DATA PREPARATION AND ANALYSIS LABORATORY

II Semester: CSE

Course Code	Category	Hours / Week			Credits	Maximum Marks		
		L	T	P	C	CIA	SEE	Total
BCSB20	Core	0	0	4	2	30	70	100
Contact Classes: Nil	Total Tutorials: Nil		Total Practical Classes: 36			Total Classes: 36		

I. COURSE OVERVIEW:

In this laboratory students will develop a solid understanding of data pre-processing, cluster analysis, genetic algorithms, data transformation, and hierarchical clustering. These skills will enable them to effectively prepare and analyze data, derive valuable insights, and make data-driven decisions in various domains.

II. OBJECTIVES

The students will try to learn:

I. The pre-processing method for multi-dimensional data
II. The Practice on data cleaning mechanisms
III. The various data exploratory analysis
IV. The visualizations for clusters or partitions

COURSE OUTCOMES:

After successful completion of the course, students should be able to:

CO 1	**Apply** pre-processing techniques for cleaning data.	Apply
CO 2	**Develop** a cluster models for categorizing data a cluster models for categorizing data	Create
CO 3	**Apply** genetic algorithms to optimization problems	Apply
CO 4	**Implement** data transformation techniques on spatial, time series and numerical data	Apply
CO 5	**Choose** clustering algorithm for implementing hierarchical clustering	Remember

IV. SYLLABUS

LIST OF EXPERIMENTS

Week-1 DATA PRE-PROCESSING AND DATA CUBE

Data preprocessing methods on student and labor datasets
Implement data cube for data warehouse on 3-dimensional data

Week-2 DATA CLEANING

Implement various missing handling mechanism, Implement various noisy handling mechanisms

Week-3 EXPLORATORY ANALYSIS

Develop k-means and MST based clustering techniques, Develop the methodology for assessment of clusters for given dataset

Week-4 ASSOCIATION ANALYSIS

Design algorithms for association rule mining algorithms

Week-5	**HYPTOTHYSIS GENERATION**
Derive the hypothesis for association rules to discovery of strong association rules; Use confidence and support thresholds.	
Week-6	**TRANSFORMATION TECHNIQUES**
Construct Haar wavelet transformation for numerical data, Construct principal component analysis (PCA) for 5-dimensional data.	
Week-7	**DATA VISUALIZATION**
Implement binning visualizations for any real time dataset, Implement linear regression techniques	
Week-8	**CLUSTERS ASSESSMENT**
Visualize the clusters for any synthetic dataset,Implement the program for converting the clusters into histograms	
Week-9	**HIERARCHICAL CLUSTERING**
Write a program to implement agglomerative clustering technique ,Write a program to implement divisive hierarchical clustering technique	
Week-10	**SCALABILITY ALGORITHMS**
Develop scalable clustering algorithms ,Develop scalable a priori algorithm	
Reference Books:	
1. Sinan Ozdemir, “Principles of Data Science”, Packt Publishers, 2016.	
Web References:	
1. https://paginas.fe.up.pt/~ec/files_1112/week_03_Data_Preparation.pdf 2. https://socialresearchmethods.net/kb/statprep.php 3. https://www.quest.com/solutions/data-preparation-and-analysis/	
SOFTWARE AND HARDWARE REQUIREMENTS FOR 18 STUDENTS: **SOFTWARE:** Open source Weka 3.8, Python **HARDWARE:** 18 numbers of Intel Desktop Computers with 4 GB RAM	