2018 Scheme

1	N	T							

NITTE MEENAKSHI INSTITUTE OF TECHNOLOGY

An Autonomous Institution Approved by UGC/AICTE/Govt. of Karnataka
Accredited by NBA (Tier – I) and NAAC 'A+' Grade
Affiliated to Visveswaraya Technological University, Belagavi
Post Box No. 6429, Yelahanka, Bengaluru – 560 064, Karnataka, India

KNOWLEDGE • CHARACTER • UNITY

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

MID SEMESTER EXAMINATION-III

Course Title with code	Data Mining, 18CS54	**Maximum Marks**	**30 Marks**
Date and Time	24/01/2022, 9.30am to 10.30am	**No. of Hours**	1.0
Course Instructor(s)	Dr. Vijaya Shetty S, Dr. Sujata Joshi, Dr. Vani V		

Instructions to Students

1. Answer any **two full questions.**
2. Any missing data may assume suitably.

Q. No	Question	MAX MARKS	CO	BL	PO/ PSO
1. a	Draw a contingency table for the rules 1. {b} → {c}, 2. {a} → {d}, 3. {b} → {d} using the transactions shown in Table 1a. Table 1a. Market basket transactions. Transaction ID | Items Bought 1 | {a, b, d, e} 2 | {b, c, d} 3 | {a, b, d, e} 4 | {a, c, d, e} 5 | {b, c, d, e} 6 | {b, d, e} 7 | {c, d} 8 | {a, b, c} 9 | {a, d, e} 10 | {b, d}	6	4	3	1,2,3 /2
1. b	Use the contingency tables in 1.a to compute and rank the rules in decreasing order according to the following measures. i)Support. ii). Confidence. iii) Interest (X →Y) iv) IS (X → Y)	9	4	3	1,2,3 /2
2. a	Compare partitioned clustering and hierarchical clustering.	5	5	2	1,2,3 /2
2. b	Consider the dataset given in the Table2.b with five objects characterized by a single continuous feature. Table2.b Dataset | | a | b | c | d | e | Feature | 1 | 2 | 4 | 5 | 6 | Apply the agglomerative algorithm with single-link cluster distance measures to produce a dendrogram tree. Note : dist(c1,c2) = abs(c1-c2)	10	5	3	1,2, 3/2

<table>
<tr>
<td>3. a</td>
<td>Construct FP tree for the Table 3a.
Table 3a. Market basket transactions
<table>
<tr><th>Transaction ID</th><th>Items Bought</th></tr>
<tr><td>1</td><td>$\{a, b, d, e\}$</td></tr>
<tr><td>2</td><td>$\{b, c, d\}$</td></tr>
<tr><td>3</td><td>$\{a, b, d, e\}$</td></tr>
<tr><td>4</td><td>$\{a, c, d, e\}$</td></tr>
<tr><td>5</td><td>$\{b, c, d, e\}$</td></tr>
<tr><td>6</td><td>$\{b, d, e\}$</td></tr>
<tr><td>7</td><td>$\{c, d\}$</td></tr>
<tr><td>8</td><td>$\{a, b, c\}$</td></tr>
<tr><td>9</td><td>$\{a, d, e\}$</td></tr>
<tr><td>10</td><td>$\{b, d\}$</td></tr>
</table>
</td>
<td>5</td><td>4</td><td>3</td><td>1,2, 3/2</td>
</tr>
<tr>
<td>3. b</td>
<td>Suppose that the data mining task is to cluster points (with (x, y) representing location) into three clusters, where the points are
A1(2, 10), A2(2, 5), A3(8, 4), A4(5, 8), A5(7, 5), A6(6, 4), A7(1, 2), A8(4, 9).
The distance function is Manhattan distance. Suppose initially we assign A1, A4, and A7 as the center of each cluster, respectively. Use the k-means algorithm to show
(a) the three cluster centers after the first round of execution.
(b) the final three clusters once the algorithm converged.</td>
<td>10</td><td>5</td><td>3</td><td>1,2, 3/2</td>
</tr>
</table>

Faculty Signature	Course Co-Ordinator/Mentor Signature	HoD Signature