

|  |  |  |  |  |  |  |  |  |  |
|--|--|--|--|--|--|--|--|--|--|
|  |  |  |  |  |  |  |  |  |  |
|--|--|--|--|--|--|--|--|--|--|

## Seventh Semester B.E. Degree Examination, June/July 2011

### Data Mining

Time: 3 hrs.

Max. Marks:100

**Note: Answer any FIVE full questions, selecting at least TWO questions from each part.**

#### PART – A

1. a. What is data mining? List the four core data mining tasks. (05 Marks)  
 b. Explain the knowledge discovery in databases (KDD) with neat diagram. (05 Marks)  
 c. For the following vectors, x and y, calculate the cosine, correlation, Euclidian and Jaccord similarity :  $x = (0, 1, 0, 1)$ ,  $y = (1, 0, 1, 0)$ . (10 Marks)
2. a. List all the data preprocessing techniques and in short explain any five. (10 Marks)  
 b. Consider a training set that contains 60 positive examples and 100 negative examples. For each of the following candidate rules,  
 Rule  $R_1$  : Covers 50 positive examples and 5 negative examples,  
 Rule  $R_2$  : Covers 2 positive examples and no negative examples,  
 determine which is the best and worst candidate rule according to, i) Rule accuracy  
 ii) FOIL's information gain iii) The likelihood ratio statistic. (10 Marks)
3. a. Explain in detail, the challenges that motivated the development of data mining. (06 Marks)  
 b. What is sampling? Explain simple random sampling v/s stratified sampling v/s progressive sampling. (10 Marks)  
 c. What is association analysis? Explain in brief, with an example. (04 Marks)
4. a. Explain DB scan algorithm with it's time and space complexity. (10 Marks)  
 b. For the following transaction data, construct the FP tree with explanation. (10 Marks)

| TID   | 1     | 2       | 3         | 4       | 5       | 6         | 7   | 8       | 9       | 10      |
|-------|-------|---------|-----------|---------|---------|-----------|-----|---------|---------|---------|
| Items | {a,b} | {b,c,d} | {a,c,d,e} | {a,d,e} | {a,b,c} | {a,b,c,d} | {a} | {a,b,c} | {a,b,d} | {b,c,e} |

#### PART – B

5. a. Explain the following ordered data with examples,  
 i) Sequential data ii) Sequence data iii) Time sequence data iv) Spatial data (08 Marks)  
 b. What is a priori principle? Explain (show) it itemset lattice. (08 Marks)  
 c. Explain basic K-means algorithm. (04 Marks)
6. a. What is cluster analysis? Explain different types of clustering in detail. (10 Marks)  
 b. List and explain any five features used to assess the data mining systems. (05 Marks)  
 c. Name any five commercial data mining systems. (05 Marks)
7. a. Explain text mining approaches and text indexing techniques in brief. (10 Marks)  
 b. Explain the social impact of data mining. (10 Marks)
8. Write a short note on the following:
  - a. Curse of dimensionality. (05 Marks)
  - b. Three standard approaches to feature selection. (05 Marks)
  - c. Foil's information gain. (05 Marks)
  - d. Spatial data mining. (05 Marks)