1.1.1 The Institution ensures effective curriculum delivery through a well planned and documented process

NEHRU INSTITUTE OF ENGINEERING AND TECHNOLOGY
An ISO 9001:2015 & 14001:2015 Certified Institution, Affiliated to Anna University, Chennai
Approved by AICTE, New Delhi, Recognized by UGC with 2(f) & 12(B)
Re-accredited by NAAC "A+", NBA Accredited (UG Courses: AERO & CSE)
Nehru Gardens, Thirumalayampalayam, Coimbatore – 641 105

Sample Course File

NEHRU INSTITUTE OF ENGINEERING AND TECHNOLOGY
An ISO 9001:2015 & 14001:2015 Certified Institution, Affiliated to Anna University, Chennai
Approved by AICTE, New Delhi, Recognized by UGC with 2(f) & 12(B)
Re-accredited by NAAC "A+", NBA Accredited (UG Courses: AERO & CSE)
Nehru Gardens, Thirumalayampalayam, Coimbatore – 641 105

Course file

Subject Name: Data warehousing and Data Mining

Subject Code: CS 8075

Department of Computer Science and Engineering

NEHRU INSTITUTE OF ENGINEERING AND TECHNOLOGY

Nehru Gardens, Thirumalayampalayam, Coimbatore – 641 105.
Approved by AICTE, New Delhi and Affiliated to Anna University, Chennai
An ISO 9001:2015 and ISO 14001:2004 Certified Institution
Accredited by NAAC and Recognized by UGC with Section 2(f) and 12(B)

Department of Computer Science & Engineering

2021 – 2022(EVEN)

Course Code/Name: CS8075 / Data Warehousing & Data Mining
Name of the faculty: JEEVANANTHAM G.
Year/Semester: 2021-2022 III / VI

COURSE FILE CONTENTS

1.	A & A book (at Top)	21.	Internal Test-II Marks Statement
2.	Question Paper Feedback Form	22.	Question paper for Internal Test – II
3.	CO/PO/PSO Attainment	23.	Review Questions for Internal Test-II
4.	University Exam Question Paper	24.	Remedial Class Details
5.	Consolidated Internal Marks and attendance Statement	25.	Micro Analysis for Test I
6.	Tutorial Sheets for the Subjects Applicable	26.	Assignment 1
7.	Course Post Analysis	27.	List of Students (A/B/C Categories)
8.	Remedial Class Details	28.	3 Sample Answer Scripts for internal test I (Above average, Average, Below Average)
9.	Micro Analysis for Test III	29.	Internal Test-I Marks Statement
10.	Assignment 3	30.	Question paper for Internal Test – I
11.	List of Students (A/B/C Categories)	31.	Review Questions for Internal Test-I
12.	3 Sample Answer Scripts for internal test III (Above average, Average, Below Average)	32.	Course Pre analysis
13.	Internal Test-III Marks Statement	33.	Course Plan
14.	Question paper for Internal Test – III	34.	Utilization Chart, Course Plan and Batch details (applicable for practical courses)
15.	Review Questions for Internal Test-III	35.	Time Table (Individual Faculty Time Table)
16.	Remedial Class Details	36.	Question Bank
17.	Micro Analysis for Test II	37.	Previous Years University Examination Question Papers
18.	Assignment 2	38.	Notes of Lesson
19.	List of Students (A/B/C Categories)	39.	Name List (Duly signed by Class Advisor and HoD)
20.	3 Sample Answer Scripts for internal test II (Above average, Average, Below Average)	40.	Syllabus copy – Duly signed by Course Instructor and HoD

HOD 30/6/22

DEAN - COMPUTING

IQAC

PRINCIPAL
30/06/2022

NEHRU INSTITUTE OF ENGINEERING AND TECHNOLOGY
T. M. Palayam, Coimbatore-641 105
(Approved by AICTE, New Delhi and Affiliated to Anna University, Chennai)
Accredited by NAAC, Recognized by UGC with Section 2(f) and 12(B)
NBA Accredited UG Courses: AERO, CSE, MECH

ANNA UNIVERSITY QUESTION PAPER - FEED BACK FORM MARCH 2022

1	Course Name	: Datawarehousing & Datamining
2	Course Code and QP Code	: CS8075 / 20394
3	Date of Exam	: 09.OT.2022.
4	Name of the Faculty Member(s) who handled Course	: 1. MY.G.JEEVANANTHAM 2. 3.
5	Are the Part-A Questions well spread (If No, Give Specific Remarks)	: Yes/No
6	Are the Part-B Questions well spread (If No, Give Specific Remarks)	: Yes/No
7	Are Questions equally covered in all Units (If No give Specific Remarks)	: Yes/No
8	Are Questions asked from outside Syllabus (If Yes, Give Details with Question Numbers)	Yes/No
9	Percentage of Marks-Questions asked from Faculty Members' Question Bank	: 95%.
10	Rank Question Paper from the Students' Perspective	: Easy/Moderate/Tough
11	Rank Question Paper from Faculty Perspective	: Easy/Moderate/Tough
12	Any Specific Remarks on Question Paper that needed to communicate to University	1. – 2. – 3. –
13	Expected Pass Percentage	: 90%.
14	Faculty Members' Specific Remarks on Question Paper (if any)	–

G. Jeev

Signature of the Course Coordinator

S. Subasree

HoD

Reg. No. : 7 2 1 4 1 7 1 0 4 0 4 8

Question Paper Code : 20394

B.E./B.Tech. DEGREE EXAMINATIONS, APRIL/MAY 2022.

Sixth Semester

Computer Science and Engineering

CS 8075 — DATA WAREHOUSING AND DATA MINING

(Common to Computer and Communication Engineering)

(Regulations 2017)

Time : Three hours Maximum : 100 marks

Answer ALL questions.

PART A — (10 × 2 = 20 marks)

1. List the major components of Data Warehousing.
2. Distinguish between OLTP and OLAP.
3. List any four reasons to perform data preprocessing.
4. Mention about data reduction technique.
5. Define the terms support and confidence
6. Represent the various aspects of pattern mining.
7. Compare clustering and classification.
8. How do you choose the best split while constructing a decision tree?
9. List the major features of the WEKA tool.
10. Mention the major advantages of WEKA in Data mining.

PART B — (5 × 13 = 65 marks)

11. (a) Draw the three-tier data warehouse architecture and explain each component in detail. (13)

Or

(b) Differentiate operational database systems and data warehouses. Explain each point in detail. (13)

12. (a) Explain the various steps of knowledge discovery from the data process with a neat diagram. (13)

Or

(b) Suppose that the data for analysis in clued the attribute age. The age values for the data types are: 13, 45. 16, 19, 52, 20, 20, 21, 22, 36, 25, 25, 33, 33, 35, 35, 35, 35, 40, 25, 30, 25, 22, 46, 16, 70, 15, 18, 19, 19

(i) Smooth the above data using bin by mean and boundaries methods. Allot values to bins by bin width technique. Consider some bins as 5. (5)

(ii) Use min-max normalization to transform the value 25 for age onto the range [0.0,1.01. (4)

(iii) Use z-score normalization to transform the value 25 for age. (4)

13. (a) A database has 5 transactions let support = 60% and confidence = 80%.

TID	Items_bought
T1	{Mango, Orange, Pineapple, Grapes, Cherry, Banana}
T2	{Lemon, Orange, Pineapple, Grapes, Cherry, Banana}
T3	{Mango, Apple, Grapes, Cherry}
T4	{Mango, Leach, Custard apple, Grapes, Banana}
T5	{Custard apple, Orange, Orange, Grapes, Watermelon, Cherry }

(i) Find all frequent itemsets using the Apriori algorithm. (7)

(ii) List all of the strong association rules. (6)

Or

(b) Discuss in detail about constraint-based frequent pattern mining with suitable examples. (13)

14. (a) State the mathematical formulation of the SVM problem. Give an outline of the method for solving the problem. (13)

Or

(b) Demonstrate how Bayesian classification helps in predicting class membership probabilities. (13)

15. (a) Discuss in detail the various data pre-processing options involved in WEKA with suitable examples. (13)

Or

(b) Demonstrate the WEKA tool with any suitable data set regarding anyone learning algorithm. (13)

NEHRU INSTITUTE OF ENGINEERING AND TECHNOLOGY

Thirumalayampalayam, Coimbatore-641 105
(Approved by AICTE, New Delhi and Affiliated to Anna University, Chennai)
An ISO 14001: 2015 & 9001:2015 Certified Institution
Accredited by NAAC, Recognized by UGC with Section 2(f) and 12(B)
NBA Accredited UG Courses: AERO, CSE, MECH

Department of Computer Science and Engineering

Teaching Learning Report

Department Name	**Computer Science and Engineering**
Academic Year	2022-2022 (EvenSemester)
Name of the Faculty I/C	Mr. G.Jeevanantham
Name of the Subject	CS8075 / DATAWAREHOUSING AND DATAMINING
Class / Semester	III /VI SEMESTER
Date / Time /Duration	01.06.2022 /10.55am-11.50am /55 minutes
Total No of Beneficiaries	57 Students
Topic of the Activity	Learning Algorithms
Innovative Pedagogy ICT Activity	Video demonstration
Tool Used in the Pedagogy Activity If any	Google Meet, PPT

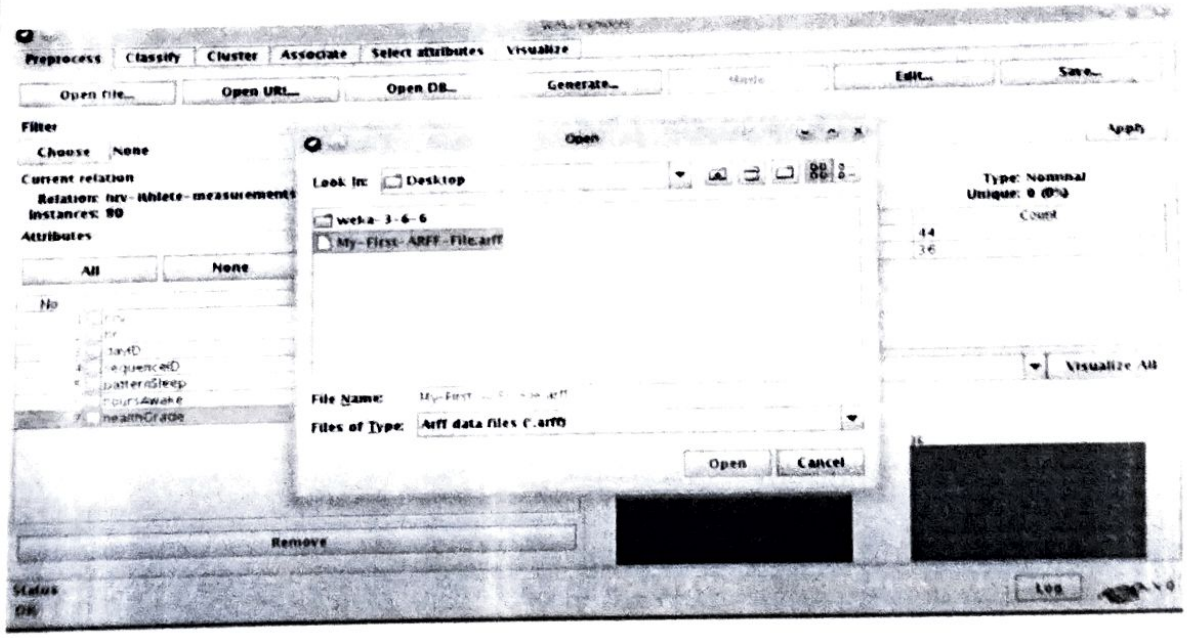
Objective of the activity:

1.To Introduce the Learning Algorithms by using WEKA Tool.

Outcome of the activity:

1.Apply the algorithms using WEKA tool.

Photo Evidence of Interactive Learning by the students



Event Coordinator

IQAC Department Coordinator

HoD

SPECIAL PROGRAMMES FOR SLOW LEARNERS	
S. No	**Description**
1	Slow Learner List
2	Identification of Slow Learner
3	Slow Learners Attendance
4	Slow Learners Report
5	Advance Learners Report

20.	721419104031	NAVANEETH KRISHNAN R V
21.	721419104037	ROHITKRISHNA S
22.	721419104038	SAHAL H
23.	721419104039	SAIRAM P
24.	721419104040	SAMSON RAJA R S
25.	721419104041	SANAM SIVA KRISHNA RAJU
26.	721419104042	SANOJ V
27.	721419104043	SARATH A
28.	721419104044	SATHISH K
29.	721419104046	SREERAG.K.DAS
30.	721419104048	SUDHAKARAN C
31.	721419104049	TAMILINBAM M
32.	721419104050	VARUNKUMAR S
33.	721419104051	VASANTH B
34.	721419104052	VASUNDHARA K
35.	721419104053	VISHWANATHAN R K
36.	721419104054	VISWA S
37.	721419104056	YOGAVIJAY E
38.	721419104302	NAYANA P
39.	721419104303	RESHMA R
40.	721419104702	HITHAYADULLAH H

Subject Incharge

HoD

IDENTIFICATION OF SLOW LEARNERS:

The process of identifying slow and advanced learners is as shown in the table. All students of a particular batch are assessed on following parameters:

S.No.	Assessment Criteria	Weightage
1.	Problem Solving Skills:	
	Previous University Exam Score	80%
	Internal Assessment	
	Attendance	
	Assignments/quizzes	
2.	Other parameters:	20%
	Ability to answer the questions in class	
	General awareness	
	Attentiveness	

SLOW LEARNERS' ATTENDANCE:

S.NO	Register No	Name of the Student	Attendance
1.	721419104002	ABISHEAK A	P
2.	721419104005	AMALDAS K K	P
3.	721419104007	ASWIN S	P
4.	721419104008	BHARATH P	P
5.	721419104009	BHUVANESH U J	P
6.	721419104011	DHANIREDDY NAGA MOHAN REDDY	P
7.	721419104012	DHANUSH S	P
8.	721419104017	HARIHARAN R	P
9.	721419104018	HEMANTH R	P
10.	721419104020	JOSHAN K	P

11.	721419104021	KAVIYA S	P
12.	721419104022	KODAVATIKANTI PRAMOD	P
13.	721419104023	KOUSIK M	P
14.	721419104024	MEGHAL R	P
15.	721419104026	MUDUGURU SRIVATSAVA	P
16.	721419104027	MUKIL VARATHAN M	P
17.	721419104028	NAGARAJAN S	P
18.	721419104029	NALLABALLE VENKATESWARLU	P
19.	721419104030	NARALA JAYADEEP	P
20.	721419104031	NAVANEETH KRISHNAN R V	P
21.	721419104037	ROHITKRISHNA S	P
22.	721419104038	SAHAL H	P
23.	721419104039	SAIRAM P	P
24.	721419104040	SAMSON RAJA R S	P
25.	721419104041	SANAM SIVA KRISHNA RAJU	P
26.	721419104042	SANOJ V	P
27.	721419104043	SARATH A	P
28.	721419104044	SATHISH K	P
29.	721419104046	SREERAG.K.DAS	P
30.	721419104048	SUDHAKARAN C	P
31.	721419104049	TAMILINBAM M	P
32.	721419104050	VARUNKUMAR S	P
33.	721419104051	VASANTH B	P

34.	721419104052	VASUNDHARA K	P
35.	721419104053	VISHWANATHAN R K	P
36.	721419104054	VISWA S	P
37.	721419104056	YOGAVIJAY E	P
38.	721419104302	NAYANA P	P
39.	721419104303	RESHMA R	P
40.	721419104702	HITHAYADULLAH H	P

REMEDIAL CLASS ATTENDANCE CS8075 / Data Warehousing and Data Mining

Slow Learner Report:

1. Conducted a coaching class after completion of the Internal I.
2. During coaching monitoring the students through the improvement.
3. Provided supplementary study materials and support system.
4. Paid individual attention through counselling based on the mentor report.

Subject Incharge

HoD

NEHRU INSTITUTE OF ENGINEERING AND TECHNOLOGY

"Nehru Gardens", Thirumalayampalayam, Coimbatore - 641 105

Approved by AICTE, New Delhi and Affiliated to Anna University, Chennai, An ISO 9001:2015 and ISO 14001:2015 Certified Institution

Accredited by NAAC and Recognized by UGC with Section 2(f) and 12(B), MBA Accredited UG Courses: AERO, CSE, MECH

Department of Computer Science and Engineering

jointly with

Artificial Intelligence and Data Science
Computer Science and Business Systems

in Association with

COMPUTER SOCIETY OF INDIA (CSI)

Two Days National Conference on

INNOVATIVE TECHNOLOGIES IN COMPUTER SCIENCE AND COMMUNICATION ENGINEERING (NCITC - 2022)

CERTIFICATE

This is to certify that

Prof. / Dr. / Mr. / Ms. PAVITHRA.V, III, CSE

from NEHRU INSTITUTE OF ENGINEERING AND TECHNOLOGY

has participated in

Two Days National Conference on

INNOVATIVE TECHNOLOGIES IN COMPUTER SCIENCE AND COMMUNICATION ENGINEERING (NCITC - 2022)

held virtually on 06th & 07th JUNE 2022

He / She also delivered a Keynote Address / an Invited Talk / Chaired a Session / Presented a Paper Entitled

AGRIBOT - AN INTELLIGENT INTERACTIVE INTERFACE TO ASSIST FARMER'S IN AGRICULTURAL ACTIVITES

Dr. S. Sivakumar	Prof. S. Venkatesh	Dr. S. Subramanian	Dr. N. R. Sakthivel	Dr. P. Manoharan
Organizing Secretary	Organizing Secretary	Convenor	Convenor	Patron

NEHRU INSTITUTE OF ENGINEERING AND TECHNOLOGY

'Nehru Gardens', Thirumalayampalayam, Coimbatore - 641 105

(Approved by AICTE, New Delhi and Affiliated to Anna University, Chennai. An ISO 9001:2015 and ISO 14001:2015 Certified Institution
Accredited by NAAC and Recognized by UGC with Section 2(f) and 12(B), NBA Accredited (UG Courses: AERO, CSE, MECH)

Department of Computer Science and Engineering

jointly with

Artificial Intelligence and Data Science
Computer Science and Business Systems

in Association with

COMPUTER SOCIETY OF INDIA (CSI)

Two Days National Conference on

INNOVATIVE TECHNOLOGIES IN COMPUTER SCIENCE AND COMMUNICATION ENGINEERING (NCITC - 2022)

CERTIFICATE

This is to certify that

~~Prof.~~ / ~~Dr.~~ / ~~Mr.~~ / Ms. FATHIMA MINHA K.S , III CSE

from NEHRU INSTITUTE OF ENGINEERING AND TECHNOLOGY

has participated in

Two Days National Conference on

INNOVATIVE TECHNOLOGIES IN COMPUTER SCIENCE AND COMMUNICATION ENGINEERING (NCITC - 2022)

held virtually on 06th & 07th JUNE 2022

He / She also delivered a Keynote Address / an Invited Talk / Chaired a Session / Presented a Paper Entitled

AGRIBOT - AN INTELLIGENT INTERACTIVE INTERFACE TO ASSIST FARMERS IN AGRICULTURAL ACTIVITIES

Dr. S. Shankumar	Prof. S. Venkatesh	Dr. S. Subasree	Dr. N. K. Sakthivel	Dr. P. Manijarasen
Organizing Secretary	Organizing Secretary	Convenor	Convenor	Patron

NEHRU INSTITUTE OF ENGINEERING AND TECHNOLOGY

COIMBATORE - 641105.

COURSE EXIT SURVEY FOR THE COURSE OUTCOME (GRADING IN A SCALE OF 5)

Name of the Department : Department of Computer Science & Engineering

Course Name : DATA WAREHOUSING AND DATA MINING

Course Code : CS8075

Class /Sem : III/V

Dear Student,
Greetings.....!!!

I would like to thank all of you for sparing your time in filling up this Course Pre-Analysis for the effective conduct of the course. As you know that this survey is meant for knowing the knowledge level of the students with respect to this course, please fill it very carefully.

At this juncture, I am glad to welcome the suggestions from you all (if any).

Questions

C315.1 Design a Data warehouse system and perform business analysis with OLAP tools.

C315.2 Apply suitable pre-processing and visualization techniques for data analysis.

C315.3 Apply frequent pattern and association rule mining techniques for data analysis.

C315.4 Apply appropriate classification and clustering techniques for data analysis.

C315.5 Apply WEGA tool and Clustering Algorithm.

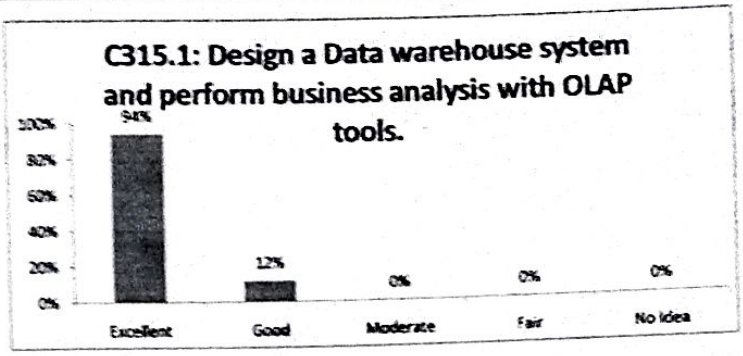
GRADING IN A SCALE OF 5				
Excellent (5)	**Good** (4)	**Moderate** (3)	**Fair** (2)	**No Idea** (1)

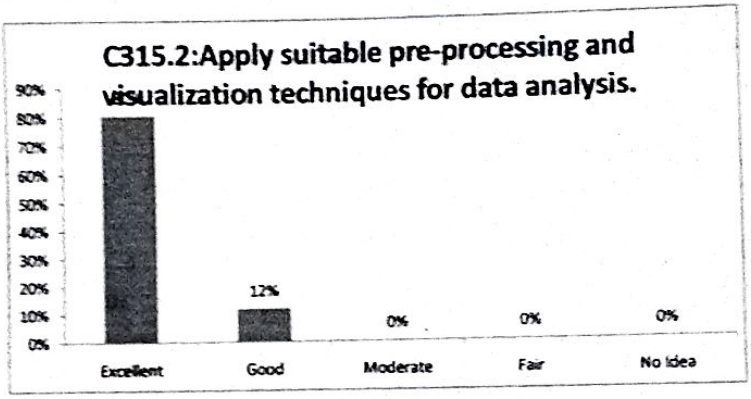
S.No	Reg No	Name of the Student	C315.1	C315.2	C315.3	C315.4	C315.5
1	721419104001	Abbereddy Kiran Mayee	5	5	5	5	5
2	721419104002	Abisheak A	5	5	5	5	5
3	721419104003	Akash Gopinath	4	5	5	5	5
4	721419104004	Allen Joshy	5	4	5	5	5
5	721419104005	Amaldas k.k	5	5	5	5	5
6	721419104006	Anupama K	5	4	5	5	4
7	721419104007	Aswin S	5	5	4	5	5
8	721419104008	Bharath P	5	5	5	5	4
9	721419104009	BHUVANESH U J	5	5	5	5	5
10	721419104011	Dhanireddy Naga Mohan Reddy	5	5	5	5	5
11	721419104012	DHANUSH S	4	5	5	5	5
12	721419104014	Fathima Minha K S	5	4	5	5	5
13	721419104015	Gopi Krishore S	5	5	5	5	5
14	721419104016	Greeshma M Nair	5	4	5	5	4
15	721419104017	Hariharan R	5	5	4	5	5
16	721419104018	Hemanth R	5	5	5	5	4
17	721419104019	JAYA SRINATH V	5	5	5	5	5
18	721419104020	Joshan K	5	5	5	5	5
19	721419104021	KAVIYA S	4	5	5	5	5
20	721419104022	Kodavatikanti Pramod	5	4	5	5	5
21	721419104023	KOUSIK M	5	5	5	5	5
22	721419104024	MEGHAL R	5	4	5	5	4
23	721419104025	MRIDANI M B	5	5	4	5	5
24	721419104026	Sri Vatsava M	5	5	5	5	4
25	721419104027	Mukil Varathan M	5	5	5	5	5
26	721419104028	Nagarajan S	5	5	5	5	5
27	721419104029	NALLABALLE VENKATESWARLU	4	5	5	5	5
28	721419104030	Jayadeep N	5	4	5	5	5
29	721419104031	Navaneeth Krishnan R V	5	5	5	5	5
30	721419104032	Pavithra V	5	4	5	5	4
31	721419104033	Penumarthi Hema Sri	5	5	4	5	5
32	721419104034	RAJEESH R	5	5	5	5	4
33	721419104035	Ramanaboina Venkata Subbaiah	5	5	5	5	5
34	721419104036	ROHAN MURALI NAIR	5	5	5	5	5
35	721419104037	Rohitkrishna S	4	5	5	5	5

