CBCS SCHEME

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

15CS651

Sixth Semester B.E. Degree Examination, June/July 2018
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 data warehouse? Elaborate data warehouse using multi-tier architecture with a neat diagram. **(10 Marks)**

b. Explain star schema and snow flake schema with examples. **(06 Marks)**

OR

2 a. What is data cube measure? How it is categorized? Explain. **(05 Marks)**

b. Explain data warehouse model with neat diagram. **(07 Marks)**

c. Briefly elaborate on typical OLAP operations on multidimensional data. **(04 Marks)**

Module-2

3 a. With respect to indexing, explain Bitmap Index and Join Index. **(04 Marks)**

b. Describe the servers involved in implementation of a warehouse server. **(08 Marks)**

c. How data mining tasks are categorized? Explain. **(04 Marks)**

OR

4 a. Describe the challenges that motivated the development of data mining. **(10 Marks)**

b. What are the properties necessary to describe attributes? Explain different types of attributes. **(06 Marks)**

Module-3

5 a. Describe frequent item set generation in Aprior algorithm with example. **(08 Marks)**

b. Explain the uses of Hash Tree in support counting. **(08 Marks)**

OR

6 a. Describe alternative methods for generating frequent item sets. **(09 Marks)**

b. Consider the following transaction dataset. Describe the construction of FP-Tree in FP-Growth algorithm. **(07 Marks)**

Tid	Items
1	{a, b}
2	{b, c, d}
3	{a, c, d, e}

Module-4

7 a. Illustrate hunt's algorithm to develop a decision tree. Consider the following training set and derive the decision tree. **(09 Marks)**

Tid	Home Owner	Marital Status	Annual Income	Defaulted Borrower
1	Yes	Single	125 K	No
2	No	Married	100 K	No
3	No	Single	70 K	No
4	Yes	Married	120 K	No
5	No	Divorced	95 K	Yes
6	No	Married	60 K	No
7	Yes	Divorced	220 K	No
8	No	Single	85 K	Yes
9	No	Married	75 K	No
10	No	Single	90 K	Yes

b. What are the characteristics of decision tree induction algorithms? **(07 Marks)**

OR

8 a. What are the characteristics of Nearest Neighbor classifiers? **(06 Marks)**

b. How Bayes theorem can be used for solving a classification problem? Explain. **(10 Marks)**

Module-5

9 a. Describe different types of clustering mechanisms. **(06 Marks)**

b. Explain DBSCAN algorithm. How the parameters are selected? **(06 Marks)**

c. List out important issues for cluster validation. **(04 Marks)**

OR

10 a. Illustrate Grid-based clustering algorithm. How clusters are formed from Dense-Grid cells. **(12 Marks)**

b. Develop DENCLUE algorithm for kernel density estimation. **(04 Marks)**

* * * * *