

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

09IS 402

10

4

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

09IS 402 - DATA WARFHOUSING AND DATA MINING

ne: 3		310 402	- DATA		r All Que		ID DA	I A IVIIIV	IIIVG	Max Mark	ks: 100
a)	What is Data M	ining?									2
b)	Describe briefly by means of a figure the architecture of a typical data mining system. Suppose a hospital tested the age and body fat for 18 randomly selected adults with the										2+6
	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 %fat 34.6	54 42.5	54 28.8	56 33.4	57 30.2	58 34.1	58 32.9	60	61 35.7		
	(i) Calculat	te the mea	n, media	an, and s	tandard o					_	5 5
a)	As per Inmon's Warehouse	definition	of Data	a Wareho	ouse desc	cribe br	iefly the	key fe	atures o	f a Data	4
b)	What are the O	nline Ana	lytical P	rocessin	g (OLAF) Serve	rs? Des	cribe th	em.		1+4
c)	Development I December 18-2 dimensions Tir	Program (20, 2012. ne, Deleg	FDP) in A data gates, R	connective warehousesearch_	ction with use for the Papers,	h Alan nis has Keynot	Turing to be contacted to be c	g's birtl onstruc kers an	n center ted cate d Acco	nary, during ring for the mmodation.	
								Bott		or merogenes	6+2
d)	What is Star Cu	ibing?									3
a)	What is Market E	Basket Ana	lysis? Ex	plain by r	means of	an exan	nple				2+4
b)		TID T100 T200 T300 T400	actions.	Let min Items bo {M O, N {D ,O,N {M ,A,K {M, U,C	sup = 6 ught (,K,E,Y) ,K,E,Y) (,E) (,K,Y)			conf= 80	0%.		
	a) b) c) a) d) a)	b) Describe briefly c) Suppose a hosp following result age 23 %fat 9.5 age 52 %fat 34.6 (i) Calculat (ii) Draw th a) As per Inmon's Warehouse b) What are the Octor The Department Development In December 18-2 dimensions Timent Draw schema of for FDP fact tale d) What is Star Cutal	a) What is Data Mining? b) Describe briefly by means. c) Suppose a hospital tested following result: age 23 23 %fat 9.5 26.5 age 52 54 %fat 34.6 42.5 (i) Calculate the mean (ii) Draw the box plot graph was a series of the control of the contro	a) What is Data Mining? b) Describe briefly by means of a fige following result: age 23 23 27 % fat 9.5 26.5 7.8 age 52 54 54 % fat 34.6 42.5 28.8 (i) Calculate the mean, media (ii) Draw the box plots for age warehouse b) What are the Online Analytical PCO The Department of Information SCO Development Program (FDP) in December 18-20, 2012. A data dimensions Time, Delegates, RCO Draw schema diagram for this differ FDP fact table and total fee conditions of the difference	a) What is Data Mining? b) Describe briefly by means of a figure the solution of Suppose a hospital tested the age and body following result: age 23 23 27 27 27 26.5 7.8 17.8 age 52 54 54 56 28.8 33.4 (i) Calculate the mean, median, and solution of Data Warehouse	a) What is Data Mining? b) Describe briefly by means of a figure the architect c) Suppose a hospital tested the age and body fat for following result: age 23 23 27 27 39 %fat 9.5 26.5 7.8 17.8 31.4 age 52 54 54 56 57 %fat 34.6 42.5 28.8 33.4 30.2 (i) Calculate the mean, median, and standard (ii) Draw the box plots for age and %fat a) As per Inmon's definition of Data Warehouse described Warehouse b) What are the Online Analytical Processing (OLAF C) The Department of Information Science and Engin Development Program (FDP) in connection with December 18-20, 2012. A data warehouse for the dimensions Time, Delegates, Research_Papers, Draw schema diagram for this data warehouse. We for FDP fact table and total fee collected for attended What is Star Cubing?	