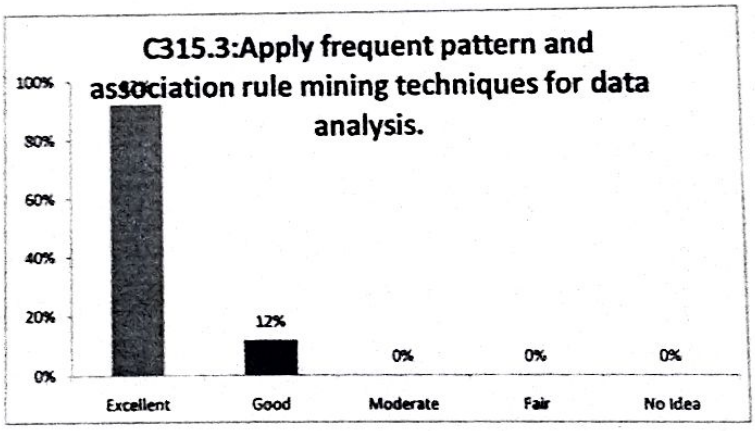
36	721419104038	SAHAL H	5	4	5	5	5
37	721419104039	Sairam P	5	5	5	5	5
38	721419104040	SAMSON RAJA R S	5	4	5	5	4
39	721419104041	Sanam Siva Krishna Raju	5	5	4	5	5
40	721419104042	Sanoj V	5	5	5	5	4
41	721419104043	Sarath A	5	5	5	5	5
42	721419104044	Sathish K	5	5	5	5	5
43	721419104045	Sijo N	4	5	5	5	5
44	721419104046	SREERAG K DAS	5	4	5	5	5
45	721419104047	Sreeram A M	5	5	5	5	5
46	721419104048	Sudakaran C	5	4	5	5	4
47	721419104049	Tamilinbam M	5	5	4	5	5
48	721419104050	Varun Kumar S	5	5	5	5	4
49	721419104051	Vasanth B	5	5	5	5	5
50	721419104052	VASUNDHARA K	5	5	5	5	5
51	721419104053	Vishwanathan R K	5	5	5	5	5
52	721419104054	VISWA S	5	4	5	5	4
53	721419104056	Yogavijay E	5	5	4	5	5
54	721419104302	NAYANA P	5	5	5	5	4
55	721419104303	RESHMA R	5	5	5	5	5
56	721419104701	HARIHARAN S	5	5	5	5	5
57	721419104702	HITHAYADULLAH H	5	5	5	5	5
		AVERAGE	5	5	5	5	5

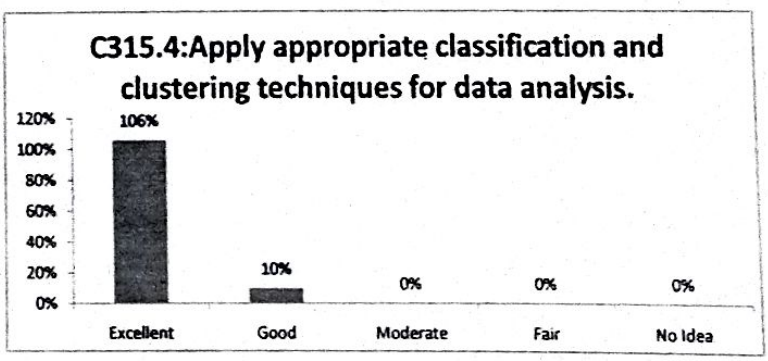
PRE ANALYSIS CHART FOR COURSE OUTCOMES

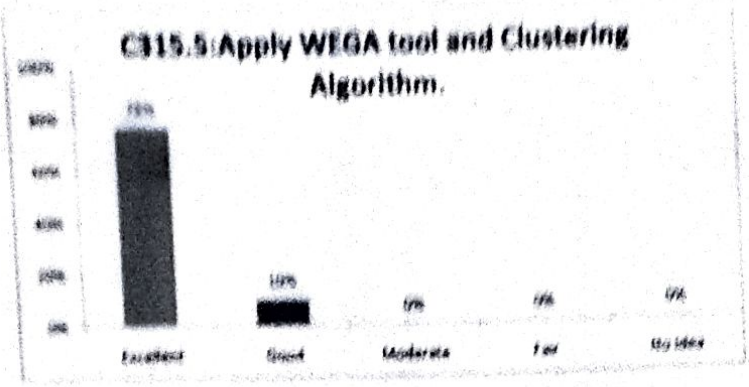
Course Outcomes	Excellent	Good	Moderate	Fair	No Idea	Excellent	Good	Moderate	Fair	No Idea
C315.1: Design a Data warehouse system and perform business analysis with OLAP tools.	48	6	0	0	0	94%	12%	0%	0%	0%
C315.2:Apply suitable pre-processing and visualization techniques for data analysis.	41	6	0	0	0	80%	12%	0%	0%	0%
C315.3:Apply frequent pattern and association rule mining techniques for data analysis.	47	6	0	0	0	92%	12%	0%	0%	0%
C315.4:Apply appropriate classification and clustering techniques for data analysis.	54	5	0	0	0	106%	10%	0%	0%	0%
C315.5:Apply WEGA tool and Clustering Algorithm.	39	5	0	0	0	76%	10%	0%	0%	0%
Overall Pre Analysis of Course outcomes	46	6	0	0	0	90%	11%	0%	0%	0%

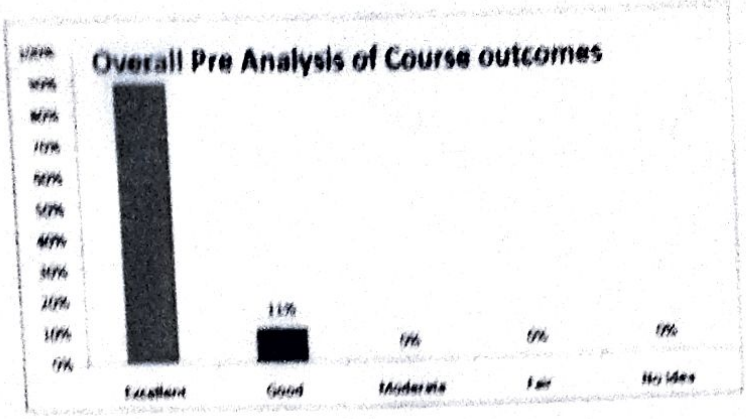












Staff Incharge

HOD

NEHRU INSTITUTE OF ENGINEERING & TECHNOLOGY

"Nehru Gardens" Thirumalayampalayam, Coimbatore – 641 105

(Approved by AICTE and Affiliated to Anna University, Chennai)

Accredited by NBA & NAAC and Recognized by UGC with 2(f) and 12(B)

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

Academic Year : 2021-2022 (Even Semester)

INTERNAL TEST III

Class/Sem : III Yr/VI Sem

Sub code /Sub : CS8075 / DATAWAREHOUSING AND DATAMINING

Course Instructor : Mr.G.Jeevanantham

Regulation : 2017

Batch: 2019-2023

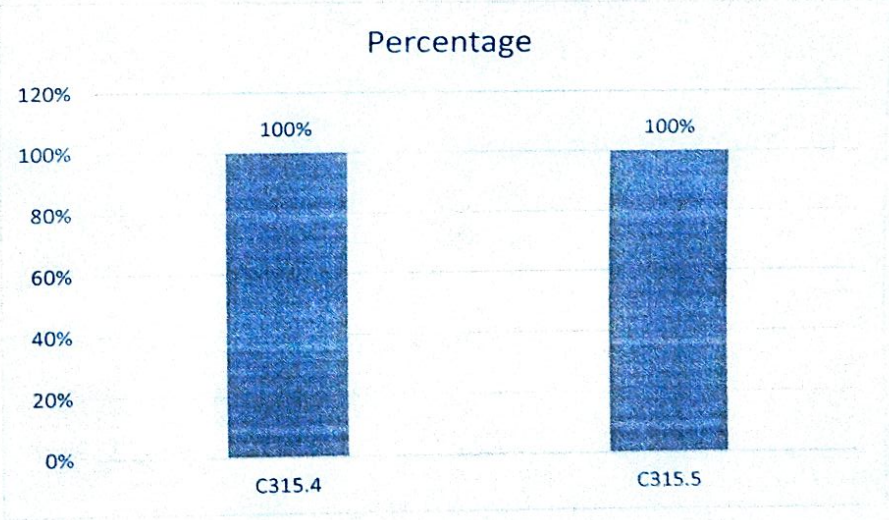
COURSE OUTCOMES	C315.4	Q1	Q2	Q3	Q8		MARKS
	C315.5	Q4	Q5	Q6	Q7		20
						TOTAL	30
							50

S. No.	Reg. No.	Name	Q1 (2)	Q2. (2)	Q3 (2)	Q4 (2)	Q5 (2)	Q6 (13)	Q7 (13)	Q8 (14)		Total (50)	C315.2 (20)	C315.3 (30)	Total (100)
1	721419104001	ABBIREDDY KIRANMAYEE	2	2	2	2	2	12	13	13	48	48	19	29	96
2	721419104002	ABISHEAK A	2	2	2	2	2	12	13	12	47	47	18	29	94
3	721419104003	AKASH GOPINATH	2	2	2	2	2	12	13	13	48	48	19	29	96
4	721419104004	ALLEN JOSHY	2	2	2	2	2	11	10	12	43	43	18	25	86
5	721419104005	AMALDAS K K	2	2	2	2	2	12	13	12.5	47.5	47.5	18.5	29	95
6	721419104006	ANUPAMA K	2	2	2	2	2	13	13	13.5	49.5	49.5	19.5	30	99
7	721419104007	ASWIN S	2	2	2	2	2	12.5	12.5	13	48	48	19	29	96
8	721419104008	BHARATH P	2	2	2	2	2	13	12	13	48	48	19	29	96
9	721419104009	BHUVANESH U J	2	2	2	2	2	12	12	13	47	47	19	28	94
10	721419104011	DHANIREDDY NAGA MOHAN REDDY	2	2	2	2	2	13	12	13	48	48	19	29	96
11	721419104012	DHANUSH S	2	2	2	2	2	13	12	13	48	48	19	29	96
12	721419104014	FATHIMA MINHA K S	2	2	2	2	2	13	12.5	13	48.5	48.5	19	30	97
13	721419104015	GOPIKISHORE S	2	2	2	2	2	13	12	13	48	48	19	29	96
14	721419104016	GREESHMA M NAIR	2	2	2	2	2	13	13	13.5	49.5	49.5	19.5	30	99
15	721419104017	HARIHARAN R	2	2	2	2	2	8	8	9	35	35	15	20	70
16	721419104018	HEMANTH R	2	2	2	2	2	12.5	12.5	13	48	48	19	29	96
17	721419104019	JAYA SRINATH V	2	2	2	2	2	12	12	12	46	46	18	28	92
18	721419104020	JOSHAN K	2	2	2	2	2	13	12	13	48	48	19	29	96

20	721419104022	KODAVATIKANTI PRAMOD	2	2	2	2	2	12	12	12.5	46.5	46.5	18.5	28	93
21	721419104023	KOUSIK M	2	2	2	2	2	10	9.5	8	37.5	37.5	14	24	75
22	721419104024	MEGHAL R	2	2	2	2	2	12	13	13	48	48	19	29	96
23	721419104025	MRIDANI M B	2	2	2	2	2	13	13	13.5	49.5	49.5	19.5	30	99
24	721419104026	SRIVATSAVA	2	2	2	2	2	12	12	13	47	47	19	28	94
25	721419104027	MUKIL VARATHAN M	2	2	2	2	2	10	9.5	8	37.5	37.5	14	24	75
26	721419104028	NAGARAJAN S	2	2	2	2	2	9	10	9	38	38	15	23	76
27	721419104029	NALLABALLE VENKATESWARLU	2	2	2	2	2	8	8	9	35	35	15	20	70
28	721419104030	NARALA JAYADEEP	2	2	2	2	2	12	12	12	46	48	18	28	96
29	721419104031	NAVANEETH KRISHNAN R V	2	2	2	2	2	10	11	12	43	43	18	25	86
30	721419104032	PAVITHRA V	2	2	2	2	2	13	13	13	49	49	19	30	98
31	721419104033	PENUMARTHI NEMA SRI	2	2	2	2	2	13	13	13.5	49.5	49.5	19.5	30	99
32	721419104034	RAJEESH R	2	2	2	2	2	13	13	13.5	49.5	49.5	19.5	30	99
33	721419104035	RAMANABOINA VENKATA SUBBAIAH	2	2	2	2	2	8	8	9.5	35.5	35	15.5	20	70
34	721419104036	ROHAN MURALI NAIR	2	2	2	2	2	11	11	12	44	44	18	26	88
35	721419104037	ROHITKRISHNA S	2	2	2	2	2	12	10	11	43	43	17	26	86
36	721419104038	SAHAL H	2	2	2	2	2	11	12	12	45	45	18	27	90
37	721419104039	SAIRAM P	2	2	2	2	2	12	10	12	44	44	18	26	88
38	721419104040	SAMSON RAJA R S	2	2	2	2	2	9	9	12	40	40	18	22	80
39	721419104041	SANAM SIVA KRISHNA RAJU	2	2	2	2	2	12	12	13	47	47	19	28	94
40	721419104042	SANOJ V	2	2	2	2	2	8	8	9	35	35	15	20	70
41	721419104043	SARATH A	2	2	2	2	2	9	11	9	39	39	15	24	78
42	721419104044	SATHISH K	2	2	2	2	2	11	11	12	44	44	18	26	88
43	721419104045	SIJO N	2	2	2	2	2	12	13	12	47	47	18	29	94
44	721419104046	SREERAG.K.DAS	2	2	2	2	2	12	12	13	47	47	19	28	94
45	721419104047	SREERAM A M	2	2	2	2	2	12	12	12	46	46	18	28	92
46	721419104048	SUDHAKARAN C	2	2	2	2	2	12	10	12	44	44	18	26	88
47	721419104049	TAMILINBAM M	2	2	2	2	2	12	12	13	47	47	19	28	94
48	721419104050	VARUNKUMAR S	2	2	2	2	2	11	12	10.5	43.5	43.5	16.5	27	87
49	721419104051	VASANTH B	2	2	2	2	2	12	11.5	12	45.5	45.5	18	28	91
50	721419104052	VASUNDHARA K	2	2	2	2	2	12	12.5	13	47.5	47.5	19	29	95

51	721419104053	VISHWANATHAN R K	2	2	2	2	2	11	11	13	45	45	19	26	90
52	721419104054	VISWA S	2	2	2	2	2	12	10	11	43	43	17	26	86
53	721419104056	YOGAVIJAY E	2	2	2	2	2	12	11.5	13	46.5	46.5	19	28	93
54	721419104302	NAYANA P	2	2	2	2	2	11	12	13	46	46	19	27	92
55	721419104303	RESHMA R	2	2	2	2	2	12	11.5	13	46.5	46.5	19	28	93
56	721419104701	HARIHARAN S	2	2	2	2	2	12	13	13	48	48	19	29	96
57	721419104702	HITHAYADULLAH H	2	2	2	2	2	11	12	12	45	45	18	27	90
NO. of Students Attended												57	57	57	57
No. of Students Gained												57	57	57	57
Percentage of CO Achieved												100	100	100	100

CO	Percentage
C315.4	100%
C315.5	100%



Course Instructor 22/6/22

HOD 22/6/22

Assignment

Name : Sri vatsava. M.

Reg no : 721419104026.

CS8075 - Data warehousing and Data mining.

T. Srivatsava
7214191040026.

Assignment.

PART-A

OLTP	OLAP
1. It is an online transactional system and manages database modification.	1. It is an online data retrieving and data analysis system.
2. OLTP and its transactions are the original source of data.	2. Different OLTP's database becomes the source of data for OLAP.
3. OLTP has short transactions.	3. OLAP has long transactions.

Drill - down vs Roll-up approach:

It refers to the process of viewing data at the level of increased detail, while roll-up refers to the process of viewing data with decreasing detail.

3) Data mining:

Data mining is the process of findi anomalies pattern and correlations within data sets to predict outcomes.

Steps:

1) Data cleaning
2) Data integration
3) Data Reduction.
4) Data Transformation.
5) Data mining
6) Pattern evaluation
7) knowledge representation.

5) Applications of Datamining:

→ Financial Analysis
→ Telecommunication Industry.
→ Intrusion Detection
→ Retail Industry
→ Higher Education.

Discrete Data	Continuous Data
1. Discrete data is one that has clear spaces between values. 2. Countable 3. It can take only distinct (or) separate values. 4. Mutually inclusive.	1. Continuous data is one that falls on a continuous sequence. 2. Measurable 3. It can take any value in some interval. 4. Mutually exclusive.

PART-B

Database Architecture:

Database architecture is depending on its design and can be of the following types:

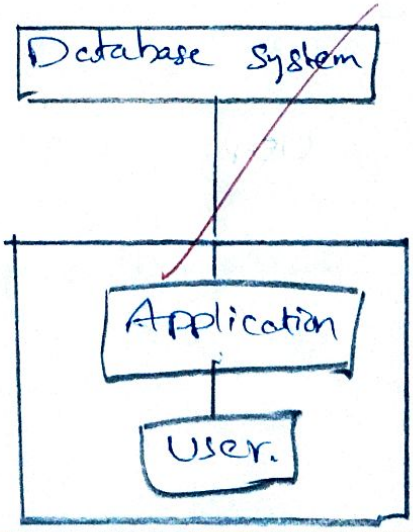
- Centralized
- Decentralized
- Hierarchical.

1-Tier Architecture:

* In this architecture, the database is directly available to the user.
* Any changes done here will directly be done on the database itself.
* It doesn't provide a handy tool for end users.

2-Tier Architecture:

* The 2-Tier architecture is same as basic client-server. In this two-tier architecture, applications on the client end can directly communicate.
* The user interfaces and application programs are run on the clientside.



DATA WAREHOUSING AND DATA MINING ASSIGNMENT

Submitted By
Navaneeth krishnan R.V.
721419104031
B.E. CSE
III YEAR.

INTERNAL TEST-II

PART-I

(1) Steps Involved in a Typical KDD process.

(a) Goal-setting and Application Understanding
(b) Data Selection and Integration.
(c) Data cleaning and preprocessing
(d) Data Transformation.
(e) Data mining.
(f) pattern Evaluation.
(g) knowledge Discovery and Use.

(2). The data set may have a large number of attributes. But some of those attributes can be irrelevant or redundant. The goal of attribute subset selection is to find a minimum set of attributes such that dropping of those irrelevant attributes does not much affect the utility of data and the cost of data analysis could be reduced.

(3) Binning: Binning method sorted data value by consulting its "neighbour-hood" that is, the values around it.

ANNA UNIVERSITY :: CHENNAI - 600 025
OFFICE OF THE CONTROLLER OF EXAMINATIONS
Assessment Details Entered - Report : By subject
APRIL / MAY EXAMINATION,2022 - EXAMINATIONS

Inst Code & Name : 7214 - NEHRU INSTITUTE OF ENGINEERING AND TECHNOLOGY

Subject Code / Name : CS8075 : Data Warehousing and Data Mining

University : AUC

Semester : 06

Register No.	Name of the Student	Attend Period1	Total Period1	Attend Period2	Total Period2	IM 2	Attend Period3	Tot Period3	IM 3	Attend Period4	Total Period4	IM 4
721419104001	ASSIREDDY KIRANMAYEE	14	14	14	14	74	14	14	96	14	14	96
721419104002	ASISHEAK A	13	14	13	14	16	13	14	98	13	14	94
721419104003	AKASH GOPINATH	14	14	14	14	61	14	14	94	14	14	96
721419104004	ALLEN JOSHY	12	14	12	14	60	13	14	94	13	14	86
721419104005	AMALDAS K K	14	14	14	14	17	14	14	98	14	14	95
721419104006	ANUPAMA K	14	14	14	14	99	14	14	99	14	14	99
721419104007	ASWIN S	12	14	12	14	50	13	14	95	13	14	96
721419104008	BHARATH P	14	14	14	14	47	14	14	93	14	14	96
721419104009	BHUVANESH U J	14	14	14	14	50	14	14	87	14	14	94
721419104011	DHANIREDDY NAGA MOHAN REDDY	12	14	12	14	50	13	14	91	13	14	96
721419104012	DHANUSH S	14	14	14	14	40	14	14	82	14	14	96
721419104014	FATHIMA MINHA K S	14	14	14	14	86	14	14	97	14	14	97
721419104015	GOPIKISHORE S	14	14	14	14	66	14	14	82	14	14	96
721419104016	GREESHMA M NAIR	14	14	14	14	96	14	14	99	14	14	99
721419104017	HARIHARAN R	14	14	14	14	50	14	14	95	14	14	70
721419104018	HEMANTH R	12	14	12	14	56	13	14	90	13	14	96
721419104019	JAYA SRINATH V	14	14	14	14	61	14	14	83	14	14	92
721419104020	JOSHAN K	14	14	14	14	50	14	14	92	14	14	96
721419104021	KAVIYA S	12	14	12	14	50	13	14	90	13	14	70
721419104022	KODAVATIKANTI PRAMOD	13	14	13	14	51	13	14	89	13	14	93
721419104023	KOUSIK M	13	14	13	14	50	13	14	90	13	14	75
721419104024	MEGHAL R	14	14	14	14	57	14	14	92	14	14	96
721419104025	MRIDANI M B	14	14	14	14	63	14	14	94	14	14	99
721419104026	MUDUGURU SRIVATSAVA	12	14	12	14	50	13	14	90	13	14	94
721419104027	MUKIL VARATHAN M	14	14	14	14	50	14	14	85	14	14	75
721419104028	NAGARAJAN S	14	14	14	14	50	14	14	85	14	14	76
721419104029	NALLABALLE VENKATESWARLU	14	14	14	14	50	14	14	92	14	14	70
721419104030	NARALA JAYADEEP	12	14	12	14	50	13	14	94	13	14	96
721419104031	NAVANEETH KRISHNAN R V	12	14	12	14	56	13	14	85	13	14	86
721419104032	PAVITHRA V	14	14	14	14	62	14	14	90	14	14	98
721419104033	PENUMARTHI HEMA SRI	14	14	14	14	86	14	14	99	14	14	99
721419104034	RAJEESH R	14	14	14	14	68	14	14	85	14	14	99
721419104035	RAMANABOINA VENKATA SUBBAIAH	14	14	14	14	80	14	14	92	14	14	70
721419104036	ROHAN MURALI NAIR	14	14	14	14	65	14	14	85	14	14	88
721419104037	ROHITKRISHNA S	14	14	14	14	36	14	14	88	14	14	86
721419104038	SAHAL H	14	14	14	14	59	14	14	92	14	14	90
721419104039	SAIRAM P	13	14	13	14	24	13	14	98	13	14	88
721419104040	SAMSON RAJA R S	14	14	14	14	36	14	14	95	14	14	80
721419104041	SANAM SIVA KRISHNA RAJU	14	14	14	14	28	14	14	98	14	14	94
721419104042	SANOJ V	14	14	14	14	50	14	14	90	14	14	70
721419104043	SARATH A	13	14	13	14	55	13	14	85	13	14	78
721419104044	SATHISH K	12	14	12	14	50	13	14	92	13	14	88
721419104045	SIJO N	13	14	13	14	86	13	14	93	13	14	94
721419104046	SREERAG.K.DAS	14	14	14	14	50	14	14	92	14	14	94
721419104047	SREERAM A M	14	14	14	14	76	14	14	92	14	14	92
721419104048	SUDHAKARAN C	14	14	14	14	34	14	14	96	14	14	88
721419104049	TAMILINBAM M	14	14	14	14	18	14	14	98	14	14	94
721419104050	VARUNKUMAR S	14	14	14	14	26	14	14	97	14	14	87
721419104051	VASANTH B	14	14	14	14	50	14	14	90	14	14	91
721419104052	VASUNDHARA K	14	14	14	14	58	14	14	84	14	14	95
721419104053	VISHWANATHAN R K	14	14	14	14	22	14	14	98	14	14	90
721419104054	VISWA S	14	14	14	14	50	14	14	90	14	14	86

ANNA UNIVERSITY :: CHENNAI - 600 025
OFFICE OF THE CONTROLLER OF EXAMINATIONS
Assessment Details Entered - Report : By subject
APRIL / MAY EXAMINATION,2022 - EXAMINATIONS

Inst Code & Name : 7214 - NEHRU INSTITUTE OF ENGINEERING AND TECHNOLOGY

7214191104056	YOGAVIJAY E	14	14	14	14	56	14	14	85	14	14	1
7214191104302	NAYANA P	14	14	14	14	30	14	14	98	14	14	1
7214191104303	RESHMA R	14	14	14	14	0	14	14	99	14	14	1
7214191104701	HARIHARAN S	14	14	14	14	72	14	14	96	14	14	1
7214191104702	HITHAYADULLAH H	12	14	12	14	26	13	14	94	13	14	1

NEHRU INSTITUTE OF ENGINEERING AND TECHNOLOGY

Nehru Gardens, Thirumalayampalayam, Coimbatore - 641 105, Tamil Nadu
ISO 14001 : 2004 Certified, Accredited by NAAC with "B++" Grade, Approved by AICTE, New Delhi
Affiliated to Anna University, Chennai, Recognized by UGC with 2(f) and 12(B), Accredited by NBA & NAAC

Anna University Examination - Internal Assessment Test

Office Use Only

Roll Number

Register Number 7 2 1 4 1 9 1 0 4 0 2 5

Read the instructions given overleaf carefully before filling in the title page (To be filled in by the candidate)

College Code	7214
College Name	Nehru Institute of Engineering and Technology, Coimbatore
Degree / Branch	BE CSE
Semester	VI
Subject Code	CS8075
Date & Session	15.06.22
Subject Title	Data warehousing & Data Mining
No. of Pages used	17

Put Tick ✓ Mark Here.

