Prasad V. Potluri Siddhartha Institute of Technology:: Vijayawada.
Department of Computer Science and Engineering
I/II M.Tech. (CSE) (First Semester)

17CSCS1L1 **DATA SCIENCE LAB** **Credits: 2**
Lab: 3 Periods/week **Internal Assessment: 25 Marks**
Semester end examination: 50 Marks

_____

Course Description

Data Science has become increasingly important in nearly every industry sector and academic field, and the discovery and forecasting of insightful patterns from "Big Data" is at the core of analytical intelligence in government, industry, and science.

Course Outcomes:
At the end of the course the student will be able to:

CO1: Gain Core Knowledge of R and Programming Concepts

CO2: Understand mathematics from a numerical point of view, including the application of these concepts root-finding, numerical integration and optimization

CO3: Understand the purpose for random variable and expectations required to understand simulations

CO4: Implement the Monte carlo and Stochastic Modelling

CO5: Work effectively in teams on data science projects using R

Experiment 1: R Environment Setup & R as calculating environment

Experiment 2: R Basic programming, Input and output

Experiment 3: Programming with functions & Sophisticated Data structures

Experiment 4: Better Graphics using Graphics parameters

Experiment 5: Frames and environments & Object –oriented Programming

Experiment 6: Numerical Accuracy and program efficiency

Experiment 7: Probability & Statistics : The law of Total probability

Experiment 8: Simulation: Monte Carlo Integration – Hit and miss method

Experiment 9:   Data Modeling: Linear and Multiple Regression Models

Case Study

Consider the data set of Ozone levels in United States for the year 2014 and do the following analysis

1. Formulate your questions
2. Read in your data
3. Check the packaging
4. Look at the top and the bottom of your data
5. Check your "n" s
6. Validate with at least one external data source
7. Make a plot
8. Follow up

Text Books:

1. Introduction to Scientific Programming and Simulation Using R, Owen Jones, Robert Maillardet and Andrew Robinson, Second Edition, CRC Press, 2014
2. The Art of Data Science: A Guide for Anyone Who Works with Data, Roger D. Peng, Elizabeth Matsui, LeanPub, 2015.
3. Data Science for Business: What You Need to Know about Data Mining and Data-analytic Thinking, Foster Provost and Tom Fawcett. 2013
4. Elements of Statistical Learning, Trevor Hastie, Robert Tibshirani and Jerome Friedman, Springer, 2009.

Reference Books:

1. Mining of Massive Datasets,Jure Leskovek, Anand Rajaraman and Jeffrey Ullman. , Cambridge University Press. 2014.
2. Machine Learning: A Probabilistic Perspective.Kevin P. Murphy, MIT Press, 2013.
3. Avrim Blum, John Hopcroft and Ravindran Kannan. Foundations of Data Science.
4. Data Mining and Analysis: Fundamental Concepts and Algorithms, Mohammed J. Zaki and Wagner Miera Jr., Cambridge University Press. 2014.
5. R Programming for Data Science, Roger D. Peng, LeanPub, 2015.
6. Python for Data Science for Dummies, Luca Massaron and John Paul Mueller, John Wiley and Sons, 2015.