Suez Canal University
Faculty of Computers & Informatics
Department of Information Systems

Answer _ Midterm Exam
First Term: 2021 /2022

Program: Information System	**Course: Data Mining**	**Course Code: IS253**
Level: 4	**Lecturer: Dr. Osama Farouk**	**Date: 25 / 11 /2021**
Total pages: 4	**Total marks: 40**	**Time allowed: 1h.**

Answer the following questions:

Question (Select the correct answer) (40 marks)

1. **Which of the following is an essential process in which the intelligent methods are applied to extract data patterns?**
   A. Warehousing B. Data Mining c. Text Mining D. Data Selection
2. **Data Matrix, Document Data and Transaction Data are examples of which type of data set?**
   A. Graph B. Record C. Numerical D. Ordered
3. **For what purpose, the analysis tools pre-compute the summaries of the huge amount of data?**
   A. In order to maintain consistency
   B. For authentication
   C. For data access
   D. To obtain the queries response
4. **What are the functions of Data Mining?**
   A. Association and correctional analysis classification
   B. Prediction and characterization
   C. Cluster analysis and Evolution analysis
   D. All of the above
5. **Which of the following statements is correct about data mining?**
   A. It can be referred to as the procedure of mining knowledge from data
   B. Data mining can be defined as the procedure of extracting information from a set of the data
   C. The procedure of data mining also involves several other processes like data cleaning, data transformation, and data integration
   D. All of the above
6. **Which of the following correctly refers the data selection?**
   A. A subject-oriented integrated time-variant non-volatile collection of data in support of management
   B. The actual discovery phase of a knowledge discovery process
   C. The stage of selecting the right data for a knowledge discovery (KDD) process
   D. All of the above
7. **Which one of the following can be considered as the correct application of the data mining?**
   A. Fraud detection
   B. Corporate Analysis & Risk management
   C. Management and market analysis
   D. All of the above
8. **Which of the following used as the first step in the knowledge discovery process?**
   A. Data selection B. Data cleaning C. Data transformation D. Data integration

9. Which of the following terms is used as a synonym for data mining?
A. knowledge discovery in databases B. data warehousing
C. regression analysis D. parallel processing in databases

10…….. is the out put of KDD
A. Query B. Useful Information C. Data D. Information

11. Data mining is ……..
A. time variant non-volatile collection of data
B. The actual discovery phase of a knowledge
C. The stage of selecting the right data
D. None of these

12. Which of the following refers to the steps of the knowledge discovery process, in which the several data sources are combined?
A. Data selection B. Data cleaning C. Data transformation D. Data integration

13. Data objects with characteristics that are considerably different than most of the other data objects in the data set refers to …… .
A. Noise B. Outliers C. Missing values D. Duplicate data

14. Data warehouse is…….
A. The actual discovery phase of a knowledge discovery process
B. The stage of selecting the right data for a KDD process
C. A subject-oriented integrated time-variant non-volatile collection of data in support of management
D. None of these

15. Spatial Data, Temporal Data, Sequential Data, and Genetic Sequence are examples of which type of data set?
A. Graph B. Record C. Numerical D. Ordered

16. An invalid signal overlapping valid data refers to …….
A. Noise B. Outliers C. Missing values D. Duplicate data

17. Data objects with characteristics that are considerably different than most of the other data objects in the data set refers to …… .
A. Noise B. Outliers C. Missing values D. Duplicate data

18. Which of the following is NOT one of the processes in Data Preprocessing?
A. Feature creation B. Sampling C. Discriminization D. Aggregation

19. Combining two or more attributes (or objects) into a single attribute (or object) explains about …… .
A. Binarization B. Aggregation
C. Dimensionality reduction D. Attribute transformation

20. All of the following are the purposes of aggregation EXCEPT:
A. More "stable" data B. Change of Scale C. Remove noise D. Data Reduction

21. The key principle for effective sampling is …… .
A. a sample will work almost as well as using the entire set if the sample is representative.
B. the appropriate ratio acceptable as sample.
C. using the correct tools and methodology to arrive at a appropriate value for sample.
D. the efficient and effective way to represent the whole population.

22……. is the main technique employed for data selection.
A. Change of scale B. Preprocessing C. Sampling D. Discretization

23. There is an equal probability of selecting any particular item refers to ……..
A. sampling with replacement B. simple random sampling
C. stratified sampling D. sampling without replacement

24. Split the data into several partitions; draw random samples from each partition refers to …… .
A. random sampling B. sampling with replacement
C. sampling without replacement D. stratified sampling

25.....refers to the mapping or classification of a class with some predefined group or class
A. Data Discrimination B. Data Characterization
C. Data Definition D. Data Visualization

26. A college professor wishes to reach a certain level of savings before her retirement. This is related to which data mining task?
A. Clustering B. Regression C. Association D. Classification

27. Suppose that the data for analysis includes the attribute age. The age values for the data tuples are (in increasing order) 13, 15, 16, 16, 19, 20, 20, 21, 22, 22, 25, 25, 25, 25, 30, 33, 33, 35, 35, 35, 35, 36, 40, 45, 46, 52, 70. Five-number summary of a distribution, Minimum, Q1, Median, Q3, Maximum are according to their order
A. 13,20,25,35,70 B. 13,35, 25, 20, 70 C.13,25,20,30,70 D. 0 ,1 ,1 ,1 ,0

28. Find cosine similarity between documents 1 and 2.
d_1 = **(5, 0, 3, 0, 2, 0, 0, 2, 0, 0)**
d_2 = **(3, 0, 2, 0, 1, 1, 0, 1, 0, 1)**
A. 0.74 B. 0.94 C. 0.84 D. 6.15

29. …….. is combines data from multiple sources into a coherent store
A. Data Discretization B. Data Cleaning C. Data Reduction D. Data integration

30. Suppose two stocks A and B have the following values in one week (2, 15), (3, 18), (5, 10), (4, 11), (6, 14), If the stocks are affected by the same industry trends, will their prices ……..
A. rise together B. fall together C. stand together D. Data integration

31. Suppose S=[2,1,4,4], by using wavelet decomposition, the detail coefficients are:
A. [0, -1, -1, 0] B. [2.75 , -1.25 , 0.5 , 0] C. [0.5 , 0, 2.75 , -1.25] D.[-1,0,0,-1]

- **Given two objects represented by the tuples (22, 1, 42, 10) and (20, 0, 36, 8):**

32. Compute the Euclidean distance between the two objects.
A. 6 B. 6.7082 C. 11 D. 6.1534

33. Compute the Manhattan distance between the two objects.
A. 6 B. 6.7082 C. 11 D. 6.1534

34. Compute the Minkowski distance between the two objects, using q = 3.
A. 6 B. 6.7082 C. 11 D. 6.1534

35. Compute the supremum distance between the two objects.
A. 6 B. 6.7082 C. 11 D. 6.1534

- **Use these methods to normalize the following group of data: (200, 300 , 400, 600 , 1000)**

36. Min-Max normalization by setting min= 0 and max=1

A. (0 , 0.125, 0.25 , 0.5 , 1)
B. (0.25 , 0.5 , 1 , 0 , 0.125)
C. (-1.06 , -0.7, -0.35 , 0.35 , 1.78)
D. (-0.35 , 0.35 , 1.78, -1.06 , -0.7)

37. z-score normalization

A. (0 , 0.125, 0.25 , 0.5 , 1)
B. (0.25 , 0.5 , 1 , 0 , 0.125)
C. (-1.06 , -0.7, -0.35 , 0.35 , 1.78)
D. (-0.35 , 0.35 , 1.78, -1.06 , -0.7)

- **Suppose a group of 9 sales price records has been sorted as follows:**

28 , 25 , 15 , 21 , 8 , 21 , 24, 4, 34

38. Partition into equal-frequency (equi-depth) bins:

Bin 1	**4**	**X**	**15**
Bin 2	**21**	**21**	**24**
Bin 3	**25**	**28**	**34**

X= ……

A. 4 B. 8 C. 34 D. 22

39. Smoothing by bin means:

Bin 1	**9**	**9**	**9**
Bin 2	**X**	**X**	**X**
Bin 3	**29**	**29**	**29**

X= ……

A. 4 B. 8 C. 34 D. 22

40. Smoothing by bin boundaries:

Bin 1	**4**	**4**	**15**
Bin 2	**21**	**21**	**24**
Bin 3	**25**	**25**	**X**

X= ……

A. 4 B. 8 C. 34 D. 22

My best wishes

Dr. Osama Farouk