II YEAR SEM-1 B.Tech CSE	SEMESTER-2	L	T	P	C
Code:CS2105	Data Analytics	1	1	0	2

Objective:

The course is aimed at introducing to data analytics providing some basic data-science. To discover patterns and laws in complex datasets will be introduced to students together with instruments to analyze, characterize, validate, parameterize and model complex data. Practical issues on business data analysis and statistics will be covered with specific case studies.

Outcomes Students will become able to analyze main statistical features of complex datasets. On successful completion of the course, a student should have a good understanding on:

1) How to analyze, characterize empirically complex data;
2) How to compute relevant statistical quantities and quantify their confidence intervals;
3) How to build sensible models and how to parameterize and validate these models;
4) How to quantify inter-dependency/causality structure between different variables;
5) How to use the outcome of data-analytics to develop better tools for forecasting.

Content

UNIT-1

Data structures in R, Arrays & Matrices, Making Tables, Lists, Data frames, Conversion of Numeric Data frames into Matrices. Reading Excel Files,and text files, Plotting function, Multiple plots, Scatter plot matrices.

UNIT-2

Essential practical familiarization with complex and big data. Typical challenges with real business data. Basics on data acquisition, manipulation, cleaning, filtering, representation and plotting.

UNIT-3
Marginal probability, joint probability and conditional probability.Empirical estimation of probability distributions.Measures of dependency.Cause and effect.transfer entropy.Spurious correlations and regularization.Forecasting and regressions. Calibration, validation hypothesis testing.

Text books:

1. Dunlop, Dorothy D., and Ajit C. Tamhane. Statistics and data analysis: from elementary to intermediate. Prentice Hall, 2000.

2. Robert Knell. “Introductory R: A Beginner's Guide to Data Visualization, Statistical Analysis and Programming in R”, 2013

Reference Books

1.Ohlhorst, Frank J. Big data analytics: turning big data into big money. John Wiley & Sons, 2012.

2. W.N. Venables, D.M Smith, "An introduction to R", Network Theory Ltd.

3. Nina Zumel, John Mount, "Practical Data Science with R", Manning Publications, 2014.

4. Mark Gardener, "Beginning R - The Statistical Programming Language", John Wiley & Sons, Inc., 2012 (Rs 509)