Enrol. No. | S | R | H | U | | | | | | | | Regn. No. | D | D | | | | | | | | | |

END SEMESTER EXAMINATION, JUNE - 2019
MCA FIFTH SEMESTER

COURSE CODE: MCA303 **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. What is the need of data mining tools?
b. Explain normalization in details?
c. Define meta data.
d. Explain DBMS schema?
e. What about data transformation?
f. Explain correlation analysis.
g. Write short note on data replication.
h. Explain the need of Data Warehouse.
i. What do you mean by data preprocessing?
j. Write short note on distributed memory architecture.

Section (B)

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

a. What are the different methods available to perform data mining, explain in brief.
b. What are relationships needed between data base to explore on data mining.
c. Write short note on: a) missing value b) noisy data
d. What are the different data cleaning approaches?
e. Explain the various types of normalization technique.
f. What do you mean by data reduction and what are the different processes of data reduction?
g. Describe the working of Partitioning Around Medoids (PAM) algorithm.

Section (C)

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

Q.3. **a.** What about genetic algorithm? Compare genetic algorithm with other on techniques and explain its use in other methods.

b. Explain mapping data warehouse with multiprocessor architecture with the concept of parallelism and data partitioning.

Q.4. **a.** Suppose that the data mining task is to cluster points (with (x, y) representing location) into three clusters, where the points are: A1(2,10), A2(2,5), A3(8,4), B1(5,8), B2(7,5), B3(6,4), C1(1,2), C2(4,9)
The distance function is Euclidean distance. Suppose initially we assign A1, B1, and C1 as the center of each cluster, respectively. Use the k-means algorithm to show only
i. The three cluster centers after the first round of execution. and **ii.** The final three clusters.

b. What about noisy data? How to remove noise from data? Discuss in detail.

Q.5. **a.** What do you mean by subject-oriented, integrated, time-variant and non-volatile collection of data in data warehousing?

b. Given a set of data : **{4,8,9,15,21,21,24,25,26,28,29,34}**
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?

Q.6. **a.** Use the two methods below to normalize the following group of data: 200, 300, 400, 600, 1000
i. min-max normalization by setting min =0 and max=1 and **ii.** z-score normalization

b. Write short notes on: **i.** Bayesian classifiers and **ii.** Bayesian net

Q.7. **a.** Define Client/Server Model. Explain second generation client/server model with suitable diagram.

b. What do you mean by decision tree? What are advantages and disadvantages of decision tree approaches over

other approaches of data mining?