1	2	3	4	5	6	7	8	9	10	11	12
H	H	H	H								

Syllabus for B. Tech. III Year I semester
Computer Science and Engineering
DATA WAREHOUSING AND DATA MINING

Code: 7EC04	**L**	**T**	**P**	**C**
	3	**1**	**0**	**4**

Prerequisite : Data structures, Database Management Systems

Course Objectives: Principles of Data Mining and Pre-processing techniques. Architecture of a Data Warehouse and OLAP operations. Concepts on Attribute Relevance Analysis and Data Mining Query Language. Implementation of Apriori and FP growth Algorithms Implementation of Classification Algorithms of Naïve Baye's & ID3 Decision Tree etc. Various categories of Clustering Algorithms

Course Outcomes: At the end of this course the student will be able to

1. Fundamentals of Data Mining and various data preprocessing techniques. and the Data Mining Query language primitives.
2. Identify the schemas used in designing Architecture of Data warehouse and OLAP operations.
3. Learn the significance and methods used for Characterization and the Analysis of Attribute Relevance..
4. Applications of Apriori and FP Growth algorithms for mining Association rules in large databases.
5. Applications of various classification models like Naïve Baye's & ID3 Decision Tree along with the prediction of the new samples.
6. Applications of clustering techniques available for numerous applications. Identify the optimal clustering technique for a particular application

UNIT – I: Introduction: Fundamentals of data mining, Data Mining Functionalities, Classification of Data Mining systems, Major issues in Data Mining.
Data Preprocessing: Needs Preprocessing the Data, Data Cleaning, Data Integration and Transformation, Data Reduction, Discretization and Concept Hierarchy Generation. Applications.

UNIT – II: Data Warehouse and OLAP Technology for Data Mining Data Warehouse, Multidimensional Data Model, Data Warehouse Architecture, Data Warehouse Implementation

UNIT – III: Concepts Description: Characterization and Comparison: Data Mining Primitives, Data Mining Query Languages, Architectures of Data Mining Systems. Data Generalization and Summarization-Based Characterization, Analytical Characterization: Analysis of Attribute Relevance, Mining Descriptive Statistical Measures in Large Databases.

UNIT – IV: Mining Association Rules in Large Databases: Association Rule Mining, Mining Single-Dimensional Boolean Association Rules from Transactional Databases, Mining Multilevel Association Rules from Transaction Databases.

UNIT – V: Classification and Prediction: Issues Regarding Classification and Prediction, Classification by Decision Tree Induction, Bayesian Classification, Classification by Back propagation, Classification Based on Concepts from Association Rule Mining, k-nearest neighbor classifier, Prediction, Classifier Accuracy.

UNIT – VI: Cluster Analysis Introduction: Introduction to machine learning, Types of Data in Cluster Analysis, A Categorization of Major Clustering Methods, Partitioning Methods, Density-Based Methods, Grid-Based Methods, Model-Based Clustering Methods, Outlier Analysis.

TEXT BOOK:

1. Data Mining – Concepts and Techniques - Jiawei Han & Micheline Kamber Harcourt India.

REFERENCES:

1. Data Mining Introductory and advanced topics –Margaret H Dunham, Pearson Education
2. Data Mining Techniques – Arun K Pujari, University Press.
3. Data Warehousing in the Real World – Sam Anahory & Dennis Murray, Pearson Education Asia.
4. Data Warehousing Fundamentals – Paulraj Ponnaiah Wiley Student Edition.
5. The Data Warehouse Life cycle Tool kit – Ralph Kimball Wiley Student Edition
6. Introduction to Data Mining - First Edition, by Pang-Ning Tan, Michael Steinbach and Vipin Kumar, ISBN-13: 978-0321321367.