**tFR. Conceicao Rodrigues College Of Engineering**

Father Agnel Ashram, Bandstand, Bandra-west, Mumbai-50

**Department of Computer Engineering**

**(2018-2019)**

**Course Outcomes & Assessment Plan**

**CLASS-SEM: T.E.- VI**

**COURSE NAME: *Data warehouse and mining* (DWM) TOTAL CREDITS: 4+1=5**

|  |  |  |
| --- | --- | --- |
| **Course Code**  | **Course Name**  | **Credits**  |
| **CSC603**  | **Data Warehousing and Mining**  | **4**  |

|  |  |  |
| --- | --- | --- |
| **Lab Code**  | **Lab Name**  | **Credits** |
| **CSL603** | **Data Warehousing and Mining Lab** | 1 |

**Professor : Dr. Sujata P. Deshmukh**

|  |  |
| --- | --- |
| **PAC Members:** | **Head Of Department:** |

**Department of Computer Engineering**

**Academic Term: Jan-May 2019(2018-2019)**

**T.E. (Computer) (semester VI)**

**Course Outcomes & Assessment Plan**

**Syllabus:**

****

****

**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 Objectives (optional):**

****

**Course Outcomes:**  **At the end of this course students should be able to**

|  |  |  |  |
| --- | --- | --- | --- |
| Co-code | CO | **Blooms Taxonomy**  | **Explanation** |
| **CSC603.1** | Understand Data Warehouse fundamentals, Data Mining Principles  | **Comprehension (**explains, gives examples, shows relationship of ) |  Understand the architecture of variant Warehouse systems. |
| **CSC603.2** | Design data warehouse with dimensional modeling and apply OLAP operations  | **Application (**applies, solves , uses, demonstrates ) | Facilitating representations for data models for data warehouse |
| **CSC603.3** | Identify appropriate techniques / algorithms to solve real world problems of data mining and data exploration. | **Knowledge (**defines, describes, recalls , labels, lists, matches, names ) |  Describe & Inference based on Classification process |
| **CSC603.4** | Compare and evaluate different data mining techniques like classification, prediction, clustering and association rule mining  | **Application (**applies, solves , uses, demonstrates )**Synthesis(**combines, compiles,composes, creates, devises, designs,diverse elements) | Analyzing Clustering as problem solving methodsApplication by integrating principles of Association Rules& implementation |
| **CSC603.5** | Describe complex data types wrt to special and web mining  | **Comprehension (**explains, gives examples, shows relationship of ) | Understanding architecture of variant DM systems |
| **CSC603.6** | Benefit the user experiences towards research and innovation. | **Create(Adapt, Build, Change, Choose)** | Apply DM algorithms and concepts to solve real world problems to find pattern/knowledge or proposing alternative solutions**.** |

**Lab Outcome: At the end of this course students should be able to**

|  |  |
| --- | --- |
| Co-code | CO |
| **CSL603.1** | 1. Design data warehouse and perform various OLAP operations. |
| **CSL603.2** |  2. Implement classification, prediction, clustering and association rule mining algorithms. |
| **CSL603.3** | 3. Demonstrate classifications, prediction, clustering and association rule mining algorithms on a given set of data sample using data mining tools. |
| **CSL603.4** | 4. Implement spatial and web mining algorithms. |

**Mapping of CO and PO/PSO**

Relationship of course outcomes with program outcomes: Indicate 1 (low importance), 2 (Moderate Importance) or 3 (High Importance) in respective mapping cell.

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | PO1(EnggKnow) | PO2(Ana) | PO3(Design) | PO4(investiga) | PO5(tools) | PO6(enggSoci) | PO7(Env) | PO8(Eth) | PO9(indTeam) | PO10(comm.) | PO11(PM) | PO12(lifeLong) |
| **CSC603.1** | 3 | - |  |  |  |  |  |  |  |  |  |  |
| **CSC603.2** | 3 | 3 | 3 |  | 2 |  |  |  | 3 | 3 | 3 | 3 |
| **CSC603.3** | 3 | 3 | 3 |  | 3 |  |  |  | 3 | 3 |  | 3 |
| **CSC603.4** | 3 | 3 | 3 |  | 3 |  |  |  | 3 | 3 |  | 3 |
| **CSC603.5** | 3 | 3 | 3 |  | 3 |  |  |  | 3 | 3 |  | 3 |
| **CSC603.6** | 3 | 3 | 3 |  |  |  |  |  | 3 | 3 |  | 3 |
| CourseTo PO | 3 | 3 | 3 |  | 2.8 |  |  |  | 3 |  |  | 3 |