a) What is Data Mining? b) Describe briefly by means of a figure the architecture of a c) Suppose a hospital tested the age and body fat for 18 rand following result: age 23 23 27 27 39 41 %fat 9.5 26.5 7.8 17.8 31.4 25.9 age 52 54 54 56 57 58 %fat 34.6 42.5 28.8 33.4 30.2 34.1 (i) Calculate the mean, median, and standard deviation (ii) Draw the box plots for age and %fat a) As per Inmon's definition of Data Warehouse describe by Warehouse b) What are the Online Analytical Processing (OLAP) Serve (C) The Department of Information Science and Engineering Development Program (FDP) in connection with Alan December 18-20, 2012. A data warehouse for this has dimensions Time, Delegates, Research_Papers, Keynoth Draw schema diagram for this data warehouse. Write a confor FDP fact table and total fee collected for attending the downwarehouse with the collected for attending th	a) What is Data Mining? b) Describe briefly by means of a figure the architecture of a typical c) Suppose a hospital tested the age and body fat for 18 randomly s following result: age 23 23 27 27 39 41 47 %fat 9.5 26.5 7.8 17.8 31.4 25.9 27.4 age 52 54 54 56 57 58 58 %fat 34.6 42.5 28.8 33.4 30.2 34.1 32.9	a) What is Data Mining? b) Describe briefly by means of a figure the architecture of a typical data m Suppose a hospital tested the age and body fat for 18 randomly selected following result: age 23 23 27 27 39 41 47 49 % fat 9.5 26.5 7.8 17.8 31.4 25.9 27.4 27.2 age 52 54 54 56 57 58 58 60 % fat 34.6 42.5 28.8 33.4 30.2 34.1 32.9 41.2 (i) Calculate the mean, median, and standard deviation of age and % (ii) Draw the box plots for age and % fat As per Inmon's definition of Data Warehouse describe briefly the key fe Warehouse Warehouse What are the Online Analytical Processing (OLAP) Servers? Describe the Warehouse Development Program (FDP) in connection with Alan Turing's birth December 18-20, 2012. A data warehouse for this has to be constructed dimensions Time, Delegates, Research_Papers, Keynote_Speakers and Draw schema diagram for this data warehouse. Write a query to get the for FDP fact table and total fee collected for attending the FDP. d) What is Market Basket Analysis? Explain by means of an example b) A database has five transactions. Let min_sup = 60 % and min_conf= 80 TID	a) What is Data Mining? b) Describe briefly by means of a figure the architecture of a typical data mining sy c) Suppose a hospital tested the age and body fat for 18 randomly selected adults w following result: age 23 23 27 27 39 41 47 49 50 96fat 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 96fat 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 (ii) Draw the box plots for age and %fat (ii) Draw the box plots for age and %fat As per Inmon's definition of Data Warehouse describe briefly the key features of Warehouse Warehouse	a) What is Data Mining? b) Describe briefly by means of a figure the architecture of a typical data mining system. 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 (ii) Draw the box plots for age and %fat (ii) Draw the box plots for age and %fat As per Inmon's definition of Data Warehouse describe briefly the key features of a Data Warehouse What are the Online Analytical Processing (OLAP) Servers? Describe them. 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. What is Star Cubing? What is Market Basket Analysis? Explain by means of an example A database has five transactions. Let min sup = 60 % and min_conf= 80%. TID

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

- What are Multi level Association Rule and Multi Dimensional Association Rule? Explain by means of examples.
- What is supervised learning and unsupervised learning?

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

USN									
-----	--	--	--	--	--	--	--	--	--

6

2+

for buying or not using Naïve Bayesian approach

Color	Туре	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

- c) Describe Support Vector Machines (SVM) using the figures for Support Vectors and Maximum Margin.
- d) What is a lazy learner? Describe briefly k-Nearest Neighbor classifier.
- 5. a) For the following points A₁(2, 10),A₂(2, 5),A_s(8, 4),B₁(5, 8),B₂(7, 5), B₃(6, 4), C₁(1, 2),C₂(4, 9), suppose initially we assign A₁, B₁, and C₁ 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.
 - b) Describe Agglomerative and Divisive Hierarchical Clustering using suitable figure.
 - c) What is Statistical Data Mining? Write briefly any one application.
 - d) Describe briefly Data Mining for Intrusion Detection with reference to Signature based detection and Anomaly based detection.