MCA 402A: Essentials of Data Science

UNIT I

Introduction: What is Data Science? - Big Data and Data Science, Statistical Inference - Populations and samples - Statistical modeling, probability distributions, fitting a model - Intro to R Language.

UNIT II

Exploratory Data Analysis and the Data Science Process - Basic tools (plots, graphs and summary statistics) of EDA - Philosophy of EDA - The Data Science Process - Case Study: RealDirect (online real estate firm)

UNIT III

Feature Generation and Feature Selection (Extracting Meaning From Data) - Motivating application: user (customer) retention - Feature Generation (brainstorming, role of domain expertise, and place for imagination) - Feature Selection algorithms – Filters; Wrappers; Decision Trees; Random Forests.

UNIT IV

Data Visualization: Basic principles, ideas and tools for data visualization 3 - Examples of inspiring (industry) projects - Exercise: create your own visualization of a complex dataset.

UNIT V

Data Science and Ethical Issues - Discussions on privacy, security, ethics - A look back at Data Science - Next-generation data scientists