Question Paper Set	A	B ✓		
Internal Test	I	II	III ✓	
Year of Study	I	II	III ✓	IV
Class	A ✓	B		

All particulars given are verified

Signature of the Hall Superintendent with date

Name of the Hall Superintendent

DO NOT WRITE THE REGISTER NUMBER, COLLEGE CODE AND THE NAME IN ANY OTHER PART OF THE ANSWER BOOK

Office Use Only

(To be filled in by the candidate)

Date : 15-06-2022 Session : Afternoon

Subject Code/Title : CS8075 Datawarehousing and Data Mining

Question Paper Set : II No. of Pages used : 17.

Office Use Only

(To be filled by the candidate)

Date : 15-06-2022 Session : Afternoon Question Paper Set : II

Subject Code / Title : CS8075 Datawarehousing and Data Mining

Instruction to the Candidate : Put a tick mark (✓) for the questions attended in the tick mark column against each question in V-1

Part - A Question No.	Part - A Marks	PART - B & C Question No.		Marks I	Marks II	Marks III	Marks Total
1	2	6	a	13			13
2	2		b				
3	2	7	a	13			13
4	2		b				
5	2	8	a	13½			13½
			b				
Total	10						39½

GRAND TOTAL (IN WORDS): Four Nine Half

GRAND TOTAL: 49½/80

Divan

Declaration by the Examiner : Verified that all the questions attended by the student are valued and the total is found to be correct

Date	Name of the Examiner	Signature of the Examiner
15/6/22	G. Saravanan	

Declaration by the Scrutiny : Verified that all the questions attended by the student are valued and the total is found to be correct

Date	Name of the Scrutiny	Signature of the Scrutiny

NEHRU INSTITUTE OF ENGINEERING AND TECHNOLOGY

'Nehru Gardens', Thirumalayampalayam, Coimbatore - 641 105, Tamil Nadu.
(ISO 14001 : 2004 Certified, Accredited by NAAC with "B++" Grade, Approved by AICTE-New Delhi,
Affiliated to Anna University, Chennai, Recognized by UGC with 2(f) and 12(B), Accreditated by NBA & NAAC

Anna University Examination - Internal Assessment Test

Read the instructions given overleaf carefully before filling in the title page. (To be filled in by the candidate)

Office Use Only

College Code	7214
College Name	Nehru Institute of Engineering and Technology, Coimbatore
Degree / Branch	B.E Computer Science
Semester	06
Subject Code	CS8075
Date & Session	15/6/2022 AN
Subject Title	Data Warehouse and Data mining
No. of Pages used	18

Roll Number: 7 2 1 4 1 9 1 0 4 0 0 7

Register Number: 7 2 1 4 1 9 1 0 4 0 0 7

Put Tick (✓) Mark Here.

Question Paper Set	A	✓ B		
Internal Test	I	II	✓ III	
Year of Study	I	II	✓ III	IV
Class	A	✓ B		

All particulars given are verified

Signature of the Hall Superintendent with date

Name of the Hall Superintendent

DO NOT WRITE THE REGISTER NUMBER, COLLEGE CODE AND THE NAME IN ANY OTHER PART OF THE ANSWER BOOK

Office Use Only

(To be filled in by the candidate)

Date : 15/6/2022 Session : AN

Subject Code/Title : CS8075 Data warehousing and Data mining

Question Paper Set : B No. of Pages used : 18

Office Use Only

Date : 15/6/2022 Session : AN Question Paper Set : B

Subject Code / Title : CS8075 Datawarehousing and Datamining

Instruction to the Candidate : Put a tick mark (✓) for the questions attended in the tick mark column against each question in V-1

Part - A Question No.	Part - A Marks	PART - B & C Question No.		Marks i	Marks ii	Marks iii	Marks Total
1	2	6	a	12			12
2	2		b				
3	2	7	a	12			12
4	2		b				
5	2	8	a	13			13
			b				
Total	10						38

GRAND TOTAL (IN WORDS): four EIGHT

GRAND TOTAL: 48/50

Declaration by the Examiner : Verified that all the questions attended by the student are valued and the total is found to be correct

Date	Name of the Examiner	Signature of the Examiner
	G. Jeevanantham.	G.

Declaration by the Scrutiny : Verified that all the questions attended by the student are valued and the total is found to be correct

Date	Name of the Scrutiny	Signature of the Scrutiny

NEHRU INSTITUTE OF ENGINEERING AND TECHNOLOGY

'Nehru Gardens', Thirumalayampalayam, Coimbatore - 641 105, Tamil Nadu
ISO 14001 : 2004 Certified, Accredited by NAAC with "B++" Grade, Approved by AICTE-New Delhi,
Affiliated to Anna University, Chennai, Recognized by UGC with 2(f) and 12(B), Accreditated by NBA & NAAC

Anna University Examination - Internal Assessment Test

Office Use Only

Read the instructions given overleaf carefully before filling in the title page. (To be filled in by the candidate)

Roll Number / Register Number: 7214910401 8

College Code	7214
College Name	Nehru Institute of Engineering and Technology, Coimbatore
Degree - Branch	BE / CSE
Semester	06
Subject Code	CS8075
Date & Session	15/6/22 AN
Subject Title	DATA WAREHOUSING AND DATA MINING
No. of Pages used	16

Put Tick (✓) Mark Here.

Question Paper Set	A	B ✓		
Internal Test	I	II	III ✓	
Year of Study	I	II	III ✓	IV
Class	A ✓	B		

All particulars given are verified

Signature of the Hall Superintendent with date

Name of the Hall Superintendent

DO NOT WRITE THE REGISTER NUMBER, COLLEGE CODE AND THE NAME IN ANY OTHER PART OF THE ANSWER BOOK

Office Use Only

(To be filled in by the candidate)

Date : 15/6/2022 Session : AN

Subject Code/Title : CS 8075 DATA WAREHOUSING AND DATA MINING

Question Paper Set : II No. of Pages used : 16

Office Use Only

(To be filled in by the candidate)

Date : 15/06/2022 Session : AN Question Paper Set : II

Subject Code / Title : CS 8075 DATA WAREHOUSING AND DATA MINING

Instruction to the Candidate : Put a tick mark (✓) for the questions attended in the tick mark column against each question in V-1

Part - A		PART - B & C						GRAND TOTAL (IN WORDS)
Question No.	Marks	Question No.		Marks I	II	III	Total	
1	✓ 2	6	a					
2	✓ 2		b	✓ 12			12	FOUR EIGHT
3	✓ 2	7	a	✓ 12			12	
4	✓ 2		b					
5	✓ 2	8	a	✓ 13			13	
			b					GRAND TOTAL
Total	10						38	48/50

Declaration by the Examiner : Verified that all the questions attended by the student are valued and the total is found to be correct

Date	Name of the Examiner	Signature of the Examiner
15/6/22	G. Jeevarathinam	G. Jmb

Declaration by the Scrutiny : Verified that all the questions attended by the student are valued and the total is found to be correct

Date	Name of the Scrutiny	Signature of the Scrutiny

(60)

NEHRU INSTITUTE OF ENGINEERING AND TECHNOLOGY

Nehru Gardens, Thirumalayampalayam, Coimbatore – 641 105
An ISO 14001: 2015 & 9001:2015 Certified Institution
Approved by AICTE, New Delhi & Affiliated to Anna University, Chennai
Accredited by NAAC, Recognized by UGC with 2(F) & 12(B)
NBA Accredited UG Courses: AERO | CSE | MECH

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

B.E./B.Tech.(Full Time) – Internal Assessment Examination - III, JUNE 2022

REGULATION – 2017 SET II

Course Instructor	:	CS8075 – DATA WAREHOUSING AND DATA MINING	Portion	:	1.5 Units
Year / Semester	:	Mr.JEEVANANTHAM G	Date / Session	:	15.06.2022 / AN
Duration	:	IV / VIII	Max. Marks	:	50
		90 minutes			

Knowledge Level	K1: Remembering	K2: Understanding	K3: Applying	K4: Analyzing	K5: Evaluating	K6: Creating
Course Outcomes	C315.4	Apply appropriate classification and clustering techniques for data analysis.				
	C315.4	Apply Weka Tool and Clustering Algorithms.				

Q.No.	PART –A (Answer ALL questions) (5×2 = 10 Marks) Questions	KL / CO
1.	Write any two weakness of decision tree approach.	K2 / C315.4
2.	State Naïve Bayesian Classifier	K2 / C315.4
3.	Give example of rule based classification	K2 / C315.4
4.	Mention the use of K-Nearest Neighbour alogithm.	K2 / C315.5
5.	What is machine learning algorithm?	K1 / C315.5

	PART-B (Answer ALL questions) (2×13 = 26 Marks)	
6. a.	Identify the Classifier are selected in WEKA.	K4 / C315.5
	(Or)	
6. b.	Categorize the Attribute Relation File.	K4 / C315.5
7. a.	What is Weka? Categorize the features of Weka.	K4 / C315.5
	(Or)	
7. b.	Classify Linear and Non Linear machine learning algorithms.	K2 / C315.5

	PART-C (1×14 = 14 Marks)	
8.a.	Explain in detail requirements of clustering in data mining.	K4 / C315.4
	(Or)	
8.b.	State Bayes' theorem of posterior probability. Outline the steps in Bayesian classification with an example.	K2 / C315.4

G. 10/06/2022
Course Coordinator

Examcell

B. Subasree 10/6/2022
HoD

Vision:
To produce highly competent and innovative Computer Professionals to meet the global demands.

Mission:
To impart quality education by creative teaching learning process.
To be technically competent, ethical and socially responsible throughout the professional career.
To inculcate leadership qualities and entrepreneurship culture to meet the global standards.

NEHRU INSTITUTE OF ENGINEERING AND TECHNOLOGY

Nehru Gardens, Thirumalayampalayam, Coimbatore – 641 105
An ISO 14001: 2015 & 9001:2015 Certified Institution
Approved by AICTE, New Delhi & Affiliated to Anna University, Chennai
Accredited by NAAC, Recognized by UGC with 2(F) & 12(B)
NBA Accredited UG Courses: AERO | CSE | MECH

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

B.E./B.Tech.(Full Time) – Internal Assessment Examination - III, June 2022
REGULATION – 2017
CS8075 – DATA WAREHOUSING AND DATA MINING

REVIEW QUESTIONS

PART –A

1.	Illustrate Data Classification with example.
2.	Mention the two setps in data classification.
3.	Identify any two reasons to explain why decision tree classifiers are so popular.
4.	What are the System requirement of WEGA.
5.	Write the names of any four types of algorithms supported by WEGA
6.	Write any two weakness of decision tree approach.
7.	State Naïve Bayesian Classifier
8.	Give example of rule based classification
9.	Mention the use of K-Nearest Neighbour alogithm.
10.	What is machine learning algorithm?

PART –B & PART –C

1.	What is Weka? Categorize the features of Weka.
2.	Categorize the Attribute Relation File.
3.	Identify the Classifier are selected in WEKA.
4.	State Bayes' theorem of posterior probability. Outline the steps in Bayesian classification with an example.
5.	Explain in detail requirements of clustering in data mining.
6.	Classify Linear and Non Linear machine learning algorithms.
7.	**Describe**in detail about the following Classification methods.(i). Bayesian classification. (ii)Classification by Back propagation.
8.	(i). Write Bayes theorem. (ii) Explain how the Bayesian Belief Networks are trained to perform classification.

10/8/22
Staff Incharge

S. Subasree
HoD
10/8/22

Vision
To produce highly competent and innovative **Computer Professionals** to meet the global demands
Mission
- To impart quality education by **creative teaching learning** process.
- To be technically **competent, ethical and socially responsible** throughout the professional career.

To inculcate **leadership qualities and entrepreneurship culture** to meet the global standards.

NEHRU INSTITUTE OF ENGINEERING & TECHNOLOGY

"Nehru Gardens" Thirumalayampalayam, Coimbatore – 641 105

(Approved by AICTE and Affiliated to Anna University, Chennai)

Accredited by NBA & NAAC and Recognized by UGC with 2(f) and 12(B)

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

Academic Year : 2021-2022 (Even Semester)

INTERNAL TEST II

Class/Sem : III Yr/VI Sem

Sub code /Sub : CS8075 / DATAWAREHOUSING AND DATAMINING

Course Instructor : Mr.G.Jeevanantham

Regulation : 2017

Batch: 2019-2023

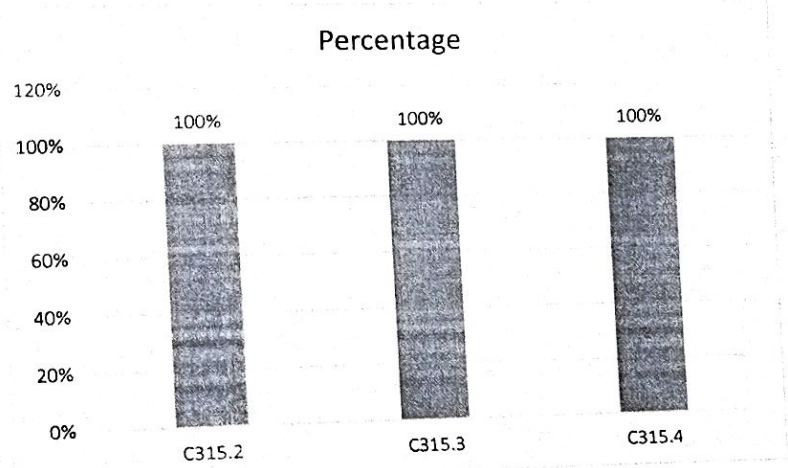
							MARKS
COURSE OUTCOMES	C315.2	Q1	Q2	Q3	Q8(i)		10
	C315.3	Q6	Q7	Q4	Q5		30
	C315.4	Q8(ii)					10
						TOTAL	50

S. No.	Reg. No.	Name	Q1 (2)	Q2 (2)	Q3 (2)	Q4 (2)	Q5 (2)	Q6 (13)	Q7 (13)	Q8(i) (4)	Q8(ii) (10)	Total (50)	C315.2 (10)	C315.3 (30)	C315.4 (10)	Total (100)
1	721419104001	ABBIREDDY KIRANMAYEE	2	2	2	2	2	12	13	3	10	48	9	29	10	96
2	721419104002	ABISHEAK A	2	2	2	2	2	13	13	3	10	49	9	30	10	98
3	721419104003	AKASH GOPINATH	2	2	2	2	2	12	12	3	10	47	9	28	10	94
4	721419104004	ALLEN JOSHY	2	2	2	2	2	12	12	3	10	47	9	28	10	94
5	721419104005	AMALDAS K K	2	2	2	2	2	13	13	3	10	49	9	30	10	98
6	721419104006	ANUPAMA K	2	2	2	2	2	13	13	3.5	10	49.5	9.5	30	10	99
7	721419104007	ASWIN S	2	2	2	2	2	12	12.5	3	10	47.5	9	29	10	95
8	721419104008	BHARATH P	2	2	2	2	2	11.5	12	3	10	46.5	9	28	10	93
9	721419104009	BHUVANESH U J	2	2	2	2	2	11.5	11	3	8	43.5	9	27	8	87
10	721419104011	DHANIREDDY NAGA MOHAN REDDY	2	2	2	2	2	12	11.5	3	9	45.5	9	28	9	91
11	721419104012	DHANUSH S	2	2	2	2	2	10	10	3	8	41	9	24	8	82
12	721419104014	FATHIMA MINHA K S	2	2	2	2	2	12	13	3.5	10	48.5	9.5	29	10	97
13	721419104015	GOPIKISHORE S	2	2	2	2	2	10	10	3	8	41	9	24	8	82
14	721419104016	GREESHMA M NAIR	2	2	2	2	2	13	13	3.5	10	49.5	9.5	30	10	99
15	721419104017	HARIHARAN R	2	2	2	2	2	12	12	3.5	10	47.5	9.5	28	10	95
16	721419104018	HEMANTH R	2	2	2	2	2	12	12	3	8	45	9	28	8	90
17	721419104019	JAYA SRINATH V	2	2	2	2	2	10	10	3	8.5	41.5	9	24	8.5	83

																9	27	10	92
18	721419104020	JOSHAN K	2	2	2	2	2	12	11	3	10	46	9	27	10	92
19	721419104021	KAVIYA S	2	2	2	2	2	12	10	3	10	45	9	26	10	90
20	721419104022	KODAVATIKANTI PRAMOD	2	2	2	2	2	12	10	3	9.5	44.5	9	26	9.5	89
21	721419104023	KOUSIK M	2	2	2	2	2	10	12	3	10	45	9	26	10	90
22	721419104024	MEGHAL R	2	2	2	2	2	12	11	3	10	46	9	27	10	92
23	721419104025	MRIDANI M B	2	2	2	2	2	12	12	3	10	47	9	28	10	94
24	721419104026	SRIVATSAVA	2	2	2	2	2	11	11	3	10	45	9	26	10	90
25	721419104027	MUKIL VARATHAN M	2	2	2	2	2	10	9.5	3	10	42.5	9	24	10	85
26	721419104028	NAGARAJAN S	2	2	2	2	2	9.5	10	3	10	42.5	9	24	10	85
27	721419104029	NALLABALLE VENKATESWARLU	2	2	2	2	2	12	11	3	10	46	9	27	10	92
28	721419104030	NARALA JAYADEEP	2	2	2	2	2	12	12	3	10	47	9	28	10	94
29	721419104031	NAVANEETH KRISHNAN R V	2	2	2	2	2	10	11	3	8.5	42.5	9	25	8.5	85
30	721419104032	PAVITHRA V	2	2	2	2	2	10	12	3	10	45	9	26	10	90
31	721419104033	PENUMARTHI NEMA SRI	2	2	2	2	2	13	13	3.5	10	49.5	9.5	30	10	99
32	721419104034	RAJEESH R	2	2	2	2	2	8.5	11	3	10	42.5	9	24	10	85
33	721419104035	RAMANABOINA VENKATA SUBBAIAH	2	2	2	2	2	11	12	3	10	46	9	27	10	92
34	721419104036	ROHAN MURALI NAIR	2	2	2	2	2	10.5	11	3	8	42.5	9	26	8	85
35	721419104037	ROHITKRISHNA S	2	2	2	2	2	9	12	3	10	44	9	25	10	88
36	721419104038	SAHAL H	2	2	2	2	2	11	12	3	10	46	9	27	10	92
37	721419104039	SAIRAM P	2	2	2	2	2	13	13	3	10	49	9	30	10	98
38	721419104040	SAMSON RAJA R S	2	2	2	2	2	12	12	3.5	10	47.5	9.5	28	10	95
39	721419104041	SANAM SIVA KRISHNA RAJU	2	2	2	2	2	13	13	3	10	49	9	30	10	98
40	721419104042	SANOJ V	2	2	2	2	2	11	11	3	10	45	9	26	10	90
41	721419104043	SARATH A	2	2	2	2	2	10	11	3	8.5	42.5	9	25	8.5	85
42	721419104044	SATHISH K	2	2	2	2	2	12	11	3	10	46	9	27	10	92
43	721419104045	SIJO N	2	2	2	2	2	12	11	3.5	10	46.5	9.5	27	10	93
44	721419104046	SREERAG.K.DAS	2	2	2	2	2	11	12	3	10	46	9	27	10	92
45	721419104047	SREERAM A M	2	2	2	2	2	12	12	3	9	46	9	28	9	92
46	721419104048	SUDHAKARAN C	2	2	2	2	2	12	13	3	10	48	9	29	10	96
47	721419104049	TAMILINBAM M	2	2	2	2	2	13	13	3	10	49	9	30	10	98

48	721419104050	VARUNKUMAR S	2	2	2	2	2	13	13	3	9.5	48.5	9	30	9.5	97
49	721419104051	VASANTH B	2	2	2	2	2	12	11	3	9	45	9	27	9	90
50	721419104052	VASUNDHARA K	2	2	2	2	2	11	11	3	7	42	9	26	7	84
51	721419104053	VISHWANATHAN R K	2	2	2	2	2	13	13	3	10	49	9	30	10	98
52	721419104054	VISWA S	2	2	2	2	2	12	10	3	10	45	9	26	10	90
53	721419104056	YOGAVIJAY E	2	2	2	2	2	11	10	3	8.5	42.5	9	25	8.5	85
54	721419104302	NAYANA P	2	2	2	2	2	13	13	3	10	49	9	30	10	98
55	721419104303	RESHMA R	2	2	2	2	2	13	13	3.5	10	49.5	9.5	30	10	99
56	721419104701	HARIHARAN S	2	2	2	2	2	12	13	3	10	48	9	29	10	96
57	721419104702	HITHAYADULLAH H	2	2	2	2	2	12	12	3	10	47	9	28	10	94
NO. of Students Attended												57	57	57	57	57
No. of Students Gained												57	57	57	57	57
Percentage of CO Achieved												100	100	100	100	100

CO	Percentage
C315.2	100%
C315.3	100%
C315.4	100%



Course Instructor 26/5/22

HOD 26/5/22

ANNA UNIVERSITY :: CHENNAI - 600 025
OFFICE OF THE CONTROLLER OF EXAMINATIONS
Assessment Details Entered - Report : By subject
APRIL / MAY EXAMINATION,2022 - EXAMINATIONS

