COMPUTER SCIENCE AND ENGINEERING

COURSE DESCRIPTION FORM

Course Title	**DATA WAREHOUSING AND DATA MINING**			
Course Code	15ACS22			
Regulation	**R15– JNTUACEP**			
Course Structure	Lectures	Tutorials	Practicals	
	4	-	-	
Course Coordinator	Mr. D. Mahendra Reddy			
Team of Instructors	Dr. G.Murali, Assistant Professor & HOD			

I. COURSEOVERVIEW:

The course addresses the concepts, skills, methodologies, and models of data warehousing. The proper techniques for designing data warehouses for various business domains, and covers concepts for potential uses of the data warehouse and other data repositories in mining opportunities are addressed. Data mining, the extraction of hidden predictive information from large databases, is a powerful new technology with great potential to help companies focus on the most important information in their data warehouses. Data mining tools predict future trends and behaviors, allowing businesses to make proactive, knowledge- driven decisions.

II. PREREQUISITE(S):

Level	Credits	Periods/ Week	Prerequisites
UG	4	4	Data warehouse and data mining

III. MARKS DISTRIBUTION:

Sessional Marks	University End Exam marks	Total Marks
Midterm Test There shall be two midterm examinations. Each midterm examination consists of essay paper, objective paper and assignment. The objective paper is for 10 marks of 20 minutes duration. It consists of 10 multiple choice and 10 fill-in-the blank questions, the student has to answer all the questions and each carries half mark. First midterm examination shall be conducted for the first two and half units of syllabus and second midterm examination shall be conducted for the remaining portion. Five marks are earmarked for assignments. There shall be two assignments in every theory course. Assignments are usually issued at the time of commencement of the semester. These are of problem solving in nature with	70	100

IV. EVALUATIONSCHEME:

S. No	Component	Duration	Marks
1.	I Mid Examination	90 minutes	30
2.	I Assignment	-	
3.	II Mid Examination	90 minutes	30
4.	II Assignment	-	
5.	External Examination	3 hours	70

V. COURSEOBJECTIVES:

At the end of the course, the students will be able to:

I. Be familiar with mathematical foundations of data miningtools.

II. Understand and implement classical models and algorithms in data warehouses and datamining.

III. Characterize the kinds of patterns that can be discovered by association rule mining, classification andclustering.

IV. Master data mining techniques in various applications like social, scientificand environmental context.

V. Develop skill in selecting the appropriate data mining algorithm for solving practicalproblems.

VI. Be familiar with the process of data analysis, identifying the problems, and choosing the relevant models and algorithms toapply.

VI. COURSEOUTCOMES:

After completing this course the student must demonstrate the knowledge and ability to:

1. **Learn** data warehouse principles, data mining concepts andworking.
2. **Understand** various data preprocessing procedures and their applicationscenarios.
3. **Discuss** the data-mining tasks like classification, regression, clustering, associationmining.
4. **Understand** the impact of machine learning solutions on the society and also the contemporaryissues.
5. **Analyze** and choose a suitable data mining task to theproblem.
6. **Visualize** and interpret the results produced by datamining.
7. **Build** statistical predictive models using various techniques such as neural networks, decision trees and logisticregression.
8. **Solve** real-world problems in business and scientific information using datamining.
9. **Acquire** hands-on experience with key components of an integrated data warehousing and business intelligence system using a leading industry commercial applicationpackage.