|  |  |  |
| --- | --- | --- |
| CO | PSO1 | PSO2 |
| **CSC603.1** | 3 |  |
| **CSC603.2** | 3 | 3 |
| **CSC603.3** | 3 | 3 |
| **CSC603.4** | 3 | 3 |
| **CSC603.5** | 3 | 3 |
| **CSC603.6** | 3 | 3 |
| Course to PSO | 3 | 3 |

**Justification:**

CO1: Understand Data Warehouse fundamentals, Data Mining Principles

PO1: Students use fundamental knowledge of data processing and DBMS to learn DWM concepts

CO2: Design data warehouse with dimensional modeling and apply OLAP operations

PO1: Students should have knowledge to use symbols of to draw Schema diagrams and to choose appropriate dimension to design the schema for chosen mini project

PO2: Students analyze the Schema diagrams and choose appropriate dimension and facts to design the dimension modeling for chosen mini projector given problem. Also verifies its correctness.

PO3: Students design DWH using dimension modeling schemas.

PO5: Students use DBMS tool (pgadmin-postgresql) or languages to design and implement a dimensional modeling given to them individually and in group.

PO9: Students analyze and design dimension modeling by discussing in group

PO10: Students communicate effectively on MINI project aspects by writing reports, and giving PPT

PO11: Students demonstrate the knowledge and concept applied to MINI project as a team leader or group member.

PO12: Students develop the ability to handle and to think independently using the concepts of dimension modeling diagram to learn futuristic subjects.

CO3: Identify appropriate data mining algorithms to solve real world problems.

PO1: Students use fundamental knowledge of mathematics, science and engineering principles to identify appropriate data mining algorithms.

PO2: Students have to analyze different algorithm for a given query and selected mini projects.

PO3: Students apply different algorithms for a given query and selected mini projects to analyse the knowledge .

PO5: Students use Weka tool or languages Python/Java to implement a data mining algorithms given to them individually and in group.

PO9: Students discuss in group for getting solution to the selected mini project.

PO10: Students communicate effectively on MINI project aspects by writing reports, and giving PPT

PO11: Students demonstrate the knowledge and concept applied to MINI project as a team leader or group member.

PO12: students develop the ability to handle and to think independently using the concepts of DWM to learn futuristic subjects like BDA.

CO4: Compare and evaluate different data mining techniques like classification, prediction, clustering and association rule mining

PO1: Students use fundamental knowledge of mathematics, science and engineering principles, to understand evaluate different data mining techniques like classification, prediction, clustering and association rule mining.

PO2: students have to analyze different data mining techniques (classification, prediction, clustering and association rule mining) and special mining functions for a given problem and selected mini projects.

PO5: Students use DM tool (Weka) or languages to write, implement, compare algorithm given to them individually and in group.

PO9: students discuss in group for getting solution to the selected mini project.

PO10: Students communicate effectively on MINI project aspects by writing reports, and giving PPT

PO11: students demonstrate the knowledge and concept applied to MINI project as a team leader or group member.

PO12: students develop the ability to handle and to think independently using the concepts of SQL to learn futuristic subjects.

CO5: Describe complex data types wrt to special and web mining

PO1: Students use fundamental knowledge of mathematics, science and engineering principles to describe complex data types wrt to special and web mining.

PO2: Students analyzing the complex data types wrt to special and web mining.

PO5: Students use Weka tool or languages Python/Java to describe complex data types given to them individually and in group.

PO9: Students discuss in group for getting solution to the selected mini project.

PO10: Students communicate effectively on MINI project aspects by writing reports, and giving PPT

PO11: students demonstrate the knowledge and concept applied to MINI project as a team leader or group member.

PO12: students develop the ability to handle and to think independently using the concepts of Normalization to learn futuristic subjects.

CO6: Benefit the user experiences towards research and innovation.

