CAP787:DATA SCIENCE TOOLBOX

L:3 T:0 P:0 Credits:3

Course Outcomes: Through this course students should be able to

CO1 :: understand use of data and data science in real time data analysis

CO2 :: practice R and Python functions that can be used in data Analysis

CO3 :: analyze the data sets using R and Python

CO4 :: develop data model using real time data sets

Unit I

Introduction to data and Data Science : • What is Data Science? • What is Data? • Getting Help • The Data Science Process • Types of Data Science Questions • Experimental Design • Download data from different sources and explore them.

Unit II

Working with R : Introduction to R programming, Data Structure of R (Vector, matrices, list, factors, and Data Frames), working with Logical statements, working with if-else statements, Working with Functions, Working with Loops

Unit III

Data Analysis Using R : Introduction to Data preprocessing (Data Wrangling, Ordering of data, Filtering of Data), Working with basic statistical functions., Visual presentation using Histograms, Box Plots, Bar Plots., Working with GG Plots, QQ plots and similar ones.

Unit IV

Working With PYTHON : Introduction to Python, Python Data Structure, Working with NumPy, SciPy, Pandas, Understanding Dataset, Handling Missing values, Data Formatting, Data Normalization, Working with sets, Understanding Binning and indicator variables

Unit V

Data Analysis using PYTHON : Pyhton Package for Data Science, Exploratory Data Analysis, Descriptive statistics, Basic Grouping, Implementing ANOVA, Understanding Correlation Analysis, Model development: Simple and multiple linear regression, Model Evaluation using visualization, Polynomial regression, Ridge regression, Grid search, and Model Refinement

Unit VI

Version Control and Github : Version Control, Github and Git, Linking Github and R Studio, Linking Github and python

Text Books:

1. PYTHON FOR DATA ANALYSIS by WES MCKENNY, O'REILLY

References:

1. R FOR DATA SCIENCE: IMPORT, TIDY, TRANSFORM, VISUALIZE, AND MODEL DATA by HADLEY WICKHAM, O'REILLY

2. GITHUB ESSENTIALS by ACHILLEAS PIPINELLIS, PACKT PUBLISHING