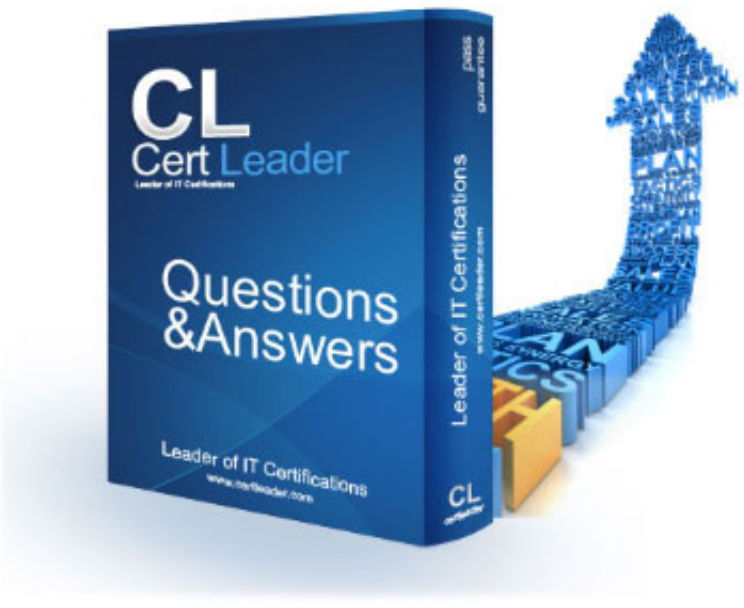
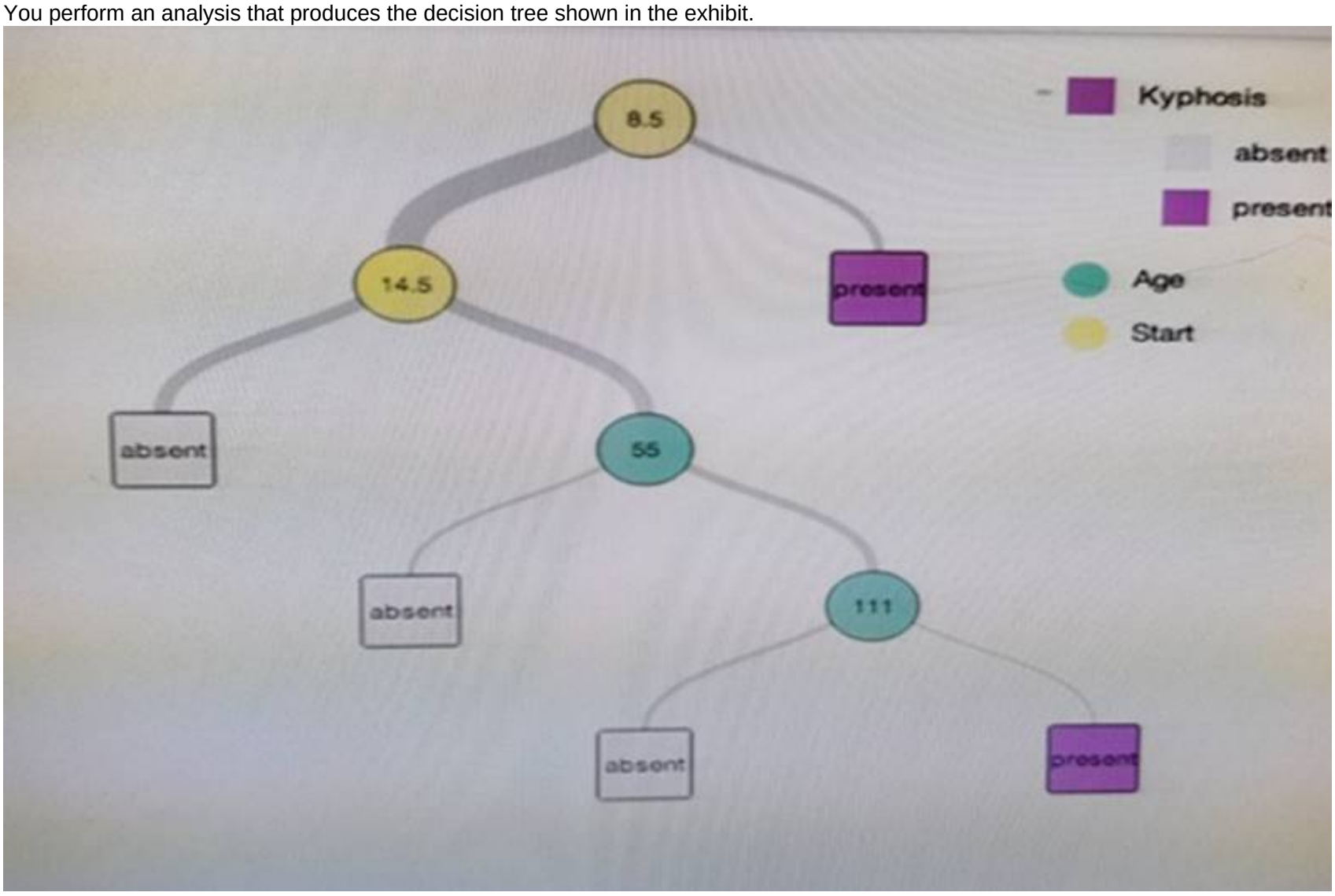
70-773 Dumps

