

END SEMESTER EXAMINATION, JUNE - 2019
MCA FOURTH SEMESTER

COURSE CODE: MCA208	COURSE TITLE: DATA WAREHOUSE & DATA MINING
----------------------------	---

[Time allotted: Three hours]**[Max. Marks: 100]***Note: Attempt all Sections.*

Section (A)

Q.1. Attempt all questions.**(2 x 10 = 20)**

- | | |
|---|---|
| a. Explain dimension and measure?
b. Define data mart.
c. What is ROLAP?
d. Write short note on histogram.
e. Differentiate classification and clustering. | f. What do you mean by data preprocessing?
g. Write short notes on: (i) Data (ii) Knowledge
h. Define the terms cleaning and transformation.
i. What is normalization in context of data mining?
j. Differentiate between database & data warehouse. |
|---|---|

Section (B)

Q.2. Attempt any 5 questions.**(7 x 5 = 35)**

- a.** Differentiate between data warehouse and data mining.
- b.** Discuss the need of human intervention in data mining process.
- c.** Write short note on: **(i)** Generalization **(ii)** Aggregation
- d.** Write down decision tree induction algorithm with a suitable example.
- e.** Explain Direct Hashing and Pruning with a suitable example.
- f.** Explain the terms overfitting and pruning.
- g.** Explain the conceptual view of a data warehouse.

Section (C)

Attempt any 3 questions.**(15 x 3 = 45)**

- | | | |
|-------------|---|--|
| Q.3. | a. What are data mining functionalities?
b. What is redundant data? How to handle data redundancy discuss in details? | 8
7 |
| Q.4. | a. Draw KDD process diagram. Discuss the five major elements of data mining in details.
b. What is noisy data? How can we handle noisy data? Explain with examples. | 8
7 |
| Q.5. | a. What is data cleaning? Explain all the phases of data cleaning.
b. Given a set of data :
{12,8,45,23,6,7,8,25,34,22,19,22,34,45,41}
Can you employ the binning method to allocate the data on the set into 3 bins by equal-frequency, Smoothing by bin means and smoothing by bin boundaries respectively? | 8

7 |
| Q.6. | a. Define the following terms: <ul style="list-style-type: none">• Min-max Normalization• Z-Score Normalization• Decimal Scaling b. Define the following terms: <ul style="list-style-type: none">• Star Schema• Snow Flake Schema• Fact Constellations |

2.5
2.5
2.5

2.5
2.5
2.5 |
| Q.7. | a. Write down apriori algorithm. Also explain an example of apriori algorithm.
b. What s clustering? Explain all the features and methods of clustering. | 8
7 |
