COMP60711 – Part 2: Data Analytics and Big Data - Lecture 1 Introduction

These multiple-choice questions may require reading in addition to the lecture notes.
There may be more than 1 correct answer.

1. The process of forming general concept definitions from examples of concepts to be learned

A. deduction
B. abduction
C. induction
D. conjunction

2. A statement to be tested is a

A. theory
B. procedure
C. principle
D. hypothesis

3. A structure designed to store data for decision support.

A. operational database
B. flat file
C. decision tree
D. data warehouse

4. Which of the following are components of data science?

A. Data Engineering
B. Advanced Computing
C. Domain expertise
D. All of the above

5. Data mining is best described as the process of

A. identifying patterns in data.
B. deducing relationships in data.
C. representing data.
D. simulating trends in data.

6. Data used to build a data mining model

A. validation data
B. training data
C. test data
D. hidden data.

7. Choose all of the following that are a prediction problem that comes under the task of classification:

A. Please tell me which of my current mortgage applicants I can safely approve for a mortgage and which are unsafe at this time to approve for any kind of mortgage.
B. Please tell me the likelihood of each of my current mortgage clients to default on their mortgage in the next 12 months.
C. Please tell me the maximum monthly payment Client X can safely make, that is avoid becoming a risk for foreclosure
D. Please tell me which of my current mortgage clients are suitable candidates for a low fixed interest rate mortgage, which for an initial low-rate, variable rate mortgage, and which for a high, fixed interest rate mortgage

8. If the classes of the training data are an unknown, we can apply which of these computational algorithms to attempt to find useable classes

A. binary sort
B. Bayesian analysis
C. pruning analysis
D. clustering

9. This approach is best when we are interested in finding all possible interactions among a set of attributes.

A. decision tree
B. association rules
C. K-Means algorithm
D. genetic learning

10. Which of the following is a common use of unsupervised clustering?

A. detect outliers
B. determine a best set of input attributes for supervised learning
C. evaluate the likely performance of a supervised learner model
D. determine if meaningful relationships can be found in a dataset
E. All of the above.

11. _____ is a methodology useful for discovering interesting relationships within large sets of data

A. big data
B. association analysis
C. data Mining
D. algorithm

12. Supervised learning and unsupervised clustering both require at least one

A. hidden attribute.
B. output attribute.
C. input attribute.
D. categorical attribute.

13. Supervised learning differs from unsupervised clustering in that supervised learning requires

A. at least one input attribute.
B. input attributes to be categorical.
C. at least one output attribute.
D. ouput attriubutes to be categorical.

14. Which of the following is the best tool for dealing with data that needs to be labeled, but there are no predetermined labels:

A. pattern Matchers
B. classifiers
C. clusterers
D. recognizers

15. _____ is a form of data analysis that extracts models describing important data classes

A. classification
B. model
C. attribute set
D. class model

16. Which of the following is not a characteristic of a data warehouse?

A. contains nonvolatile data
B. is subject oriented
C. supports data processing, collection and management
D. stores data to be reported on, analyzed and tested

17. Which of the following is not a characteristic of a data warehouse?

A. contains historical data
B. designed for decision support
C. stores data in normalized tables
D. promotes data redundancy

18. _____ assesses the feasibility of clustering analysis on a data set and the quality of the results generated by a clustering method.

A. grid-based method
B. analytics
C. clustering evaluations

19. Sets of frequent items hidden in large data sets are called _____

A. association rules.
B. big data.
C. binary representation.
D. analysis.

20. The strength of an association rule can be measured in terms of its____ and _____

A. support
B. finances
C. confidence
D. size

21. Data classification is a _____ step process

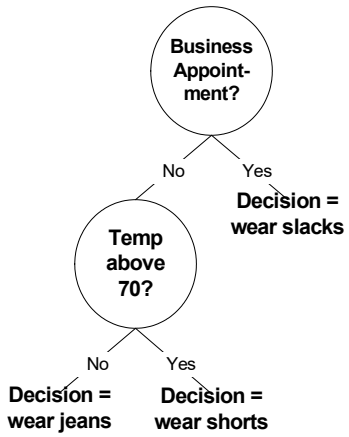
A. three
B. one
C. two
D. four

22. Which of these terms describes the first major task in the data classification process:

A. classify
B. learning
C. choose training data
D. analyze the training data for possible "noise" in the database
E. None of the above

23. Which of the following is a valid rule for the decision tree below?

A. IF Business Appointment = No & Temp above 70 = No
THEN Decision = wear slacks
B. IF Business Appointment = Yes & Temp above 70 = Yes
THEN Decision = wear shorts
C. IF Temp above 70 = No
THEN Decision = wear shorts
D. IF Business Appointment= No & Temp above 70 = No
THEN Decision = wear jeans



Matching Questions

Determine which is the best approach for each problem.

a. *supervised learning*
b. *unsupervised clustering*
c. *data query*

24. What is the average weekly salary of all employees under forty years of age?
25. Develop a profile for credit card customers likely to carry an average monthly balance of more than $1000.00.
26. Determine the characteristics of a successful used car salesperson.
27. What attribute similarities group customers holding one or several insurance policies?
28. Do meaningful attribute relationships exist in a database containing information about credit card customers?
29. Do young people play more tennis than older people?
30. Determine whether a credit card transaction is valid or fraudulent.

31. How not to get shot!

Please read the following references and attempt the *brain teaser* exercise in Part 2.

The Cowboy Problem: How not to get shot – Part 1; Prad Upadrashta, April 2020
https://blogs.mastechinfotrellis.com/the-cowboy-problem-how-not-to-get-shot-part-1

The Cowboy Problem: How not to get shot – Part 2; Prad Upadrashta, April 2020
https://blogs.mastechinfotrellis.com/the-cowboy-problem-how-not-to-get-shot-part-2

BLANK PAGE – answers on next page

MCQ Answers

1. C
2. D
3. D
4. D
5. A
6. B
7. A, D
8. D
9. B
10. E
11. B
12. C
13. C
14. C
15. A
16. C
17. C
18. C
19. A
20. A, C
21. A
22. B
23. D

Matching Answers

24. C
25. A
26. A
27. A
28. B
29. C
30. A

Acknoledgement is made to various surces.