# 20CS3101 - DATA WAREHOUSING AND MINING

|  |  |  |  |
| --- | --- | --- | --- |
| Course Category: | Professional 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 successful completion of the course, the students will be able to: | |
| CO1 | Understand the basic concepts of data warehouse |
| CO2 | Understand the basic concepts of ETL Process |
| CO3 | Apply pre-processing techniques for data cleaning |
| CO4 | Analyze and evaluate performance of algorithms for Association Rules |
| CO5 | Analyze Classification and Bayes Theorems |
| CO6 | Analyze Classification and Clustering algorithms |
| Course Content | UNIT-I  **Introduction to Data Warehousing**: Introduction, Data Warehouse, Data Warehousing, Difference between OLAP&OLTP.  **Kimball’s DW/BI Architecture:** Operational Source Systems, Extract, Transformation and Load Systems, Presentation area to support Business intelligence, Business intelligence Applications.  **Alternative DW/BI Architectures:** Independent Data Mart Architecture.  UNIT-II  **Kimball’s Dimensional Modelling Techniques Overview:** Basic Fact Table Techniques, Basic Dimensional Table Techniques, Dealing with Slowly Changing Dimension Attributes.  UNIT-III  **Introduction to Data Mining**: Data Mining, Architecture of Data Mining, Data Mining Functionalities, Classification of Data Mining Systems, Major issues in Data Mining.  **Data Pre-processing**: Data Cleaning, Data Integration and Transformation, Data Reduction, Discretization and Concept Hierarchy Generation.  UNIT-IV  **Attribute-Oriented Induction:** Attribute-Oriented Induction for Data Characterization.  **Mining Class Comparisons:** Discriminating between Different Classes.  **Mining Frequent Patterns, Associations and Correlations:** Basic Concepts, Efficient and Scalable Frequent Item set Mining Methods, Mining various kinds of Association Rules, From Association Mining to Correlation Analysis, Constraint-Based Association Mining.  UNIT-V  **Classification and Prediction:** Issues Regarding Classification and Prediction, Classification by Decision Tree Induction.  **Bayesian Classification Methods:** Bayes Theorem, Naive Bayesian Classification, Rule-Based Classification, Support Vector Machines, Lazy Learners.  UNIT-VI  **Cluster Analysis Introduction:** Types of Data in Cluster Analysis, A Categorization of Major Clustering Methods.  **Partitioning Methods:** Classical Partitioning Methods: k-Means and k-Medio’s, Hierarchical Methods, Density-Based Methods, Grid-Based Methods. | |
| Text Books &  Reference  Books | **TEXT BOOKS:**   1. The Data Warehouse Toolkit, Ralph Kimball Margy Ross, Third Edition. 2. Data Mining: Concepts and Techniques, Jiawei Han and Micheline Kamber, Morgan Kaufmann Publishers, Elsevier, Third Edition, 2012.   **REFERENCE BOOKS:**   1. Introduction to Data Mining – Pang-Ning Tan, Michael Steinbach and Vipin Kumar, Pearson Education. 2. Data Warehousing in the Real World, Sam Aanhory & Dennis Murray Pearson | |
| E-Resources | 1. <https://aatinegar.com/wp-content/uploads/2016/05/Kimball_The-Data-Warehouse-Toolkit-3rd-Edition.pdf> 2. <http://myweb.sabanciuniv.edu/rdehkharghani/files/2016/02/The-Morgan-Kaufmann-Series-in-Data-Management-Systems-Jiawei-Han-Micheline-Kamber-Jian-Pei-Data-Mining.-Concepts-and-Techniques-3rd-Edition-Morgan-Kaufmann-2011.pdf> 3. <https://freevideolectures.com/university/iitm> | |