|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| |  |  |  | | --- | --- | --- | | 13CS4101 | - | DATA WAREHOUSING AND DATA MINING | | | | | | | | |
|  |  | |  | | | | |
| Hours / Week | : | 4 | |  | Sessional Marks | : | 40 |
| Credits | : | 4 | |  | End Examination Marks | : | 60 |

|  |
| --- |
| **UNIT – I** |
| **Introduction to Data Warehousing**: Introduction: Data Warehouse, Data Warehousing, Framework of the Data Warehouse, Data Warehouse Options, Developing Data Warehouses, The Business Driven Approach, The DWRM Technique, Requirements Management Control, The Data Warehouse Development Life Cycle, Data Warehouse Development Methodologies. |
|  |
| **UNIT – II** |
| **Data Warehouse Design & Modeling**: Defining Dimensional Model, Granularity of Facts, Additives of Facts, Functional Dependency of the Data, Slowly changing dimensions types, implementing rapidly changing Dimensions, Multi-use Dimensions, Designing: Identifying the source, Data Warehouse Architecture (ETL process) |
|  |
| **UNIT – III** |
| **Introduction to Data Mining**: Data Mining, Architecture of Data Mining, Data Mining Functionalities, Interestingness of a pattern, Classification of Data Mining Systems, Major issues in Data Mining. Data Mining Task Primitives.  **Data Pre-processing**: Data Cleaning, Data Integration and Transformation, Data Reduction, Discretization and Concept Hierarchy Generation. |
|  |
| **UNIT – IV** |
| **Concept Description**: Characterization and Discrimination: Attribute-Oriented Induction for Data Characterization, mining class discriminations, presentation of both characterization and discrimination.  **Mining Frequent Patterns, Associations and Correlations**: Basic concepts and a Road Map, the Apriori Algorithm, Generating Association Rules, improving the efficiency of Apriori, Mining Frequent Item sets without Candidate Generation, From Association Analysis to Correlation Analysis. |
|  |
| **UNIT – V** |
| **Classification and prediction**: Issues regarding Classification and prediction, Decision Tree induction, Bayes’ theorem, Naive Bayesian classification, Linear Regression, Nonlinear Regression, Other Regression-Based Methods.  **Cluster Analysis**: Types of Data in cluster analysis, A categorization of Major Clustering Methods, Partitioning Methods.  **Outlier Analysis**: Distance-Based Outlier Detection, Density-Based Local Outlier Detection. |
|  |
|  |
| TEXT BOOKS |
| 1. Data Mining Concepts and Techniques, Jiawei Han and Micheline Kamber , Morgan Kaufman Publications. 2. Data Warehousing Design, Development and Best Practices, Soumendra Mohanty, TMH. |
|  |
| REFERENCE BOOKS |
| 1. Data Mining Introductory and Advanced Topics, Margaret H Dunhan, Pearson Education. 2. Data Mining, Ian H. Witten Eibe Frank, Morgan Kaufman Publications. 3. The Data Warehouse Life Cycle Toolkit, Ralph Kimball, WILEY Computer publishing. |