Inst Code & Name : 7214 - NEHRU INSTITUTE OF ENGINEERING AND TECHNOLOGY

Subject Code / Name : CS8075 : Data Warehousing and Data Mining

University : AUC

Semester : 06

Register No.	Name of the Student	Attend Period1	Total Period1	Attend Period2	Total Period2	IM 2	Attend Period3	Tot Period3	IM 3	Attend Period4	Total Period4	IM 4
7214191040 01	ABBIREDDY KIRANMAYEE	14	14	14	14	74	14	14	96	14		
7214191040 02	ABISHEAA A	13	14	13	14	16	13	14	96	13		
7214191040 03	AKASH GOPINATH	14	14	14	14	61	14	14	94	14		
7214191040 04	ALLEN JOSHY	12	14	12	14	60	13	14	94	13		
7214191040 05	AMALDAS K K	14	14	14	14	17	14	14	98	14		
7214191040 06	ANUPAMA K	14	14	14	14	99	14	14	99	14		
7214191040 07	ASWIN S	12	14	12	14	50	13	14	95	13		
7214191040 08	BNARATH P	14	14	14	14	47	14	14	93	14		
7214191040 09	BHUVANESH G J	14	14	14	14	50	14	14	87	14		
7214191040 11	DHANIREDDY NAGA MOHAN REDDY	12	14	12	14	50	13	14	91	13		
7214191040 12	DHANUSH S	14	14	14	14	40	14	14	82	14		
7214191040 14	FATHIMA MINHA K S	14	14	14	14	86	14	14	97	14		
7214191040 15	GOPIKISHORE S	14	14	14	14	66	14	14	82	14		
7214191040 16	GREESHMA M NAIR	14	14	14	14	96	14	14	99	14		
7214191040 17	HARINARAN R	14	14	14	14	50	14	14	95	14		
7214191040 18	HEMANTH R	12	14	12	14	56	13	14	90	13		
7214191040 19	JAYA SRINATH V	14	14	14	14	61	14	14	83	14		
7214191040 20	JOSHAN K	14	14	14	14	50	14	14	92	14		
7214191040 21	KAVIYA S	12	14	12	14	50	13	14	90	13		
7214191040 22	KODAVATIKANTI PRAMOD	13	14	13	14	51	13	14	89	13		
7214191040 23	KOUSIK M	13	14	13	14	50	13	14	90	13		
7214191040 24	MEGHAL R	14	14	14	14	57	14	14	92	14		
7214191040 25	MRIDANI M B	14	14	14	14	63	14	14	94	14		
7214191040 26	MUDUGURU SRIVATSAVA	12	14	12	14	60	13	14	90	13		
7214191040 27	MUKIL VARATHAN M	14	14	14	14	50	14	14	85	14		
7214191040 28	NAGARAJAN S	14	14	14	14	50	14	14	85	14		
7214191040 29	NALLABALLE VENKATESWARLU	14	14	14	14	50	14	14	92	14		
7214191040 30	NARALA JAYADEEP	12	14	12	14	50	13	14	94	13		
7214191040 31	NAVANEETH KRISHNAN R V	12	14	12	14	56	13	14	85	13		
7214191040 32	PAVITHRA V	14	14	14	14	62	14	14	90	14		
7214191040 33	PENUMARTHI HEMA SRI	14	14	14	14	86	14	14	99	14		
7214191040 34	RAJEESH R	14	14	14	14	68	14	14	85	14		
7214191040 35	RAMANABOINA VENKATA SUBBAIAH	14	14	14	14	80	14	14	92	14		
7214191040 36	ROHAN MURALI NAIR	14	14	14	14	65	14	14	85	14		
7214191040 37	ROHITKRISHNA S	14	14	14	14	36	14	14	88	14		
7214191040 38	SAHAL H	14	14	14	14	59	14	14	92	14		
7214191040 39	SAIRAM P	13	14	13	14	24	13	14	98	13		
7214191040 40	SAMSON RAJA R S	14	14	14	14	36	14	14	95	14		
7214191040 41	SANAM SIVA KRISHNA RAJU	14	14	14	14	28	14	14	98	14		
7214191040 42	SANOJ V	14	14	14	14	50	14	14	90	14		
7214191040 43	SARATH A	13	14	13	14	55	13	14	85	13		
7214191040 44	SATHISH K	12	14	12	14	50	13	14	92	13		
7214191040 45	SIJO N	13	14	13	14	86	13	14	93	13		
7214191040 46	SREERAG.K.DAS	14	14	14	14	50	14	14	92	14		
7214191040 47	SREERAM A M	14	14	14	14	76	14	14	92	14		
7214191040 48	SUDHAKARAN C	14	14	14	14	34	14	14	96	14		
7214191040 49	TAMILINBAM M	14	14	14	14	18	14	14	98	14		
7214191040 50	VARUNKUMAR S	14	14	14	14	26	14	14	97	14		
7214191040 51	VASANTH B	14	14	14	14	50	14	14	90	14		
7214191040 52	VASUNDHARA K	14	14	14	14	58	14	14	84	14		
7214191040 53	VISHWANATHAN R K	14	14	14	14	22	14	14	98	14		
7214191040 54	VISWA S	14	14	14	14	50	14	14	90	14		

ANNA UNIVERSITY :: CHENNAI - 600 025
OFFICE OF THE CONTROLLER OF EXAMINATIONS

Assessment Details Entered - Report : By subject

APRIL / MAY EXAMINATION,2022 - EXAMINATIONS

Inst Code & Name : 7214 - NEHRU INSTITUTE OF ENGINEERING AND TECHNOLOGY

721419104056	YOGAVIJAY E	14	14	14	14	98	14	14	88	14
721419104302	HAYANA P	14	14	14	14	30	14	14	98	14
721419104303	RESHMA R	14	14	14	14	0	14	14	89	14
721419104701	HARIHARAN S	14	14	14	14	72	14	14	98	14
721419104702	HITHAYADULLAH H	12	14	12	14	28	13	14	94	13

26/8/22

NEHRU INSTITUTE OF ENGINEERING AND TECHNOLOGY

Nehru Gardens, Thirumalayampalayam, Coimbatore - 641 105, Tamil Nadu
(An ISO 14001 : 2004 Certified, Accredited by NAAC with "B++" Grade, Approved by AICTE, New Delhi
Affiliated to Anna University, Chennai, Recognized by UGC with 2(f) and 12(B), Accreditated by NBA & NAAC)

Anna University Examination - Internal Assessment Test

Read the instructions given overleaf carefully before filling in the title page (To be filled in by the candidate)

College Code	7214
College Name	Nehru Institute of Engineering and Technology, Coimbatore
Degree / Branch	BE Computer Science and Engineering
Subject Code	CS8075
Subject Title	Data warehousing & Data Mining
Semester	06
Date & Session	20/5/22 AN
No. of Pages used	23

Roll Number

Register Number: 4 2 1 4 1 0 4 0 1 6

Question Paper Set	A	B		
Internal Test	I	II	III	
Year of Study	I	II	III	IV
Class	A	B		

All particulars given are verified

Signature of the Hall Superintendent with date

Name of the Hall Superintendent: PRIYAS

DO NOT WRITE THE REGISTER NUMBER, COLLEGE CODE AND THE NAME IN ANY OTHER PART OF THE ANSWER BOOK

Office Use Only

(To be filled in by the candidate)

Date: 20/5/22 Session: AN

Subject Code/Title: CS8075 Datawarehousing and Data Mining

Question Paper Set: II No. of Pages used: 23

Office Use Only

Date: 20/5/22 Session: AN Question Paper Set: II

Subject Code / Title: CS8075 Datawarehousing and Data Mining

Office Use Only

Instruction to the Candidate : Put a tick mark (✓) for the questions attended in the tick mark column against each question in V-1

Part - A Question No.	Part - A Marks	PART - B & C Question No.		Marks I	Marks II	Marks III	Marks Total
1	✓2	6	a				
2	✓2		b	✓12			12
3	✓2	7	a				
4	✓2		b	✓2			12
5	✓2	8	a	✓3	✓10		13
			b				
Total	10						37

GRAND TOTAL (IN WORDS): FORTY SEVEN AND A HALF

GRAND TOTAL: 47½/50

Declaration by the Examiner : Verified that all the questions attended by the student are valued and the total is found to be correct

Date	Name of the Examiner	Signature of the Examiner
23/5/22	G. Jeyasanthan	G. Jeyasanthan

Declaration by the Scrutiny : Verified that all the questions attended by the student are valued and the total is found to be correct

Date	Name of the Scrutiny	Signature of the Scrutiny

NEHRU INSTITUTE OF ENGINEERING AND TECHNOLOGY

'Nehru Gardens', Thirumalayampalayam, Coimbatore - 641 105, Tamil Nadu.
ISO 14001 : 2004 Certified, Accredited by NAAC with "B++" Grade, Approved by AICTE-New Delhi,
Affiliated to Anna University, Chennai, Recognized by UGC with 2(f) and 12(B), Accreditated by NBA & NAAC

Anna University Examination - Internal Assessment Test

Office Use Only

Read the instructions given overleaf carefully before filling in the title page. (To be filled in by the candidate)

Roll Number: 7214106022

Register Number: 212141106022

College Code	7214
College Name	Nehru Institute of Engineering and Technology, Coimbatore
Degree / Branch	BE CSE
Semester	VI
Subject Code	CS8075
Date & Session	25/5/22
Subject Title	DWDM
No. of Pages used	12

Put Tick (✓) Mark Here:

Question Paper Set	A	B		
Internal Test	I	II	III	
Year of Study	I	II	III	IV
Class	A	B		

All particulars given are verified

Signature of the Hall Superintendent with date

Name of the Hall Superintendent

DO NOT WRITE THE REGISTER NUMBER, COLLEGE CODE AND THE NAME IN ANY OTHER PART OF THE ANSWER BOOK

Office Use Only

(To be filled in by the candidate)

Date : 25/5/22 Session : .....

Subject Code/Title : CS8075 DWDM

Question Paper Set : .....

No. of Pages used : 12

Office Use Only

(To be filled in by the candidate)

Date : 25/5/22 Session ..... Question Paper Set .....

Subject Code / Title CS8075 DWDM

Instruction to the Candidate : Put a tick mark (✓) for the questions attended in the tick mark column against each question in V-1

Part - A Question No.	Part - A Marks	PART - B & C Question No.		Marks I	Marks II	Marks III	Marks Total	GRAND TOTAL (IN WORDS)
1	2	6	a	12			12	
2	✓		b					
3	✓	7	a	10			10	Four Four Half
4	✓		b					
5	✓	8	a	12½			12½	
			b					GRAND TOTAL
							34½	44½ / 50
Total	10							

Declaration by the Examiner : Verified that all the questions attended by the student are valued and the total is found to be correct

Date	Name of the Examiner	Signature of the Examiner
28/5/22	G. Selvavinayagam	

Declaration by the Scrutiny : Verified that all the questions attended by the student are valued and the total is found to be correct

Date	Name of the Scrutiny	Signature of the Scrutiny

NEHRU INSTITUTE OF ENGINEERING AND TECHNOLOGY

'Nehru Gardens', Thirumalayampalayam, Coimbatore - 641 105, Tamil Nadu.
ISO 14001 : 2004 Certified, Accredited by NAAC with "B++" Grade, Approved by AICTE-New Delhi,
Affiliated to Anna University, Chennai, Recognized by UGC with 2(f) and 12(B), Accreditated by NBA & NAAC

Anna University Examination - Internal Assessment Test

Read the instructions given overleaf carefully before filling in the title page. (To be filled in by the candidate)

College Code	7214
College Name	Nehru Institute of Engineering and Technology, Coimbatore
Degree / Branch	BE COMPUTER SCIENCE AND ENGINEERING
Semester	06
Subject Code	CS8075
Date & Session	25/05/22 AN
Subject Title	Data warehousing and Data mining
No. of Pages used	10

Roll Number

Register Number: 7214191040311

Put Tick (✓) Mark Here:

Question Paper Set	A	B		
Internal Test	I	II ✓	III	
Year of Study	I	II	III ✓	IV
Class	A	B		

All particulars given are verified

Signature of the Hall Superintendent with date

Name of the Hall Superintendent

DO NOT WRITE THE REGISTER NUMBER, COLLEGE CODE AND THE NAME IN ANY OTHER PART OF THE ANSWER BOOK

Office Use Only

(To be filled in by the candidate)

Date : 25-05-22 Session : AN

Subject Code/Title : CS8075 Data warehousing and Data mining

Question Paper Set : B No. of Pages used : 10

Office Use Only

(To be filled by the candidate)

Date : 25-05-22 Session AN Question Paper Set B

Subject Code / Title CS8075 Data warehousing and Data mining

Office Use Only

Instruction to the Candidate : Put a tick mark (✓) for the questions attended in the tick mark column against each question in V-1

Part - A		PART - B & C						GRAND TOTAL (IN WORDS)
Question No.	Marks	Question No.		Marks I	II	III	Total	
1	✓ 2	6	a	✓ 10			10	
2	✓ 2		b					
3	✓ 2	7	a	✓ 11			11	
4	✓ 2		b					
5	✓ 2	8	a	✓ 11½	✓		11½	
			b					
								GRAND TOTAL
								42½
Total	10						32½	

V-1

Declaration by the Examiner : Verified that all the questions attended by the student are valued and the total is found to be correct

Date	Name of the Examiner	Signature of the Examiner
25/5/22	G. Jeevarathnam	G. Jeev

Declaration by the Scrutiny : Verified that all the questions attended by the student are valued and the total is found to be correct

Date	Name of the Scrutiny	Signature of the Scrutiny

60

NEHRU INSTITUTE OF ENGINEERING AND TECHNOLOGY

Nehru Gardens, Thirumalayampalayam, Coimbatore – 641 105
An ISO 14001: 2015 & 9001:2015 Certified Institution
Approved by AICTE, New Delhi & Affiliated to Anna University, Chennai
Accredited by NAAC, Recognized by UGC with 2(F) & 12(B)
NBA Accredited UG Courses: AERO | CSE | MECH

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

B.E./B.Tech.(Full Time) – Internal Assessment Examination - II, May 2022
REGULATION – 2017 SET II
CS8075 – DATA WAREHOUSING AND DATA MINING

Course Instructor	:	Mr.JEEVANANTHAM G	Portion	:	1.5 Units
Year / Semester	:	IV / VIII	Date / Session	:	2005.2022 / AN
Duration	:	90 minutes	Max. Marks	:	50

Knowledge Level	K1: Remembering	K2: Understanding	K3: Applying	K4: Analyzing	K5: Evaluating	K6: Creating
Course Outcomes	C315.2	Apply suitable pre-processing and visualization techniques for data analysis.				
	C315.3	Apply frequent pattern and association rule mining techniques for data analysis.				
	C315.4	Apply appropriate classification and clustering techniques for data analysis.				

PART –A (Answer ALL questions) (5×2 = 10 Marks)

No.	Questions	KL / CO
1.	List the steps involved in the process of KDD. How does it relate to data mining?	K1 / C315.2
2.	Show how the attribute selection set is important in data reduction.	K3 / C315.2
3.	Define an efficient procedure for cleaning the noisy data.	K1 / C315.2
4.	How would you evaluate accuracy of a classifier?	K1 / C315.3
5.	Illustrate support vector machine with example.	K3 / C315.3

PART-B (Answer ALL questions) (2×13 = 26 Marks)

a.	**Explain** and Apply the Apriori algorithm for discovering frequent item sets of the table. Trans ID / Items Purchased 101 / Milk,bread,eggs 102 / Milk,juice 103 / Juice,butter 104 / Milk,bread,eggs 105 / Coffee,eggs 106 / Coffee 107 / Coffee,Juice 108 / Milk,bread,cookies,eggs 109 / Cookies,butter 110 / Milk,bread Use 2 for the minimum support value.Illustrate each step of the AprioriAlgorithm.	K4 / C315.3
	(Or)	
b.	Compare Pattern Mining in Multilevel and Multidimensional Space.	K4 / C315.1
a.	Explain in detail about Constraint Based Frequent Pattern Mining.	K4 / C315.1
	(Or)	
b.	Classify the Use of Frequent Pattern in Data Mining.	K2 / C315.1

PART-C (1×14 = 14 Marks)

8.a.	(i)List the application area of data mining.(4)	K2 / C315.2
	(ii)Explain in detail about Decision Tree Induction.(10)	K3 / C315.4
	(Or)	
8.b.	(i)What is data? How different type of data and attributes can be designed?	K2 / C315.2
	(ii)Discuss in detail about Bayesian Classification.(10)	K2 / C315.4

Course Coordinator Examcell HoD 10/5/2022

Vision:
To produce highly competent and innovative Computer Professionals to meet the global demands.

Mission:
To impart quality education by creative teaching learning process.
To be technically competent, ethical and socially responsible throughout the professional career.
To inculcate leadership qualities and entrepreneurship culture to meet the global standards.

NEHRU INSTITUTE OF ENGINEERING AND TECHNOLOGY

Nehru Gardens, Thirumalayampalayam, Coimbatore – 641 105
An ISO 14001: 2015 & 9001:2015 Certified Institution
Approved by AICTE, New Delhi & Affiliated to Anna University, Chennai
Accredited by NAAC, Recognized by UGC with 2(F) & 12(B)
NBA Accredited UG Courses: AERO | CSE | MECH

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

B.E./B.Tech.(Full Time) – Internal Assessment Examination - II, May 2022
REGULATION – 2017 SET I
CS8075 – DATA WAREHOUSING AND DATA MINING

REVIEW QUESTIONS

PART –A

1.	Evaluate the major tasks of data preprocessing.
2.	Classify different types of reductions.
3.	Distinguish between data cleaning and noisy data.
4.	Formulate the principle frequent itemset and closed itemset.
5.	Discuss association rule mining
6.	List the steps involved in the process of KDD. How does it relate to data mining?
7.	Show how the attribute selection set is important in data reduction.
8.	Define an efficient procedure for cleaning the noisy data.
9.	How would you evaluate accuracy of a classifier?
10.	Illustrate support vector machine with example.

PART –B & PART –C

6.	Find all frequent item sets for the given training set using Apriori Algorithm respectively. TID ITEMS BROUGHT 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 }
7.	Classify the Pattern Evaluation Method in detail.
8.	Illustrate the use of Mining Frequent Patterns.
9.	Identify the use of Associations and Correlations.
10.	(i)State and Explain the various classification of data mining systemswith example.(4) (ii)Explain the Classification by back propagation.(10)
11.	Generalize in detail how data mining system are classified.(4) Discuss in detail about Support Vector Machine.(10)
12.	**Explain** and Apply the Apriori algorithm for discovering frequent item sets of the table.

Trans ID	Items Purchased
101	Milk,bread,eggs
102	Milk,juice
103	Juice,butter
104	Milk,bread,eggs
105	Coffee,eggs
106	Coffee
107	Coffee,Juice
108	Milk,bread,cookies,eggs
109	Cookies,butter
110	Milk,bread

Vision
To produce highly competent and innovative **Computer Professionals** to meet the global demands
Mission
- To impart quality education by **creative teaching learning** process.
- To be technically **competent, ethical and socially responsible** throughout the professional career.

To inculcate **leadership qualities and entrepreneurship culture** to meet the global standards.

	Use 2 for the minimum support value.Illustrate each step of the AprioriAlgorithm.
13.	Compare Pattern Mining in Multilevel and Multidimensional Space.
14.	Explain in detail about Constraint Based Frequent Pattern Mining.
15.	Classify the Use of Frequent Pattern in Data Mining.
16.	(i)List the application area of data mining.(4) (ii)Explain in detail about Decision Tree Induction.(10)
17.	(i)What is data? How different type of data and attributes can be designed? (ii)Discuss in detail about Bayesian Classification.(10)

Staff Incharge

NEHRU INSTITUTE OF ENGINEERING & TECHNOLOGY

"Nehru Gardens" Thirumalayampalayam, Coimbatore – 641 105

An ISO 14001: 2015 & 9001:2015 Certified Institution

(Approved by AICTE and Affiliated to Anna University, Chennai)

Accredited by NAAC and Recognized by UGC under section 2(f) and 12(B)

NBA Accredited UG Courses: AERO, CSE, MECH

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

Academic Year : 2021-2022 (Even Semester)

INTERNAL TEST I

Class/Sem : III Yr / V Sem

Sub code /Sub : CS8075 /DATAWAREHOUSING AND DATAMINING

Course Instructor : Mr.G.Jeevanantham

Regulation : 2017

Batch: 2019-2023

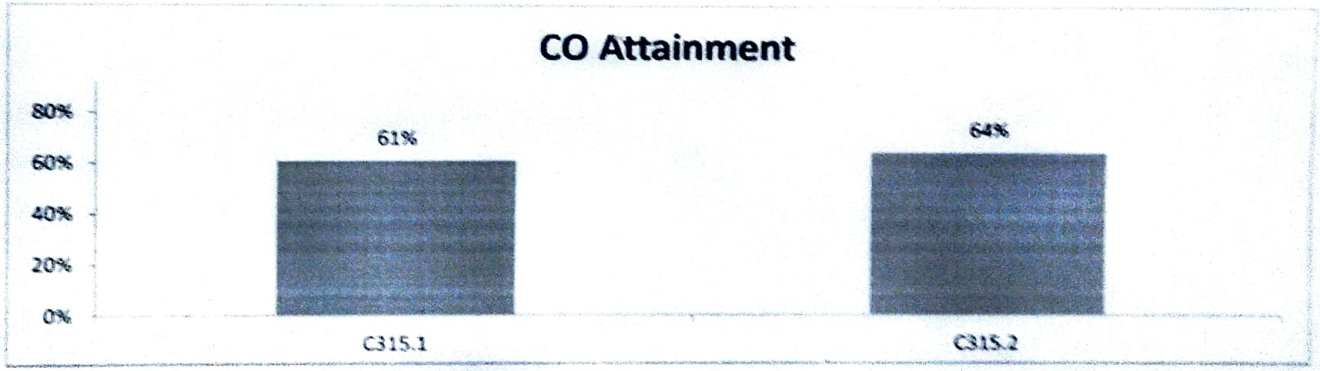
COURSE OUTCOMES							MARKS
	C315.1	Q1	Q2	Q6			30
	C315.2	Q3	Q4	Q5			20
						TOTAL	50

S. No.	Reg. No.	Name	Q1 (2)	Q2 (2)	Q3 (2)	Q4 (2)	Q5 (2)	Q6 (13)	Q7 (13)	Q8 (14)	Total (50)	C315.1 (30)	C315.2 (20)	Total (100)
1	721419104001	ABBIREDDY KIRANMAYEE	2	2	2	2	2	11	10	6	37	25	12	74
2	721419104002	ABISHEAK A	1	0	1	0	0	3	3	0	8	7	1	16
3	721419104003	AKASH GOPINATH	1.5	0	0	0	0	10	9	10	30.5	20.5	10	61
4	721419104004	ALLEN JOSHY	2	2	2	1	0	11	0	12	30	15	15	60
5	721419104005	AMALDAS K K	2	0	2	2	1.5	0	0	1	8.5	2	6.5	17
6	721419104006	ANUPAMA K	2	2	2	2	2	13	13	13.5	49.5	30	19.5	99
7	721419104007	ASWIN S	2	2	2	0	2	11	0	6	25	15	10	50
8	721419104008	BHARATH P	0	2	1	1.5	0	10	9	0	23.5	21	2.5	47
9	721419104009	BHUVANESH U J	1.5	2	2	2	2	6	8	1.5	25	17.5	7.5	50
10	721419104011	DHANIREDDY NAGA MOHAN REDDY	1.5	2.5	2	2	2	6	4	5	25	14	11	50
11	721419104012	DHANUSH S	2	0	2	2	2	4	1	7	20	7	13	40
12	721419104014	FATHIMA MINHA K S	2	2	2	2	2	11	11	11	43	26	17	86
13	721419104015	GOPIKISHORE S	1	2	1	2	2	9	10	6	33	22	11	66
14	721419104016	GREESHMA M NAIR	2	2	2	2	2	13	12	13	48	29	19	96
15	721419104017	HARIHARAN R	2	2	2	2	2	2	4	9	25	10	15	50
16	721419104018	HEMANTH R	2	2	2	1	1	10	0	10	28	14	14	56

20	721419104022	KODAVATIKANTI PRAMOD	2	2	2	2	2	7	1.5	7	25.5	12.5	13	51
21	721419104023	KOUSIK M	2	2	1	1	2	3	10	4	25	17	8	50
22	721419104024	MEGHAL R	1.5	2	2	2	2	8	4	7	28.5	15.5	13	57
23	721419104025	MRIDANI M B	1.5	1	2	2	1	4	10	10	31.5	16.5	15	63
24	721419104026	MUDUGURU SRIVATSAVA	2	2	2	2	2	5	4	5	25	13	11	50
25	721419104027	MUKIL VARATHAN M	2	2	2	2	2	5	5	5	25	14	11	50
26	721419104028	NAGARAJAN S	2	2	2	2	1	6	5	5	25	15	10	50
27	721419104029	NALLABALLE VENKATESWARLU	2	1	2	1	2	5	6	6	25	14	11	50
28	721419104030	NARALA JAYADEEP	2	2	1	2	1	5	6	6	25	15	10	50
29	721419104031	NAVANEETH KRISHNAN R V	2	2	1	0	0	8	8	7	28	20	8	56
30	721419104032	PAVITHRA V	2	0	1	0	0	10	6	12	31	18	13	62
31	721419104033	PENUMARTHI NEMA SRI	2	2	2	2	2	11	11	11	43	26	17	86
32	721419104034	RAJEESH R	2	2	1	2	2	11	4	10	34	19	15	68
33	721419104035	RAMANABOINA VENKATA SUBBAIAH	2	2	2	2	2	11	8	11	40	23	17	80
34	721419104036	ROHAN MURALI NAIR	1.5	1	1	0	0	11	10	8	32.5	23.5	9	65
35	721419104037	ROHITKRISHNA S	0	2	2	2	2	10	0	0	18	12	6	36
36	721419104038	SAHAL H	0.5	0	0	0	0	11	9	9	29.5	20.5	9	59
37	721419104039	SAIRAM P	1	0	0	0	0	7	4	0	12	12	0	24
38	721419104040	SAMSON RAJA R S	2	2	2	2	2	8	0	0	18	12	6	36
39	721419104041	SANAM SIVA KRISHNA RAJU	1	1	1	1	1	9	0	0	14	11	3	28
40	721419104042	SANOJ V	2	2	1	2	1	6	5	6	25	15	10	50
41	721419104043	SARATH A	2	2	0	0.5	0	11	3	9	27.5	18	9.5	55
42	721419104044	SATHISH K	2	2	1	2	1	5	6	6	25	15	10	50
43	721419104045	SIJO N	2	2	2	2	2	12	10	11	43	26	17	86
44	721419104046	SREERAG K DAS	2	1	1	1	1	6	8	5	25	17	8	50
45	721419104047	SREERAM A M	2	2	1	1	0	10	11	11	38	25	13	76
46	721419104048	SUDHAKARAN C	1	0	2	2	2	10	0	0	17	11	6	34
47	721419104049	TAMILINBAM M	0	1	1	2	2	1	0	2	9	2	7	18
48	721419104050	VARUNKUMAR S	0	2	1	0	0	0	0	10	13	2	11	26
49	721419104051	VASANTH B	2	2	1	1	2	6	6	5	25	16	9	50

50	721419104052	VASUNDHARA K	2	2	2	2	2	11	0	8	29	15	14	58
51	721419104053	VISHWANATHAN R K									11	0	0	22
52	721419104054	VISWA S	2	2	2	2	1	6	5	5	25	15	10	50
53	721419104056	YOGAVIJAY E	2	2	1	1	2	8	4	8	28	16	12	56
54	721419104302	NAYANA P	2	2	0	2	2	4	3	0	15	11	4	30
55	721419104303	RESHMA R	0	0	0	0	0	0	0	0	0	0	0	0
56	721419104701	HARIHARAN S	2	2	2	2	2	2	11	4	36	17	10	72
57	721419104702	HITHAYADULLAH H	0	0	0	0	0	0	0	2	13	0	2	26
NO. of Students Attended												56	56	56
No. of Students Gained												34	36	
Percentage of CO Achieved												61%	64%	

CO	Percentage
C315.1	61%
C315.2	64%



15/4/22
Course Instructor

HOD

Data Warehousing And Data Mining

Assignment - I

10

G. Sudha

E. Yogavijay

7214191040 56

Data warehouse schema for decision support :

Schema is a logical discription of the entire database. It includes the name and description of records of all record types including all aesociated data. items and aggregrates a database uses relational model, while a data warehouse uses Star, snowflake and Fact Consellation Schema.

