

Course code	Course Name	L-T-P - Credits	Year of Introduction
IT304	Data Warehousing and Mining	3-0-0-3	2016
Prerequisite: CS208 Principles of data base design			
Course Objectives - To understand Data Mining, its origin, taxonomy and applications - To understand types of data and to improve the quality of data and efficiency and the ease of the mining process. - To understand the supervised learning that is Classification, its applications and approaches - To understand how to identify associations among objects and to learn various algorithms to find them - To understand methods and need for finding complex Association Rules - To learn the unsupervised learning to identify the relation among the objects and to understand applications and algorithms for Clustering			
Syllabus Data Mining, Applications, Data Mining Models, Data Warehousing and OLAP, Challenges, Tools, Data Mining Principles, Data Preprocessing: Data Preprocessing Concepts, Data Visualization, Data Sets and Their Significance, Classification Models, Multi Resolution Spatial Data Mining, Classifiers, Association Rules Mining, Cluster Analysis, Practical Data Mining Tools, Advanced Data Mining Techniques, Web Mining, Text Mining, CRM Applications and Data Mining, Data warehousing.			
Expected outcome . The student will understand the concept of data mining, association rule mining and data classification methods			
Text Book: - Jaiwei Han, Micheline Kamber, "Data Mining Concepts and Techniques", Elsevier, 2006. M. Sudeep Elayidom, "Data Mining and Warehousing", 1st Edition, 2015 Cengage Learning India Pvt. Ltd.			
References: - Pang-Ning Tan, Michael Steinbach, "Introduction to Data Mining", Addison Wesley, 2006. - Dunham M H, "Data Mining: Introductory and Advanced Topics", Pearson Education, New Delhi, 2003. - Mehmed Kantardzic, "Data Mining Concepts, Methods and Algorithms", John Wiley and Sons, USA, 2003.			
Course Plan			
Module	Contents	Hours	Sem. Exam Marks
I	Data Mining: Concepts: Concepts, Data Mining Applications, Data Mining Stages, Data Mining Models, Data Warehousing and OLAP, Need for Data Warehousing, Challenges, Application of Data Mining Principles, Machine Learning and Statistics, Ethics of Data Mining, Popular Tools. OLTP Vs DWH, Applications of DWH	8	15%
II	Data Preprocessing: Data Preprocessing Concepts, Data Cleaning, Handling Missing Data, Data Transformation and Discretization, Data Visualization. UCI Data Sets and Their Significance	6	15%
FIRST INTERNAL EXAMINATION			
III	Classification Models: Introduction to Classification Models,	6	15%

	Decision Tree, Neural Networks		
IV	Naive Bayes Classifier, Support Vector Machines. Prediction Models, Issues regarding classification and prediction.	7	15%
SECOND INTERNAL EXAMINATION			
V	Association Rules Mining: Concepts, Apriori Algorithm. Cluster Analysis: Introduction, Concepts, K-Means Clustering, Density-Based Clustering, Weighted Graph Partitioning, Hypergraph Partitioning,	8	20%
VI	Practical Data Mining Tools: Weka, R Package for Data Mining. Advanced Data Mining Techniques: Introduction, Web Mining- Web Content Mining, Web Structure Mining, Web Usage Mining. Text Mining, CRM Applications and Data Mining, CRM Data Mining Models. Data Warehousing with Oracle BI	7	20%
END SEMESTER EXAM			

QUESTION PAPER PATTERN

Maximum Marks: 100

Exam Duration: 3 hours

The question paper shall consist of Part A, Part B and Part C.

Part A shall consist of three questions of 15 marks each uniformly covering Modules I and II. The student has to answer any two questions ($15 \times 2 = 30$ marks).

Part B shall consist of three questions of 15 marks each uniformly covering Modules III and IV. The student has to answer any two questions ($15 \times 2 = 30$ marks).

Part C shall consist of three questions of 20 marks each uniformly covering Modules V and VI. The student has to answer any two questions ($20 \times 2 = 40$ marks).

Note : Each question can have a maximum of 4 subparts, if needed