FR. Conceicao Rodrigues College Of Engineering

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

Department of Computer Engineering

(2022-2023)

Course Outcomes & Assessment Plan

CLASS-SEM: T.E.- V

COURSE NAME: *Data warehouse and mining* (DWM) **TOTAL CREDITS: 3+1=4**

Course Code	Course Name	Credits
CSC504	Data Warehousing and Mining	3

Lab Code	Lab Name	Credits
CSL503	Data Warehousing and Mining Lab	1

Professor : Dr. Sujata P. Deshmukh

Department of Computer Engineering
Academic Term: Jul-Dec 2022(2022-2022)
T.E. (Computer) (semester V)

Course Outcomes & Assessment Plan

Syllabus:

Module	Content	Hrs
1	**Data Warehousing Fundamentals**	8
	Introduction to Data Warehouse, Data warehouse architecture, Data warehouse versus Data Marts, 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. Major steps in ETL process, OLTP versus OLAP, OLAP operations: Slice, Dice, Rollup, Drilldown and Pivot.	
2	**Introduction to Data Mining, Data Exploration and Data Pre-processing**	8
	Data Mining Task Primitives, Architecture, KDD process, Issues in Data Mining, Applications of Data Mining, Data Exploration: Types of Attributes, Statistical Description of Data, Data Visualization, Data Preprocessing: Descriptive data summarization, Cleaning, Integration & transformation, Data reduction, Data Discretization and Concept hierarchy generation.	
3	**Classification**	6
	Basic Concepts, Decision Tree Induction, Naive Bayesian Classification, Accuracy and Error measures, Evaluating the Accuracy of a Classifier: Holdout & Random Subsampling, Cross Validation, Bootstrap.	
4	**Clustering**	6
	Types of data in Cluster analysis, Partitioning Methods (*k*-Means, *k*-Medoids), Hierarchical Methods (Agglomerative, Divisive).	
5	**Mining frequent patterns and associations**	6
	Market Basket Analysis, Frequent Item sets, Closed Item sets, and Association Rule, Frequent Pattern Mining, Apriori Algorithm, Association Rule Generation, Improving the Efficiency of Apriori, Mining Frequent Itemsets without candidate generation, Introduction to Mining Multilevel Association Rules and Mining Multidimensional Association Rules.	
6	**Web Mining**	5
	Introduction, Web Content Mining: Crawlers, Harvest System, Virtual Web View, Personalization, Web Structure Mining: Page Rank, Clever, Web Usage Mining.	

Textbooks:	
1	Paulraj Ponniah, *"Data Warehousing: Fundamentals for IT Professionals"*, Wiley India.
2	Han, Kamber, *"Data Mining Concepts and Techniques"*, Morgan Kaufmann 2nd edition.
3	M.H. Dunham, *"Data Mining Introductory and Advanced Topics"*, Pearson Education.
References:	
1	Reema Theraja, *"Data warehousing"*, Oxford University Press 2009.
2	Pang-Ning Tan, Michael Steinbach and Vipin Kumar, *"Introduction to Data Mining"*, Pearson Publisher 2nd edition.
3	Ian H. Witten, Eibe Frank and Mark A. Hall, *"Data Mining"*, Morgan Kaufmann 3rd edition.

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.

Lab Code	Lab Name	Credit
CSL503	Data Warehousing and Mining Lab	1

Prerequisite: Database Concepts	
Lab Objectives:	
1.	Learn how to build a data warehouse and query it.
2.	Learn about the data sets and data preprocessing.
3.	Demonstrate the working of algorithms for data mining tasks such Classification, clustering, Association rule mining & Web mining.
4.	Apply the data mining techniques with varied input values for different parameters.
5.	Explore open source software (like WEKA) to perform data mining tasks.
Lab Outcomes: At the end of the course, the student will be able to	
1.	Design data warehouse and perform various OLAP operations.
2.	Implement data mining algorithms like classification.
3.	Implement clustering algorithms on a given set of data sample.
4.	Implement Association rule mining & web mining algorithm.

