|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Course Title** | **Data Warehouse and Data Mining** | | | | | |
| **Course Code** | **CSH313B-T&P** | | | | | |
| **Credits** | Lecture/Tutorial:3 ; Lab:2,Outcome:0; (TOTAL:4) | | | | | |
| **Course Coordinator /Instructor(s):** | Deepanshi Gupta/Anu Priya Sharma | | | | | |
| **Course type:** | Lecture/Tutorial/Lab/Outcome (3-0-2-0) | | | | | |
| **Core or Elective:** | Domain Elective | | | | | |
| **Term Offered:** | Even Semester (6th semester) | | | | | |
| **Course Schedule:** | **Lecture:**  3 hrs per week. | |  | | **Lab:** 2 hrs/week Hands-on Session. |  |
| **Course Assessment:** | **Formal Assessment:**  **Theory, Practical :**2 mid-term and 1 end semester test | | | **Informal Assessment**  **Theory:** Assignments, Class tests and problem solving in lectures,  **Practical:** Viva, Lab Participation, | | |
| **Relationship to other courses** | **Course Prerequisites** | Database Management Systems | | | | |
| **Assumed Knowledge** | The student will be able to design and represent multi-dimensional models and obtain business intelligence from them. | | | | |
| **Following Courses** | Big Data | | | | |
| **Objective** | **Student (A) would be able to effectively apply Data Warehouse and Data Mining skills (B) using Data Mining tools and techniques (C) to solve real-world problems (D).** | | | | | |
| **Learning Outcomes (COs)** | By the end of this course, you should be able to:  CSH313B.1: Acquiring the basic knowledge of data warehouse, data mining and Data Decision.  CSH313B.2: Illustrating the types of data warehouses.  CSH313B.3: Appreciate the strengths and limitations of various data mining and data warehousing models.  CSH313B.4: Describe and analyze different methodologies used in data mining and data ware housing.  CSH313B.5: Validating techniques to solve real world problems. | | | | | |

**Syllabus**

|  |  |
| --- | --- |
| **Course Title/ Code** | **Data Warehousing and Data Mining(CSH313B) T & P** |
| **Course Type:** | ELECTIVE (Departmental) |
| **Course Nature:** | HARD |
| **L-T-P-O Structure** | (3-0-2-0) |
| **Objectives** | The student will be able to design and represent multi-dimensional models and obtain business intelligence from them. |

|  |  |  |
| --- | --- | --- |
| **Syllabus** | **Sections** | **Weightage** |
| A | 25% |
| B | 25% |
| C | 25% |
| D | 25% |
| **TOTAL** | **100%** |

# Section-A

**Introduction to data ware house and decision making:** Need for data warehousing, Escalating Need for strategic information, Decision making, failures of past decision-support systems, operational versus decision-support systems, data warehousing – the only viable solution. Applications of Data ware house**:** Operational System and Business Intelligence.

**Data ware house**: Data warehouse definition, types of data warehouses and data marts, types of data marts, ETL process, 3 – Tier data warehouse architecture, Meta data, role of meta data repository in data warehouse, .

**Section-B**

**Multidimensional Data Model:** Difference between Database System and Data Warehouse, Multidimensional data model, Facts, Dimensions, Measures, Data cubes, Schemas for Multidimensional Database (Stars, snowflakes and fact constellations) defining schemas. OLAP Technology: Starnet query model, Concepts Hierarchies, Partitioning strategies, OLAP operations: Slice, Dice. Roll up, Drill down, Pivot etc., Types of OLAP servers: ROLAP, MOLAP, HOLAP. Data warehouse implementation: Computation of data cubes, Partial Materialization, Indexing OLAP data, and Efficient Processing of OLAP queries. Tuning and testing of data warehouse.

**Section-C**

**Data Mining:**Data mining definition & task, KDD versus data mining, Data Mining Applications , Data preprocessing, Classification of data mining systems, Data mining task primitives, data mining techniques, Data mining query languages .Data mining techniques: Mining frequent Patterns, Association, and correlation.

# Section-D

**Classification & Prediction:** Decision tree knowledge discovery, Bayesian Classification, Neural Networks (MLP) &Support Victor Machines (SVM). Classification Accuracy and error measures, evaluating the accuracy of Classifier.

**Clustering Analysis:** K-means and K-Medoids, outlier analysis. Mining complex data object: (Spatial databases, Multimedia databases, Time series and Sequence data mining Text Databases and mining Word Wide Web).

