Devi Ahilya University, Indore, India Institute of Engineering & Technology				IV Year B.E. (INFORMATION TECHNOLOGY)			
Subject Code & Name	**Instructions Hours per Week**			**Credits**			
8 ITRC 1 **Data warehousing and Mining**	**L**	**T**	**P**	**L**	**T**	**P**	**Total**
	3	**1**	**2**	**3**	**1**	**1**	**5**
Duration of Theory Paper: 3 Hours							

Learning Objectives:

- To learn the latest development of Data warehousing & data mining concepts and techniques.
- To understand and apply theories and wide range of data mining algorithms.
- To develop skills for using recent data mining software, including WEKA, SPSS, or the R language to solve practical problems in a variety of disciplines.
- To develop skill in selecting the appropriate data mining algorithm for solving practical problems.

Pre-requisites : Basic knowledge of a programming language and Basic knowledge of probabilities and statistics is required.

Course Outcomes (CO) and Program Outcomes (PO) Mapping

CO No.	Course Outcome	Program Outcomes (PO)
CO1	Theoretical Foundations - Acquire a solid understanding of basics of data mining	PO1, PO2, PO12
CO2	Learn different techniques of principles of data mining algorithms	PO2, PO4, PO5, PO10
CO3	Problem identification and solution – understanding of problems and designing of data mining	PO2, PO4, PO5
CO4	Problem-Solving and analysis using data mining software's	PO1, PO2, PO4, PO5, PO9, PO10, PO11, PO12
CO5	Learning and understanding of advance data mining tools	PO3, PO4, PO9, PO12

CO-PO Relationship Matrix

CO	PO -1	PO -2	PO -3	PO -4	PO -5	PO -6	PO -7	PO -8	PO -9	PO -10	PO -11	PO -12
CO 1	3	2	-	-	-	-	-	-	-	-	-	2
CO 2	-	2	-	2	3	-	-	-	-	2	-	-

CO	PO -1	PO -2	PO -3	PO -4	PO -5	PO -6	PO -7	PO -8	PO -9	PO -10	PO -11	PO -12
CO 3	-	3	-	3	3	-	-	-	-	-	-	-
CO 4	3	3	-	3	3	-	-	-	2	2	2	3
CO 5	-	2	3	3	-	-	-	-	3	-	-	3

COURSE CONTENTS

UNIT-I

Data Warehouse: Introduction to Data Warehouse, Differences between OLAP and OLTP. Data Warehouse characteristics, Data Warehouse Architecture and its components, Extraction-Transformation Loading, Data Modeling, Schema Design, star and snow-Flake Schema, Fact Constellation, Fact Table, Fully Addictive, Semi-Addictive, Non-Addictive Measures; Dimension Table characteristics; OLAP cube, OLAP Operations, OLAP Server Architecture- ROLAP, MOLAP and HOLAP.

UNIT-II

Introduction to Data Mining: Introduction to Data Mining, Definition, KDD, Challenges, Data Mining Tasks, Data Preprocessing- Data Cleaning, Missing Data, Dimensionality Reduction, Feature Subset Selection, Discretization and Binarization, Data Transformation; Measures of similarity and dissimilarity-Basics.

UNIT-III

Classification: Problem definition, General Approaches to solving a classification problem, Issues regarding classification, Classification techniques, Decision trees-Decision Tree Construction, Measures for Selecting the Best split, Algorithm for Decision tree Induction, Naïve-Bayes classification, Bayesian Rule-based classification, Classification by back-propagation.

UNIT-IV

Association Rules: Introduction to Association Rule, Frequent Item Set Generation, The APRIORI Principle, Support and Confidence Measures, Association Rule Generation, APRIORI Algorithm, The Partition Algorithms, FP-Growth Algorithms, Frequent Item Set-Maximal Frequent Item Set, Closed Frequent Item Set.

UNIT-V

Clustering: Clustering overview, Evaluation of clustering algorithms, Partitioning clustering K-Means Algorithm, K-Means Additional Issues, PAM Algorithm, Hierarchical Clustering-Algorithm Agglomerative Methods and Divisive Methods, Basic Agglomerative Hierarchical Clustering Algorithm, Specific techniques, Key Issues in Hierarchical Clustering, Strengths and weakness, Outlier Detection.

Learning Outcomes:

Upon Completing the Course, students will have knowledge of Data Warehousing and various Data Mining Algorithm useful for solving the real world problems.

BOOKS RECOMMENDED:

[1] Sam Anahory and Dennis Murray, Data Warehousing in the real World, Pearson Education
Asia, 2000.
[2] P. Tan, M. Steinback and V. Kumar, Introduction to Data Mining, Addison Wesley, Second Edition, 2016.
[3] J. W. Han and M. Kamber, Data Mining: Concepts and Techniques, Morgan Kaufmann Publishers, San Francisco, CA, Third Edition, 2011.

List of Practical Assignment:

- Problem based on different Data Mining algorithm
- Works on different Data Mining Algorithm
- Case Study on different data sets