Mahavir Education Trust's

SHAH & ANCHOR KUTCHHI ENGINEERING COLLEGE

Chembur, Mumbai - 400 088

UG Program in Computer Engineering

Course Code	CSC603	**Course Name**	Data Warehousing and Mining
Academic Year	2020-2021	**Semester**	VI
Class	TE-3,TE-4	**Course Incharge**	Ms. Pinki Vishwakarma

Course Outcomes (CO)

CO No.	**CO Statement** (At the end of the course, students will be able to …)	BL
1	Understand Data Warehouse fundamentals, Data Mining Principles	2
2	Design data warehouse with dimensional modelling and apply OLAP operations.	3
3	Identify appropriate data mining algorithms to solve real world problems	2
4	Compare and evaluate different data mining techniques like classification, prediction, clustering and association rule mining	4
5	Describe complex data types with respect to spatial and web mining.	2
6	Benefit the user experiences towards research and innovation.	4

Assignment-1

Date:23/02/2021 **Submission Date:3/03/2021**

Q. No.	Question	Marks	BL	CO	PSO	PI
1	In real-world data, tuples with missing values for some attributes are a common occurrence. Describe various methods for handling this problem.	5	2	CO1	1,2	1.4.1,2.1.2,2.2.2, 2.1.3,
2	Suppose that a data warehouse consists of three dimensions time, doctor and patient, and the two measures count and charge where charge is the fee doctor charges a	5	3	CO2	1,2	2.2.2,2.2.3, 10.1.1,10.1.2

<table>
<tr><td></td><td>patient for visit. Draw a star schema diagram for the above data warehouse.</td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td>3</td><td>The following table consists of training data from an employee database. The data have been generalized. For example, “31…35” for age represents the age range of 31 to 35. For a given row entry, count represents the number of data tuples having the values for department, status, age and salary given in that row.
<table>
<tr><th>dept</th><th>status</th><th>age</th><th>salary</th><th>count</th></tr>
<tr><td>sales</td><td>senior</td><td>31…35</td><td>46K50K</td><td>30</td></tr>
<tr><td>sales</td><td>junior</td><td>26…30</td><td>26K…30K</td><td>40</td></tr>
<tr><td>sales</td><td>junior</td><td>31…35</td><td>31K…35K</td><td>40</td></tr>
<tr><td>systems</td><td>junior</td><td>21…25</td><td>46K…50K</td><td>20</td></tr>
<tr><td>systems</td><td>senior</td><td>31…35</td><td>66K…70K</td><td>5</td></tr>
<tr><td>systems</td><td>junior</td><td>26…30</td><td>46K…50K</td><td>3</td></tr>
<tr><td>systems</td><td>senior</td><td>41…45</td><td>66K…70K</td><td>3</td></tr>
<tr><td>marketing</td><td>senior</td><td>36…40</td><td>46K…50K</td><td>10</td></tr>
</table></td><td>5</td><td>2</td><td>CO3</td><td>1,2</td><td>1.4.1,2.1.2,
2.1.3,2.2.3,2.3.1, 4.2.1</td></tr>
</table>

	<table><tr><td>marketing</td><td>junior</td><td>31 … 35</td><td>41K…45K</td><td>4</td></tr><tr><td>secretary</td><td>senior</td><td>46 … 50</td><td>36K…40K</td><td>4</td></tr><tr><td>secretary</td><td>junior</td><td>26 … 30</td><td>26K…30K</td><td>6</td></tr></table> Let status be the class label attribute. a) Given a data tuple having the values "systems," "26 . . . 30," and "46–50K" for the attributes department, age, and salary, respectively, what would a naive Bayesian classification of the status for the tuple be?							
4	Show the dendrogram using complete link hierarchical clustering algorithm. The table below comprises sample data items indicating. the distance between the elements (adjacency matrix). <table><tr><td>Item</td><td>E</td><td>A</td><td>C</td><td>B</td><td>D</td></tr><tr><td>E</td><td>0</td><td>1</td><td>2</td><td>2</td><td>3</td></tr><tr><td>A</td><td>1</td><td>0</td><td>2</td><td>5</td><td>3</td></tr><tr><td>C</td><td>2</td><td>2</td><td>0</td><td>1</td><td>6</td></tr><tr><td>B</td><td>2</td><td>5</td><td>1</td><td>0</td><td>3</td></tr><tr><td>D</td><td>3</td><td>3</td><td>6</td><td>3</td><td>0</td></tr></table>	5	2	CO3	1,2	1.4.1,2.1.2 2.1.3,2.2.3,2.3.1, 4.2.1

Name: **Signature:**

Date: