

SREENIVASA INSTITUTE of TECHNOLOGY and MANAGEMENT STUDIES

II MCA - II Semester

L	P	C
4	0	4

16MCA221

DATA WAREHOUSING & DATA MINING

Course Objectives:

- To understand and implement classical models and algorithms in data warehousing and data mining.
- To analyze the data, identify the problems, and choose the relevant models and algorithms to apply.
- To familiarize the student with the concepts of data warehouse and data mining,
- To make the student acquaint with the tools and techniques used for Knowledge Discovery in Databases, and other data repositories.
- To evaluate different models used for OLAP and data pre-processing.
- To Categorize and carefully differentiate between situations for applying different data mining techniques: mining frequent pattern, association, correlation, classification, prediction, and cluster analysis.
- To design and implement systems for data mining.
- To evaluate the performance of different data mining algorithms.
- To propose data mining solutions for different applications.
- To assess the strengths and weaknesses of various methods and algorithms and to analyze their behavior.

Syllabus:

UNIT I : Introduction and Data Pre-Processing

Motivation and Importance- What is Data Mining- Data Mining on What Kind of data Data Mining Functionalities-Classification of Data Mining Systems- Data Mining Task Primitives Major Issues in Data Mining.

Why Preprocess the data-Data Cleaning-Data Integration- Data Transformation- Data Reduction.

UNIT II : Data Warehouse and OLAP Technology

What is a Data Warehouse-Multi dimensional data model- schemas for multi dimensional database- OLAP operations in the Multi dimensional data model-Data Warehouse Architecture-Data Warehouse Implementation-From Data Warehousing to Data Mining.

UNIT III : Mining Frequent Patterns , Associations and Correlations

Basic Concepts and a Road Map, Efficient and Scalable Frequent Itemset Mining Methods, The Apriori Algorithm: Finding Frequent Itemsets Using Candidate Generation, Generating Association Rules from Frequent Itemsets, Mining Frequent Itemset without Candidate Generation, Mining various Kinds of Association Rules, From Association Mining to Correlation Analysis, Constraint-Based Association Mining.

UNIT IV : Classification & Prediction

What is Classification, What is Prediction, Classification by Decision Tree Induction, Bayesian Classification, Rule-Based Classification, Classification by Backpropagation, Support Vector Machines, Lazy Learners, Prediction, Linear Regression and non-Linear Regression, Accuracy and Error Measures.

UNIT V : Cluster Analysis, Advanced Techniques and Applications

What is Cluster Analysis, Types of Data in Cluster Analysis, Partitioning Methods, K-Means Method, Hierarchical Methods, Agglomerative & Divisive Hierarchical Clustering, Density Based Methods, DBSCAN: A Density – Based Clustering Method Based on Connected Regions with sufficiently High Density, Grid-Based Methods, STING: Statistical Information Grid.

Mining Time- Series Data, Spatial Data Mining, Multimedia Data Mining, Text Mining, Data Mining Applications.

Course Outcome:

- Define what knowledge discovery and data mining.
- Define the concept, structure of data warehousing.
- Apply multi-dimensional modeling techniques in designing data warehouses.
- Apply the online analytical processing (OLAP) technology for decision support.
- Apply data preprocessing and data quality techniques.
- Design of data mart or data warehouse for any organization.
- Apply data cubing techniques using knowledge discovery in data warehouses.
- Extract knowledge using data mining techniques.
- Adapt to new data mining tools.
- Explore recent trends in data mining such as web mining, spatial-temporal mining.

TEXT BOOKS :

1. Data Mining: Concepts and Techniques, 2/e, 2009, Jiawei Han and Micheline Kamber, Morgan Kaufmann Publishers, New Delhi, India.
2. Introduction to Data Mining, 2006, Pang - Ning Tan, Micael Steinbach and Vipin Kumar, Pearson education, New Delhi, India.

REFERENCE BOOKS :

1. Data Mining: Introductory and Advanced topics, 2008, Margaret H. Dunham, Pearson Education, New Delhi, India.
2. Building the Data Warehouse, 4/e, 2008, W.H.Inmon , Wiley-India, Fourth Edition, New Delhi, India.
3. Insight into Data Mining Theory and Practice, 2010, K.P. Soman, Shyam Diwaker, V. Ajay, PHI Learning Private Limited, New Delhi, India.
4. Data Mining Techniques, 2007, Arun K. Pujari, Tata Mc- Graw Hill, New Delhi, India.
5. Data Mining A tutorial Based Primer, 2003, Richard J. Roiger and Michael W. Geatz, Pearson Education, Singapore.

Dr. S .Jyothi

Professor, Dept. of Computer Science,
Sri Padmavathi Mahila University,
Tirupathi

University BOS Member

Dr. N. Ch. S. N. Iyengar

Sr. Professor,
School of SCSE,
VIT University,
Vellore, T.N.

Academic Expert member

Dr. A. Rama Mohan Reddy

Professor,
Dept. of Computer Science & Engineering,
S.V. University,
Tirupathi.

Academic Expert member

Mr. D. Babu Rao,

Director of Engineering,
Aricent Technologies,
Bangalore

Industrial Expert member