PVP 14

Code: CS6T4

III B.Tech - II Semester – Regular/Supplementary Examinations
March 2020

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 OLAP?
   b) Define a data warehouse model.
   c) What is data mining?
   d) Define a pattern.
   e) What is a frequent item set?
   f) Recall Rule based system.
   g) What is a dissimilarity matrix?
   h) In k-fold cross validation, what is the highest value of k and why?
   i) What is a mediod?
   j) What is an outlier detection method?
   k) List four complex mining datatypes.

PART – B

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

3 x 16 = 48 M

2. a) Explain Star Schema in multidimensional database with a neat diagram. 8 M

   b) Briefly explain about efficient computation of Data Cubes. 8 M

3. a) Explain the different data mining tasks. 8 M

   b) Describe Data Transformation & Data Discretization. 8 M

4. a) How tree pruning in decision tree induction is useful? Explain various methods for pruning decision trees. 8 M

   b) Explain about Bayesian belief networks. 8 M

5. a) Explain Agglomerative Hierarchical Clustering algorithm with example. 8 M

   b) Differentiate between density based and grid-based methods. 8 M

6. a) Discuss any two outlier detection techniques in detail. 8 M

   b) Discuss about mining text databases. 8 M