Suggested List of Experiments	
Sr. No.	**Title of Experiment**
1	One case study on building Data warehouse/Data Mart • Write Detailed Problem statement and design dimensional modelling (creation of star and snowflake schema)
2	Implementation of all dimension table and fact table based on experiment 1 case study
3	Implementation of OLAP operations: Slice, Dice, Rollup, Drilldown and Pivot based on experiment 1 case study
4	Implementation of Bayesian algorithm
5	Implementation of Data Discretization (any one) & Visualization (any one)
6	Perform data Pre-processing task and demonstrate Classification, Clustering, Association algorithm on data sets using data mining tool (WEKA/R tool)
7	Implementation of Clustering algorithm (K-means/K-medoids)
8	Implementation of any one Hierarchical Clustering method
9	Implementation of Association Rule Mining algorithm (Aprion)
10	Implementation of Page rank/HITS algorithm

Term Work:	
1	Term work should consist of 10 experiments.
2	Journal must include at least 1 assignment on content of theory and practical of "Data Warehousing and Mining"
3	The final certification and acceptance of term work ensures that satisfactory performance of laboratory work and minimum passing marks in term work.
4	Total 25 Marks (Experiments: 15-marks, Attendance (Theory & Practical): 05-marks, Assignments: 05-marks)
Oral & Practical exam	
	Based on the entire syllabus of CSC504 : Data Warehousing and Mining

Course Objectives (optional):

Course Code:	Course Title	Credit
CSC504	Data Warehousing and Mining	3

Prerequisite: Database Concepts	
Course Objectives:	
1.	To identify the significance of Data Warehousing and Mining.
2.	To analyze data, choose relevant models and algorithms for respective applications.
3.	To study web data mining.
4.	To develop research interest towards advances in data mining.
Course Outcomes: At the end of the course, the student will be able to	
1.	Understand data warehouse fundamentals and design data warehouse with dimensional modelling and apply OLAP operations.
2.	Understand data mining principles and perform Data preprocessing and Visualization.
3.	Identify appropriate data mining algorithms to solve real world problems.
4.	Compare and evaluate different data mining techniques like classification, prediction, clustering and association rule mining
5.	Describe complex information and social networks with respect to web mining.

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

Co-code	CO	Blooms Taxonomy	Explanation
CSC504.1	Understand Data Warehouse and Data Mining principles	**Comprehension** (explains, gives examples, shows relationship of)	Understand the architecture of variant Warehouse systems.
CSC504.2	Design data warehouse with dimensional modeling and apply OLAP operations	**Application** (applies, solves , uses, demonstrates)	Facilitating representations for data models for data warehouse
CSC504.3	Demonstrate data mining principles and perform Data preprocessing and Visualization	**Knowledge** (defines, describes, recalls , labels, lists, matches, names) Evaluate –(Determine, Compare, Interpret, conclude, Select)	Describe & Inference based on Classification process **Evaluate -Present and defend opinions by making judgments about information, validity of ideas, or quality of work based on a set of criteria**
CSC504.4	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 methods Application by integrating principles of Association Rules& implementation
CSC504.5	Describe complex data types wrt to special and web mining	**Comprehension** (explains, gives examples, shows relationship of)	Understanding architecture of variant DM systems
CSC504.6	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
CSL503.1	1. Design data warehouse and perform various OLAP operations.
CSL503.2	2. Implement data mining algorithms like classification.
CSL503.3	3. Implement clustering algorithms on a given set of data sample.
CSL503.4	4. Implement Association rule mining & web mining algorithm

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 (Engg Know)	PO2 (Ana)	PO3 (Design)	PO4 (investiga)	PO5 (tools)	PO6 (engg Soci)	PO7 (Env)	PO8 (Eth)	PO9 (ind Team)	PO10 (comm.)	PO11 (PM)	PO12 (life Long)
CSC504.1	3	-										
CSC504.2	3	3	3		2				3	3	3	3
CSC504.3	3	3	3		3				3	3		3
CSC504.4	3	3	3		3				3	3		3
CSC504.5	3	3	3		3				3	3		3
CSC504.6	3	3	3						3	3		3
Course To PO	3	3	3		2.75				3	3		3

CO	PSO1	PSO2
CSC504.1	3	
CSC504.2	3	
CSC504.3	3	
CSC504.4	3	1
CSC504.5	3	
CSC504.6	3	1
Course to PSO	3	1

CO Assessment Tools:

CSC504.1 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)	
CSC504.1= 0.8* CSC504.1dm + 0.2* CSC504.1idm	

CSC504.2 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)	
CSC504.2= 0.8* CSC504.2dm + 0.2* CSC504.2idm	

CSC504.3 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)	
CSC504.3= 0.8* CSC504.3dm + 0.2* CSC504.3idm	

CSC504.4 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)**	
CSC504.4= 0.8* CSC504.4dm + 0.2* CSC504.4idm	

CSC504.5 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)**	
CSC504.5= 0.8* CSC504.5dm + 0.2* CSC504.5idm	

CSC504.6 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)**	
CSC504.6= 0.8* CSC504.6dm + 0.2* CSC504.6idm	

Rubrics for the Lab Experiments:

Rubrics for Experiment Grading of Expt.

