

Total No. of Questions : 8]

SEAT No. :

[Total No. of Pages : 3

**P2355**

**[4937]-2003**

**M.Sc.**

**COMPUTER SCIENCE**

**CS - 203 : Data Mining and Data Warehousing**

**(2013 Pattern) (Semester - II)**

*Time : 3 Hours]*

*[Max. Marks : 50*

*Instructions to the candidates:*

- 1) *Answer any five questions.*
- 2) *Neat diagrams must be drawn wherever necessary.*
- 3) *Figures to the right indicate full marks.*

**Q1)** a) “In association rule mining FP tree algorithm is more efficient than Apriori”  
Comment. [4]

b) What is Web Mining? Write short note on web taxonomy. [4]

c) What is summerization? [2]

**Q2)** a) What are social implications of data mining? [4]

b) What are issues in decision tree classification? [4]

c) Write any two applications of Graph Mining. [2]

**Q3)** a) Perform Apriori algorithm to find frequent patterns with minimum support = 2  
[4]

Tid	Items
1	A,B,C D
2	A,B,D
3	A,D
4	A,C
5	B,C
6	B,D
7	A,C,D

- b) Write note on linear classifiers. [4]
- c) What is pattern matching? [2]

**Q4)** a) Suppose a data warehouse for Big University Consists of following dimensions

- Student
- Course
- Semester
- Instructor

The measures considered are score and average grade.

Draw star schema diagram for the data ware house. [4]

- b) Explain tree pruning with suitable example. [4]
- c) Write formulas to calculate sensitivity and specificity. [2]

**Q5)** a) What are advantages of having data warehouse? Draw suitable diagram of architecture of data warehouse. [4]

- b) Write note on K means clustering. [4]
- c) What is use of dimensionality reduction? [2]

**Q6)** a) What are issues considered during data integration? Explain binning with suitable example. [4]

- b) Explain following terms [4]
  - Confusion Matrix
  - F measure
- c) What is divisive clustering. [2]

**Q7) a)** Explain text mining and its applications. **[5]**

**b)** Consider the database given below: **[5]**

Patient AGE	Disease	Sugar level	Survival chances
Small	serious	high	yes
Medium	normal	low	yes
Senior	life time	normal	yes
Small	life time	high	no
Small	normal	high	yes
Senior	serious	normal	no
Medium	serious	low	yes
Senior	normal	low	no
Medium	life time	normal	yes
Medium	Serious	high	No
Senior	normal	low	no

Find out class label of the given tuple using Bayesian Classification.

$\langle \text{age : Senior, Disease, normal, Sugar level - normal} \rangle$ .

**Q8) a)** Explain support vector machine for linearly inseparable data. **[5]**

**b)** Explain the terms: **[5]**

i) Tree Mining

ii) Sequence Mining.

Write applications of frequent subgraph mining.

