

# Microsoft

## Exam 70-774

### Perform Cloud Data Science with Azure Machine Learning

Version: Demo

[ Total Questions: 10 ]

**Question No : 1**

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: B,D,E**

**Question No : 2**

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**

**Question No : 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 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

**Answer: B**

**Explanation:**

References:

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

**Question No : 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 connect the Score Model modules from each trained model as inputs for the Evaluate Model module, and use the Execute R Script module.

Doe

s this meet the goal?

- A. Yes
- B. No

**Answer: B**

**Explanation:**

References:

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

**Question No : 5**

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

**Answer: H**

**Explanation:**

References:

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

**Question No : 6**

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

**Answer: B**

**Explanation:**

References:

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

You have an Azure Machine Learning experiment.

You discover that a model causes many errors in a production dataset. The model causes only few errors in the training data.

What is the cause of the errors?

- A. overfitting
- B. generalization
- C. underfitting
- D. a simple predictor

**Answer: A**

**Question No : 8**

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.

**Answer: A**

**Question No : 9**

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**

**Question No : 10 DRAG DROP**

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:**

## 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")
```