

--	--	--	--	--	--	--	--	--	--

Seventh Semester B.E. Degree Examination, May/June 2010
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 classification? How is data mining related to business intelligence? (08 Marks)
b. Discuss the issues to be considered during data integration. (06 Marks)
c. List any three application areas of data mining. (06 Marks)
- 2 a. What is data quality? What are the dimensions that assess the data quality? (08 Marks)
b. Distinguish between:
i) Noise and outliers.
ii) Jaccard coefficient and SMC
iii) Discretization and binarization. (12 Marks)
- 3 a. Explain the cosine similarity for calculating the similarity of two documents, with an example. (10 Marks)
b. Explain why attribute relevance analysis is needed. How can it be performed using entropy and gini index? (10 Marks)
- 4 a. Define a frequent pattern tree. Discuss the method of computing a FP-tree, with an algorithm. (10 Marks)
b. Give an example to show that items in a strong association rule may actually be negatively correlated. (10 Marks)

PART – B

- 5 a. A database has five transactions. Let min-sup = 60% and min-conf = 80%.

TID	Items bought
T1	{M, O, N, K, E, Y}
T2	{D, O, N, K, E, Y}
T3	{M, A, K, E}
T4	{M, U, C, K, Y}
T5	{C, O, O, K, I, E}

Find all frequent item sets using Apriori and FP growth respectively. (14 Marks)

b. Explain the method of candidate generation and proving, with an example. (06 Marks)
- 6 a. Explain the K-means clustering method. (10 Marks)
b. Is data missing a threat to privacy and data security? Discuss. (10 Marks)
- 7 a. Discuss the social impacts of data mining. Discuss spatial data mining. (10 Marks)
b. What is a multimedia database? Explain the methods of mining multimedia database. (10 Marks)
- 8 Write short notes on: (20 Marks)
 - a. Outlier analysis
 - b. Nearest neighbor classifier
 - c. DBSCAN
 - d. Skewed support distribution.