Analyzing Big Data with Microsoft R (beta)

https://www.certleader.com/70-773-dumps.html



NEW QUESTION 1
You perform an analysis that produces the decision tree shown in the exhibit.



How many leaf nodes are there on the tree?

A. 2
B. 3
C. 5
D. 7

Answer: B

NEW QUESTION 2
Note: This question Is part of a series of questions that use the same or similar answer choice. 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 generate a residual based on two columns. The solution must build a trend indicator.
Which function should you use?

A. rxPredict
B. rxLogit
C. Summary
D. rxLinMod
E. rxTweedie
F. stepAic
G. rxTransform
H. rxDataStep

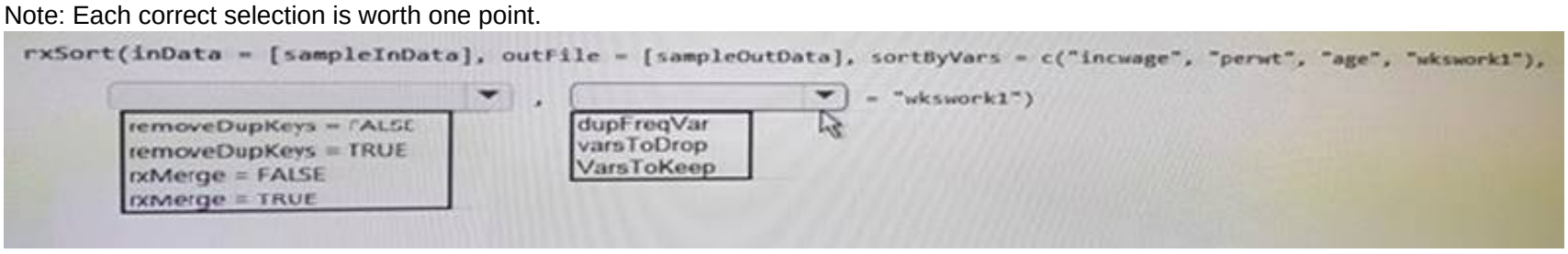
Answer: C

NEW QUESTION 3
HOTSPOT
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 are developing a Microsoft R Open solution that will leverage the computing power of the database server for some of your datasets.
You are performing feature engineering and data preparation for the datasets. The following is a sample of the dataset.

```
rxGetInfo(df)
head(df)
```

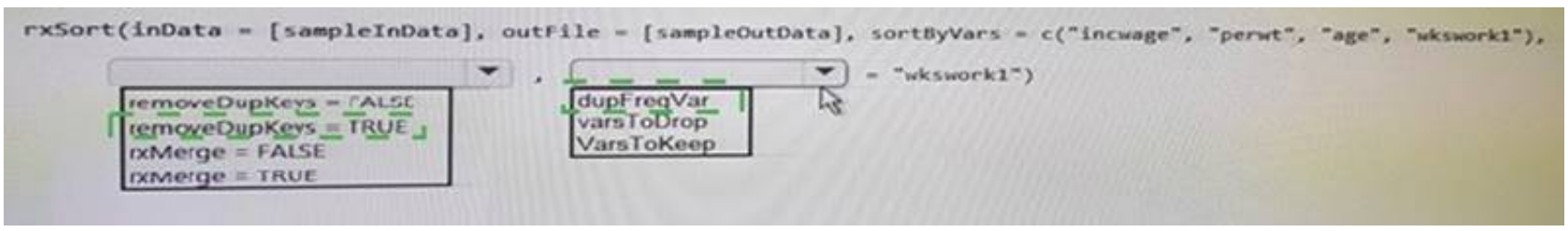
	age	incwage	perwt	wkswork1	state
1	50	9000	30	48	Indiana
2	41	35000	20	48	Indiana
3	55	40400	21	49	Indiana
4	56	45000	30	52	Indiana
5	46	17200	60	51	Indiana
6	49	35000	21	52	Indiana

End of repeated scenario
You need to sort the data from the dataset sample and to remove duplicates by using wkswork1.
Which R code segment should you use? to answer, select the appropriate options in the answer area.
Note: Each correct selection is worth one point.



