E20-007Q&As

Data Science and Big Data Analytics

Pass EMC E20-007 Exam with 100% Guarantee

Free Download Real Questions & Answers **PDF** and **VCE** file from:

https://www.certbus.com/e20-007.html

100% Passing Guarantee
100% Money Back Assurance

Following Questions and Answers are all new published by EMC Official Exam Center

QUESTION 1

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

Correct Answer: A

QUESTION 2

Which chart type is the most effective way to show trends over time?

A. Line Chart

B. Bar Chart

C. Stacked Bar Chart

D. Histogram

Correct Answer: A

QUESTION 3

Review the following code:

SELECT pn, vn, sum(prc*qty)

FROM sale

GROUP BY CUBE(pn, vn)

ORDER BY 1, 2, 3;

Which combination of subtotals do you expect to be returned by the query?

A. (pn, vn)

B. ((pn, vn), (pn))

C. ((pn, vn) , (pn), (vn))

D. ((pn, vn) , (pn), (vn) , ( ))

Correct Answer: D

QUESTION 4

On analyzing your time series data you suspect that the data represented as y1, y2, y3, ... , yn-1, yn may have a trend component that is quadratic in nature. Which pattern of data will indicate that the trend in

the time series data is quadratic in nature?

A. (y3-y2) ?(y2-y1) = ...= (yn-yn-1)-(yn-1-yn-2)

B. (y2-y1) = (y3-y2) = ... = (yn-yn-1)

C. ((y2-y1) /y1) * 100% = ...((yn-yn-1)/yn-1) * 100%

D. (y4-y2) ?(y3-y1) = ...= (yn-yn-2)-(yn-1-yn-3)

Correct Answer: A

QUESTION 5

You have fit a decision tree classifier using 12 input variables. The resulting tree used 7 of the 12 variables, and is 5 levels deep. Some of the nodes contain only 3 data points. The AUC of the model is

0.85. What is your evaluation of this model?

A. The tree is probably overfit. Try fitting shallower trees and using an ensemble method.

B. The AUC is high, and the small nodes are all very pure. This is an accurate model.

C. The tree did not split on all the input variables. You need a larger data set to get a more accurate model.

D. The AUC is high, so the overall model is accurate. It is not well-calibrated, because the small nodes will give poor estimates of probability.

Correct Answer: A

QUESTION 6

You submit a MapReduce job to a Hadoop cluster and notice that although the job was successfully submitted, it is not completing. What should you do?

