USN | | | | | | | | | | |

PES

PES Institute of Technology, Bangalore
(Autonomous Institute under VTU, Belgaum)

CS 363

SEMESTER END EXAMINATION (SEE) B. E. 6th SEMESTER – MAY 2010

CS 363 – DATA MINING AND WAREHOUSING

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

1 a) What is Data Mining?

a) Describe briefly by means of a figure the architecture of a typical data mining system 2+

b) Suppose a hospital tested the age and body fat for 18 randomly selected adults with the following result:

age	23	23	27	27	39	41	47	49	50
%fat	9.5	26.5	7.8	17.8	31.4	25.9	27.4	27.2	31.2

age	52	54	54	56	57	58	58	60	61
%fat	34.6	42.5	28.8	33.4	30.2	34.1	32.9	41.2	35.7

(i) Calculate the mean, median, and standard deviation of age and %fat 5

(ii) Draw the box plots for age and %fat 5

2. a) As per Inmon's definition of Data Warehouse describe briefly the key features of a Data Warehouse

b) What are the Online Analytical Processing (OLAP) Servers? Describe them 1+

c) The Department of Information Science and Engineering will be conducting the Second International Conference on Multimedia and CBIR (ICMCBIR) in July 2010. A data warehouse for this has to be constructed catering for the dimensions Time, Delegates, Papers and Accommodation. Draw schema diagram for this data warehouse. The fact table has to contain information on the count and the amount_received.

(d) What is Star Cubing?

3. a) What is Market Basket Analysis ? Give an example. 2+

b) A database has six transactions of purchase of books from a bookshop as given
t_1 = {ANN, CC, TC,CG} , t_2 = {CC, D ,CG} t_3 = {ANN, D, CC, TC},
t_4 = {ANN, CC, D,CG }, t_5 = { ANN, CC, D, TC,CG}, t_6 = {CC, D ,TC}
Let X= {CC, TC} and Y= { ANN,TC,CC}
Find the confidence and support of the association rule X -> Y and inverse rule Y-> X

c) A 2 X 2 contingency table summarizing the transactions w.r.t to game and video purchases

	game	$\overline{\text{game}}$	Σ_{row}
video	4000	3500	7500
$\overline{\text{Video}}$	2000	500	2500
Σ_{column}	6000	4000	10000

Compute the expected values and χ^2 value

d) Give an example of uniform support 4

USN | | | | | | | | | | |

the basic principle of back propagation by means of a diagram 2+

the criteria for comparing and evaluating Classification and Prediction Methods?

table for X (years of experience) and Y (corresponding salary of engineers in ...ands) predict the salary of engineer with 15 years of experience using linear ... technique

1	20
3	36
6	43
8	57

...matically Bayes Theorem and briefly explain.

...NES and DIANA clustering algorithm.

...mal table where patients are described by binary attributes is given

gender	fever	cough	test-1	test-2	test-3	test-4
M	Y	Y	P	N	N	N
	Y	N	P	N	P	N
M	Y	Y	N	N	N	N

...e the distance between the each pair of the three patients Ram, Riya and

...objects represented by tuples (22, 1, 42, 10) and (20, 0, 36, 8)

...te the following

Euclidean distance between the two objects.

Manhattan distance between the two objects.

Minkowski distance between the two objects, using the power q=3

...briefly any one data mining application