**Department of Computer Engineering**

**(Academic Year :2023-2024)**

|  |  |
| --- | --- |
| **Course Code: CSC504** | |
| **Course Name: Data Warehousing and Mining** | |
| **Course Teacher: Dr. Sujata P. Deshmukh** | |
| **Course Outcomes (CO):** ***At the End of the course students will be able to*** | |
| **CSC504.1** | Explain the need of Data warehouse and Mining Principles |
| **CSC504.2** | Design a Data-warehouse Using Dimension Modelling and apply OLAP operations |
| **CSC504.3** | Select appropriate Data Pre-processing Technique and apply appropriate Data Mining Algorithm to the given Real World Problem. |
| **CSC504.4** | Compare and Evaluate the different Data Mining techniques like Regression, classification, Clustering and Association Rule Mining. |
| **CSC504.5** | Explain the concepts of Web Mining |
|  |  |

**Course Lesson Plan**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Sr. No.** | **Proposed Date** | **Actual Date** | **Topics** | **CO** | **Teacher’s Remark** | **HoD’s Remark** |
| 1 | 12-07-23 |  | Prerequisite,  CO Discussion  **Moving Distributed DBMS to DWM-**The Need for Data Warehousing; Increasing Demand for Strategic Information; Inability of Past Decision Support System; Operational V/s Decisional Support System; | CO1 |  |  |
| 2 | 13-07-23 |  | Introduction to Data Warehouse, Data warehouse architecture, Data warehouse versus Data Marts, Different approaches and layered Architecture | CO1 |  |  |
| 3 | 14-07-23 |  | E-R Modeling versus Dimensional Modeling, Information Package Diagram, Data Warehouse Schemas; Star Schema, Snowflake Schema, Factless Fact Table, Fact Constellation Schema. Update to the dimension tables. | CO2 |  |  |
| 4 | 19-07-23 |  | OLTP versus OLAP, OLAP operations: Slice, Dice, Rollup, Drilldown and Pivot. OLAP Models: MOLAP, ROLAP, HOLAP, DOLAP, Definition of Schema using DMQL, Examples on OLAP | CO2 |  |  |
| 5 | 20-07-23 |  | Data Warehouse Modeling Vs Operational Database Modeling; Dimensional Model Vs ER Model; Features of a Good Dimensional Model; The Star Schema; How Does a Query Execute? |  |  |  |
| 6 | 21-07-23 |  | Keys in DW, Snowflake Schema, Fact Constellation Schema or Families of Star. | CO2 |  |  |
| 7 | 24-07-23 |  | The Factless Fact, Table; Aggregate Table, Updates To Dimension Tables: Slowly Changing Dimensions, Type 1 Changes, Type 2 Changes, Type 3 Changes, Large Dimension Tables, Rapidly Changing or Large Slowly Changing Dimensions, Junk Dimensions | CO2 |  |  |
| 8 | 26-07-23 |  | Major steps in ETL process, Challenges in ETL Functions; Data Extraction; Identification of Data Sources, Issues in Data Cleansing, Extracting Data: Immediate Data Extraction, Deferred Data Extraction  Data Transformation: Tasks Involved in Data Transformation | CO2 |  |  |
| 9 | 28-07-23 |  | Problems based on DWH design and OLAP operation | CO2 | Assignment 1 |  |
| 10 | 31-07-23 |  | What is Data Mining; Knowledge Discovery in Database (KDD), What can be Data to be Mined, Related Concept to Data Mining, Data Mining Task Primitives, Architecture, KDD process, | CO1 |  |  |
| 11 | 01-08-23 |  | Issues in Data Mining, Applications of Data Mining, Comparison of DM with ML and BDA | CO3 |  |  |
| 12 | 04-08-23 |  | Data Exploration: Types of Attributes, Statistical Description of Data, | CO3 |  |  |
| 13 | 07-08-23 |  | Why Pre-processing? Data Cleaning; Data Integration; Data Reduction: Attribute subset selection, Histograms, Clustering and Sampling | CO3 |  |  |
| 14 | 08-08-23 |  | Guest Lecture-Role of Analytics from placement perspective | CO3,CO4 |  |  |
| 15 | 11-08-23 |  | Data Transformation & Data Discretization: Normalization | CO3,CO4 |  |  |
| 16 | 14-08-23 |  | Data Discretization and Concept hierarchy generation-. Binning, Histogram Analysis and Concept hierarchy generation and description | CO3 | Assignment 2 |  |
