RCSE-001

Ph.D. IN COMPUTER SCIENCE (PHDCS)

⌂⌂2⌂5   **Term-End Examination**

June, 2018

RCSE-001 : DATA MINING

Time : 3 hours   *Maximum Marks : 100*
(Weightage : 50%)

*Note : Question no. 1 is **compulsory**. Answer any **three** questions from the rest.*

1. (a) A bank has data of its customers for the last 10 years. The customers may have availed many services of the bank like loan, saving account, checking account, locker facilities, etc. The information about account holders includes place of stay, place of work, gender, educational qualification, income tax deduction, etc. Perform the following tasks for this bank (make and state assumptions).

(i) Suggest (at least five) different data pre-processing methods that can be used for the bank. *10*

(ii) Suggest a suitable multi-dimensional data model, that will be useful for the bank. You must give reasons for your selection. *5*

(ii) Describe how classification and prediction techniques may help the bank. *5*

(b) Explain the concept of Association Rules, with the help of an example. *5*

(c) What is Bayes' theorem ? Explain. *5*

(d) Explain linear regression with the help of an example. How is it used in prediction ? *10*

2. (a) What is OLAP in the context of data warehousing ? Explain the OLAP operations – Roll up, Drill-down, Slice and dice with the help of an example. *10*

(b) Explain the three-tier data warehouse architecture with the help of a diagram. *10*

3. (a) What is frequent item-set ? Explain the terms support and confidence in the context of frequent item-set, with the help of an example. *8*

(b) Demonstrate the use of Apriori algorithm to determine frequent item-sets for the following transactions : *12*

Transaction ID	Item List Purchased
S001	P1, P2, P5
S002	P2, P3
S003	P1, P2, P3, P5
S004	P2, P3
S005	P1, P3
S006	P2, P3
S007	P1, P3
S008	P1, P2, P3
S009	P2, P4, P5

Use a minimum support count of 2. Make suitable assumptions, if any.

4. (a) Discuss the major issues and challenges in data mining. *10*

(b) What is meant by Rule-based classification ? How can you use If-then rules for classification ? Explain with the help of an example. *10*

5. Explain any *four* of the following : $4 \times 5 = 20$

(a) Outlier Analysis

(b) Concept Hierarchy in the Context of Data Warehouse

(c) Dimensionality Reduction

(d) Correlation Analysis

(e) Decision Tree

(f) Support Vector Machines