CAP788:DATA SCIENCE TOOLBOX-LABORATORY

Course Outcomes: Through this course students should be able to

CO1 :: understand data and data science

CO2 :: use R and R studio to data science

CO3 :: apply Python for data science

CO4 :: manage and control version using Git and Github

List of Practicals / Experiments:

Introduction to Data Science Toolbox

- •Types of Data Science Questions•Experimental Design•Download data from different sources and explore them.

R and R studio

- •Installing R•Installing R Studio•R Studio Tour•Loading data and basic analysis of data •R Packages•Projects in R•R Markdown

Python Basics

- •Python Data structure•Programming Fundamental•Working with data in python•Working with numpy Array

Basic Data Analysis Using Python

- •Introduction: Understanding the Dataset, Python package for data science, Importing and Exporting Data in Python, Basic Insights from Datasets.•Data Wrangling: Identify and Handle Missing Values, Data Formatting, Data Normalization, Sets, , Binning, Indicator variables,

Advance Data analysis Using Python

- •Exploratory Data Analysis: Descriptive Statistics, Basic of Grouping, ANOVA, Correlation, Correlation 2.•Model Development: Simple and Multiple Linear Regression, Model Evaluation using Visualization, Polynomial Regression and Pipelines, R-squared and MSE for In-Sample Evaluation, Prediction and Decision Making, • Working with Data in Python: Model Evaluation , Over Fitting, Under fitting and Model Selection , Ridge Regression, Grid Search , Model Refinement ,

Version Control Using Git and Github

- •Version Control•Github and Git•Linking Github and R Studio•Linking Github and python•Projects under Version Control

Text Books:

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

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

References:

1. GITHUB ESSENTIALS by ACHILLEAS PIPINELLIS, PACKT PUBLISHING