CBCS SCHEME

USN | | | | | | | | | | 15CS651

Sixth Semester B.E. Degree Examination, Dec.2019/Jan.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 Data warehouse? Explain three tier architecture of data warehouse. **(08 Marks)**
b. Explain the schemas of multidimensional data models. **(08 Marks)**

OR

2 a. What is Data cube measure? Explain the categorization of measures. **(08 Marks)**
b. Explain data cube operations with examples. **(08 Marks)**

Module-2

3 a. Explain data cube computation and curse of dimensionality. **(08 Marks)**
b. Explain different methods of indexing OLAP data. **(08 Marks)**

OR

4 a. State and explain various data mining tasks. **(08 Marks)**
b. Define Similarity and dissimilarity between the objects. Find SMC and Jaccord's coefficient of two binary vectors.
X = (1, 0, 0, 0, 0, 0, 0, 0, 0, 0) Y = (0, 0, 0, 0, 0, 0, 1, 0, 0, 1). **(08 Marks)**

Module-3

5 a. What is Association Analysis? Explain Association rule , Support and Confidence. **(08 Marks)**
b. State Apriori principle. Write apriori algorithm for frequent itemset. **(08 Marks)**

OR

6 a. Construct an FP tree for the following dataset. **(08 Marks)**

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

b. Explain the strategies used in frequent itemset generation. **(08 Marks)**

Module-4

7 a. Explain the general approach for solving classification problem. **(08 Marks)**
b. Write the algorithm for decision tree induction. **(08 Marks)**

OR

8 a. Explain the methods of comparing classifiers. (08 Marks)
b. Write the characteristics of nearest neighbor classifier. (08 Marks)

Module-5

9 a. Explain the requirements of cluster analysis. (08 Marks)
b. State and explain K – means algorithm. (08 Marks)

OR

10 a. Write DBSCAN clustering algorithm and estimate time and space complexity. (08 Marks)
b. State and explain the issues in cluster evaluation. (08 Marks)

* * * * *

CBCS SCHEME

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

15CS651

Sixth Semester B.E. Degree Examination, June/July 2019
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. Describe a 3 – tier data warehousing architecture. **(06 Marks)**

b. Compare OLTP and OLAP Systems. **(06 Marks)**

c. What is a Data warehouse and what are its four key features? **(04 Marks)**

OR

2 a. Explain with suitable examples the various OLAP operations in a multidimensional data model. **(07 Marks)**

b. Explain the following terms with examples : i) Snowflake schema ii) Fact constellation schema iii) Star schema **(09 Marks)**

Module-2

3 a. Describe ROLAP , MOLAP , HOLAP. **(06 Marks)**

b. What is Data Mining? With a neat diagram, explain the KDD process in data mining. **(06 Marks)**

c. For the following vectors X and Y, calculate the cosine similarity, where
X = {3 2 0 5 0 0 0 2 0 0} , Y = {1 0 0 0 0 0 0 1 0 2}. **(04 Marks)**

OR

4 a. Describe the various types of attributes and data sets. **(08 Marks)**

b. Define Data preprocessing. Mention the steps involved in it. Explain any 2 steps in detail. **(08 Marks)**

Module-3

5 a. Briefly explain the Apriori Algorithm for frequent itemset generation. **(05 Marks)**

b. Explain the following terms with example :
i) Rule – generation ii) Computational complexity. **(06 Marks)**

c. Generate frequent itemset for the given data with support = 50%. **(05 Marks)**

TID	100	200	300	40
Items	{1, 3, 4}	{2, 3, 5}	{1, 2, 3, 5}	{2, 5}

OR

6 a. Consider the following transaction data set :
i) Construct an FP tree ii) Generate the list of frequent itemset. **(09 Marks)**
Ordered by their corresponding suffixes.

TID	1	2	3	4	5	6	7
Items	{a, b}	{b, c, d}	{a, c, d, e}	{a, d, e}	{a, b, c}	{a, b, c, d}	{a}

8	9	10
{a, b, c}	{a, b, d}	{b, c, e}

b. Briefly explain the candidate generation procedure using $F_{k-1} \times F_{k-1}$ Merging strategy. **(07 Marks)**

Module-4

7 a. Explain how decision tree induction algorithm works. Give example. **(08 Marks)**
b. List and explain the different characteristics of decision tree induction. **(08 Marks)**

OR

8 a. Describe the nearest neighbour classification technique. **(09 Marks)**
b. Write a note on Bayesian classifier. **(07 Marks)**

Module-5

9 a. What is Cluster analysis? Describe the different types of clustering techniques with example. **(08 Marks)**

b. Explain the following terms :
i) K – means clustering ii) Graph based clustering. **(08 Marks)**

OR

10 a. What are the basic approaches used for generating a agglomerative hierarchical clustering? **(08 Marks)**

b. Explain D B Scan algorithm, with example. **(08 Marks)**

* * * * *