CMPT 353 - D100 Computational Data Science

Instructor(s): Gregory Baker

Calendar Objective/Description:
Basic concepts and programming tools for handling and processing data. Includes data acquisition, cleaning data sources, application of machine learning techniques and data analysis techniques, large-scale computation on a computing cluster.

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.

Prerequisites:
CMPT 225 and (STAT 101, STAT 270, BUED 232, ENSC 280, or MSE 210).

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:
Details to be announced in first week of class. Will include weekly exercises, a project, quizzes, and exams.
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 ).