

**Code No: 844AD****JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY HYDERABAD****MCA IV Semester Examinations, October/ November - 2020****DATA WAREHOUSING AND DATAMINING****Time: 2 Hours****Max.Marks:75**

Answer any five questions
All questions carry equal marks

1. Define data mining. Explain the steps in applying data mining to any database or data warehouse and the significance of data mining primitives. [15]
- 2.a) What is the use of principal component analysis in data preprocessing?
b) How to handle missing values in data mining process? [7+8]
- 3.a) With the help of a neat sketch describe the three-tier architecture of data warehouse.
b) Differentiate between dependent and independent data marts. [7+8]
4. Why are data models insufficient for data warehouse? Explain snowflake schema and fact constellation schema for data warehouse. [15]
5. Find strong association rules from the following transactional data using Apriori algorithm. Let support=30%, confidence=40% [15]

TID	List of Items
T1	Soap, shampoo, body lotion, toothpaste
T2	Mug, bucket, soap, scrub
T3	Dettol liquid, bucket, shower caps, soap
T4	Tooth brush, foot scrapper, shower caps
T5	Harpic, Dettol liquid, shower gel, shower caps
T6	Soap, body wash, toothpaste

- 6.a) What is the role of information gain in decision tree induction?
b) Why is k-nearest neighbor classifier known as lazy learner? [7+8]
7. What is an outlier? What are the applications of outlier analysis? Demonstrate deviation based approach for outlier detection. [15]
8. What is the objective of Page Ranking algorithm? Explain it as web structure mining algorithm. [15]

---oo0oo---