CMPT 353 - D100 Computational Data Science

Instructor(s): Gregory Baker

Calendar Objective/Description:
Computational Data Science

Instructor’s Objectives:
This course will be an introduction to the tools and techniques in data science. We will explore common challenges and solutions used in analysis of data.

Online offering notes: you will need a computer with a webcam and reliable Internet access. The computer should be powerful enough to run a virtual machine: at least 8 GB memory, 20 GB disk, and a reasonably decent processor. There will be 3-4 in-class activities during the semester which must be completed during the lecture time. Otherwise, lectures will be posted as a “watch party” where we can watch together (and ask questions in a forum), but they can also be viewed later.

Prerequisites:
see go.sfu.ca

Topics:
- Basics of data science: concepts, goals, motivation, expectations.
- Introduction to selected data processing tools: Python with numpy and pandas.
- Working with data. Cleaning data; extract, transform, load tasks; applying concepts from statistics.
- Machine learning basics with existing implementations (such as scikit-learn).
- Data analysis strategies: selecting techniques from statistics and machine learning.
- Big data tools.
- Data visualization and summarizing results.

Grading:
Will include weekly exercises, quizzes (in lecture time), and a project. Details will be discussed in the first week of class.

Students must attain an overall passing grade on the weighted average of exams in the course in order to obtain a clear pass (C- or better).

Academic Honesty Statement:
Academic honesty plays a key role in our efforts to maintain a high standard of academic excellence and integrity. Students are advised that ALL acts of intellectual dishonesty will be handled in accordance with the SFU Academic Honesty and Student Conduct Policies (http://www.sfu.ca/policies/gazette/student.html).