**LIST OF EXPERIMENTS:**

1. To generate a data table in MS Access and perform various tasks on the data.
2. To Implement Pivot Table, Report Generation for the table data.
3. To study /implementation of various task on data using MS Excel.
4. To implement Pivot Chart , Pivot table of a given Data in table using MS Excel.
5. Introduction to data mining tool.
6. To implement Classification.
7. To Implement Association rule.
8. To implement Clustering.
9. To study PDI(Pentaho Data Integration) IDE.
10. To Create new repository in PDI and performing various tasks.

**Text Books:**

1. Data Warehousing In the Real World; Sam Anahory& Dennis Murray; 1997, PearsonNo.of copies in the Library.
2. Data Mining- Concepts & Techniques; Jiawei Han &MichelineKamber- 2001, Morgan Kaufmann.No.of copies in the Library.
3. Data Mining Techniques; ArunPujari; 2001, University Press; Hyderabad.No.of copies in the Library.

**Reference Book:**

1. Paul Raj Poonia, “Fundamentals of Data Warehousing”, John Wiley & Sons, 2003.No.of copies in the Library.
2. Sam Anahony, “Data Warehousing in the real world: A practical guide for building decision support systems”, John Wiley, 2004.
3. W. H. Inmon, “Building the operational data store”, 2nd Ed., John Wiley, 1999.
4. Mattison R., Web Warehousing and Knowledge Management, Tat McGraw-Hill .
5. Ponniah P., Data Warehousing, Wiley.

**LESSON PLAN**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Section** | **Topics** | **L** | **T** | **P** |
| **TOPIC** | **Section A: Introduction to data ware house and decision making:** | 15 | 0 | 4 |
|  | | | |
| **Section B: Multidimensional Data Model** | 12 | 0 | 3 |
|  | | | |
| **Section C: Data Mining** | 6 | 0 | 2 |
| **Sction D: Classification & Prediction and Cluster** | 9 | 0 | 3 |
|  | | | |
| **TOTAL** | **42** | **0** | **12** |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Week** | **Day** | **L** | **P** | **Course Outcome** | **Blooms Taxonomy** | **Program Outcome** |
| **I** | **Day1** | Need for data warehousing | **Lab 0: Implementation of basic functions using excel** | **CO1** | **B T1/BT2** | **P01,PO2,PO3,PO4,PO4,PO6** |
| **Day2** | Escalating Need for strategic information | **CO1** |
| **Day3** | Decision making | **CO1** |
| **II** | **Day1** | Failures of past decision-support systems | **Lab 1: Implementation of advance functions using excel** | **CO1** | **BT1/BT2/BT3** | **PO1,PO7,PO10,PO11** |
| **Day2** | Operational versus decision-support systems | **CO1** |
| **Day3** | Data warehousing – the only viable solution | **CO1** |
| **III** | **Day1** | Applications of Data ware house**:** Operational System and Business Intelligence | **Lab 2: Generation of Pivot table** | **CO1** | **BT2/BT3/BT4** | **PO1,PO7,PO4,PO5,PO6,PO9** |
| **Day2** | Data warehouse definition, types of data warehouses | **CO2** |
| **Day3** | Data marts, types of data marts. | **CO2** |
| **IV** | **Day1** | ETL process | **Lab 3: Extraction of data sets using various tools/sites** | **CO2** | **BT1/BT2/BT4** | **PO1, PO2,PO3, PO7, PO8,PO9** |
| **Day2** | 3 – Tier data warehouse architecture | **CO2** |
| **Day3** | Meta data, role of meta data repository in data warehouse. | **CO2** |
| **V** | **Day1** | Distributed and virtual data warehouses. | **Lab 4: Transform and loading of data in excel** | **CO2** | **BT1/BT2/BT3/BT4** | **PO1,PO2,PO3,PO4,PO5,PO6,PO7,PO8** |
| **Day2** | Difference between Database System and Data Warehouse | **CO2** |
| **CO2** |
| **Day3** | Revision |  |  |  |
|  |  |  |
| **VI** | **Day1** | Multidimensional data model | **Lab 5: PT1** | **CO3** | **BT1/BT4/BT5/BT6** | **PO1, PO6,PO7,PO8,PO9** |
| **Day2** | Facts, Dimensions, Measures, Data cubes | **CO3** |
| **CO3** |
| **Day3** | Schemas for Multidimensional Database (Stars) defining schemas | **CO3** |
| **CO3** |
| **VII** | **Day1** | Schemas for Multidimensional Database(snowflakes and fact constellations) | **Lab 6: Introduction to weka tool with various functionality** | **CO3** | **BT1/BT4/BT5/BT6** | **PO1,PO7, PO8,PO9,PO10** |
| **CO3** |
| **Day2** | OLAP Technology: Starnet query model | **CO3** |
| **CO3** |
| **Day 3** | Concepts Hierarchies, Partitioning strategies | **CO3** |
| **VIII** | **Day 1** | OLAP operations: Slice, Dice. Roll up, Drill down, Pivot etc. | **Lab 7: Preparation of dataset** | **CO3** | **BT1/BT4/BT5/BT6** | **PO1,PO7, PO8,PO9,PO10** |
| **Day 2** | Types of OLAP servers: ROLAP, MOLAP, HOLAP | **CO3** |
| **Day 3** | Data warehouse implementation: Computation of data cubes | **CO3** |
| **IX** | **Day 1** | Partial Materialization, Indexing OLAP data | **Lab 8: Implementation of Preprocessing** | **CO3** | **BT1/BT2/BT3** | **PO1,PO2,PO3,PO7,PO8** |
| **Day 2** | Efficient Processing of OLAP queries | **CO3** |
| **Day 3** | Tuning and testing of data warehouse | **CO3** |
| **X** | **Day 1** | Data mining definition & task, KDD versus data mining, Data Mining Applications | **Lab 9: PT2** | **CO4** | **BT1/BT4/BT5/BT6** | **PO1,PO7, PO8,PO9,PO10** |
| **Day 2** | Data preprocessing, Classification of data mining systems | **CO4** |
| **Day 3** | Data mining task primitives, data mining techniques | **CO4** |
| **XI** | **Day 1** | Data mining query languages | **Lab 10: Data visualization/feature selection using weka** | **CO4** | **BT1/BT4/BT5/BT6** | **PO1,PO7, PO8,PO9,PO10** |
| **Day 2** | Data mining techniques: Mining frequent Patterns | **CO4** |
| **Day 3** | Data mining techniques: Association, and correlation | **CO4,CO5** |
| **XII** | **Day 1** | Decision tree knowledge discovery | **Lab 11: Implementation of Association and co relation** | **CO4,CO5** | **BT1/BT4/BT5/BT6** | **PO1,PO7, PO8,PO9,PO10** |
| **Day 2** | Bayesian Classification | **CO4,CO5** |
| **Day 3** | Neural Networks (MLP) &Support Victor Machines (SVM) | **CO4,CO5** |
| **XIII** | **Day 1** | Classification Accuracy and error measures, evaluating the accuracy of Classifier | **Lab 12: Implementation of classification and regression** | **CO4,CO5** | **BT1/BT4/BT5/BT6** | **PO1,PO7, PO8,PO9,PO10** |
| **Day 2** | K-means and K-Medoids | **CO4,CO5** |
| **Day 3** | Outlier analysis | **CO4,CO5** |
| **XIV** | **Day 1** | Mining complex data object: (Spatial databases, Multimedia databases,) | **Lab 13: Implementation cluster using K-Means** | **CO4** | **BT1/BT2** | **PO1,PO7, PO8,PO9,PO10** |
| **Day 2** | Mining complex data object: (Time series and Sequence data mining Text Databases) | **CO4** |
| **Day 3** | Mining complex data object: (Mining Word Wide Web). | **CO4** |

