Duration:(3 Hours)   **[80 Marks]**

N.B. 1) Question No. 1 is compulsory.
    2) Attempt any Three questions out of the remaining.
    3) Assume suitable data wherever necessary and state them clearly.

Q.1 Solve any four of the following   (20)
- A. Compare OLTP vs OLAP systems.
- B. Explain the KDD process of data mining.
- C. Explain any two methods of evaluating the accuracy of a Classifier.
- D. Explain K-means clustering algorithm and draw flowchart.
- E. Explain multilevel association rule mining with example.
- F. Write a short note on web usage mining.

Q.2 A. Consider the following transaction database with minimum support 50% and minimum confidence 66%. Find the frequent patterns and strong association rules.   (10)

Tid	Items
10	A,C,D
20	B,C,E
30	A,B,C,E
40	B,E

Q.2 B. Explain different steps involved in data preprocessing.   (10)

Q.3 A. Find the clusters for the following dataset using a single link technique. Use Euclidean distance and draw the dendrogram.   (10)

Sample No	X	Y
P1	0.40	0.53
P2	0.22	0.38
P3	0.35	0.32
P4	0.26	0.19
P5	0.08	0.41
P6	0.45	0.30

Q.3.B. The college wants to record the Marks for the courses completed by students using the dimensions: I) Course, II)Student, III) Time & a measure Aggregate marks .
Create a cube and describe following OLAP operations :
I) Slice II) Dice III) Roll up IV) Drill Down V)Pivot   (10)

Q.4.A. What is dimensional modeling? Design the data warehouse dimensional model for a wholesale furniture Company. The data warehouse has to analyze the company's situation at least with respect to the Furniture, Customer and Time. Moreover, the company needs to analyze: The furniture with respect to its type, category and material. The customer with respect to their spatial location, by considering at least cities, regions and states. The company is interested in learning the quantity, income and discount of its sales..   (10)

Q.4 B. A data sample is given below. Find whether Patient X has flu or not using Naïve Bayes classifier.

If X= (chills=Y, runny nose=N, headache=Mild, fever=Y, flu=?)   (10)

chills	Runny nose	headache	fever	Flu
Y	N	Mild	Y	N
Y	Y	No	N	Y
Y	N	Strong	Y	Y
N	Y	Mild	Y	Y
N	N	No	N	N
N	Y	Strong	Y	Y
N	Y	Strong	N	N
|Y	Y	Mild	|Y	Y

Q.5 A.Explain Page Rank algorithm with example.   (10)
    B. Explain different data visualization techniques.   (10)

Q.6. Write short notes on following: (20)
    A. Applications of Data Mining.
    B. FP Tree
    C. Web content Mining
    D. Techniques of data Loading

****

B9EAC3297E8D354CAC0A4DDB91967A02