USN | | | | | | | | | |

PES

PES Institute of Technology, Bangalore

(Autonomous Institute under VTU, Belgaum)

10IS 402

SEMESTER END EXAMINATION (SEE) B. E. 7TH SEMESTER –Dec. 2013

10IS 402 - DATA WAREHOUSING AND DATA MINING

Time: 3 Hrs | Answer All Questions | Max Marks: 100

1.	a)	How is data warehouse different from database? How are they similar?	5
	b)	List the issues of Data Mining. Explain about mining dynamic, networked and global data repositories.	2+3
	c)	The test scores in Mathematics of 12 students are as below. Compute the Five Number Summary and draw the Box Plot for the scores 70,80,73,69,88,100,79,77,71,85,65,75.	5
	d)	You are given the speed of 5 cars (in Kmph) as 35, 36,46,68,70. Normalize this group of data by (i) min-max normalization by setting min=0 and max= 1 (ii) z-score normalization. Which of these give a better representation?	2+2+1
2.	a)	What are the general strategies for Data Cube Computation? Describe any two of them.	5
	b)	Write the star schema for IPL6 cricket taking into account the spectator, location, game, date for the centralized sales table. Starting with the base cuboid [date, spectator, location, game] what specific OLAP operations should one perform in order to get the total charges paid by spectators of Black Dog Pavilion at Chinnaswamy Stadium in 2013?	5+4
	c)	Describe Enterprise Warehouse, Data Mart and Virtual Warehouse.	6
3.	a)	What is Market Basket Analysis? Explain by means of an example	2+4
	b)	The items purchased by customers in GK Vale Co. are as per the following transactions. TID: T100 — Camera, Lens, Binoculars T200 — Lens, Tripod, Camera Bag T300 — Lens, Flash, Filter T400 — Camera, Lens, Flash, Screen Guards T500 — Camera, Lens, Tripod, Lens Cleaner The minimum support = 30%. You may use the alphabets B for Binoculars, C for Camera, F for Flash, L for Lens, T for Tripod, CB for Camera Bag, FI for Filter, SG for Screen Guards and LC for Lens Cleaner. Generate the Frequent Itemsets using Apriori algorithm. 3 Write the FP Tree and corresponding conditional pattern bases along with conditional FP Tree. Compare the efficiency of the two mining processes. 7	
	c)	What are Multi level Association Rule and Multi Dimensional Association Rule? Explain by means of examples.	4
4	a)	What is Bayesian Belief Networks? Explain the concept by means of an example.	5
4	b)	The opinions of seven persons have been taken for an item. This is in terms of the attributes namely size, colour and shape, as per the following table.	

Num	Size	Colour	Shape	Satisfied
1	med	blue	brick	yes
2	small	red	wedge	no
3	small	red	sphere	yes
4	large	red	wedge	no
5	large	green	pillar	yes
6	large	red	pillar	no
7	large	green	sphere	yes

Based on the principle of Information Gain determine which attribute has to be used first for the split operation of the Decision Tree ($\log_2 3 = 1.58$ and $\log_2 7 = 2.81$) — 10

c) Describe Support Vector Machines (SVM) using the figures for Support Vectors and Maximum Margin. What is Kernel trick? — 5

5. a) Consider the following data set consisting of the scores of two variables on each of seven subjects in the range 1 t0 7.

Subject	A	B
1	1.0	1.0
2	1.5	2.0
3	3.0	4.0
4	5.0	7.0
5	3.5	5.0
6	4.5	5.0
7	3.5	4.5

This data set is to be grouped into two clusters. As a first step 1 and 4 define the initial cluster means. Using K – means algorithms generate the initial and final clusters. — 4+6

b) Describe briefly Social Network Model. — 4

c) Describe briefly Data Mining for Intrusion Detection with reference to Signature based detection and Anomaly based detection. — 6