**PROGRAM EDUCATIONAL OBJECTIVES:**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Data Warehouse and Data Mining** |  | | | | |
| **Preparation** | **Core Competence** | **Breadth** | **Professionalism** | **Learning Environment** |
| **√** | **√** | **√** | **√** | **√** |

**Teaching Methodologies:**

1. Lectures will be delivered in interactive mode.
2. Students will work individually and also in groups to solve different kinds of problems.
3. Students can work in groups in designing different circuits in labs.
4. Home assignments will help the learners in applying and designing of circuits.

**Evaluation scheme:**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Theory(Formal: Informal)** | |  | **Lab (Formal: Informal)** | | |
| **Evaluation criteria (formal)** | **Marks** |  | **Evaluation criteria** | **Marks** | |
| Test 1 (T1) | 30 |  | Test 1 (PT1) | 10 |
| Test 2 (T2) | 30 |  | Test 2 (PT2) | 10 | |
| PROJECT | 80 |  | Test 3 (PT3) | 20 | |
| **Evaluation criteria (Informal)** | **Marks** |  | **Evaluation criteria (Informal)** | **Marks** | |
| PRESENTATION1 | 20 |  | Lab REPORT | 10 | |
| PRESENTATION 2 | 20 |  |
| ASSIGNMENT | 20 |  |
| **Total** | **200** |  |  | **50** | |

**Text Books:**

1. Data Warehousing In the Real World; Sam Anahory& Dennis Murray; 1997, PearsonNo.of copies in the Library.
2. Data Mining- Concepts & Techniques; Jiawei Han &MichelineKamber- 2001, Morgan Kaufmann.No.of copies in the Library.
3. Data Mining Techniques; ArunPujari; 2001, University Press; Hyderabad.No.of copies in the Library.

