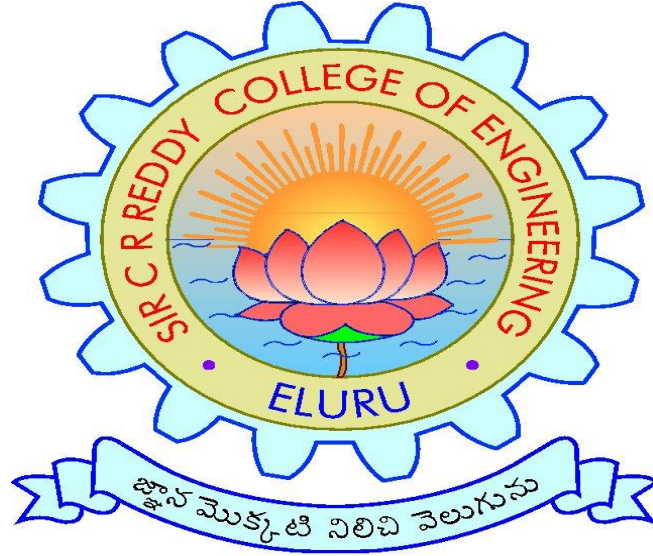


**SIR C.R.REDDY COLLEGE OF ENGINEERING, ELURU**  
**DEPARTMENT OF INFORMATION TECHNOLOGY**  
**COURSE HANDOUT**



**SUBJECT: DATA MINING TECHNIQUES**  
**CLASS: III/IV B.Tech., I SEMESTER, A.Y.2022-23**  
**INSTRUCTOR: Sri. P.Rajendra Kumar**

### Course Handout Index

S.No	Description
1	College Vision & Mission
2	Department Vision & Mission
3	Program Educational Objectives (PEOs)
4	Program Outcomes (POs)
5	Program Specific Outcomes (PSOs)
6	JNTUK Academic Calendar
7	Department Academic Calendar
8	Course Description
9	Course Objectives
10	Course Outcomes
11	Lesson Plan
12	Evaluation Pattern
13	Timetable
14	Unit wise Questions

### **COLLEGE VISION**

To emerge as a premier institution in the field of technical education and research in the state and as a home for holistic development of the students and contribute to the advancement of society and the region

### **COLLEGE MISSION**

To provide high quality technical education through a creative balance of academic and industry oriented learning; to create an inspiring environment of scholarship and research; to instill high levels of academic and professional discipline; and to establish standards that inculcate ethical and moral values that contribute to growth in career and development of society in general.

### **DEPARTMENT VISION**

To be a premier Department in the region in the field of Information Technology through academic excellence and research that enable graduates to meet the challenges of industry and society

### **DEPARTMENT MISSION**

- ❖ To Provide dynamic teaching-learning environment to make the students industry ready and advancement in career
- ❖ To inculcate professional and leadership quality for better employability and entrepreneurship
- ❖ To make high quality professional with moral and ethical values suitable for industry and society

### **PROGRAM EDUCATIONAL OBJECTIVES (PEOs)**

**PEO1:** Solve real world problems through effective professional skills in Information Technology industry and academic research.

**PEO2:** Analyze and develop applications in Information Technology domain and adapt to changing technology trends with continuous learning.

**PEO3:** Practice the profession in society with ethical and moral values.

## **PROGRAM OUTCOMES (POs)**

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

**PO2: Problem Analysis:** Identify, formulate, research literature, and analyze complex engineering problems reaching substantiated conclusions using the first principles of mathematics, natural sciences, and engineering sciences.

**PO3: Design/Development of Solutions:** Design solutions for complex engineering problems and system components or processes that meet the specified needs with appropriate consideration for the public health and safety, and the cultural, society, 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 society 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 multi-disciplinary settings.

**PO10: Communication:** Communicate effectively on complex engineering activities with the engineering community and with 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 multi-disciplinary environments.

**PO12: Life-long Learning:** Recognize 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 (PSOs)**

**PSO1: Design Skill:** Design and develop software in the area of relevance under realistic constraints.

**PSO2: New Technology:** Adapt new and fast emerging technologies in the field of Information Technology.

Website: [www.jntuk.edu.in](http://www.jntuk.edu.in)  
Email: [dap@jntuk.edu.in](mailto:dap@jntuk.edu.in)



Phone: 0884-2300991

**Directorate of Academic Planning**  
JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY KAKINADA  
KAKINADA-533003, Andhra Pradesh, INDIA  
(Established by AP Government Act No. 30 of 2008)

Lr. No. DAP/AC/III Year /B. Tech/B. Pharmacy/2022

Date 14.07.2022

**Dr. KVSG Murali Krishna,**  
*M.E. Ph.D.*  
Director, Academic Planning  
JNTUK, Kakinada

To  
All the Principals of Affiliated Colleges,  
JNTUK, Kakinada.

