# 19CS41P2 - DATA ANALYTICS LABORATORY

|  |  |  |  |
| --- | --- | --- | --- |
| **Course Category:** | Professional Core | **Credits** | 1.5 |
| **Course Type:** | Laboratory | **Practical:** | 0-0-3 |
| **Prerequisite:** | Require the basics of Database Management Systems and Knowledge of Probability and Statistics | **Sessional Evaluation:****Univ. Exam Evaluation:****Total Marks:** | 4060100 |
| **Course Objectives** | * To implement Map Reduce programs for processing big data
* To realize storage of big data using H base
* To analyze big data using linear models
* To analyze big data using machine learning techniques such as Decision tree classification and clustering
 |

|  |  |
| --- | --- |
| **Course Outcomes** | Upon the successful completion of the course, the students will be able to: |
| CO1 | * Process big data using Hadoop framework
* Build and apply linear and logistic regression models
* Perform data analysis with machine learning methods
* Perform graphical data analysis
 |
| **Course Content** | **Hadoop** 1. Install, configure and run Hadoop and HDFS2. Implement word count / frequency programs using MapReduce 3. Implement an MR program that processes a weather dataset R **R**4. Implement Linear and logistic Regression5. Implement Decision tree classification techniques 6. Implement clustering techniques7. Visualize data using any plotting framework 8. Implement an application that stores big data in Hbase / R  |
| **Text Books and References** | **Text Books**:1. Data Science & Big Data Analytics: Discovering, Analyzing, Visualizing and Presenting Data by EMC2 Education Services, Published by John Wiley & Sons, Inc.
2. Tom White “Hadoop: The Definitive Guide” Third Edition, O’reily Media, 2012.

**Reference books:**1. Michael Berthold, David J. Hand, "Intelligent Data Analysis”, Springer, 2007.
2. Jay Liebowitz, “Big Data and Business Analytics” Auerbach Publications, CRC press (2013).
3. Tom Plunkett, Mark Hornick, “Using R to Unlock the Value of Big Data: Big Data Analytics with Oracle R Enterprise and Oracle R Connector for Hadoop”, McGraw-Hill/Osborne Media (2013), Oracle press.
4. SeemaAcharya, SubhasiniChellappan, "Big Data Analytics" Wiley 2015
 |
| **E-Resources** | 1. https://nptel.ac.in/courses
2. https://freevideolectures.com/university/iitm
 |