VII. HOW PROGRAM OUTCOMES AREASSESSED:

Program Outcomes		Level	Proficiency assessed by
PO1	**Engineering knowledge**: Apply the knowledge of mathematics, science, engineering fundamentals, and an engineering specialization to the solution of complex engineering problems.	H	Assignments, Tutorials
PO2	**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.	H	Lectures, Assignments, Exams
PO3	**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.	S	Mini Projects
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.	S	Projects
PO5	**Modern tool usage**: Create, select, and apply appropriate techniques, resources, and modern engineeringand IT tools including prediction and modeling to complex engineering activities with an understanding of the limitations.	S	Projects
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.	N	--
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.	N	--
PO8	**Ethics**: Apply ethical principles and commit to professional ethics and responsibilities and norms of the engineering practice.	S	Discussions
PO9	**Individual and team work**: Function effectively as an individual, and as a member or leader in diverse teams, and in multidisciplinary settings.	N	--
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.	N	--
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.	N	--
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.	S	Projects

N-None **S-Supportive** **H - HighlyRelated**

VIII. HOW PROGRAM SPECIFIC OUTCOMES AREASSESSED:

Program Specific Outcomes		Level	Proficiency assessed by
PSO1	**Professional Skills:** The ability to research, understand and implement computer programs in the areas related toalgorithms, system software, multimedia,webdesign, bigdataanalytics, and networking for efficient analysis and design of computer-based systems of varying complexity.	H	Lectures, Assignments
PSO2	**Problem-Solving Skills:** The ability to apply standardpractices and strategies in software project development using open-ended programming environments to deliver a quality product for business success.	H	Projects
PSO3	**Successful Career and Entrepreneurship:** The ability to employ modern computer languages, and environments,platforms in creating innovative career paths, to be an entrepreneur, and a zest for higherstudies.	S	Guest Lectures

N-None **S-Supportive** **H - HighlyRelated**

IX. SYLLABUS:

UNIT I: Introduction to Data Warehousing

Evolution of Decision Support Systems- Data warehousing Components –Building a Data warehouse, Data Warehouse and DBMS, Data marts, Metadata, Multidimensional data model, OLAP vs OLTP, OLAP operations, Data cubes, Schemas for Multidimensional Database: Stars, Snowflakes and Fact constellations.

UNIT II: Data Warehouse Process and Architecture

Types of OLAP servers, 3–Tier data warehouse architecture, distributed and virtual data warehouses. Data warehouse implementation, tuning and testing of data warehouse. Data Staging (ETL) Design and Development, data warehouse visualization, Data Warehouse Deployment, Maintenance, Growth, Business Intelligence Overview- Data Warehousing and Business Intelligence Trends - Business Applications- tools-SAS.

UNIT III: Introduction to Data Mining

Data mining-KDD versus data mining, Stages of the Data Mining Process-task primitives, Data Mining Techniques -Data mining knowledge representation – Data mining query languages, Integration of a Data Mining System with a Data Warehouse – Issues, Data pre-processing – Data cleaning, Data transformation, Feature selection, Dimensionality reduction, Discretization and generating concept hierarchies-Mining frequent patterns- association-correlation.

UNIT IV: Classification and Clustering

Decision Tree Induction - Bayesian Classification – Rule Based Classification –Classification by Back propagation – Support Vector Machines – Associative Classification – Lazy Learners – Other Classification Methods - Clustering techniques – , Partitioning methods- k-means Hierarchical Methods - distance-based agglomerative and divisible clustering, Density-Based Methods – expectation maximization -Grid Based Methods – Model-Based Clustering Methods – Constraint – Based Cluster Analysis – Outlier Analysis.

UNIT V: Data Warehousing and Data Mining Software's and Applications

Mining complex data objects, spatial databases, temporal databases, Multimedia databases, Time series and Sequence data; Text Mining –Graph mining-web mining-Application and Trends in data mining.

TEXT BOOKS:

1. Jiawei Han and MichelineKamber, "Data Mining: Concepts and Techniques", Morgan Kaufmann Publishers, third edition2011, ISBN: 1558604898.
2. Alex Berson and Stephen J. Smith, "Data Warehousing, Data Mining & OLAP", TataMc Graw Hill Edition, Tenth Reprint 2007.
3. G. K. Gupta, "Introduction to Data Min Data Mining with Case Studies", Easter Economy Edition, Prentice Hall of India, 2006.

