EMC

E20-007 Exam

Data Science Associate Exam

Thank you for Downloading E20-007 exam PDF Demo

You can Buy Latest E20-007 Full Version Download

https://www.certkillers.net/Exam/E20-007

Version: 8.0

Question: 1

You are using MADlib for Linear Regression analysis. Which value does the statement return?
SELECT (linregr(depvar, indepvar)).r2 FROM zeta1;

A. Goodness of fit
B. Coefficients
C. Standard error
D. P-value

Answer: A

Question: 2

Which data asset is an example of quasi-structured data?

A. Webserver log
B. XML data file
C. Database table
D. News article

Answer: A

Question: 3

What would be considered "Big Data"?

A. An OLAP Cube containing customer demographic information about 100, 000, 000 customers
B. Daily Log files from a web server that receives 100, 000 hits per minute
C. Aggregated statistical data stored in a relational database table
D. Spreadsheets containing monthly sales data for a Global 100 corporation

Answer: B

Question: 4

A data scientist plans to classify the sentiment polarity of 10, 000 product reviews collected from the Internet. What is the most appropriate model to use? Suppose labeled training data is available.

A. Naïve Bayesian classifier
B. Linear regression
C. Logistic regression

D. K-means clustering

Answer: A

Question: 5

In which lifecycle stage are test and training data sets created?

A. Model building
B. Model planning
C. Discovery
D. Data preparation

Answer: A

Question: 6

When creating a presentation for a technical audience, what is the main objective?

A. Show that you met the project goals
B. Show how you met the project goals
C. Show if the model will meet the SLA
D. Show the technique to be used in the production environment

Answer: B

Question: 7

Your company has 3 different sales teams. Each team's sales manager has developed incentive offers to increase the size of each sales transaction. Any sales manager whose incentive program can be shown to increase the size of the average sales transaction will receive a bonus.
Data are available for the number and average sale amount for transactions offering one of the incentives as well as transactions offering no incentive.
The VP of Sales has asked you to determine analytically if any of the incentive programs has resulted in a demonstrable increase in the average sale amount. Which analytical technique would be appropriate in this situation?

A. One-way ANOVA
B. Multi-way ANOVA
C. Student's t-test
D. Wilcoxson Rank Sum Test

Answer: A

Question: 8

In data visualization, what is used to focus the audience on a key part of a chart?

A. Emphasis colors
B. Detailed text
C. Pastel colors
D. A data table

Answer: A

Question: 9

Which word or phrase completes the statement? Data-ink ratio is to data visualization as _____ .

A. Confusion matrix is to classifier
B. Data scientist is to big data
C. Seasonality is to ARIMA
D. K-means is to Naive Bayes

Answer: A

Question: 10

Consider a database with 4 transactions:
Transaction 1: {cheese, bread, milk}
Transaction 2: {soda, bread, milk}
Transaction 3: {cheese, bread}
Transaction 4: {cheese, soda, juice}
You decide to run the association rules algorithm where minimum support is 50%. Which rule has a confidence at least 50%?

A. {cheese} => {bread}
B. {juice} => {cheese}
C. {milk} => {soda}
D. {soda} => {milk}

Answer: A

Question: 11

You are using the Apriori algorithm to determine the likelihood that a person who owns a home has a good credit score. You have determined that the confidence for the rules used in the algorithm is > 75%. You calculate lift = 1.011 for the rule, "People with good credit are homeowners". What can you determine from the lift calculation?

A. Support for the association is low

B. Leverage of the rules is low
C. The rule is coincidental
D. The rule is true

Answer: C

Question: 12

Consider a database with 4 transactions:
Transaction 1: {cheese, bread, milk}
Transaction 2: {soda, bread, milk}
Transaction 3: {cheese, bread}
Transaction 4: {cheese, soda, juice}
The minimum support is 25%. Which rule has a confidence equal to 50%?

A. {bread, milk} => {cheese}
B. {bread} => {milk}
C. {juice} => {soda}
D. {bread} => {cheese}

Answer: A

Question: 13

Under which circumstance do you need to implement N-fold cross-validation after creating a regression model?

A. There is not enough data to create a test set.
B. The data is unformatted.
C. There are missing values in the data.
D. There are categorical variables in the model.

Answer: A

Question: 14

What is an appropriate data visualization to use in a presentation for an analyst audience?

A. Pie chart
B. Area chart
C. Stacked bar chart
D. ROC curve

Answer: D

Question: 15

When would you use GROUP BY ROLLUP clause in your OLAP query?

A. where all subtotals and grand totals are to be included in the output
B. where only the subtotals are to be included in the output
C. where only the grand totals are to be included in the output
D. where only specific subtotals and grand totals for a combination of variables are to be included in the output

Answer: A

Question: 16

Which type of numeric value does a logistic regression model estimate?

A. Probability
B. A p-value
C. Any integer
D. Any real number

Answer: A

Question: 17

Your colleague, who is new to Hadoop, approaches you with a question. They want to know how best to access their dat
a. This colleague has a strong background in data flow languages and programming.
Which query interface would you recommend?
A. Pig
B. Hive
C. Howl
D. HBase

Answer: A

Question: 18

The web analytics team uses Hadoop to process access logs. They now want to correlate this data with structured user data residing in a production single-instance JDBC database. They collaborate with the production team to import the data into Hadoop. Which tool should they use?

A. Sqoop
B. Pig
C. Chukwa
D. Scribe

Answer: A

Question: 19

What does the R code
z <- f[1:10,]
do?

A. Assigns the first 10 rows of f to the vector z
B. Assigns the 1st 10 columns of the 1st row of f to z
C. Assigns a sequence of values from 1 to 10 to z
D. Assigns the 1st 10 columns to z

Answer: A

Question: 20

In R, functions like plot() and hist() are known as what?

A. generic functions
B. virtual methods
C. virtual functions
D. generic methods

Answer: B

Thank You for trying E20-007 PDF Demo

To Buy Latest E20-007 Full Version Download visit link below

https://www.certkillers.net/Exam/E20-007

Start Your E20-007 Preparation

[Limited Time Offer] Use Coupon "CKNET" for Further discount on your purchase. Test your E20-007 preparation with actual exam questions.