

Question Paper

Exam Date & Time: 25-Jun-2022 (02:00 PM - 05:00 PM)



MANIPAL ACADEMY OF HIGHER EDUCATION

Manipal School of Information Sciences, Manipal
Second Semester Master of Engineering - ME (Cloud Computing) Degree Examination - June 2022

Data Warehousing and Data Mining Elective -2 [CDC 5232]

Marks: 100

Duration: 180 mins.

Saturday, June 25, 2022

Answer all the questions.

- 1) Discuss what is Multi-dimensional Schema in a DWH? Describe Three-Tier Data Warehouse Architecture (TLO 1.1) (10)
- 2) Describe what is KDD process with a neat diagram and briefly write about various steps of a KDD PROCESS (TLO 1.2) (10)
- 3) Given the Initial attribute set: {A1, A2, A3, A4, A5, A6} Discuss Attribute subset selection for the following methods (TLO 1.2) (10)
 1. Step-wise forward selection
 2. Step-wise backward elimination
 3. Decision-tree induction
- 4) Illustrate and Generate the frequent item sets using the Apriori Algorithm for the following database with (TLO 2.1) (10)

$$Sup_{min} = 2$$

TID	Items
100	1 3 4
200	2 3 5
300	1 2 3 5
400	2 5

- 5) Examine and explain Constraint based association mining? Briefly describe five types of (10)

constraints? (TLO 2.1)

- 6) Analyze and Explain Classification-A Two-Step Process with a neat diagram of model construction process 1 and 2 using the following training and testing Dataset TLO 2.2 (10)

Training Dataset

NAME	RANK	YEARS	TENURED
Mike	Assistant Prof	3	no
Mary	Assistant Prof	5	yes
Bill	Professor	2	yes
Jim	Associate Prof	6	yes
Dave	Assistant Prof	4	no
Anne	Associate Prof	3	no

Testing Dataset

NAME	RANK	YEARS	TENURED
Tom	Assistant Prof	2	no
Merlisa	Associate Prof	7	no
George	Professor	5	yes
Joseph	Assistant Prof	7	yes

- 7) Analyze and Describe What are support vector machines used for? What is a support vector in SVM? (TLO 2.2) (10)
- 8) Define what is Cluster Analysis? Give a writeup on Major Clustering Approaches? TLO 3.1 (10)
- 9) Give an example of how specific clustering methods may be integrated, for example, where one clustering algorithm is used as a preprocessing step for another. In addition, provide reasoning on why the integration of two methods may sometimes lead to improved clustering quality and efficiency TLOO 3.1 (10)
- 10) Name and Discuss at least 4 Major application areas for data mining TLOO 3.2 (10)

-----End-----