REFERENCES:

1. Mehmedkantardzic, "Dataminingconcepts, models, methods, and algorithms", Wiley Interscience, 2003.
2. Ian Witten, Eibe Frank, Data Mining; Practical Machine Learning Tools and Techniques, Third edition, Morgan Kaufmann, 2011.
3. George M Marakas, Modern Data Warehousing, Mining and Visualization, Prentice Hall, 2003.

X. COURSEPLAN:

At the end of the course, the students are able to achieve the following course learning outcomes:

Lecture No.	Topics to be covered	Course Learning Outcomes	References
Lecture 1,2 (1-7-19)	UNIT I Introduction to data ware house	**Outline** the importance of data Warehouse and databases.	T1:125 T2:4-15 R3:4
Lecture 3 (2-7-19)	Evolution data ware house and decision support system	**Identify** decision support system in data warehouse .	T1: 126 T2:169-185 R3:6
Lecture 4 (3-7-19)	data ware house architecture,	**Explain** architecture of data warehouse	T1: 127 T2:151-153 R3:51-53
Lecture 5,6 (8-7-19)	data ware house components, building data warehouse.	Explain data warehouse components	T1: 127 T2:115-127
Lecture 7 (9-7-19)	Building a Data warehouse, Data Warehouse and DBMS	**Difference** between data warehouse and dbms	T1: 128-130 T2:130-149 R3:180-186
Lecture 8 (10-7-19)	. Data marts, Metadata, Multidimensional data model	**Define** multidimentional data model	T1: 133-135 T2:205-219 R3:17-19
Lecture 9,10 (15-7-19)	OLAP vs OLTP, OLAP operations,	**Explain OLAP and OLTP**	T1: 135 T2:247-265
Lecture 11 (16-7-19)	Data cubes, Schemas for Multidimensional Database	**Explain** data cubes	T1: 136-139 T2:251-255
Lecture 12 (17-7-19)	Stars, Snowflakes design	**Explain** multidimensional DB	T1: 139-142 T2:248-256
Lecture 13,14 (22-7-19)	Fact constellation https://www.youtube.com/watch?v	**Learn design of data warehouse**	T1:143
Lecture 15 (23-7-19)	UNIT II Types of OLAP servers	**Explain** types of OLAP	T1: 131-133 T2:251-255
Lecture 16 (25-7-19)	3–Tier data warehouse architecture, distributed and virtual data warehouses	**Illustrate** 3 tier data warehouse and virtual data	T1: 148-150

Lecture 17,18 (29-7-19)	Data warehouse implementation, tuning and testing of data warehouse.	**Define** implementation and testing	T1: 6.1- 6.2
Lecture 19 (30-7-19)	Data Staging (ETL) Design and Development,	**Illustrate** development ETL design	T1: 150-154
Lecture 20 (1-8-19)	data warehouse visualization	**Understand** visualization	T1: 7.3
Lecture 21,22 (5-8-19)	Data Warehouse Deployment	**Understand** deployment	T1: 8.1
Lecture 23 (6-8-19)	DW Maintenance, Growth, Business Intelligence Overview	**Explain** maintenance nad growth	T1: 8.2
Lecture 24 (7-8-19)	DW Business Intelligence Overview	**Explain** business intelligence	T1: 8.3-8.4
Lecture 25,26 (12-8-19)	Data Warehousing and Business Intelligence Trends	**Define** business intelligence trends	T1: 9.2
Lecture 27 (13-8-19)	Business Applications	**Define** DW application	T1: 9.3
Lecture 28 (14-8-19)	DW tools-SAS.	**Define** DW tools	T1: 9.5-9.6
Lecture 29,30 (26-8-19)	Data warehouse applications https://www.youtube.com/watch?v=m-aKj5ovDfg	**Explain** example application	T1: 5.4
Lecture 31 (27-8-19)	UNIT III Data mining-KDD versus data mining	**Explain** data mining and KDD	T1: 1-5 T2:331-336 R3:81-83

