

2024/25
K. J. Somaiya Institute of Technology, Sion, Mumbai-22
(Autonomous College Affiliated to University of Mumbai)

Nov – Dec 2024

(B. Tech Program: Computer Engineering B. Tech Scheme:II)
Regular Examination TY Semester: V

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

Date of Exam: 2nd /12/2024

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.

Q. No.	Question	Max. Marks	CO	BT level
Q 1	Solve any two questions out of three: (05 marks each)	10		
a)	Differentiate between OLTP and OLAP. (Min. 5 points)		CO1	U
b)	For the following data for the attribute age: 13, 15, 16, 16, 19, 20, 23, 29, 33, 41, 44, 53, 62, 69, 72 i) Use Min-Max Normalization to transform the values: 45 onto the range [0.0,1.0] and ii) Z-Score Normalization to transform: 32 & 45 where the standard deviation of age is 20.64 years.		CO2	AP
c)	Explain different accuracy and error measures in classification.		CO3 CO4	U
Q 2	Solve any two questions out of three: (05 marks each)	10		
a)	Explain k-means clustering algorithm in brief.		CO3 CO4	U
b)	Illustrate the working of market basket analysis with an example.		CO3 CO4	U
c)	Explain different types of attributes with examples.		CO2	U
Q.3	Solve any two questions out of three. (10 marks each)	20		
a)	Design a Star Schema and Snowflake Schema for a Supermarket Chain, with dimensions, namely Product, store, time, promotion and three measures unit sales, dollars_sales and dollar_cost.		CO1	AP
b)	Suppose that the data for analysis includes the attribute salary (in thousands of dollars), shown in increasing order: 30, 36, 47, 50, 52, 52, 56, 60, 63, 70, 70, 110. i. What are the mean, median, mode and midrange of the data? ii. Find the first quartile(Q1) & the third quartile(Q3) of the data.		CO2	AP

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

Nov – Dec 2024

(B. Tech Program: Computer Engineering B. Tech Scheme:II)