**REFERECE BOOKS:**

1. Paul Raj Poonia, “Fundamentals of Data Warehousing”, John Wiley & Sons, 2003.No.of copies in the Library.
2. Sam Anahony, “Data Warehousing in the real world: A practical guide for building decision support systems”, John Wiley, 2004.
3. W. H. Inmon, “Building the operational data store”, 2nd Ed., John Wiley, 1999.
4. Mattison R., Web Warehousing and Knowledge Management, Tat McGraw-Hill .
5. Ponniah P., Data Warehousing, Wiley.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| |  |  |  |  |  |  |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | | **COURSE ASSESSMENT PLAN (CAP)** | | | | | | | | | | | | | | **FACULTY NAME: Gunjan** | | | **NAME OF COURSE COORDINATOR: Deepanshi Gupta** |  |  |  |  |  |  |  |  |  | | **COURSE NAME: Data Warehouseing and Data Mining** | **COURSE CODE: CSH313B-T&P** | **CREDITS: 4** | CSH313B.1: Acquiring the basic knowledge of data warehouse, data mining and Data Decision. CSH313B.2: Illustrating the types of data warehouses. CSH313B.3: Appreciate the strengths and limitations of various data mining and data warehousing models. CSH313B.4: Describe and analyze different methodologies used in data mining and data ware housing. CSH313B.5: Validating techniques to solve real world problems.   **ASSESSMENT METHOD** | | | | | | | | | | | **COURSE OUTCOME:** | **SECTION AS PER SYLLABUS** | **T1 (30)** | **T2 (30)** | **presentation 1(20)** | **presentation 2(20)** | **ASSIGNMENT(20)** | **project(80)** | **PT1 (10)** | **PT2(10)** | **LAB REPORT[10]** | **PT3(20)** | | | **CSH313B.1** | **A,B,C,D** | √ | √ | √ | √ | v | √ | √ | √ | √ | √ | | | **CSH313B.2** | **A** | √ |  |  |  | v |  | √ |  | √ |  | | | **CSH313B.3** | **B** |  | √ | √ |  |  |  |  | √ | √ |  | | | **CSH313B.4** | **C,D** |  |  |  |  |  | √ |  |  |  | √ | | | **CSH313B.5** | **D** |  |  |  | √ |  | √ |  |  |  | √ | | | **TOTAL WEIGHTAGE** | | 12 | 12 | 8 | 8 | 8 | 32 | **4** | **4** | **4** | **8** | | |

**Programme Specific Outcomes (PSOs)**

PSO1:    Attain the ability to design and develop computer programs and possess acquaintance with emerging technologies and open source platforms in the area of mobile app development, artificial intelligence, machine learning, web development, data analytics, cloud computing and networking to build effective computer-based systems

PSO2:  Acquire technical competency to deliver computer based innovative and effective solutions to tackle business and societal challenges, for pursuing successful career, entrepreneurship, research and higher studies.

**Bloom’s Taxonomy (BTs)**

**BT1- Knowledge:** Knowledge involves recognizing or remembering facts, terms, basic concepts, or answers without necessarily understanding what they mean.

**BT2-Comprehension:** Comprehension involves demonstrating understanding of facts and ideas by organizing, comparing, translating, interpreting, giving descriptions, and stating the main ideas.

**BT3-Applications:** Applying involves using acquired knowledge—solving problems in new situations by applying acquired knowledge, facts, techniques and rules. Learners should be able to use prior knowledge to solve problems, identify connections and relationships and how they apply in new situations.

**BT4-Analysis:** Analyzing involves examining and breaking information into component parts, determining how the parts relate to one another, identifying motives or causes, making inferences, and finding evidence to support generalizations.

**BT5-Synthesis:** Synthesizing involves building a structure or pattern from diverse elements; it also refers to the act of putting parts together to form a whole.

**BT6-Evaluation:** Evaluating involves presenting and defending opinions by making judgments about information, the validity of ideas, or quality of work based on a set of criteria.

**Program Outcomes (POs)**

**PO1.** **Engineering Knowledge:** Apply the knowledge of mathematics, science, engineering fundamentals and computer engineering specialization to the solutions of complex engineering problems

**PO2.** **Problem Analysis:** Identify, formulate, do research literature and analyze engineering problems to arrive at substantiated conclusions using the fundamental principles of mathematics, natural and engineering sciences

**PO3.** **Design/development of solutions:** Design Computer Engineering solutions for core and interdisciplinary problems and design system components, processes to meet the specifications with consideration for the public health and safety, cultural, societal and environmental aspects

**PO4.** **Conduct investigation of complex problems:** Use research-based knowledge including design of experiments, analysis and the interpretation of data, and synthesis of the information to provide valid conclusions.