PO1: Students use fundamental knowledge of mathematics, science and engineering principles to design DWH and to mine data using DM principles and algorithms

PO2: Students analyzing the data to to design DWH and to mine data using DM principles and algorithms

PO9: students discuss in group for getting solution to the selected mini project.

PO10: Students communicate effectively on MINI project aspects by writing reports, and giving PPT

PO11: students demonstrate the knowledge and concept applied to MINI project as a team leader or group member.

PO12: students develop the ability to handle and to think independently using the concepts of transaction to learn futuristic subjects.

Justification

PO1: All COs are mapped to PO1 because engineering graduates will be able to apply the knowledge of DWM fundamentals to solve complex engineering problems.

PO2: CSC603.2-CSC603.6, are mapped to PO2 because the students analyze a DW system.

PO3: CSC603.2-CSC603.6, are mapped to PO3 because the students design a DW system using dimensional Modeling .

PO5: CSC603.2-CSC603.6 is mapped to this PO5 because the students use the postgresql database software to design DWH and WEKA/rapid miner/ different languages to mine knowledge from data

PO9: CSC603.2-CSC603.6, are mapped to this PO9 because the students work in a team to develop the mini project DWM system.

PO10: CSC603.2-CSC603.6, are mapped to this PO10 because Students communicate effectively on MINI project aspects by writing reports, and giving PPT

PO11: CSC603.2-CSC603.6, are mapped to this PO11 because students demonstrate the knowledge and concept applied to MINI project as a team leader or group member.

PO12: CSC603.2-CSC603.6, are mapped to this PO12 because students develop the ability to handle and to think independently using the concepts of dimension modeling and DM algorithms to learn futuristic subjects.

PSO1: All COs are mapped to PSO1 because the graduates will be able to apply fundamental knowledge of DWM to mine datasets for real world problems.

PSO2: CSC603.2-CSC603.5, are mapped to PSO2 because the students design and implement the mini project using the knowledge of database concepts of dimension modeling, DM algorithms.

**CO Assessment Tools:**

**CSC603.1 Understand Data Warehouse fundamentals, Data Mining Principles**

|  |  |
| --- | --- |
|  **Direct Method Tools (dm)** | **Wt=80%** |
| Lab (**lab)** | 0.1 |
| Assignment1**(assign)** | 0.2 |
| Mini project  | 0.1 |
| UnitTest1**(ut1)** | 0.2 |
| End Sem Marks(Theory) **(uth)** | 0.2 |
| End Sem Marks(practical and oral) **(utpra)** | 0.2 |
| **Indirect Method Tools(idm)** | **Wt=20%** |
| Course Exit Survey **(C03idm)** |  |
|  **CSC603.1= 0.8\*** **CSC603.1dm + 0.2\* CSC603.1idm** |

***CSC603.2*** Design data warehouse with dimensional modelling and apply OLAP operations

|  |  |
| --- | --- |
| **Direct Method Tools (dm)** | **Wt=80%** |
| Lab (**lab)** | 0.2 |
| Assignment1**(assign)** | 0.1 |
| Mini project  | 0.1 |
| UnitTest1**(ut1)** | 0.2 |
| End Sem Marks(Theory) **(uth)** | 0.2 |
| End Sem Marks(practical) **(utpra)** | 0.2 |
| **Indirect Method Tools(idm)** | **Wt=20%** |
| Course Exit Survey **(C03idm)** |  |
| **CSC603.2= 0.8\*** **CSC603.2dm + 0.2\* CSC603.2idm** |

***CSC603.3 Identify appropriate data mining algorithms to solve real world problems.***

|  |  |
| --- | --- |
| **Direct Method Tools (dm)** | **Wt=80%** |
| Lab (**lab)** | 0.2 |
| Assignment2 **(assign)** | 0.1 |
| Mini project  | 0.1 |
| UnitTest1**(ut1)** | 0.2 |
| End Sem Marks(Theory) **(uth)** | 0.2 |
| End Sem Marks(practical) **(utpra)** | 0.2 |
| **Indirect Method Tools(idm)** | **Wt=20%** |
| Course Exit Survey **(C03idm)** |  |
| **CSC603.3= 0.8\*** **CSC603.3dm + 0.2\* CSC603.3idm** |

