



COURSE HANDOUT

PART-A

Name of Course Instructor: Michael Sadgun Rao Kona

Course Name & Code : Data Warehousing and Data mining & 20CS01

L-T-P Structure : 3-0-0

Credits: 3

Program/Sem/Sec : B.Tech/IV/A

A.Y.: 2021-22

PREREQUISITE : DBMS and Probability and Statistics

COURSE EDUCATIONAL OBJECTIVES (CEOs): The Objective of the course is to introduce the concepts of data warehouse and data mining, which gives a complete description about the principles, used, architectures, applications, design and implementation of data mining and data warehousing concepts.

COURSE OUTCOMES (COs): At the end of the course, student will be able to

CO1	Summarize the architecture of data warehouse.(Understand- L2)
CO2	Apply different preprocessing methods, Similarity, Dissimilarity measures for any given raw data.(Apply - L3)
CO3	Construct a decision tree and resolve the problem of model over fitting.(Analyze- L4)
CO4	Compare Apriori and FP-growth association rule mining algorithms for frequent itemset generation.(Apply - L3)
CO5	Apply suitable clustering algorithm for the given data set.(Apply - L3)

COURSE ARTICULATION MATRIX (Correlation between COs, POs & PSOs):

COs	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12	PSO1	PSO2	PSO3
CO1	-	3	-	-	-	-	-	-	-	-	-	-	2	-	-
CO2	3	1	-	-	-	-	-	-	-	-	-	-	3	-	-
CO3	-	-	3	2	-	-	-	-	-	-	-	-	-	3	-
CO4	-	-	3	2	-	-	-	-	-	-	-	-	-	-	2
CO5	-	-	3	2	-	-	-	-	-	-	-	-	-	2	-
1 - Low			2 -Medium			3 - High									

TEXTBOOKS:

- T1** Introduction to Data Mining : Pang-Ning Tan & Michael Steinbach, Vipin Kumar, Fifth Impression, Pearson, 2015.
- T2** Data Mining concepts and Techniques, 3rd Edition, Jiawei Han, Michel Kamber, Elsevier, 2011.

REFERENCE BOOKS:

- R1** Data Mining Techniques and Applications: An Introduction, Hongbo Du, Cengage Learning, 2010.
- R2** Data Mining : Introductory and Advanced topics : Dunham, First Edition, Pearson, 2020
- R3** Data Warehousing Data Mining & OLAP, Alex Berson, Stephen Smith, TMH, 2008.

PART-B

COURSE DELIVERY PLAN (LESSON PLAN):

UNIT-I: Data Warehouse and OLAP Technology

S. No.	Topics to be covered	No. of Classes Required	Tentative Date of Completion	Actual Date of Completion	Teaching Learning Methods	HOD Sign Weekly
1.	Introduction to Course and COs	1	07.03.2022		TLM1,2	
2.	Introduction to Unit-I	1	09.03.2022			
3.	Data Warehouse and OLAP Technology: An Overview: Data Warehouse	1	10.03.2022		TLM1,2	
4.	Data Warehouse and OLAP Technology: A Multidimensional Data Model	1	11.03.2022		TLM1,2	
5.	Data Warehouse and OLAP Technology: Data Warehouse Architecture	1	14.03.2022		TLM1,2	
6.	Data Warehouse and OLAP Technology: Data Warehouse Implementation	2	16.03.2022, 21.03.2022		TLM1,2	
7.	Data Warehouse and OLAP Technology: From Data Warehousing to Data Mining.	2	23.03.2022, 25.03.2022		TLM1,2	
8.	Revision	1	28.03.2022		TLM1,2	
No. of classes required to complete UNIT-I: 10				No. of classes taken:		