A. Ensure that the TaskTracker is running.

B. Ensure that the JobTracker is running

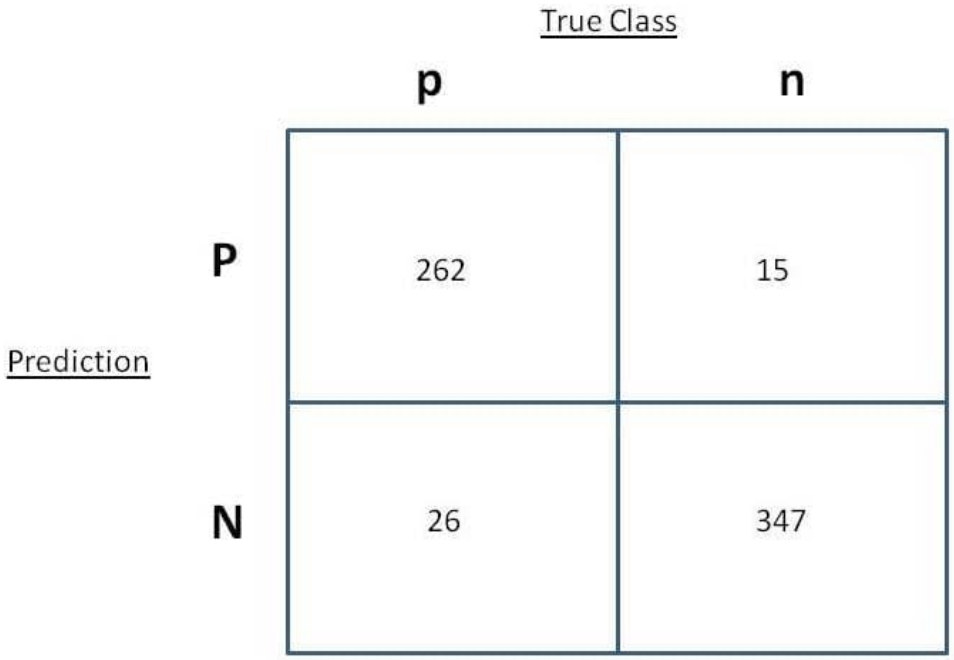
C. Ensure that the NameNode is running

D. Ensure that a DataNode is running

Correct Answer: A

QUESTION 7

Refer to the exhibit.



You have scored your Naive bayesian classifier model on a hold out test data for cross validation and determined the way the samples scored and tabulated them as shown in the exhibit.

What are the the False Positive Rate (FPR) and the False Negative Rate (FNR) of the model?

A. FPR = 15/262 FNR = 26/288

B. FPR = 26/288 FNR = 15/262

C. FPR = 262/15 FNR = 288/26

D. FPR = 288/26 FNR = 262/15

Correct Answer: A

QUESTION 8

Which data type value is used for the observed response variable in a logistic regression model?

A. Any positive real number

B. Any integer

C. A binary value

D. Any real number

Correct Answer: C

QUESTION 9

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

Correct Answer: A

QUESTION 10

You have been assigned to do a study of the daily revenue effect of a pricing model of online transactions. All the data currently available to you has been loaded into your analytics database; revenue data, pricing data, and online transaction data. You find that all the data comes in different levels of granularity. The transaction data has timestamps (day, hour, minutes, seconds), pricing is stored at the daily level, and revenue data is only reported monthly. What is your next step?

A. Report back to the business owner that the current data model does not support the business question.

B. Interpolate a daily model for revenue from the monthly revenue data.

C. Aggregate all data to the monthly level in order to create a monthly revenue model.

D. Disregard revenue as a driver in the pricing model, and create a daily model based on pricing and transactions only.

Correct Answer: A

QUESTION 11

You are studying the behavior of a population and are provided with multi-dimensional data at the individual level. You have identified four specific individuals who are valuable to your study. You would like to find all users who are most similar to each individual.

Which algorithm is most appropriate for this study?

A. K-means clustering

B. Linear regression

C. Association rules

D. Decision trees

Correct Answer: A

QUESTION 12

Refer to the exhibit.

	FREE HOUSING	HOME OWNER	RENTER	TOTAL
BAD CREDIT	54	476	270	800
GOOD CREDIT	75	1245	460	1780
TOTAL	129	1721	730	2580

Click on the calculator icon in the upper left corner. You are going into a meeting where you know your manager will have a question on your dataset -- specifically relating to customers that are classified as renters with good credit status.

In order to prepare for the meeting, you create a rule: RENTER => GOOD CREDIT. What is the confidence of the rule?

A. 63%

B. 41%

C. 18%

D. 73%

Correct Answer: A

QUESTION 13

How does Pig\\'s use of a schema differ from that of a traditional RDBMS?

A. Pig\\'s schema is optional

B. Pig\\'s schema requires that the data is physically present when the schema is defined

C. Pig\\'s schema is required for ETL

D. Pig\\'s schema supports a single data type

Correct Answer: A

QUESTION 14

How are window functions different from regular aggregate functions?

A. Rows retain their separate identities and the window function can access more than the current row.

B. Rows are grouped into an output row and the window function can access more than the current row.

C. Rows retain their separate identities and the window function can only access the current row.

D. Rows are grouped into an output row and the window function can only access the current row.

Correct Answer: A

QUESTION 15

Which word or phrase completes the statement?

Business Intelligence is to ad-hoc reporting and dashboards as Data Science is to _____ .

A. Optimization and Predictive Modeling

B. Alerts and Queries

C. Structured Data and Data Sources

D. Sales and profit reporting

Correct Answer: A