***CSC603.4 Compare and evaluate different data mining techniques like classification, prediction, clustering and association rule mining***

|  |  |
| --- | --- |
| **Direct Method Tools (dm)** | **Wt=80%** |
| Lab (**lab)** | 0.4 |
| Assignment 2 | 0.1 |
| Mini project | 0.2 |
| UnitTest1**(ut1)** | 0.1 |
| End Sem Marks(Theory) **(uth)** | 0.1 |
| End Sem Marks(practical) **(utpra)** | 0.1 |
| **Indirect Method Tools(idm)** | **Wt=20%** |
| Course Exit Survey **(C03idm)** |  |
| **CSC603.4= 0.8\*** **CSC603.4dm + 0.2\* CSC603.4idm** |

***CSC603.5 Describe complex data types wrt to special and web mining.***

|  |  |
| --- | --- |
| **Direct Method Tools (dm)** | **Wt=80%** |
| Assignment3 (**ASG3)** | 0.1 |
| Lab work  | 0.2 |
| Mini project | 0.1 |
| UnitTest2**(ut2)** | 0.2 |
| End Sem Marks(Theory) **(uth)** | 0.2 |
| End Sem Marks(practical) **(utpra)** | 0.2 |
| **Indirect Method Tools(idm)** | **Wt=20%** |
| Course Exit Survey **(C03idm)** |  |
| **CSC603.5= 0.8\*** **CSC603.5dm + 0.2\* CSC603.5idm** |

***CSC603.6 Benefit the user experiences towards research and innovation.***

|  |  |
| --- | --- |
| **Direct Method Tools (dm)** | **Wt=80%** |
| Mini project | 0.5 |
| Lab work | 0.1 |
| End Sem Marks(Theory) **(uth)** | 0.2 |
| End Sem Marks(practical) **(utpra)** | 0.2 |
| **Indirect Method Tools(idm)** | **Wt=20%** |
| Course Exit Survey **(C03idm)** |  |
| **CSC603.6= 0.8\*** **CSC603.6dm + 0.2\* CSC603.6idm** |

**Course Outcomes Target:**

*Upon completion of this course students will be able to:*

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| CO-CODE | Course Outcomes  | Blooms T. | Level | Target  | Achieved  |
| CSC603.1  | Understand Data Warehouse fundamentals, Data Mining Principles  | Comprehension | L1, L2 | 2.5 | **2.52** |
| CSC603.2 | Design data warehouse with dimensional modelling and apply OLAP operations  |  Design | L5 | 2.5 | **2.52** |
| CSC603.3 | Identify appropriate data mining algorithms to solve real world problems  | Apply and Design | L4,L5 | 2.5 | **2.68** |
| CSC603.4 | Compare and evaluate different data mining techniques like classification, prediction, clustering and association rule mining  | Analyze, Apply | L4, L3 | 2.5 | **2.84** |
| CSC603.5 | Describe complex data types wrt to special and web mining  | Analyze, Apply | L4, L3 | 2.4 | **2.68** |
| CSC603.6 | Benefit the user experiences towards research and innovation. | Comprehension, apply | L1, L2, L4 | 2.4 | **2.68** |

**Content Beyond Syllabus:**

 1. Moving from data mining to big data analytics- Limitations of data mining

**Curriculum Gap:**

No Curriculum Gap

**Rubrics for the Lab Experiments:**

**Rubrics for Experiment Grading of Expt.**

|  |  |  |  |
| --- | --- | --- | --- |
| Indicator | Average | Good | Excellent  |
| Timeline * On time Completion & Submission (02)
 | Late submission (0) | 01 (On Time ) | 02 (Before deadline ) |
| Completeness and neatness* Complete all parts of schema diagram / OLAP / Algorithm (2)
 | < 60% complete (0) | < 80% complete (1) |  100% complete (2) |
| Implementation * Extent of coding (4)
 | < 60% complete (2) | < 80% complete (3) |  100% complete (4) |
| Knowledge* In depth knowledge of the post assignment questions (2)
 | Unable to answer 2 questions(0) | Unable to answer 1 question (1) | Able to answer 2 questions (2) |
| **Completeness and neatness (2)** | **Implementation (4)** | **Timeline (2)** | **Knowledge (2)** |
|  |  |  |  |