Expt1-2

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)**

Expt3-10

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)**

Assignments4

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(3)	Understands the problem (2)	Understands enough to solve part of the problem or to get part of the solution (1)
Describe complex data types (4)	Explains how the complex data types are appropriate in given problem (4)	Applies complex data types in given problem (3)	Applies some complex data types given problem (2)
Timeline (3)	**Understands the Problem(3)**	**Applies Appropriate Symbols (4)**	**Total (10)**

FR. Conceicao Rodrigues College Of Engineering

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

Department of Computer Engineering
(2022-2023)
Lesson Plan

CLASS-SEM: T.E.- V **Credits-4**
SUBJECT: ***Data warehouse and mining (DWM)*** **SUBJECT CODE- CSC503**

Professor: Dr. Sujata P. Deshmukh

Modes of Content Delivery:

i	Google classroom-online-meet 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	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;	19/07/2022	19/07/2022	Classroom Teaching -PPT
2	1.2	Introduction to Data Warehouse, Data warehouse architecture, Data warehouse versus Data Marts, Different approaches and layered Architecture	20/07/2022	20/07/2022	Classroom Teaching
3	1.3	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.	22/07/2022	22/07/2022	Classroom Teaching
4	1.4	OLTP versus OLAP, OLAP operations: Slice, Dice, Rollup, Drilldown and Pivot. OLAP	25/07/2022	25/07/2022	Classroom Teaching

		Models: MOLAP, ROLAP, HOLAP, DOLAP, Definition of Schema using DMQL, Examples on OLAP			
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?	27/07/2022	27/07/2022	Classroom Teaching
6	1.6	Keys in DW, Snowflake Schema, Fact Constellation Schema or Families of Star.	29/07/2022	29/07/2022	Classroom Teaching
7	1.7	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	01/08/2022	26/07/2022	Classroom Teaching
8	1.8	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	02/07/2022	28/07/2022	Classroom Teaching
9	2.1	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, Issues in Data Mining, Applications of Data Mining, Comparison of DM with ML and BDA	03/08/2022	30/07/2022	Classroom Teaching

10	2.2	Data Exploration: Types of Attributes, Statistical Description of Data,	09/08/22	09/08/22	Classroom Teaching
11	2.3	Data Visualization; Measuring similarity and dissimilarity	10/08/22	11/08/22	Classroom Teaching
12	2.4	Data Preprocessing: Descriptive data summarization, Cleaning, Integration & transformation, Data reduction,	12/08/22	12/08/22	Classroom Teaching
13	2.5	Data exploration: Types of Attributes; Statistical Description of Data;	23/08/22	09/08/22	Classroom Teaching
14	2.6	Why Preprocessing? Data Cleaning; Data Integration; Data Reduction: Attribute subset selection, Histograms, Clustering and Sampling	24/08/22	24/08/22	Classroom Teaching
15	2.7	Data Transformation & Data Discretization: Normalization	25/08/22	25/08/22	Classroom Teaching
16	2.8	Data Discretization and Concept hierarchy generation-. Binning, Histogram Analysis and Concept hierarchy generation and description	29/08/22	29/08/22	Classroom Teaching
17	3.1	Basic Concepts, Decision Tree Induction: Attribute Selection Measures, Tree pruning.	30/08/22 (mid term break-31-04)	30/08/22	Classroom Teaching
18	3.2	Decision Tree Induction and problems	08/09/22	08/09/22	Classroom Teaching
20	3.3	Naïve Bayesian Classification and problems	13/09/22	13/09/22	Classroom Teaching
20	3.4	Naïve Bayesian Classification, Accuracy and Error measures,	14/09/22	14/09/22	Classroom Teaching
22	3.5	Model Evaluation & Selection: Accuracy and Error measures, Holdout, Random Sampling, Cross Validation, Bootstrap.	16/09/22	16/09/22	Classroom Teaching
22	3.6	Problems based on classification	20/09/22	06/09/22	Classroom Teaching
23	4.1	What is clustering? Types of data, Partitioning Methods (K-Means)	21/09/22	08/09/22	Classroom Teaching

