INSTITUTE OF AERONAUTICAL ENGINEERING

(Autonomous)

Dundigal, Hyderabad - 500 043

Department of Computer Science and Engineering

COURSE DESCRIPTION FORM

Course Title	**DATA WAREHOUSING AND DATA MINING**			
Course Code	**A70520**			
Regulation	**R15 – JNTUH**			
Course Structure	Lectures	Tutorials	Practicals	Credits
	4	-	-	4
Course Coordinator	Dr. K Suvarchala , Professor, CSE			
Team of Instructors	Dr. M. Madhubala, Professor, CSE Ms. B. Padmaja Associate Professor, CSE Mr. P Anjaiah , Assistant Professor, CSE			

I. COURSE OVERVIEW:

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	Database Management Systems, Probability & Statistics

III. MARKS DISTRIBUTION:

Sessional Marks	University End Exam Marks	Total Marks
Mid Semester Test There shall be 2 midterm examinations. Each midterm examination consists of subjective type and Objective type tests. The subjective test is for 10 marks, with duration of 1 hour. Subjective test of each midterm exam shall contain 4 questions. The student has to answer 2 questions, each carrying 5 marks. The objective type test is for 10 marks with duration of 20minutes. It consists of 10 Multiple choice and 10 objective type questions. The student has to answer all the questions and each carries half mark. First midterm examination shall be conducted for the first 2 ½ units of syllabus and second midterm examination shall be conducted for the remaining 2 ½ units.	75	100

Seasonal Marks	University End Exam Marks	Total Marks
Five marks are earmarked for assignments. There shall be two assignments in every theory course. Marks shall be awarded considering the average of two assignments in each course reason whatsoever, will get zero marks(s).		

IV. EVALUATION SCHEME:

S. No	Component	Duration	Marks
1	I Mid Examination	90 minutes	20
2	I Assignment	-	5
3	II Mid Examination	90 minutes	20
4	II Assignment	-	5
5	External Examination	3 hours	75

V. COURSE OBJECTIVES:

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

I. Learn mathematical foundations of data mining tools.
II. Understand and implement classical models and algorithms in data warehouses and data mining.
III. Analyze the kinds of patterns that can be discovered by association rule mining, classification and clustering.
IV. Apply data mining techniques in various applications like social, scientific and environmental context.
V. Develop skill in selecting the appropriate data mining algorithm for solving practical problems.
VI. Understand with the process of data analysis, identifying the problems, and choosing the relevant models and algorithms to apply.

VI. COURSE OUTCOMES:

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

1. Learn data warehouse principles, data mining concepts and working.
2. Understand various data preprocessing procedures and their application scenarios.
3. Discuss the data-mining tasks like classification, regression, clustering, association mining.
4. Understand the impact of machine learning solutions on the society and also the contemporary issues.
5. Explore a suitable data mining task to the problem.
6. Visualize and interpret the results produced by data mining.
7. Build statistical predictive models using various techniques such as neural networks, decision trees and logistic regression.
8. Solve real-world problems in business and scientific information using data mining.
9. Acquire hands-on experience with key components of an integrated data warehousing and business intelligence system using a leading industry commercial application package.
10. Analyze and compare various data mining techniques based on different parameters.

VII. HOW PROGRAM OUTCOMES ARE ASSESSED:

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, Analyze complex engineering problems reaching substantiated conclusions using first principles of mathematics, natural sciences, and engineering sciences.	H	Assignments
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.	H	Projects
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.	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	Oral 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.	S	Presentations
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.	H	Development of Prototype, Projects
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	Seminars, Discussions

N - None **S - Supportive** **H - Highly Related**

VIII. HOW PROGRAM SPECIFIC OUTCOMES ARE ASSESSED:

Program Specific Outcomes		Level	Proficiency Assessed by
PSO1	**Professional Skills:** An ability to understand the basic concepts in Electronics & Communication Engineering and to apply them to various areas, like Electronics, Communications, Signal processing, VLSI, Embedded systems etc., in the design and implementation of complex systems.	H	Lectures, Assignments
PSO2	**Problem-Solving Skills:** An ability to solve complex Electronics and communication Engineering problems, using latest hardware and software tools, along with analytical skills to arrive cost effective and appropriate solutions.	S	Tutorials
PSO3	**Successful career and Entrepreneurship:** An understanding of social-awareness & environmental-wisdom along with ethical responsibility to have a successful career and to sustain passion and zeal for real-world applications using optimal resources as an Entrepreneur.	S	Seminars and Projects

N - None **S - Supportive** **H - Highly Related**

IX. SYLLABUS:

UNIT - I

Data warehouse: Introduction to Data warehouse, Difference between operational database systems and data warehouses, Data warehouse characteristics, Data warehouse Architecture and its components, Extraction-Transformation-Loading, Logical (Multi-Dimensional), Data Modeling, Schema Design, Star and show-Flake Schema, Fact Consultation, Fact Table, Fully Addictive, Semi-Addictive, Non Addictive Measures; Fact-Less Facts, Dimension Table Characteristics; OLAP Cube, OLAP Operations, OLAP Server Architecture-ROLAP, MOLAP and HOLAP.

UNIT - II

Introduction to Data Mining: Introduction, What is Data Mining, Definition, KDD, Challenges, Data Mining Tasks, Data Processing, Data Cleaning, Missing data, Dimensionality Reduction, Feature Subset Selections, Discretization and Binaryzation, Data Transformation; Measures of Similarity and Dissimilarity-Basics.

UNIT - III

Association Rules: Problem Definition, Frequent item set generation, The APRIORI Principle, support and confidence measures, association rule generation; APRIORI algorithm. The Partition Algorithms, FP-Growth Algorithms, Compact Representation of Frequent item Set-Maximal Frequent item set, closed frequent item set.