**Rubrics for Experiment Grading of Mini Project**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Indicator** | ***Poor*** | ***Average*** | ***Good*** | ***Excellent*** |
| ***Timeline(2)*** | More than two sessions late (0) | Two sessions late (0) | One session late (1) | Early or on time (2) |
| ***Identification and implementation of DM algorithms(3)*** | No (0) | < 60% complete (1) | < 80% complete (2) |  100% complete (3) |
| ***Selection of problem for Mini Project and data set***  | Selection of problem for Mini Project and missing data set(0) | Selection of problem for Mini Project and data set (1) | Selection of problem for Mini Project and improper data set(3) | Selection of problem for Mini Project and proper data set(4) |

**Rubrics for the Assignments:**

**Assignments1**

|  |  |  |  |
| --- | --- | --- | --- |
| **Indicator** | Excellent | Good | Average |
| **Timeline** **(3)** |  submitted on time or early (3) | Submitted in same week (2) | Submitted in next week (1) |
| **Understands the Problem(3)** | Identifies special factors that influences the approach before starting the problem | Understands the problem | Understands enough to solve part of the problem or to get part of the solution |
| **Applies Appropriate technique (schema Diagram/OLAP/data preprocessing) (4)** | Explains how the techniques are appropriate in given problem (4) | Applies completely appropriate techniques in given problem (3) | Applies some appropriate techniques in given problem (2) |
| **Timeline** **(3)** | **Understands the Problem(3)** | **Applies Appropriate Symbols (4)** | **Total (10)** |
|  |  |  |  |

**Assignments2 and 3**

|  |  |  |  |
| --- | --- | --- | --- |
| **Indicator** | Excellent | Good | Average |
| **Timeline** **(2)** |  submitted on time or early (2) | Submitted in same week (1) | Submitted in next week (0) |
| **Understands the DM algorithm (4)** | Explains why procedures are appropriate DM algorithm (4) | Applies Appropriate DM algorithm (3) | Applies Appropriate DM algorithm <50% (2) |
| **Applies Appropriate procedure (4)** | Explains why procedures are appropriate for SQL(4) | Applies Appropriate SQL (3) | Applies some appropriate SQL (2) |
| **Timeline** **(2)** | **Understands the Problem(4)** | **Applies Appropriate Procedures(4)** | **Total (10)** |
|  |  |  |  |

