VCE & PDF
https://www.passapply.com

Passapply.com

E20-065^Q&As

Advanced Analytics Specialist Exam for Data Scientists

Pass EMC E20-065 Exam with 100% Guarantee

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

https://www.passapply.com/e20-065.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

Which HDFS feature protects against user errors causing accidental loss of data?

A. Encryption

B. Replication

C. Namenode federation

D. Snapshots

Correct Answer: B

QUESTION 2

What is a key beneficial characteristic of the Random Forest algorithm?

A. Provides and explanatory model

B. Distinguishes categorical from continuous variables

C. Support for unstructured data

D. Resiliency to complex, non-linear variable interactions

Correct Answer: D

QUESTION 3

How is the relative value of a node visualized in a sunburst?

A. Color

B. Area

C. Gradient

D. Position

Correct Answer: A

QUESTION 4

What is a random subspace of features, as used by Random Forests?

A. A random subset of features that are chosen at each split in the decision tree

B. Filtration of data that does not meet a pre-defined weighting thrsehold

C. The creation of out-of-bag (OOB) data that is used to select features

D. Removal of highly correlated variables to randomize the features

Correct Answer: A

QUESTION 5

What are the major components of the YARN architecture?

A. ResourceManager and NodeManager

B. Task Tracker and NameNode

C. HDFS, Tez, and Spark

D. Avro, ZooKeeper, and HDFS

Correct Answer: A

QUESTION 6

What is a characteristic of stemming?

A. Reduces words of variant forms to their base forms based on a set of heuristics

B. Can be performed by calling the stemming!) function on a lemma in NLTK

C. Can be performed by calling the stemming() function on a synset in NLTK

D. Reduces words of variant forms to their base forms based on a dictionary

Correct Answer: A

QUESTION 7

What is a characteristic of stop words?

A. Used in term frequency analysis

B. Include words such as "a", "an", and "the"

C. Meaningful words requiring a parser to stop and examine them

D. Don\\'t occur often in text

Correct Answer: B

QUESTION 8

A simul-ation to compare two different sales models yields different results for the same set of input

variables in different runs.

What is the likely cause?

A. bit operating system was used

B. The same number of trials was used.

C. A linear congruenlial generator (LCG) was used for pseudo-random number generation.

D. Different seeds forthe random number generator were used

Correct Answer: C

QUESTION 9

What elements are needed to determine the time complexity of finding all the cliques of size k in social network analysis?

A. Eigenvector centrality and betwenness

B. Clique size and total number of nodes in the network

C. Number of edges in the network and centrality measure of the cliques

D. Clique size and betweenness centrality

Correct Answer: B

QUESTION 10

In the graph, which edge would be considered a weak lie? Refer to the exhibit.

A. C-E

B. E-F

C. B-C

D. G-I

Correct Answer: D

QUESTION 11

What is the maximum degree of a node in an undirected graph with 50 nodes\\'?

A. 49

B. 50

C. 1250

D. 2500

Correct Answer: A

QUESTION 12

What best describes tokenization?

A. Adding lexical relations to the raw text

B. Converting text into the list of terms

C. Converting text into a list of unique terms

D. Reducing variant forms of tokens to their base forms

Correct Answer: B

QUESTION 13

What does YARN provide over and above MapReduce?

A. Separate cluster and resource management

B. Parallelized processing

C. Serialized processing

D. Access to HDFS data

Correct Answer: A

QUESTION 14

Which scenario is a proper use case for multinomial logistic regression?

A. A marketing firm wants to estimate the personal income of a group of potential customers. Using inputs such as age, education, marital status, and credit card expenditures, a data scientist is building a model that will estimate a person\\'s income

B. A logistic distribution company wants to minimize the distance traveled by its delivery trucks. A data scientist is building a model to determine the optimal route for each of tis trucks

C. To improve the initial routing of a loan application, a financial institution plans to classify a loan application as Approve, Reject, or Possibly_Approve. Based on the company\\'s historical loan application data, a data scientist is building a model to assign one of these three outcomes to each submitted application.

D. A manufacturer plans to determine the optimal number of workers to employ in an assembly line process. Utilizing the observed distributions of the task durations of each process step, a data scientist is building a model to mimic the interactions and dependencies between each stage in the manufacturing process.

Correct Answer: C

QUESTION 15

What describes how nodes in a social network are similar to each other in characteristics?

A. Community clustering

B. Modularity

C. Homophily

D. Strongly tied network

Correct Answer: C

E20-065 VCE Dumps   E20-065 Practice Test   E20-065 Braindumps