

OCT/NOV-2022

Total No. of Questions : 5]

SEAT No. :

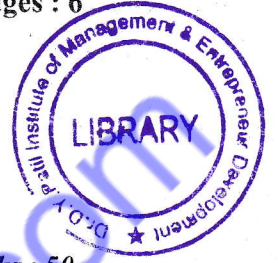
PA-2562

[Total No. of Pages : 6

[5948]-302

M.C.A. (Management)

IT - 32 : DATA WAREHOUSING AND DATA MINING  
(2020 Pattern) (Semester - III)



Time : 2½ Hours]

[Max. Marks : 50

Instructions to the candidates:

- 1) All questions are compulsory.
- 2) Draw neat & labelled diagrams wherever necessary.

Q1) Answer the following multiple-choice questions. [20×½=10]

- a) The functions, transfer of data from source to destination, recommending rules for extraction are performed by
  - i) ETL Tools
  - ii) Database tools
  - iii) Mining Tools
  - iv) All of the mentioned
- b) Text mining tasks involves \_\_\_\_\_
  - i) Text Categorization
  - ii) Text clustering
  - iii) DFD
  - iv) Both (i) and (ii)
- c) Web page contents used for mining of data are collected from \_\_\_\_\_
  - i) Log files
  - ii) Web page contents
  - iii) Images
  - iv) Audio
- d) Google PR checker, Lin viewer are tools for \_\_\_\_\_
  - i) Text mining
  - ii) Web content mining
  - iii) Web structure mining
  - iv) Web usage mining
- e) K in the K-means Algorithm stands for \_\_\_\_\_
  - i) Data set
  - ii) Number of clusters
  - iii) Error function
  - iv) Knowledge
- f) Agglomerative clustering falls under which type of clustering method?
  - i) Partition method
  - ii) Hierarchical method
  - iii) Both (i) and (ii)
  - iv) Neither (i) and (ii)
- g) In Bayes is Theorem Class conditional probability is called as \_\_\_\_\_
  - i) Evidence
  - ii) Likelihood
  - iii) Prior
  - iv) Posterior

P.T.O.

- h) K-means squared error function is related with which of the following?
- i) Manhattan distance
  - ii) Hamming distance
  - iii) Euclidean distance
  - iv) Minkowski distance
- i) \_\_\_\_\_ collects all the information and the subjects about an entire organization.
- i) Data Mart
  - ii) Virtual warehouse
  - iii) Data warehouse view
  - iv) enterprise warehouse
- j) A star schema has which types of relationship between dimensional and fact tables?
- i) Many to many
  - ii) One to one
  - iii) One to many
  - iv) All of the mentioned
- k) \_\_\_\_\_ maps the core warehouse metadata to business concepts, familiar and is useful to end users.
- i) Application level metadata
  - ii) User level metadata
  - iii) End user level metadata
  - iv) Core level metadata
- l) Which of the following is not a component of a data warehouse?
- i) Metadata
  - ii) Current detailed data
  - iii) Lightly summarized data
  - iv) Component key
- m) Which is NOT a basic conceptual schema in Data warehouses?
- i) Star schema
  - ii) Tree schema
  - iii) Snowflake Schema
  - iv) Fact constellation schema
- n) The Extract process is \_\_\_\_\_
- i) Capturing all of the data contained in various operational system
  - ii) Capturing sub set of the data contained in various operational system
  - iii) Capturing all of the data contained in various decision support systems
  - iv) Capturing of sub set of the data contained in various decision support systems.



- o) An array in which data is stored are characterized by multiple dimension is a \_\_\_\_\_
- i) Table ii) Cube  
iii) Schema iv) Collection
- p) Which of the following is NOT an advantage of ROLAP?
- i) High data efficiency ii) Scalability  
iii) High utilization of resources iv) Flexible
- q) Data environment of a \_\_\_\_\_ is in the 3<sup>rd</sup> normal form
- i) OLAP ii) OLTP  
iii) Data warehouse iv) both (i) & (ii)
- r) \_\_\_\_\_ operation gets data from coarser granularity to fine granularity
- i) Roll up ii) Dice  
iii) Pivot iv) Drill Down
- s) Data mining contributes to \_\_\_\_\_
- i) Data ware house ii) Data stores  
iii) Knowledge base iv) Pattern finding
- t) Data mining can be performed on \_\_\_\_\_
- i) Spatial database ii) Temporal data  
iii) Text database iv) All of the mentioned

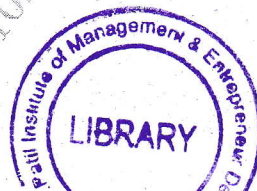
Q2) a) Discuss the schemas in Data warehousing with the help of employee database example. [5]

b) Give 5 differences between transactional data and warehouse data based on their characteristics. [5]

OR

a) Explain the architecture of a Data warehouse with a neat diagram. [5]

b) Name the different OLAP architectures. Pick any two (2) and describe in detail with advantage. [5]



- Q3) a)** What are Discretization and concept Hierarchy generation process? Give an example for each. [5]
- b) Explain the tools used for data warehouse development. [5]

OR

- a) Explain the different data sources for data warehouse and methods of data collection. [5]
- b) Explain the different steps of creating on OLAP, with examples. [5]

- Q4) a)** Consider the data set given below, compute the support for item sets {e}, {b, d} and {b, d, e} [5]

customer ID	Transaction ID	Items Brought
1	0001	{a, d, e}
1		
1	0024	{a, b, c, e}
2	0012	{a, b, d, e}
2	0031	{a, c, d, e}
3	0015	{b, c, e}
3	0022	{b, d, e}
4	0029	{c, d}
4	0040	{a, b, c}
5	0033	{a, d, e}
5	0038	{a, b, e}

- b) Using the result from problem a, above, compute the confidence for the association rules {b, d} → {e} and {e} → {b, d} [5]

OR



- a) A consultancy wants to categories MCA students into classes as Excellent, Good, and Average. The data collected from students are their average percentage in MCA- I year and result of the aptitude test conducted by the consultancy.

Solve the problem using decision tree Algorithm.

[5]

- b) Using Bayes an classification to classify the sample data: {6, 43}. As male or female. Training data is given.

[5]

Person	Height	Weight
Male	6.2	82
Male	5.11	65
Male	5.7	58
Male	5.11	55
Female	4.10	42
Female	5.5	50
Female	5.0	43
Female	5.75	50

- Q5) a) Construct a FP-Tree Algorithm, to find frequency patterns for the given data.

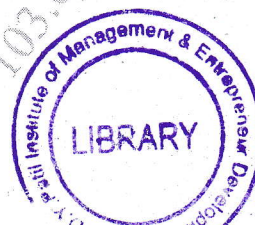
[5]

Transaction ID	Item Bought
100	{f, a, c, d, g, i, m, p}
200	{a, b, c, f, l, m, o}
300	{b, f, h, j, o}
400	{b, c, k, s, p}
500	{a, f, c, e, l, p, m, n}

- b) Explain Hierarchical clustering using examples.

[5]

OR



- a) Perform K. means clustering and show all the calculations at each iteration, to form the final cluster. Assume the initial clusters are A, E and H. [5]

Points	X1	X2
A	3	3
B	8	5
C	4	4
D	2	4
E	7	7
F	5	8
G	3	5
H	4	8
I	6	9
J	9	6

- b) What are agent based and database based approaches in web mining? Explain with example. [5]

