

- I. Course Title** : **Big Data Analytics**
II. Course Code : **CSE 501**
III. Credit Hours : **2+1**

IV. Aim of the course

To understand principles of analyzing and mining big data and to use simple tools to extract useful information from big data sets.

V. Theory

Unit I

Data analysis, data matrix attributes. Data: Algebraic and geometric view, probabilistic view.

Unit II

Basics of data mining and CRISP-DM, organizational and data understanding, purposes, Intents and limitations of data mining, database, data warehouse, data mart and data set, types of data, privacy and security, data preparation, collation and data scrubbing.

Unit III

Data mining models and methods, correlation, association rules, k-means, clustering understanding of concept, preparation and modeling.

Unit IV

Discriminant analysis, linear regression, logistic regression, understanding, preparation and modeling.

Unit V

Decision trees, neural networks, understanding, preparation and modeling.

VI. Practical

Introduction to OpenOffice and RapidMiner in data analytics and mining. Preparing RapidMiner, Importing data, handling missing data, data reduction, handling Inconsistent data, attribute reduction. Performing different analysis using RapidMiner or suitable software.

VII. Learning outcome

Capability to understand the principles behind analysis of big data and apply the same using simple tools.

VIII. Lecture Schedule

S. No.	Topic	No of Lectures
1.	Data analysis, data matrix attributes	2
2.	Algebraic and geometric view, probabilistic view.	3
3.	Basics of data mining and CRISP-DM	1
4.	Organizational and data understanding	2
5.	Intents and limitations of data mining, database, data warehouse data mart and data set	3
6.	Types of data, privacy and security, data preparation, collation and data scrubbing	3
7.	Data mining models and methods, correlation, association rules	5
8.	K-means, clustering understanding of concept, preparation and modeling.	4
9.	Discriminant analysis, linear regression, logistic regression, understanding, preparation and modeling	4
10.	Decision trees, neural networks, understanding, preparation and modeling.	5
	Total	32

IX. List of Practicals

S. No.	Topic	No of Practicals
1	Working of OpenOffice and RapidMiner	4
2	Preparing RapidMiner Dataset	3
3	Handling the inconsistent data, missing data, attribute reduction	4
4	Performing analysis on dataset using RapidMiner	5
Total		16

X. Suggested Reading

- Dr Matthew North *Data Mining for the Masses A Global Text Project Book* ISBN: 0615684378 ISBN-13: 978-0615684376.
- Mohammed J Z, Troy and Wagner M Jr. *Data Mining and Analysis: Fundamental Concepts and Algorithms*. Universidade Federal de Minas Gerais, Brazil. Cambridge University Press ISBN 978-0-521-76633-3 Hardback