

Con. 5770-09.

B.E. (COMP) Sem VIII (R)

SP-6857

Data Warehousing & Mining

17/12/09

(3 Hours)

[Total Marks : 100

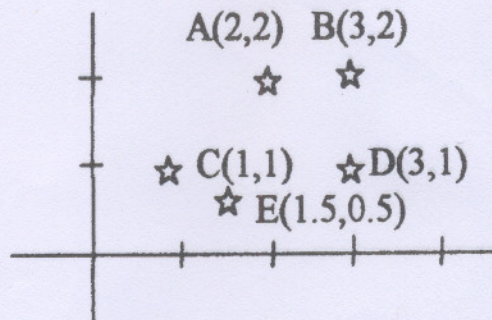
- N.B.** (1) Question No.1 is compulsory.
 (2) Attempt any **four** questions out of the remaining **six** questions.
 (3) **Figures** to the **right** indicate **full** marks.

1. (a) A manufacturing company has a huge sales network. To control the sales, it is divided in the regions. Each region has multiple zones. Each zone has different cities. Each sales person is allocated different cities. The object is to track sales figure at different granularity levels of region. Also to count no. of products sold. Create data warehouse schema to take into consideration of above granularity levels for region, sales person and the quarterly, yearly and monthly sales. **10**
- (b) Consider the transaction database in the **Appendix B**. Use Apriori Algorithm with minimum support of 30% and Confidence of 75% **10**

Appendix B

TID	Items
01	A,B,D
02	B,C,D
03	A,B
04	B,D
05	A,B,C

2. (a) Use data set in **Appendix A**. Create adjacency matrix. Use single link or complete link algorithm to cluster given data set. Draw dendrogram. **(3 + 7)**
- (b) With neat diagram explain the process of KDD giving emphasis on SELECTION pre-processing phase. **10**
3. (a) Use data set in **Appendix A**. Use K-means algorithm to create two clusters. **10**
- (b) List and describe five primitives for specifying data mining task. **10**

Appendix A

4. (a) What is web structure mining ? What are the techniques used for it ? What are Authorative and Hub pages. **10**
- (b) What are the applications of web usage mining ? What is web log ? Give typical structure of web log . **10**

[TURN OVER

Con. 5770-SP-6857-09.

2

5. (a) Explain Extraction Process in detail. 10
 (b) What is slowly changing dimension ? How this problem is solved ? Give examples. 10
6. (a) Define Classification. Explain K-nearest neighbour classification algorithm. 10
 (b) Using table in **Appendix C**. Create classification model using decision-tree and hence classify following tuple. 10

Appendix C

	Income	Age	Own House
1.	Very High	Young	Yes
2.	High	Medium	Yes
3.	Low	Young	Rented
4.	High	Medium	Yes
5.	Very High	Medium	Yes
6.	Medium	Young	Yes
7.	High	Old	Yes
8.	Medium	Medium	Rented
9.	Low	Medium	Rented
10.	Low	Old	Rented
11.	High	Young	Yes

7. Write short notes on any **two** of the following :— 20
- (a) Information Package Diagram
 (b) Project Planning and management of data warehouse
 (c) OLAP Servers.