PES

PES Institute of Technology, Bangalore

(Autonomous Institute under VTU, Belgaum)

09IS 402

SEMESTER END EXAMINATION (SEE) B. E. 7TH SEMESTER –Dec. 2012

09IS 402 - DATA WAREHOUSING AND DATA MINING

Time: 3 Hrs Answer All Questions Max Marks: 100

1. a) What is Data Mining? 2

b) Describe briefly by means of a figure the architecture of a typical data mining system. 2+6

c) 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 4

b) What are the Online Analytical Processing (OLAP) Servers? Describe them. 1+4

c) The Department of Information Science and Engineering will be conducting a 3 day Faculty Development Program (FDP) in connection with Alan Turing's birth centenary, during December 18-20, 2012. A data warehouse for this has to be constructed catering for the dimensions Time, Delegates, Research_Papers, Keynote_Speakers and Accommodation. Draw schema diagram for this data warehouse. Write a query to get the count of delegates for FDP fact table and total fee collected for attending the FDP. 6+2

d) What is Star Cubing? 3

3. a) What is Market Basket Analysis? Explain by means of an example 2+4

b) A database has five transactions. Let min_sup = 60 % and min_conf= 80%.

TID	Items bought
T100	{M O, N,K,E,Y}
T200	{D ,O,N,K,E,Y}
T300	{M ,A,K,E}
T400	{M, U,C,K,Y}
T500	{C ,O,O,K,I,E}

Find all frequent itemsets(individual alphabets) using Apriori and FP growth, respectively. Compare the efficiency of the two mining processes. 10

c) What are Multi level Association Rule and Multi Dimensional Association Rule? Explain by means of examples. 4

4 a) What is supervised learning and unsupervised learning? 4

b) You are given a set of 10 examples for cars of different colors, types and origin. Determine whether the car with Color = "Red" , type= "Sports" and Origin= "Domestic" is considered for buying or not using Naïve Bayesian approach

Color	Type	Origin	Buying?
Red	Sports	Domestic	Yes
Red	Sports	Domestic	No
Red	Sports	Domestic	Yes
Yellow	Sports	Domestic	No
Yellow	Sports	Imported	Yes
Yellow	SUV	Imported	No
Yellow	SUV	Imported	Yes
Yellow	SUV	Domestic	No
Red	SUV	Imported	No
Red	Sports	Imported	Yes

6

c) Describe Support Vector Machines (SVM) using the figures for Support Vectors and Maximum Margin. 5

d) What is a lazy learner? Describe briefly k-Nearest Neighbor classifier. 5

5. a) For the following points $A_1(2, 10)$,$A_2(2, 5)$,$A_s(8, 4)$,$B_1(5, 8)$,$B_2(7, 5)$, $B_3(6, 4)$, $C_1(1, 2)$,$C_2(4, 9)$, suppose initially we assign A_1, B_1, and C_1 as the center of each cluster, respectively. Use the k-means algorithm to show *only* the three cluster centers after the first round of execution. 4

b) Describe Agglomerative and Divisive Hierarchical Clustering using suitable figure. 2+

c) What is Statistical Data Mining? Write briefly any one application. 4

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