Course Code: IT-32
Course Name: Data Warehousing and Data Mining

Credit Scheme			Evaluation Scheme				
Lecture	Practical	Credit	Internal			External	Total
			Written	Practical	Tutorial		
3 Hrs./Week	-	3	10	10	5	50	75

Course Description:

Course Objectives:

1. To introduce the concepts, techniques and applications of data warehousing and data mining.
2. To understand how to Preprocess, understand and analyze various kinds of data
3. To Study data warehouse Concepts, architectures, OLAP and the project planning aspects in building a data warehouse
4. To enable students to understand and implement various techniques of association, classification and clustering in data mining
5. To enable students to understand and implement the concepts of Web mining and Text Mining in data mining

Course Outcomes:

Student will be able to

CO1: Understand Data warehouse concepts, architecture and models (Understand)
CO2: Learn and understand techniques of preprocessing on various kinds of data (Understand)
CO3: Apply association Mining and Classification Techniques on Data Sets (Apply)
CO4: Apply Clustering Techniques and Web Mining on Data Sets (Apply)
CO5: Understand other approaches of Data mining (Understand)

Course Structure:

Unit No.	Topics Details	Weightage in %	No of Sessions
1	1. Data Warehouse Fundamentals 1.1. Introduction to Data Warehouse, OLTP Systems; Differences between OLTP Systems and Data Warehouse: 1.2. Characteristics of Data Warehouse; Functionality of Data Warehouse: 1.3. Advantages and Applications of Data Warehouse; Top- Down and Bottom-Up Development Methodology: 1.4. Tools for Data warehouse development: Data	15	6

	Warehouse Types 1.5. Planning and Project Management in constructing Data warehouse: Data Warehouse Project; 1.6. Data Warehouse development Life Cycle, Kimball Lifecycle Diagram		
2	2. Data Warehouse Architecture 2.1. Introductions, Components of Data warehouse Architecture 2.2. Technical Architectures; Federated Data Warehouse Architecture: Tool selection; 2.3. Dimensional Modeling: E-R Modeling VS Dimensional Modeling 2.4. Data Warehouse Schemas; Star Schema, Inside Dimensional Table, Inside Fact Table, Fact Less Fact Table, Granularity, Star Schema Keys: Snowflake Schema, Fact Constellation Schema 2.5. Introduction to Metadata : Categorizing Metadata: 2.6. Metadata management in practice; Meta data requirements gathering, Metadata classification, Meta data collection strategies, Tools for Metadata Management	15	6
3	3. Data Preprocessing and ETL 3.1. Data Pre-processing: Data Cleaning tasks 3.2. Data Integration and Data Reduction 3.3. Discretization and Concept Hierarchy Generation 3.4. Data Transformation; Basic Tasks in Transformation, Major Data Transformation Types 3.5. Introduction to ETL(Extract, Transform and Load) 3.6. ETL requirements and steps: Data Extraction; Extraction Methods, Logical Extraction Methods, Physical Extraction Methods 3.7. Data loading; Data Loading Techniques, ETL Tools	20	7
4	4. Data Warehouse & OLAP: 4.1. Introduction: What is OLAP?; Characteristics of OLAP, 4.2. Steps in the OLAP Creation Process, OLAP operations, Advantages of OLAP: Multidimensional Data: 4.3. OLAP Architectures; MOLAP, ROLAP, HOLAP: 4.4. Data Warehouse and OLAP: Hypercube & Multicubes	10	5
5	5. Introduction to Data Mining: 5.1. Introduction and Scope of Data Mining 5.2. How does Data Mining Works, Predictive Modeling 5.3. Data Mining and Data Warehousing 5.4. Architecture for Data Mining	5	4

	5.5. Profitable Applications: Data Mining Tools:		
6	6. Data Mining Techniques 6.1. An Overview: Introduction, Data Mining, Data Mining Versus Database Management System, 6.2. Data Mining Techniques- Association rules(Apriori, FP Tree algorithms) 6.3. Classification (Decision Tree induction, Bayesian classification, SVM, KNN) 6.4. Clustering, Neural networks. 6.5. Evaluating Association rules , Classification model	15	6
7	7. Clustering 7.1. Introduction to Clustering, Cluster Analysis 7.2. Clustering Methods- K means, Hierarchical clustering, Agglomerative clustering, Divisive clustering, 7.3. clustering and segmentation software 7.4. Evaluating clusters 7.5. Data Mining trends and Applications	10	5
8	8. Web Mining 8.1. Introduction, Terminologies 8.2. Categories of Web Mining: Web Content Mining, Web Structure Mining, Web Usage Mining 8.3. Applications of Web Mining, and Agent based and Database approaches, Web mining Software/Tools. 8.4. Text Mining: process and types, steps in Text Mining, applications and tools of Text Mining 8.5. Data visualization, Dashboard- KPI, Business Intelligence and its future.	10	6
	Total:	**100**	**45**

List of Practicals (if any)

1. Creating a simple data warehouse & performing OLAP operations using simple tools
2. Extracting data from any Operational database (ETL) and performing pre-processing tasks
3. Performing association mining on large data sets and extracting best possible rules / a case study
4. Performing classification and evaluating the efficient model / a case study
5. A case study on finding efficient Clusters on very large set of documents data
6. A case study on Web mining and Text mining using software tools

Students may practice or implement Data warehouse, ETL & Data mining concepts on the following software/ tools (Students versions) at on premise / Cloud based platform

1) Data warehouse - My-SQL, MongoDB / Google BigQuery / Amazon Redshift / Talend
2) ETL Tools : Pentaho Kettle / Talend-Open Studio / Apache Kafka / Informatica Power Center
3) BI and Analytics tools : Python / XL-Miner, R-Studio / Rapid-Miner Studio
4) Visualization Tools : Tableau / Power-BI / Qlick sense

Course References:

Recommended Books:

Text Books:

1. Data Mining: Introductory and Advanced Topics, by Margaret Dunham, Pearson Education
2. Data Mining by Arun K. Pujari – University Press.

Reference Books:

1. DATAWAREHOUSING FUNDAMENTALS: A COMPREHENSIVE GUIDE FOR IT PROFESSIONALS, by Paulraj Poonniah, Latest Edition
2. Building the Data Warehouse, 3rd edition by W. H. Inmon
3. Data Mining concepts and Techniques by Jiawei Han, Micheline Kambler –Elsevier.
4. Data Mining practical Machine Learning Tools and Techniques by Ian H. Witten Eibe Frank Mark Hall - Elsevier publication
5. Introduction to Data Mining with Case Studies by G. K. Gupta, Prentice Hall

Web Reference:

1. www.ibm.com/in/en/
2. www.pentaho.com/
3. www.jaspersoft.com/
4. www.amazon.com/Data-Mining-Business-Intelligence-Applications
5. www.ibm.com/insights/in
6. www.sas.com
7. Weka– Data Mining with Open Source Machine Learning Software, www.cs.waikato.ac.nz/ml/weka.
8. https://cloud.google.com/bigquery/
9. https://www.rstudio.com/ 10.https://aws.amazon.com/redshift/