CBCS SCHEME

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

15CS651

Sixth Semester B.E. Degree Examination, Aug./Sept.2020
Data Mining and Data Warehousing

Time: 3 hrs. Max. Marks: 80

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

Module-1

1 a. What is a Data Warehouse? Discuss various usage and trends in data warehousing. **(08 Marks)**

b. Explain in detail the three – tier Data warehouse architecture. **(08 Marks)**

OR

2 a. Discuss the concept of star, snowflake and galaxy schemas for multidimensional databases. **(08 Marks)**

b. Summarize the various OLAP operations in the multidimensional data model. **(08 Marks)**

Module-2

3 a. Identify the different indexing method used for OLAP data with brief explanation. **(08 Marks)**

b. Differentiate ROLAP , MOLAP and HOLAP servers. **(08 Marks)**

OR

4 a. What is Data Mining? Explain various data mining tasks, with suitable examples. **(08 Marks)**

b. Explain different steps involved in preprocessing steps. Write any 3 challenges faced in Data Mining. **(08 Marks)**

Module-3

5 a. What is Apriori Algorithm? How it is used to find frequent item sets? Explain. **(08 Marks)**

b. Illustrate the advantages of using closed frequent itemsets, with an example. Show the relationships among frequent, maximal frequent and closed frequent itemsets. **(08 Marks)**

OR

6 a. Explain FP – growth algorithm for discovering frequent itemsets. **(08 Marks)**

b. Briefly explain the Objective Measures of Interestingness for evaluating association patterns. **(08 Marks)**

Module-4

7 a. Define Classification. With a neat figure, explain the general approach for solving classification model. **(08 Marks)**

b. Consider the following data set for a binary.

Instance	a_1	a_2	a_3	Target class
1	T	T	1.0	+
2	T	T	6.0	+
3	T	F	5.0	-
4	F	F	4.0	+
5	F	T	7.0	-
6	F	T	3.0	-
7	F	F	8.0	-
8	T	F	7.0	+
9	F	T	5.0	-

Classification problem :

i) What is the entropy of this collection of training examples, with respect to the positive class?

ii) What are the information gains of a_1 and a_2 relative to these training examples?

iii) For a_3, which is a continuous attribute, compute the information gain for every possible split. **(08 Marks)**

OR

8 a. What is a rule – based classifier? Explain sequential covering algorithm in rule – based classifier. **(08 Marks)**

b. Write an algorithm for K – Nearest Neighbor (KNN) classification. List the characteristics of Nearest Neighbor classifiers. **(08 Marks)**

Module-5

9 a. What is Cluster analysis? Discuss the different types of clusters with examples. **(08 Marks)**

b. Describe K – means clustering algorithm. What are its limitations? **(08 Marks)**

OR

10 a. Discuss DBSCAN algorithm for clustering. **(08 Marks)**

b. Explain Agglomerative Hierarchical Clustering Algorithm, with different proximity between clusters. **(08 Marks)**

* * * * *