**FR. Conceicao Rodrigues College Of Engineering**

Father Agnel Ashram, Bandstand, Bandra-west, Mumbai-50

**Department of Computer Engineering**

**(2018-2019)**

**Lesson Plan**

**CLASS-SEM: T.E.- VI Credits-4**

**SUBJECT: *Data warehouse and mining (DWM)* SUBJECT CODE- CSC603**

**Professor : Dr. Sujata P. Deshmukh**

**Modes of Content Delivery:**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| i | Class Room Teaching | v | Self Learning Online Resources | Ix | Industry Visit |
| ii | Tutorial | vi | Slides, PPT | X | Group Discussion |
| iii | Remedial Coaching | vii | Simulations/Demonstrations | xi | Seminar |
| iv | Lab Experiment | viii | Expert Lecture | xii | Case Study |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Lect.****No.** | **Chap.no** | **Portion to be covered** | **Planned date** | **Actual date** | **Content Delivery Method/Learning Activities** |
| 1 | 1.1 | The Need for Data Warehousing; Increasing Demand for Strategic Information; Inability of Past Decision Support System; Operational V/s Decisional Support System; | 01/01/2019 | 01/01/2019 | Class Room Teaching, PPT |
| 2 | 1.2 | Data Warehouse Definition; Benefits of Data Warehousing; Features of a Data Warehouse; The Information Flow Mechanism; Architecture of DW, Metadata | 02/01/2019 | 02/01/2019 | Class Room Teaching, PPT |
| 3 | 1.3 | DWH Vs data marts, EDWH | 03/01/2019 | 03/01/2019 | Class Room Teaching, PPT |
| 4 | 1.4 | Different approaches and layered Architecture | 04/01/2019 | 04/01/2019 | Class Room Teaching, PPT |
| 5 | 1.5 | 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? | 07/01/2019 | 07/01/2019 | Class Room Teaching, Lab Experiment |
| 6 | 1.6 | Keys in DW, Snowflake Schema, Fact Constellation Schema or Families of Star. | 08/01/2019 | 08/01/2019 | Class Room Teaching, Lab Experiment |
| 7 | 1.7 | The Factless Fact, Table; Aggregate Table, Information Package Diagram | 09/01/2019 | 08/01/2019 | Class Room Teaching, Lab Experiment |
| 8 | 1.8 | 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 | 10/01/2019 | 09/01/2019 | Class Room Teaching, Lab Experiment |
| 9 | 2.1 | Challenges in ETL Functions; Data Extraction; Identification of Data Sources, Issues in Data Cleansing. | 14/01/2019 | 14/01/2019 | Class Room Teaching, PPT |
| 10 | 2.2 | Extracting Data: Immediate Data Extraction, Deferred Data Extraction | 16/01/2019 | 16/01/2019 | Class Room Teaching |
| 11 | 2.3 | Data Transformation: Tasks Involved in Data Transformation, | 17/01/2019 | 17/01/2019 | Class Room Teaching |
| 12 | 2.4 | Data Loading: Techniques of Data Loading, Loading the Fact Tables and Dimension Tables Data Quality;  | 18/01/2019 | 18/01/2019 | Class Room Teaching |
| 13 | 2.5 | Need for Online Analytical Processing; OLTP V/s OLAP; OLAP and Multidimensional Analysis  | 21/01/2019 | 21/01/2019 | Class Room Teaching, PPT, Lab Experiment |
| 14 | 2.6 | Hypercubes; OLAP Operations in Multidimensional Data Model;  | 23/01/2019 | 23/01/2019 | Class Room Teaching, Lab Experiment, PPT |
| 15 | 2.7 | OLAP Models: MOLAP, ROLAP, HOLAP, DOLAP, Definition of Schema using DMQL, Examples on OLAP | 24/01/2019 | 24/01/2019 | Class Room Teaching, Lab Experiment, PPT |
| 16 | 2.8 | Problems based on OLAP Operations | 25/01/2019 | 25/01/2019 | Class Room Teaching, Lab Experiment |
| 17 | 3.1 | What is Data Mining; Knowledge Discovery in Database (KDD) | 28/01/2019 | 28/01/2019 | Class Room Teaching, Lab Experiment |
| 18 | 3.2 | What can be Data to be Mined, Related Concept to Data Mining, | 30/01/2019 | 30/01/2019 | Class Room Teaching, Lab Experiment |
| 19 | 3.3 | Data Mining Technique, Application and Issues in Data Mining | 01/02/2019 | 01/02/2019 | Class Room Teaching, Lab Experiment, PPT |
| 20 | 3.4 | Comparison of DM with BDA | 07/02/2019 | 03/02/2019 | Class Room Teaching, self learning online resources |
| 21 | 3.5 | Data exploration: Types of Attributes; Statistical Description of Data;  | 08/02/2019 | 07/02/2019 | Class Room Teaching, Lab Experiment |
