Subject: Data Science: Data Warehousing & Data Mining

Code:DIP15220

Credits-3 | Semester V

A. INTRODUCTION:

- Introduce students to the domain of Data Warehousing and Data Mining

B. COURSE OUTCOMES: At the end of the course, students will be able to

[CO1]Student will have general idea about Data Warehousing and Data Mining techniques, will be able to explore further and effectively use related tools.

C. ASSESSMENT PLAN:

Criteria	Description	Maximum Marks
Continuous Internal Assessment (CIA)	Internal Examination	20
	Attendance	5
	Assignment	5
End Semester Examination(ESE)	End Semester Examination	70
Total		100
Attendance	A minimum of 75% Attendance is required to be maintained by a student to be qualified for taking up the End Semester examination. The allowance of 25% includes all types of leaves including medical leaves.	

D. SYLLABUS

INTRODUCTION: Motivation, Importance, Definitions, Kind of Data, Data Mining Functionalities, Kinds of Patterns, Classification of Data Mining Systems.Data Mining Task Primitives, Integration of A Data Mining System with A Database or Data Warehouse System.Major Issues in Data Mining, Types of Data Sets and Attribute Values, Basic Statistical Descriptions of Data, Data Visualization, Measuring.Data Similarity. PREPROCESSING Data Quality, Major Tasks in Data Preprocessing, Data Reduction, Data Transformation and Data Discretization, Data Cleaning and Data Integration.

DATA WAREHOUSING AND ON-LINE ANALYTICAL PROCESSING: Data Warehouse basic concepts, Data Warehouse Modeling - Data Cube and OLAP, Data Warehouse Design and Usage.Data Warehouse Implementation, Data Generalization by Attribute-Oriented Induction, Data Cube Computation.

PATTERNS, ASSOCIATIONS AND CORRELATIONS: Mining Frequent Patterns, Associations and Correlations: Basic Concepts, Efficient and Scalable Frequent Itemset Mining Methods, Pattern Evaluation Methods, Applications of frequent pattern and associations.Frequent Patterns and Association Mining: A Road Map, Mining Various Kinds of Association Rules.Constraint-Based Frequent Pattern Mining Extended Applications of Frequent Patterns.

CLASSIFICATION: Basic Concepts, Decision Tree Induction, Bayesian Classification Methods, Rule-Based Classification, Model Evaluation and Selection.Techniques to Improve Classification Accuracy: Ensemble Methods, Handling Different Kinds of Cases in Classification, Classification by Neural Networks, Support Vector Machines, Pattern-Based Classification, Lazy Learners (or Learning from Your Neighbors).

CLUSTER ANALYSIS: Basic Concepts of Cluster Analysis, Clustering Structures, Major Clustering Approaches, Partitioning Methods, Hierarchical Methods, Density-Based Methods.Model-Based Clustering, Why outlieranalysis,Identifying and handling of outliers, Outlier Detection Techniques. WEB MINING: Basic concepts of web mining, different types of web mining, PAGE RANK Algorithm, HITS Algorithm.

E. TEXT BOOKS

T1. Jiawei Han, Micheline Kamber, Jian Pei, Data Mining: Concepts and Techniques, Elsevier

T2. Margaret H Dunham, Data Mining Introductory and Advanced Topics, Pearson Education

F. REFERENCE BOOKS

R1. Amitesh Sinha, Data Warehousing, Thomson Learning, India.

R2. Xingdong Wu, Vipin Kumar, the Top Ten Algorithms in Data Mining, CRC Press, UK.

G. Course Articulation Matrix: (Mapping of COs with POs)

CO	STATEMENT	CORRELATION WITH PROGRAM OUTCOMES										CORRELATION WITH PROGRAM SPECIFIC OUTCOMES	
		PO 1	PO 2	PO 3	PO 4	PO 5	PO 6	PO 7	PO 8	PO 9	PO 10	PSO 1	PSO 2
[CO1]	Student will have general idea about Data Warehousing and Data Mining techniques, will be able to explore further and effectively use related tools.				2							1	

1- Low Correlation; 2- Moderate Correlation; 3- Substantial Correlation