|  | 15-08-23 |  | **Independence Day** |  |  |  |
|  | 16-08-23 |  | **Parsi New Year** |  |  |  |
| 17 | 18-08-23 |  | **Innovative Teaching Learning Activity** and problem based on data cleaning and transformation. | CO3,CO4 | Assignment 1 test and discussion |  |
| 18 | 19-08-23(online-Saturday) |  | Basic Concepts, Decision Tree Induction: Attribute Selection Measures, Tree pruning.Naïve Bayesian Classification and problems, Decision Tree Induction and problems Naïve Bayesian Classification, Accuracy and Error measures, | CO4 |  |  |
| 19 | 21-08-23 |  | Model Evaluation & Selection: Accuracy and Error measures, Holdout, Random Sampling, Cross Validation, Bootstrap. | CO3 |  |  |
| 20 | 22-08-23 |  | Problems based on classification | CO3 |  |  |
| 21 | 25-08-23 |  | What is clustering? Types of data, Partitioning Methods (K-Means)Partitioning Methods (KMedoids) | CO3 |  |  |
|  | 28-08-23 to  30-08-23 |  | **Unit Test-1** |  |  |  |
| 22 | 01-09-23 |  | Hierarchical Methods(Agglomerative , Divisive, BRICH), | CO3 |  |  |
|  | 02-09-23(Saturday –online) |  | Remedial Session |  |  |  |
| 23 | 04-09-23 |  | Hierarchical Methods(Agglomerative , Divisive, BRICH), | CO3 |  |  |
| 24 | 05-09-23 |  | Problems based on clustering | CO4 |  |  |
| 25 | 08-09-23 |  | Comparison of clustering with classification and more problems | CO4 | Assignment 3 |  |
| 26 | 11-09-23 |  | Market Basket Analysis, Frequent Item sets, Closed Item sets, and Association Rule, Frequent Pattern Mining, |  |  |  |
| 27 | 12-09-23 |  | The Apriori Algorithm for finding Frequent Itemsets Using Candidate Generation | CO4 | Cross Word |  |
| 28 | 15-09-23 |  | Generating Association Rules from Frequent Itemsets, Improving the Efficiency of Apriori, | CO4 |  |  |
| 29 | 18-09-23 |  | A pattern growth approach for mining Frequent Itemsets; Mining Frequent itemsets using vertical data formats; Mining closed and maximal patterns; | CO4 |  |  |
|  | 19-09-23 to 22-09-23 |  | Shri Ganesh Festival |  |  |  |
| 30 | 25-09-23 |  | Problems based on FP tree, Vertical DB, APriori algo for practise | CO4 |  |  |
| 31 | 26-09-23 |  | From Association Mining to Correlation Analysis, Pattern Evaluation Measures; | CO4 |  | 28-09 Holiday |
|  | 28-09-23 |  | Anant Chaturdashi |  |  |  |
| 32 | 29-09-23 |  | Introduction to Mining Multilevel Association Rules and Multidimensional Association Rules; | CO4 |  |  |
|  | 02-10-23 |  | Gandhi Jayanti |  |  |  |
| 33 | 03-10-23 |  | Introduction to web content Mining | CO5 |  |  |
| 34 | 06-10-23 |  | Crawlers ,Harvest System, Virtual web View | CO5 |  |  |
|  | 09-10-23 to 11-10-23 |  | Unit Test-2 |  |  |  |
| 35 | 13-10-23 |  | Personalisation | CO5 |  |  |
| 36 | 13-10-23 |  | Web Structure Mining, Page Rank | CO5 |  |  |
| 37 | 16-10-23 |  | Web usage Mining | CO5 | Assignment4 |  |
| 38 | 17-09-23 |  | Remedial Session |  |  |  |
| 39 | 20-10-23 |  | Remedial Session /Guest Lecture- Power BI/ Tableau |  |  |  |
| 40 | 21-10-23 |  | Remedial Session |  |  |  |
|  |  |  | **University ESE Examination** |  |  |  |

**TEXT BOOKS**

1. PaulrajPonniah, ―Data Warehousing: Fundamentals for IT Professionals‖, Wiley India.

2. Han, Kamber, "Data Mining Concepts and Techniques", Morgan Kaufmann 3rd edition.

3. ReemaTheraja ―Data warehousing‖, Oxford University Press.

4. M.H. Dunham, "Data Mining Introductory and Advanced Topics", Pearson Education.

**Reference Books:**

1. Ian H. Witten, Eibe Frank and Mark A. Hall " Data Mining ", 3rd Edition Morgan kaufmann publisher.

2. Pang-Ning Tan, Michael Steinbach and Vipin Kumar, Introduction to Data Mining", Person Publisher.

3. R. Chattamvelli, "Data Mining Methods" 2nd Edition NarosaPublishing House.

**Course Instructor: Dr. Sujata Deshmukh**