70-774.exam.20q

Number: 70-774
Passing Score: 0
Time Limit: 120 min

Website: https://vceplus.com
VCE to PDF Converter: https://vceplus.com/vce-to-pdf/
Facebook: https://www.facebook.com/VCE.For.All.VN/
Twitter : https://twitter.com/VCE_Plus

https://vceplus.com/

70-774

Perform Cloud Data Science with Azure Machine Learning

Exam A

QUESTION 1
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 consolidate the output of the Score Model modules by using the Add Rows module, and then use the Execute R Script module.

Does this meet the goal?

A. Yes
B. No

Correct Answer: B
Section: (none)
Explanation

Explanation/Reference:
Explanation:

References: https://msdn.microsoft.com/en-
us/library/azure/dn905915.aspx **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 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 then save the results as a dataset.

Does this meet the goal?

A. Yes
B. No

Correct Answer: A
Section: (none)
Explanation

Explanation/Reference:
Explanation:

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

QUESTION 3
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

Correct Answer: B
Section: (none)
Explanation

Explanation/Reference:
Explanation:

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

QUESTION 4
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 save the output of the Score Model modules as a combined set, and then use the Project Columns module to select the MAE.

Does this meet the goal?

A. Yes
B. No

Correct Answer: B
Section: (none)
Explanation

Explanation/Reference:
Explanation:

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

QUESTION 5
You have data about the following:
▪ Users
▪ Movies
▪ User ratings of the movies

You need to predict whether a user will like a particular movie.

Which Matchbox recommender should you use?

A. Item Recommendation
B. Related Items
C. Rating Prediction
D. Related Users

Correct Answer: C
Section: (none)
Explanation

Explanation/Reference:
Explanation:

References: https://msdn.microsoft.com/en-
us/library/azure/dn905970.aspx#RatingPredictionOptions

QUESTION 6
You have the following three training datasets for a restaurant:
▪ User features
▪ Item features
▪ Ratings of items by users

You must recommend restaurant to a particular user based only on the users features.

You need to use a Matchbox Recommender to make recommendations.

How many input parameters should you specify?

A. 1
B. 2
C. 3
D. 4

Correct Answer: B

Section: (none)
Explanation

Explanation/Reference:
Explanation:

References: https://msdn.microsoft.com/en-
us/library/azure/dn905987.aspx

QUESTION 7
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: ▪ DepartureDate: The departure date aggregated at a per hour granularity
▪ Carrier: The code assigned by the IATA and commonly used to identify a carrier
▪ OriginAirportID: An identification number assigned by the USDOT to identify a unique airport (the flight's origin)
▪ DestAirportID: An identification number assigned by the USDOT to identify a unique airport (the flight's destination)
▪ DepDel: The departure delay in minutes
▪ DepDel30: A Boolean value indicating whether the departure was delayed by 30 minutes or more (a value of 1 indicates that the departure was delayed by 30 minutes or more)

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

Correct Answer: C

Section: (none)
Explanation

Explanation/Reference:
Explanation:

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

QUESTION 8
DRAG DROP

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: ▪ DepartureDate: The departure date aggregated at a per hour granularity
▪ Carrier: The code assigned by the IATA and commonly used to identify a carrier
▪ OriginAitportID: An identification number assigned by the USDOT to identify a unique airport (the flight's origin)
▪ DestAirportID: An identification number assigned by the USDOT to identify a unique airport (the flight's destination)
▪ DepDel: The departure delay in minutes
▪ DepDel30: A Boolean value indicating whether the departure was delayed by 30 minutes or more (a value of 1 indicates that the departure was delayed by 30 minutes or more)

The weather data contains the following attributes: AirportID, ReadingDate (YYYY/MM/DD HH), SkyConditionVisibility, WeatherType, WindSpeed, StationPressure, PressureChange, and HourlyPrecip.

You need to remove the bias and to identify the columns in the input dataset that have the greatest predictive power.

Which module should you use for each requirement? To answer, drag the appropriate modules to the correct requirements. Each module 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:

Modules

Cross-validate Model
Evaluate Model
Filter and Sample
Filter Based Feature Selection Module
Parameter Sweep
Tune Model Hyperparameters

Answer Area

Remove bias:	Module

Identify the columns that have the greatest predictive power:	Module

Correct Answer:

Modules

Evaluate Model
Filter and Sample
Filter Based Feature Selection Module
Parameter Sweep

Answer Area

Remove bias:	Cross-validate Model

Identify the columns that have the greatest predictive power:	Tune Model Hyperparameters

Section: (none)
Explanation

Explanation/Reference:
References:
https://gallery.cortanaintelligence.com/Experiment/Binary-Classification-Flight-delay-prediction-3 https://msdn.microsoft.com/library/azure/038d91b6-c2f2-42a1-9215-1f2c20ed1b40

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.

