Roll No .....

MCSE-301(A)

M.E./M.Tech., III Semester Examination, June 2020

Data Warehousing and Mining

(Elective - I)

Time : Three Hours

Maximum Marks: 70

Note: i) Attempt any five questions.
ii) All questions carry equal marks.

1. a) Suppose that you are employed as a data mining consultant for an E-commerce company. Describe how data mining and social media analysis can help the company by giving specific examples of how techniques, such as clustering, classification, association rule mining, and anomaly detection can be applied.

   b) List and explain various types of data set which are used for analysis.

2. a) Explain the different challenges that motivated the development of the data mining technologies.

   b) Define data warehouse. Explain what the need for developing a data warehouse and hence its architecture.

3. a) What are advantages of data mining? What is non linear regression method? Give two examples.

   b) What is metadata and why is it important? Discuss the multidimensional data.

4. a) Explain K-means logarithm with help of example. Differentiate between k-means and k-medoids partitioning methods.

   b) For the following vectors, x and y, calculate the indicated similarity or distance measures.

      i) x = (1,1,1,1,), y = (2,2,2,2) cosine, correlation, Euclidean.

      ii) x = (0,1,0,1), y=(1,0,1,0) cosine, correlation, Euclidean, Jaccard.

5. Consider the following data set for a binary class problem:

A	B	Class Label
T	F	+
T	T	+
T	T	+
T	F	-
T	T	+
F	F	-
F	F	-
F	F	-
T	T	-
T	F	-

a) Calculate the information gain when splitting on A and B. Which attribute would the decision tree induction algorithm choose?

b) Calculate the gain in the Gini index when splitting on A and B. Which attribute would the decision tree induction algorithm choose?

6. a) What is meant by classification? Justify why clustering is said to be supervised learning. How the classifier accuracy determined and also explains its various types.

b) Explain time series analysis with real life example.

7. a) Explain content based image and video retrieval with help of example.

b) What is crisp set? How it is different with fuzzy set? Explain genetic algorithm.

8. Write short notes on :

a) OLAP, MOLAP, HOLAP

b) Social impact of data mining.

******