UNIT-II: Data Mining & Data Preprocessing

S. No.	Topics to be covered	No. of Classes Required	Tentative Date of Completion	Actual Date of Completion	Teaching Learning Methods	HOD Sign Weekly
1.	Data Mining: Introduction to Data Mining	1	30.03.2022		TLM1,2	
2.	Motivating challenges, The origins of Data Mining,	1	01.04.2022		TLM1,2	
3.	Data MiningTasks, Types of Data, Data Quality.	2	04.04.202,2 06.04.2022		TLM1,2	
4.	Data Preprocessing: Aggregation	1	07.04.2022		TLM1,2	
5.	Data Preprocessing: Sampling, Dimensionality Reduction, Feature Subset Selection	1	08.04.2022		TLM1,2	
6.	Data Preprocessing: Feature creation	1	11.04.2022		TLM1,2	
7.	Data Preprocessing: Discretization and Binarization	1	13.04.2022		TLM1,2	
8.	Data Preprocessing: Variable Transformation	1	14.04.2022		TLM1,2	
9.	Data Preprocessing: Measures of Similarity and Dissimilarity	1	18.04.2022		TLM1,2	
10.	Revision	1	20.04.2022		TLM1,2	
No. of classes required to complete UNIT-II: 11				No. of classes taken:		

UNIT-III: Classification & Model Over fitting

S. No.	Topics to be covered	No. of Classes Required	Tentative Date of Completion	Actual Date of Completion	Teaching Learning Methods	HOD Sign Weekly
1.	Classification: Basic Concepts, General Approach to solving a classification problem	1	21.04.2022	22.04.2022	TLM1,2	
2.	Decision Tree Induction: Working of Decision Tree, building a decision tree	1	22.04.2022	22.04.2022	TLM1,2	
Mid - I Examinations from 25.04.2022 to 30.04.2022						
3.	methods for expressing an attribute test conditions, measures for selecting the best split	1	02.05.2022		TLM1,2	
4.	Algorithm for decision tree induction.	1	04.05.2022		TLM1,2	
5.	Model Overfitting: Due to presence of noise, due to lack of representation samples	1	05.05.2022		TLM1,2	
6.	Evaluating the performance of classifier: holdout method, random sub sampling, cross-validation, bootstrap.	2	06.05.2022, 09.05.2022		TLM1,2	
7.	BayesTheorem	2	11.05.2022, 12.05.2022		TLM1,2	
8.	Naïve Bayes Classifier	2	13.05.2022, 16.05.2022		TLM1,2	
No. of classes required to complete UNIT-III: 11				No. of classes taken:		

UNIT-IV: Association Analysis

S. No.	Topics to be covered	No. of Classes Required	Tentative Date of Completion	Actual Date of Completion	Teaching Learning Methods	HOD Sign Weekly
1.	Association Analysis: Basic Concepts	1	18.05.2022		TLM1,2	
2.	Algorithms: Problem Definition, Frequent Item Set Generation	2	19.05.2022, 20.05.2022		TLM1,2	
3.	Apriori Principle, Apriori Algorithm	2	23.05.2022, 25.05.2022		TLM1,2	
4.	Rule Generation, Compact Representation of Frequent Itemsets	2	26.05.2022, 27.05.2022		TLM1,2	
5.	FPGrowth Algorithm	2	30.05.2022, 31.05.2022		TLM1,2	
No. of classes required to complete UNIT-IV: 09				No. of classes taken:		

