Name…… Roll No……..

ST. JOSEPH'S COLLEGE OF ENGINEERING AND TECHNOLOGY, PALAI.

(An ISO 9001:2015 Certified College)

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

B. TECH DEGREE SECOND INTERNAL TEST- JUNE 2021

EIGHTH SEMESTER -Batch A & B

CS 402 DATA MINING AND WAREHOUSING

Time: 2 Hours Max. Marks: 50

PART A

Answer all questions

Qn No.	Questions	Marks	K level	CO & KL
1	a) Discriminate the conflict resolution strategy in rule-based classification.	2	K1	CO3&K3
	b) How does backpropagation algorithm work?	3	K1	CO3&K3
2	The following table shows the midterm and final exam grades obtained for students in a database course. *x* and *y* have a linear relationship. *x* Midterm exam / *y* Final exam 72 84 50 63 81 77 74 78 94 90 86 75 59 49 83 79 65 77 33 52 88 74 81 90			
	a) Use the *method of least squares* to find an equation for the prediction of a student's final exam grade based on the student's midterm grade in the course.	3	K3	CO3&K3
	b) Predict the final exam grade of a student who received an 86 on the midterm exam.	2	K3	CO3&K3
3	The data mining task is to cluster the following eight points (with (*x*, *y*) representing location) into three clusters: *A*1(2, 10), *A*2(2, 5), *A*3(8, 4), *B*1(5, 8), *B*2(7, 5), *B*3(6, 4), *C*1(1, 2), *C*2(4, 9) The distance function is Euclidean distance. Initially we assign *A*1, *B*1, and *C*1 as the center of each cluster, respectively. Use the *k-means* algorithm to show			
	a) The three cluster centers after the first-round execution.	2.5	K2	CO6&K2
	b) The final three clusters.	2.5	K2	CO6&K2
4	a) What are the characteristics of social networks?	2	K1	CO7&K2
	b) We have 4 points, (3,5) (2,3) (4,3) & (1,5) in a cluster C1. Calculate the Cluster Feature (CF).	3	K2	CO6&K2

Name…… Roll No……..

PART B

Answer all questions

<table>
<tr><th>Qn No.</th><th>Questions</th><th>Marks</th><th>K level</th><th>CO & KL</th></tr>
<tr><td rowspan="5">5</td><td>a) Compare eager and lazy classification.</td><td>4</td><td>K2</td><td>CO3&K3</td></tr>
<tr><td>b) A confusion matrix for the classes buys computer = yes and buys computer = no has given.

<table><tr><th>Classes</th><th>buys_computer = yes</th><th>buys_computer = no</th><th>Total</th><th>Recognition (%)</th></tr><tr><td>buys_computer = yes</td><td>6,954</td><td>46</td><td>7,000</td><td>99.34</td></tr><tr><td>buys_computer = no</td><td>412</td><td>2,588</td><td>3,000</td><td>86.27</td></tr><tr><td>Total</td><td>7,366</td><td>2,634</td><td>10,000</td><td>95.52</td></tr></table>
Calculate</td><td></td><td></td><td></td></tr>
<tr><td>a) Accuracy</td><td>2</td><td>K3</td><td>CO4&K3</td></tr>
<tr><td>b) Precision &</td><td>2</td><td>K3</td><td>CO4&K3</td></tr>
<tr><td>c) Recall</td><td>2</td><td>K3</td><td>CO4&K3</td></tr>
<tr><td>6</td><td>Consider the transaction database given below. Which consists of 5 items such as-Asparagus (A), Corn (C), Beans (B), Tomatoes (T) & Squash (S)

Set minimum support count as 2 and minimum confidence threshold as 70%. Find all frequent patterns using FP-growth algorithm.

<table><tr><th>Transaction ID</th><th>List of items in the transaction</th></tr><tr><td>T1</td><td>B , A , T</td></tr><tr><td>T2</td><td>A , C</td></tr><tr><td>T3</td><td>A , S</td></tr><tr><td>T4</td><td>B , A , C</td></tr><tr><td>T5</td><td>B , S</td></tr><tr><td>T6</td><td>A , S</td></tr><tr><td>T7</td><td>B , S</td></tr><tr><td>T8</td><td>B , A , S , T</td></tr><tr><td>T9</td><td>B , A , S</td></tr></table></td><td>10</td><td>K3</td><td>CO5&K3</td></tr>
<tr><td rowspan="2">7</td><td>a) Differentiate between Agglomerative and Divisive hierarchical clustering method with an example.</td><td>5</td><td>K2</td><td>CO6&K2</td></tr>
<tr><td>b) Explain Apriori based frequent subgraph mining.</td><td>5</td><td>K2</td><td>CO7&K2</td></tr>
</table>