

Total No. of Questions : 5]

SEAT No. : 

PB838

[Total No. of Pages : 3

[6204]-32

S.Y. M.C.A. (Management)

IT-32 : DATA WAREHOUSING AND DATA MINING

(Revised 2020 Pattern) (Semester-III)



Time : 2½ Hours]

[Max. Marks : 50

Instructions to the candidates:

- 1) All questions are compulsory.
- 2) Draw neat labeled diagram wherever necessary.

- Q1)** a) Discuss advantages & applications of data warehouse. [5]  
 b) Define Schema. Describe fact constellation Schema with suitable example. [5]

OR

- c) What is OLAP? Explain ROLAP & MOLAP. [5]  
 d) Differentiate between data warehouse and OLTP systems. [5]

- Q2)** a) Explain the need for data transformation & the steps involved in data transformation during ETL process. [5]  
 b) Describe dimensionality reduction in data pre-processing. [5]

OR

- c) Write types of attributes used in data pre-processing task & explain properties of attribute values. [5]  
 d) Explain discretization & concept hierarchy generation steps in ETL. [5]

- Q3)** a) What is KDD? Explain how KDD is used in data mining. [5]  
 b) Define text mining. Discuss types & steps used in text mining. [5]

OR

- c) Write a note on any two data mining tools. [5]  
 d) Discuss web usage mining with its applications. [5]

P.T.O.

Q4) a) Consider the following data set & draw FP tree for minimum support=3. [5]

T.ID	Items
1	ABDE
2	ABCE
3	BCE
4	ABDE
5	BCD
6	ABCDE

b) Consider the same data set in Q.4. 9 & calculate the support & confidence for following rules. [5]

- i)  $\{A,B\} \rightarrow \{C,D,E\}$
- ii)  $\{C, D\} \rightarrow \{B\}$
- iii)  $\{B,C\} \rightarrow \{E\}$
- iv)  $\{A,B\} \rightarrow \{C\}$
- v)  $\{A, E\} \rightarrow \{B\}$

OR

c) Apply Naive Baye's Classifier on Below dataset for the instance=[Plant-Type="Tree", Flowering="No", Origin="east"] [5]

ID	Plant type	Flowering	Origin	Medicinal-prop
1	Tree	Yes	East	Yes
2	Tree	Yes	East	No
3	Tree	Yes	East	Yes
4	Herb	Yes	East	No
5	Herb	Yes	west	Yes
6	Herb	No	west	No
7	Herb	No	west	Yes
8	Herb	No	East	No
9	Tree	No	west	No
10	Tree	Yes	East	Yes



d) Write the algorithm for decision tree. Generate decision tree for classification of credit card approval or rejection based on the attributes [5]

- i) Age
- ii) Income-Group
- iii) CIB12-Score

Q5) a) Let  $s$  be a dataset containing speed in km/hr.  
 $S = \{15, 18, 25, 30, 32, 35, 22, 24, 28, 30, 40, 48, 42, 34, 36, 45, 12, 20, 22, 46\}$  Apply K-Means clustering for  $K=2$  on the dataset [5]

b) Write a note on Agglomerative clustering. [5]

OR

c) Give the dataset "w" containing weight of persons,  $W = \{15, 20, 30, 35, 40, 42, 43, 18, 60, 25, 28, 32, 45, 48, 50, \dots\}$ . Apply K-means algorithm for  $K=2$ . [5]

d) Draw and explain Kimball's life cycle diagram. [5]

