

Dy Pooja Sharma

2/5/2023

12

23

Department of Computer Science and Engineering
National Institute of Technology Hamirpur
Datawarehouse and Data Mining (CS-421)
B. Tech (Computer Science & Engineering/Dual Degree), 8th Semester
End Term Examination (May 2023)

Max Marks: 50

Time: 3 hours

Note: Attempt all the questions.

Q1. What is Clustering Technique? Discuss the Agglomerative algorithm with the following data and plot a Dendrogram using single link approach. The table below comprises sample data items indicating the distance between the elements.

Item	E	A	C	B	D
E	0	1	2	2	3
A	1	0	2	5	3
C	2	2	0	1	6
B	2	5	1	0	3
D	3	3	6	3	0

(10)

Q2. A database has six transactions. Let min-support = 50% and min- confidence = 60%. Find all frequent item sets by using Apriori Algorithm T_ID is the transaction ID.

Transaction	List of items
T1	I1,I2,I3
T2	I2,I3,I4
T3	I4,I5
T4	I1,I2,I4
T5	I1,I2,I3,I5
T6	I1,I2,I3,I4

(10)

24

Q3. Consider the weather dataset given below:

Day	Weather	Temperature	Humidity	Wind	Play?
1	Sunny	Hot	High	Weak	No
2	Cloudy	Hot	High	Weak	Yes
3	Sunny	Mild	Normal	Strong	Yes
4	Cloudy	Mild	High	Strong	Yes
5	Rainy	Mild	High	Strong	No
6	Rainy	Cool	Normal	Strong	No
7	Rainy	Mild	High	Weak	Yes
8	Sunny	Hot	High	Strong	No
9	Cloudy	Hot	Normal	Weak	Yes
10	Rainy	Mild	High	Strong	No

Make a classification model using ID-3 decision tree approach to predict whether the game will be played or not on a particular day. (10)

Q4. Describe the various functionalities of Data mining as a step in the process of knowledge Discovery. (5)

Q5. In what way ETL cycle can be used in typical data ware house. Explain with suitable instance. (5)

Q6. Why naïve Bayesian classification is called “naive”. Briefly outline the major ideas of naïve Bayesian classification. (5)

Q7. Consider following dimensions for a Hypermarket chain: Product, Store, Time and Promotion. With respect to this business scenario, answer the following questions, Clearly state any reasonable assumptions you make- Design a star schema, Whether the star schema can be converted to snowflake schema? Justify your answer and draw snowflake schema for the data warehouse (clearly mention the Fact table(s), Dimension table(s), their attributes and measures). (5)