**Academic Calendar for III Year - B. Tech/B. Pharmacy for the AY 2022-23  
(2020-21 Admitted Batch)**


I SEMESTER			
Description	From	To	Weeks
Community Service Project	15.07.2022	30.07.2022	2W
I Unit of Instruction	01.08.2022	24.09.2022	8W
I Mid Examinations	26.09.2022	01.10.2022	1W
II Unit of Instructions	03.10.2022	26.11.2022	8W
II Mid Examinations	28.11.2022	03.12.2022	1W
Preparation & Practicals	05.12.2022	10.12.2022	1W
End Examinations	12.12.2022	25.12.2022	2W

\* As per the AFSCH Guidelines Out of the Total 180 hours of Community Service Project leading to 4 Credits, two weeks will be offline and remaining project work can be done during the III-I semester weekends and holidays.

  
Director,  
Academics & Planning, JNTUK  
14-7-22  
Director  
Academic Planning  
JNTUK Kakinada

Copy to the Secretary to the Hon'ble Vice Chancellor, JNTUK  
Copy to Rector, Registrar, JNTUK  
Copy to Director Academic Audit, JNTUK  
Copy to Director of Evaluation, JNTUK

## Department Academic Calendar

		Department of Information Technology III/IV B.Tech Academic Calendar for 2022-23																																		
		2022-23	S	M	T	W	T	F	S	S	M	T	W	T	F	S	S	M	T	W	T	F	S	S	M	T	W	T	F	S	S	M				
Jul 22						1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31
Aug 22		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31				
Sep 22					1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30		
Oct 22						1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31
Nov 22			1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30				
Dec 22					1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	
Jan 23	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31					
Feb 23				1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28					
Mar 23				1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31		
Apr 23					1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30		
May 23		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31				
Jun 23					1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30		
<b>List of Holidays</b>	Oct 9: <u>Maulud Nabi</u>					Mar 22: <u>Usadhi</u>					Mid exams																									
	July 10: <u>Bakrid</u>					Oct 24: <u>Diwali</u>					Mar 30: <u>Srirama Navami</u>					End Examinations																				
	Aug 9: <u>Moharum</u>					Dec 25: <u>Christmas</u>					Apr 5: <u>Babu Jagjivan Ram Jayanti</u>					Commencement of Class work																				
	Aug 15: <u>Independence day</u>					Jan 14-16: <u>sankranti</u>					Apr 7: <u>Good friday</u>					Workshops																				
	Aug 31: <u>Ganesh Chaturdi</u>					Jan 26: <u>Republic Day</u>					Apr 14: <u>Ambekar Jayanthi</u>					Department fest Elite																				
	Oct 2: <u>Gandhi jayanthi</u>					Feb 18: <u>Sivaratri</u>					Jun 29: <u>Bakrid</u>					<b>HoD</b>																				
	Oct 5: <u>Vijayadasami</u>					Mar 8: <u>holi</u>										Department of IT																				

## Course Description

Data Mining Techniques studies algorithms and computational paradigms that allow computers to find patterns and regularities in databases, perform prediction and forecasting, and generally improve their performance through interaction with data. It is currently regarded as the key element of a more general process called Knowledge Discovery that deals with extracting useful knowledge from raw data. The knowledge discovery process includes data selection, cleaning, coding, using different statistical and machine learning techniques, and visualization of the generated structures. The course will cover all these issues and will illustrate the whole process by examples. Special emphasis will be give to the Machine Learning methods as they provide the real knowledge discovery tools. Important related technologies, as data warehousing and on-line analytical processing (OLAP) will be also discussed.

## Course Objectives

The main objectives of this course are given below:

1. Introduce basic concepts and techniques of data warehousing and data mining
2. Examine the types of the data to be mined and apply pre-processing methods on raw data

- Discover interesting patterns, analyze supervised and unsupervised models and estimate the accuracy of the algorithms.