UNIT - IV

Classification: Problem Definition, General Approaches to solving classification problem, Evaluation of Classifiers, Classification techniques, Decision Trees-Decision trees construction, Methods, for expressing attribute test conditions, Measures for selecting the best split, Algorithm for Decision tree induction; Naive- Bayes Classifier, Bayesian Belief Network; K-Nearest neighbor classification-Algorithm and characteristics.

UNIT - V

Clustering: Problem Definition, Clustering Overview, Evaluation of Clustering algorithms, partitioning clustering-K-Means Algorithm, K-Means Additional issues, PAM Algorithm; Hierarchical Clustering-Agglomerative Methods and divisive methods, Basic Agglomerative Hierarchical Clustering Algorithms Specific techniques, Key issues in Hierarchical Clustering, Strengths and weakness; outlier detection.

X. TEXT BOOKS:

1. Jiawei Han and Micheline Kamber, "Data Mining: Concepts and Techniques", 2e, Elsevier, 2008.
2. Pangning Tan Vipin Kumar Micaelsteinbanch, "Introduction to data mining ", Pearson Education.

XI. REFERENCE BOOKS:

1. Margaret H Dunham, "Data Mining Introductory and Advanced Topics", 2e, Pearson Education, 2006.
2. Amitesh Sinha, "Data Warehousing", Thomson Learning, 2007.
3. Arun K pujari, "Data Mining Techniques", 3e, Universities Press.
4. VikramPudi, P Radha Krishna, "Data Mining", Oxford University Press.

XII. COURSE PLAN:

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

Lecture No.	Unit No	Course Learning Objective	Topics Covered	Text Book / Reference
1-3	I	Distinguish between data warehouse from other databases.	Introduction to Data warehouse, Difference between operational database systems and data warehouses, Data warehouse characteristics.	T1: 3.1
4-6		Prepare Data warehouse architecture.	Data warehouse Architecture and its components, Extraction-Transformation-Loading, Logical (Multi-Dimensional).	T1: 3.3
7-10		Design Multidimensional Data Model.	Modeling, Schema Design, Star and show-Flake Schema, Fact Consultation, Fact Table, Fully Addictive, Semi-Addictive, Non Addictive Measures; Fact-Less Facts, Dimension Table	T1: 3.2
11-14		Implementation of Data Warehouse.	OLAP Cube, OLAP Operations, OLAP Server Architecture-ROLAP, MOLAP and HOLAP.	T1: 3.4-3.5
15-18	II	Outline the importance of data mining in big data technology.	Introduction, Fundamentals of Data Mining, Definition, KDD, Challenges, Data Mining Tasks.	T1: 1.1-1.7
19-23		List the data Preprocessing techniques.	Data Processing, Data Cleaning, Missing data, Dimensionality Reduction, Feature Subset Selections, Data Transformation.	T1: 2.1-2.5
24-26		Define data Discretization.	Discretization and Measures of Similarity and Dissimilarity-Basics.	T1: 2.6
27-28	III	Illustrate the process of association rule mining	Association Rules: Problem definition, Frequent item set generation, The APRIORI Principle.	T1: 5.3
29-30		Define basic concepts of Apriori Algorithm.	Support and confidence measures, association rule generation; APRIORI algorithm.	T1: 5.2
31-33		Define basic concepts of frequent pattern mining.	The partition algorithms, fp-growth Algorithm.	T1: 5.2
34-37		Illustrate frequent item set.	Compact Representation of Frequent item Set-Maximal Frequent item set, closed frequent item set.	T1: 5.1
38-41	IV	Describe Classification.	Classification Problem Definition, General Approaches to solving classification problem, Evaluation of Classifiers, Classification techniques.	T1: 6.1-6.2

42-45		Construct an optimal decision tree for a given dataset.	Decision Trees-Decision trees construction, Methods, for expressing attribute test conditions, Measures for selecting the best split, Algorithm for Decision tree induction.	T1: 6.3
46-48		Classify Bayesian methods.	Naive-bayes classifier, Bayesian belief network.	T1: 6.4
49-50		Generalize the learning from your neighbors.	K-Nearest neighbor classification-Algorithm and characteristics.	T1: 6.9
51-52	V	Differentiate classification and clustering.	Clustering Problem Definition, Clustering Overview, Evaluation of Clustering algorithms.	T1: 7.1-7.3
53-55		Understand partitioning methods used for clustering.	Partitioning clustering-K-Means Algorithm, K-Means Additional issues, PAM Algorithm.	T1: 7.4
56-58		Identify various types of Hierarchical clustering techniques	Hierarchical Clustering-Agglomerative Methods and divisive methods, Basic Agglomerative Hierarchical Clustering Algorithms Specific techniques.	T1: 7.5
59-60		Analyze outlier detection methods.	Key issues in Hierarchical Clustering, Strengths and weakness; outlier detection.	T1: 7.11

XIII. MAPPING COURSE OBJECTIVES LEADING TO THE ACHIEVEMENT OF PROGRAM OUTCOMES AND PROGRAM SPECIFIC OUTCOMES:

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	
IV	S	S	S	S	H								S	H	
V		S	S				S						S	H	S
VI	H	S													S

S – Supportive **H - Highly Related**

XIV. MAPPING COURSE OUTCOMES LEADING TO THE ACHIEVEMENT OF PROGRAM OUTCOMES AND PROGRAM SPECIFIC OUTCOMES:

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
10	H		H	H		S	S				S	H	H	S	H

S – Supportive **H - Highly Related**

Prepared by: Dr. K Suvarchala, Professor, CSE

HOD, CSE