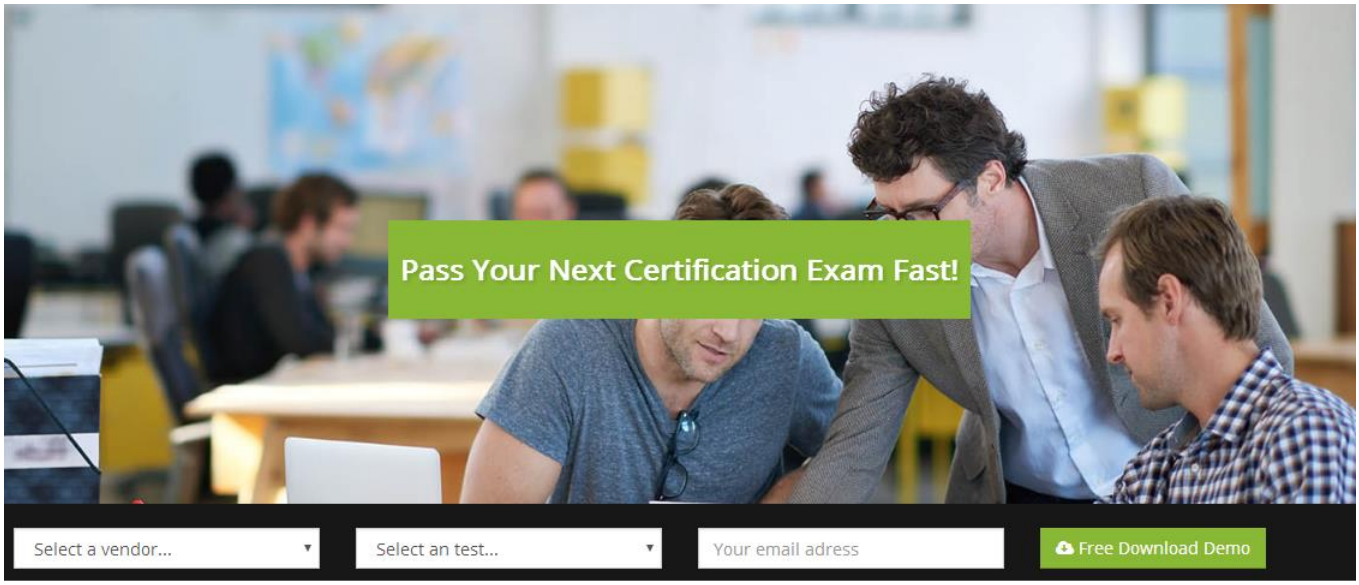
ITPassLeader



Instant Download | 365 Days Free Updates | Money Back Guarantee | Security & Privacy

Choose the version that fits your needs	PDF Version	Desktop Test Engine	Online Test Engine
Latest and Up-to-Date exam dumps with real exam questions answers.	✓	✓	✓
Get 12-Months free updates without any extra charges.	✓	✓	✓
Experience same exam environment before appearing in the certification exam.	✗	✓	✓
100% exam passing guarante in the first attempt.	✓	✓	✓
20% discount on more than one license and 30% discount on 5+ license purchases.	✗	✓	✓
100% secure purchase on SSL.	✓	✓	✓
Completely private purchase without sharing your personal info with anyone.	✓	✓	✓

http://www.itpassleader.com

High-praise Exam Dumps Questions grant you success by high pass rate - ITPassLeader

Exam : E20-007

Title : Data Science and Big Data Analytics

Vendors : EMC

Version : DEMO

NO.1 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

NO.2 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

NO.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

NO.4 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

NO.5 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

NO.6 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

NO.7 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

NO.8 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

NO.9 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. Howl
D. HBase
Answer: A

NO.10 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

NO.11 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

NO.12 In which lifecycle stage are test and training data sets created?
A. Model building
B. Model planning
C. Discovery
D. Data preparation
Answer: A

NO.13 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

NO.14 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

NO.15 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: D

NO.16 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. Na ve Bayesian classifier
B. Linear regression
C. Logistic regression
D. K-means clustering
Answer: A

NO.17 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

NO.18 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

NO.19 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

NO.20 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