DATA 5000: Introduction to Data Science

Carleton University, Ottawa, Canada

Course outline: The course covers topics relevant to data science: working with data, exploratory data analysis, data mining, machine learning. The concepts are illustrated using the R language. Students also receive introduction to IBM Cognos Workspace and IBM SPSS Modeler. Term projects.

Lectures: Tuesdays from 11:35 to 14:25 in 2202 Canal Building

Instructor: Boyan Bejanov

Contact: boyanbejanov@cemail.carleton.ca

Office hours: Tuesdays 14:30 to 15:30 in HP5331 or by appointment.

Marking: Students will work on projects in small teams of 2 or 3. Teams where all members are from the same departments will not be allowed.

- 10% Project proposal, due 27 January 2015
- 10% Presentation outline, due 17 March 2015
- 30% Presentation in class on 7 April 2015
- 50% Project paper, due TBA

Resources: The following books are suggested but not required.

- *Doing Data Science: Straight Talk From the Frontline*, by Cathy O'Neil and Rachel Schutt, O'Reilly Media, 2013.
- *Data Science for Business: what you need to know about data mining and data-analytic thinking*, by Foster Provost and Tom Fawcett, O'Reilly Media, 2013.

The following books are good references for data mining and machine learning algorithms

- *An Introduction to Statistical Learning: with Applications in R*, by Gareth James, Daniela Witten, Trevor Hastie and Robert Tibshirani, Springer, 2013.

The following are good references for R (just to name a few)

- *Cookbook for R* by Winston Chang.
- *The R Inferno* by Patrick Burns.
- *Quick-R*.