Subject Code	Subject Name	Category	L	T	P	C
AI19P62	DATA ANALYSIS AND DATA MINING	PE	2	0	2	3

Objectives:	
●	To learn the introduction of Data Warehouse and Data Mining.
●	To understand the concepts of clustering analysis.
●	To learn the basics of mining text data.
●	To acquire the basics of mining spatial data.
●	To study the basic concepts of mining web data.

UNIT-I	INTRODUCTION TO DATA WAREHOUSE AND DATA MINING	6
Data Warehouse: Characteristics of Data Warehouse - Data Warehouse Components - Designing the Data Warehouse - Data Warehouse Architecture - Getting Heterogeneous Data into the Warehouse - Getting Multidimensional Data out of the Warehouse. **Data Mining:** Definition – Architecture – data mining: on what kind of data? - Data mining functionalities. (T2: Chapter – 1 and 2)		
UNIT-II	**CLUSTERING ANALYSIS**	**6**
Introduction – Feature selection for clustering – Representative based algorithms – Hierarchical clustering algorithms – probabilistic model based algorithms – Grid based and density based algorithms – Graph based algorithms – non negative matrix factorization – clustering validation. (T1: Chapter – 6)		
UNIT-III	**MINING TEXT DATA**	**6**
Document Preparation and Similarity computation – Specialized clustering methods for text – topic modeling – Specialized Classification Methods for Text – Novelty and First Story Detection. (T1: Chapter –13)		
UNIT-IV	**MINING SPATIAL DATA**	**6**
Mining with Contextual Spatial Attributes – Trajectory mining – Equivalence of Trajectories and Multivariate Time Series – Converting Trajectories to Multi dimensional Data – Trajectory Pattern Mining – Trajectory Clustering – Trajectory Outlier Detection – Trajectory Classification. (T1: Chapter –16)		
UNIT-V	**MINING WEB DATA**	**6**
Web crawling and Resource Discovery – Search Engine Indexing and Query Processing – Ranking Algorithm – Recommender Systems – Web Usage Mining. (T1: Chapter –18)		
	Contact Hours : 30	

List of Experiments				
	In H_2O implement the following			
1	Perform the basic pre-processing operations on data relation such as removing an attribute and filter attribute bank data			
2	To predict the Numerical Values in the given Data Set is using Regression Methods.			
3	To predict with the smallest total error using rules based on One attribute			
4	To understand the theoretical aspects and build a hierarchy of clusters using hierarchical clustering techniques			
5	To Demonstrate Clustering features in Large Databases with noise			
6	Generate association rule for the credit card promotion dataset using a priory algorithm with the support range 40% to 100% confidence as 10% incremental decrease as 5% and generate 6 rules			
		Contact Hours	**:**	**30**
		Total Contact Hours	**:**	**60**

Course Outcomes: On completion of the course, the students will be able to	
●	Explain the introduction of Data Warehouse and Data Mining.
●	Apply the concepts of clustering analysis.
●	Analyze the basics of mining text data.
●	Integrate the concepts of mining spatial data.
●	Demonstrate the basic concepts of mining web data.

Text Books:	
1	Charu C. Aggarwal, Data Mining: The Textbook, Springer 2015 Edition, Kindle Edition.
2	Sartaj Singh “Data Warehousing and Data Mining”, Lovely Professional University, Phagwara.

Reference Books:	
1	Usama M. Fayyad, Gregory Piatetsky - Shapiro, Padhrai Smyth, and Ramasamy Uthurusamy, "Advances In Knowledge Discovery And Data Mining", The M.I.T Press, 1996.
2	N. J. Nilsson, "Principles of Artificial Intelligence", Narosa Publishing House, 1980.

CO - PO – PSO matrices of course

PO/PSO 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	PSO 1	PSO 2	PSO 3
AI19P62.1	3	3	-	-	-	-	-	-	-	-	-	-	3	-	1
AI19P62.2	3	3	2	-	-	-	-	-	-	-	-	-	3	-	1
AI19P62.3	-	2	3	3	2	-	3	2	3	-	3	3	-	3	3
AI19P62.4	-	3	3	3	2	-	3	3	3	-	3	3	-	3	3
AI19P62.5	-	3	3	3	3	-	3	3	3	-	3	3	-	3	3
Average	1.2	2.8	2.2	1.8	1.4	-	1.8	1.6	1.8	-	1.8	1.8	1.2	1.8	2.2

Correlation levels 1, 2 or 3 are as defined below:
1: Slight (Low) 2: Moderate (Medium) 3: Substantial (High)
No correlation: “-”