

Total No. of Printed Pages:02

SUBJECT CODE NO:- H-306
FACULTY OF SCIENCE AND TECHNOLOGY
B.E. (CSE/IT)
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 and standard deviation. 07
 b) Explain proximity measure for nominal, binary and ordinal attributes with example. 08
- Q.4 a) Consider the following 2-D dataset 08

	A ₁	A ₂
X ₁	1.5	1.7
X ₂	2	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 A₁ and A₂.

- b) What are the different method for handling missing values in the tuples. 07
- Q.5 a) What is OLAP? Explain various operation that can be performed on data cube. 08
 b) Find chi square value for following contingency table for two attributes gender & reading type 07

	Male	Female	Total
Action	300	400	700
Non-fiction	100	800	900
	400	1200	1600

Section B

- Q.6 a) Define classification, prediction, clustering and regression. 05
 b) What is market basket analysis? Explain. 05

- Q.7 a) Consider the following dataset with 7 transactions. (Let min-sup=60% and min-coef=80%) 10

Transaction	Item-bought
101	I ₁ ,I ₂ ,I ₅
102	I ₂ ,I ₄
103	I ₂ ,I ₃
104	I ₁ ,I ₂ ,I ₄
105	I ₁ ,I ₂ ,I ₃
106	I ₁ ,I ₃
107	I ₁ ,I ₂

- 1) Find all frequent item set with MSC=2.
 2) Find the strongest association rules. 05
 b) Explain rule based classifier with example. 05

- Q.8 a) Consider the following data item in cluster. 08
 (2,4,8,10,12,3,20,30,11,13,25)
 And K=2 then determine two cluster using K-mean clustering.

- b) Draw and explain BI architecture with all its components. 07

- Q.9 a) Apply the decision tree (ID3) algorithm on following dataset and find decision tree 10

Name	Experience	Salary	Qualification	Post
Manish	6	70000	M.E	Associate
Nisha	5	30000	B.E	Assistant
Rutu	6	40000	B.E	Assistant
Rahul	4	50000	M.E	Assistant
Seema	5	35000	B.E	Assistant
Rushi	7	80000	M.E	Associate

- b) Explain business-process-response support model. 05

- Q.10 a) Explain linear and non-linear regression in detail. 08
 b) Write a note “transaction processing versus analytics processing.” 07