### **Course Outcomes**

Students are able to

CO No's	COs	Level
CO1	The student will be able to <b>Understand</b> the process of Knowledge Discovery of Databases and the architectures.	L2
CO2	The student will be able to <b>Apply</b> the preprocessing techniques like cleaning, integration, reduction, transformation and discretization.	L3
CO3	The student will be able to <b>Apply</b> the various data mining techniques like frequent pattern and association rule mining techniques, classification and clustering techniques for the given data to be mined.	L3
CO4	The student will be able to <b>Analyze</b> the given data using various mining methodologies and techniques to mine and discover interesting patterns for decision support.	L4

S.No	Unit	Description	Teaching Aids	CO
1	I	Definition of data ware house, subject oriented integrated, time variant, nonvolatile collection of data	BB	CO1
2		Data modeling-data cube what is OLAP and OLTP	BB	CO1
3		Multi-dimensional data modeling-differences	BB	CO1
4		Schemas, snow flake schema	BB	CO1
5		Fact constellation schema	BB	CO1
6		Star schema	BB	CO1
7		OLAP operations-slice, dice	BB	CO1
8		Pivot, rollup, drilldown based on concept hierarchy	BB	CO1
9		Architectural frame work of data ware house 3 tier architecture	BB	CO1
10		Introduction of data mining and What is data mining	BB	CO1
11		kinds of data need to be mined and patterns can be mined	BB	CO4
12		Technologies and kinds of applications are targeted	PPT	CO1
13		Data mining functionalities	BB	CO1
14		Data mining issues & applications	BB	CO1

15	II	Data preprocessing, why we need preprocessing,	BB	CO1
16		Quality data	BB, PPT	CO1
17		Data cleaning, missing values, noisy data	BB	CO1
18		Binning, clustering, combined computed, human Regression inspection	BB	CO1
19		Data integration-redundancy of data core relation analysis	BB	CO4
20		Data transformation-smoothing,	BB	CO1
21		Aggregation	BB	CO1
22		Generalization	BB	CO1
23		Normalization-MinMax	BB	CO1
24		Z-score	BB	CO1
25		Decimal scaling	BB	CO1
26		Data reduction-strategies ,data cube aggregation	BB	CO1
27		Dimensionality reduction,	BB	CO1
28		Numerosity reduction	BB	CO1
29		Histogram, clustering, sampling	BB	CO1
30		Data discretization, concept hierarchy	BB	CO1
31		Entropy based discretization	BB	CO1
32		Segmentation by natural partition	BB	CO1
33	III	Basic Concepts of classification	BB	CO1
34		General Approach to solving a classification problem	BB	CO3
35		Different types of classification algorithm	BB	CO1
36		Attribute Selection Measures, Tree Pruning	BB	CO1
37		Scalability and Decision Tree Induction	BB	CO1
38		Visual Mining for Decision Tree Induction.	BB	CO1
39		Bayes classification with example	BB	CO3
40		Basic concepts and definitions of classification and prediction	BB	CO1
41	IV	Frequent Item set Generation	BB	CO3
42		Rule Generation:	BB	CO1
43		Confident Based Pruning,	BB	CO1
44		Apriori algorithm	BB	CO1
45		Apriori algorithm with example problem explanation	BB	CO4
46		Improving efficiency of Apriori	BB	CO1
47		Rule Generation in Apriori Algorithm,	BB	CO1
48		Compact Representation of frequent item sets	BB	CO1
49		FP-Growth Algorithm.	BB	CO1
50		Overview, Basics and Importance of Cluster Analysis,	BB	CO1
51	V	Clustering techniques Different Types of Clusters	PPT	CO1
52		Partition based clustering	BB	CO1
53		Hierarchy based clustering	BB	CO1



54		Density based clustering	BB	CO1
55		Grid based clustering	BB	CO1
56		K-means: The Basic K-means Algorithm	BB	CO1
57		K-means Additional Issues	BB	CO3
58		Bi-secting K Means Algorithm	BB	CO1
		<b>Total Classes</b>	<b>58</b>	

	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12	PSO1	PSO2
<b>CO1</b>														
<b>CO2</b>	3													
<b>CO3</b>	3								2	2			1	2
<b>CO4</b>		3							2	2			1	2

### Evaluation Pattern

S. No	Components	Internal	External	Total
1	Theory	30	70	100
2	Engineering Graphics/Design/Drawing	30	70	100
3	Practical	15	35	50
4	Mini Project/Internship/Industrial Training/ Skill Development programmes/Research Project	-	50	50
5	Project Work	60	140	200

Marks Range Theory (Max – 100)	Marks Range Lab (Max – 50)	Level	Letter Grade	Grade Point
≥ 90	≥ 45	Outstanding	A+	10
≥80 to <89	≥40 to <44	Excellent	A	9
≥70 to <79	≥35 to <39	Very Good	B	8
≥60 to <69	≥30 to <34	Good	C	7
≥50 to <59	≥25 to <29	Fair	D	6
≥40 to <49	≥20 to <24	Satisfactory	E	5
<40	<20	Fail	F	0
-		Absent	AB	0