STAR SCHEMA :

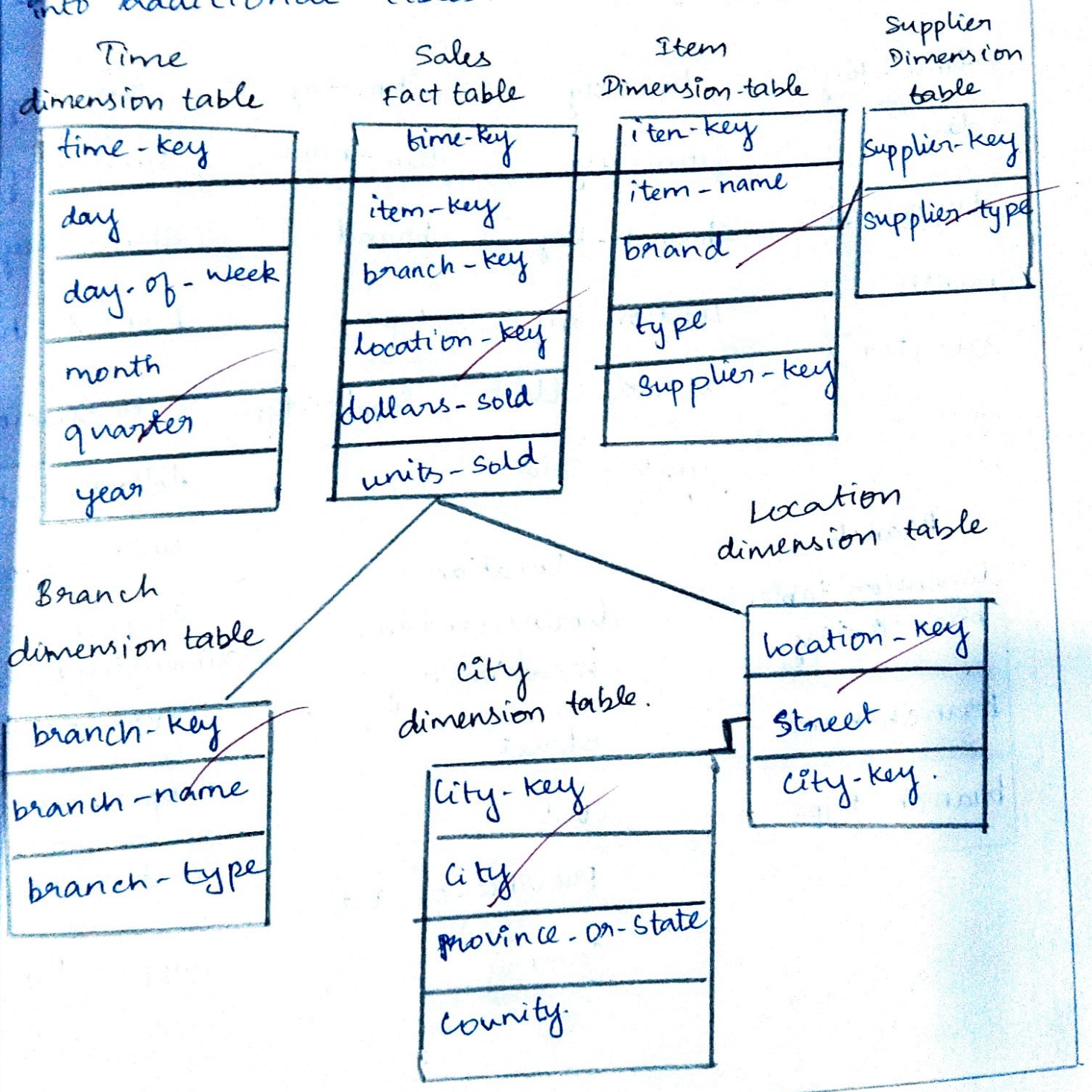
Each dimension in a star schema is represented with only one dimension table.

This dimension table contains the set of attributes.

Snow flake schema:-

Some dimensional tables in the snow flake schema are normalized.

The normalization splits up the data into additional tables.



NEHRU INSTITUTE OF ENGINEERING & TECHNOLOGY

"Nehru Gardens" Thirumalayampalayam, Coimbatore – 641 105

An ISO 14001: 2015 & 9001:2015 Certified Institution

(Approved by AICTE and Affiliated to Anna University, Chennai)

Accredited by NAAC and Recognized by UGC under section 2(f) and 12(B)

NBA Accredited UG Courses: AERO, CSE, MECH

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

Academic Year : 2021-2022 (Even Semester)

INTERNAL TEST I

Class/Sem : III Yr / V Sem

Sub code /Sub : CS8075 /DATAWAREHOUSING AND DATAMINING

Course Instructor : Mr.G.Jeevanantham

Regulation : 2017

Batch: 2019-2023

CATEGORY A

S.No	Reg.No	Student Name	Marks
1	721419104001	ABBIREDDY KIRANMAYEE	74
2	721419104006	ANUPAMA K	99
3	721419104014	FATHIMA MINHA K S	86
4	721419104016	GREESHMA M NAIR	96
5	721419104033	PENUMARTHI NEMA SRI	86
6	721419104035	RAMANABOINA VENKATA SUBBAIAH	80
7	721419104045	SIJO N	86
8	721419104047	SREERAM A M	76
9	721419104701	HARIHARAN S	72

CATEGORY B

S.No	Reg.No	Student Name	Marks
10	721419104003	AKASH GOPINATH	61
11	721419104004	ALLEN JOSHY	60
12	721419104007	ASWIN S	50
13	721419104009	BHUVANESH U J	50
14	721419104011	DHANIREDDY NAGA MOHAN REDDY	50
15	721419104015	GOPIKISHORE S	66
16	721419104017	HARIHARAN R	50
17	721419104018	HEMANTH R	56
18	721419104019	JAYA SRINATH V	61
19	721419104020	JOSHAN K	50
20	721419104021	KAVIYA S	50
21	721419104022	KODAVATIKANTI PRAMOD	51
22	721419104023	KOUSIK M	50
23	721419104024	MEGHAL R	57
24	721419104025	MRIDANI M B	63
25	721419104026	MUDUGURU SRIVATSAVA	50
26	721419104027	MUKIL VARATHAN M	50
27	721419104028	NAGARAJAN S	50
28	721419104029	NALLABALLE VENKATESWARLU	50
29	721419104030	NARALA JAYADEEP	50
30	721419104031	NAVANEETH KRISHNAN R V	56
31	721419104032	PAVITHRA V	62
32	721419104034	RAJEESH R	68

NEHRU INSTITUTE OF ENGINEERING AND TECHNOLOGY

'Nehru Gardens', Thirumalayampalayam, Coimbatore - 641 105, Tamil Nadu.
ISO 14001 : 2004 Certified, Accredited by NAAC with "B++" Grade, Approved by AICTE-New Delhi,
Affiliated to Anna University, Chennai, Recognized by UGC with 2(f) and 12(B), Accreditated by NBA & NAAC

Anna University Examination - Internal Assessment Test

Read the instructions given overleaf carefully before filling in the title page. (To be filled in by the candidate)

Office Use Only

College Code	7214
College Name	Nehru Institute of Engineering and Technology, Coimbatore
Degree / Branch	B.E computer science and engineering
Semester	VI
Subject Code	CS8075
Date & Session	08.04.22 AN
Subject Title	Data warehousing and data mining
No. of Pages used	19

Roll Number

Register Number: 7 2 1 4 1 9 1 0 4 0 0 6

Put Tick (✓) Mark Here:

Question Paper Set	A	B		
Internal Test	✓	II	III	
Year of Study	I	II	✓ III	IV
Class	A	B		

All particulars given are verified

Signature of the Hall Superintendent with date

Name of the Hall Superintendent

DO NOT WRITE THE REGISTER NUMBER, COLLEGE CODE AND THE NAME IN ANY OTHER PART OF THE ANSWER BOOK

Office Use Only

(To be filled in by the candidate)

Date : 08.04.2022 Session : AN

Subject Code/Title : CS8075 Data Warehousing and Data Mining

Question Paper Set :

No. of Pages used : 19

Office Use Only

(To be filled in by the candidate)

Date : 08.04.2022 Session AN Question Paper Set

Subject Code / Title CS8075 Data Warehousing and Data Mining

Instruction to the Candidate : Put a tick mark (✓) for the questions attended in the tick mark column against each question in V-1

Part - A Question No.	Part - A Marks	PART - B & C Question No.		Marks i	Marks ii	Marks iii	Marks Total	GRAND TOTAL (IN WORDS)
1	2	6	a					
2	2		b	11			11	four five
3	2	7	a	12			12	
4	2		b					
5	2	8	a	12			12	
			b					GRAND TOTAL
								45/50
Total	10						35	

Declaration by the Examiner : Verified that all the questions attended by the student are valued and the total is found to be correct

Date	Name of the Examiner	Signature of the Examiner
8/4/22	G. Selvanathan	G. Selvanathan

Declaration by the Scrutiny : Verified that all the questions attended by the student are valued and the total is found to be correct

Date	Name of the Scrutiny	Signature of the Scrutiny

DATA WAREHOUSING AND DATA MINING ASSIGNMENT

9

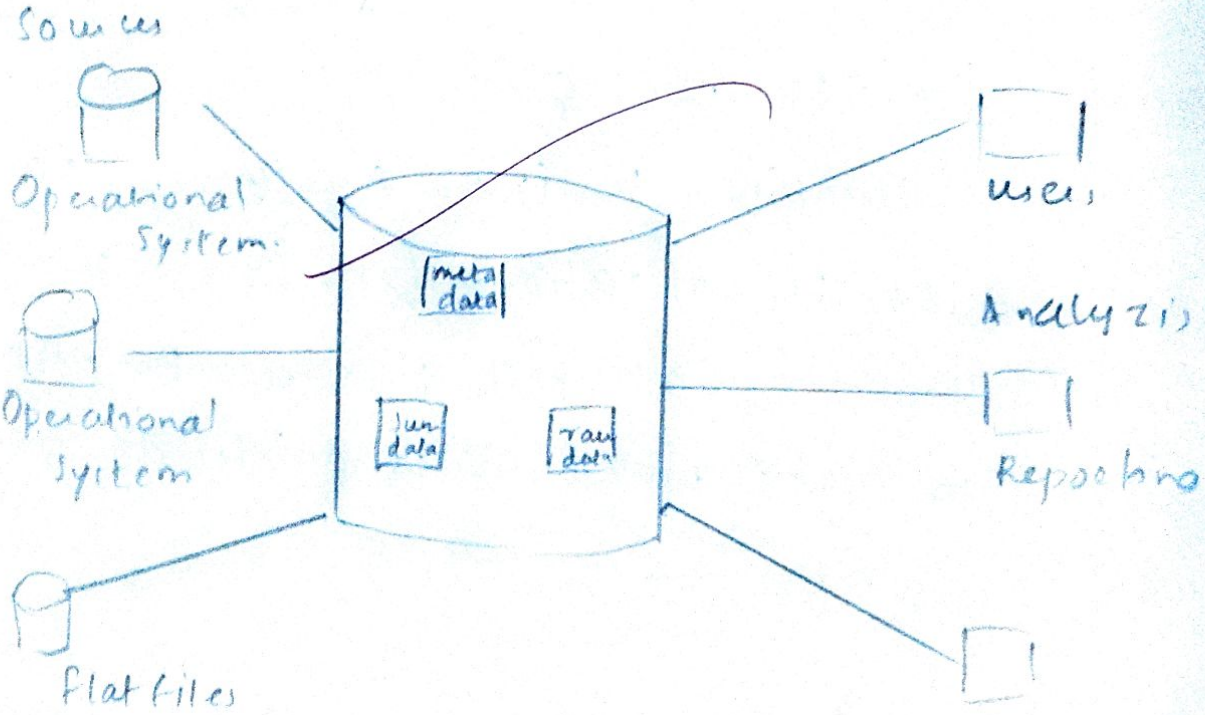
G. Anupama

ANUPAMA.K
721419104006
IIIrd YEAR
B.E CSE

PART A

DataMart	Metadata
- Data may or maynot be informative	- Metadata is always informative
- Data may or maynot been processed	- It is always a processed data
- Any type of information that is solved in computer	- It describes relevent information about the data

Architecture of Datawarehouse



PART B

Three-tier Data warehouse Architecture

Data warehouse usually have a three tier architecture that includes:

1. Bottom Tier (Data warehouse server)
2. Middle Tier (OLAP server)
3. Top tier (Front end tools)

A bottom tier that consist of the data warehouse server which is almost always an RDBMS. It may include several specialized data marts and a metadata repository.

Data from operational databases and external sources (such as user profile data provided by external consultant) are extracted using application program interface called a gateway.

Examples of gateways contains ODBC (Open database connection) and OLE-DB (Open linking and embedding for databases)

A middle tier which consist of an OLAP server for fast querying of the data warehouse.

Data base Architecture of parallel processing

There are 3 DBMS software architecture styles for parallel processing

1. shared memory architecture
2. shared disk architecture
3. shared nothing architecture.

shared Memory Architecture

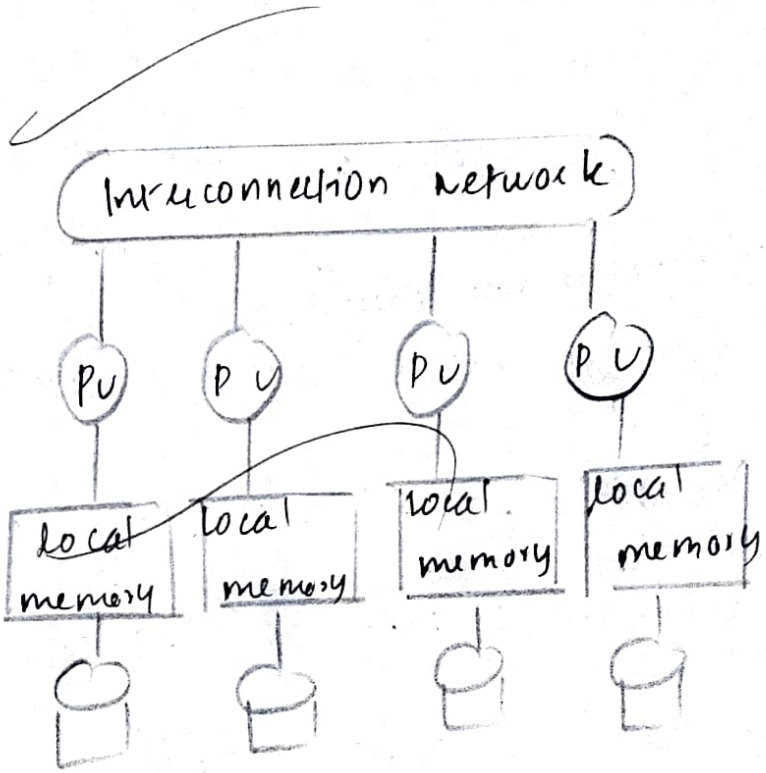
→ Multiple PUs share memory

→ Each PU has full access to all shared memory through a common bus.

Tightly coupled shared memory systems illustrated in following figure,

Shared Nothing Architecture

- shared nothing systems provide for incremental growth
- system growth is practically unlimited
- MPP, are good for read only databases and decision support application



NEHRU INSTITUTE OF ENGINEERING AND TECHNOLOGY

Nehru Gardens, Thirumalayampalayam, Coimbatore - 641 105, Tamil Nadu
ISO 14001 : 2004 Certified, Accredited by NAAC with "B++" Grade, Approved by AICTE-New Delhi,
Affiliated to Anna University, Chennai, Recognized by UGC with 2(f) and 12(B), Accreditated by NBA & NAAC

Anna University Examination - Internal Assessment Test

Office Use Only

Read the instructions given overleaf carefully before filling in the title page. (To be filled in by the candidate)

College Code	7214
College Name	Nehru Institute of Engineering and Technology, Coimbatore
Degree / Branch	B.E - C.S.E
Semester	6
Subject Code	CS8075
Date & Session	8/4/22-AN
Subject Title	Data Warehouse & Data mining
No. of Pages used	14

Roll Number: 7214191040001

Register Number: 7214191040001

Put Tick (✓) Mark Here

Question Paper Set	A ✓	B		
Internal Test	I	II	III ✓	
Year of Study	I	II	III ✓	IV
Class	A	B		

All particulars given are verified

Signature of the Hall Superintendent with date

Name of the Hall Superintendent

DO NOT WRITE THE REGISTER NUMBER, COLLEGE CODE AND THE NAME IN ANY OTHER PART OF THE ANSWER BOOK

Office Use Only

(To be filled in by the candidate)

Date : 08.04.2022 Session : After Noon

Subject Code/Title : CS8075 Data Warehouse and Data mining.

Question Paper Set :

No. of Pages used : 14

Office Use Only

(To be filled by the candidate)

Date : 08.04.2022 Session : After Noon Question Paper Set

Subject Code / Title : CS8075 Data Warehouse and Data Mining.

Instruction to the Candidate : Put a tick mark (✓) for the questions attended in the tick mark column against each question in V-1

Part - A		PART - B & C					GRAND TOTAL (IN WORDS)
Question No.	Marks	Question No.		Marks i	ii	iii	Total	
1 ✓	2	6	a					Thirty seven
2 ✓	2		b ✓	11			11	
3 ✓	2	7	a ✓	10			10	
4 ✓	2		b					
5 ✓	2	8	a ✓	6			6	
			b					GRAND TOTAL
								37/50
Total	10						27	

Declaration by the Examiner : Verified that all the questions attended by the student are valued and the total is found to be correct

Date	Name of the Examiner	Signature of the Examiner
8/4/22	G. Jeevanantham	G. Jeeva

Declaration by the Scrutiny : Verified that all the questions attended by the student are valued and the total is found to be correct

Date	Name of the Scrutiny	Signature of the Scrutiny

NEHRU INSTITUTE OF ENGINEERING AND TECHNOLOGY

'Nehru Gardens', Thirumalayampalayam, Coimbatore - 641 105, Tamil Nadu.
ISO 14001 : 2004 Certified, Accredited by NAAC with "B++" Grade, Approved by AICTE-New Delhi,
Affiliated to Anna University, Chennai, Recognized by UGC with 2(f) and 12(B), Accreditated by NBA & NAAC

Anna University Examination - Internal Assessment Test

Read the instructions given overleaf carefully before filling in the title page. (To be filled in by the candidate)

Office Use Only

College Code	7214
College Name	Nehru Institute of Engineering and Technology, Coimbatore
Degree / Branch	BE / CSE
Semester	VI
Subject Code	CS8075
Date & Session	8/4/22 AN
Subject Title	Data warhousing and data mining
No. of Pages used	

Roll Number: 7214191040053

Register Number: 7214191040053

Put Tick (✓) Mark Here:

Question Paper Set	A	B		
Internal Test	I	II ✓	III	
Year of Study	I	II	III ✓	IV
Class	A	B		

All particulars given are verified

Signature of the Hall Superintendent with date

Name of the Hall Superintendent

DO NOT WRITE THE REGISTER NUMBER, COLLEGE CODE AND THE NAME IN ANY OTHER PART OF THE ANSWER BOOK

(To be filled in by the candidate)

Date : 8/4/22 Session : AN

Subject Code/Title : CS8075 Data warhousing and mining

Question Paper Set :

No. of Pages used :

Office Use Only

(To be filled by the candidate)

Date : 8/4/22 Session : AN Question Paper Set

Subject Code / Title : CS8075 Data warhousing and mining

Office Use Only

Instruction to the Candidate : Put a tick mark (✓) for the questions attended in the tick mark column against each question in V-1

Part - A Question No.	Part - A Marks	PART - B & C Question No.		PART - B & C Marks i	ii	iii	Total	GRAND TOTAL (IN WORDS)
1	✓	6	a	✓			✓	ONE ONE
2	✓		b	✓			✓	
3	✓	7	a	✓			✓	
4	2		b	✓			✓	
5	1	8	a	8			8	GRAND TOTAL
			b					
Total	03						08	11/50

Declaration by the Examiner : Verified that all the questions attended by the student are valued and the total is found to be correct

Date	Name of the Examiner	Signature of the Examiner
8/4/22	Q. Jeyanthi	

Declaration by the Scrutiny : Verified that all the questions attended by the student are valued and the total is found to be correct

Date	Name of the Scrutiny	Signature of the Scrutiny

ANNA UNIVERSITY :: CHENNAI - 600 025
OFFICE OF THE CONTROLLER OF EXAMINATIONS
Assessment Details Entered - Report : By subject
APRIL / MAY EXAMINATION,2022 - EXAMINATIONS

Inst Code & Name : 7214 - NEHRU INSTITUTE OF ENGINEERING AND TECHNOLOGY

Subject Code / Name : CS8075 : Data Warehousing and Data Mining

University : AUC

Semester : 06

Register No.	Name of the Student	Attend Period1	Total Period1	Attend Period2	Total Period2	IM 2	Attend Period3	Tot Period3	IM 3	Attend Period4	Total Period4	IM 4
721419104001	ABBIREDDY KIRANMAYEE	14	14	14	14	74						
721419104002	ABISHEAK A	13	14	13	14	18						
721419104003	AKASH GOPINATH	14	14	14	14	61						
721419104004	ALLEN JOSHY	12	14	12	14	60						
721419104005	AMALDAS K K	14	14	14	14	17						
721419104006	ANUPAMA K	14	14	14	14	99						
721419104007	ASWIN S	12	14	12	14	50						
721419104008	BHARATH P	14	14	14	14	47						
721419104009	BHUVANESH U J	14	14	14	14	50						
721419104011	DHANIREDDY NAGA MOHAN REDDY	12	14	12	14	50						
721419104012	DHANUSH S	14	14	14	14	40						
721419104014	FATHIMA MINHA K S	14	14	14	14	86						
721419104015	GOPIKISHORE S	14	14	14	14	68						
721419104016	GREESHMA M NAIR	14	14	14	14	96						
721419104017	HARIHARAN R	14	14	14	14	50						
721419104018	HEMANTH R	12	14	12	14	56						
721419104019	JAYA SRINATH V	14	14	14	14	61						
721419104020	JOSHAN K	14	14	14	14	50						
721419104021	KAVIYA S	12	14	12	14	50						
721419104022	KODAVATIKANTI PRAMOD	13	14	13	14	51						
721419104023	KOUSIK M	13	14	13	14	50						
721419104024	MEGHAL R	14	14	14	14	57						
721419104025	MRIDANI M B	14	14	14	14	63						
721419104026	MUDUGURU SRIVATSAVA	12	14	12	14	50						
721419104027	MUKIL VARATHAN M	14	14	14	14	50						
721419104028	NAGARAJAN S	14	14	14	14	50						
721419104029	NALLABALLE VENKATESWARLU	14	14	14	14	50						
721419104030	NARALA JAYADEEP	12	14	12	14	50						
721419104031	NAVANEETH KRISHNAN R V	12	14	12	14	56						
721419104032	PAVITHRA V	14	14	14	14	62						
721419104033	PENUMARTHI HEMA SRI	14	14	14	14	86						
721419104034	RAJEESH R	14	14	14	14	68						
721419104035	RAMANABOINA VENKATA SUBBAIAH	14	14	14	14	80						
721419104036	ROHAN MURALI NAIR	14	14	14	14	65						
721419104037	ROHITKRISHNA S	14	14	14	14	36						
721419104038	SAHAL H	14	14	14	14	59						
721419104039	SAIRAM P	13	14	13	14	24						
721419104040	SAMSON RAJA R S	14	14	14	14	36						
721419104041	SANAM SIVA KRISHNA RAJU	14	14	14	14	28						
721419104042	SANOJ V	14	14	14	14	50						
721419104043	SARATH A	13	14	13	14	55						
721419104044	SATHISH K	12	14	12	14	50						
721419104045	SIJO N	13	14	13	14	86						
721419104046	SREERAG.K.DAS	14	14	14	14	50						
721419104047	SREERAM A M	14	14	14	14	76						
721419104048	SUDHAKARAN C	14	14	14	14	34						
721419104049	TAMILINBAM M	14	14	14	14	18						
721419104050	VARUNKUMAR S	14	14	14	14	26						
721419104051	VASANTH B	14	14	14	14	50						
721419104052	VASUNDHARA K	14	14	14	14	58						
721419104053	VISHWANATHAN R K	14	14	14	14	22						
721419104054	VISWA S	14	14	14	14	50						

ANNA UNIVERSITY : CHENNAI - 600 025
OFFICE OF THE CONTROLLER OF EXAMINATIONS

Assessment Details Entered Report By subject

APRIL / MAY EXAMINATIONS 2022 EXAMINATIONS

Inst Code & Name 7514 NEHRU INSTITUTE OF ENGINEERING AND TECHNOLOGY

(67)

Reg. No.