Lecture 32 (28-8-19)	Stages of the Data Mining Process-task primitives	**Understand** stages of data mining	T1:7-9 T2:340-342
Lecture 33,34 (29-9-18)	Data Mining Techniques -Data mining knowledge representation	**Understand** techniques and representation	T1:10-12 R3:78-80
Lecture 35 (4-9-18)	Data mining query languages, Integration of a Data Mining System with a Data Warehouse	**Analyze** query language	T1: 10.5
Lecture 36 (5-9-18)	Issues, Data pre-processing – Data cleaning, Data transformation,	**Learn** data cleaning and data transformation	T1: 84-86
Lecture 37,38 (9-9-19)	Feature selection, Dimensionality reduction	**Learn** dimension reduction	T1: 88-91
Lecture 39 (10-9-19)	Dimensionality reduction, Discretization and generating concept hierarchies	Learn discretization	T1: 111-118
Lecture 39, (11-9-19)	Mining frequent patterns-association-correlation. Bits for unit III	Learn patterns and associtio n	T1: 17-19
Lecture 40,41 (16-9-19)	Unit IV Decision Tree Induction - Bayesian Classification	Learn decision tree	T1: 330-348 T2:351-355

Lecture 42 (17-9-19)	Rule Based Classification – Classification by Back propagation	**Explain** classification and back propogation	T1:355-359
Lecture 43, (18-9-19)	Support Vector Machines – Associative Classification – Lazy Learners – Other Classification Methods	Explain associative classification method	T1: 408-413
Lecture 44,45 (23-9-19)	Clustering techniques – , Partitioning methods- k-means Hierarchical Methods	Learn k-means	T1: 497-505 T2:431-435
Lecture 46 (24-9-19)	based agglomerative and divisible clustering, Density-Based Methods	Learn clustering	T1:508-519 T2:409-412
Lecture 47, (25-9-19)	expectation maximization - Grid Based Methods – Model-Based Clustering	Learn grid based methods	T1:479-481
Lecture 49,50 30-10-19)	Constraint – Based Cluster Analysis	Learn cluster analysis	T1:497

Lecture 51 (1-10-19)	Outlier Analysis.	Learn analysis	T1:544-548
Lecture 52 (2-10-19)	UNIT V Mining complex data objects, spatial databases, temporal databases	Learn data mining application	T1:585-595
Lecture 53,54 (7-10-19)	Multimedia databases, Time series and Sequence data	Learn multimedia database	R3-244-246
Lecture 55 (8-10-19)	Text Mining –Graph mining-web mining	Learn graph mining and text mining	R3:242-243
Lecture 56 (9-10-19)	Application and Trends in data mining.	Learn trends in data mining	T1:607-615
Lecture 57,58 (14-10-18)	MCQ for unit V	Learn all topics in DWH and DM applications	

XI. MAPPING COURSE OBJECTIVES LEADING TO THE ACHIEVEMENTOF PROGRAM OUTCOMES AND PROGRAM SPECIFICOUTCOMES:

Course Objectives	Program Outcomes												Program Specific Outcomes		
	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12	PSO1	PSO2	PSO3
I	H								S		S		H		S
II	H	S		S	H								H	S	
III	H	S		S	H								H	S	
V	S	S	S	S	H								S	H	
V		S	S				S						S	H	S
VI	H	S													S

S–Supportive **H - HighlyRelated**

XII. MAPPING COURSE OUTCOMES LEADING TO THE ACHIEVEMENT OFPROGRAM OUTCOMES AND PROGRAM SPECIFICOUTCOMES:

Course Outcomes	Program Outcomes												Program Specific Outcomes		
	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12	PSO1	PSO2	PSO3
1	H							S				S	H		S
2	H	S											H		
3	H	S											H		
4	H	S											H	S	
5			H	S	S								S	H	
6		H		S	H								S		
7			H	H	H			S					H	H	
8			H	H	H		S	S					H	H	S
9		H		H	H		S	S				S	H	H	S

S–Supportive **H-HighlyRelated**