



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.pass4lead.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 Consider the training data set shown in the exhibit. What are the classification ($Y = 0$ or 1) and the probability of the classification for the tuple $X(0, 0, 1)$ using Naive Bayesian classifier?

X1	X2	X3	Y
1	1	1	0
1	1	0	0
0	0	0	0
0	1	0	1
1	0	1	1
0	1	1	1

- A. Classification $Y = 1$, Probability = $4/54$
- B. Classification $Y = 0$, Probability = $1/54$
- C. Classification $Y = 1$, Probability = $1/54$
- D. Classification $Y = 0$, Probability = $4/54$

Correct Answer: A

QUESTION 2

What is Hadoop?

- A. Java classes for HDFS types and MapReduce job management and HDFS



- B. Java classes for HDFS types and MapReduce job management and the MapReduce paradigm
- C. MapReduce paradigm and HDFS
- D. MapReduce paradigm and massive unstructured data storage on commodity hardware

Correct Answer: A

QUESTION 3

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 4

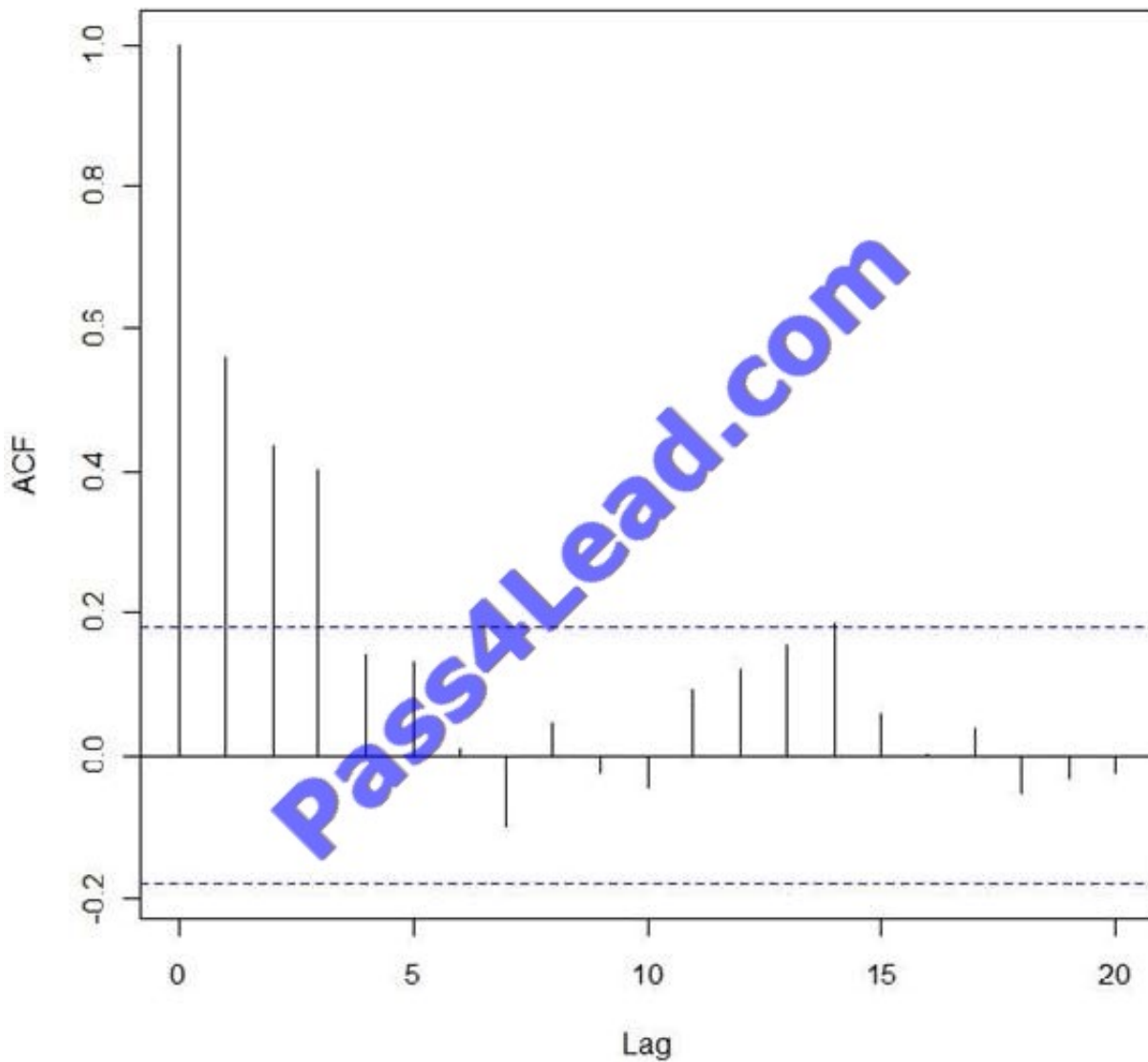
Which data asset is an example of semi-structured data?

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

Correct Answer: A

QUESTION 5

Refer to the exhibit.



In the exhibit, a correlogram is provided based on an autocorrelation analysis of a sample dataset.

What can you conclude from only this exhibit?

- A. There is significant autocorrelation through lag 3
- B. There is no structure left to model in the data
- C. Lag 7 has a significant negative autocorrelation
- D. Differencing is required before proceeding with any analysis

Correct Answer: A

QUESTION 6

When is a Wilcoxon Rank-Sum test used?