NEHRU INSTITUTE OF ENGINEERING AND TECHNOLOGY

Nehru Gardens, Thirumalayampalayam, Coimbatore – 641 105
Approved by AICTE, New Delhi & Affiliated to Anna University, Chennai
Accredited by NAAC, Recognized by UGC with 2(F) & 12(B)
NBA Accredited UG Courses: AERO | CSE | MECH

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

B.E./B.Tech.(Full Time) – Internal Assessment Examination - I, APRIL 2022

REGULATION - 2017

CS8075 – DATA WAREHOUSING AND DATA MINING

Course Instructor	:	Mr. JEEVANANTHAM G	Portion	:	1.5 Units
Year / Semester	:	IV / VIII	Date / Session	:	08.04.2022 / AN
Duration	:	90 minutes	Max. Marks	:	50

Knowledge Level	K1. Remembering	K2. Understanding	K3. Applying	K4. Analyzing	K5. Evaluating	K6. Creating
Course Outcomes	C315.1	Design a Data warehouse system and perform business analysis with OLAP tools.				
	C315.2	Apply suitable pre-processing and visualization techniques for data analysis.				

PART –A (Answer ALL questions) (5×2 = 10 Marks)

Q.No.	Questions	Knowledge Level / Course Outcomes
1.	Differentiate metadata and data mart	K2 / C315.1
2.	Design the data warehouse architecture.	K6 / C315.1
3.	State the use of Machine Learning Techniques.	K1 / C315.2
4.	Indicate the steps involved in Knowledge Discovery Process.	K2 / C315.2
5.	Discuss the term Business Intelligence.	K2 / C315.2

PART-B (Answer ALL questions) (2×13 = 26 Marks)

6.	a.	What is data warehouse? Give the steps for design and construction of Data Warehouses and explain with three tier architecture diagram.	K2 / C315.1
		(Or)	
	b.	Demonstrate in detail about Database architecture for parallel processing.	K3 / C315.1
7.	a.	Explain in detail about OLAP.	K4 / C315.1
		(Or)	
	b.	Discuss in brief about Multidimensional Data Model.	K2 / C315.1

PART-C (1×14 = 14 Marks)

8.	a.	State and Explain the various classifications of data mining Techniques with example.	K3 / C315.2
		(Or)	
	b.	i) Describe the issues of data mining. ii) Describe in detail about the applications of data mining	K2 / C315.2

Course Coordinator Examcell HoD

Vision
To produce highly competent and innovative Computer Professionals to meet the global demands.

Mission
To impart quality education by creative teaching learning process.
To be technically competent, ethical and socially responsible throughout the professional career.
To inculcate leadership qualities and entrepreneurship culture to meet the global standards.

Reg. No.

NEHRU INSTITUTE OF ENGINEERING AND TECHNOLOGY

Nehru Gardens, Thirumalayampalayam, Coimbatore – 641 105
Approved by AICTE, New Delhi & Affiliated to Anna University, Chennai
Accredited by NAAC, Recognized by UGC with 2(F) & 12(B)
NBA Accredited UG Courses: AERO | CSE | MECH

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

B.E./B.Tech.(Full Time) – Internal Assessment Examination - I, APRIL 2022

REGULATION - 2017

CS8075 – DATA WAREHOUSING AND DATA MINING / REVIEW QUESTIONS

PART –A

1.	Differentiate metadata and data mart
2.	Design the data warehouse architecture.
3.	State the use of Machine Learning Techniques.
4.	Indicate the steps involved in Knowledge Discovery Process.
5.	Discuss the term Business Intelligence.
6.	Compare OLTP and OLAP system.
7.	Compare drill down with roll up approach.
8.	Define Data mining. List out the steps in data mining.
9.	Compare Discrete versus Continuous Attributes.
10.	Give the applications of Data Mining.

PART –B & PART –C

1.	What is data warehouse? Give the steps for design and construction of Data Warehouses and explain with three tier architecture diagram.
2.	Demonstrate in detail about Database architecture for parallel processing.
3.	Explain in detail about OLAP.
4.	Discuss in brief about Multidimensional Data Model.
5.	State and Explain the various classifications of data mining Techniques with example.
6.	i) Describe the issues of data mining. ii) Describe in detail about the applications of data mining
7.	Examine the relevant examples discuss multidimensional online analytical processing and multi relational online analytical processing.
8.	Explain about Database architecture.
9.	Describe the need for OLAP and their guidelines.
10.	Illustrate and describe the architecture of MOLAP and ROLAP.
11.	State and Explain Knowledge Discovery Process.
12.	State and Explain the various classifications of data mining systems data mining functionalities in detail.

Staff Incharge

HoD 7/6/22

Vision
To produce highly competent and innovative Computer Professionals to meet the global demands.

Mission
- To impart quality education by creative teaching learning process.
- To be technically competent, ethical and socially responsible throughout the professional career.
- To inculcate leadership qualities and entrepreneurship culture to meet the global standards.

NEHRU INSTITUTE OF ENGINEERING AND TECHNOLOGY

COIMBATORE - 641105.

COURSE EXIT SURVEY FOR THE COURSE OUTCOME (GRADING IN A SCALE OF 5)

Name of the Department : **Department of Computer Science & Engineering**

Course Name : **DATA WAREHOUSING AND DATA MINING**

Course Code : **CS8075**

Class /Sem : **III/V**

Dear Student,
Greetings.....!!!

I would like to thank all of you for sparing your time in filling up this Course Pre-Analysis for the effective conduct of the course. As you know that this survey is meant for knowing the knowledge level of the students with respect to this course, please fill it very carefully.

At this juncture, I am glad to welcome the suggestions from you all (if any).

Questions

C315.1 Design a Data warehouse system and perform business analysis with OLAP tools.

C315.2 Apply suitable pre-processing and visualization techniques for data analysis.

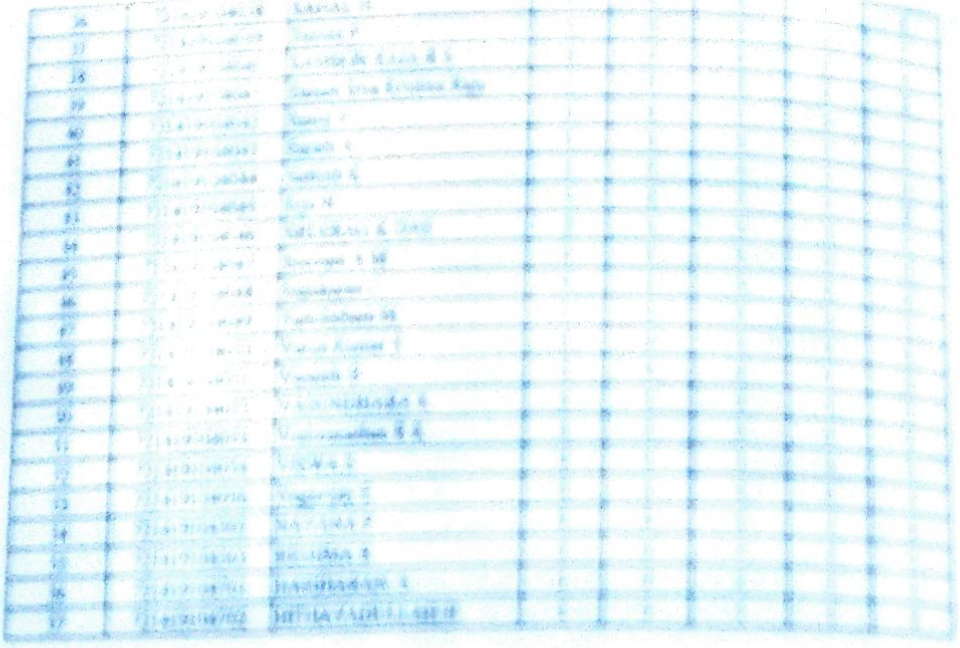
C315.3 Apply frequent pattern and association rule mining techniques for data analysis.

C315.4 Apply appropriate classification and clustering techniques for data analysis.

C315.5 Apply WEGA tool and Clustering Algorithm.

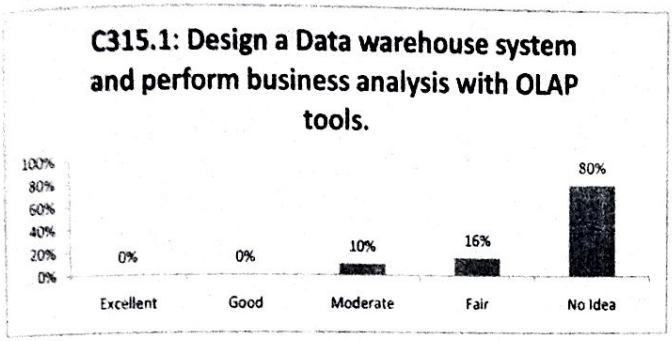
GRADING IN A SCALE OF 5				
Excellent (5)	**Good** (4)	**Moderate** (3)	**Fair** (2)	**No Idea** (1)

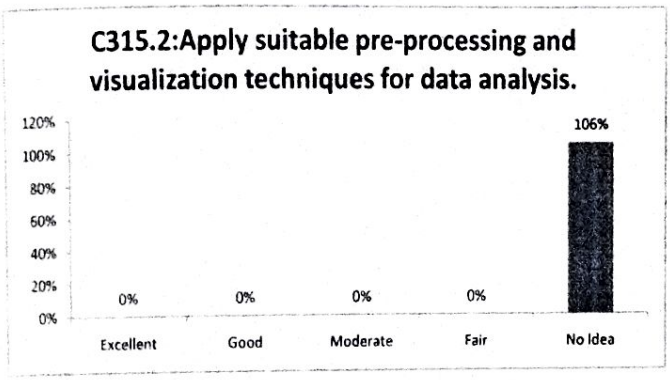
S.No	Reg No	Name of the Student	C315.1	C315.2	C315.3	C315.4	C315.5
1	721419104001	Abbereddy Kiran Mayee	2	1	1	1	1
2	721419104002	Abisheak A	2	1	1	1	1
3	721419104003	Akash Gopinath	3	1	1	1	2
4	721419104004	Allen Joshy	3	1	1	1	2
5	721419104005	Amaldas k.k	1	1	1	1	2
6	721419104006	Anupama K	1	1	1	1	1
7	721419104007	Aswin S	1	1	1	1	2
8	721419104008	Bharath P	1	1	1	1	1
9	721419104009	BHUVANESH U J	1	1	1	1	1
10	721419104011	Dhanireddy Naga Mohan Reddy	1	1	1	1	1
11	721419104012	DHANUSH S	1	1	1	1	1
12	721419104014	Fathima Minha K S	1	1	1	1	1
13	721419104015	Gopi Krishore S	3	1	1	1	2
14	721419104016	Greeshma M Nair	1	1	1	1	2
15	721419104017	Hariharan R	1	1	1	1	1
16	721419104018	Hemanth R	1	1	1	1	1
17	721419104019	JAYA SRINATH V	1	1	1	1	1
18	721419104020	Joshan K	1	1	1	1	1
19	721419104021	KAVIYA S	1	1	1	1	1
20	721419104022	Kodavatikanti Pramod	1	1	1	1	1
21	721419104023	KOUSIK M	1	1	1	1	1
22	721419104024	MEGHAL R	1	1	1	1	1
23	721419104025	MRIDANI M B	1	1	1	1	1
24	721419104026	Sri Vatsava M	1	1	1	1	1
25	721419104027	Mukil Varathan M	2	1	1	1	1
26	721419104028	Nagarajan S	1	1	1	1	1
27	721419104029	NALLABALLE VENKATESWARLU	3	1	2	1	2
28	721419104030	Jayadeep N	2	1	2	1	3
29	721419104031	Navaneeth Krishnan R V	1	1	1	1	2
30	721419104032	Pavithra V	2	1	1	1	1
31	721419104033	Penumarthi Hema Sri	1	1	1	1	1
32	721419104034	RAJEESH R	3	1	1	1	2
33	721419104035	Ramanaboina Venkata Subbaiah	2	1	1	1	1
34	721419104036	ROHAN MURALI NAIR	2	1	1	1	1
35	721419104037	Rohitkrishna S	1	1	1	1	1

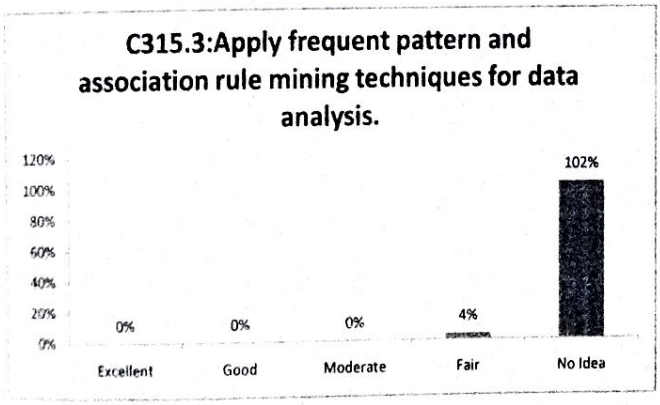


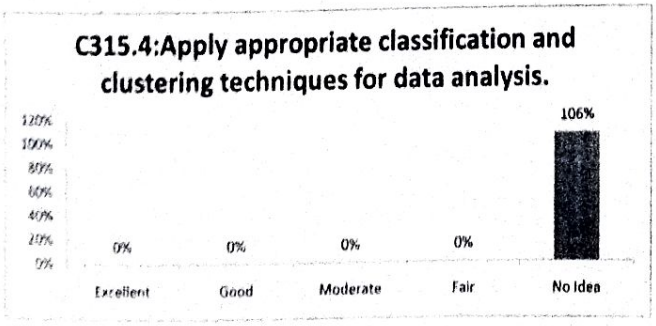
PRE ANALYSIS CHART FOR COURSE OUTCOMES

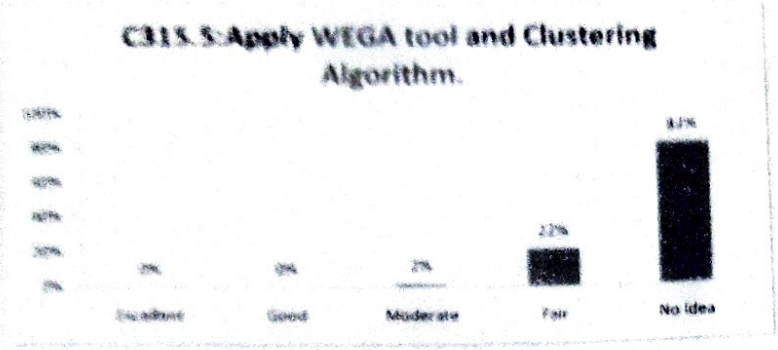
Course Outcomes	Excellent	Good	Moderate	Fair	No Idea	Excellent	Good	Moderate	Fair	No Idea
Design a Data warehouse system and perform business	0	0	5	8	41	0%	0%	10%	16%	80%
Apply suitable pre-processing and visualization	0	0	0	0	54	0%	0%	0%	0%	106%
Apply frequent pattern and association rule mining	0	0	0	2	52	0%	0%	0%	4%	102%
Apply appropriate classification and clustering	0	0	0	0	54	0%	0%	0%	0%	106%
Apply WEKA Tool and Clustering Algorithm	0	0	1	11	42	0%	0%	2%	22%	82%
Pre Analysis of Course outcomes	0	0	1	4	49	0%	0%	2%	8%	95%

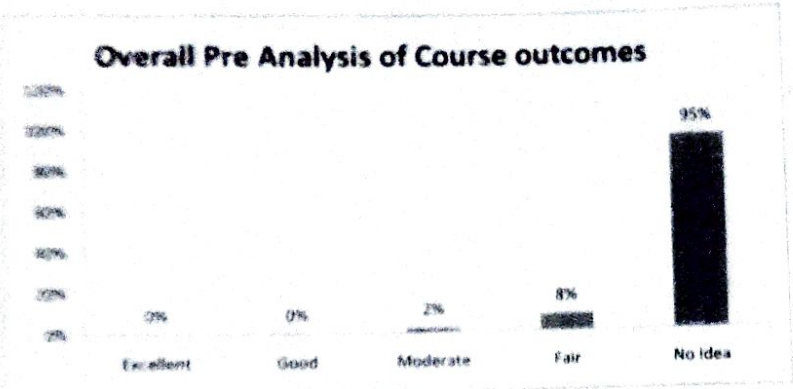









CO1.5.5:Apply WEGA tool and Clustering Algorithm.
0%
0%
2%
22%
83%
Excellent
Good
Moderate
Fair
No idea


Overall Pre Analysis of Course outcomes
0%
0%
2%
8%
95%
Excellent
Good
Moderate
Fair
No idea

HoD

NEHRU INSTITUTE OF ENGINEERING AND TECHNOLOGY

"Nehru Gardens" T. M. PALAYAM, COIMBATORE-105
(Approved by AICTE and Affiliated to Anna University, Chennai)
(Accredited by NAAC, Recognized by UGC with section 2(f) and 12(B))
NBA Accredited UG Courses: Aero, CSE, Mech

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

Course Code & Title : **CS8075 & DATA WAREHOUSING AND DATA MINING**
Course Instructor : Mr. G.Jeevanantham
Class : III Year BE - CSE Semester : VII
Regulation : R2017 Academic Year : 2021-2022 Even Sem

Dear Student,
Greetings....!!!

I would like to thank all of you for sparing your time in filling up this Course Pre-Analysis survey for the effective conduct of the **DATA WAREHOUSING AND DATA MINING** Course. As you know that this survey is meant for knowing the knowledge level of the students with respect to this course, please fill it very carefully.

At this juncture, I am glad to welcome the suggestions from you all (if any).Rate your prior knowledge against the topic mentioned (Please Select the option)

Course Outcomes	Description	Rate your prior knowledge against the topic mentioned				
		Excellent (4)	Good (3)	Moderate (2)	Fair (1)	No idea (0)
C315.1	Your Knowledge about Data warehouse system and perform business analysis with OLAP tools				✓	
C315.2	Your knowledge about suitable pre-processing and visualization techniques for data analysis				✓	
C315.3	Your knowledge frequent pattern and association rule mining techniques for data analysis				✓	
C315.4	Your knowledge on appropriate classification and clustering techniques for data analysis.				✓	
C315.5	Apply Learning Tools and Learning Algorithms			✓		
Suggestions (If any)						

Date: 07/3/2022 Name of the student : SIJO.N Signature of the student

Vision
To produce highly competent and innovative **Computer Professionals** to meet the global demands
Mission
- To impart quality education by **creative teaching learning** process.
- To be technically **competent, ethical and socially responsible** throughout the professional career.
- To inculcate **leadership qualities and entrepreneurship culture** to meet the global standards.

NEHRU INSTITUTE OF ENGINEERING AND TECHNOLOGY

"Nehru Gardens" T. M. PALAYAM, COIMBATORE-105
(Approved by AICTE and Affiliated to Anna University, Chennai)
(Accredited by NAAC, Recognized by UGC with section 2(f) and 12(B))
NBA Accredited UG Courses: Aero, CSE, Mech

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

Course Code & Title : **CS8075 & DATA WAREHOUSING AND DATA MINING**
Course Instructor : Mr. G.Jeevanantham
Class : III Year BE - CSE Semester : VII
Regulation : R2017 Academic Year : 2021-2022 Even Sem

Dear Student,
Greetings....!!!
I would like to thank all of you for sparing your time in filling up this Course Pre-Analysis survey for the effective conduct of the **DATA WAREHOUSING AND DATA MINING** Course. As you know that this survey is meant for knowing the knowledge level of the students with respect to this course, please fill it very carefully.
At this juncture, I am glad to welcome the suggestions from you all (if any).Rate your prior knowledge against the topic mentioned (Please Select the option)

Course Outcomes	Description	Rate your prior knowledge against the topic mentioned				
		Excellent (4)	Good (3)	Moderate (2)	Fair (1)	No idea (0)
C315.1	Your Knowledge about Data warehouse system and perform business analysis with OLAP tools			✓		
C315.2	Your knowledge about suitable pre-processing and visualization techniques for data analysis				✓	
C315.3	Your knowledge frequent pattern and association rule mining techniques for data analysis				✓	
C315.4	Your knowledge on appropriate classification and clustering techniques for data analysis.				✓	
C315.5	Apply Learning Tools and Learning Algorithms				✓	
Suggestions (If any)						

Date: 07/03/2022 Name of the student Sreeraj.K.Das Signature of the student

Vision
To produce highly competent and innovative **Computer Professionals** to meet the global demands
Mission
- To impart quality education by **creative teaching learning** process.
- To be technically **competent, ethical and socially responsible** throughout the professional career.
- To inculcate **leadership qualities and entrepreneurship culture** to meet the global standards.

NEHRU INSTITUTE OF ENGINEERING AND TECHNOLOGY

"Nehru Gardens" T. M. PALAYAM, COIMBATORE-105
(Approved by AICTE and Affiliated to Anna University, Chennai)
(Accredited by NAAC, Recognized by UGC with section 2(f) and 12(B))
NBA Accredited UG Courses: Aero, CSE, Mech

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

Course Code & Title : **CS8075 & DATA WAREHOUSING AND DATA MINING**
Course Instructor : Mr. G.Jeevanantham
Class : III Year BE - CSE Semester : VII
Regulation : R2017 Academic Year : 2021-2022 Even Sem

Dear Student,
Greetings....!!!
I would like to thank all of you for sparing your time in filling up this Course Pre-Analysis survey for the effective conduct of the **DATA WAREHOUSING AND DATA MINING** Course. As you know that this survey is meant for knowing the knowledge level of the students with respect to this course, please fill it very carefully.
At this juncture, I am glad to welcome the suggestions from you all (if any).Rate your prior knowledge against the topic mentioned (Please Select the option)

Course Outcomes	Description	Rate your prior knowledge against the topic mentioned				
		Excellent (4)	Good (3)	Moderate (2)	Fair (1)	No idea (0)
C315.1	Your Knowledge about Data warehouse system and perform business analysis with OLAP tools		✓			
C315.2	Your knowledge about suitable pre-processing and visualization techniques for data analysis				✓	
C315.3	Your knowledge frequent pattern and association rule mining techniques for data analysis				✓	
C315.4	Your knowledge on appropriate classification and clustering techniques for data analysis.				✓	
C315.5	Apply Learning Tools and Learning Algorithms			✓		
Suggestions (If any)						

Date: 07/03/2022 Name of the student [signature] Signature of the student Rajeev L. K

Vision
To produce highly competent and innovative **Computer Professionals** to meet the global demands
Mission
- To impart quality education by **creative teaching learning** process.
- To be technically **competent, ethical and socially responsible** throughout the professional career.
- To inculcate **leadership qualities and entrepreneurship culture** to meet the global standards.