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: ▪ DepartureDate: The departure date aggregated at a per hour granularity
▪ Carrier: The code assigned by the IATA and commonly used to identify a carrier
▪ OriginAitportID: An identification number assigned by the USDOT to identify a unique airport (the flight's origin)
▪ DestAirportID: An identification number assigned by the USDOT to identify a unique airport (the flight's destination)
▪ DepDel: The departure delay in minutes
▪ DepDel30: A Boolean value indicating whether the departure was delayed by 30 minutes or more (a value of 1 indicates that the departure was delayed by 30 minutes or more)

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

Correct Answer: A
Section: (none)
Explanation
Explanation/Reference:
Explanation:

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

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: ▪ DepartureDate: The departure date aggregated at a per hour granularity
▪ Carrier: The code assigned by the IATA and commonly used to identify a carrier
▪ OriginAitportID: An identification number assigned by the USDOT to identify a unique airport (the flight's origin)
▪ DestAirportID: An identification number assigned by the USDOT to identify a unique airport (the flight's destination)
▪ DepDel: The departure delay in minutes
▪ DepDel30: A Boolean value indicating whether the departure was delayed by 30 minutes or more (a value of 1 indicates that the departure was delayed by 30 minutes or more)

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

Correct Answer: C
Section: (none)
Explanation
Explanation/Reference:
Explanation:

QUESTION 11
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 need to remove rows that have an empty value in a specific column. 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

Correct Answer: H
Section: (none)
Explanation

Explanation/Reference:
Explanation:

References: https://blogs.msdn.microsoft.com/azuredev/2017/05/27/data-cleansing-tools-in-azure-machine-learning/

QUESTION 12
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

Correct Answer: E
Section: (none)
Explanation

Explanation/Reference:
Explanation:

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

QUESTION 13
From the Cortana Intelligence Gallery, you deploy a solution.

You need to modify the solution.

What should you use?

A. Azure Stream Analytics
B. Microsoft Power BI Desktop
C. Azure Machine Learning Studio
D. R Tools for Visual Studio

Correct Answer: C
Section: (none)
Explanation

Explanation/Reference:
Explanation:

References: https://docs.microsoft.com/en-us/azure/machine-learning/studio/gallery-experiments

QUESTION 14
You are building an Azure Machine Learning workflow by using Azure Machine Learning Studio.

You create an Azure notebook that supports the Microsoft Cognitive Toolkit.

You need to ensure that the stochastic gradient descent (SGD) configuration maximizes the samples per second and supports parallel modeling that is managed by a parameter server.

Which SGD algorithm should you use?

A. DataParallelASGD
B. DataParallelSGD
C. ModelAveragingSGD
D. BlockMomentumSGD

Correct Answer: B
Section: (none)
Explanation

Explanation/Reference:
Explanation:

QUESTION 15
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

Correct Answer: C
Section: (none)
Explanation

Explanation/Reference:
Explanation:

References: https://azure.microsoft.com/en-gb/blog/general-availability-of-hdinsight-interactive-query-blazing-fast-data-warehouse-style-queries-on-hyper-scale-data-2/

QUESTION 16
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.

Correct Answer: BDE
Section: (none)
Explanation

Explanation/Reference:
Explanation:

QUESTION 17
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 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

Correct Answer: D
Section: (none)
Explanation

Explanation/Reference:
Explanation:

QUESTION 18
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

Correct Answer: A
Section: (none)
Explanation

Explanation/Reference:
Explanation:

QUESTION 19
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

Correct Answer: A

Section: (none)
Explanation

Explanation/Reference:
Explanation:

QUESTION 20
You have a dataset that is missing values in a column named Column3. Column3 is correlated to two columns named Column4 and Column5.

You need to improve the accuracy of the dataset, while minimizing data loss.

What should you do?

A. Replace the missing values in Column3 by using probabilistic Principal Component Analysis (PCA).
B. Remove all of the rows that have the missing values in Column4 and Column5.
C. Replace the missing values in Column3 with a mean value.
D. Remove the rows that have the missing values in Column3.

Correct Answer: A
Section: (none)
Explanation

Explanation/Reference:
Explanation: