استخلاص البيانات وذكاء الأعمال
3:5
26/3/2021 الجمعة
أ.د/مرغنى حسن محمد

Faculty of Computers & Information, Assiut University
4th Level
Final Exam
Duration: 2 hours

Hi Mohamed, when you submit this form, the owner will be able to see your name and email address.

* Required

1. الإسم الرباعي (بالعربي فقط) *

محمد منصور عيد رياض

2. رقم الجلوس *

162014220

3. المستوي *

◯ الاول

- ○ الثاني
- ○ الثالث
- ○ الرابع 2013
- ◉ الرابع 2014
- ○ الرابع 2015
- ○ الرابع 2016
- ○ الرابع 2017

4. البرنامج *

- ◉ عام
- ○ بايو
- ○ هندسة

5. رقم المعمل *

د٣

6. رقم الكمبيوتر *

11

7. كود المراقب *

DE23

8. إسم الملف بالمسار (تملي بمعرفت المراقب) *

Enter your answer

9. Test set is independent of training set, otherwise over-fitting will occur

a. True
b. False *
(1 Point)

- ○ a
- ◉ b

10. What is true about data mining?

A. Data Mining is defined as the procedure of extracting information from huge sets of data
B. Data mining also involves other processes such as Data Cleaning, Data Integration, Data Transformation
C. Data mining is the procedure of mining knowledge from data.
D. All of the above *
(1 Point)

- ○ a
- ○ b
- ○ c
- ◉ d

11. A frequent set is a closed frequent set if it is a frequent set and no superset of this is a frequent set.
a. True
b. False *
(1 Point)

- (●) a
- ( ) b

12. Which one of the following is alternative search strategies for mining multiple-level associations with reduced support?
a) Level – by level independent
b) Level – cross-filtering by a single item
c) Level – cross-filtering by k – itemset:
d) All the above *
(1 Point)

- ( ) a
- ( ) b
- ( ) c
- (●) d

13. Association rule support is defined as
a. the percentage of instances that contain the antecendent conditional items listed in the association rule.
b. the percentage of instances that contain the consequent conditions listed in the association rule.
c. the percentage of instances that contain all items listed in the association rule.
d. the percentage of instances in the database that contain at least one of the antecendent conditional items listed in the association rule. *
(1 Point)

- (●) a
- ( )

b

c

d

14. Data Mining System Classification consists of?
A. Database Technology
B. Machine Learning
C. computer Vision
D. All of the above *
(1 Point)

a

b

c

d

15. Classification accuracy is
A. A subdivision of a set of examples into a number of classes
B. Measure of the accuracy, of the classification of a concept that is given by a certain theory
C. The task of assigning a classification to a set of examples
D. None of these *
(1 Point)

a

b

c

d

16. To improve accuracy, data mining programs are used to analyze audit data and extract features that can distinguish normal activities from intrusions.
a.true
b.false *
(1 Point)

- ○ a
- ◉ b

17. A type of data that represents the numeric values of specific variables. for example age number of children etc
a. Ratio data
b. Numeric data
c. Nominal data
d. Interval data *
(1 Point)

- ○ a
- ◉ b
- ○ c
- ○ d

18. Noise is a random error or variance in measured variables.
a.True
b.False *
(1 Point)

- ◉ a
- ○ b

19. Consider discretizing a continuous attribute whose values are listed below:
3, 4, 5, 10, 21, 32, 43, 44, 46, 52, 59, 67
Using equal-width partitioning and four bins, how many values are there in the first bin (the bin with small values)?
A. 1
B. 2
C. 3
D. 4 *
(1 Point)

- ○ a
- ○ b
- ○ c
- ◉ d

20. Prediction can be viewed as the construction and use of a model to assess the class of an unlabeled sample, or to assess the value or value ranges of an attribute that a given sample is likely to have.
a. true
b.false *
(1 Point)

- ◉ a
- ○ b

21. The information gain measure is used to select the test attribute at each node in the tree.
a.true
b,false *
(1 Point)

- ◉ a
- ○ b

22. Which statement about outliers is true?
    a. Outliers should be identified and removed from a dataset.
    b. Outliers should be part of the training dataset but should not be present in the test data.
    c. Outliers should be part of the test dataset but should not be present in the training data.
    d. The nature of the problem determines how outliers are used. *
    (1 Point)

- (●) a
- ( ) b
- ( ) c
- ( ) d

23. Classification methods is robustness if it can handle noise and missing values

    a. True
    b. False *
    (1 Point)

- (●) a
- ( ) b

24. A class of learning algorithm that tries to find an optimum classification of a set of examples using the probabilistic theory is named as ...
    a) Bayesian classifiers
    b) Dijkstra classifiers
    c) doppler classifiers
    d) all of these *
    (1 Point)

- (●) a
- ( ) b
- ( ) c

d

25. The first steps involved in the knowledge discovery is?
A. Data Integration
B. Data Selection
C. Data Transformation
D. Data Cleaning *
(1 Point)

a

b

c

d

26. Data used to build a data mining model.
a. validation data
b. training data
c. test data
d. hidden data *
(1 Point)

a

b

c

d

27. The problem of finding hidden structure in unlabeled data is called
A. Supervised learning
B. Unsupervised learning
C. Reinforcement learning *
(1 Point)

a

- (●) b
- ( ) c

28. Data mining can be used to improve _____.
a) Efficiency
b) Quality of data
c) Marketing
d) All the above *
(1 Point)

- ( ) a
- ( ) b
- ( ) c
- (●) d

29. Which of the following is not a data mining functionality?

A) Characterization and Discrimination
B) Classification and regression
C) Selection and interpretation
D) Clustering and Analysis *
(1 Point)

- ( ) a
- ( ) b
- (●) c
- ( ) d

30. The most commonly used algorithm to discover association rules by recursively identifying frequent item sets
    a. Apriori algorithm
    b. Ordinal data
    c. Nominal data
    d. Categorical data *
    (1 Point)

- ◯ a
- ◉ b
- ◯ c
- ◯ d

31. "Efficiency and scalability of data mining algorithms" issues comes under?
    A. Mining Methodology and User Interaction Issues
    B. Performance Issues
    C. Diverse Data Types Issues
    D. None of the above *
    (1 Point)

- ◯ a
- ◉ b
- ◯ c
- ◯ d

32. Relevance analysis (feature selection) remove the irrelevant or redundant attributes
    a. True
    b. False *
    (1 Point)

- ◉ a
- ◯ b

33. Which of the following is true for Classification?
    a) A subdivision of a set
    b) A measure of the accuracy
    c) The task of assigning a classification
    d) All of these *
    (1 Point)

- ○ a
- ○ b
- ○ c
- ◉ d

34. Redundancies can be detected by correlation analysis
    a. True
    b. False *
    (1 Point)

- ◉ a
- ○ b

35. In decision tree Induction, tree is constructed in a top-down recursive divide-and-conquer manner
    a. True
    b. False *
    (1 Point)

- ◉ a
- ○ b

36. This technique uses mean and standard deviation scores to transform real-valued attributes.
    a. decimal scaling
    b. min-max normalization
    c. z-score normalization
    d. logarithmic normalization *
    (1 Point)

- [ ] a
- [ ] b
- [x] c
- [ ] d

37. Scalability method refers to the ability to construct the model efficiently given a large amount of data.
    a. true
    b.false *
    (1 Point)

- [x] a
- [ ] b

38. In which step of Knowledge Discovery, multiple data sources are combined?
    A. Data Cleaning
    B. Data Integration
    C. Data Selection
    D. Data Transformation *
    (1 Point)

- [ ] a
- [x] b
- [ ] c
- [ ] d

39. Discovery is

A. It is hidden within a database and can only be recovered if one is given certain clues (an example IS encrypted information).

B. The process of executing implicit previously unknown and potentially useful information from data

C. An extremely complex molecule that occurs in human chromosomes and that carries genetic information in the form of genes.

D. None of these *

(1 Point)

- ○ a
- ○ b
- ○ c
- ◉ d

40. 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 *

(1 Point)

- ○ a
- ◉ b
- ○ c
- ○ d

41. This step of the KDD process model deals with noisy data.
    a. Creating a target dataset
    b. data preprocessing
    c. data transformation
    d. data mining *
    (1 Point)

- ○ a
- ○ b
- ○ c
- ◉ d

42. Naïve Bayesian prediction requires each conditional prob. be zero
    a. True
    b.False *
    (1 Point)

- ○ a
- ◉ b

43. This data transformation technique works well when minimum and maximum values for a real-valued attribute are known.
    a. min-max normalization
    b. decimal scaling
    c. z-score normalization
    d. logarithmic normalization *
    (1 Point)

- ◉ a
- ○ b
- ○ c
- ○ d

44. Given a rule of the form IF X THEN Y, rule confidence is defined as the conditional probability that
    a. Y is true when X is known to be true.
    b. X is true when Y is known to be true.
    c. Y is false when X is known to be false.
    d. X is false when Y is known to be false. *
    (1 Point)

- ◯ a
- ◉ b
- ◯ c
- ◯ d

45. Attibutes may be eliminated from the target dataset during this step of the KDD process.
    a. creating a target dataset
    b. data preprocessing
    c. data transformation
    d. data mining *
    (1 Point)

- ◉ a
- ◯ b
- ◯ c
- ◯ d

46. Gain ratio tends to prefer unbalanced splits in which one partition is much smaller than the others
    a. True
    b. False *
    (1 Point)

- ◉ a

○ b

47. Which of the following statements about Naive Bayes is incorrect?
    A. Attributes are equally important.
    B. Attributes are statistically dependent of one another given the class value.
    C. Attributes are statistically independent of one another given the class value.
    D. All of the above *
    (1 Point)

○ a

○ b

◉ c

○ d

48. Data mining can be used to help predict future patient behaviour and to improve treatment programs
    a.true
    b.false *
    (1 Point)

◉ a

○ b

49. Information gain measure is not biased towards attributes with a large number of values

    a,True
    b.False *
    (1 Point)

◉ a

○ b

50. An itemset X is closed if X is frequent and there exists super-pattern Y ↄ X, with the same support as X
a. true
b.false *
(1 Point)

- [x] a
- [ ] b

51. Task of inferring a model from labeled training data is called
A. Unsupervised learning
B. Supervised learning
C. Reinforcement learning *
(1 Point)

- [ ] a
- [x] b
- [ ] c

52. In association rule mining the generation of the frequent itermsets is the computational intensive step
a.True
b. False *
(1 Point)

- [x] a
- [ ] b

53. FP–Tree Growth Algorithm can be implemented in tow Phases.
a. True
b. False *
(1 Point)

- [x] a

b

54. The term data mining was originally used to _____.
    a. include most forms of data analysis in order to increase sales
    b. describe the prices through which previously unknown patterns in data were discovered
    c. describe the analysis of huge datasets stored in data warehouses
    d. All of the above *
    (1 Point)

a

b

c

d

55. The basic algorithm for decision tree induction is a greedy algorithm
    a.true
    b.false *
    (1 Point)

a

b

56. The correlation coefficient for two real-valued attributes is –0.85. What does this value tell you?
    a. The attributes are not linearly related.
    b. As the value of one attribute increases the value of the second attribute also increases.
    c. As the value of one attribute decreases the value of the second attribute increases.
    d. The attributes show a curvilinear relationship. *
    (1 Point)

a

- (●) b
- ( ) c
- ( ) d

57. In Smoothing by bin means each value in a bin is replaced by the mean value of the bin.
a. True
b.False *
(1 Point)

- ( ) a
- (●) b

58. 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. *
(1 Point)

- ( ) a
- (●) b
- ( ) c
- ( ) d

Submit

password.