Answer:

Explanation:



NEW QUESTION 4
Note: This Question is part of a series of Questions that use the same or similar answer choices. An answer choice may be correct 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 data source that is larger than memory.
You need to visualize the distribution of the values for a variable in the data source. What should you use?

A. the Describe package
B. the rxHistogram function
C. the rxSummary function
D. the rxQuantile function
E. the rxCube function
F. the summary function
G. the rxCrossTabs function
H. the ggplot2 package

Answer: B

NEW QUESTION 5
You have an Apache Hadoop Hive data warehouse. RevoScaleR is not installed. You need to sort the data according to the variables in the dataset.
What should you do?

A. Connect to the database by using an ODBC connection, and then use the rxSort function.
B. Create a table in the ORC file format.
C. Connect to the database by using an ODBC connection, and then use the rxDataStep function.
D. Execute a Hive query that sorts the data, and then reads the results.

Answer: D

NEW QUESTION 6
You have a dataset that has multiple blocks and only numeric variables. You are computing in a local compute context.
You plan to lag a variable named x to create a new variable named x_lagged by using a transform function. You will create a new element in the output of the function.
You need to minimize the number of missing values.
Which three actions should you perform? Each correct answer presents part of the solution.
NOTE: Each correct selection is worth one point.

A. Assign a value to the first value of x_lagged in the current block.
B. Use rxSet to store the last value of x_lagged in the current block.

C. Use rxSet to store the last value of x in the current block.
D. Use rxGet to retrieve the first value of x in the next block to be processed.
E. Use rxGet to retrieve a value stored in processing of the prior block.

Answer: ACD

NEW QUESTION 7
You need to build a model that looks at the probability of an outcome. You must regulate between L1 and L2.
Which classification method should you use?

A. Two-Class Neural Network
B. Two-Class Support Vector Machine
C. Two-Class Decision Forest
D. Two-Class Logistic Regression

Answer: A

NEW QUESTION 8
Note: This question Is part of a series of questions that use the same or similar answer choice. 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 evaluate the significance of coefficient that are produced by using a model that was estimated already.
Which function should you use?

A. rxPredict
B. rxLogit
C. Summary
D. rxLinMod
E. rxTweedie
F. stepAic
G. rxTransform
H. rxDataStep

Answer: D

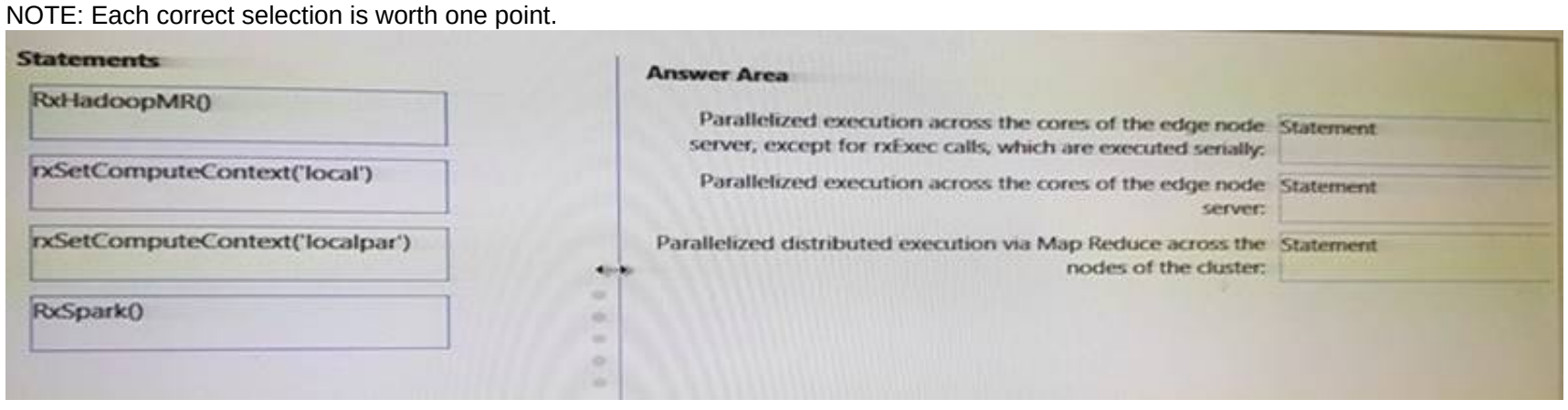
Explanation: https://docs.microsoft.com/en-us/r-server/r/how-to-revoscaler-linear-model

