Exam.Code:0924
Sub. Code: 6851

2062

B.E. (Information Technology) Sixth Semester
IT-601: Data Warehouse and Data Mining

Time allowed: 3 Hours

Max. Marks: 50

NOTE: *Attempt* <u>*five*</u> *questions in all, including Question No. 1 which is compulsory and selecting two questions from each Unit. Any missing or misprinted data may be assumed suitably.*

x-x-x

I. Attempt the following:-

a. How measures of central tendency help to study a dataset?

b. Define and explain Holistic measure.

c. List and elaborate parallel RDBMS features.

d. What is/are the importance of *apriori property* ?

e. Compare structured Vs Unstructured data and their effects of any Data Mining process.

(5x2)

UNIT -I

II. a) Is there any need of Data warehouse when we already have RDBMS? Justify your answer. [5]

b) *An OLTP system is customer-oriented and An OLAP system is market-oriented.* Explain. [5]

III. a) List and explain ETL tools used in Data Mining? [5]

b) Discuss architecture of data warehouse. [5]

IV. a) *"A data cube is a lattice of cuboids."* Discuss with the help of suitable example. [5]

b) Refer a sample Task relevant data set as shown in below figure. (Imagine it as data of an electronic store dealing with items related to computers and accessories. You can expand the dataset while keeping in mind the items usually available in such kind of store). [5]

Draw Concept Hierarchy diagram for *items* in the dataset.

Task-Relevant Data, D	
TID	**Items Purchased**
T100	Apple 17″ MacBook Pro Notebook, HP Photosmart Pro b9180
T200	Microsoft Office Professional 2010, Microsoft Wireless Optical Mouse 5000
T300	Logitech VX Nano Cordless Laser Mouse, Fellowes GEL Wrist Rest
T400	Dell Studio XPS 16 Notebook, Canon PowerShot SD1400
T500	Lenovo ThinkPad X200 Tablet PC, Symantec Norton Antivirus 2010
...	...

P.T.O.

UNIT - II

V. a) List and explain various data mining applications? Does mining data helped during COVID-19? [5]

b) What are the steps involved in data pre-processing? Explain in brief. [5]

VI. A database has five transactions. Let min sup = 60% and min confidence = 80%. [10]

TID	items_bought
T100	{M, O, N, K, E, Y}
T200	{D, O, N, K, E, Y }
T300	{M, A, K, E}
T400	{M, U, C, K, Y}
T500	{C, O, O, K, I ,E}

Find all frequent item sets using FP Growth algorithm.

Also, Explain why FP Growth can perform better than Apriori on some parameters.

VII. a) Perform a class comparison on the following. Suppose that you would like to compare the general properties of the *graduate* and *undergraduate* students at Panjab University, given the attributes **name, gender, major, birth place, birth date, residence, phone#, and gpa**. (You can prepare sample database yourself. Assume target and contrasting classes yourself, Also Assume missing data, if any, yourself). [6]

b) What is Embedded Data Mining? In what scenario this mining may be used? [4]

x-x-x