

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

Seventh Semester B.E. Degree Examination, June / July 2014
Data Warehouse & Data Mining

Time: 3 hrs.

Max. Marks:100

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

PART – A

- 1 a. Explain ODS (Operational Data Store) and its structure with a neat figure. (07 Marks)
 b. Explain the implementation steps for data warehouse. (07 Marks)
 c. Write the differences between OLTP and data warehouse. (06 Marks)
- 2 a. Explain the characteristics of OLAP systems and write the comparison of OLTP and OLAP. (12 Marks)
 b. Explain ROLAP and MOLAP. (08 Marks)
- 3 a. Explain four types of attributes with statistical operations and examples. (06 Marks)
 b. Explain the steps applied in data preprocessing. (10 Marks)
 c. Two binary vectors are given below:
 $X = (1, 0, 0, 0, 0, 0, 0, 0, 0, 0)$
 $Y = (0, 0, 0, 0, 0, 0, 1, 0, 0, 1)$
 Calculate (i) SMC (ii) Jaccard similarity coefficient and hamming distance. (04 Marks)
- 4 a. Consider the following transaction data set 'D' shows 9 transactions and list of items using Apriori algorithm find frequent itemset min-support = 2. (10 Marks)

Tid	T ₁	T ₂	T ₃	T ₄	T ₅	T ₆	T ₇	T ₈	T ₉
List of items	I ₁ , I ₂ , I ₅	I ₂ , I ₄	I ₂ , I ₃	I ₁ , I ₂ , I ₄	I ₁ , I ₃	I ₂ , I ₃	I ₁ , I ₃	I ₁ , I ₂ , I ₃ , I ₅	I ₁ , I ₂ , I ₃

- b. For the following transaction data set table construct an FP-tree and explain stepwise for all the transactions. (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. Define classification. Draw a neat figure and explain general approach for solving classification model. (06 Marks)
 b. Mention the three impurity measures for selecting best splits. (04 Marks)
 c. Consider a training set that contains 60 +ve examples and 100 -ve examples, for each of the following candidate rules.
 Ruel r₁: Covers 50 +ve examples and 5 -ve examples.
 Ruel r₂: Covers 2 the examples and No -ve examples.
 Determine which is the best and worst candidate rule according to,
 i) Rule accuracy
 ii) Likelihood ratio statistic.
 iii) Laplace measure. (10 Marks)

- 6 a. For the given confusion matrix below for three classes. Find sensitivity and specificity metrics to estimate predictive accuracy of classification methods. (10 Marks)

Predicted class	True class		
	1	2	3
1	8	1	1
2	2	9	2
3	0	0	7

Table : Confusion matrix for three classes

- b. Explain with example the two approaches for extending the binary classifiers to handle multiclass problem. (10 Marks)
- 7 a. Explain K means clustering method and algorithm. (10 Marks)
- b. What is Hierarchical clustering method? Explain the algorithms for computing distances between clusters. (10 Marks)
- 8 Write short notes on the following:
- Web content mining.
 - Text mining.
 - Spatial data mining.
 - Spatio temporal data mining.
- (20 Marks)

* * * * *