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