| 22 | 3.6 | Data Visualization; Measuring similarity and dissimilarity | 11/02/2019 | 08/02/2019 | Class Room Teaching, Lab Experiment |
| 23 | 3.7 | Why Preprocessing? Data Cleaning; Data Integration; Data Reduction: Attribute subset selection, Histograms, Clustering and Sampling; | 12/02/2019 | 11/02/2019 | Class Room Teaching, self learning online resources |
| 24 | 3.8 | Data Transformation & Data Discretization: Normalization | 18/02/2019 | 18/02/2019 | Class Room Teaching, self learning online resources |
| 25 | 3.9 | Binning, Histogram Analysis and Concept hierarchy generation and description  | 20/02/2019 | 20/02/2019 | Class Room Teaching |
| 26 | 4.1 | Basic Concepts; Decision Tree Induction: Attribute Selection Measures, Tree pruning. | 21/02/2019 | 21/02/2019 | Class Room Teaching |
| 27 | 4.2 | Bayesian Classification: Naïve Bayes’ Classifier. | 22/02/2019 | 22/02/2019 | Class Room Teaching  |
| 28 | 4.3 | Prediction: Structure of regression models; Simple linear regression, Multiple linear regressions. | 25/02/2019 | 27/02/2019 | Class Room Teaching |
| 29 | 4.4 | Model Evaluation & Selection: **A**ccuracy and Error measures, Holdout, Random Sampling, Cross Validation | 27/02/2019 | 28/02/2019 | Open Discussion with Students  |
| 30 | 4.5 | Problems based on classification  | 28/02/2019 | 28/02/2019 | Class Room Teaching, Lab Experiment |
| 31 | 4.6 | What is clustering? Types of data, Partitioning Methods (K-Means) | 01/03/2019 | 28/02/2019 | Class Room Teaching |
| 32 | 4.7 | Partitioning Methods (K-Means, KMedoids) | 06/03/2019 | 01/03/2019 | Class Room Teaching |
| 33 | 4.8 | Hierarchical Methods(Agglomerative , Divisive, BRICH), | 07/03/2019 | 06/03/2019 | Class Room Teaching |
| 34 | 4.9 | Hierarchical Methods(Agglomerative , Divisive, BRICH), | 08/03/2019 | 07/03/2019 | Class Room Teaching |
| 35 | 4.10 | Problems based on clustering  | 11/03/2019 | 08/03/2019 | Class Room Teaching |
| 36 | 5.1 | Market Basket Analysis, Frequent Itemsets, Closed Itemsets, and Association Rules; | 13/03/2019 | 11/03/2019 | Class Room Teaching |
| 37 | 5.2 | Frequent Pattern Mining, Efficient and Scalable Frequent Itemset Mining Methods,  | 14/03/2019 | 13/03/2019 | Class Room Teaching, Lab Experiment |
| 39 | 5.3 | The Apriori Algorithm for finding Frequent Itemsets Using Candidate Generation  | 15/03/2019 | 14/03/2019 | Class Room Teaching, Lab Experiment |
| 40 | 5.4 | Generating Association Rules from Frequent Itemsets, Improving the Efficiency of Apriori, | 18/03/2019 | 18/03/2019 | Class Room Teaching |
| 41 | 5.5 | A pattern growth approach for mining Frequent Itemsets; Mining Frequent itemsets using vertical data formats; Mining closed and maximal patterns;  | 20/03/2019 | 20/03/2019 | Class Room Teaching |
| 42 | 5.6 | Introduction to Mining Multilevel Association Rules and Multidimensional Association Rules; | 22/03/2019 | 22/03/2019 | Class Room Teaching |
| 43 | 5.7 | From Association Mining to Correlation Analysis, Pattern Evaluation Measures; | 25/03/2019 | 25/03/2019 | Class Room Teaching |
| 44 | 5.8 | Introduction to Constraint-Based Association Mining., Examples | 27/03/2019 | 27/03/2019 | Class Room Teaching |
| 45 | 6.1 | Special vs classical data mining, data structures,  | 28/03/2019 | 28/03/2019 | Class Room Teaching |
| 46 | 6.2 | Mining co location and special patterns , Special clustering techniques | 29/03/2019 | 29/03/2019 | Class Room Teaching |
| 47  | **6.3** | Web Content Mining, Web structure Mining | 01/04/19 | 01/04/19 | PPT |
| 48 | 6.4 | Content beyond syllabus-Moving from data mining to BDA Analytics | 03/04/19 | 03/04/19 | Test followed discussion  |
| **Remedial**  |
| 49 | r.1 | DWH revision and discussion | 04/04/19 | 05/04/19 | PPT, discussion  |
| 50 | r.2 | Data mining revision and discussion | 04/04/19 | 05/04/19 | PPT |
| 51 | r.3 | Different algorithms discussion  | 05/04/19 | 11/04/19 | PPT |
| 52 | r.4 | Special and web mining  | 06/04/19 | 12/04/19 | PPT |

