Data Science Syllabus

Module1: Introduction to Data Science Introduction: What is Data Science? Big Data and Data Science hype – and getting past the hype, Why now? – Datafication, Current landscape of perspectives, Skill sets. Needed Statistical Inference: Populations and samples, Statistical modelling, probability distributions, fitting a model.

Module2: 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: Real Direct (online real estate firm). Three Basic Machine Learning Algorithms: Linear Regression, k-Nearest Neighbours (k- NN), k-means.

Module3: 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. Recommendation Systems: Building a User-Facing Data Product, Algorithmic ingredients of a Recommendation Engine, Dimensionality Reduction, Singular Value Decomposition, Principal Component Analysis, Exercise: build your own recommendation system**.**

Text Book: Doing Data Science, **Cathy O'Neil** and **Rachel Schutt**, O'Reilly Media, Inc O'Reilly Media, Inc, 2013

Text Book URL: **Doing Data Science**

Chapter1: **Introduction: What Is Data Science?**

Authors

Rachel Schutt is the Senior VP of Data Science at News Corp. She is also an adjunct professor of **statistics** at Columbia University and a founding member of the CU's Education Committee for the Institute for Data Sciences and Engineering.

 Cathy O'Neil is a senior data scientist at Johnson Research. She holds a Ph.D. in math from Harvard and was previously a postdoc in the math department at MIT and a professor at Barnard College.