Program Structure for Third Year Computer Engineering

UNIVERSITY OF MUMBAI (With Effect from 2021-2022)

Semester V

Course Code	Course Name	Teaching Scheme (Contact Hours)		Credits Assigned		
		Theory	Pract.	Theory	Pract.	Total
CSC501	Theoretical Computer Science	3	--	3	--	3
CSC502	Software Engineering	3	--	3		3
CSC503	Computer Network	3	--	3	--	3
CSC504	Data Warehousing & Mining	3	--	3	--	3
CSDLO501x	Department Level Optional Course- 1	3	--	3	--	3
CSL501	Software Engineering Lab	--	2	--	1	1
CSL502	Computer Network Lab	--	2	--	1	1
CSL503	Data Warehousing & Mining Lab	--	2	--	1	1
CSL504	Business Comm. & Ethics II	--	2*+2	--	2	2
CSM501	Mini Project: 2 A	--	4$	--	2	2
Total		**15**	**14**	**15**	**07**	**22**

Course Code	Course Name	Examination Scheme							
		Theory					Term Work	Pract &oral	Total
		Internal Assessment			End Sem Exam	Exam. Duration (in Hrs)			
		Test 1	Test 2	Avg					
CSC501	Theoretical Computer Science	20	20	20	80	3	25	--	125
CSC502	Software Engineering	20	20	20	80	3	--	--	100
CSC503	Computer Network	20	20	20	80	3	--	--	100
CSC504	Data Warehousing & Mining	20	20	20	80	3	--	--	100
CSDLO501x	Department Level Optional Course -1	20	20	20	80	3	--	--	100
CSL501	Software Engineering Lab	--	--	--	--	--	25	25	50
CSL502	Computer Network Lab	--	--	--	--	--	25	25	50
CSL503	Data Warehousing & Mining Lab	--	--	--	--	--	25	25	50
CSL504	Business Comm. & Ethics II	--	--	--	--	--	50	--	50
CSM501	Mini Project : 2A	--	--	--	--	--	25	25	50
Total		**--**	**--**	**100**	**400**	**--**	**175**	**100**	**775**

* Theory class to be conducted for full class and $ indicates workload of Learner (Not Faculty), students can form groups with minimum 2(Two) and not more than 4(Four). Faculty Load: 1hour per week per four groups.

Program Structure for Computer Engineering

UNIVERSITY OF MUMBAI (With Effect from 2021-2022)

Department Optional Courses

Department Level Optional Courses	Semester	Code & Course
Department Level Optional Course -1	V	CSDLO5011: Probabilistic Graphical Models CSDLO5012: Internet Programming CSDLO5013: Advance Database Management System

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.

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, Naïve 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.

Assessment:	
Internal Assessment:	
Assessment consists of two class tests of 20 marks each. The first-class test is to be conducted when approx. 40% syllabus is completed and second-class test when additional 40% syllabus is completed. Duration of each test shall be one hour.	
End Semester Theory Examination:	
1	Question paper will comprise of total six questions.
2	All question carries equal marks
3	Questions will be mixed in nature (for example, If Q.2 part (a) from module 3 then part (b) can be from any module other than module 3)
4	Only Four questions need to be solved.
5	In question paper weightage of each module will be proportional to the number of respective lecture hours as mentioned in the syllabus.
Useful Links	
1	https://onlinecourses.nptel.ac.in/noc20_cs12/preview
2	https://www.coursera.org/specializations/data-mining

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 (Apriori)
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