24	4.2	Partitioning Methods (KMedoids)	22/09/22	10/09/22	Classroom Teaching
25	4.3	Hierarchical Methods(Agglomerative , Divisive, BRICH),	27/09/22	15/09/22	Classroom Teaching
26	4.4	Hierarchical Methods(Agglomerative , Divisive, BRICH),	28/09/22	17/09/22	Classroom Teaching
27	4.5	Problems based on clustering	29/09/22	20/09/22	Classroom Teaching
28	4.6	Comparison of clustering with classification and more problems	22/09/22	22/09/22	Classroom Teaching
29	5.1	Market Basket Analysis, Frequent Item sets, Closed Item sets, and Association Rule, Frequent Pattern Mining,	24/09/22	24/09/22	Open Discussion with Students with teaching
30	5.2	The Apriori Algorithm for finding Frequent Itemsets Using Candidate Generation	24/09/22	24/09/22	Google classroom-online-meet Teaching, Lab Experiment
31	5.3	Generating Association Rules from Frequent Itemsets, Improving the Efficiency of Apriori,	24/09/22	24/09/22	Google classroom-online-meet Teaching
32	5.4	A pattern growth approach for mining Frequent Itemsets; Mining Frequent itemsets using vertical data formats; Mining closed and maximal patterns;	27/09/22	27/09/22	Google classroom-online-meet Teaching
33	5.5	From Association Mining to Correlation Analysis, Pattern Evaluation Measures;	28/09/22	28/09/22	Google classroom-online-meet Teaching
34	5.6	Introduction to Mining Multilevel Association Rules and Multidimensional Association Rules;	29/09/22	29/09/22	Google classroom-online-meet Teaching
35	6.1	Introduction, Web Content Mining: Crawlers, Harvest System, Virtual Web View, Personalization,	04/10/22	04/10/22	Google classroom-online-meet Teaching
36	6.2	Web Structure Mining: Page Rank and its Problems, Clever,	05/10/22	05/10/22	Google classroom-online-meet Teaching

37	**6.3**	Web Structure Mining: Page Rank and its Problems, Clever	07/10/22	07/10/22	Google classroom-online-meet Teaching, Lab Experiment
39	6.4	Web Taxonomy	11/10/22	11/10/22	Google classroom-online-meet Teaching
40	6.5	Content beyond syllabus-Moving from data mining to BDA Analytics	12/10/22	12/10/22	Google classroom-online-meet Teaching
EVALUATION PRESENTATION and submission					
41	r.1	BATCH1	31/10/22	31/10/22	Evaluation Presentation
42	r.2	BATCH2	31/10/22	31/10/22	Evaluation Presentation
43	r.3	BATCH3	31/10/22	31/10/22	Evaluation Presentation
44	r.4	BATCH4	31/10/22	31/10/22	Evaluation Presentation

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)
NPTEL :: Computer Science and Engineering - NOC:Data Mining
Data Mining | Sloan School of Management | MIT OpenCourseWare

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 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	**CSC504.1, CSC504.2**	**CSL503.1**
	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	**CSC504.1, CSC504.2**	**CSL503.1**
3	To perform data exploration and data cleaning in python /R	**CSC504.3, CSC504.4**	**CSL503.2, CSL503.3.**
4	Implementation of Association Rule Mining algorithm (Apriori in java/python).	**CSC504.3, CSC504.4**	**CSL503.2, CSL503.3.**
5	Implementation of Clustering algorithm (K-means in java/python).	**CSC504.3, CSC504.4**	**CSL503.2, CSL503.3.**
6	Implementation of Linear Regression.	**CSC504.3, CSC504.4**	**CSL503.2, CSL503.3.**
7	Use WEKA to implement classification (Part1, Part2)	**CSC504.3, CSC504.4**	**CSL503.2, CSL503.3**
8	Implement spatial and web mining algorithms.	**CSC504.5, CSC504.6**	**CSL503.4,**
9	Use WEKA to implement Association Mining and Clustering algorithm	**CSC504.5, CSC504.6**	**CSL503.3**
10	Mini project/Case study	**CSC504.6**	**All**
Assignments			
11	**Assignment1-Data ware housing**	**CSC504.1, CSC504.2**	
12	**Assignment2-Data Exploration**	**CSC504.3, CSC504.4**	
13	**Assignment3-DM Algorithms**	**CSC504.3, CSC504.4**	
14	**Assignment4-Social mining**	**CSC504.5**	

Lab Plan

Sr. no.	No. Title of Experiments	Schedule
1	Build Data Warehouse/Data Mart for a given problem statement	Week1
2	To perform various OLAP operations such as slice, dice, drilldown, rollup, pivot	Week2
3	To perform data exploration and data cleaning in python /R	Week3
4	Implementation of Association Rule Mining algorithm (Apriori)	Week4
5	Perform data Pre-processing task and Demonstrate performing Classification, Clustering, Association algorithm on data sets using data mining tool (WEKA/R tool)	Week3
6	Implementation of Clustering algorithm (K-means/K-medoids)	Week5
7	Implementation of any Hierarchical Clustering method	Week5
8	Implementation of Bayesian algorithm	Week6
9	Implementation of Page rank/HITS algorithm	Week6
1	Implementation of Text mining	Week7
11	Innovative experiment-ML/DEEP learning Algorithms.(NN---)	Week8
12	Mini project/Case study(submission)	Week9