Train Model
E. Score Model

Answer: A

Explanation: References:
https://msdn.microsoft.com/en-us/library/azure/dn905852.aspx

NEW QUESTION 36
You are building an Azure Machine Learning experiment.
You need to transform a string column that has 47 distinct values into a binary indicator column. The solution must use the One-vs-All Multiclass model.
Which module should you use?

A. Select Column Transform
B. Convert to Indicator Values
C. Group Categorical Values
D. Edit Metadata

Answer: B

NEW QUESTION 40
You need to integrate code and formatted text into an Azure Machine Learning experiment that enables interactive execution.
What should you use?

A. a Jupyter notebook
B. Azure Stream Analytics
C. an Execute Python Script module
D. an Execute R Script module

Answer: A

NEW QUESTION 43
You have an Apache Spark cluster in Azure HDinsight. The cluster includes 200 TB in five Apache Hive tables that have multiple foreign key relationships.
You have an Azure Machine Learning model that was built by using SPARK Accelerated Failure Time (AFT) Survival Regression Model (spark-survreg).
You need to prepare the Hive data into a single table as input for the Machine Learning model. The Hive data must be prepared in the least amount of time possible.
What should you use to prepare the data?

A. a Hive user-defined function (UDF)
B. Spark SQL

C. the GPU
D. Java Mapreduce jobs

Answer: A

NEW QUESTION 46
Note: This question is part of a series of questions that use the same scenario. For your convenience, the scenario is repeated in each question. Each question presents a different goal and answer choices, but the text of the scenario is exactly the same in each question in this series.
Start of repeated scenario
You plan to create a predictive analytics solution for credit risk assessment and fraud prediction in Azure Machine Learning. The Machine Learning workspace for the solution will be shared with other users in your organization. You will add assets to projects and conduct experiments in the workspace.
The experiments will be used for training models that will be published to provide scoring from web services. The experiment for fraud prediction will use Machine Learning modules and APIs to train the models and will predict probabilities in an Apache Hadoop ecosystem.
End of repeated scenario.
The users will use different data sources that follow a standard format. The users will receive results in a standard format by using the fraud prediction web service. The results will be saved to a location specified by the users.
You need to provide the users with the ability to get results for different risk tolerances without affecting the calculation of the model. Which three modules should be configured to use the Web Service Parameters? Each correct answer presents part of the solution.
NOTE: Each correct selection is worth one point.

A. Evaluate Model
B. Import Data
C. Select Columns in Dataset
D. Export Data
E. Time Series Anomaly Detection

Answer: ABD

NEW QUESTION 47
You have an Execute R Script module that has one input from either a Partition and Sample module or a Web service input module.
You need to preprocess tweets by using R. The solution must meet the following requirements:
How should you complete the R code? To answer, drag the appropriate values to the correct targets. Each value may be used once, more than once, or not at all. You may need to drag the split bar panes or scroll to view content.
NOTE: Each correct selection is worth one point.

Values

```
dataset[[1]]
```

```
dataset[[2]]
```

```
gsub("[^a-z]", " ", tweet_text, ignore.case = FALSE)
```

```
maml.mapInputPort(1)
```

```
sapply(tweet_text, tolower)
```

● ● ● ●

Answer area

```
dataset <-           Value

tweet_text <-        Value

tweet_text <-        Value

tweet_text <-        Value

data.set <- as.data.frame(tweet_text, stringsAsFactors=FALSE)

maml.mapOutputPort("data.set")
```


Answer:

Explanation:

Values

```
dataset[[1]]
```

```
dataset[[2]]
```

```
gsub("[^a-z]", " ", tweet_text, ignore.case = FALSE)
```

```
maml.mapInputPort(1)
```

```
sapply(tweet_text, tolower)
```

• • • •

Answer area

```
dataset <-       maml.mapInputPort(1)

tweet_text <-  dataset[[1]]

tweet_text <-  sapply(tweet_text, tolower)

tweet_text <-  gsub("[^a-z]", " ", tweet_text, ignore.case = FALSE)

data.set <- as.data.frame(tweet_text, stringsAsFactors=FALSE)

maml.mapOutputPort("data.set")
```

NEW QUESTION 49

You work for a company that has retail department stores.

You are developing an Azure Machine Learning experiment to predict seasonal sales. You need to address a model overfitting issue by using the following two solutions:

• Solution 1: Controls the penalty for complexity, which, when successful, prevents overfitting

• Solution 2: Separates model selection from testing, causing a more conservative estimate of generalization Which method should you use for each solution? To answer, drag the appropriate methods to the correct

solutions. Each method 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.

Methods				Answer Area
Cross-validation	Generalization	Normalization	Regularization	Solution 1: Method
				Solution 2: Method

Answer:

Explanation:

Methods				Answer Area
Cross-validation	Generalization	Normalization	Regularization	Solution 1: Cross-validation
				Solution 2: Cross-validation

NEW QUESTION 54

You are building an Azure Machine Learning solution for an online retailer.

