PVP 14

Code: CS6T4

III B.Tech - II Semester – Regular/Supplementary Examinations
AUGUST - 2021

DATA WAREHOUSING AND DATA MINING
(COMPUTER SCIENCE & ENGINEERING)

Duration: 3 hours Max. Marks: 70

PART – A

Answer *all* the questions. All questions carry equal marks

11x 2 = 22 M

1.

a) What is Online Analytical Processing?
b) What are the ideas of subject-oriented and non-volatile features of data warehouse?
c) List the data mining functionalities
d) Compute the mid-range of following data
18, 5, 20, 22, 67, 82, 44, 25, 43, 88, 76, 99, 56, 73, 81, 34, 90, 62, 14, 33
e) What is an association rule? When an association rule becomes a strong association rule?
f) State Bayes Theorem.
g) How is dissimilarity computation is done for ordinal variables and ratio-scaled variables?
h) Define cluster analysis.
i) What is outlier.
j) Outline the complex data types.

k) Write the formula for variance of a dataset. Explain each term in the formula.

PART – B

Answer any *THREE* questions. All questions carry equal marks.

3 x 16 = 48 M

2. a) Explain the data warehouse implementation. 8 M

b) What are the various OLAP operations are used in the multidimensional data model? Explain them in detail with an example. 8 M

3. a) What are the major issues in data mining? Explain. 8 M

b) Normalize the following group of data by using the following techniques.
200, 300, 400, 600, 1000
i. min-max normalization technique
ii. z-score normalization
iii. Decimal scaling.
Write your observations on the above techniques. 8 M

4. Consider the following transactional data for a commercial shop.

TID	List of Items with Ids
T1	i2, i4
T2	i1, i2, i5
T3	i2, i3
T4	i1, i3
T5	i1, i2, i4
T6	i2, i3
T7	i1, i3
T8	i1, i2, i3
T9	i1, i2, i3, i5

Generate all the frequent itemsets using apriori algorithm. Consider the minimum support count is 2. Clearly show your computational steps. 16 M

5. a) Explain about the k-means clustering algorithm. 8 M

b) What do you mean by grid-based clustering method? Explain in detail. 8 M

6. a) What are the challenges of outlier detection? Explain in detail. 8 M

b) Explain in detail about distance-based outlier detection method. 8 M