

Code: CS6T4

III B.Tech-II Semester–Regular/Supplementary Examinations–March 2019

**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 a data warehouse?
- b) List OLAP Operations.
- c) Write a note on OTLP.
- d) Write the equations for Support and Confidence in Association.
- e) How datasets are classified?
- f) Define pre-processing.
- g) Pattern measurements are only possible in clustering, Justify.
- h) Identify the issues in bayes classification.
- i) List two major usage of clustering.
- j) Identify the types of outliers.
- k) Mention the usage of outlier detection

PART – B

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

3 x 16 = 48 M

2. a) Differentiate between operational database system and data warehouse. 8 M

b) Explain the basic elements of Data warehouse with a neat sketch. 8 M

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

b) Differentiate between data reduction and dimensionality reduction for data discretization. 8 M

4. a) Differentiate between Rule based and Bayes Classification. 8 M

b) Apply Apriori Algorithm in tracing all the frequent item datasets

Transactions	Itemset
T100	1 2 3
T200	2 3 5
T300	1 2 3 5
T400	2 5
T500	1 3 5

#Hint: Consider appropriate minimal support and minimal confidence values for generating the rules. 8 M

5. a) Explain K Means clustering method. 8 M

b) Apply agglomerative hierarchical clusters for the given distance matrix. 8 M

	1	2	3	4	5
1	0				
2	9	0			
3	3	7	0		
4	6	5	9	0	
5	11	10	2	8	0

6. a) Explain in detail multivariate outlier . 8 M

b) Describe and list all the common courses of outliers on a dataset. 8 M