Regular Examination TY Semester: V

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

Date of Exam: 2 /12/2024

Duration: 2.5 Hours

Max. Marks: 60

iii. Calculate the standard deviation. iv. Show the boxplot of the data.																																																																																														
c)	Apply the Naïve Bayes classification algorithm and classify the tuple: X= (Age = Young, Income = Medium, Student = Yes, Credit rating = fair) <table><tr><th>R/D</th><th>age</th><th>income</th><th>student</th><th>credit_rating</th><th>Class: buys_computer</th></tr><tr><td>1</td><td>youth</td><td>high</td><td>no</td><td>fair</td><td>no</td></tr><tr><td>2</td><td>youth</td><td>high</td><td>no</td><td>excellent</td><td>no</td></tr><tr><td>3</td><td>middle_aged</td><td>high</td><td>no</td><td>fair</td><td>yes</td></tr><tr><td>4</td><td>senior</td><td>medium</td><td>no</td><td>fair</td><td>yes</td></tr><tr><td>5</td><td>senior</td><td>low</td><td>yes</td><td>fair</td><td>yes</td></tr><tr><td>6</td><td>senior</td><td>low</td><td>yes</td><td>excellent</td><td>no</td></tr><tr><td>7</td><td>middle_aged</td><td>low</td><td>yes</td><td>excellent</td><td>yes</td></tr><tr><td>8</td><td>youth</td><td>medium</td><td>no</td><td>fair</td><td>no</td></tr><tr><td>9</td><td>youth</td><td>low</td><td>yes</td><td>fair</td><td>yes</td></tr><tr><td>10</td><td>senior</td><td>medium</td><td>yes</td><td>fair</td><td>yes</td></tr><tr><td>11</td><td>youth</td><td>medium</td><td>yes</td><td>excellent</td><td>yes</td></tr><tr><td>12</td><td>middle_aged</td><td>medium</td><td>no</td><td>excellent</td><td>yes</td></tr><tr><td>13</td><td>middle_aged</td><td>high</td><td>yes</td><td>fair</td><td>yes</td></tr><tr><td>14</td><td>senior</td><td>medium</td><td>no</td><td>excellent</td><td>no</td></tr></table>	R/D	age	income	student	credit_rating	Class: buys_computer	1	youth	high	no	fair	no	2	youth	high	no	excellent	no	3	middle_aged	high	no	fair	yes	4	senior	medium	no	fair	yes	5	senior	low	yes	fair	yes	6	senior	low	yes	excellent	no	7	middle_aged	low	yes	excellent	yes	8	youth	medium	no	fair	no	9	youth	low	yes	fair	yes	10	senior	medium	yes	fair	yes	11	youth	medium	yes	excellent	yes	12	middle_aged	medium	no	excellent	yes	13	middle_aged	high	yes	fair	yes	14	senior	medium	no	excellent	no		CO3 CO4	AP
R/D	age	income	student	credit_rating	Class: buys_computer																																																																																									
1	youth	high	no	fair	no																																																																																									
2	youth	high	no	excellent	no																																																																																									
3	middle_aged	high	no	fair	yes																																																																																									
4	senior	medium	no	fair	yes																																																																																									
5	senior	low	yes	fair	yes																																																																																									
6	senior	low	yes	excellent	no																																																																																									
7	middle_aged	low	yes	excellent	yes																																																																																									
8	youth	medium	no	fair	no																																																																																									
9	youth	low	yes	fair	yes																																																																																									
10	senior	medium	yes	fair	yes																																																																																									
11	youth	medium	yes	excellent	yes																																																																																									
12	middle_aged	medium	no	excellent	yes																																																																																									
13	middle_aged	high	yes	fair	yes																																																																																									
14	senior	medium	no	excellent	no																																																																																									
Q.4	Solve any two questions out of three. (10 marks each)	20																																																																																												
a)	Apply the k-Means clustering algorithm and find the clusters for the given data. If we have several objects (4 types of Medicines) and each object have 2 attributes of features as shown in the table. Group these two objects into k=2 groups of medicine based on 2-features (PH & Weight Index). <table><tr><th>Object</th><th>X(Wt index)</th><th>Y(Ph)</th></tr><tr><td>Medicine A</td><td>1</td><td>1</td></tr><tr><td>Medicine B</td><td>2</td><td>1</td></tr><tr><td>Medicine C</td><td>4</td><td>3</td></tr><tr><td>Medicine D</td><td>5</td><td>4</td></tr></table>	Object	X(Wt index)	Y(Ph)	Medicine A	1	1	Medicine B	2	1	Medicine C	4	3	Medicine D	5	4		CO3 CO4	AP																																																																											
Object	X(Wt index)	Y(Ph)																																																																																												
Medicine A	1	1																																																																																												
Medicine B	2	1																																																																																												
Medicine C	4	3																																																																																												
Medicine D	5	4																																																																																												

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

Nov – Dec 2024

(B. Tech Program: Computer Engineering B. Tech Scheme:II)

Regular Examination TY Semester: V

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

Date of Exam: 2. /12/2024

Duration: 2.5 Hours

Max. Marks: 60

b)	<p>Apply the Apriori algorithm to find frequent patterns and strong association rules for the following database with min. support=50% and min. confidence 66%.</p> <table><tr><th>Tid</th><th>Items</th></tr><tr><td>10</td><td>A, C, D</td></tr><tr><td>20</td><td>B, C, E</td></tr><tr><td>30</td><td>A, B, C, E</td></tr><tr><td>40</td><td>B, E</td></tr></table>	Tid	Items	10	A, C, D	20	B, C, E	30	A, B, C, E	40	B, E	CO3 CO4	AP
Tid	Items												
10	A, C, D												
20	B, C, E												
30	A, B, C, E												
40	B, E												
c)	<p>The adjacency matrix for a graph of four vertices {n1 to n4} is given as follows:</p> $A = \begin{bmatrix} 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix}$ <p>Calculate the authority and hub scores for this graph using the HITS algorithm with $k = 3$ and identify best authority and hub nodes.</p>	CO5	AP										
