Total No. of Printed Pages:02

SUBJECT CODE NO:- H-306 FACULTY OF SCIENCE AND TECHNOLOGY B.E. (CSE/IT) Data Worshawing & Data Mining

Data Warehousing & Data Mining (REVISED)

[Time: Three Hours] [Max. Marks: 80]

Please check whether you have got the right question paper.

N.B

- 1. Q.1. and Q.6. are compulsory.
- 2. Solve any two from questions 2,3,4,5 and any two from question 7, 8, 9, 10.

Section A

- Q.1 a) Define Data warehouse? Explain the need for data warehouse. 03 b) What is KDD? Explain various components of KDD. 03 c) Explain various steps in data cleaning. 04 Q.2 a) What is an attribute? Explain all types of attributes. 07 b) What is data mining? What kind of patterns can be mined with data mining? Explain in detail. 08 Q.3 a) Consider the following data: 36, 30,50,47,52,56,52,60. Find mean, median, mode, variance 07 and standard deviation. b) Explain proximity measure for nominal, binary and ordinal attributes with example. 08
- Q.4 a) Consider the following 2-D dataset

200	A_1	A_2
X_1	1.5	
X_2	2000	1.9
X_3	1.6	1.8
X_4	1.2	1.5
X_5	1.5	1.0

Compute Euclidean distance, Manhattan distance, minkowski distance & cosine similarity Between A_1 and A_2 .

b) What are the different method for handling missing values in the tuples.

07

08

08

Q.5 a) What is OLAP? Explain various operation that can be performed on data cube.

b) Find chi square value for following contingency table for two attributes gender & reading type 07

1. 12. 12. 13. 15. 15. 15. 15. 15. 15. 15. 15. 15. 15	Male	Female	Total
Action	300	400	700
Non-fiction	100	800	900
90000	400	1200	1600

Examination NOV/DEC 2018

H-306

a)	Consider the	e following datas	set with 7 trans	sactions. (Let min-su	up=60% and min-coef=80%)
ĺ	Transaction	Item-bought		25 40 X 20	
ŀ	101	I_1, I_2, I_5		200 9 7 9 VOV	
	102	I_2,I_4			
	103	I_2,I_3		10 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	2, 10, 2, 20, 20, 21, 21, 20, 21, 21, 21, 21, 21, 21, 21, 21, 21, 21
	104	I_1, I_2, I_4			
	105	I_1, I_2, I_3	200		
	106	I_1,I_3	877		1 1 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2
	107	I_1,I_2			
-	1) Find	all frequent item	cot with MSC	(2) \ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \	
	1) Fillu	an nequent nen	i set with Misc	(2 (4)	7/ , QV 7/ , YN 702, YED Y) , QN
		the strongest ass			
a)	2) Find Explain rule Consider the (2,4,8,10,12 And K=2 the	the strongest ass based classifier e following data ,3,20,30,11,13,2 en determine tw	ociation rules. with example item in cluster 5) o cluster using	. K-mean clustering.	
a)	2) Find Explain rule Consider the (2,4,8,10,12 And K=2 the	the strongest ass based classifier e following data ,3,20,30,11,13,2	ociation rules. with example item in cluster 5) o cluster using	. K-mean clustering.	
a)	2) Find Explain rule Consider the (2,4,8,10,12 And K=2 the Draw and ex	the strongest ass based classifier e following data ,3,20,30,11,13,2 en determine two splain BI archite	ociation rules. with example item in cluster 5) o cluster using cture with all i	K-mean clustering. ts components.	nd find decision tree
a) b)	2) Find Explain rule Consider the (2,4,8,10,12 And K=2 the Draw and ex	the strongest ass based classifier e following data ,3,20,30,11,13,2 en determine two splain BI archite	ociation rules. with example item in cluster 5) o cluster using cture with all i	K-mean clustering. ts components.	nd find decision tree Post
a) b)	2) Find Explain rule Consider the (2,4,8,10,12 And K=2 the Draw and example the details of the consideration of th	the strongest asset based classifier to be following data 1,3,20,30,11,13,2 then determine two splain BI archite the ecision tree (ID3 Experience 6	ociation rules. with example item in cluster 5) o cluster using cture with all i	K-mean clustering. ts components.	
a) b)	2) Find Explain rule Consider the (2,4,8,10,12 And K=2 the Draw and exply the de Name	the strongest asset based classifier e following data 1,3,20,30,11,13,2 en determine two explain BI archite ecision tree (ID3 Experience	ociation rules. with example item in cluster 5) o cluster using cture with all i	K-mean clustering. ts components. following dataset a Qualification	Post
a) b)	2) Find Explain rule Consider the (2,4,8,10,12) And K=2 the Draw and explain the dependence of the control of	the strongest asset based classifier to be following data 1,3,20,30,11,13,2 then determine two splain BI archite the ecision tree (ID3 Experience 6	ociation rules. with example item in cluster 5) o cluster using cture with all i	K-mean clustering. ts components. following dataset a: Qualification M.E	Post Associate
a) b)	2) Find Explain rule Consider the (2,4,8,10,12 And K=2 th Draw and explain the decomposition of the control of	the strongest assabased classifier e following data 3,20,30,11,13,2 en determine two explain BI archite ecision tree (ID3 Experience 6 5	ociation rules. with example item in cluster 5) o cluster using cture with all i 1) algorithm or Salary 70000 30000	K-mean clustering. ts components. following dataset a: Qualification M.E B.E	Post Associate Assistant
a) b)	2) Find Explain rule Consider the (2,4,8,10,12) And K=2 the Draw and explain the decomposition of the control	the strongest asset based classifier to be following data 1,3,20,30,11,13,2 ten determine two explain BI archites the ecision tree (ID3 Experience 6 5 6	ociation rules. with example item in cluster 5) o cluster using cture with all i algorithm or Salary 70000 30000 40000	K-mean clustering. ts components. following dataset a: Qualification M.E B.E B.E B.E	Post Associate Assistant Assistant
a) b)	2) Find Explain rule Consider the (2,4,8,10,12) And K=2 the Draw and explain rule Apply the de Name Manish Nisha Rutu Rahul	the strongest asses based classifier to be following data 1,3,20,30,11,13,2 to determine two explain BI archites the ecision tree (ID3 Experience 6 5 6 4	item in cluster 5) o cluster using cture with all i 3) algorithm or Salary 70000 30000 40000 50000	K-mean clustering, ts components. following dataset a: Qualification M.E B.E B.E M.E	Post Associate Assistant Assistant Assistant