**PO5.** **Modern tool usage:** Create, select, and apply appropriate techniques, resources, and modern engineering and IT tools including prediction and modeling to varied engineering activities with an understanding of the limitations

**PO6.** **The engineer and society:** Apply reasoning informed by the contextual knowledge to assess societal, health, safety, legal and cultural issues and the consequent responsibilities relevant to the professional engineering practice.

**PO7.** **Environment and sustainability:** Understand the impact of the professional engineering solutions in societal and environmental contexts, and demonstrate the knowledge of, and need for sustainable development.

**PO8.** **Ethics:** Apply ethical principles and commit to professional ethics and responsibilities and norms of the Computer Engineering practice.

**PO9.** **Individual and team work:** Function effectively as an individual, and as a member or leader in teams, and in multidisciplinary settings.

**PO10. Communication:** Communicate effectively with the Engineering community and with society at large. Be able to comprehend and write effective reports documentation. Make effective presentations, and give and receive clear instructions.

**PO11. Project management and finance:** Demonstrate knowledge and understanding of the engineering and management principles and apply these to one’s own work, as a member and leader in a team, to manage projects in multidisciplinary environments.

**PO12.** **Recognize the need** for, and have the preparation and ability to engage in independent and life-long learning in the broadest context of technological change.

**Mapping of Course Outcomes and Program Outcomes**

CO-PO-PSO Mapping Matrix:

1: Low Mapping 2: Medium Mapping 3: High Mapping“-“ :No Mapping

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Course Outcomes** | **Program Outcomes** | | | | | | | | | | | | | **PSO1** | | **PSO2** | |
| **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** | **9** | **10** | **11** | **12** |  | |  | |
| CO1: Acquiring the basic knowledge of Data warehouse and Data Decision. | 2 | 3 | - | - | 1 | 1 | - | - | - | - | 1 | 1 | 1 | | 2 | |
| CO2: Illustrating the types of Data warehouses. | 2 | 2 | 1 | 3 | 2 | 1 | - | - | - | - | - | 1 | 1 | | 2 | |
| CO3: Differentiation of different types of database systems. | 2 | 2 | 3 | 1 | 3 | - | - | - | - | 1 | - | 3 | 3 | | 2 | |
| CO4: Focus on analytical processes. | 3 | 1 | 3 | 1 | 3 | - | - | - | - | - | - | - | 3 | | 2 | |
| CO5: Classification of data mining systems. | 2 | 3 | 2 | 1 | 1 | 1 | - | - | 2 | - | 2 | 3 | 3 | | 3 | |
| **CO6:** Use of different types of algorithm in data mining. | 2 | 2 | 3 | 2 | 3 | 2 | - | 1 | 2 | 1 | 2 | 1 | 3 | | 2 | |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  | |  | |

**Once the student has successfully completed the course, students must be able to answer the following questions of perform/ demonstrate the following:**

* + - 1. What is the significance of Data warehouse and how its different from DBMS
      2. Will be able to understand the architecture of Data Warehouse
      3. Will be able to provide Analytical solutions
      4. Will be to use Data Mining techniques for problem solving
      5. Will be able to work on different kind of complex data Objects.

**Delivery Methodology - Course Outcome**

|  |  |  |  |
| --- | --- | --- | --- |
| **Course Outcome** | **Method** | **Supporting Tools** | **Demonstration** |
| CO1 | * Lecture * Discussion | * Online Lecture * Discussion * Assignment/Quiz/Test | * Will be able to illustrate the features of Data Warehouse |
| CO2 | * Lecture * Discussion | * Online Lecture * Discussion * Assignment/Quiz/Test | * Will be able to explain the difference between Data Base and Data warehouse |
| CO3 | * Lecture * Discussion | * Online Lecture * Discussion * Assignment/Quiz/Test | * Will be able to understand the architecture of Data Warehouse |
| CO4 | * Lecture * Discussion | * Online Lecture * Discussion * Assignment/Quiz/Test | * Will be able to provide Analytical solutions |
| CO5 | * Lecture * Discussion | * Online Lecture * Discussion * Assignment/Quiz/Test | * Will be to use Data Mining techniques for problem solving |
| CO6 | * Lecture * Discussion | * Online Lecture * Discussion * Assignment/Quiz/Test | * Will be able to work on different kind of complex data Objects |