

CE350: DATA WAREHOUSING & DATA MINING

Credits and Hours:

Teaching Scheme	Theory	Practical	Tutorial	Total	Credit
Hours/week	3	2	0	5	4
Marks	100	50	0	150	

A. Objective of the Course:

The main objectives for offering the course Data Warehousing and Data Mining are:

- To understand the concept of Data Warehousing and mining.
- To have hands on experience on various Data Mining tools.
- To have insight of Data Pre-processing, Classification, Association and Prediction and various data mining related advance topics like text mining, web mining etc.

B. Outline of the Course:

Sr. No.	Title of the unit	Minimum number of Hours
1.	Introduction to Data Warehousing	06
2.	Concepts and techniques in Data Warehousing	04
3.	Introduction to data mining (DM)	04
4.	Data Pre-processing	08
5.	Concept Description & Association Rule Mining	08
6.	Classification and Prediction	06
7.	Clustering	04
8.	Advance topics	05

Total Hours (Theory): 45

Total Hours (Lab): 30

Total Hours: 75

C. Detailed Syllabus:

1. Overview and Concepts Data Warehousing 06 Hours 14 %
 - 1.1 What is data warehousing - The building Blocks
 - 1.2 Defining Features – Data warehouses and data marts

1.3	Overview of the components		
1.4	Metadata in the data warehouse		
1.5	Need for data warehousing		
1.6	Basic elements of data warehousing		
1.7	Trends in data warehousing.		
2.	OLAP in Data Warehouse	04 Hours	10 %
2.1	OLAP (Online analytical processing) definitions		
2.2	Difference between OLAP and OLTP		
2.3	Dimensional analysis - What are cubes?		
2.4	Drill-down and roll-up - slice and dice or rotation		
2.5	OLAP models		
2.6	ROLAP versus MOLAP		
2.7	defining schemas: Stars, snowflakes and fact constellations		
3.	Introduction to data mining (DM)	04 Hours	10 %
3.1	DM Functionalities		
3.2	Classification of DM Systems		
3.3	Issues in DM – KDD Process		
4.	Data Pre-processing	08 Hours	18 %
4.1	Why to pre-process data?		
4.2	Data cleaning: Missing Values, Noisy Data		
4.3	Data Integration and transformation		
4.4	Data Reduction: Data cube aggregation, Dimensionality reduction		
4.5	Data Compression		
4.6	Numerosity Reduction		
4.7	Data Mining Primitives		
4.8	Languages and System Architectures: Task relevant data		
4.9	Kind of Knowledge to be mined		
4.10	Discretization and Concept Hierarchy.		
5.	Concept Description and Association Rule Mining	08 Hours	18 %
5.1	What is concept description?		
5.2	Data Generalization and summarization-based characterization, Attribute Oriented Induction for Characterization &		

Discrimination		
5.3 Attribute relevance		
5.4 class comparisons		
5.4 Association Rule Mining		
5.4 Market basket analysis- basic concepts		
5.5 Finding frequent item sets: Apriori algorithm - generating rules		
6. Classification and Prediction	06 Hours	14 %
6.1 What is classification and prediction?		
6.2 Issues, Classification using Decision trees		
6.3 Performance Evaluation Criteria, Accuracy & Error Measures		
6.4 Linear and nonlinear regression		
6.5 Introduction of tools such as DBMiner /WEKA DM Tools		
7. Clustering	04 Hours	10 %
7.1 Introduction to Clustering		
7.2 Types of Data in Cluster Analysis		
7.3 Major Clustering Methods: Partitioning Methods, Hierarchical Methods, Density Based Methods, Grid Based Methods, Outlier Analysis		
8. Advance topics	05 Hours	06 %
8.1 Text mining, Spatial mining		
8.2 Web mining		
8.3 Social Networking Analysis		
8.4 Big Data: Introduction to Big Data and Hadoop, What is Big Data and what are the challenges to process Big Data, What technologies supports Big Data, Hadoop introduction and its history, Hadoop vs RDBMS, Introduction of Hadoop Distributed File System, Introduction of Map-Reduce		

D. Instructional Method and Pedagogy:

- At the start of course, the course delivery pattern, prerequisite of the subject will be discussed.
- Lectures will be conducted with the aid of multi-media projector, black board, OHP etc.
- Attendance is compulsory in lectures and laboratory which carries 5 Marks weightage.

- Two internal exams will be conducted and average of the same will be converted to equivalent of 15 Marks as a part of internal theory evaluation.
- Assignments based on course content will be given to the students at the end of each unit/topic and will be evaluated at regular interval. It carries a weightage of 5 Marks as a part of internal theory evaluation.
- Surprise tests/Quizzes/Seminar will be conducted which carries 5 Marks as a part of internal theory evaluation.
- The course includes a laboratory, where students have an opportunity to build an appreciation for the concepts being taught in lectures.
- Experiments/Tutorials related to course content will be carried out in the laboratory.

E. Student Learning Outcome:

Upon completion of this course, students will be able to do the following:

- Students will be able to understand importance of data mining and its various concepts like data preprocessing, various classification algorithms etc.
- Students will be able to use knowledge of the subject in higher semester for subjects like Advance Data Warehousing and Data Mining concepts.
- Student will be able to develop a reasonably sophisticated data mining application

F. Recommended Study Material:

❖ Text Books:

1. “Data Mining Concepts and Techniques”, J. Han, M. Kamber, Morgan Kaufmann
2. “Data Warehousing Fundamentals”, Paulraj Ponnian, John Wiley.

❖ Reference Books:

1. “Data mining: Concepts, models, methods and algorithms, M. Kantardzic, John Wiley & Sons Inc.
2. “Data Mining: Introductory and Advanced Topics”, M. Dunham, Pearson Education.
3. “Data Mining”, Pieter Adriaans, Dolf Zantinge, Pearson Education Asia

❖ Web Materials:

1. <http://www.dataminingblog.com>
2. <http://www.kdnugget.com>