Institute of Science and Technology
Bachelor of Science in Computer Science & Information Technology
Model Question

Course Title: Data Warehousing and Data Mining   **Time:** 3 hours
Full Marks: 60   **Pass Marks:** 24
Course Code: CSC410   **Semester:** VII

Group 'A'

Attempt any TWO Questions. (2 × 10 = 20)

1. Explain the different components of data warehouse. How data cube precomputation is performed? Describe. (5 + 5)
2. Write the limitation of Apriori algorithm. Given the objects P1(2,3), P2(4,5), P3(10,40), P4(60,55), P5(70,80), apply K-means algorithm (K = 2) to show the final clusters after 2 iterations. Assume P1 and P3 as initial cluster centroids. (2 + 8)
3. Consider the following training data set. (10)

Tid	Attrib1	Attrib2	Attrib3	Class
1	Yes	Large	125K	No
2	No	Medium	100K	No
3	No	Small	70K	No
4	Yes	Medium	120K	No
5	No	Large	95K	Yes
6	No	Medium	60K	No
7	Yes	Large	220K	No
8	No	Small	85K	Yes
9	No	Medium	75K	No
10	No	Small	90K	Yes

Now classify the data →

Tid	Attrib1	Attrib2	Attrib3	Class
11	No	Small	55K	?
12	Yes	Medium	80K	?
13	Yes	Large	110K	?
14	No	Small	95K	?
15	No	Large	67K	?

Group 'B'

Attempt any EIGHT Questions. (8 × 5 = 40)

4. List any two challenge of multimedia mining. Differentiate between web usage mining and web content mining. (2 + 3)
5. How trust and distrust propagate in social network Explain. (5)
6. Why data preprocessing is mandatory? Justify. (5)
7. Describe any five types of OLAP operations. (5)
8. Given the following data set, find the frequent itemset using Apriori algorithm with minimum support 3. (5)

T1	{A, B, C, D, E, F}
T2	{B, C, D, E, F, G}
T3	{A, D, E, H}
T4	{A, D, F, I, J}
T5	{B, D, E, K}

9. Illustrate the hierarchical clustering with an example. (5)
10. Discuss about overfitting and underfitting. How precision and recall is used to evaluate classifier. (2 + 3)
11. What is the concept mini batch k-means? How DBSCAN works? (2 + 3)
12. How beam search and logic programming is used to mine graph? Explain. (5)