**Online Resources :**

1) https://www.tutorialspoint.com/dwh/ (weak students)

2) https://www.coursera.org/learn/dwdesign

3 https://sites.google.com/site/dataminingandbisem6/home/materials/notes

4) https://www.postgresql.org/docs/9.5/cube.html , grouping sets, rollup

5) https://www.postgresql.org/docs/9.5/cube.html (strong students)

6) https://www.kaggle.com/kernals (strong students)

**Online Resources for datasets:**

1. https://storm.cis.fordham.edu/~gweiss/data-mining/datasets.html- **Sample Weka Data Sets**
2. https://storm.cis.fordham.edu/~gweiss/data-mining/weka-data/contact-lenses.arff
3. http://repository.seasr.org/Datasets/UCI/arff/
4. https://www.kaggle.com/datasets
5. **Stanford Large Network Dataset Collection-https://snap.stanford.edu/data/(strong students)**

**List of Experiments, Assignments and mapping**

|  |  |  |  |
| --- | --- | --- | --- |
| Sr. no. | No. Title of Experiments | Course co mapping | Lab co mapping |
| 1 |  Build Data Warehouse/Data Mart for a given problem statement | **CSC603.1,****CSC603.2** | **CSL603.1** |
| i) Identifying the source tables and populating sample data |
| ii) Design dimensional data model i.e. Star schema, Snowflake schema and Fact Constellation schema (if applicable) |
| iii) implementation of DWH for a given problem |
| 2 |  To perform various OLAP operations such as slice, dice, drilldown, rollup, pivot | **CSC603.1,****CSC603.2** | **CSL603.1** |
| 3 | To perform data exploration and data cleaning in python /R | **CSC603.3,****CSC603.4** | **CSL603.2,****CSL603.3.** |
| 4 | Implementation of Association Rule Mining algorithm (Apriori in java/python). | **CSC603.3,****CSC603.4** | **CSL603.2,****CSL603.3.** |
| 5 | Implementation of Clustering algorithm ( K-means in java/python). | **CSC603.3,****CSC603.4** | **CSL603.2,****CSL603.3.** |
| 6 | Implementation of Linear Regression. | **CSC603.3,****CSC603.4** | **CSL603.2,****CSL603.3.** |
| 7 |  Use WEKA to implement classification (Part1, Part2) | **CSC603.3,****CSC603.4** | **CSL603.2,****CSL603.3** |
| 8 | Implement spatial and web mining algorithms. | **CSC603.5,****CSC603.6** | **CSL603.4,** |
| 9 |

|  |
| --- |
| Use WEKA to implement Association Mining and Clustering algorithm |

 | **CSC603.5,****CSC603.6** | **CSL603.3** |
| 10 | Mini project/Case study | **CSC603.6** | **All**  |
| **Assignments** |
| 11 | **Assignment1-Data ware housing** | **CSC603.1,****CSC603.2** |  |
| 12 | **Assignment2-Data Exploration** | **CSC603.3,****CSC603.4** |  |
| 13 | **Assignment3-DM Algorithms** | **CSC603.3,****CSC603.4** |  |
| 14 | **Assignment4-Social mining** | **CSC603.5** |  |

Lab Plan

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Sr. no. | No. Title of Experiments | Batch A | Batch B | Batch C | Batch D |
| 1 |  Build Data Warehouse/Data Mart for a given problem statement | 14/01/2019 | 18/01/2019 | 16/01/2019 | 17/01/2019 |
| 2 |  To perform various OLAP operations such as slice, dice, drilldown, rollup, pivot | 21/01/2019 | 25/01/2019 | 23/01/2019 | 22/01/2019 |
| 3 | To perform data exploration and data cleaning in python /R | 28/01/2019 | 01/02/2019 | 30/01/2019 | 29/01/2019 |
| 4 | Implementation of Association Rule Mining algorithm (Apriori in java/python). | 18/02/2019 | 08/02/2019 | 20/02/2019 | 28/02/2019 |
| 5 | Implementation of Clustering algorithm ( K-means in java/python). | 25/02/2019 | 21/02/2019 | 27/02/2019 | 28/02/2019 |
| 6 | Implementation of Linear Regression. | 11/03/2019 | 28/02/2019 | 06/03/2019 | 06/03/2019 |
| 7 |  Use WEKA to implement classification (Part1, Part2) | 18/03/2019 | 08/03/2019 | 13/03/2019 | 13/03/2019 |
| 8 | Implementation of Page Rank  | 25/03/2019 | 22/03/2019 | 20/03/2019 | 26/03/2019 |
| 9 | WEKA to implement Association Mining and Clustering algorithm | 01/04/2019 | 29/03/2019 | 27/03/2019 | 02/04/2019 |
| 10 | Mini project/Case study | 05/04/2019 | 05/04/2019 | 03/04/2019 | 05/04/2019 |