

E20-007^{Q&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.leads4pass.com/e20-007.html>

100% Passing Guarantee
100% Money Back Assurance

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

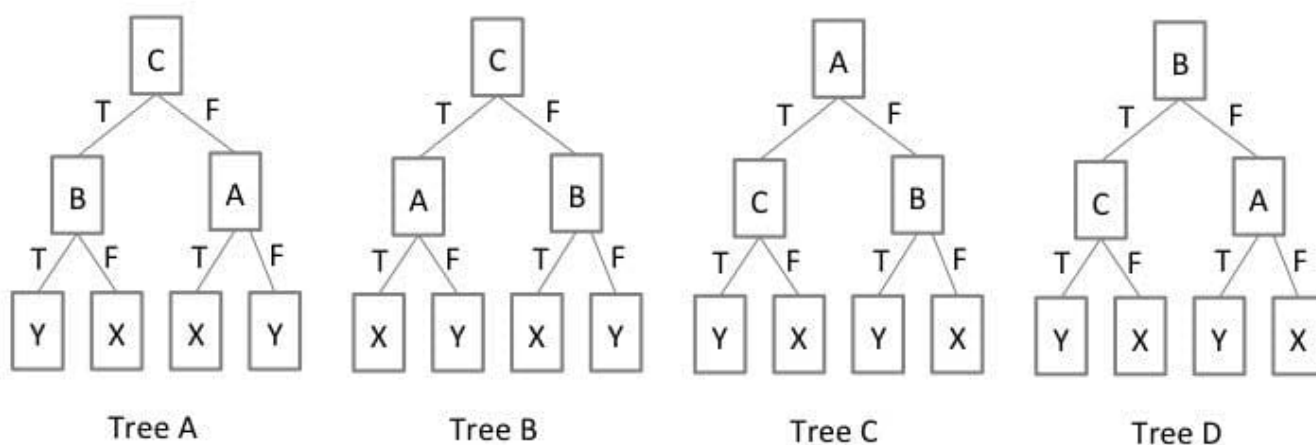
- ⚙ **Instant Download** After Purchase
- ⚙ **100% Money Back** Guarantee
- ⚙ **365 Days** Free Update
- ⚙ **800,000+** Satisfied Customers



QUESTION 1

Refer to the Exhibit.

C	B	A	CLASS
T	T	T	X
T	T	F	Y
T	F	T	X
F	F	F	Y
F	T	T	X
F	F	T	Y



In the Exhibit, the table shows the values for the input Boolean attributes "A", "B", and "C". It also shows the values for the output attribute "class". Which decision tree is valid for the data?

- A. Tree B
- B. Tree A
- C. Tree C
- D. Tree D

Correct Answer: A

QUESTION 2

Consider the following itemsets: (hat, scarf, coat)

(hat, scarf, coat, gloves)

(hat, scarf, gloves)

(hat, gloves)

(scarf, coat, gloves)

What is the confidence of the rule (hat, scarf) => gloves?

A. 40%

B. 50%

C. 60%

D. 66%

Correct Answer: D

QUESTION 3

You have scored your Naïve Bayesian Classifier model on "hold out" test data for cross validation. You have determined the way the samples scored and have tabulated them as shown in the exhibit.

		<u>True Class</u>	
		p	n
<u>Prediction</u>	P	262	15
	N	26	347

What are the Precision and Recall rates of the model?

- A. Precision = 262/277 Recall = 262/288
- B. Precision = 262/288 Recall = 262/277
- C. Precision = 277/262 Recall = 288/262
- D. Precision = 288/262 Recall = 277/262

Correct Answer: A

QUESTION 4

You have run the association rules algorithm on your data set, and the two rules {banana, apple} => {grape} and {apple, orange}=> {grape} have been found to be relevant. What else must be true?

- A. {grape, apple, orange} must be a frequent itemset.
- B. {banana, apple, grape, orange} must be a frequent itemset.
- C. {grape} => {banana, apple} must be a relevant rule.
- D. {banana, apple} => {orange} must be a relevant rule.

Correct Answer: A

QUESTION 5

Which word or phrase completes the statement; "A theater actor is to `artistic and expressive\' as a data scientist is to _____."?

- A. Communicative and collaborative
- B. Introverted and technical
- C. Logical and steadfast
- D. Independent and intelligent

Correct Answer: A

QUESTION 6

What is an appropriate data visualization to use in a presentation to a project sponsor?

- A. Bar Chart
- B. Pie Chart
- C. Box and Whisker Plot

D. Density Plot

Correct Answer: A

QUESTION 7

What is LOESS used for?

- A. It fits a smoothed curve to scatterplot data, to give a general sense of the data's behavior.
- B. It is a significance test for the correlation between two variables.
- C. It plots a continuous variable versus a discrete variable, to compare distributions across classes.
- D. It is run after a one-way ANOVA, to determine which population has the highest mean value.

Correct Answer: A

QUESTION 8

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

- A. Pig
- B. Hive
- C. Hwi
- D. HBase

Correct Answer: A

QUESTION 9

When is the GROUP BY ROLLUP clause used in an OLAP query?

- A. All subtotals and grand totals are to be included in the output
- B. Subtotals are only to be included in the output
- C. Grand totals are only to be included in the output
- D. Specific subtotals and grand totals for a combination of variables are only to be included in the output

Correct Answer: A

QUESTION 10

Which word or phrase completes the statement? A Data Scientist would consider that a RDBMS is to a Table as R is to a _____ .

- A. Data frame
- B. List
- C. Matrix
- D. Array

Correct Answer: A

QUESTION 11

In a Student's t-test, what is the meaning of the p-value?

- A. it is the area under the appropriate tails of the Student's distribution
- B. it is the "power" of the Student's t-test
- C. it is the mean of the distribution for the null hypothesis
- D. it is the mean of the distribution for the alternate hypothesis

Correct Answer: A

QUESTION 12

Your colleague, who is new to Hadoop, approaches you with a question. They want to know how best to access their data. This colleague has previously worked extensively with SQL and databases.

Which query interface would you recommend?

- A. Hive
- B. Pig
- C. Howl
- D. HBase

Correct Answer: A

QUESTION 13

What is the primary bottleneck in text classification?

- A. The availability of tagged training data.
- B. The ability to parse unstructured text data.

- C. The high dimensionality of text data.
- D. The fact that text corpora are dynamic.

Correct Answer: A

QUESTION 14

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

Correct Answer: B

QUESTION 15

You have been assigned to do a study of the daily revenue effect of a pricing model of online transactions. When have you completed the analytics lifecycle?

- A. You have written documentation, and the code has been handed off to the Data Base Administrator and business operations.
- B. You have a completely developed model, and the results have shown statistically acceptable results.
- C. You have presented the results of the model to both the internal analytics team and the business owner of the project.
- D. You have a completely developed model based on both a sample of the data and the entire set of data available.

Correct Answer: A

[E20-007 Practice Test](#)

[E20-007 Study Guide](#)

[E20-007 Exam Questions](#)