NEW QUESTION 9
You have cloud and on-premises resources that include Microsoft SQL Server and a big data environment in Apache Hadoop.
You have 50 billion fact records.
You need to build time series models to execute forecasting reports on the fact records. What should you use?

A. RxSpark on the Hadoop cluster
B. RxHadoopMR on the Hadoop cluster
C. RxLocalseq on the SQL Server database
D. RxLocalParallel on the SQL Server database

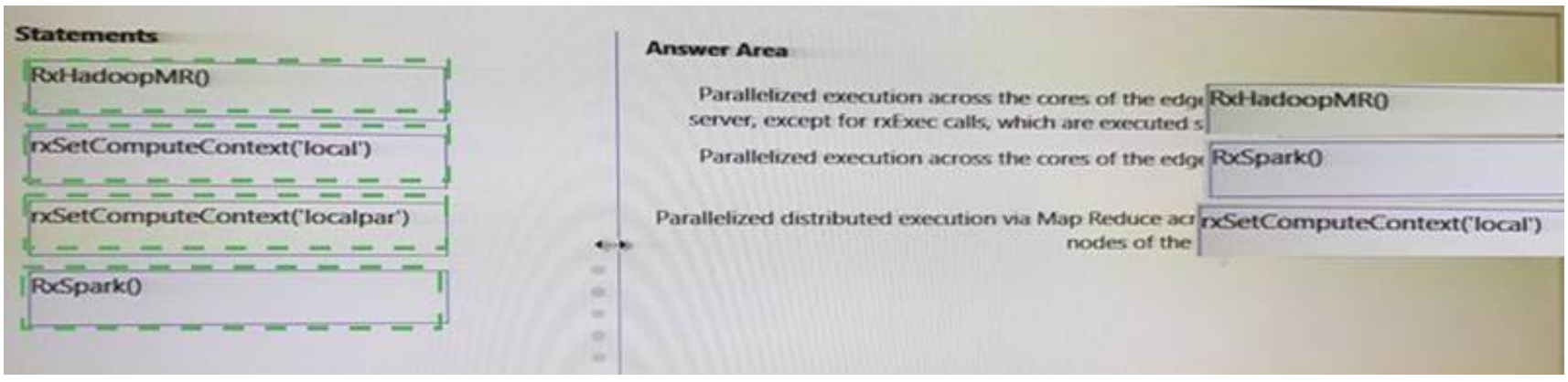
Answer: A

NEW QUESTION 10
DRAG DROP
You need to set the compute context for three different target environments.
Which Statement should you use for each environment? To answer, drag the appropriate statements to the correct execution contexts. Each statement 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.



Answer:

Explanation:



NEW QUESTION 10
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, whale others might not have a correct solution-After you answer a question in this section, you will NOT be able to return to it- As a result, these questions will not appear in the review screen.
You use dplyrXdf and you discover that after you exit the session, the output files that were created were deleted. You need to prevent the files from being deleted.
Solution: You remove all instances of the file.remove method.
Does this meet the goal?

A. Yes
B. No

Answer: B

NEW QUESTION 14
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, whale others might not have a correct solution-After you answer a question in this section, you will NOT be able to return to it- As a result, these questions will not appear in the review screen.
You have a Microsoft SQL Server instance that has R Services (In-Database) installed. You need to monitor the R jobs that are sent to SQL Server.
Solution: You create an events trace configuration file and place the file in the same directory as the BXLServer process.
Does this meet the goal?

A. Yes
B. No

Answer: B

NEW QUESTION 19
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, whale others might not have a correct solution-After you answer a question in this section, you will NOT be able to return to it- As a result, these questions will not appear in the review screen.
You have a Microsoft SQL Server instance that has R Services (In-Database) installed. You need to monitor the R jobs that are sent to SQL Server.
Solution: You call a function from the RevoPemaR Package.
Does this meet the goal?

A. Yes
B. No

Answer: B

NEW QUESTION 21
You need to use the ScaleR distributed processing in an Apache Hadoop environment. Which data source should you use?

