PAPER ID-310614

Roll No: | | | | | | | | | | | | | |

MCA
(SEM III) THEORY EXAMINATION 2024-25
DATA WAREHOUSING & DATA MINING

TIME: 3 HRS **M.MARKS: 100**

Note: Attempt all Sections. In case of any missing data; choose suitably.

SECTION A

1. Attempt *all* questions in brief. **2 x 10 = 20**

Q no.	Question	CO	Level
a.	Though we can work with database, why Data Warehouse is required?	1	K2
b.	What do you mean by data warehouse schema?	1	K2
c.	Explain Data Cube.	2	K2
d.	How data mining is different from data warehousing.	2	K2
e.	Discuss difference between OLTP and OLAP.	3	K2
f.	Give the reason for noisy data?	3	K2
g.	Difference between clustering and classification.	4	K2
h.	Discuss basic assumption in Naïve Bayes Classifiers.	4	K2
i.	What do you mean by data visualization?	5	K2
j.	Explain Spatial Data with example.	5	K2

SECTION B

2. Attempt any *three* of the following: **10 x 3 = 30**

a.	Describe Three-tier data warehouse architecture with a neat diagram.	1	K2
b.	Explain Star schema & snowflake schema with suitable example.	2	K2
c.	Discuss K means algorithm with proper example.	3	K2
d.	Describe different types of OLAP operations with examples.	4	K2
e.	Explain web mining in detail	5	K2

SECTION C

3. Attempt any *one* part of the following: **10 x 1 = 10**

a.	Why data cleaning is required? Discuss the methods of data cleaning.	1	K2
b.	Describe in detail about the Hardware and Operating Systems for Data Warehousing.	1	K2

4. Attempt any *one* part of the following: **10 x 1 = 10**

a.	Discuss the KDD process in detail.	2	K2
b.	Assume we have the following dataset of 6 samples, each with 2 features, and a binary class label:	2	K3

Sample	Age	Income	Class
1	25	High	No
2	30	High	No
3	35	Medium	Yes
4	40	Medium	Yes
5	50	Low	Yes
6	60	Low	No

The entropy of the entire data set is 1, Entropy of the splitting attribute Age is 1, Weighted Entropy of the splitting attribute Age is 1, Entropy of the splitting attribute Income is 0, Weighted Entropy of the splitting attribute Income is .33. Draw the decision tree.

PAPER ID-310614

Roll No: | | | | | | | | | | | | | |

MCA
(SEM III) THEORY EXAMINATION 2024-25
DATA WAREHOUSING & DATA MINING

TIME: 3 HRS **M.MARKS: 100**

5. Attempt any *one* part of the following: 10 x 1 = 10

a.	How distance-based technique is used for data mining?	3	K2
b.	<table><tr><th>Transaction</th><th>List of Items</th></tr><tr><td>T1</td><td>I1, I2, I3, I5</td></tr><tr><td>T2</td><td>I2, I3, I4</td></tr><tr><td>T3</td><td>I4, I5</td></tr><tr><td>T4</td><td>I2, I4, I5</td></tr><tr><td>T5</td><td>I1, I2, I3, I5</td></tr><tr><td>T6</td><td>I1, I2, I3, I4</td></tr></table> Support = 50%, Confidence = 60% Generate association rule for given items using Apriori Algorithms.	3	

6. Attempt any *one* part of the following: 10 x 1 = 10

a.	Discuss Hierarchical Clustering in detail.	4	K2
b.	Write notes on i) STING ii) CLIQUE	4	K2

7. Attempt any *one* part of the following: 10 x 1 = 10

a.	Discuss in detail the security issues in Data Warehousing.	5	K2
b.	What is temporal mining? Explain in detail	5	K2