Name: **Enrolment No:**	

UPES

End Semester Examination, December 2023

Course: DATA MINING AND BUSINESS INTELLIGENCE **Semester : 1**

Program: M. Tech (CSE) **Time : 03 hrs.**

Course Code: CSDA 7001 **Max. Marks: 100**

Instructions:

SECTION A
(5Qx4M=20Marks)

S. No.		Marks	CO
Q 1	What is the apriori property? Write down the disadvantages of the apriori algorithm.	**2+2=4**	**CO3**
Q 2	Explain with example, what is content description. What are the steps of content description? Explain.	**2+2=4**	**CO2**
Q 3	Why is data normalization done? Let the income range $10,000 to $1,70,000 be normalized to [0.0, 1.0]. What value will $97,000 be mapped to? (Show steps).	**2+2=4**	**CO2**
Q 4	With example, explain incomplete, noisy, intentionally dirty, and inconsistent data.	**4**	**CO2**
Q 5	What is data integration? Why is data integration done? Provide examples.	**1+3=4**	**CO2**
SECTION B **(4Qx10M= 40 Marks)**			
Q 6	Explain the nominal, numeric, and ordinal data discretization. Why do we need dimensionality reduction? Explain linear regression with example.	**4+2+4=10**	**CO2**
Q 7	What is an FP-tree? What are the benefits of using FP-Growth algorithm over apriori algorithm? Explain the steps of FP-Growth algorithm.	**2+2+6=10**	**CO3**

<table>
<tr><td>Q 8</td><td>Define linear and non-linear regression using figures. Calculate the value of Y for X=100 using the linear regression prediction method. 4+6

<table><tr><th>X</th><th>Y</th></tr><tr><td>4</td><td>390</td></tr><tr><td>9</td><td>580</td></tr><tr><td>10</td><td>650</td></tr><tr><td>14</td><td>730</td></tr><tr><td>4</td><td>410</td></tr><tr><td>7</td><td>530</td></tr><tr><td>12</td><td>600</td></tr><tr><td>22</td><td>790</td></tr><tr><td>1</td><td>350</td></tr><tr><td>3</td><td>400</td></tr><tr><td>8</td><td>590</td></tr><tr><td>11</td><td>640</td></tr></table>
<table><tr><td>5</td><td>450</td></tr><tr><td>6</td><td>520</td></tr><tr><td>10</td><td>690</td></tr><tr><td>11</td><td>690</td></tr><tr><td>16</td><td>770</td></tr><tr><td>13</td><td>700</td></tr><tr><td>13</td><td>730</td></tr><tr><td>10</td><td>640</td></tr></table></td><td>4+6=10</td><td>CO4</td></tr>
<tr><td>Q 9</td><td>Why is knowledge representation an integral part of business intelligence? Explain with examples four different methods of knowledge representation. 2+8=10

OR

Enlist data preprocessing steps. Explain any one of the preprocessing steps in detail. 4+6=10</td><td>10</td><td>CO4</td></tr>
<tr><td colspan="3">SECTION-C
(2Qx20M=40 Marks)</td></tr>
<tr><td>Q 10</td><td>What is OLAP? What is the advantage of using OLAP over DBMS? Explain with figures different OLAP operations. List the differences between ROLAP, MOLAP, and HOLAP over technology, data volume, memory, and response time.</td><td>1+1+10+8 = 20</td><td>CO1</td></tr>
<tr><td>Q 11</td><td>Consider the list of transactions. Let min_sup=30% and min_confidence=60%. Find out all frequent items using apriori algorithm. List ‘strong’ association rules. Write a note on incremental ARM. What are the metrics of associated rule learning? 10+4+3+3=20</td><td>20</td><td>CO4</td></tr>
</table>

TID	items bought
T1	pen, pencil
T2	book, eraser, pencil
T3	book, chalk, eraser, pen
T4	chalk, eraser, pen
T5	book, pen, pencil
T6	book, eraser, pen, pencil
T7	ink, pen
T8	book, pen, pencil
T9	eraser, pen, pencil
T10	book, chalk, pencil

OR

What is classification and prediction? List out issues regarding classification and prediction. Describe four different measures of predictive accuracy. Briefly describe the life cycle of Data analytics, and role of data scientists. 2+3+8+7=20.