NEHRU INSTITUTE OF ENGINEERING AND TECHNOLOGY

"Nehru Gardens" T. M. PALAYAM, COIMBATORE-105
(Approved by AICTE and Affiliated to Anna University, Chennai)
(Accredited by NAAC, Recognized by UGC with section 2(f) and 12(B))
NBA Accredited UG Courses: Aero, CSE, Mech

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

Course Code & Title : **CS8075 & DATA WAREHOUSING AND DATA MINING**
Course Instructor : Mr. G.Jeevanantham
Class : III Year BE - CSE Semester : VII
Regulation : R2017 Academic Year : 2021-2022 Even Sem

Dear Student,
Greetings....!!!
I would like to thank all of you for sparing your time in filling up this Course Pre-Analysis survey for the effective conduct of the **DATA WAREHOUSING AND DATA MINING** Course. As you know that this survey is meant for knowing the knowledge level of the students with respect to this course, please fill it very carefully.
At this juncture, I am glad to welcome the suggestions from you all (if any).Rate your prior knowledge against the topic mentioned (Please Select the option)

Course Outcomes	Description	Rate your prior knowledge against the topic mentioned				
		Excellent (4)	Good (3)	Moderate (2)	Fair (1)	No idea (0)
C315.1	Your Knowledge about Data warehouse system and perform business analysis with OLAP tools				✓	
C315.2	Your knowledge about suitable pre-processing and visualization techniques for data analysis				✓	
C315.3	Your knowledge frequent pattern and association rule mining techniques for data analysis				✓	
C315.4	Your knowledge on appropriate classification and clustering techniques for data analysis.				✓	
C315.5	Apply Learning Tools and Learning Algorithms				✓	
Suggestions (If any)						

Date: 21/3/2022 Name of the student Penumarthi Hema Sri Signature of the student P. Hema Sri

Vision
To produce highly competent and innovative **Computer Professionals** to meet the global demands
Mission
- To impart quality education by **creative teaching learning** process.
- To be technically **competent, ethical and socially responsible** throughout the professional career.
- To inculcate **leadership qualities and entrepreneurship culture** to meet the global standards.

NEHRU INSTITUTE OF ENGINEERING AND TECHNOLOGY

"Nehru Gardens" T. M. PALAYAM, COIMBATORE-105
(Approved by AICTE and Affiliated to Anna University, Chennai)
(Accredited by NAAC, Recognized by UGC with section 2(f) and 12(B))
NBA Accredited UG Courses: Aero, CSE, Mech

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

COURSE PLAN

Course code and title	**CS8075 & DATA WAREHOUSING AND DATA MINING**		**L T P C** **3 0 0 3**
Class	IV Year B.E. Computer Science and Engineering	**Semester**	VII
Regulation	R-2017	**Academic year**	2021-2022 (Even)
Course prerequisites	1) CS8391 - Data Structures 2) CS8492 –Database Management Systems 3) CS8591 – Computer Networks		
Course objectives	Students will be made to **OBJ1:** To understand data warehouse concepts, architecture, business analysis and tools. **OBJ2:** To understand data pre-processing and data visualization techniques. **OBJ3:** To study algorithms for finding hidden and interesting patterns in data. **OBJ4:** To understand and apply various classification and clustering techniques using tools.		

COURSE OUTCOMES

C315.1	Design a Data warehouse system and perform business analysis with OLAP tools.
C315.2	Apply suitable pre-processing and visualization techniques for data analysis.
C315.3	Apply frequent pattern and association rule mining techniques for data analysis.
C315.4	Apply appropriate classification and clustering techniques for data analysis.

MAPPING OF PROGRAM OUTCOMES

PO1: Engineering knowledge: Apply the knowledge of mathematics, science, engineering fundamentals, and an engineering specialization for the solution of complex engineering problems.

PO2: Problem analysis: Identify, formulate, research literature, and analyse complex engineering problems reaching substantiated conclusions using first principles of mathematics, natural sciences, and engineering sciences.

PO3: Design/development of solutions: Design solutions for complex engineering problems and design system components or processes that meet t h e specified needs with appropriate consideration for public health and safety, and cultural, societal, and environmental considerations

PO4: Conduct investigations of complex problems: Use research-based knowledge and research methods including design of experiments, analysis and interpretation of data, and synthesis of the information to provide valid conclusions.

PO5: Modern tool usage: Create, select, and apply appropriate techniques, resources, and modern engineering and IT tools including prediction and modeling to complex engineering activities with an understanding of the limitations.

PO6: The engineer and society: Apply reasoning informed by the contextual knowledge to assess societal, health, safety, legal and cultural issues and the consequent responsibilities relevant to the professional engineering practice.

PO7: Environment and sustainability: Understand the impact of the professional engineering solutions in societal and environmental contexts, and demonstrate the knowledge of, and need for sustainable development.

PO8: Ethics: Apply ethical principles and commit to professional ethics and responsibilities and norms of the engineering practice.

PO9: Individual and team work: Function effectively as an individual, and as a member or leader in diverse teams, and in multidisciplinary settings.

PO10: Communication: Communicate effectively on complex engineering activities with the engineering the community and with the society at large, such as, being able to comprehend and write effective reports and design documentation, make effective presentations, and give and receive clear instructions.

PO11: Project management and finance: Demonstrate knowledge and understanding of the engineering and management principles and apply these to one's own work, as a member and leader in a team, to manage projects and in multidisciplinary environments.

PO12: Life-long learning: Recognise the need for, and have the preparation and ability to engage in independent and life-long learning in the broadest context of technological change.

PROGRAM SPECIFIC OUTCOMES

The students of Computer Science and Engineering Programme will be able to

PSO1	**Professional Skills:** Acquaint in-depth knowledge on the basic and advanced computer science domains like Data Sciences, Cryptography, Cloud and Distributed Computing, Neural Networks and Artificial Intelligence.
PSO2	**Entrepreneurship and Successful Career:** Apply the standard practices to have successful career path in the field of information and communication technology and entrepreneurship.

PROGRAM EDUCATIONAL OUTCOMES

The graduates of Computer Science and Engineering Programme will be able to

PEO1	Acquire and Apply knowledge in Computer Science, Mathematics, Science and inter-disciplinary engineering principles in order to excel in computer professional career.
PEO2	Analyze real life problems adapting to new Computing Technologies for professional excellence and ethical attitude, in order to provide economically feasible engineering solutions.
PEO3	Carry out complex engineering problems with best practices exhibiting communication skills, team work and interpersonal skills to enable continued computer professional development through life-long learning.

TEXT BOOK:

1. Jiawei Han and Micheline Kamber, —Data Mining Concepts and Techniques‖, Third Edition, Elsevier, 2012.

REFERENCES:

1. Alex Berson and Stephen J.Smith, —Data Warehousing, Data Mining & OLAP‖, Tata McGraw – Hill Edition, 35th Reprint 2016.
2. K.P. Soman, Shyam Diwakar and V. Ajay, —Insight into Data Mining Theory and Practice‖, Eastern Economy Edition, Prentice Hall of India, 2006.
3. Ian H.Witten and Eibe Frank, —Data Mining: Practical Machine Learning Tools and Techniques‖, Elsevier, Second Edition.

e-learning resources	
Mode of Evaluation	Internal Examination (20%) End Semester Examination (80%)
Faculty	Mr. Jeevanantham G, Assistant Professor, CSE
e-mail id	nietjeevanantham.g@nehrucolleges.com

NEHRU INSTITUTE OF ENGINEERING AND TECHNOLOGY

"Nehru Gardens" T. M. PALAYAM, COIMBATORE-105
(Approved by AICTE and Affiliated to Anna University, Chennai)
(Accredited by NAAC, Recognized by UGC with section 2(f) and 12(B))
NBA Accredited UG Courses: Aero, CSE, Mech

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

COURSE PLAN CS8075 & DATA WAREHOUSING AND DATA MINING

No of lecture hours	Topics to be covered	Reference Book No.	Scheduled on date	Delivered on date
UNIT I DATA WAREHOUSING, BUSINESS ANALYSIS AND ON-LINE ANALYTICAL PROCESSING				
1	Basic Concepts	T1, R1	09.03.2022	11/3/22
2	Data Warehousing Components	T1, R1	09.03.2022	11/3/22
3	Building a Data Warehouse	T1	10.03.2022	11/3/22
4,5	Database Architectures for Parallel Processing	T1	11.03.2022	23/3/22
6,7	Parallel DBMS Vendors, Multidimensional Data Model	T1	16.03.2022 16.03.2022	23/3/22
8	Data Warehouse Schemas for Decision Support, Concept Hierarchies	T1	17.03.2022	23/3/22
9	Characteristics of OLAP Systems	T1	18.03.2022	24/3/22
10	Typical OLAP Operations, OLAP and OLTP	T1	18.03.2022	24/3/22
UNIT II DATA MINING – INTRODUCTION				
11	Introduction to Data Mining Systems	T1, R2	23.03.2022	25/3/22
12	Knowledge Discovery Process	T1, R2	23.03.2022	28/3/22
13	Data Mining Techniques, Issues – applications	T1	24.03.2022	30/3/22
14	Data Objects and attribute types	T1, R2	25.03.2022	30/3/22
15	Statistical description of data	T1, R2	25.03.2022	31/3/22
16	Data Preprocessing	T1	30.03.2022	1/4/22
17	Cleaning, Integration, Reduction, Transformation and discretization	T1	30.03.2022	2/4/22
18	Data Visualization	T1	31.03.2022	20/4/22
19	Data similarity and dissimilarity measures	T1	01.04.2022	21/4/22
UNIT III DATA MINING - FREQUENT PATTERN ANALYSIS				
20	Mining Frequent Patterns	T1, R1	06.04.2022	21/4/22
21	Associations and Correlations	T1, R3	06.04.2022	22/4/22
22	Mining Methods	T1	07.04.2022	22/4/22
23	Pattern Evaluation Method	T1, R1	08.04.2022	23/4/22
24	Pattern Mining in Multilevel	T1	08.04.2022	23/4/22

Vision
To produce highly competent and innovative **Computer Professionals** to meet the global demands

Mission
- To impart quality education by **creative teaching learning** process.
- To be technically **competent, ethical and socially responsible** throughout the professional career.
- To inculcate **leadership qualities and entrepreneurship culture** to meet the global standards.

25	Multi Dimensional Space	T1	13.04.2022	27/4/22
26	Constraint Based Frequent Pattern Mining	T1	13.04.2022	27/4/22
27	Classification using Frequent Patterns	T1, R1	14.04.2022	28/4/22
	UNIT IV CLASSIFICATION AND CLUSTERING			
29	Decision Tree Induction	T1	15.04.2022	4/5/22
30	Bayesian Classification	T1	15.04.2022	4/5/22
31	Rule Based Classification	T1	20.04.2022	5/5/22
32	Classification by Back Propagation – Support Vector Machines	T1	20.04.2022	6/5/22
33	Lazy Learners – Model Evaluation and	T1	21.04.2022	9/5/22
34	Techniques to improve Classification Accuracy	T1	22.04.2022	11/5/22
35	Clustering Techniques – Cluster analysis- Partitioning Methods	T1	22.04.2022	11/5/22
36	Hierarchical Methods – Density Based Methods - Grid Based Methods	T1	27.04.2022	13/5/22
37	Evaluation of clustering – Clustering high dimensional data- Clustering with constraints,	T1	27.04.2022	13/5/22
38	Outlier analysis-outlier detection methods	T1	28.04.2022	20/5/22
	UNIT V WEKA TOOL			
39	Datasets – Introduction, Iris plants database	T1	29.04.2022	20/5/22
40	Breast cancer database	T1	04.05.2022	25/5/22
41	Auto imports database	T1	04.05.2022	25/5/22
42	Introduction to WEKA	T1	11.05.2022	26/5/22
43	The Explorer – Getting started	T1	11.05.2022	27/5/22
44	Exploring the explorer	T1	18.05.2022	1/6/22
45	Learning algorithms	T1	18.05.2022	1/6/22
46	Clustering algorithms	T1	19.05.2022	2/6/22
47	Association–rule learners	T1	20.05.2022	3/6/22

NEHRU INSTITUTE OF ENGINEERING AND TECHNOLOGY

"Nehru Gardens" T. M. PALAYAM, COIMBATORE-105
(Approved by AICTE and Affiliated to Anna University, Chennai)
(Accredited by NAAC, Recognized by UGC with section 2(f) and 12(B))
NBA Accredited UG Courses: Aero, CSE, Mech

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

Course outcome –Program outcome Mapping Table

CW075 & DATA WAREHOUSING AND DATA MINING	Cognitive level	Program outcomes Low Correlation-1, Moderate Correlation – 2; High Correlation –3													
		PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12	PSO1	PSO2
DIRECT METHOD															
Design a Data warehouse system and perform business analysis with OLAP	Remember, understand	2	1	3	1	1	1	2	1	2	2	1	1	2	1
Apply suitable pre-processing and visualization techniques for data analysis	Remember, understand	2	2	2	3	1	2	2	1	2	2	3	1	2	1
Apply frequent pattern and association rule mining techniques for data analysis	Apply, Design, Manage	3	3	3	3	3	3	2	1	2	2	3	2	3	3
Apply appropriate classification and clustering techniques for data analysis	Apply, Design, Manage	3	3	3	3	3	3	2	1	2	2	3	2	3	3

Date of creation:

Course Instructor

DQAC Member

HoD/CSE 29/12/2022

Vision
To produce highly competent and innovative **Computer Professionals** to meet the global demands

Mission
- To impart quality education by **creative teaching learning** process.
- To be technically **competent, ethical and socially responsible** throughout the professional career.
- To inculcate **leadership qualities and entrepreneurship culture** to meet the global standards.

NEHRU INSTITUTE OF ENGINEERING AND TECHNOLOGY

T. M. Palayam, Coimbatore-641 105
(Approved by AICTE, New Delhi and Affiliated to Anna University, Chennai)
Accredited by NAAC, Recognized by UGC under Section 2(f) and 12(B)
NBA Accredited UG Courses: AERO, CSE & MECH

Class Time Table – Academic Year: 2021- 22 (EVEN Semester)

Department: CSE
Degree and Branch: BE., CSE
Class Advisor(s): Ms. S. Priya

Class Committee Chairman: Dr. N. K. Sakthivel

Class / Sem/Section: III/ VI
Class Room : 113
W.E.F: 07.03.2022

Period DAY	1 9:10 AM – 09:55 AM	2 09:55 AM- 10:40 AM	Break (10:40- 10:55)	3 10:55 AM- 11:40 AM	4 11:40 AM- 12:25 PM	5 12:25 PM- 01:10 PM	Lunch Break (01:10- 01:55)	6 01:55 PM – 02:45 PM	7 02:45 PM – 03:35 PM	8 03:35 PM – 04:25 PM
Monday	CS8603 DS	CS8601 MC		CS8651 IP	CS8691 AI	SEMINAR		CS8691 AI	CS8602 CD	CS8651 IP
Tuesday	CS8651 IP	CS8602 CD		CS8603 DS	CS8601 MC	APT		(Lab 1) CS8661 IP LAB/ (Lab 10) CS8662 MAD LAB		
Wednesday	CS8075 DW&DM	CS8691 AI		HS8581 PC LAB		CS8603 DS		CS8602 CD	CS8075 DW&DM	CS8601 MC
Thursday	CS8602 CD	CS8651 IP		CS8603 DS	CS8075 DW&DM	LIB		(Lab 10) CS8662 MAD LAB/ (Lab 1)CS8661 IP LAB		
Friday	CS8691 AI	CS8601 MC		CS8075 DW&DM	CS8602 CD	SOFTSKILL		(Lab 10) CS8611 MINI PROJECT		

Course Code	Course Title	L	T	P	C	Name of the Faculty	Department
CS8651	Internet Programming	3	0	0	3	R.Senthil Kumar	CSE
CS8691	Artificial Intelligence	3	0	0	3	Dr.D.Sathishkumar	CSE
CS8601	Mobile Computing	3	0	0	3	Dr.S.Subasree	CSE
CS8602	Compiler Design	3	0	2	4	Ms. S. Priya	CSE
CS8603	Distributed Systems	3	0	0	3	Mr.M.Madan Mohan	CSE
CS8075	Data Warehousing and Data Mining	3	0	0	3	Dr. N. K. Sakthivel	CSE
CS8661	Internet Programming Laboratory	0	0	4	2	Mr.G.Jeevanantham / Mr. S. Venkatesh/ Dr. S. Siva Kumar	CSE
CS8662	Mobile Application Development Laboratory	0	0	4	2	Mr. A.Wasim Raja/ Mrs. S. Priya	CSE
CS8611	Mini Project	0	0	2	1	Mr. S. Venkatesh	CSE
HS8581	Professional Communication	0	0	2	1	Mrs. Susmi Mariyam Vargees	S & H/ English
LIB/ MENTOR	Library/ Mentor					Ms. S. Priya	CSE
SEMINAR	SEMINAR					Dr.S.Subasree	CSE
SOFT SKILL	SOFT SKILL					Mr.G.Jeevanantham	CSE
APT	APTITUDE					Dr. S. Sivakumar	CSE

Time Table Coordinator

HoD 4/3/2022

Dean- Computing

IQAC Coordinator

Principal
Dr. P. MANIARASU 4/3/2022
Principal
Nehru Institute of Engg. & Technology
T.M. Palayam, Coimbatore-641 105

DEPARTMENT OF
COMPUTER SCIENCE AND ENGINEERING
QUESTION BANK

VI SEMESTER

CS8075-DATAWARE HOUSING AND DATA MINING

Regulation – 2017

Academic Year 2021– 22

Prepared by

Mr.G.Jeevanantham, Assistant Professor/CSE

29/8/2022

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

QUESTION BANK

SUBJECT : CS8075-DATAWAREHOUSING AND DATAMINING

SEM / YEAR : VI - Third year

UNIT -I- DATA WAREHOUSING

Data warehousing Components –Building a Data warehouse — Mapping the Data Warehouse to Multiprocessor Architecture – DBMS Schemas for Decision Support – Data Extraction, Cleanup, and Transformation Tools –Metadata.

PART A

Q.NO	QUESTIONS	COMPETENCE	LEVEL
1.	How is data ware house different from a database? **Identify** the similarity.	Remember	BTL-1
2.	**Differentiate** metadata and data mart.	Understand	BTL-2
3.	**Analyze** why one of the biggest challenges when designing a data ware house is the data placement and distribution strategy.	Analyze	BTL-4
4.	How would you **evaluate**the goals of data mining?	Evaluate	BTL-5
5.	**List** the two ways the parallel execution of the tasks within SQL statements can be done.	Remember	BTL-1
6.	What elements would you use to **relate**the design of data warehouse?	Apply	BTL-3
7.	**Define**Data mart	Remember	BTL-1
8.	**Define**starschema	Remember	BTL-1
9.	What is Data warehousing? **Explain**the benefits of Data warehousing.	Evaluate	BTL-5
10.	Why data transformation is essential in the process of Knowledge discovery?**Describe** it.	Remember	BTL-1
11.	**Describe** the alternate technologies used to improve the performance in data warehouse environment	Understand	BTL-2
12.	**Distinguish**STARjoin and STARindex.	Understand	BTL-2
13.	Analyse the types of data mart.	Apply	BTL-3
14.	**Formulate**what is data discretization.	Create	BTL-6
15.	Point out the major differences between the star schema and the snowflake schema	Analyze	BTL-4
16.	**Pointout** the features of Metadata repository in data warehousing	Analyze	BTL-4

17.	**Define** Metadata repository	Remember	BTL-1
18.	**Discuss**metadata with an example.	Understand	BTL-2
19.	**Illustrate** the benefits of metadata repository.	Apply	BTL-3
20.	**Design** the data warehouse architecture.	Create	BTL-6

PART B

1	What is data warehouse?**Give** the Steps for design and construction of Data Warehouses and explain with three tier architecture diagram.(13)	Understand	BTL-2
2	Diagrammatically **illustrate** and discuss the following preprocessing techniques: (i) Data cleaning (3) (ii) Data Integration (3) (iii) Data transformation (3) (iv) Data reduction (4)	Apply	BTL-3
3	(i).Draw the data warehouse architecture and **explain** its components.(7) (ii).**Explain** the different types of OLAP tools. (6)	Analyze	BTL-4
4	**(i).Describe** in detail about Mapping the Data warehouse to a multiprocessor architecture (8) (ii).**Describe** in detail on data warehouse Metadata. (5)	Remember	BTL-1
5	(i).**Explain** the steps in building a data warehouse. (8) (ii).**Analyze** the information needed to support DBMS schemas for Decision support. (5)	Analyze	BTL-4
6	(i).**Discuss**in detail about access tools types?(6) ii) **Describe** the overall architecture of data warehouse? (7)	understand	BTL-2
7	(i) Discuss the different types of data repositories on which mining can be performed. (7) (ii) Differentiate tangible and intangible benefits of data warehouse. (6)	Understand	BTL-2
8	i) **Describe** in detail about data extraction (6) ii) **Describe** in detail about transformation tools (7)	Remember	BTL-1
9	(i).Suppose that a data warehouse consists of four dimensions customer, product, salesperson and sales time, and the three measure sales Amt(in rupees), VAT(in rupees) and payment type(in rupees). Draw the different classes of schemas that are popularly used for modeling data warehouses and **explain** it. (7) (ii).How would you **explain** Metadata implementation with examples? (6)	Evaluate	BTL-5
10	**Describe** in detail abouti) Bitmapped indexing (6) ii)STARjoin and index. (7)	Remember	BTL-1
11	(i).What is data Preprocessing? **Explain** the various data pre-processing techniques. (7) (ii).**Explain** the basic methods for data cleaning. (6)	Analyze	BTL-4
12	**Describe** with diagrammatic representation the relationship between operational data, a data warehouse and data marts. (13)	Remember	BTL-1
13	i)**Demonstrate** in detail about Data marts . (6) ii)**Demonstrate** data warehouse administration and management. (7)	Apply	BTL-3

14	(i)**Generalize**the potential performance problems with star schema.(6) (ii).**Design** and discuss about the star and snowflake schema models of a Data warehouse. (7)	Create	

PART C

1	**Explain** mapping datawarehouse with multiprocessor architecture with the concept of parallelism and data partitioning (15)	Analyze	B
2	**Design** a star-schema , snow-flake schema and fact- constellation schema for the datawarehouse that consists of the following four dimensions(Time , Item, Branch And Location). Include the appropriate measures required for the schema.	Create	BT
3	i)**Generalize** why we need data preprocessing step in datawarehousing. (8) ii)**Explain** the various methods of data cleaning and data reduction technique(7)	Create	BT
4	i)**Compare** the similarities and differences between the database and datawarehouse. (8) ii)**Explain** what is data visualization. How it helps in data warehousing.(7)	Evaluate	BT

UNIT 2- BUSINESS ANALYSIS

Reporting and Query tools and Applications – Tool Categories – The Need for Applications – Cog Impromptu – Online Analytical Processing (OLAP) – Need – Multidimensional Data Model – OL Guidelines – Multidimensional versus Multirelational OLAP – Categories of Tools – OLAP Tools a the Internet.