A. Microsoft SQL Server database
B. XDF data files
C. ODBC data
D. Teradata database

Answer: B

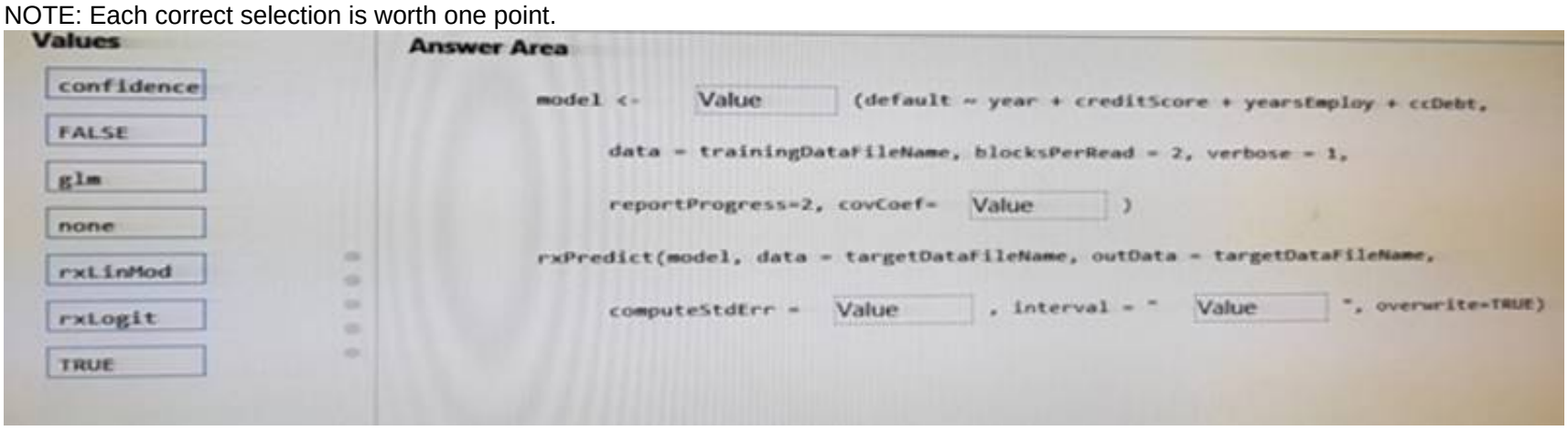
NEW QUESTION 25
You have a dataset that has a character variable. You need to create a bag of counts of n-grams. Which function should you use?

A. featurizeText0
B. categoricalHash0
C. concat0
D. selcctFeatures0
E. categorical0

Answer: A

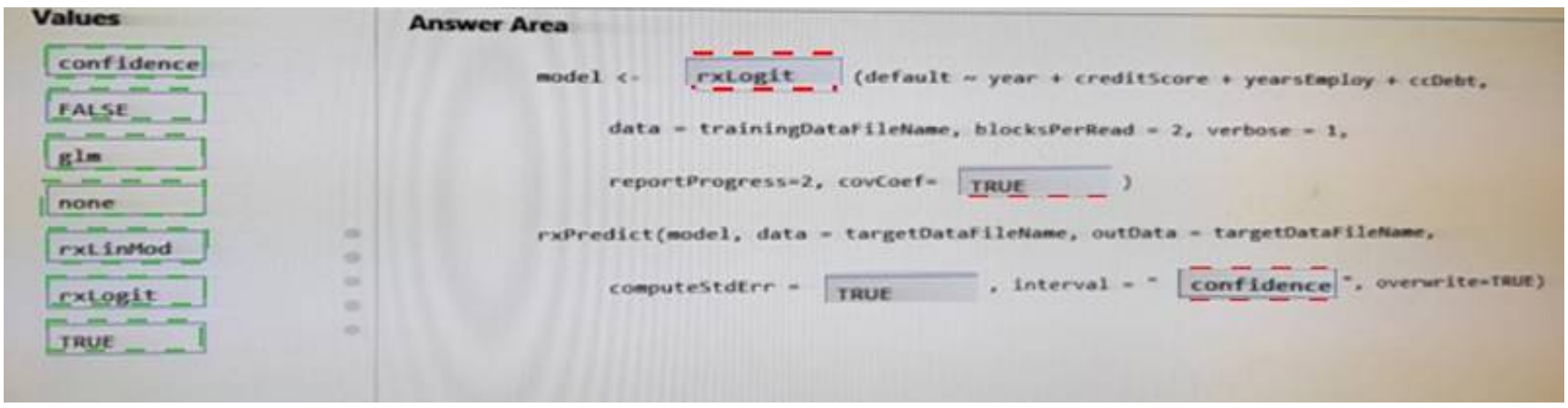
Explanation: featurizeText: Produces a bag of counts of sequences of consecutive words, called n-grams, from a given corpus of text. It offers language detection, tokenization, stopwords removing, text normalization and feature generation.