- A. When an assumption about the distribution of the populations cannot be made
- B. When the data can be easily sorted
- C. When the populations represent the sums of other values
- D. When the data cannot be easily sorted

Correct Answer: A

QUESTION 7

The web analytics team uses Hadoop to process access logs. They now want to correlate this data with structured user data residing in their massively parallel database. Which tool should they use to export the structured data from Hadoop?

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

Correct Answer: A

QUESTION 8

You are building a logistic regression model to predict whether a tax filer will be audited within the next two years. Your training set population is 1000 filers. The audit rate in your training data is 4.2%. What is the sum of the probabilities that the model assigns to all the filers in your training set that have been audited?

- A. 42.0
- B. 4.2
- C. 0.42
- D. 0.042

Correct Answer: A

QUESTION 9

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 10

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 11

Refer to the exhibit.



Table A: Term inverse document frequencies

	Solid	State	Disk
IDF	2	1.5	0.5

Table B: Term per-document frequencies

Document	Tf		
	Solid	State	Disk
A	0.1	0.1	0.1
B	0.2	0.2	0.1
C	0.1	0.1	0.5
D	0.1	0.3	0.1

Click on the calculator icon in the upper left corner. An analyst is searching a corpus of documents for the topic "solid state disk". In the Exhibit, Table A provides the inverse document frequency for each term across the corpus. Table B provides each term's frequency in four documents selected from corpus. Which of the four documents is most relevant to the analyst's search?

- A. Document B
- B. Document A
- C. Document C
- D. Document D

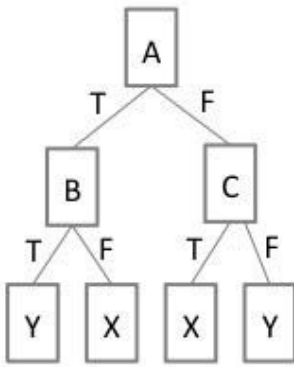
Correct Answer: A

QUESTION 12

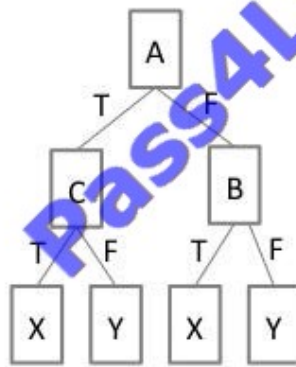
Refer to the Exhibit.



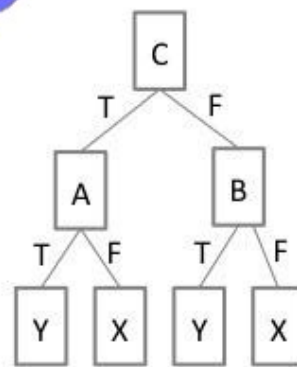
A	B	C	CLASS
T	T	T	X
T	T	F	Y
T	F	T	X
F	F	F	Y
F	T	T	X
F	F	T	Y



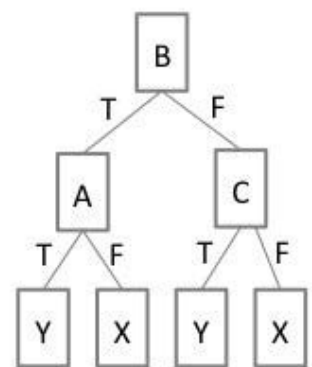
Tree A



Tree B



Tree C



Tree D

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



To Read the [Whole Q&As](#), please purchase the [Complete Version](#) from [Our website](#).

Try our product !

100% Guaranteed Success

100% Money Back Guarantee

365 Days Free Update

Instant Download After Purchase

24x7 Customer Support

Average 99.9% Success Rate

More than 800,000 Satisfied Customers Worldwide

Multi-Platform capabilities - [Windows](#), [Mac](#), [Android](#), [iPhone](#), [iPod](#), [iPad](#), [Kindle](#)

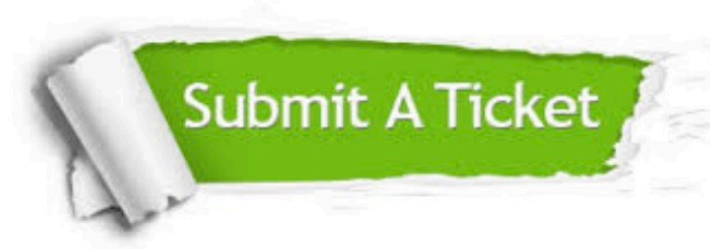
We provide exam PDF and VCE of Cisco, Microsoft, IBM, CompTIA, Oracle and other IT Certifications. You can view Vendor list of All Certification Exams offered:

<https://www.pass4lead.com/allproducts>

Need Help

Please provide as much detail as possible so we can best assist you.

To update a previously submitted ticket:



 <p>One Year Free Update Free update is available within One Year after your purchase. After One Year, you will get 50% discounts for updating. And we are proud to boast a 24/7 efficient Customer Support system via Email.</p>	 <p>Money Back Guarantee To ensure that you are spending on quality products, we provide 100% money back guarantee for 30 days from the date of purchase.</p>	 <p>Security & Privacy We respect customer privacy. We use McAfee's security service to provide you with utmost security for your personal information & peace of mind.</p>
---	---	--

Any charges made through this site will appear as Global Simulators Limited.

All trademarks are the property of their respective owners.

Copyright © pass4lead, All Rights Reserved.