Day/Time	09.00- 09.50	09.50- 10.40	11.00- 11.50	11.50- 12.40	01.40- 02.30	02.30- 03.20	03.20- 04.10	04.10- 05.00
<b>Mon</b>		B-Sec.		A-Sec.	B-Sec. – Lab.			
<b>Tue</b>	B-Sec.		A-Sec.					
<b>Wed</b>		A-Sec. – Lab.			B-Sec.			
<b>Thu</b>				A-Sec.		B-Sec.		
<b>Fri</b>			A-Sec.					
<b>Sat</b>	A-Sec.			B-Sec.				

### **UNIT WISE Important Questions**

#### UNIT-1

1. Write the differences between designing a data ware house and OLTP systems.
2. Explain the OLAP operations in a multi dimensional data model.
3. Discuss the motivation behind OLAP mining.
4. Explain the Architectural framework approach of a data warehouse.
5. Describe multitiered architecture data ware house.
6. Describe star schema for multi dimensional data model.
7. Differentiate operational data base systems and data ware hosing.
8. Discuss Star, Snowflake and Fact Constellation Schemas with suitable examples.
9. Describe the star cubing algorithm for computing iceberg cubes
10. What is multi dimensional data model?
11. Discuss the Database Architectures for parallel processing
12. With a neat sketch explain the 3 –tier architecture of Data warehouse
13. Discuss the various basic statistical measures used in data analysis.
14. Describe why concept hierarchies are useful in data mining
15. What is multidimensional analysis? Explain with example.
16. Give generalization based mining of plan databases by divide and conquer.

#### UNIT-2

1. Explain need of preprocessing technique and also explain different forms of data preprocessing briefly
2. Explain the basic methods for data cleaning
3. Explain various methods of data cleaning in detail
4. Explain about data integration and data transformation
5. Explain about data transformation by normalization.
6. Explain dimensionality reduction
7. Briefly define the major 4 types of concept hierarchies

8. Explain the Numerosity reduction techniques
9. Explain about concept hierarchy generation for categorical data.
10. Explain about Discretization and concept hierarchy generation

### **UNIT-3**

1. Illustrate Decision Tree Induction Algorithm using proper attribute selection measures.
2. Discuss
  - i. Naïve Bayesian classification
  - ii. Multilayer Feed-forward Neural Networks.
3. Compare classification and prediction methods.
4. What is decision tree induction? Discuss about different attribute selection measures considered during decision tree construction
5. Explain the algorithm for constructing decision tree from training samples.
6. Compare the advantages and disadvantages of eager classification versus lazy classification
7. Why naïve Bayesian classification called naïve? Briefly discuss major ideas of naïve Bayesian classification
8. Describe back propagation algorithm for neural network based classification of data?
9. Write short notes on Bayesian classification.
10. Explain Baye's theorem
11. What are the Bayesian classifiers? Explain about a) naïve Bayesian classification  
(b) Linear and multiple regression.
12. Discuss about Bagging and Boosting techniques for improving classifier and predictor accuracy.
13. Explain classification by case based reasoning
14. Define regression.
15. Write about classifier accuracy measures in detail and also explain predictor error measures

### **UNIT-4**

1. Discuss FP Tree Growth algorithm with an example.
2. Illustrate the method of generating association rules from frequent itemsets.
3. Explain the terms support and confidence.
4. Explain the associate rule mining using market basket analysis
5. Explain about apriori algorithm.
6. Can we design a method that mines the complete set of frequent item sets without candidate generation? If yes, explain with example (FP growth association rule mining algorithm).
7. How do you classify frequent pattern mining? Write the apriori algorithm for discovering frequent item sets for mining Boolean association rules
8. Write and explain FP growth association rule mining algorithm without candidate generation.
9. Define the following a) frequent item set b) Iceberg queries c) strong association rules.
10. Write a note on multidimensional association rule mining.

## UNIT-5

1. What is cluster analysis? What are some typical applications of clustering? What are some typical requirements of clustering in data mining?
2. Explain the following clustering algorithms
  - i. K-means
  - ii. K-medoids
  - iii. DB-Scan
3. Explain about cluster analysis. Explain different types of data in cluster analysis
4. Explain about different major categorization of clustering methods.
5. Explain the clustering methods
  - a) K-means
  - b) DBSCAN
6. a) Explain what is cluster analysis  
b) Discuss about model based clustering methods
6. Explain multiphase hierarchical clustering using dynamic modeling.