

DATA PREPARATION AND ANALYSIS

II Semester: CSE								
Course Code	Category	Hours / Week			Credits	Maximum Marks		
BCSC15	Elective	L	T	P	C	CIA	SEE	Total
		3	0	0	3	30	70	100
Contact Classes: 45	Total Tutorials: Nil	Total Practical Classes: Nil			Total Classes: 45			
I. COURSE OVERVIEW:								
This course introduces the process of cleaning and transforming raw data before it is processed and analyzed. Students gain knowledge regarding the process involving reformatting the data, making adjustments to the data, and combining data sets to enrich the data.								
II. COURSE OBJECTIVES:								
The students will try to learn:								
I. The data for analysis and develop meaningful data visualizations.								
II. How to perform exploratory analysis with the help of descriptive and perceptive statistics.								
III. Illustrate data visualizations for understanding and presenting.								
III. COURSE OUTCOMES:								
After successful completion of the course, students should be able to								
CO 1	Select appropriate data preparation techniques to transform raw data into a standard format.						Understand	
CO 2	Apply data cleaning methods on real-time data for usage of data in analytics						Understand	
CO 3	Make use of statistical methods for performing exploratory analysis						Apply	
CO 4	Infer complex data models with respect to time series and geographical data mining						Apply	
CO 5	Identify the effective visualization techniques for data communication						Understand	
IV. SYLLABUS:								
MODULE - I DATA GATHERING AND PREPARATION (09)								
Data formats, parsing and transformation, Scalability and real-time issues.								
MODULE-II: DATA CLEANING (09)								
Consistency checking, Heterogeneous and missing data, Data Transformation and segmentation.								
MODULE-III: EXPLORATORY ANALYSIS (09)								
Descriptive and comparative statistics, Clustering and association, Hypothesis generation.								
MODULE-IV: VISUALIZATION -1 (09)								
Designing visualizations, Time series, Geo located data, Correlations and connections.								
MODULE-V: VISUALIZATION -2 (09)								
Hierarchies and networks, interactivity.								
V. TEXT BOOKS:								
1. GlennJ. Myatt, "Making sense of Data: A practical Guide to Exploratory Data Analysis and Data Mining", Wiley publishers, 2007.								
VI REFERENCE BOOKS:								

1. D. Pyle, Data Preparation for Data Mining. Morgan Kaufmann, 1999.
2. Ian H. Witten, Eibe Frank. Data Mining: Practical Machine Learning Tools and Techniques(Second Edition) Morgan Kaufmann, 2005.

VII Web References:

1. <http://www.sctie.iitkgp.ernet.in/>
2. <http://www.rkala.in/softcomputingvideos.php>
3. <http://www.sharbani.org/home2/soft-computing-1>
4. http://www.myreaders.info/html/soft_computing.html

VIII. E-TEXT BOOKS:

1. <https://www.books.google.co.in/books?id=bVbj9nhvHd4C>
2. <https://www.books.google.co.in/books?id=GrZHPgAACAAJ&dq=1.+J.S.R.Jang,+C.T.Sun+and+E.Mizutani,+Neuro,+Fuzzy+and+Soft+Computing,+PHI,+2004,Pearson+Education.>