No. of Printed Pages : 4

RCSE–001

Ph. D. IN COMPUTER SCIENCE

(PHDCS)

Term-End Examination

December, 2024

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) Compare classification and prediction. Discuss the methods used for classification and methods used for prediction. List the steps involved in preparing the data for classification and prediction. 10

   (b) *Books N Books* is a large distributor with domestic and international distribution channels. The company orders from

publishers and distributes publications to all the leading booksellers. You want to build a Data Warehouse (DW) to analyze shipments that are made from the company's warehouse. Prepare IPD showing all the facts, dimensions and the hierarchies within each dimension. Convert it to STAR SCHEMA. 10

(c) A market sales basket has five transactions :

TID	Items Brought
T1	ABCDEF
T2	BCDEFG
T3	ADEH
T4	ADFIJ
T5	BDJKL

Find all frequent item sets and association rules using Apriori algorithm. Let the minimum support be 40% and minimum confidence be 40%. 10

(d) Define Support Vector Machine (SVM). Discuss the applicability of SVM when the data is linearly separable. Give suitable example in support of your discussion. 10

2. (a) What is a Frequent Item Set ? Which of the following two algorithms : F_P tree and Apriori is more efficient for finding frequency item sets ? Justify your answer. 10

(b) What is Regression ? How is it different from Correlation ? Is regression a technique for classification or prediction ? Justify your answer. Briefly discuss the terms linear regression and non-linear regression. 10

3. (a) With the help of an example schema for each, explain the following multidimensional data models along with their *pros* and *cons* : 10

(i) Snowflake schema

(ii) Fact constellations

(b) What is Data Mining ? What is the role of data warehouse in data mining ? Explain the types of problems that can be addressed using data mining. 10

4. (a) Explain different issues and challenges in data mining. Describe different misconceptions about data mining in brief. 10

(b) Describe the steps for the design and architecture of a 3-tier data warehouse from e-commerce company along with necessary considerations. Draw the architectural diagram also. 10

5. (a) Discuss the following algorithms : 5+5

(i) ID3

(ii) C4.5

(b) Write short notes on the following : 5+5

(i) ROC curves

(ii) Data clustering

× × × × × × ×