CBCS SCHEME

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

18CS641

Sixth Semester B.E. Degree Examination, June/July 2023
Data Mining and Data Warehousing

Time: 3 hrs. Max. Marks: 100

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

Module-1

1 a. Define Datawarehouse. Explain Multitier Architecture of Data Warehousing with diagram. **(10 Marks)**

b. Explain ETL process with the neat diagram. **(10 Marks)**

OR

2 a. Explain the schemas of multi-dimensional data models. **(10 Marks)**

b. Explain data cube operations with example. **(10 Marks)**

Module-2

3 a. Explain different methods of indexing OLAP data. **(10 Marks)**

b. Explain the following preprocessing techniques :
(i) Feature subset selection. (ii) Sampling. **(10 Marks)**

OR

4 a. List the different types of Dataset and explain with an example. **(10 Marks)**

b. Consider X = (0, 1, 0, 1), Y = (1, 0, 1, 0). Find cosine, correlation, Euclidean, Jaccard and SMC. **(10 Marks)**

Module-3

5 a. A database has five transactions. Let min sup = 60% and min conf = 80%.

Table Q5 (a)

TID	Items.bought
T_{100}	{M,O,N,K,E,Y}
T_{200}	{D,O,N,K,E,Y}
T_{300}	{M,A,K,E}
T_{400}	{M,U,C,K,Y}
T_{500}	{C,O,O,K,I,E}

(i) Find all frequent itemsets using Apriori Algorithm.
(ii) List all the strong association Rules. **(12 Marks)**

b. Identify and explain the alternative methods for generating frequent itemsets. **(08 Marks)**

OR

6 a. Construct FP tree by showing tree separately after reading each transaction and find the frequent itemset generation. Consider the transaction dataset :

Table : Q6 (a)

TID	Items.bought
100	{f, a, c, d, g, i, m, p}
200	{a, b, c, f, l, m, o}
300	{b, f, h, j, o}
400	{b, c, k, s, p}
500	{a, f, c, e, l, p, m, n }

Let min-support = 3 **(12 Marks)**

b. Explain the various measures of evaluating association patterns. **(08 Marks)**

Module-4

7 a. Explain the general approach for solving classification problem. **(08 Marks)**

b. Build a decision tree using Hun't Algorithm for the given dataset.

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

Table Q7 (b) **(12 Marks)**

OR

8 a. Explain K-nearest neighbor classification algorithm with example. **(08 Marks)**

b. State Bays theorem and explain how bayes theorem is used in the Naïve Bayesian classifier with example. **(12 Marks)**

Module-5

9 a. Calculate the cluster for the following 8 points (x, y). Represents into 3 clusters $A_1(2, 10)$ $A_2(2, 5)$ $A_3(8, 4)$ $B_1(5, 8)$ $B_2(7, 5)$ $B_3(6, 4)$ $C_1(1, 2)$ $C_2(4, 9)$ **(10 Marks)**

b. Explain the following :
(i) Density based clustering.
(ii) Graph based clustering. **(10 Marks)**

OR

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

b. Explain DBSCAN algorithm with example. **(10 Marks)**

* * * * *