MSCS205 DATA WAREHOUSING AND DATA MINING (3-0-0)

Module I: 9 Hours

Introduction to Data Warehousing and Data Mining: Data Warehouse Defined, Features of a Data Warehouse, Data Granularity, The Information Flow Mechanism, Metadata, Two Classes of Data, The Lifecycle of Data, Data Flow from Warehouse to Operational Systems, Failures of Past Decision-Support Systems, Operational Versus Decision-Support Systems, Data Warehouse v/s Data Mining, Data Mining Process, Data Mining Functionalities, Data Pre-processing – Descriptive Data Summarization, Data Cleaning, Integration and Transformation, Reduction

Module II: 9 Hours

The Building Blocks of a Data Warehouse and Data Warehouse Schema: Data Warehouse Architecture Goals, Data Warehouse Architecture, Data Warehouse and Data Mart, Issues in Building Data Marts, Building Data Marts, Other Data Mart Issues, Overview of the Components, Data Warehouse Schema: The Star Schema, The Snowflake Schema, Aggregate Tables, Fact Constellation Schema or Families of Star, Keys in the Data Warehouse Schema

Module III: 9 Hours

Data Warehouse Modelling and Online Analytical Processing: Building the Fact Tables and Dimension Tables, Characteristics of a Dimension Table, Characteristics of a Fact Table, The Factless Fact Table, Updates To Dimension Tables, Cyclicity of Data - Wrinkle of Time, Dimensional Modeling, Strengths of Dimensional Modeling,Data Warehouse and the Data Model, Enhancing the Data Warehouse Performance Data Warehouse Design,

Usage and Implementation: Data Warehouse Design Process, Data Warehouse Usage for Information Processing

Module IV: 9 Hours

Efficient Data Cube Computation, Data Cube and OLAP, Typical OLAP Operations, From Online Analytical Processing to Multidimensional Data Mining,

Data Mining Techniques:A Statistical Perspective on Data Mining, Classification, Issues in Classification, Statistical-Based Algorithms, Distance-Based Algorithms, Decision Tree-Based Algorithms, Prediction – Prediction techniques, Linear and Non-Linear Regression. Clustering: Applications of clustering, Categorization of Major Clustering Methods: Partitioning Methods, Hierarchical Methods, Density-Based Methods, Grid-Based Methods, Outlier Detection.

Text Books:

1. Jiawei Han and Micheline Kamber, "Data Mining Concepts & Techniques", 3rd Edition, Elsevier Pub, 2021.

2. M. H. Dunham. Data Mining, "Introductory and Advanced Topics", 1st Edition, Pearson Education Publisher, 2006

Reference Books:

1. Reema Thareja, "Data Warehousing",2nd Edition, Oxford University Press, 2009.

2. Paulraj Ponniah, "Data Warehousing Fundamentals", 2nd Edition, John Wiley & Sons, Inc, 2010.