Comp
EXTRA

K. J. Somaiya Institute of Technology, Sion, Mumbai-22
(Autonomous College Affiliated to University of Mumbai)

Nov – Dec 2023
B.Tech Program: Computer Engineering Scheme: II
Examination: TY Semester: V
Course Code: CEC504 and Course Name: Data Warehousing and Mining
Date of Exam: 07/12/2023 Duration: 2.5 Hours Max. Marks: 60

Instructions:
(1)All questions are compulsory.
(2)Draw neat diagrams wherever applicable.
(3)Assume suitable data, if necessary.

		Max. Marks	CO	BT level
Q1	**Solve any six questions out of eight:**	**12**		
i)	Draw and explain bottom-up approach to build data warehouse with its advantages.	02	CO1	U
ii)	List and explain any two methods used for dealing with missing values in Data Cleaning process.	02	CO2	U
iii)	How FP growth tree overcome the drawback of Apriori algorithm. Justify?	02	CO3, CO4	An
iv)	Difference between Classification and Clustering in data mining(4 points)	02	CO3, CO4	U
v)	Explain with example concept of Market Basket Analysis	02	CO3, CO4	U
vi)	What is meant by Web Structure Mining?	02	CO5	U
vii)	Comparison between OLTP and OLAP (4 points)	02	CO1	U
viii)	List and explain any four applications of data mining.	02	CO2	U
Q.2	**Solve any four questions out of six.**	**16**		
i)	Suppose that a data warehouse consists of the four dimensions date, spectator, location, and game, and the two measures count and charge, where charge is the fare that a spectator pays when watching a game on a given date. Spectators may be students, adults, or seniors, with each category having its own charge rate. (a) Draw a snowflake schema diagram for the data warehouse in above scenario.	04	CO1	Ap
ii)	(i) Explain Normalization method used in Data Transformation process(02M) (ii) Suppose income range from \$10,000 to \$95,000 is normalized to [0.0,1.0]. Using min-max normalization, specify normalized value for income of \$64,300? (02M)	04	CO2	Ap
iii)	Explain in detail any two method used to evaluate accuracy of a classifier	04	CO3, CO4	U

K. J. Somaiya Institute of Technology, Sion, Mumbai-22
(Autonomous College Affiliated to University of Mumbai)

Nov – Dec 2023
B.Tech Program: Computer Engineering Scheme: II
Examination: TY Semester: V
Course Code: CEC504 and Course Name: Data Warehousing and Mining
Date of Exam: 07/12/2023 Duration: 2.5 Hours Max. Marks: 60

iv)	Define k-mean clustering algorithm. (01M) Use k-means algorithm to create 3-clusters for given set of values:{2,3,6,8,9,12,15,18,22} with initial mean values as m_1=6, m_2=12 and m_3=22 (03M)	04	CO3, CO4	Ap
v)	Explain in brief Multilevel association rule with suitable example.	04	CO3,CO4	U
vi)	Compare HITS and PageRank Algorithm used for web structure mining (08 points)	04	CO5	U
Q.3	**Solve any two questions out of three.**	**16**		
i)	Explain ETL process of Data Warehousing in detail.	08	CO1	U
ii)	Using the following training dataset, create classification model using Decision Tree and draw the tree upto second iteration.	08	CO3,CO4	Ap

Tid	Income	Age	Own House
1	Very High	Young	Yes
2	High	Medium	Yes
3	Low	Young	Rented
4	High	Medium	Yes
5	Very High	Medium	Yes
6	Medium	Young	Yes
7	High	Old	Yes
8	Medium	Medium	Rented
9	Low	Medium	Rented
10	Low	Old	Rented
11	High	Young	Yes
12	Medium	Old	Rented

K. J. Somaiya Institute of Technology, Sion, Mumbai-22
(Autonomous College Affiliated to University of Mumbai)

Nov – Dec 2023

B.Tech Program: Computer Engineering — Scheme: II

Examination: TY — Semester: V

Course Code: CEC504 and Course Name: Data Warehousing and Mining

Date of Exam: 07/12/2023 — Duration: 2.5 Hours — Max. Marks: 60

iii)	Show the dendrogram created by complete linkage hierarchical clustering algorithm for given set of data points **Data Points** | **X** | **Y** P1 | 2 | 4 P2 | 8 | 2 P3 | 9 | 3 P4 | 1 | 5 P5 | 8.5 | 1	08	CO3, CO4	Ap
Q.4	**Solve any two questions out of three.**	**16**		
i)	Draw and elaborate steps involved in Knowledge Discovery in Database(KDD) process of Data Mining.	08	CO2	U
ii)	Generate Frequent Pattern Tree for following transaction with 30% minimum support **Transaction ID** | **Items** T1 | E, A, D, B T2 | D, A, C, E, B T3 | C, A, B, E T4 | B, A, D T5 | D T6 | D,B T7 | A, D, E T8 | B, C	08	CO3, CO4	Ap
iii)	Let the adjacency matrix for a graph of four vertices {n1 to n4} is given as below: A = [0 1 1 1 0 0 1 1 1 0 0 1 0 0 0 1] Calculate the authority and hub scores for this graph using HITS algorithm with k = 3(iterations) and identify best authority and hub nodes	08	CO5	Ap

****