

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

Fifth Semester MCA Degree Examination, December 2011

Data Mining

Time: 3 hrs.

Max. Marks:100

Note: Answer any FIVE full questions.

1.
 - a. Explain data mining. Describe the knowledge discovery process in data mining. (05 Marks)
 - b. Discuss the different types of data sets in data mining. (10 Marks)
 - c. Explain the various types of data mining tasks. (05 Marks)
2.
 - a. Define data preprocessing. Mention the approaches of preprocessing data. (05 Marks)
 - b.
 - i) Calculate SMC and Jaccard similarity for the binary vectors :
 $X = (1, 0, 0, 0, 0, 0, 0, 0, 0, 0)$ and $Y = (0, 0, 0, 0, 0, 0, 1, 0, 0, 1)$
 - ii) Calculate cosine similarity for the data objects :
 $X = (3, 2, 0, 5, 0, 0, 0, 2, 0, 0)$ and $Y = (1, 0, 0, 0, 0, 0, 0, 1, 0, 2)$ (10 Marks)
 - c. Discuss the dissimilarities between data objects. (05 Marks)
3.
 - a. Explain classification process. Write algorithm for decision tree induction technique. (10 Marks)
 - b. Explain the characteristics of rule base classifier. (05 Marks)
 - c. Write an algorithm for K-nearest neighbor classifier and explain. (05 Marks)
4.
 - a. Write apriori algorithm for finding frequent item sets. Find the frequent pattern generated using apriori for the following set and transactions : (10 Marks)

TID	List of item - IDs
T ₁₀₀	L ₁ , L ₂ , L ₅
T ₂₀₀	L ₂ , L ₄
T ₃₀₀	L ₂ , L ₃
T ₄₀₀	L ₁ , L ₂ , L ₄
T ₅₀₀	L ₁ , L ₃
T ₆₀₀	L ₂ , L ₃
T ₇₀₀	L ₁ , L ₃
T ₈₀₀	L ₁ , L ₂ , L ₃ , L ₅
T ₉₀₀	L ₁ , L ₂ , L ₃

- b. Draw F-P tree for the given set of transactions. Find all frequent item sets using F-P growth. (05 Marks)

TID	List of item - IDs
T ₁₀₀	{M, O, N, K, E, Y}
T ₂₀₀	{D, O, N, K, E, Y}
T ₃₀₀	{M, A, K, E}
T ₄₀₀	{M, U, C, K, Y}
T ₅₀₀	{C, O, O, K, I, E}

- c. Write a note on alternate methods for generating frequent item sets. (05 Marks)
5.
 - a. Write a note on evaluation of association patterns. (05 Marks)
 - b. Explain the effect of skewed support distribution. (05 Marks)
 - c. Describe the ripper algorithm. (05 Marks)
 - d. Explain merging sequence criteria, with an example. (05 Marks)

Important Note : 1. On completing your answers, compulsorily draw diagonal cross lines on the remaining blank pages.
2. Any revealing of identification, appeal to evaluator and /or equations written eg, 42+8 = 50, will be treated as malpractice.

- 6** a. Describe cluster analysis. Briefly explain different types of clustering. (10 Marks)
b. Write algorithm for density based clustering and explain in brief. (10 Marks)
- 7** a. Explain data mining for biomedical and DNA data analysis. (10 Marks)
b. Briefly describe the various trends in data mining. (10 Marks)
- 8** Write short notes on the following :
a. Origins of data mining
b. Association rule mining
c. Dimensionality reduction
d. Spatial data mining. (20 Marks)

* * * * *