# 19CS3203 - DATA WAREHOUSING AND MINING

|  |  |  |  |
| --- | --- | --- | --- |
| **Course Category:** | Program Core | **Credits:** | 3 |
| **Course Type:** | Theory | **Lecture – Tutorial – Practical:** | 3-0-0 |
| **Prerequisite:** | Require knowledge on Database Management System concepts. | **Sessional Evaluation:**  **Univ. Exam Evaluation:**  **Total Marks:** | 40  60  100 |
| **Objectives** | * Learn the basic concepts of data warehousing and data preprocessing concepts * Explore the data mining and data classification, prediction and clustering techniques for various applications. | | |

|  |  |  |
| --- | --- | --- |
| **Course Outcomes** | Upon the successful completion of the course, the students will be able to: | |
| CO1 | Know the importance of Data warehousing and Development methodologies |
| CO2 | Explore on Data warehousing design and modeling concepts for application development |
| CO3 | Learn the Implementation of ETL Process. |
| CO4 | Understand data mining and data-preprocessing techniques |
| CO5 | Identify the importance of Mining Frequent Patterns. to study various regression methods |
| CO6 | Study the classification and prediction, Cluster analysis and supporting major categories including outer analysis concepts to explore on various applications |
| **Course Content** | UNIT – I  **Introduction to Data Warehousing**: Introduction, Data Warehouse, Data Warehousing, Framework of the Data Warehouse, Data Warehouse Options, Developing Data Warehouses, The Data Warehouse Development Life Cycle, Data Warehouse Development Methodologies.  UNIT – II  **Data Warehouse Design & Modeling**: Defining Dimensional Model, Granularity of Facts, Additivity 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  **Implementation of ETL Process:** Extract Processing, Interface Processing, Load into Staging, Transform and Prepare for Load, Load Process, Post Load Processing, Generic warehouse Population function.  UNIT – IV  **Introduction to Data Mining**: Data Mining, Architecture of Data Mining, Data Mining Functionalities, Classification of Data Mining Systems, Data Mining Task Primitives, Major issues in Data Mining.  **Data Pre-processing**: Data Cleaning, Data Integration and Transformation, Data Reduction, Discretization and Concept Hierarchy Generation.  UNIT – V  **Attribute-Oriented Induction**: 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.  UNIT – VI  **Classification and Prediction**: Issues regarding Classification and prediction, Decision Tree induction, Bayes’ theorem, Naive Bayesian classification, Regression Methods.  **Cluster Analysis**: Types of Data in clustering, Classification of Major Clustering Methods.  **Outlier Analysis**: Types of Outliers, Classification of Major Outlier Analysis. | |
| **Text Books and References** | **TEXT BOOKS:**   1. Data Warehousing Design, Development and Best Practices, Soumendra Mohanty, TMH. 2. Data Mining Concepts and Techniques, Jiawei Han and Micheline Kamber, Morgan Kaufman Publications, 2nd Edition.   **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. | |
| **E-Resources** | 1. <https://nptel.ac.in/courses> 2. <https://freevideolectures.com/university/iitm> | |