Vendor:EC-COUNCIL

Exam Code:E20-007

Exam Name: Data Science and Big Data Analytics

Version: Demo

QUESTION 1

Before you build an ARMA model, how can you tell if your time series is weakly stationary?

A. There appears to be a constant variance around a constant mean.

B. The mean of the series is close to 0.

C. The series is normally distributed.

D. There appears to be no apparent trend component.

Correct Answer: A

QUESTION 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

Correct Answer: C

QUESTION 3

Which process in text analysis can be used to reduce dimensionality?

A. Stemming

B. Parsing

C. Digitizing

D. Sorting

Correct Answer: A

QUESTION 4

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(1, 0, 0) 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 = 0, Probability = 4/54

B. Classification Y = 1, Probability = 4/54

C. Classification Y = 0, Probability = 1/54

D. Classification Y = 1, Probability = 1/54

Correct Answer: A

QUESTION 5

What is the purpose of the process step "parsing" in text analysis?

A. imposes a structure on the unstructured/semi-structured text for downstream analysis

B. performs the search and/or retrieval in finding a specific topic or an entity in a document

C. executes the clustering and classification to organize the contents

D. computes the TF-IDF values for all keywords and indices

Correct Answer: A

QUESTION 6

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 7

In which lifecycle stage are initial hypotheses formed?

A. Discovery

B. Model planning

C. Model building

D. Data preparation

Correct Answer: A

QUESTION 8

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 9

Which R data structure allows elements to have different data types?

A. List

B. Vector

C. Matrix

D. Array

Correct Answer: A

QUESTION 10

You are attempting to find the Euclidean distance between two centroids:

Centroid A\\'s coordinates: (X = 2, Y = 4)

Centroid B\\'s coordinates (X = 8, Y = 10)

Which formula finds the correct Euclidean distance?

A. SQRT((2-8)2+(4-10)2) or 8.49

B. SQRT(((2-8) x 2) + ((4-10) x 2)) or 12.17

C. ((2-8)2+(4-10)2) or 72

D. ((2-8) x 2 + (4-10) x 2) or 148

Correct Answer: A

QUESTION 11

You are testing two new weight-gain formulas for puppies. The test gives the results: Control group: 1% weight gain
Formula A. 3% weight gain

Formula B. 4% weight gain A one-way ANOVA returns a p-value = 0.027 What can you conclude?

A. Either Formula A or Formula B is effective at promoting weight gain.

B. Formula B is more effective at promoting weight gain than Formula A.

C. Formula A and Formula B are both effective at promoting weight gain.

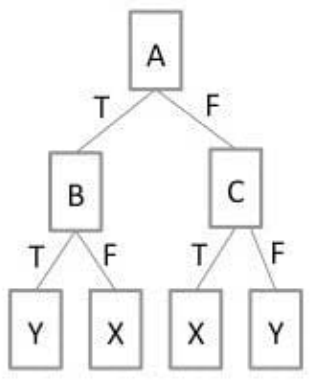D. Formula A and Formula B are about equally effective at promoting weight gain.
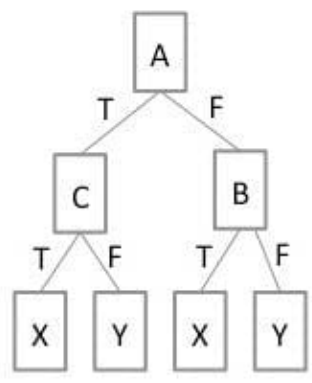
Correct Answer: A

QUESTION 12

Based on the exhibit, the table shows the values for the input Boolean attributes A, B, and

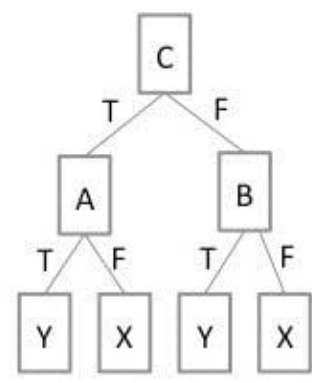C. In addition, the exhibit shows the values for the output attribute "class".

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

```
        A                    A                    C                    B
      T / \ F              T / \ F              T / \ F              T / \ F
     B     C              C     B              A     B              A     C
   T/ \F  T/ \F         T/ \F  T/ \F         T/ \F  T/ \F         T/ \F  T/ \F
   Y   X  X   Y         X   Y  X   Y         Y   X  Y   X         Y   X  Y   X

     Tree A               Tree B               Tree C               Tree D
```
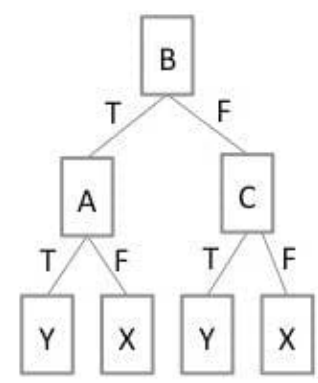
Which decision tree is valid for the data?

A. Tree A

B. Tree B

C. Tree C

D. Tree D

Correct Answer: B