

DATA PREPARATION AND ANALYSIS LABORATORY

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

COURSE OBJECTIVES:

The course should enable the students to:

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

COURSE OUTCOMES(CO_s):

CO 1: Evaluate the data pre-processing methods by implementing data cube for data warehouse on 3- dimensional data.

CO 2: Impart the architectural concepts of data warehousing for handling consistent data.

CO 3: Construct various methods for data pre-processing techniques.

CO 4: Visualize the data interpretations for real time data sets.

CO 5: Implement best practices and techniques for data preparation efficiently.

COURSE LEARNING OUTCOMES(CLO_s):

1. Analyze various data preprocessing methods on different data sets. Analyze the role of Information Systems in an organization.
2. Describe the fundamentals of data cleaning and implement various missing and noisy handling mechanisms. Apply the basic concepts of MS Excel –worksheet management, cell referencing and range formulas.
3. Gain knowledge to identify appropriate clustering techniques, and develop clusters for given dataset.
4. Identify the association rule mining techniques, based on the requirements of the problem.
5. Derive the hypothesis for association rules to discovery of strong association rules.
6. Understand the concept of transformation techniques for numerical datasets.
7. Learn various data visualization techniques and use them to solve statistical problems.
8. Visualize the cluster datasets and convert the clusters into histograms.
9. Understand hierarchical clustering and solve the problem for the given related datasets.
10. Understand how scalability clustering done for apriori algorithm.

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 mechanisms ,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