CBCS SCHEME

USN | | | | | | | | | | |

16/17MCA442

Fourth Semester MCA Degree Examination, Dec.2019/Jan.2020

Data Warehousing and Data Mining

Time: 3 hrs.

Max. Marks: 80

Note: *Answer FIVE full questions, choosing ONE full question from each module.*

Module-1

1 a. What is ETL? Differentiate between ODS and Datawarehouse. **(08 Marks)**

b. Explain Inmon's definition of a datawarehouse. Discuss the star schema used in data warehouse design with a suitable example. **(08 Marks)**

OR

2 a. Explain the guidelines for the implementation of a datawarehouse. **(08 Marks)**

b. Discuss the typical OLAP operations with an example. **(08 Marks)**

Module-2

3 a. What is data preprocessing? Explain any two data pre-processing steps in detail. **(12 Marks)**

b. Calculate SMC and Jaccard in the following binary vectors :
$x = (1, 0, 0, 0, 0, 0, 0, 0, 0, 0)$ and $y = (0, 0, 0, 0, 0, 0, 1, 0, 0, 1)$ **(04 Marks)**

OR

4 a. Explain the different types of datasets in data mining. **(10 Marks)**

b. What is data mining? Discuss the KDD process. **(06 Marks)**

Module-3

5 a. Illustrate the process of reducing the number of candidate itemsets using Apriori principle with a suitable example. **(10 Marks)**

b. Define : i) Support ii) Confidence iii) Frequency item set. **(06 Marks)**

OR

6 a. State FP – growth algorithm. Construct FP – tree for the following transaction data set.

Tid	1	2	3	4	5
Item	{a, b}	{b, c, d}	{a, c, d, e}	{a, d, e}	{a, b, c}
Tid	6	7	8	9	10
Item	{a, b, d}	{a}	{a, b, c}	{a, b, d}	{b, c, e}

Show the trees separately after reading each transaction. **(10 Marks)**

b. Explain maximal frequency and closed frequency itemsets. **(06 Marks)**

Module-4

7 a. Explain Hunt's algorithm for a decision tree induction and illustrate it using a training set. **(08 Marks)**

b. Write an algorithm for K-NN classifiers. Discuss the characteristics of K-NN. **(08 Marks)**

OR

8 a. Explain the various measures for selecting the best split with an example, for each attribute type. **(08 Marks)**

b. Explain Bayesian classification in detail. **(08 Marks)**

Important Note : 1. On completing your answers, compulsorily draw diagonal cross lines on the remaining blank pages.
2. Any revealing of identification, appeal to evaluator and /or equations written eg, 42+8 = 50, will be treated as malpractice.

Module-5

9 a. What is cluster analysis? Explain the basic K-means clustering technique with an example. **(10 Marks)**

b. Write a note on three types of cluster evaluation. **(06 Marks)**

OR

10 a. List and explain the various approaches to anomaly detection. **(08 Marks)**

b. Illustrate with an example how agglomerative clusterings method can be implemented. **(08 Marks)**

* * * * *