NEW QUESTION 26
DRAG DROP
You are using rxPredict for a logistic regression model.
You need to obtain prediction standard errors and confidence intervals. Which R code segment should you use?
To answer, drag the .impropriate values to the correct targets. Each value may be used once, more than once, or not .You may need to drag the split bar between panes or scroll to view content.
NOTE: Each correct selection is worth one point.



Answer:

Explanation:



NEW QUESTION 29
You need to run a larger data tree model by using rsDForest. The model must use cross validation.
Which rxDForest option should you use?

A. maxSurrogate
B. maxNumBins
C. maxDepth
D. maxCompete
E. xVal

Answer: E

Explanation: https://docs.microsoft.com/en-us/r-server/r/how-to-revoscaler-decision-tree

NEW QUESTION 34
You have a dataset.
You need to repeatedly split randomly the dataset so that 80 percent of the data is used as a training set and the remaining 20 percent is used as a test set.
Which method should you use?

A. threshold
B. binary classification
C. imputation
D. cross validation
E. pruning

Answer: D

NEW QUESTION 39
You are planning the compute contexts for your environment. You need to execute rx-function calls in parallel.
What are three possible compute contexts that you can use to achieve this goal? Each correct answer presents a complete solution.
NOTE: Each correct selection is worth one point.

A. local parallel
B. Spark
C. local sequential
D. Map Reduce
E. SQL

Answer: ABC

Explanation: https://docs.microsoft.com/en-us/azure/hdinsight/hdinsight-hadoop-r-server-compute-contexts

NEW QUESTION 40
Note: This Question is part of a series of Questions that use the same or similar answer choices. An answer choice may be correct 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 the physical characteristics of people.
You need to visualize a relationship between height and weight for a subset of observations in the dataset.
What should you use?

A. the Describe package
B. the rxHistogram function
C. the rxSummary function
D. the rxQuantile function
E. the rxCube function
F. the summary function
G. the rxCrossTabs function
H. the ggplot2 package

Answer: E

NEW QUESTION 41
Note: This question Is part of a series of questions that use the same or similar answer choice. 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 build a model that uses xyz regression.
You need to estimate a model that predicts a binary variable.
Which function should you use?

A. rxPredict
B. rxLogit
C. Summary
D. rxLinMod
E. rxTweedie
F. stepAic
G. rxTransform
H. rxDataStep

Answer: B

Explanation: https://docs.microsoft.com/en-us/r-server/r/how-to-revoscaler-logistic- regression

NEW QUESTION 45
You plan to read data from an Oracle database table and to store the data in the file system for later processing by dplyrXdf, The size of the data is larger than the memory on the server to used for modelling.
You need to ensure that the data can be processed by dplyrXdf in the least amount of time possible.
How should you transfer the data from the Oracle database?

A. Use the RODBC library, connect to the Oracle database server by using odbcConnec
B. and then use rxDataStep to export the data to a comma-separated values (CSV) file.
C. Define a data source to the Oracle database server by using RxOdbcData, and then use rxlmport to save the data to an XDF file.
D. Use the RODBC library, connect to the Oracle database server by using odbcConnec
E. and then use rxSplit to save the data to multiple comma-separated values (CSV) files.

Answer: C

NEW QUESTION 47
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, whale others might not have a correct solution-After you answer a question in this section, you will NOT be able to return to it- As a result, these questions will not appear in the review screen.
You have a Microsoft SQL Server instance that has R Services (In-Database) installed. You need to monitor the R jobs that are sent to SQL Server.
Solution: You register an Extended Events package.
Does this meet the goal?

A. Yes
B. No

Answer: A

NEW QUESTION 49
.....

Thank You for Trying Our Product

* 100% Pass or Money Back

All our products come with a 90-day Money Back Guarantee.

* One year free update

You can enjoy free update one year. 24x7 online support.

* Trusted by Millions

We currently serve more than 30,000,000 customers.

* Shop Securely

All transactions are protected by VeriSign!

100% Pass Your 70-773 Exam with Our Prep Materials Via below:

https://www.certleader.com/70-773-dumps.html