When a customer selects a product, you need to recommend products that the customer might like to purchase at the same time. The recommendation should be

based on what other customers purchased when they purchased the same product.
Which model should you use?

A. Collaborative filtering
B. Boosted Decision Tree Regression model
C. Two-Class boosted decision tree
D. K-Means Clustering

Answer: A

NEW QUESTION 58

You have an Azure Machine Learning environment. You are evaluating whether to use R code or Python.
Which three actions can you perform by using both R code and Python in the Machine Learning environment? Each correct answer presents a complete solution.
NOTE: Each correct selection is worth one point.

A. Preprocess, cleanse, and group data.
B. Score a training model.
C. Create visualizations.
D. Create an untrained model that can be used with the Train Model module.
E. Implement feature ranking.

Answer: ABC

NEW QUESTION 60

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 sections, you will NOT be able to return to it. As a result, these questions will not appear in the review screen.
You are designing an Azure Machine Learning workflow.
You have a dataset that contains two million large digital photographs. You plan to detect the presence of trees in the photographs.
You need to ensure that your model supports the following:
Solution: You create a Machine Learning experiment that implements the Multiclass Neural Network module. Does this meet the goal?

A. Yes
B. No

Answer: A

NEW QUESTION 64

Note: This question is part of a series of questions that use the same scenario. For your convenience, the scenario is repeated in each question. Each question presents a different goal and answer choices, but the text of the scenario is exactly the same in each question in this series.
A travel agency named Margie's Travel sells airline tickets to customers in the United States.
Margie's Travel wants you to provide insights and predictions on flight delays. The agency is considering implementing a system that will communicate to its customers as the flight departure nears about possible delays due to weather conditions. The flight data contains the following attributes:
The weather data contains the following attributes: AirportID, ReadingDate (YYYY/MM/DD HH), SkyConditionVisibility, WeatherType, WindSpeed, StationPressure, PressureChange, and HourlyPrecip.
You plan to predict flight delays that are 30 minutes or more.
You need to build a training model that accurately fits the data. The solution must minimize over fitting and minimize data leakage.
Which attribute should you remove?

A. OriginAirportID
B. DepDel
C. DepDel30
D. Carrier
E. DestAirportID

Answer: C

NEW QUESTION 66

Note: This question is part of a series of questions that use the same scenario. For your convenience, the scenario is repeated in each question. Each question presents a different goal and answer choices, but the text of the scenario is exactly the same in each question in this series.
You plan to create a predictive analytics solution for credit risk assessment and fraud prediction in Azure Machine Learning. The Machine Learning workspace for the solution will be shared with other users in your organization. You will add assets to projects and conduct experiments in the workspace.
The experiments will be used for training models that will be published to provide scoring from web services. The experiment for fraud prediction will use Machine Learning modules and APIs to train the models and will predict probabilities in an Apache Hadoop ecosystem.
You finish training the model and are ready to publish a predictive web service that will provide the users with the ability to specify the data source and the save location of the results. The model includes a Split Data module.
Which two actions should you perform to convert the Machine Learning experiment to a predictive web service? To answer, drag the appropriate actions to the correct targets. Each action 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.

Actions

Click Set Up Web Service for the training experiment.
Configure a web service endpoint for input and output, and then specify the parameters.
Remove the Split Data module.
Replace the Machine Learning algorithm and the train model by using a saved training model.
Save the trained model.

Answer Area

First action:

Action

Second action:

Action

Answer:

Explanation: References:
https://docs.microsoft.com/en-us/azure/machine-learning/studio/convert-training-experiment-to-scoring-experim

NEW QUESTION 69
You have the following HiveQL query in an Import Data module.

```
    from Student_Tabl
  ) a
where state_rank <= state_cnt*'${hiveconf:sampleRate}'


(
select
    field1, field2, field3, ..., fieldN, state,
    count(*) over (partition by state) as state_cnt,
    rank() over (partition by state order by rand()) as state_rank

    from Student_Tabl
  ) a
where state_rank <= state_cnt*'${hiveconf:sampleRate}'
```

Which type of operation is being performed?

A. sampling a bucketized table
B. random sampling by groups
C. uniform random sampling
D. stratified sampling

Answer: D

NEW QUESTION 71
......

Thank You for Trying Our Product

We offer two products:

1st - We have Practice Tests Software with Actual Exam Questions

2nd - Questons and Answers in PDF Format

70-774 Practice Exam Features:

* 70-774 Questions and Answers Updated Frequently

* 70-774 Practice Questions Verified by Expert Senior Certified Staff

* 70-774 Most Realistic Questions that Guarantee you a Pass on Your FirstTry

* 70-774 Practice Test Questions in Multiple Choice Formats and Updatesfor 1 Year

100% Actual & Verified — Instant Download, Please Click
Order The 70-774 Practice Test Here