UNIT-V: Memory System Design, Peripheral Devices and their characteristics

S. No.	Topics to be covered	No. of Classes Required	Tentative Date of Completion	Actual Date of Completion	Teaching Learning Methods	HOD Sign Weekly
1.	Cluster Analysis: Basic Concepts and Algorithms: Preliminaries	1	01.06.2022		TLM1,2	
2.	Different Types ofClustering, Different Types of Clusters;	1	02.06.2022		TLM1,2	
3.	K-means: The Basic K-means Algorithm	1	03.06.2022		TLM1,2	
4.	K-means Additional Issues, Bisecting K-means, Strengths and Weaknesses;	1	06.06.2022		TLM1,2	
5.	Exercise problems on K-means	1	08.06.2022		TLM1,2	
6.	Agglomerative Hierarchical Clustering: Basic Agglomerative Hierarchical Clustering Algorithm	1	09.06.2022		TLM1,2	
7.	Exercise problems on Agglomerative Hierarchical Clustering Algorithm	1	10.06.2022		TLM1,2	
8.	DBSCAN: Traditional Density Center-Based Approach,DBSCAN Algorithm, Strengths and Weaknesses.	1	13.06.2022		TLM1,2	
9.	Exercise problems on DBSCAN Algorithm	1	15.06.2022		TLM1,2	
No. of classes required to complete UNIT-V: 09				No. of classes taken:		

CONTENT BEYOND THE SYLLABUS:

S. No.	Topics to be covered	No. of Classes Required	Tentative Date of Completion	Actual Date of Completion	Teaching Learning Methods	HOD Sign Weekly
1.	Regression Analysis - I (Linear Regression)	1	16.06.2022		TLM1,2	
2.	Regression Analysis - II (Logistic Regression)	1	17.06.2022		TLM1,2	

Teaching Learning Methods					
TLM1	Chalk and Talk	TLM4	Demonstration (Lab/Field Visit)		
TLM2	PPT	TLM5	ICT (NPTEL/Swayam Prabha/MOOCs)		
TLM3	Tutorial	TLM6	Group Discussion/Project		

PART-C

EVALUATION PROCESS (R20 Regulation):

Evaluation Task	Marks
Assignment-I (Units-I, II & UNIT-III (Half of the Syllabus))	A1=5
I-Descriptive Examination (Units-I, II & UNIT-III (Half of the Syllabus))	M1=15
I-Quiz Examination (Units-I, II & UNIT-III (Half of the Syllabus))	Q1=10
Assignment-II (Unit-III (Remaining Half of the Syllabus), IV & V)	A2=5
II- Descriptive Examination (UNIT-III (Remaining Half of the Syllabus), IV & V)	M2=15
II-Quiz Examination (UNIT-III (Remaining Half of the Syllabus), IV & V)	Q2=10
Mid Marks =80% of Max ((M1+Q1+A1), (M2+Q2+A2)) + 20% of Min ((M1+Q1+A1), (M2+Q2+A2))	M=30
Cumulative Internal Examination (CIE): M	30
Semester End Examination (SEE)	70
Total Marks = CIE + SEE	100

PART-D**PROGRAMME OUTCOMES (POs):**

PO 1	Engineering knowledge: Apply the knowledge of mathematics, science, engineering fundamentals, and an engineering specialization to the solution of complex engineering problems.
PO 2	Problem analysis: Identify, formulate, review research literature, and analyze complex engineering problems reaching substantiated conclusions using first principles of mathematics, natural sciences, and engineering sciences.
PO 3	Design/development of solutions: Design solutions for complex engineering problems and design system components or processes that meet the specified needs with appropriate consideration for the public health and safety, and the cultural, societal, and environmental considerations.
PO 4	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.
PO 5	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.
PO 6	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.
PO 7	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.
PO 8	Ethics: Apply ethical principles and commit to professional ethics and responsibilities and norms of the engineering practice.
PO 9	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 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 multidisciplinary environments.
P012	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

PROGRAMME SPECIFIC OUTCOMES (PSOs):

PSO 1	The ability to apply Software Engineering practices and strategies in software project development using open-source programming environment for the success of Organization.
PSO 2	The ability to design and develop computer programs in networking, web applications and IoT as per the society needs
PSO 3	To inculcate an ability to analyze, design and implement database applications.

Title	Course Instructor	Course Coordinator	Module Coordinator	Head of the Department
Name of the Faculty	Mr.Michael Sadgun Rao.K	Dr.K.Anu Priya	Dr.K. Lavanya	Dr.B.Srinivasa Rao
Signature				