PART A

1	Can you **list** the categories of tools in business analysis	Remember	BTL-1
2	How would you **describe** the basic functions used in reporting tools?	Remember	BTL-1
3	Can you **explain** the tool for corporate data.	Analyze	BTL-4
4	**Define** the various access types to the data stored in a data warehouse	Remember	BTL-1
5	What would **demonstrate** the result if an organization uses a familiar approach to build a query and reporting environment for the data warehouse?	Apply	BTL-3
6	What is a reporting tool? **Give** examples for managed query tools.	Understand	BTL-2
7	**Explain**the need of Cognos Impromptu.	Analyze	BTL-4
8	**Compare**OLTP and OLAP system.	Analyze	BTL-4
9	**Illustrate** about reporting tools.	Apply	BTL-3
10	**Discuss** the need for OLAP.	Understand	BTL-2
11	How would you **classify** the ideas of multidimensional data model with multirelational OLAP?	Apply	BTL-3
12	Which parameter and **design** choices determine the MOLAP?	Create	BTL-6

3	**Summarize** the distinct features of OLTP with OLAP.	Understand	BTL-2
4	List any four tools for performing OLAP.	Remember	BTL-1
5	**Explain**about query tools with example.	Evaluate	BTL-5
6	**Generalize** the function of OLAP tools in the internet.	Create	BTL-6
7	**Describe** MOLAP and ROLAP.	Understand	BTL-2
8	How would you **explain** the key features of OLAP components?	Evaluate	BTL-5
9	**Describe**about the Internet tools.	Remember	BTL-1
20	**List** out the key features in business applications using OLAP.	Remember	BTL-1

PART B

1	What is business analysis? **List** and discuss the basic features that are provided by reporting and query tools used for business analysis. (13)	Remember	BTL-1
2	**Explain the** features ofCognos Impromptu business analysis tool. (13)	Analyze	BTL-4
3	(i).**Compare** OLTP and OLAP systems. (7) (ii).**Summarize** the various OLAP operations in the Multidimensional Data Model. (6)	Evaluate	BTL-5
4	i)How would you **describe** in detail about reporting query classification ? (7) ii) **Describe** in detail about application of OLAP tools. (6)	Understand	BTL-2
5	**Describe** in detail about tool categories in business analysis. (13)	Remember	BTL-1
6	**Examine**with an example the reporting and query tools in business analysis.(13).	Apply	BTL-3
7	**(i). Describe** the need for OLAP. (6) **(ii).Discuss** in detail about the OLAP guidelines. (7)	Understand	BTL-2
8	(i). Diagrammatically illustrate and **describe** the architecture of MOLAP and ROLAP. (7) (ii). **Identify** the Major difference between MOLAP and ROLAP. (6)	Remember	BTL-1
9	**(i).** Write about the multidimensional data model. **Describe** how it is used in data warehousing. (7) **(ii). Describe** in detail about tool categories in datawarehouse. (6)	Remember	BTL-1
10	**i). Discuss** in detail about the OLAP tools. (6) ii) **Discuss** in detail about application in the internet (7)	Understand	BTL-2
11.	**Explain** in detail about the characteristics of OLAP cube. (13)	Analyze	BTL-4
12	**Examine** the relevant examples discuss multidimensional online analytical processing and multi relational online analytical processing. (13)	Apply	BTL-3
13	**Generalize** the topic on (i). Operations in Multidimensional Data Model. (7)	Create	BTL-6

	(ii). OLAP Tools and the Internet. (6)		
14	**Explain** in detail about the features of OLTP and OLAP. (13)	Analyze	BT
	PART C		
1	**Pointout** and **Illustrate** the various multidimensional data modeling to design a data warehouse . (15)	Analyze	BT
2	**Summarize** different tool categories in data warehouse business analysis. (15)	Create	BT
3	(i) Design a multidimensional cube with your own example. (5) (ii) Suppose that a data warehouse consists of the four dimensions date, spectator, location, and game, and the two measures count and charge, where charge is the fare that a spectator pays when watching a game on a given date spectators may be students, adults, or seniors, with each category having its own charge rate. (1) Draw a star schema diagram for the data warehouse. (4) (2) Starting with a base cuboid [date, spectator, location, game], what specific OLAP operations should one perform in order to list the total charge paid by student spectators at GM Place in 2000? (4).	Create	BT
4	i)Depict the 3 tier data warehousing architecture and **explain**its features in detail. (8) ii).**Explain** the different types of OLAP servers (7)	Evaluate	BTL-5
	UNIT 3- DATA MINING		
	Introduction – Data – Types of Data – Data Mining Functionalities – Interestingness of Patterns Classification of Data Mining Systems – Data Mining Task Primitives – Integration of a Data Mining System with a Data Warehouse – Issues –Data Preprocessing.		
	PART A		
1	**Define** Data mining. **List** out the steps in data mining?	Remember	BTL-1
2	**List** the steps involved in the process of KDD. How does it relate to data mining?.	Remember	BTL-1
3	**List** the ways in which interesting patterns should be mined.	Remember	BTL-1
4	**Compare** drill down with roll up approach.	Analyze	BTL-4
5	**Describe** what are the other kinds of data in data mining.	Understand	BTL-2
6	How would you **illustrate**Handling outlier or incomplete data?	Apply	BTL-3
7	**Analyze** data characterization related to data discrimination.	Analyze	BTL-4
8	**Define** association and correlations.	Remember	BTL-1
9	**List** the five primitives for specification a data mining task.	Remember	BTL-1
10	**Evaluate** the major tasks of data preprocessing.	Evaluate	BTL-5

11	Are all patterns generated are interesting and useful? **Give** reasons to justify.	Understand	BTL-2
12	**Classify** different types of reductions.	Apply	BTL-3
13	**Distinguish** between data cleaning and noisy data.	Understand	BTL-2
14	**Explain** the principle elements of missing values in data cleaning.	Analyze	BTL-4
15	**Discuss** the roles of noisy data in data preprocessing.	Analyze	BTL-2
16	**Consider** that the minimum and maximum values for the attribute "salary" are 12,000 and 98,000 respectively and the mapping range of salary is [0.0 , 1.0]. Find the transformation for the salary 73,600 using min-max normalization.	Create	BTL-6
17	**Show** how the attribute selection set is important in data reduction.	Apply	BTL-3
18	Consider the following set of data X = {15,27,62,35,39,50,44,44,22,98} Do preprocessing using smoothing by bin means and bin boundary to smooth the data, using a bin of depth 3.**Evaluate** it.	Evaluate	BTL-5
19	**Formulate** why do we need data transformation. Mention the ways by which data can be transformed.	Create	BTL-6
20	**Define** an efficient procedure for cleaning the noisy data.	Remember	BTL-1

PART B

1	i) **Demonstrate** in detail about data mining steps in the process of knowledge discovery? (8) ii) List the application area of data mining? (5)	Apply	BTL-3
2	**Explain** in detail about the Evolution of Database Technology. (13)	Evaluate	BTL-5
3	(i).What is data? How different type of data and attributes can be **designed**? (6) (ii). **Design** and discuss in detail about Primitives for specifying a data mining task (7)	Create	BTL-6
4	(i).**Discuss** whether or not each of the following activities is a data mining task.(5) 1. Credit card fraud detection using transaction records. 2. Dividing the customers of a company according to their gender. 3. Computing the total sales of a company 4. Predicting the future stock price of a company using historical records. 5. Monitoring seismic waves for earthquake activities. (ii). **Discuss** on descriptive and predictive data mining tasks with illustrations.(8)	Understand	BTL-2
5	**(i).State and Explain** the various classification of data mining systems with example. (7) **(ii). Explain** the various data mining functionalities in detail. (6)	Analyze	BTL-4

6	Suppose that the data for analysis include the attributed age.The age values for the data tuples are 13,15,16,19,20,20,21,22,22,25,25,25,25,30,33,33,35,35,35,35, 36,40,45,46,52,70. (i).use smoothing by bin means to smooth the above data using a bin depth of 3. **Illustrate** your steps. (6) (ii). **Classify** the various methods for data smoothing. (7)	Apply	BTL
7	Sketch the various phases of data mining and explain the **different** steps involved in preprocessing with their significance before mining, Give an example for each process. (13)	Understand	BTL
8	**Describe** in detail about the issues of data mining. (13)	Remember	BTL
9	**Describe** in detail about datareduction in data preprocessing (13)	Remember	BTL
10	**Describe** in detail about various data transformation techniques (13)	Remember	BTL
11	**List** and explain the primitives for specifying a data mining task.(13)	Remember	BTL
12	(i). How will you handle missing value in a dataset before mining process?. (4) (ii).**Give** the architecture of a typical data mining system. (9)	Understand	BTL
13	i)**Explain** how integration is done with a database or data warehouse system. ii)**Consider** the following data for the attribute AGE:4,8,21,5,21,24,34,28,25. Perform smoothing by bin means and bin boundaries using a bin depth of 3	Analyze	BTL
14.	**Analyze**Using Equi-depth binning method, partition the data given below into 4 bins and perform smoothing according to the following methods.(8) 1. Smoothing by bin means 2. Smoothing by bin median 3. Smoothing by bin boundaries 24,25,26,27,28,56,67,70,70,75,78,89,89,90,91,94,95,96,100,102,103,107,109,112. (ii).What motivated data mining? Why is it important? (5)	Analyze	BTL

PART C

1	**Describe** the Major issues in datawarehousing and data mining. (15)	Analyze	BTL
2	i)What is interestingness of a pattern ? (5) ii)**Summarize** the integration of data mining system with a data warehouse.(10)	Create	BTL
3	List the major data preprocessing techniques and **explain** in detail with examples.(15)	Analyze	BTL
4	i)**Generalize** in detail how data mining system are classified (5) ii)Discuss each classification with an example. (10)	Create	BTL

UNIT-4- ASSOCIATION RULE MINING AND CLASSIFICATION

Mining Frequent Patterns, Associations and Correlations – Mining Methods – Mining various Kinds ofAssociation Rules – Correlation Analysis – Constraint Based Association Mining – Classification andPrediction - Basic Concepts - Decision Tree Induction : Bayesian Classification – Rule BasedClassification – Classification by Back propagation – Support Vector Machines

	AssociativeClassification – Lazy Learners – Other Classification Methods – Prediction.		
	PART A		
1	**Define**correlation and market basket analysis.	Remember	BTL-1
2	**Formulate** the principle frequent itemset and closed itemset.	Create	BTL-6
3	How would you **explain** the principle of Apriori algorithm? How can the efficiency of an Apriori algorithm be improved?	Evaluate	BTL-5
4	**Define** Data pruning. State the need for pruning phase in decision tree construction.	Remember	BTL-1
5	**Compare**the advantages of FP growth algorithm over apriori algorithm.	Analyze	BTL-4
6	**Explain** how will you generate association rules from frequent itemsets	Analyze	BTL-4
7	What is naïve Bayesian classification? How is it **differing** from Bayesian classification?	Analyze	BTL-4
8	**Discuss**association rule mining .	Understand	BTL-2
9	**Describe** the uses correlation	Understand	BTL-2
10	**Discuss** the features of Decision tree induction.	Understand	BTL-2
11	How would you **evaluate** accuracy of a classifier?	Evaluate	BTL-5
12	**List** the two interesting measures of an association rule.	Remember	BTL-1
13	**Define** Back propagation.	Remember	BTL-1
14	**Illustrate**support vector machine with example.	Apply	BTL-3
15	How would you **show** your understanding about rule based classification?	Apply	BTL-3
16	**Discuss why** pruning is needed in decision tree.	Understand	BTL-2
17	What inference can you **formulate**with Bayes theorem?	Create	BTL-6
18	**Demonstrate**the Bayes classification methods.	Apply	BTL-3
19	**Define** Lazy learners with an example.	Remember	BTL-1
20	**What** are eager learners?	Remember	BTL-1
	PART B		
1	(i).**Compare** Classification and Prediction. (3) (ii).**Explain** the issues regarding classification and prediction(3). (iii).Write and **explain** the algorithm for mining frequent item sets without candidate generation. (7)	Analyze	BTL-4
2	i)How would you **summarize** in detail about mining methods?(6) ii)Summarize in detail about various kinds of association rules. (7)	Understand	BTL-2
3	**Describe** in detail about constraint and correlation based association mining. (13)	Remember	BTL-1
4	(i).Develop an algorithm for classification using decision trees. **Illustrate** the algorithm with a relevant example. (7) (ii).What approach would you use to **apply** decision tree induction? (6)	Apply	BTL-3

5	What is Classification? What are the features of Bayesian classification?**Explain** in detail with an example.(13)	Evaluate	BT
6	(i). **Giving** concrete example , explain a method that performs frequent itemset mining by using the prior knowledge of frequent item set properties. (7) (ii). **Discuss** in detail the constraint based association mining. (6).	Understand	BT
7	**(i).Examine** in detail about Lazy learners with examples. (4) (ii). **Describe**about the process of multi-layer feed-forward neural network classification using back propagation learning.(9)	Remember	BT
8	**(i).Describe** in detail about frequent pattern classification. (7) (ii). Write an algorithm for FP-Tree Construction and **discuss** how frequent itemsets are generated from FP-Tree.(6)	Understand	BT
9	Consider a home finance loan to predict the housing loan payment. Design a general hierarchical astructure and analyze the factors using rule discovery techniques to accurately predict the number of loan payments in a given quarter/year. Loan is availed for a period of 20 to 25 years, but an average life span of the loan exists for only 7 to 10 years due to payment. Make necessary assumptions: Maintenance record of the customer details and details of the prevailing interest rates, borrower characteristics, account dare, fine tune loan prepayment such as interest rates and fees in order to maximize the profits of the company. Elaborately discuss the association rule mining issues. Also **Examine** on the multi level association rules and find if you could relate any relation on from the above application. (13)	Apply	BT
10	**Generalize** the Bayes theorem of posterior probability and explain the working of a Bayesian classifier with an example.(13)	Create	BT
11.	**Explain** and Apply the Apriori algorithm for discovering frequent item sets of the table. (13)	Analyze	BT

Trans ID	Items Purchased
101	Milk,bread,eggs
102	Milk,juice
103	Juice,butter
104	Milk,bread,eggs
105	Coffee,eggs
106	Coffee
107	Coffee,Juice
108	Milk,bread,cookies,eggs
109	Cookies,butter
110	Milk,bread

Use 0.3 for the minimum support value.Illustrate each step of the Apriori Algorithm.

12.	(i).**Define** classification? With an example explain how support vector machines can be used for classification.(7) (ii). What are the prediction techniques supported by a data mining systems? (6)	Remember	BT
13.	(i). Write Bayes theorem.(4) (ii) Explain how the Bayesian Belief Networks are trained to perform classification.(9)	Analyze	BT
14.	.**Describe**in detail about the following Classification methods. (i). Bayesian classification(6) (ii)Classification by Back propagation.(7)	Remember	BT

PART C

1	Find all frequent item sets for the given training set using Apriori and FP growth respectively. **Compare** the efficiency of the two mining processes (15) TID ITEMS BROUGHT 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 }	Create	BTL-4
2	**Generalize** and Discuss about constraint based association rule mining with examples and state how association mining to correlation analysis is dealt with. (15)	Create	BTL-6
3	Discuss the single dimensional Boolean association rule mining for transaction database. **Evaluate** the below transaction database. (15) Transaction ID Items Bought 2000 A,B,C 1000 A,C 4000 A,D 5000 B,E,F Let minimum support 50% and minimum confidence 50% We have A=>C (50% , 66.6%) C=>A (50%, 100 %)	Evaluate	BTL-5
4	Construct the decision tree for the following training dataset using decision tree algorithm. (13)	Create	BTL-6

Age	Income	Student	Credit_rating	Buys_Computer
<=30	High	No	Fair	No
<=30	High	No	excellent	No
31..40	High	No	Fair	Yes
>40	Medium	No	Fair	Yes
>40	low	Yes	Fair	Yes
>40	low	Yes	excellent	No
31..40	low	Yes	excellent	Yes
<=30	Medium	No	Fair	No
<=30	low	Yes	Fair	Yes
>40	Medium	Yes	Fair	Yes
<=30	Medium	Yes	excellent	Yes
31..40	Medium	No	excellent	Yes
31..40	High	Yes	Fair	Yes
>40	Medium	No	excellent	No

UNIT 5- CLUSTERING AND TRENDS IN DATA MINING

Cluster Analysis - Types of Data – Categorization of Major Clustering Methods – K-means–Partitioning Methods – Hierarchical Methods - Density-Based Methods –Grid Based Methods – Model-Based Clustering Methods – Clustering High Dimensional Data - Constraint – Based ClusterAnalysis – Outlier Analysis – Data Mining Applications.

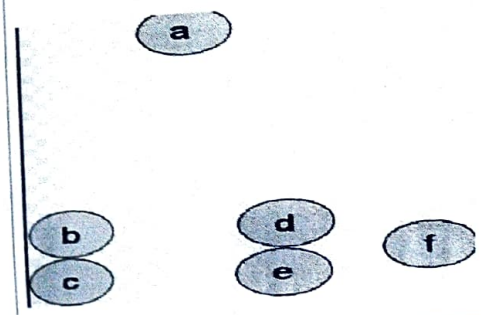
PART A

1	**Identify** what changes would you make to solve the problem in cluster analysis.	Remember	BTL-1

2	**Define** K-means partitioning.	Remember	
3	**List** the major clustering methods.	Remember	
4	**Explain** why a cluster has to be evaluated.	Analyze	
5	**Illustrate** the intrinsic methods in cluster analysis.	Apply	
6	How do you **explain** the similarityin clustering?	Evaluate	
7	**Define** what is meant by K nearest neighbor algorithm.	Remember	
8	**Illustrate** some applications of clustering.	Apply	
9	Whatthe services are provided by grid based clustering .	Apply	
10	**Formulate** challenges in clustering.	Create	
11	Organize the **design** goals of constraint based clustering methods.	Create	
12	**Classify** the hierarchical clustering methods.	Analyze	
13	**Distinguish** between density based clustering and grid based clustering.	Understand	
14	**Define** outlier. How will you determine outliers in the data?	Remember	
15	**Discuss** the challenges of outlier detection.	Understand	
16	**Distinguish** between Classification and clustering.	Understand	
17	**Evaluate** what information is used by outlier detection method.	Evaluate	
18	**Give** the methods of clustering high dimensional data.	Understand	
19	**List** out the difference between characterization and clustering.	Remember	
20	**Explain**the typical phases of outlier detection methods.	Analyze	
	PART B		
1	(i).**Analyze** the Requirements of clustering in Data Mining(8). ii)Analyse the desirable properties of Clustering algorithm.(5)	Analyze	
2	(i).**Describe**in detail about categorization of major clustering methods . (8) (ii).List out the General applications of Clustering. (5)	Remember	
3	**What is clustering? Describe** in detail about the features of K-means partitioning method. (13)	Remember	
4	**i) Explain** in detail about hierarchical based method. (7) **ii) Explain** in detail about density based methods. (6)	Analyze	
5	What is grid based clustering? With an example explain an algorithm for grid based clustering.(13)	Remember	
6	**(i) Demonstrate** in detail about model based clustering methods.(7) (ii).Illustrate the topic on (6) 1. CLIQUE 2. DBSCAN	Apply	
7	**(i).Demonstrate** on clustering high dimensional data. (6) (ii).Consider five points { X_1, X_2,X_3, X_4, X_5} with the following coordinates as a two dimensional sample for clustering:	Apply	

	X_1 = (0,2,5); X_2 = (0,0); X_3= (1.5,0); X_4 =(5,0); X_5 = (5,2) **Illustrate** the K-means partitioning algorithm using the above data set.(7)		
8	i)How would you **discuss** the outlier analysis in detail ? (7) ii) **Discuss** in detail about the various detection techniques in outlier.(6)	Understand	BTL-2
9	(i).**Explain** in detail about data mining applications (5). (ii). With relevant examples **summarize** in detail about constraint based cluster analysis. (8)	Evaluate	BTL-5
10	**Design** statistical approaches in outlier detection with neat design and with examples. (13)	Create	BTL-6
11	**Discuss** the various clustering method in Datamining. With an example (13)	Understand	BTL-2
12.	(i). **Disucss** in detail about the different types of data in cluster analysis.(5) (ii). **Discuss** the following clustering algorithm using examples.(8) 1. K.means 2. K-medoid.	Understand	BTL-2
13.	Describe the applications and trends in data mining in detail(13)	Analyze	BTL-4
14	What is outlier mining important? Briefly **describe** the different approaches behind statistical –based outlier detection, distance based outlier detection and deviation based outlier detection. (13)	Remember	BTL-1

PART C

1	Explain hierarchical clustering in detail. **Analyse** the below diagram and draw the dendrogram using hierarchical clustering algorithm. (15) 	Analyze	BTL-4
2	Consider that the data mining task is to cluster the following eight points A1,A2,A3,B1,B2,B3,C1AND C2(with (X,Y) representing location) into three clusters A1(2,10) , A2(2,5) , A3(8,4) , B1(5,8) , B2(7,5) , B3(6,4) , C1(1,2) , C2(4,9). The distance function is Euclidean distance.Suppose initially we assign A1, B1 and C1 as the center of each cluster, respectively. Use the K-means algorithm to show the three cluster centres after the first round of execution and the final tree clusters. (15)	Create	BTL-6
3	Discuss the steps in K-means algorithm and **evaluate** the following table using K-means. (15)	Evaluate	BTL-5

Subject	A	B
1	1.0	1.0
2	1.5	2.0
3	3.0	4.0
4	5.0	7.0
5	3.5	5.0
6	4.5	5.0
7	3.5	4.5

4	**Analyze**and elaborate the current trends in data mining in any **three**fields. 3x5=15 1. Financial data analysis 2. Biological data analysis 3. Telecommunication industry 4. Intrusion detection 5. Retail industry	Create	B1

Reg. No. :

Question Paper Code : X10301

B.E./B.Tech. DEGREE EXAMINATIONS, APRIL/MAY 2021
Sixth Semester
Computer Science and Engineering
CS 8075 – DATA WAREHOUSING AND DATA MINING
(Common to Computer and Communication Engineering)
(Regulations 2017)

Time : Three Hours | Maximum : 100 Marks

Answer ALL questions

PART – A **(10×2=20 Marks)**

1. Define data warehouse metadata.
2. Name the schemas a multidimensional data model can adopt.
3. What is data integration ?
4. Name the steps in the process of knowledge discovery.
5. What is an association rule ? Give example.
6. Present an outline of frequent pattern growth method.
7. What is decision tree induction ?
8. Outline how tree pruning algorithms attempt to improve accuracy.
9. How the Euclidean distance between two points or tuples, say, $X_1 = (x_{11}, x_{12}, ...x_{1n})$ and $X_2 = (x_{21}, x_{22}, \ldots x_{2n})$ is calculated ?
10. What is a rule-based classifier ?

PART – B **(5×13=65 Marks)**

11. a) What is lattice of cuboids ? Construct a lattice of cuboids forming a data cube for the representation of sales data, according to the dimensions time, item, location and supplier. **(13)**

(OR)

b) Outline the typical OLAP operations for multidimensional data with an example and relevant diagrams. **(13)**

12. a) How is an attribute normalized ? When is normalization particularly useful ? Outline min-max normalization with an example. **(13)**

(OR)

b) What is data reduction ? Outline how sampling can be used as a data reduction technique with an example. **(13)**

13. a) What is correlation analysis ? Outline the steps in correlation analysis using lift with an example. **(13)**

(OR)

b) Outline the steps in classification based on multiple association rules with an example. **(13)**

14. a) State Bayes' theorem of posterior probability. Outline the steps in Bayesian classification with an example. **(13)**

(OR)

b) What is a dendrogram ? How to derive clusters from dendrogram ? Outline with an example. **(13)**

15. a) What is Weka ? Highlight the features of Weka. **(13)**

(OR)

b) Represent the data presented in the following table a IF-THEN rules, decision tree and neural network. **(13)**

Age	Income	Class
Youth	High	A
Youth	Low	B
Middle aged	High	C
Middle aged	Low	C
Senior	High	C
Senior	Low	C

PART – C (1×15=15 Marks)

16 a) Apply the Apriori algorithm for discovering frequent item sets to the following data set :

Trans ID	Items Purchased
101	Mulberry, Raspberry, Cherry
102	Mulberry, Papaya
103	Papaya, Mango
104	Mulberry, Raspberry, Cherry
105	Passion Fruit, Cherry
106	Passion Fruit
107	Passion Fruit, Papaya
108	Mulberry, Raspberry, Guava, Cherry
109	Guava, Mango
110	Mulberry, Raspberry

Use 0.3 for the minimum support value. **(15)**

(OR)

b) Consider five points $\{X_1, X_2, X_3, X_4, X_5\}$ with the following coordinates as a two-dimensional sample for clustering:

$X_1 = (0, 2)$, $X_2 = (1, 0)$, $X_3 = (2, 1)$, $X_4 = (4, 1)$ and $X_5 = (5, 3)$.

Illustrate the K-means algorithm on the above data set. The required number of clusters is two and initially, clusters are formed from random distribution of samples : $C_1 = \{X_1, X_2, X_4\}$ and $C_2 = (X_3, X_5)$. **(15)**

———————

UNIT I

DATA WAREHOUSING

Data warehousing Components -Building a Data warehouse - Mapping the Data Warehouse to a Multiprocessor Architecture - DBMS Schemas for Decision Support - Data Extraction, Cleanup, and Transformation Tools -Metadata.

Data Warehouse Introduction

A data warehouse is a collection of data marts representing historical data from different operations in the company. This data is stored in a structure optimized for querying and data analysis as a data warehouse. Table design, dimensions and organization should be consistent throughout a data warehouse so that reports or queries across the data warehouse are consistent.

A data warehouse can also be viewed as a database for historical data from different functions within a company. The term Data Warehouse was coined by Bill Inmon in 1990, which he defined in the following way: "A warehouse is a subject-oriented, integrated, time-variant and non-volatile collection of data in support of management's decision making process". He defined the terms in the sentence as follows:

- *Subject Oriented:* Data that gives information about a particular subject instead of about a company's ongoing operations.
- *Integrated:* Data that is gathered into the data warehouse from a variety of sources and merged into a coherent whole.
- *Time-variant:* All data in the data warehouse is identified with a particular time period.
- *Non-volatile:* Data is stable in a data warehouse. More data is added but data is never removed. This enables management to gain a consistent picture of the business. It is a single, complete and consistent store of data obtained from a variety of different sources made available to end users in what they can understand and use in a business context. It can be Used for decision Support, Used to manage and control business, Used by managers and end-users to understand the business and make judgments.